

Genetic Factors and Genes Underpinning Drought Response in Wheat

By

Hollie Webster BRurSc(Hons)

The thesis is presented in fulfilment of the requirements for the degree of Doctor
of Philosophy

School of Veterinary and Life Sciences,
Murdoch University, Western Australia

March 2013

This thesis was supported by Murdoch University (Western Australia), via infrastructure at the State Agricultural Biotechnology Centre, and funding from an Australian Post Graduate Award. A co-contribution was provided by a Grains Industry Research and Development Corporation grant (GRS10028).

Declaration

Except where otherwise indicated, all work in this thesis is based on work carried out by me at the State Agricultural Biotechnology Centre (SABC) and the Australia China Centre for Wheat Improvement (ACCWI) Murdoch University, Australia. I declare the content of this thesis is my own account of my research and has not been previously submitted for a degree at any tertiary education centre. To the best of my knowledge, all work performed by others, published or unpublished, has been duly acknowledged.

Hollie Webster

28/03/2014

Abstract

Pollen fertility is one of the main factors limiting yield in crops such as rice and wheat, both highly inbreeding species (Wang *et al.*, 2003). Water deficit during reproductive stage growth can result in pollen sterility due to impaired carbohydrate supply (sucrose/glucose) to pollen and supporting tissues such as the tapetum. In cereal species, and monocotyledons in general, viable pollen development is accompanied by starch accumulation in amounts that are sufficient to support pollen germination and pollen tube growth (Franchi *et al.*, 1996). Cereal pollen grains rendered sterile by the down regulation of the sugar transport gene *IVR1* lack starch (Sheoran and Saini 1996) and have failed or impaired intine formation (Lalonde *et al.*, 1997). *IVR1* down regulation results in incomplete or total absence of sucrose cleavage to the hexose sugars glucose and fructose (the final energy substrates used in plant metabolism to support pollen development), resulting an accumulation of sucrose despite the high energy demand of the developing tissues (Dorion *et al.*, 1996).

The mechanisms underlying the sensitivity of pollen and the tapetum to abiotic stress (water deficit) provide a basis for developing molecular approaches aimed at increasing stress tolerance (Parish and Li 2010). This thesis has characterized the cell wall invertase gene family in detail in order to provide DNA sequence signatures that allow the expression of specific genes to be followed. The study demonstrated *IVR1.1-3B* expression was confined to leaves while *IVR1-4A* and *IVR1-5B* represented genes expressed during early head development. Two double haploid lines identified from a population of 225 lines derived from a cross between the varieties Westonia and Kauz showed significant differences in response to water deficit stress induced grain set reduction and expression of *IVR1* isoforms *IVR1-4A* and *IVR1-5B*. The differences in expression were investigated in more detail using a large-scale RNASeq study, which indicated there was a significant expression response to water deficit in a suite of genes involved in carbohydrate metabolism. In addition *IVR1* expression analysis was validated using SQ-PCR, which highlighted significant differences in response to

water deficit stress. Rice was used as a benchmark for selecting genes in this study.

The external phenotype, penultimate leaf internode auricle distance 5cm (Zadoks growth stages Z39/40), was validated as an indicator of pollen meiosis completion. A significant developmentally related drought escape QTL was identified on chromosomes 5B and 5D. High resolution mapping of the group 5 chromosomes (using a 90K SNP chip) enabled a high density of markers to be located across the chromosome 5B QTL in particular, and identified water deficit responsive UDP-glucose 6-dehydrogenase (from the KEGG carbohydrate metabolism pathway) as a gene of interest. A global analysis using RNASeq data identified a more extensive suite of anther-specific genes that were water deficit responsive. In turn this suit of genes provided the basis for defining a set of molecular markers to screen for variation in drought responsiveness in wheat varieties.

Although this work has focussed on a developmental stage specific water deficit response, it is evident the thesis contributes to the wider body of information becoming available for wheat, much of which now indicates it should be feasible to define a haplotype for wheat using the allelic variation in genes associated with drought tolerance in different environments. Results from this study have shown that reproductive stage water deficit tolerance is a complex quantitative trait and that it may be more useful to quantify a risk factor value using a suite of markers, rather than a plus/minus analysis of single markers. We propose the network of carbohydrate metabolism genes defined in this thesis would make significant contributions to risk factor analyses in early selection or backcrossing for sensitivity in particular environments.

Acknowledgements

Work presented in this thesis was made possible by collaborative and supportive contributions made by numerous individuals. First and foremost I wish to express my sincere gratitude to my primary supervisor Prof. Rudi Appels, for his research wisdom, genomics expertise, and tireless support for the duration of the PhD project. Rudi's efforts in this regard facilitated a truly exceptional learning experience and enabled me to benefit to the fullest extent from the rigorous discovery process that was this PhD. I will feel forever grateful for the numerous opportunities Rudi has extended to me, none more so than his enabling me to feel a very welcome and valued member of the global wheat research community, of which I'm immeasurably proud to be a part. To my other supervisors, Prof. Bernard Dell and Dr. John Fosu-Nyarko, I owe enormous thanks for many aspects of their respective expert contributions in molecular biology and plant physiology but most of all I must acknowledge how being challenged by them to improve my interdisciplinary thinking has facilitated my growth as a research scientist.

Importantly I must acknowledge the collective support of key members of the Stage Agricultural Biotechnology Centre (SABC), the facility in which a large proportion of the PhD project was carried out. Individuals to thank in this regard included Prof. Michael Jones, David Berryman, Francis Brigg, and Bee Lay Addis. Having these world class facilities and resources made available to me for my research undoubtedly underpinned the success of the early part of the PhD project.

Two of my colleagues, namely Associate Prof. Dean Diepeveen and Gabriel Keeble-Gagnere, in the research group to which I belong, the Australia China Centre for Wheat Improvement (ACCWI), have made significant informatics contributions to all chapters of the thesis. In addition to contributing in context with their core disciplines these two individuals have always offered friendly and supportive advice on how to manage some of the common problems PhD students encounter. Additional members of the ACCWI team to thank include Dr.

Dora Lee and Dr. Jingzhan Zhang, for their valuable advice on how to trouble shoot problems with experiments not working properly in the lab. Two colleagues in the Victorian Department of Environment and Primary Industries, Dr. Matthew Hayden and Dr. Josquin Tibbits, were also instrumental to the success of RNASeq studies in this thesis.

I would also like to thank two technical experts from Murdoch University's Department of Biological Sciences for the invaluable contributions, namely Gordon Thompson and Jose Minetto. Gordon for his supervision of me during the of cytology component of the PhD project, and Jose for practical knowledge and advice on how to manage large-scale glasshouse trails. Much valued student administrative support has also been conferred to me during the course of the PhD by Karen Olkowski and Carolyn Jones (also of the Department of Biological Sciences).

Lastly I wish to pay a very special thank you to my mother, Julie-Anne Webster, for her unwavering support throughout my PhD candidature. Without her encouragement to pursue my passion to make a difference in the world through science, this thesis would undoubtedly not have eventuated.

List of abbreviations

Abbreviation	Description
ABA	Abscisic acid
ABARE	Australian Bureau of Agricultural and Resource Economics
ABS	Australian Bureau of Statistics
AC	Anthesis commenced
ACCWI	Australia China Centre for Wheat Improvement
AD	Auricle distance
AFV	Awns first visible
AMS	Aborted microspore
ANOVA	Analysis of variance
APW	Australian prime hard
AT	Ambient temperature
AUD	Australian dollar
BAC	Bacterial artificial chromosome
BBCH	Biologische Bundesanstalt Bundessortment Scale
BP	Bicellular pollen
BS	Boot swollen
C-INV	Cytoplasmic invertase
CCG	Centre for Comparative Genomics
CDS	Coding deoxyribonucleic acid sequence
CIMMYT	International Maize and Wheat Improvement Centre
cM	Centimorgan
CNV	Copy number variation
CPM	Counts per million
CW-INV	Cell wall invertase
DAFWA	Department of Agriculture and Food Western Australia
DAPI	4',6-diamidino-2-phenylindole
DAS	Days after sowing
DEPI	Department of Environment and Primary Industry (Victoria)
DF	Degrees of freedom
DH	Double haploid
DNA	Deoxyribonucleic acid
EC	Enzyme commission number
ER	Endoplasmic reticulum
EST	Expressed sequence tag
ET	Evapotranspiration
FAO	Food and Agriculture Organization of the United Nations
FDR	False discovery rate
FEH	Fructan exohydrolase
FHB	Fusarium head blight
fl-cDNA	Full length cDNA
FPC	Fingerprint contig

Fru	Fructose
GO	Gene ontology
GSS	Genome survey sequence
GWAS	Genome wide association study
HCS	High confidence score
HICF	High information content fingerprinting
H XK	Hexokinase
ICIM	Inclusive composite interval mapping
ID	Identity description
IPCC	Intergovernmental Panel on Climate Change
IRGSP	International Rice Genome Sequencing Project
IWGSC	International Wheat Genome Sequencing Consortium
JCVI	J. Craig Ventur Institute
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LAI	Leaf area index
LB	Lysogeny broth
LCS	Low confidence score
LEA	Late embryogenesis abundant
LOC	Locus
LOD	Logarithm of edits
LTC	Linear topological contig
MAB	Marker assisted backcross
MAGIC	Multi-parent advanced generation inter-cross
MAS	Marker assisted selection
Mbp	Mega base pairs
MDS	Multi dimensional scaling
MIPS	Munich Information Centre for Protein Sequence
MMC	Microspore mother cell
MP	Mature pollen
MS	Mean sums of squares
NA	Not applicable
NB	nota bene/take note
NCBI	National Centre for Biotechnology Information
NSW	New South Wales
PCA	Principal component analysis
PCD	Programmed cell death
PCP	Principal component biplot
PCR	Polymerase chain reaction
PDB	Protein data bank
PPP	Pentose phosphate pathway
PH	Plant height
Ppd	Photoperiod
PVE	Phenotypic variation explained
PWP	Permanent wilting point
QC	Quality control

QLD	Queensland
QTL	Quantitative trait loci
REML	Restricted maximum likelihood
RH	Relative humidity
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RSWD	Reproductive stage water deficit stress
RT	Reverse transcription
SABC	State Agricultural Biotechnology Centre
SCW	Seedling cuticle waxiness
SNP	Single nucleotide polymorphism
SQ-PCR	Semi quantitative polymerase chain reaction
SS	Sums of squares
SSR	Simple sequence repeat
Suc	Sucrose
SucTP	Sucrose transporter
SVWC	Soil volumetric water content
TD-PCR	Touch down polymerase chain reaction
TDR	Tapetal degradation retardation
THT	The hordeum toolbox
TMM	Trimmed means of m
UGPase	Glucose-1-phosphate
UN	United Nations
USA	United States of America
USD	United States of America dollar
UV	Ultra violet
V-INV	Vacuolar invertase
Vrn	Vernalization
WxK	Westonia x Kauz
Z	Zadoks growth stage scale

Table of Contents

DECLARATION	III
ABSTRACT	IV
ACKNOWLEDGEMENTS	VI
LIST OF ABBREVIATIONS	VIII
CHAPTER 1: INTRODUCTION	14
1.1 TOLERANCE OF REPRODUCTIVE GROWTH WATER DEFICIT IN WHEAT	14
1.2 LIMITATIONS ON CURRENT UNDERSTANDING OF GENETIC FACTORS DETERMINING WATER DEFICIT INDUCED POLLEN STERILITY	16
CHAPTER 2: LITERATURE REVIEW	19
2.1 WHEAT	19
2.1.1 <i>Economic and social importance</i>	19
2.1.2 <i>Taxonomy and genome structure</i>	20
2.1.3 <i>Definitions of drought</i>	23
2.1.4 <i>The significance of water deficit during reproductive growth in wheat</i>	24
2.1.5 <i>Regulation of sugar translocation to developing fruit and seed</i>	26
2.1.6 <i>Sugar signaling</i>	29
2.2 MORPHOLOGY OF MALE REPRODUCTIVE STRUCTURES IN WHEAT	31
2.2.1 <i>Pollen development</i>	32
2.2.2 <i>Tapetal development</i>	35
2.2.3 <i>Gene network regulation of during pollen development</i>	36
2.3 WATER DEFICIT	40
2.3.1 <i>The effect on pollen development</i>	40
2.3.2 <i>The effect of sugar accumulation in developing anthers</i>	42
2.3.3 <i>How water deficit is sensed by the reproductive tissues</i>	43
2.4 TRANSLOCATION OF COMPLEX CARBOHYDRATES TO DEVELOPING POLLEN.....	44
2.4.1 <i>The role of cell wall invertase in carbohydrate delivery to pollen</i>	44
2.4.2 <i>Cell wall invertases</i>	47
2.4.3 <i>Cell wall invertase three dimensional structure</i>	48
2.5 FACTORS REGULATING THE EXPRESSION OF CELL WALL INVERTASES.....	49
2.5.1 <i>The role of abscisic acid in reproductive tissues sensing water deficit</i>	50
2.6 COMPARATIVE GENOMICS	53
2.7 BREEDING FOR REPRODUCTIVE STAGE DROUGHT TOLERANCE	54
2.7.1 <i>Plant genetic improvement</i>	55
2.7.2 <i>The phenotype and environment</i>	58
2.7.3 <i>Comparative phenomics</i>	60
2.7.4 <i>Changing agendas to meet new challenges and conclusions</i>	62
CHAPTER 3: GENERAL MATERIALS AND METHODS	63
3.1 PLANT MATERIALS AND GROWTH IN THE GLASSHOUSE	63
3.1.1 <i>Selection of population structure and variety cross</i>	64
3.2 CHARACTERIZING THE WxK DH POPULATION	65
3.2.1 <i>Phenotyping – phenology and morphology</i>	65
3.2.2 <i>Phenotyping - cytology</i>	66
3.2.3 <i>Pollen sterility</i>	69
3.2.4 <i>Genotyping and selection of WxK lines for reproductive stage water deficit stress testing</i>	70
3.3 EXPERIMENTAL DESIGN	72
3.3.1 <i>Water deficit treatment</i>	73
3.3.2 <i>Soil volumetric water content (SVMC)</i>	74
3.3.3 <i>Plant water status Ψ_w</i>	75

3.3.4	<i>Relative humidity and ambient temperature</i>	77
3.3.5	<i>Sampling schedule and tissue collection</i>	77
3.4	MOLECULAR PROTOCOLS	78
3.4.1	<i>Extraction of genomic DNA from leaf and developing head tissue</i>	78
3.4.2	<i>Extraction of RNA from leaf and developing head tissue</i>	79
3.4.3	<i>Polymerase chain reaction (PCR)</i>	79
3.4.4	<i>Reverse transcription (RT)</i> ,.....	80
3.4.5	<i>Primer design for RT, PCR, and DNA sequencing</i>	80
3.4.6	<i>Agarose gel electrophoresis</i>	81
3.4.7	<i>Cloning of PCR products</i>	81
3.4.8	<i>Restriction endonuclease digestion</i>	82
3.4.9	<i>Transformation of competent E.coli cells</i>	82
3.4.10	<i>Heat shock transformation and bacterial growth</i>	83
3.4.11	<i>Selection for transformants</i>	83
3.4.12	<i>Analysis of transformants by PCR</i>	84
3.4.13	<i>Analysis of transformants by plasmid DNA mini-prep restriction digestion</i>	84
3.4.14	<i>DNA purification from agarose gels</i>	84
3.4.15	<i>Plasmid DNA purification</i>	85
3.4.16	<i>DNA sequencing</i>	85
3.4.17	<i>RNaseq</i>	86
CHAPTER 4: GENOME-LEVEL IDENTIFICATION OF CELL WALL INVERTASE GENES IN WHEAT FOR THE STUDY OF DROUGHT TOLERANCE		87
4.1	ABSTRACT	88
4.2	INTRODUCTION	89
4.3	MATERIALS AND METHODS.....	92
4.3.1	<i>Wheat BAC sequencing</i>	92
4.3.2	<i>Identification of IVR1 isoforms and sequence annotation</i>	92
4.3.3	<i>Plant material and DNA extraction</i>	93
4.3.4	<i>IVR1 sequencing and analysis</i>	93
4.4	RESULTS	95
4.4.1	<i>The genome level structure of wheat IVR1 isoforms on chromosome 3B</i>	95
4.4.2	<i>The IVR1 gene family in wheat</i>	97
4.4.3	<i>Comparative genomics of cell wall invertases in wheat and rice</i>	103
4.5	DISCUSSION.....	104
CHAPTER 5: INVESTIGATING A WESTONIA X KAUZ MAPPING POPULATION FOR DOUBLE HAPLOID LINES SUITABLE FOR STUDYING THE EFFECTS OF REPRODUCTIVE STAGE WATER DEFICIT		108
5.1	ABSTRACT	108
5.2	INTRODUCTION	109
5.3	MATERIALS AND METHODS.....	115
5.3.1	<i>Marker analysis and WxK DH genetic map construction</i>	115
5.3.2	<i>ICIM QTL analysis</i>	115
5.4	RESULTS	116
5.4.1	<i>A genetic resource for defining doubled haploid lines for detailed analysis of response to reproductive stage water deficit stress</i>	116
5.4.2	<i>Selection of two double haploid lines from the Westonia x Kauz double haploid population for water deficit stress studies</i>	126
5.4.3	<i>Investigation of the chromosomal location of the drought escape QtAD5 QTL and cell wall invertase (IVR1) genes</i>	134
5.5	DISCUSSION.....	146
CHAPTER 6: GENOTYPE SPECIFIC GENE EXPRESSION IN THE DEVELOPING SPIKE OF DOUBLE HAPLOID WHEAT LINES UNDER WATER DEFICIT STRESS		152

6.1	ABSTRACT	152
6.2	INTRODUCTION	153
6.3	MATERIALS AND METHODS.....	157
6.3.1	<i>Samples submitted for RNASeq transcriptome analysis</i>	157
6.3.2	<i>Raw transcriptome data</i>	157
6.3.3	<i>Quality control of short reads</i>	158
6.3.4	<i>Preparation of alignment references for transcript assembly construction.....</i>	158
6.3.5	<i>Alignment of short reads to genome reference and generating counts</i>	158
6.3.6	<i>Identifying outliers.....</i>	159
6.3.7	<i>PCA analysis.....</i>	159
6.3.8	<i>Total expressed transcripts and assigning of GO terms.....</i>	159
6.3.9	<i>Comparison between expressed transcripts in 2011 and 2012 experiments</i>	160
6.3.10	<i>Identification of stable expression transcripts.....</i>	160
6.3.11	<i>Global data set analysis for statistical calculation of differential transcript expression.....</i>	160
6.3.12	<i>Calculation of differential transcript expression (pairwise comparisons between single transcripts).....</i>	161
6.3.13	<i>Investigation of KEGG starch sucrose metabolism pathway</i>	163
6.4	RESULTS	163
6.4.1	<i>Illumina sequencing and assemble construction using the MIPS wheat GSS reference.....</i>	163
6.4.2	<i>Analysis of stable expression genes for use as house keeping genes.....</i>	170
6.4.3	<i>Global analysis for transcriptome wide changes in expression.....</i>	175
6.4.4	<i>IVR1 transcript expression in developing spikes in response to plus and minus water stress during reproductive stage growth.....</i>	182
6.4.5	<i>Expression profiling of QtAD5-5BL associated transcripts in response to plus and minus reproductive stage water deficit stress</i>	197
6.4.6	<i>Gene expression analysis of the KEGG starch-sucrose metabolism network.</i>	205
6.5	DISCUSSION.....	211
6.5.1	<i>Global analysis of transcriptome data</i>	211
6.5.2	<i>Cell wall invertase IVR1 and its relative contribution to starch accumulation in pollen and the pollen fertility trait per se</i>	216
6.5.3	<i>UDP-glucose 6-dehydrogenase as a key enzyme in spike development.....</i>	221
6.5.4	<i>Key drivers of carbohydrate metabolism in developing anthers in relation to spike fertility under water deficit conditions.....</i>	223
CHAPTER 7: GENERAL DISCUSSION AND CONCLUSIONS.....		233
7.1	CANDIDATE GENE IDENTIFICATION AND SELECTION	233
7.2	GENE ANNOTATION, MAPPING/QTL ANALYSIS, AND COMPARATIVE GENOMICS	234
7.3	GENE EXPRESSION	236
7.4	SUMMARY OF THESIS FINDINGS AND INTEGRATION INTO MOLECULAR BREEDING.....	238
CHAPTER 8:.....		241
8.1	APPENDIX I	241
8.2	APPENDIX II.....	244
8.3	APPENDIX III	248
REFERENCES.....		250

Chapter 1: Introduction

1.1 Tolerance of reproductive growth water deficit in wheat

Wheat provides approximately one-fifth of the total calorific input of the world's population and is a crop of global significance particularly with regards to short-term food availability and long-term food security (Nellemann 2009). It is widely accepted that global demand for wheat is increasing at a faster rate than genetic gains being achieved (Fischer 2007). Hence, the medium term challenge is to identify and successfully utilize wheat breeding methodologies which will generate new varieties capable of meeting the dietary requirements of the predicted world population of 9.3 billion by 2050 (UN 2011). To achieve the necessary yield productivity gains by 2050, yield potential must increase at a rate of 2% per annum for the duration until that time (Gill *et al.*, 2004).

Globally, water deficit stress is the single most yield limiting factor in crop production and presents the greatest challenge faced by plant breeders when trying to increase yield potential under abiotic stress conditions (Slafer and Garaus 2007). Among the crops of global human dietary importance, wheat is particularly susceptible to drought induced yield penalties due to the fact it is predominantly grown in rainfed marginal areas throughout the world (Collins *et al.*, 2008). The urgency for developing new high yielding drought tolerant wheat varieties has been further intensified by unprecedented yield losses over the last decade which are attributed to increasing intensity and frequency of abiotic stress conditions under climate change (IPCC 2007).

Over the last fifty years significant gains in yield potential of wheat have been achieved through conventional breeding methodologies based on selection for reduced plant height and vernalisation requirement, which has seen the global average per annum yield increase from 1 ton/ha to 2.5 ton/ha (William *et al.*, 2007). However, conventional breeding methodologies have had little success in introgressing drought tolerance traits primarily due to their quantitative nature,

low heritability and variation in expression under different environments (Fleury *et al.*, 2010)

It is widely recognised that molecular genetic approaches which enable the dissection of complex quantitative traits, such as drought tolerance, allow for the individual genetic determinants of the trait to be identified and thus potentially enabling them to be utilized for marker assisted selection (MAS) (Morgante and Salamini 2003). Despite the potential to utilize molecular breeding to increase the rate of genetic gain in wheat, it is important to note that to date the application of the technology in drought breeding programs has been minimal (Tuberosa and Salvi 2006) primarily due to the complexity of identifying and measuring physiological determinants of yield under drought conditions (Cattivelli *et al.*, 2008; Passioura 2007).

It has long been understood that grain number and size are the two primary determinants of genetic yield potential (Peltonen-Sainio *et al.*, 2007; Reynolds *et al.*, 2009(A)). In the last decade considerable focus has been given to improving yield potential under drought through increasing partitioning of assimilates to the developing grain during grain fill, and thus contributing to stabilized grain size (Xue *et al.*, 2008). However, Dolferus *et al.*, (2011) reported it is the reduction in grain number (as a factor of water deficit during reproductive development), rather than grain size that has the most detrimental effect on cereal yields. These claims are consistent with findings that in wheat reproductive stage growth is the second most water deficit sensitive developmental stage after seedling establishment (Lalonde *et al.*, 1997).

There is now clear evidence that the well-being of reproductive structures, which provide structural and nutritive support to developing pollen during meiosis, is a critical yield determinant (Dolferus *et al.*, 2011; Ji *et al.*, 2010). Continued partitioning of assimilates to reproductive structures, under abiotic stress, is generally matched by adequate pollen fertility (Dolferus *et al.*, 2011; Reynolds *et al.*, 2005).

1.2 Limitations on current understanding of genetic factors determining water deficit induced pollen sterility

The importance of the *IVR1* gene in maintaining the carbohydrate supply to developing pollen and optimising grain set has been proposed (Dorion *et al.*, 1996; Ji *et al.*, 2010; Koonjul *et al.*, 2005; Saini and Lalonde 1998) (Figure 1.1). The fact that the levels of *IVR1* expression have also been found to show varietal variation, has established its credentials as a key target for reproductive stage drought tolerance (Ji *et al.*, 2010). More detailed investigation of a target such as the *IVR1* gene(s) requires the study of genetically structured populations (usually from a biparental cross) in order to associate genetic factors with aspects of the drought phenotype (Maccaferri *et al.*, 2011).

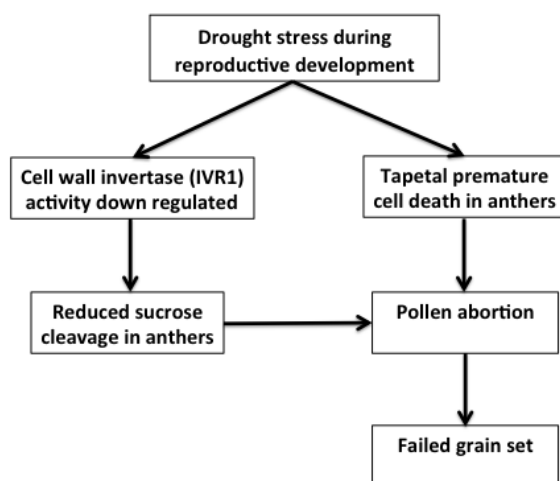


Figure 1.1 Proposed model for the effect of reproductive stage water deficit stress (RSWD) on pollen fertility and grain set (Dolferus *et al.*, 2013).

There are limitations on drawing definitive conclusion from previous *IVR1* gene expression studies due to:

- lack of tissue specific gene expression data, the variation in methods utilized to impose RSWD stress, and
- difficulty in discriminating between different versions of published *IVR1* gene sequences.

Working within these limitations means that in the literature the causal relationship between abiotic stress and changes in physiological processes and gene expression has been ambiguous (Chaves *et al.*, 2009). Given there are a number of aspects relating to the regulation of *IVR1* expression, which are yet to

be elucidated, it is hoped that this thesis will generate findings which will assist in the development of molecular approaches seeking to increase stress tolerance.

The working hypothesis of the thesis is that maintaining carbohydrate transport to developing pollen under water deficit confers tolerance to drought by maintaining viable pollen development, facilitating optimal fertilization rates and thus final grain number.

The aims of the thesis are to:

1. Characterize the *IVR1* gene family to define the isoform(s) that are active early in head development
2. Investigate the gene network regulation mechanisms determining gene expression under water deficit and non limited conditions
3. Integrate the findings and build a strategy for tackling drought tolerance in pre-breeding programs to evaluate the relative merits of developing molecular markers for genes in marker assisted breeding programs

The structure of the thesis and the interconnectedness of its individual chapters is shown in Figure 1.2. Key drivers in this thesis were the identification of candidate genes associated with a well defined feature of the RSWD stress phenotype.

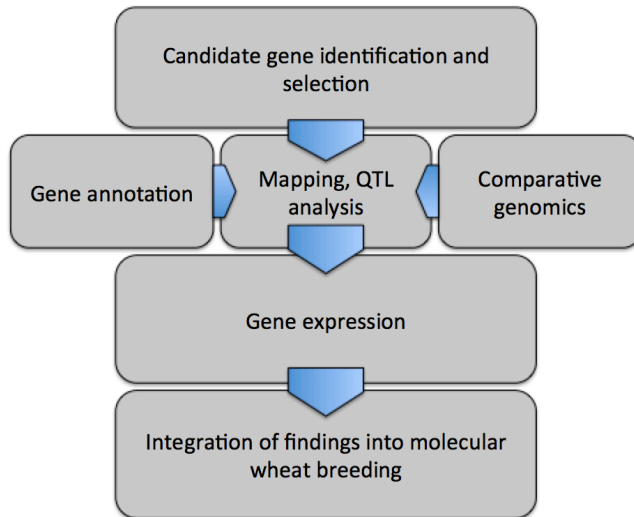


Figure 1.2 Thesis project pipeline.

The experiments were carried out in 2011 and 2012 (Figure 1.3). The objective in the two successive experiments was the same but the experimental design and methodology in 2012 was refined from 2011 to optimize the ability to detect significant differences in gene expression.

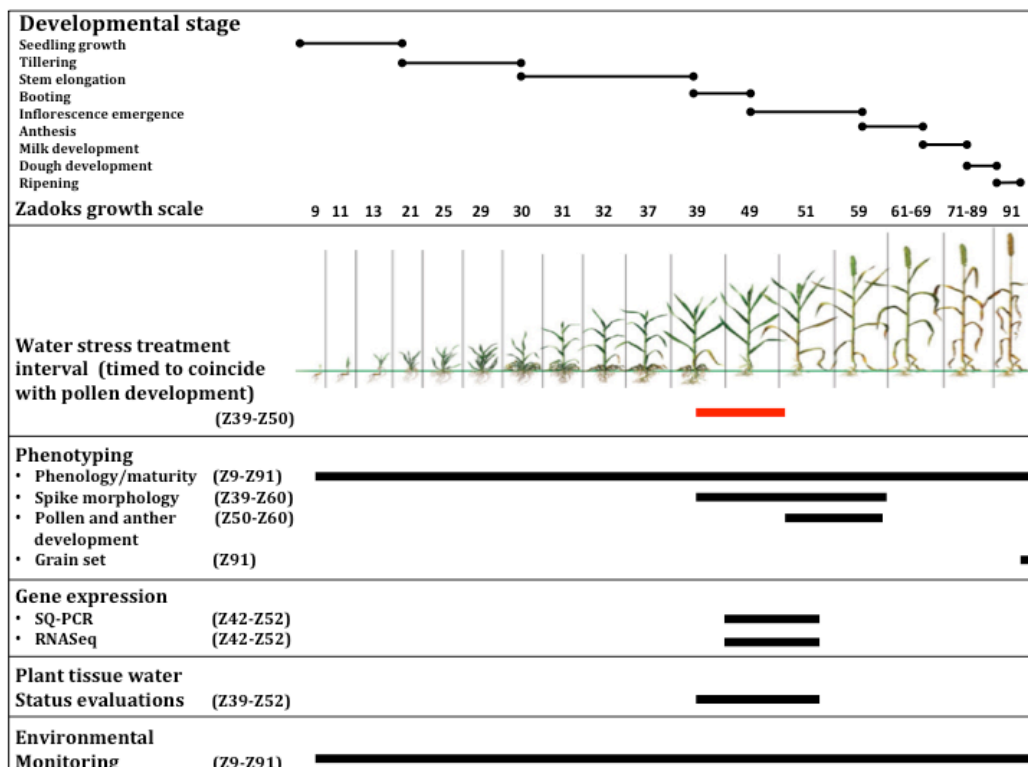


Figure 1.3 The experimental approach in this thesis is summarized. A key feature of the design was that the biological samples and analyses were carried out in triplicate.

Chapter 2: Literature review

2.1 Wheat

2.1.1 Economic and social importance

Among plants providing food for humans and animals, one of the oldest and most widespread is wheat (*Triticum aestivum*) (Paux *et al.*, 2008). The net energy value of global yield output from wheat, rice and maize provides in excess of 45% of the world population daily calorific requirement (Naylor 2009).

On average wheat is grown on more than 200 million hectares of land worldwide as compared to 147 and 139 million for rice and maize respectively (FAO, 2003). Developing countries (excluding those in Eastern Europe) account for approximately 50% of the world wheat production, the leading being China, India, Turkey, Pakistan, and Argentina (Acquaah 2007). *Triticum aestivum* is a temperate crop species and is grown in both hemispheres within the range of 65°North in Norway extending to 45° South to Argentina (Kilian *et al.*, 2009).

The average annual global yield is 620 million tonnes, of which 95% is bread wheat which is utilized in bread making, pastries, and noodles. The remaining 5% is durum wheat which is the main constituent in pasta and semolina food stuffs (Nesbitt and Samuel 1996; Zohary and Hopf 2000). The global trade value of bread wheat production exceeds all other cereals (including the two other major cereal crops rice and maize) and is estimated at a value of 33 billion (USD) per annum (FAO 2010). Utilization of wheat for animal nutrition in intensive livestock production is growing throughout South East Asia, particularly in China where by 2020 its estimated the nation will have to import 2 million tonnes annually to meet the additional demand from the livestock sector (Zhou *et al.*, 2012).

Wheat is the largest grain crop in Australia with an average national yield of 22 million tonnes, of which 70% is sold into the export market. However, the prevailing drought conditions experienced across a number of the major grain

production areas in Australia between 2007-2009 seasons significantly reduced grain production outputs and reduced Australia's national exports by 50% (ABARE 2009). The export value of which is exports approximately \$4 billion (AUD) per annum, and accounts for 14 per cent of the world's export demand (ABS 2006).

2.1.2 Taxonomy and genome structure

Wheat belongs to the grass family *Poaceae* (*Gramineae*) within the tribe *Triticeae* (= *Hordeae*) and together with rye (*Secale*), *Aegilops*, *Agropyron*, *Eremopyron*, and other *Triticineae* genera form the subtribe *Triticeae* (Acquaah 2007).

Depending on its ploidy level wheat is classified into one of three categories: (i) diploid $2n = 14$ = einkorn wheat; (ii) tetraploid $4n = 28$ = emmer wheat; (iii) hexaploid wheat $6n = 42$ = bread wheats (Kilian *et al.*, 2009). The *T. aestivum* species is further classified as: (i) free threshing forms including *T. compactum* (club wheat), *T. spehaerococcum*, (Indian dwarf wheat), *T. petropavovskiyi* (rice wheat), or (ii) hulled *T. spelta* (Dvorak *et al.*, 1998). It is generally accepted the cross between emmer wheat, Durum, and *A. tauschii* occurred in the fertile crescent approximately 8000 years ago, which after undergoing primitive selective breeding, gave rise to the early landraces from which modern bread wheat descends (Dubcovsky and Dvorak 2007; Kilian *et al.*, 2009; Salamini *et al.*, 2002).

Modern bread wheat evolved from three ancestral genomes: A, B, and D, each of which contain seven groups of chromosome ($2n = 6x = 42$, AABBDD)(Francki and Appels 2002). Each chromosome group in turn contains three homologous chromosomes which are derived from respective A, B, D genome donors (Gill *et al.*, 2004) (Figure 2.1). Allohexaploid wheat (*T. aestivum*) has the largest genome among cultivated crop species at approximately 16,000Mb, which is approximately 40 fold and 8 fold times larger than rice or maize respectively (Arumuganathan and Earle 1991). The wheat genome DNA is highly repetitive with approximately 70% being conserved in transposable elements that are often but not always associated with gene free segments of the sequence(Choulet *et al.*, 2010; Paux *et al.*, 2008). The distribution of genes is not equal along

chromosomes but rather small gene islands exist (gene rich regions) which are interspersed between the transposable elements (Breen *et al.*, 2010(B); Feuillet *et al.*, 2012).

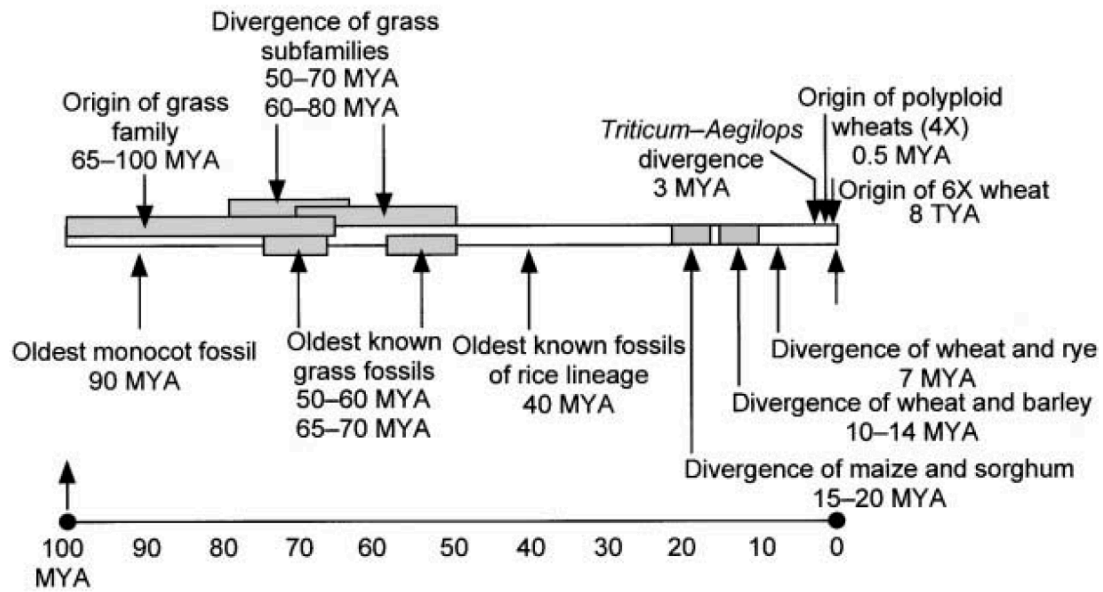


Figure 2.1 Grasses as a single genetic system and recent coevolutionary history of cereals and humans. Grasses originated 55-75 million years ago and now dominate 20% of the land area. The three major cereals (rice, maize, and wheat), which diverged from a common ancestor (≈ 40 million years ago), provide most of the food for humans. Humans and wheat share a remarkably parallel evolutionary history. About 3 million years ago, humans diverged from apes, and diploid A, B, and D progenitor species of wheat diverged from a common ancestor. About 200,000 years ago, at around the same time that modern humans originated in Africa, two diploid grass species hybridized to form polyploidy wheat in the Middle East. Humans domesticated wheat $\approx 10,000$ years ago in the fertile crescent marking the dawn of modern civilisation by cultivating plant species for food (Gill *et al.*, 2004).

Grasses as a single genetic system and recent coevolutionary history of cereals and humans. Grasses originated 55-75 million years ago and now dominate 20% of the land area. The three major cereals (rice, maize, and wheat), which diverged from a common ancestor (≈ 40 million years ago), provide most of the food for humans. Humans and wheat share a remarkably parallel evolutionary history. About 3 million years ago, humans diverged from apes, and diploid A, B, and D progenitor species of wheat diverged from a common ancestor. About 200,000 years ago, at around the same time that modern humans originated in Africa, two diploid grass species hybridized to form polyploid wheat in the Middle East. Humans domesticated wheat $\approx 10,000$ years ago in the fertile crescent marking the dawn of modern civilisation by cultivating plant species for food (Gill *et al.*, 2004).

2.1.3 Definitions of drought

On a global scale, drought (soil and/or atmospheric water deficits), in conjunction with coincident high temperature and radiation, poses the most important environmental constraints to plant survival and to crop productivity (Boyer 1982). Research undertaken to better understand the effects of drought on agricultural crops at a physiological and molecular level has defined the response categories (escape, avoidance, and tolerance) for the purpose of minimizing the 'yield gap' associated with the respective response mechanisms (Cattivelli *et al.*, 2008).

Research into escape mechanisms have focussed on the genetic mechanisms controlling photoperiod (i.e. long/short day) and vernalisation requirement as a means of altering the timing at which the transition from vegetative to reproductive growth occurs (Amasino 2010; Jung and Muller 2009; Trevaskis 2010). Theoretically manipulating phenological development can enable plants to escape drought by ensuring flowering occurs either prior to or after a known period in the growing season that is consistently associated with drought. However, the effectiveness of this approach to mitigate against drought stress is limited based on the fact it does not confer adaptive mechanisms capable of tolerating transient and/or expected drought episodes that may occur during reproductive development (Dolferus *et al.*, 2013).

Drought avoidance mechanisms confer the ability to maintain adequate plant water status during stress through minimizing water loss. Research into drought avoidance has targeted a number of aspects of leaf physiology and morphology to reduce the rate and duration of gas exchange via the stomates including; stomatal conductance, leaf surface area, leaf cuticle type, and cues for senescence of mature leaves (Barnabas *et al.*, 2008). Drought tolerance mechanisms are most commonly associated with tolerance to low water potential and the capacity to maintain critical physiological processes at tissue hydration levels close to or at those which would be consistent with permanent wilting point (PWP) in less tolerant species (Sairam and Saxena 2000).

A range of variables need to be considered in drought tolerance studies, including: (i) genetic factors relating to heritability traits, epistasis, and quantitative trait loci (QTL), (ii) the length and intensity of drought, and the relative contribution from heat and/or water deficit stress effects, (iii) the growth stage at which drought occurs, (iv) the organ, tissue, and subcellular level affected, and (v) the extent of genetic x environment interactions (Piepho 2000).

2.1.4 The significance of water deficit during reproductive growth in wheat

Productivity in agricultural ecosystems is severely reduced by various biotic and abiotic stresses (Boyer 1982). The extent to which crop productivity is affected depends largely on the stage of development at which the plants encounter stress. Water deficit ranks as the most important biotic factor limiting growth. In cereals the most water stress sensitive period after seed germination is reproductive development (Salter and Goode 1967). The high level of sensitivity to water deficit during reproductive stage growth is in part due to: the increased rate of transpiration and thus water demand by the plant required to support increased metabolic activity, and in Mediterranean climate the increased likelihood of water deficits coinciding with this stage due to decreasing rainfall and increasing temperatures (Saini and Lalonde 1998).

Developing reproductive organs require access to photosynthates previously destined for storage in vegetative tissue. Upon the plant sensing water deficit, photosynthesis in leaves is arrested, while respiration continues. During periods of photosynthetic in-activity the plant must utilize its stored carbohydrate reserves until resumption of photosynthesis (Ruan *et al.*, 2010). If photosynthetic in-activity extends for a number of days the plant can go from a position of excess carbohydrates to shortage. The depletion of photosynthetic products results in the portioning of assimilates to feed parental respiration at the exclusion of the energy requirements of the developing florets. Developing anthers in turn must meet their own respiration energy requirements solely from assimilates which had been stored in their own tissue (Figure 2.2),

particularly the highly metabolically active tapetal tissue, prior to photosynthesis being arrested (Ruan *et al.*, 2010).

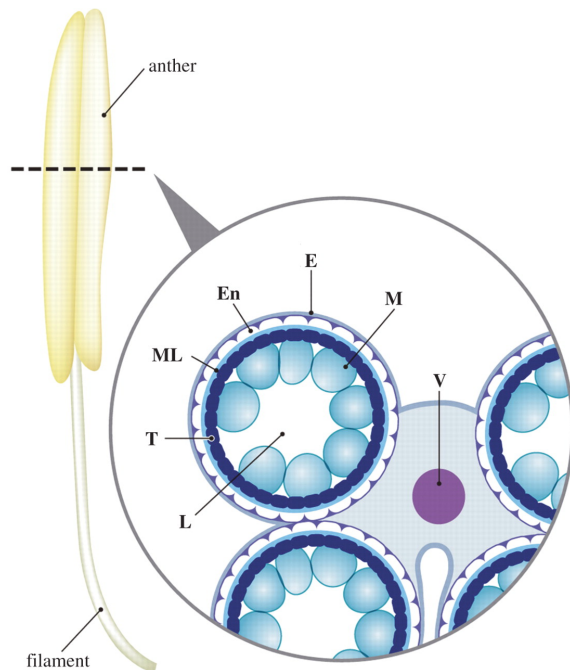


Figure 2.2 Scheme of a cross-section of a rice anther containing immature microspores L = locule; T = tapetum; ML = middle layer; En = endothecium; E = epidermis; M = microspore; V = vascular bundle (Suwabe *et al.*, 2008).

The effect of water deficit (from the period of stamen initiation to anthesis) on viable pollen production has been studied in many cereal species and has consistently been shown to result in yield loss (Saini 1997). In wheat the developmental period when stress sensitivity is most acute has been identified as extending from meiosis to tetrad break-up in the anthers (Dembinska *et al.*, 1992). Saini *et al.*, (1984) reported that in wheat when plants are subjected to a transitory meiotic-stage episode of water deficit stress, microspores develop normally until approximately the first pollen grain mitosis (PGM-1), but then become dislodged from their normal peripheral location and lose contact with the tapetum, claims which were also supported by studies conducted by Ji *et al.*, (2010).

The sensitivity of female reproductive organs is far less than the male gametophyte (Saini and Aspinall 1981). In cross pollinating species it is possible this adaptive tolerance of female gametes to water deficit has evolutionary significance, which would allow the potential for reduced grain set from sterile

pollen production to be offset by cross-cross pollination from excess pollen produced from surrounding plants (Saini 1997). In the wheat ovary the highest period of sensitivity to water deficit stress has been identified at fertilization and grain filling (Zinselmeier *et al.*, 1995(a)). Thus it is plausible to infer the reduced sensitivity of wheat ovaries at the young microspore stage may be attributable to its lower assimilate dependence at this developmental stage.

2.1.5 Regulation of sugar translocation to developing fruit and seed

Higher plants develop from an embryo which depends on heterotrophic metabolism of storage products. With the growth and development of photosynthetically active tissue (leaves) the process of carbohydrate export to (and storage in) less photosynthetically active tissue (fruit and or seed) commences (Gupta and Kaur 2005). This export process begins with sucrose being uploaded into the phloem of leaf tissue either by apoplastic or symplastic flow. The movement of sucrose follows a pressure gradient from the site of uploading in the source tissue to tissues requiring the sucrose (Rosenquist 2007). Sucrose unloading in the fruit/seed tissue may also be via symplastic or apoplastic transfer. While in the majority of plant tissues symplastic transfer predominates there some tissue which are symplastically isolated such as developing pollen cells in the anther where there is no symplastic connection to maternal tissues and thus sucrose transport must follow the apoplastic pathway (Patrick 1997). Symplastic isolation occurs shortly before meiosis, where plasmodesmatal connections between the anther wall and the tapetal tissue (which encapsulate the developing pollen) disappear (Mamun *et al.*, 2005). In the apoplastic pathway the cell wall invertases (CW-INV) play essential roles in hydrolysing sucrose to glucose and fructose, both compounds of which are subsequently transported into the cells via hexose transporters located in the plasma membrane (Bush 1999). Experimental evidence suggests sucrose cleavage by CW-INV in the apoplastic space is one of the key rate limiting steps in carbohydrate metabolism determining the capacity of tissue to attract photosynthetic assimilates for storage (Eschrich 1980).

Apoplastic transport of sucrose provides a transport pathway for sucrose directly into the cell, either the cytoplasm or the vacuole, and hydrolysis by cytoplasmic invertase and vacuolar invertase (Figure 2.3). The cleavage of sucrose in the cytoplasm can also be carried out by the enzyme sucrose synthase (SuSy). The end products of SuSy cleavage of sucrose are different to INV's in that SuSy produces glucose and UDP-glucose, while INV's produce glucose and fructose (Bush 1999). UDP-glucose and fructose can be further metabolised by conversion into glucose-1-phosphate (via UGPase) and fructose-6-phosphate (via fructokinase) respectively (Hawker *et al.*, 1991). When sucrose arrives in the fruit/seed tissue there are five possible metabolic pathways by which it can be hydrolysed (Figure 2.3.), namely, three invertase pathways, a SuSy pathway, and a SST and FFT based pathway. It is rare for all pathways to be active in all tissues.

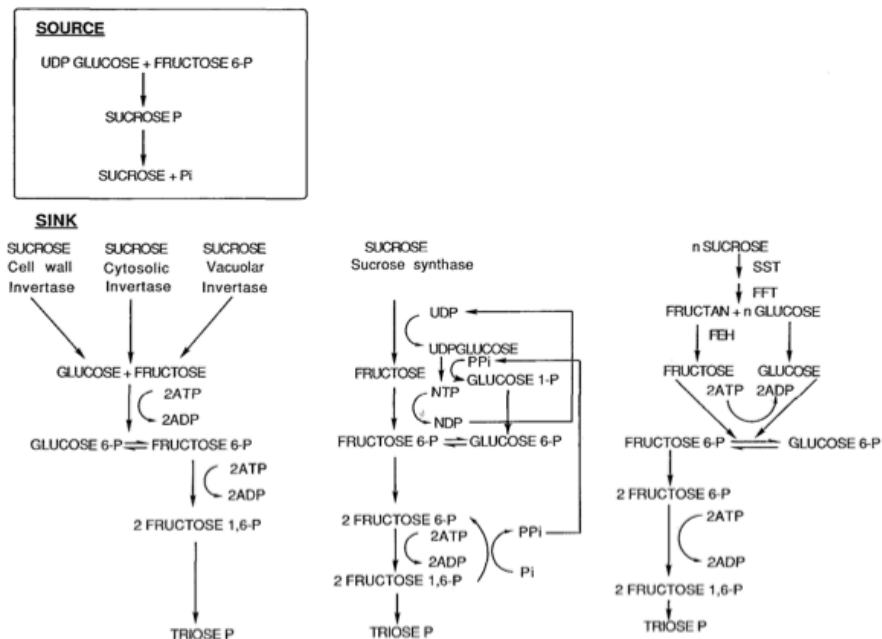


Figure 2.3 Synthesis of sucrose in the leaf and degradation of sucrose fruit/seed. The synthesis of fructan by SST and FFT is not followed immediately by fructan breakdown by FEH, the glucose moiety is available for immediate use as an energy source (Hawker *et al.*, 1991).

The translocation of complex sugars between the leaf and fruit/seed is subject to the physiological demands of different tissue types and growth stages which can alter the number of tissues competing for a common pool of photosynthetic

assimilates. Additional determinants of assimilate portioning within the plant include abiotic stress and pathogen infection (Roitsch 1999). It is widely accepted that there is an interdependence of these environmental and genetic determinants on the partitioning of photosynthetic assimilates. More recent studies into these diverse signalling processes have focused on understanding the crosstalk between the various regulatory mechanisms and their individual signalling pathways (Figure 2.4).

Water deficit stress is one of the major abiotic stress determinants of frequency and throughput of carbohydrate translocation. Under water deficit conditions there is a marked reduction in photosynthetic activity in the leaves that directly reduces the amount of assimilates exported to the fruit/seed tissue (Figure 2.5).

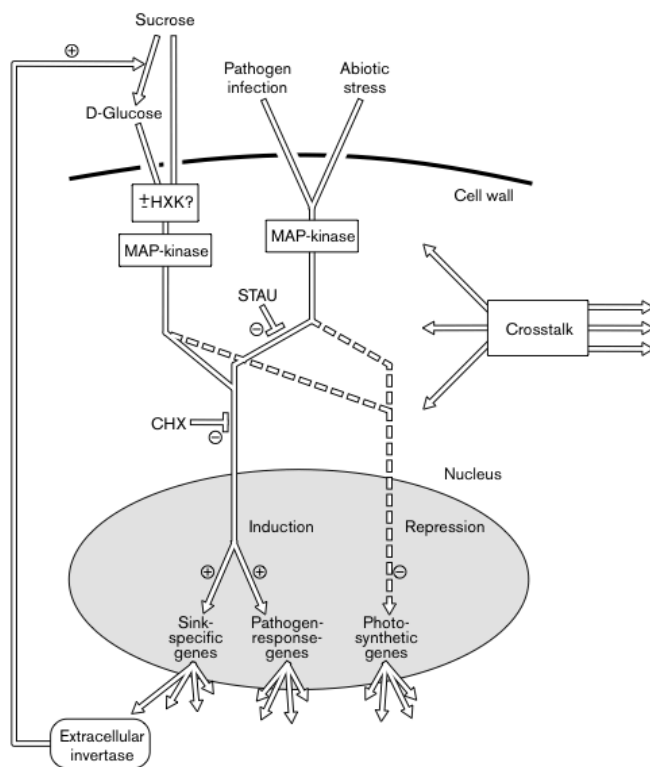


Figure 2.4 Model for the regulation of metabolism, photosynthesis and defence responses by sugars and stress-related stimuli. Sugars and stress-related stimuli activate different signal transduction pathways that are ultimately integrated to coordinately regulate gene expression. Intracellular signalling may involve hexokinase and MAP kinases and is modulated by other signal transduction pathways. Any signal that upregulated extracellular invertase will be amplified and maintained via the sugar-induced expression of this enzyme (Roitsch 1999).

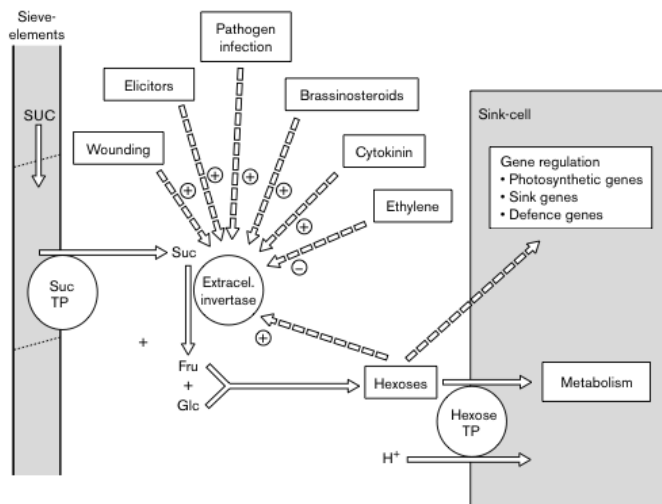


Figure 2.5 Model for apoplastic unloading in fruit/seed tissues. Sucrose from the sieve elements in the phloem are translocated into the apoplast by a sucrose transporter (Suc TP). The disaccharide (Suc) is cleaved by an extracellular invertase and the hexose monomers (Fru and Glc) are taken up by monosaccharide transporters. Extracellular invertase is regulated by glucose and phytohormones. The hexose monomers are the substrates for heterotrophic growth and function as signals for gene regulation (Roitsch 1999).

2.1.6 Sugar signaling

Sugar signaling is an important regulatory mechanism not only in metabolic processes but also plant stress response, growth, and development (Moghaddam and Van den Ende 2012; Rolland *et al.*, 2006; Ruan 2012). The signaling involves three phases; sugar sensing, signal transduction, and target gene expression (Rosenquist 2007). The sugar status of plants has significant implications for the sink strength of developing tissues and their relative ability to attract and assimilate these energy substrates for cell differentiation and growth (Sturm 1999). The status of sugar levels within plants is sensed by sensor proteins which initiate a cascade of gene expression and enzymatic activities (Gupta and Kaur 2005). Sucrose is the primary sugar complex involved in signaling and photosynthetic assimilate transport. However, glucose and fructose, the hydrolytic products of sucrose cleavage, may also act as signaling agents (Rolland *et al.*, 2006).

The molecular mechanisms influencing sucrose transport comprise three broad sucrose sensing mechanisms (Figure 2.6), which determine the direction of assimilate transport (Moghaddam and Van den Ende 2012; Rolland *et al.*, 2006; Ruan 2012); (1) sucrose sensors including hexokinase (HXK) have been found to

interact strongly with plant hormones (ABA, ethylene, auxin, and cytokins) to regulate sugar translocation under stress response (Rolland *et al.*, 2006), and the G-protein complex, also involved in the abiotic and biotic stress response; and (Nilson and Assmann 2010); (2) SnRK1 protein kinases have been found to influence gene expression and enzyme activity in the glucose-6-phosphate (Toroser *et al.*, 2000) and trehalose-6-phosphate pathways (Schluepmann *et al.*, 2004; Zhang *et al.*, 2009(B)); and (3) a sucrose specific pathway also involving SnRK1 protein kinases which acts to control the basic leucine zipper (bZIP)-type transcription factor (Wiese *et al.*, 2004). The distinction within these three regulatory mechanisms is that the first is HXK dependent while the latter two are hexokinase independent. The apparent HXK independent regulation pathways were discovered in a study of the glucose analogue 6-deoxy glucose, which is transported across the plasma membrane but cannot be phosphorylated by HXK, activated the expression of CW-INV and sucrose synthase (SuSy) (Godt and Roitsch 1997; Roitsch *et al.*, 1995).

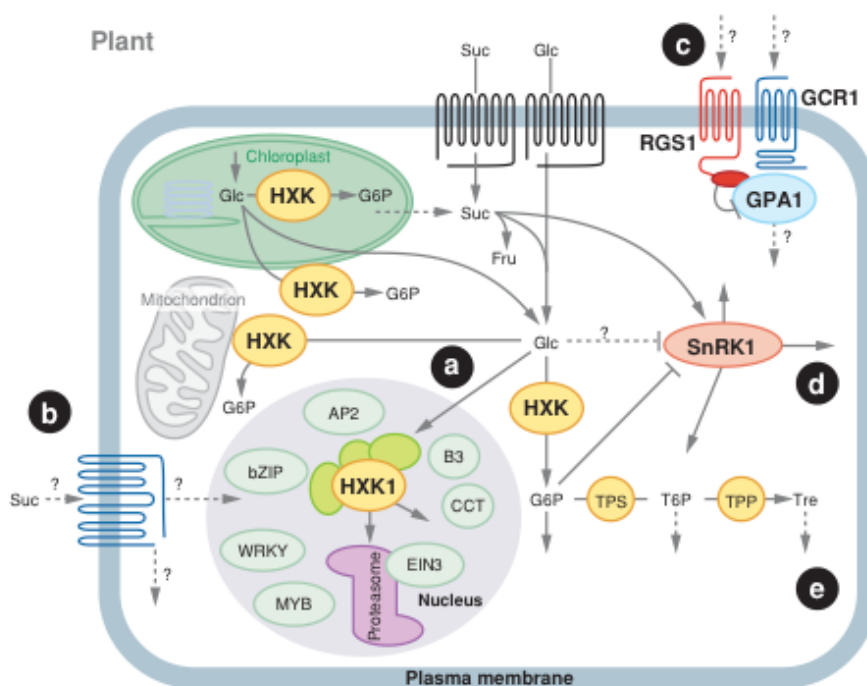


Figure 2.6 Model of sugar-sensing mechanisms in plants. (a) The HXK1 glucose sensor is mainly associated with mitochondria, possibly as part of a glycolytic metabolism. In addition, HXK1 is found in high-molecular-weight complexes in the nucleus where it controls transcription and proteasome-mediated degradation of the EIN3 TF. Other HXK and HKL proteins are also associated with the outer membrane of plastids, including chloroplasts, or cytosol. HXK can also be found in the chloroplast stroma. (b) Sucrose (and other disaccharides) appears to be sensed at the plasma membrane, possibly by transporter homologs. Monosaccharide transporters might have similar functions as

membrane sensors. (c) G-protein coupled receptor signaling by RGS1 and GPA1 is involved in glucose control of seed germination and seedling development, possibly in a hexokinase-independent way. (d) SnRK1 proteins play an important role in plant sugar and starvation signaling, although the significance of the regulation of these proteins by sucrose (Suc) and G6P is still unclear. (e) Important regulatory effects are reported for trehalose (Tre) and T6P, apparently downstream of SnRK1. In the nucleus, several types of transcription factors are involved in sugar-regulated transcription (Rolland *et al.*, 2006).

2.2 Morphology of male reproductive structures in wheat

Reproductive structures of plants provide a major source of food for humans and animals. The ability to control male fertility in reproductive development is very important for plant breeding in agriculture because (1) male sterility in dryland wheat crops results in reduced yield resulting from reduced grain set, and (2) in hybrid breeding systems male sterility can greatly enhance productivity through enhanced control of gene flow in cross breeding (Ma 2005). Because of the economic implications of water deficit induced pollen sterility on reduced grain set there has been extensive cytological observation of the floral development (male and female reproductive structure) to better understand the structural changes in these tissues which render them infertile (Bennett *et al.*, 1973).

In flowering plants male reproductive development begins in the sporophytic generation with the initiation and formation of the male reproductive organ, the stamen in the flower. The development of stamens involves the formation of an anther that has multiple specialized cell types and that house male meiotic cells and a filament that supports the anther (Ma 2005). While stamen development differs slightly between species the basic morphology is similar. To date the majority of cytological observations of stamen development has been performed in Arabidopsis. Floral development extending from the initiation of stamen primordia through to its completion after meiosis has been characterised at length and has been divided into 14 stages (Table 2.1, Figure 2.7) (Zhang and Wilson 2009).

Table 2.1 Stages of cellular development in the anther (Adapted from Ma 2005).

Stage	Developmental process
1	L1, L2, L3 layers divide resulting in formation of anther primordia
2	L1 differentiates to form epidermis L3 differentiates in connective and vascular tissue L2 periclinal and anticlinal cell divisions lead to formation of four archesporial
3	Archesporial cells divide to form primary parietal layer and sporogenous cells
4	Secondary parietal layer differentiates to form endothecium and middle layer
5	Four individual lobes are distinct and anther morphogenesis is complete. The anther now has 4 non-reproductive layers: epidermis, endothecium, the middle layer and the tapetum. Within these layers are the microspore mother cells (MMC).
6	MMC's undergo meiosis. When meiosis is complete MMC's dissociate from the tapetum and each other. Resulting in a space called the locule.
7	Completion of meiosis and formation of tetrad of haploid microspores. Microspores are released from the tetrad following degradation of the callose wall.
9-12	Microspores develop into pollen grains
10-11	Tapetal degeneration
12	Pollen mitotic division occurs
13-14	Pollen dehiscence

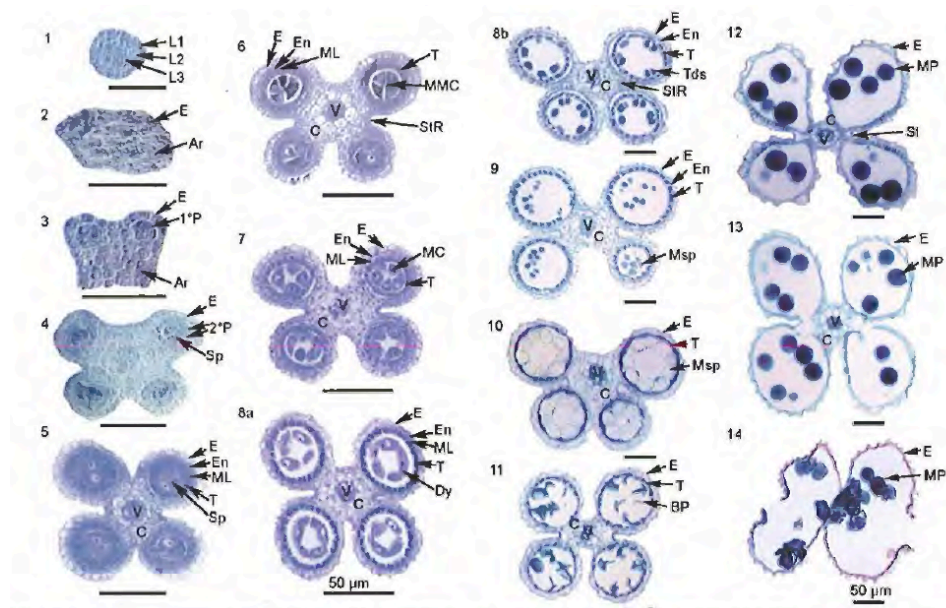


Figure 2.7 Development stages of the rice anther. Ar, archesporial cell; C, connective tissue; BP bicellular pollen; Dy, dyad cell; E, epidermis; En, endothecium; L1, L2, and L3, the three cell-layers in stamen primordia; MC, meiotic cell; MMC microspore mother cell; MP, mature pollen; MSp, microspore; parietal cell; 1°P, primary parietal layer; 2°P, secondary parietal cell layer; Sp, sporeogenous cell; St, stomium; StR, stomium region; T, tapetum; Tds, tetrads (Zhang and Wilson 2009).

2.2.1 Pollen development

The male gametophyte (pollen and later the pollen tube) in flowering plants is a microscopic structure that completes its early development within the

sporophytic tissue of the anther (Mascarenhas 1990). The male pollen tube participates in a process of double fertilization with the female ovule, resulting in the formation of the zygote and primary endosperm. The individual stages of microsporogenesis have been described by Ma (2005) during the early stage of pollen development, the microspore becomes vacuolated after release from the tetrad; the formation of a large centralized vacuole is accompanied by the migration of the microspore nucleus to one side of the cell. The first mitosis in pollen development is asymmetric, producing a large vegetative cell and a small generative cell. The vegetative cell contains an uncondensed nucleus and most of the cytoplasm from the microspore, whereas the generative cell has highly condensed chromatin and very little cytoplasm. Subsequently, the generative cell is completely engulfed by the cytoplasm of the vegetative cell; during mitosis II two sperm cells form from the generative cell. During this process it is evident that transport of carbohydrate across cell walls to the developing pollen is a significant variable. Different tissues have different requirements for assimilates and two possible pathways that can be utilized use sucrose synthase or cell wall invertase to hydrolyze sucrose. An important distinction between these two pathways is that hydrolysis of sucrose (intended for conversion into starch) by cell wall invertase is more energetically expensive and less efficient in the conversion of sucrose into hexoses than via the sucrose synthase route, with the former requiring an additional ATP unit (Ferreira and Sonnewald 2012).

Dispersal of mature pollen grains is usually by wind, water, or insects, which facilitate their deposition on the stigma. Germination of the pollen commences with the pollen tube extruding through a germ pore in the pollen wall. The pollen tube (carrying the sperm cells) grows through the style until it reaches the embryo sac. The sperm cells are then discharged into the synergid. Thereafter one sperm cell fertilizes the egg to form the zygote (diploid) and the second sperm cell fuses with the central cell (diploid) of the embryo sac to form the triploid primary endosperm cell (Mascarenhas 1990; McCormick 1993; McCormick 2004)

The role of the male gametophyte is the production of two sperm cells and their transport within the pollen tube through the tissues of the style and into the embryo sac in the ovule. The stages of pollen development have been described by Ma (2005). In the double fertilization that follows, fusion of one sperm with the egg and of the second sperm with the central cell results in the formation of the zygote and a primary endosperm cell respectively.

The events that culminate in the formation and release of the pollen grain from the plant are inherently linked to meiosis and tapetal cell development and involve an intricate and tightly controlled set of structural and molecular changes, requiring gene expression in both the gametophytic and sporophytic tissues in the anther Figure 2.8 (McCormick 1993).

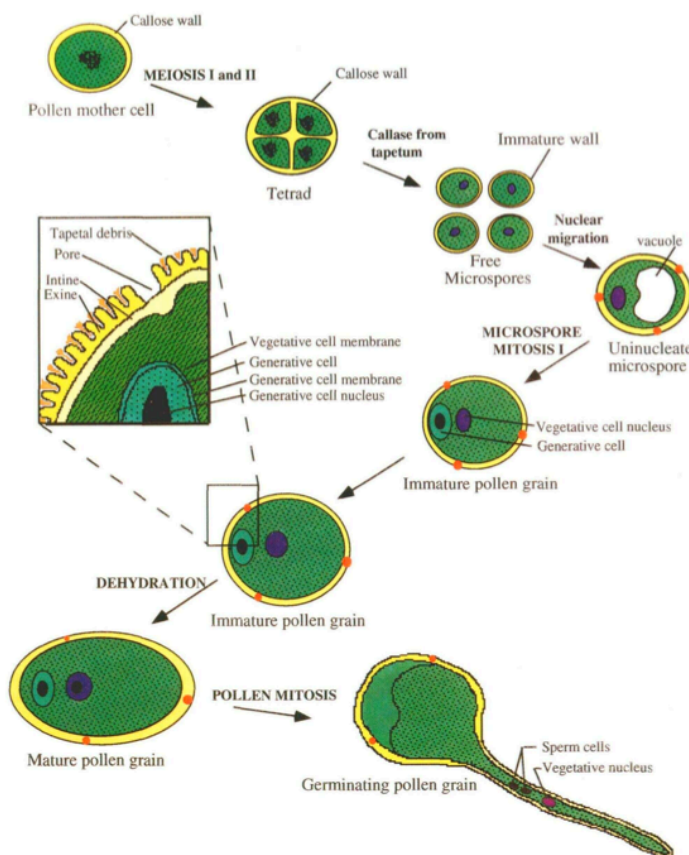


Figure 2.8 Schematic representation of microsporogenesis typical in most angiosperms (McCormick 1993).

2.2.2 Tapetal development

The tapetum is the tissue in the anther that is in closest contact with the developing pollen during microsporogenesis. Early research into the importance of the tapetum to pollen development focused on understanding how its properties provided essential structural support. However, there is now a large body of evidence that indicates the tapetum has additional functionality. These functions are believed to be essential to the viable development and maturation of pollen, specifically (1) supply of complex carbohydrates required to support pollen growth, and (2) also as a metabolic signal sensory structure involved in the sugar transduction pathway (Golberg *et al.*, 1993). Evidence first supporting these claims was provided by studies which isolated microsporocytes and early microspores of *Trillium* and grew them to maturity in tissue culture without the tapetum and other anther tissues (Takegami *et al.*, 1981). These results paved the way for future research which sought to test the theory the tapetum function was not in fact to supply critical morphogenetic components essential for the progress of pollen differentiation but rather concerned with the mobilization and production of simple nutrients and the necessary physical conditions for development (Mascarenhas 1990).

Studies into molecular cell ablation of tapetal tissue have demonstrated the timing of programmed cell death (PCD) is critical to pollen development, where by PCD must occur in synchronisation with the onset of late gametogenesis at the PMI stage (Varnier *et al.*, 2005). In the event that these processes are asynchronous then significant levels of pollen sterility can result in the developing reproductive structure and thus crop yield loss (Goldberg *et al.*, 1995; Parish and Li 2010). Tapetal development is highly sensitive to abiotic stress and the cytological changes that result have been characterised in a number of crops including wheat (Lalonde *et al.*, 1997; Saini *et al.*, 1984) rice (Ji *et al.*, 2010; Jin *et al.*, 2013; Mamun *et al.*, 2006; Oshino *et al.*, 2007) barley (Abiko *et al.*, 2005; Oshino *et al.*, 2007) and arabidopsis (Kim *et al.*, 2001).

There are two types of tapetum; secretory and amoeboid. Secretory is the form where the tapetal cells remain intact for the majority of pollen development and

secrete nutrients into the anther locule. Conversely amoeboid tapetum, in addition to providing nutrients supporting pollen growth, combine with the outer pollen wall layer which forms the pollen exine. Amoeboid tapetum is the form present in wheat and rice (Taylor *et al.*, 2012). In wheat and rice it is believed that orbicles, which are globular lipid cores, are essential to pollen development potentially as an energy source from the sporophytically produced proteins they carry (Parish and Li 2010; Wang *et al.*, 2003). Normal tapetal PCD under non stress conditions has also been studied in a number of other species including *Oryza sativa* (Li *et al.*, 2006), *Tillandsia albida* (Papini *et al.*, 1999), *Lilium longiflorum* (Papini 1990; Reznickova and Willemse 1980), *Beta vulgaris* (Majewska-Sawka *et al.*, 1993), *Brassica spp.* (Huysmans *et al.*, 1998), and *Arabidopsis thaliana* (Owen and Makaroff 1995; Zhang *et al.*, 2002). While there is some temporal and special differences in the processes of PCD between secretory tapetal species the primary characteristics are conserved. These common processes include; vacuolarization at microspore mother cell meiosis, loss of the cell wall at the tetrad stage, shrinkage of the nucleus and peripheral chromatic condensation, reduction in size and persistence of mitochondria, and tapetal disintegration immediately prior to meiosis (Parish and Li 2010).

2.2.3 Gene network regulation of during pollen development

In a recent review by Parish and Li (2010) the conclusion was drawn from the wide body of evidence that, of the types of programmed cell death evident in plants and animals, the cellular processes involved in tapetal degradation most strongly resembled apoptosis like PCD (Figure 2.9).

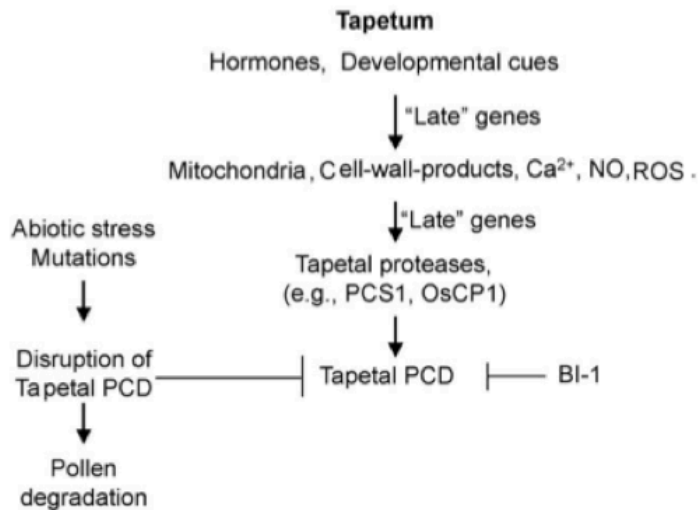


Figure 2.9 In the suggested tapetal PCD pathway (proposed by Parish *et al.*, (2010)), a differentiation program activates “late” tapetal genes. The cell death program includes modification of hormone levels, increased mitochondrial permeability and changes in cytoplasmic free Ca²⁺ concentrations. Breakdown products of the tapetal cell wall elicit aspects of the program and might involve nitric oxide and ROS production. Tapetal proteases resulting from the activation of “late” tapetal genes may directly or indirectly influence PCD.

Kapoor *et al.*, (2002) suggested that both constructive and degenerative forces are acting on tapetal tissue at various stages of development; and when the gene or genes responsible for the constructive forces are down regulated as part of development, the degenerative forces proceed unchecked. This hypothesis is gaining support due to the fact that it not only accounts for single gene effects in tapetal degradation but also the interdependent genetic cues that regulate the balance between cellular development, maintenance, and degradation (Kapoor *et al.*, 2002).

To date the majority of tapetal development studies have been conducted in *Arabidopsis thaliana*. The output from these works has enabled a number of genes involved in tapetal cell development to be identified. Inroads have also been made toward determining their relative levels of activation or repression during tapetal development (Ma 2005). Of particular relevance to better understanding water deficit induced tapetal degradation the next step in the research should be focussed on categorise the functional role of the individual genes as role as being either constructive or degenerative how their expression is influenced by water deficit stress (Parish and Li 2010). A number of

transcription factors involved in *A. thaliana* anther development have been identified (Wilson and Zhang 2009). Of specific relevance to this thesis are the proteins coding for AMS, AtMyB33, AtMyB36 (Table 2.2).

Table 2.2 Transcription factors and cell receptors involved in *Arabidopsis thaliana* anther development. Adapted from Parish and Li (2010).

Protein	Expression Pattern	Mutant phenotype	Putative function in anther development
<i>AtMYB33</i> and <i>AtMYB65</i>	Initially in all anther cell layers; following meiosis, largely confined to tapetum.	Redundant genes; abnormal tapetum at meiosis, undergoes hypertrophy at pollen mother cell stage.	Tapetal differentiation.
ABORTED MICROSPORES (<i>AMS</i>) (MYC class of bHLH)	Tapetum, post meiosis. (Possibly also in microspores).	Male sterile. Tapetal cells become large and vacuolated. Premature tapetum and microspore degeneration.	Tapetal and microspore development.

In *A. thaliana* the expression of *AtMYB33* and *AtMYB65* genes (members of the GAMB-like gene family) have been found to be tapetal specific and temporal expression profiles indicate this occurs within the developmental stages just prior to meiosis (Figure 2.10). The highest expression occurs during the mobilization of starch reserves in the tapetum. Double mutant expression was also found to coincide with higher levels of soluble carbohydrate. In light of these results tentative functionality for the genes was assigned as being involved in starch mobilization (Gocal *et al.*, 2001; Millar and Gubler 2005). In rice GAMB mutants have also been shown to be clocked at a similar stage to *AtMYB33* and *AtMYB65* and to undergo tapetal hypertrophy (Kaneko *et al.*, 2004). Thus the expression and function of GAMBY's appear to be highly conserved between Arabidopsis, rice, and close cereal family members such as wheat (Millar and Gubler 2005).

In Arabidopsis the aborted microspore (*AMS*) gene encodes a basic-helix—loop (bHLH0 protein belonging to the MYC class family of transcription factors. Studies of *AMS* mutants in Arabidopsis have shown degradation of microspores occurs prior to nuclear migration (Sorensen *et al.*, 2003). Interestingly comparative analysis of *AMS* to close gene relatives in rice has also provided evidence to

suggest there is functional conversation of this family of genes between plant species (Parish and Li 2010). The Tapetum Degradation Retardation gene in rice encodes a bHLH transcription factor most similar to the Arabidopsis *AMS* protein. Similar to *AMS*, the *TDR* gene is predominantly expressed in the tapetum at the young microspore stage. Studies of the *TDR* mutants gave rise to completely male sterile microspores (Li *et al.*, 2006). Parish *et al.*, (2010) suggested the changes in gene expression of the tapetum during PCD may also be influenced by hormones and proteases particularly hormones ethylene and ABA. Figure 2.10.

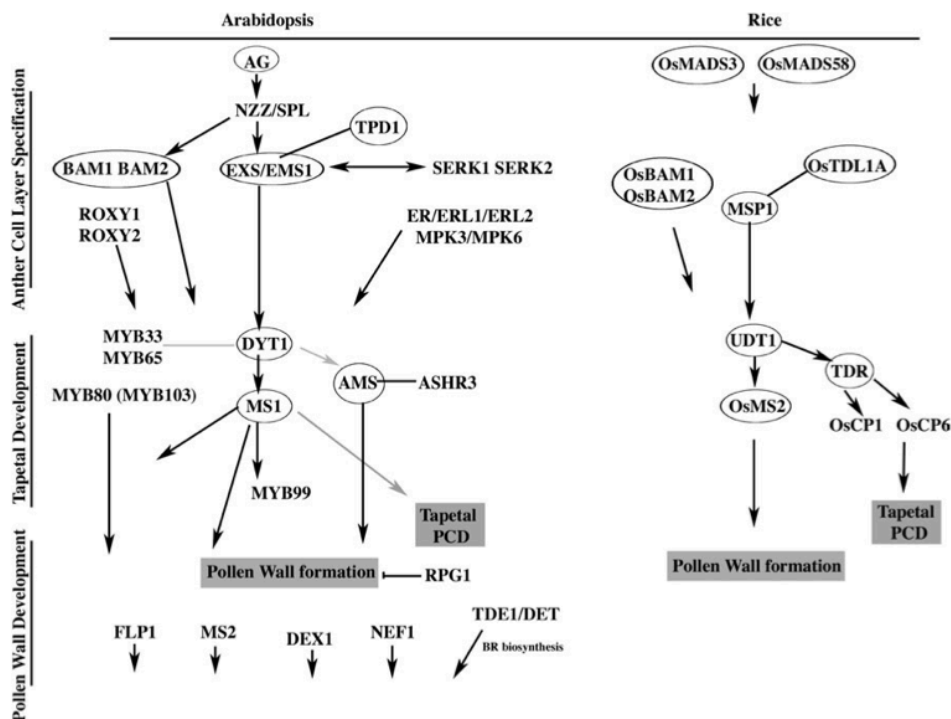


Figure 2.10 Comparisons between the network of pollen development in Arabidopsis and Rice. (AG, AGAMOUS; AMS, Aborted microspore; BAM1, Barley any meristem 1; BAM2, Barley any meristem 2; DET, De-etiolated; DEX1, Defective exine patterning 1; DYT1, Dysfunctional tapetum 1; EXS/EMS1, Extra sporogenous cells/excess micro-sporocytes 1; FLP1, Faceless pollen 1; MS1, Male sterility 1; MS2, Male sterility 2; MSP1, Multiple sporocyte 1; NEF1, No exine formation 1; NZZ/SPL, Nozzle/sporocyteless; RPG1, Ruptured pollen grain 1; SERK1, Somatic embryogenesis receptor-like kinase 2; TDE1, Transient defective exine 1; TDR, Tapetum degradation retardation; TPD1, Tapetal determinant 1; UDT1, Undeveloped tapetum 1)(Wilson and Zhang 2009).

2.3 Water deficit

2.3.1 The effect on pollen development

Grain development begins with the process of double fertilization, hence the physiological processes ensuring viable development and fusion of the gametes is the first and most important rate limiting steps determining final grain yield (Barnabas *et al.*, 2008). Pollen development is critically sensitive to drought stress, to a much greater extent than the ovary, and in particular the developmental transition phase between meiosis to microspore (Figure 2.11 from Koonjul *et al.*, 2005) has been identified as most vulnerable (Dolferus *et al.*, 2013). In a recent review by Storme and Geelen (2013) evidence compiled from a number of studies prompted the authors to conclude that there is compelling grounds to suggest impeded tapetal development, caused by drought stress, is in the primary cause of the pollen sterility further downstream in the drought response pathway.

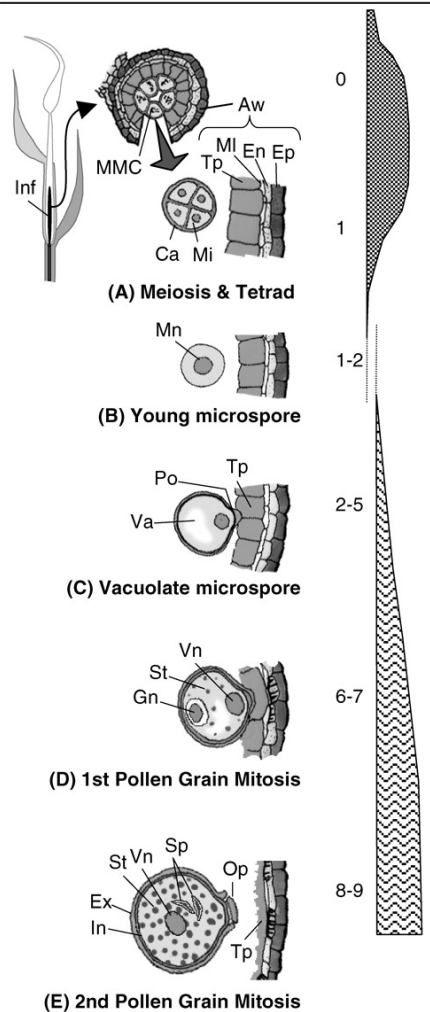
Stages during normal development	Days	Normal Development	Development following water stress
 <p>(A) Meiosis & Tetrad</p> <p>(B) Young microspore</p> <p>(C) Vacuolate microspore</p> <p>(D) 1st Pollen Grain Mitosis</p> <p>(E) 2nd Pollen Grain Mitosis</p>	0 1 1-2 2-5 6-7 8-9	<p>Inflorescence protected by two uppermost leaf sheaths. A callose envelope isolates each MMC, which undergoes meiosis. Starch in anther wall.</p> <p>Meiosis produces a tetrad of microspores.</p> <p>Callose envelope dissolves, releasing thin-walled microspores, which align along tapetal cells. Degeneration of tapetum begins.</p> <p>Microspores with a prominent vacuole, always in contact with tapetum via a pore. Tapetum degenerating, microspore-wall forming. Starch disappears from anther-wall. Anthers start enlarging.</p> <p>Microspore nucleus divides into vegetative and generative nuclei. A separate cell forms around the generative nucleus. Pollen grains begin accumulating starch. Rapid pollen-wall formation. Pore remains positioned towards tapetum.</p> <p>Generative nucleus divides to form two ovoid sperms. Exine and intine formation advanced, pollens full of starch, and tapetal degeneration nearly complete. Pollen ready for shedding within a day.</p>	<p>In plants stressed during meiosis and then re-watered: Callose envelope present, meiosis is completed.</p> <p>Anther development normal. Tetrads formed.</p> <p>Occasional microspore disorientation or abnormal vacuolation of tapetum, otherwise development proceeding normally.</p> <p>Tapetal vacuolation progresses. Microspores losing contact with tapetum, which degenerates too fast or persists. <i>No more sensitivity to concurrent water deficit.</i></p> <p>Pollen grains fail to accumulate starch, have poorly-formed or no intine. All abnormal pollen grains are detached from tapetum. Anther smaller and paler than normal.</p> <p>Shriveled pollen grains, devoid of starch and with thin or no intine. Pollen grains non-viable and fail to germinate.</p>

Figure 2.11 A summary of the development of normal wheat anthers, together with the timing and sequence of events during the failure of pollen development in response to a transitory episode of water deficit during meiosis (meiosis to tetrad). The formation is adapted from Saini *et al.*, (1984) and Lalonde *et al.*, (1997). (Filled bar) relative sensitivity to water deficit; (hatched bar) relative symptoms of developmental disruption. Abbreviations denote: Aw, anther wall; Ca, callose; En, endothecium; Ep, epidermis; Ex, exine; Gn, generative nucleus; In, intine; Inf, inflorescence; Mi, microspore; MI, middle layer, MMC, microspore mother cell; Mn, microspore nucleus, Op, operculum; Po, pore; Sp, sperm; St, starch; Tp, tapetum; Va, vacuole; Vn, vegetative nucleus (Koonjul *et al.*, 2005).

As discussed in previous sections of the literature review, drought causes asynchrony between the occurrence of tapetal PCD and pollen meiosis, which in turn leads to impeded microspore development and ultimately pollen sterility. The drought induced developmental abnormalities and cellular processes underlying altered onset of tapetal PCD have been evaluated in rice (Gothandam *et al.*, 2007; Ku *et al.*, 2003; Suzuki *et al.*, 2001) and beans (Suzuki *et al.*, 2001),

and the results showed the endoplasmic reticulum (ER) is the cellular organelle in which functionality is critically compromised. In plants the subcellular ER has a number of important functions, including: (1) mediating post-translational processing of newly formed membrane-specific proteins, which act as the building blocks of tissue deposition in pollen development and exine layer formation, (2) acting as a quality control system to eliminate misfolded proteins (Helenius *et al.*, 1993; Hurtley and Helenius 1989), and (3) generating essential callose-degrading enzymes, which are critical to downstream pollen wall formation (Wu and Yang 2005). It is generally understood that excessive overload and/or prolonged dysfunction of the ER leads to the altered timing of PCD and/or a combination of cellular defects during the formation of the pollen wall. These results have been consistently reported in a number of studies evaluating a range of plant species. These observations prompted Storme and Geelen (2013) to postulate in their recent review that the tapetal ER constitutes an important centre of drought sensitivity and mediates the stress-induced male sterility through a variety of biological processes.

2.3.2 The effect of sugar accumulation in developing anthers

Abiotic stress induced alterations in anther sugar content and carbohydrate profile have been widely studied in a range of crops including wheat (Dorion *et al.*, 1996), rice (Fu *et al.*, 2011), sorghum (Jain *et al.*, 2007), and tomato (Pressman *et al.*, 2002). However, comparisons between these studies highlights significant variability in their respective findings, which points to the fact there is still a significant gap in the collective research communities understanding of how carbohydrates metabolism in plants is affected by abiotic stress. In one study of wheat sugars accumulation in the anthers in response to water deficit stress it was reported that sugar accumulation was upregulated (Saini 1997; Saini and Westgate 1999), conversely in a similar study by Lalonde *et al.*, (1997) accumulation was observed to be significantly repressed. Interestingly in a recent water deficit stress study in rice it was found that drought-tolerant rice varieties show an increased sugar content in the anthers, compared to low sugar levels in susceptible varieties (Fu *et al.*, 2011). In a review by Dolferus *et al.*, (2013) the authors claimed accumulation of sugars in water deficit stressed

anthers in cereal crop species indicated that aborted pollen was not due to repressed production of photo-assimilates under drought conditions. The alternative conclusion drawn by Ji *et al.*, (2010) was that sugars are accumulated and stored in the anthers as starch, rather than being transported to the developing pollen. The inference being that sugar metabolism in vegetative tissue and subsequent transport to anthers is not repressed by water deficit stress but sugar transport within the anther is.

Likely causes of the discrepancies between the results from various studies relate to differences in experimental approach for cytological observations, including: (1) what cellular component or tissue the sugar profiling data is accumulated on (i.e. tapetum, locular space, and the pollen itself), (2) at what stage during the anther development cycle the sugar profile observations are made, and (3) which carbohydrate substrates are measured (i.e. sucrose, glucose, fructose, and starch). It is known that under non stress conditions starch begins to accumulate only a few days before anthesis but for this to occur sucrose (the metabolite pre-cursor for starch) must be transported to the pollen where it undergoes conversion to hexose sugars (glucose and fructose) and then finally to starch (Raghavan 1988). It is, therefore, reasonable to suggest that studies evaluating the effects of water deficit stress on carbohydrate profiles in developing anthers as a measure of pollen sterility need to observe all sugar substrates, in the full range of pollen development stages, and in all tissues and cellular compartments where storage and synthesis occurs.

2.3.3 How water deficit is sensed by the reproductive tissues

In a review of studies observing the effect of water deficit on the fertility of male gametophytes in cereals, Saini (1997) concluded that water deficit stress induced grain set does not correlate with the water status of reproductive organs. It has been reported that while grain set is a direct measure of male fertility, grain set is not reduced until xylem water potential (Ψ_w) falls below -1.2MPa (Fischer 1973). However, in this study further reduction in Ψ_w below -1.2 resulted in a linear decline in grain set until xylem Ψ_w dropped to -2.4MPa, at which point grain set is zero. The Fischer (1973) study leaves open the

question of how is the environmental cue for the stress sensed, and by what molecular mechanism is the signal received in the reproductive tissue, to ultimately causes failed of carbohydrate delivery to the developing pollen. In rice and wheat a number of studies have observed the Ψ_w of floral organs under water deficit in tandem with flag leaf Ψ_w (Dorion *et al.*, 1996; Tsuda and Takami 1992; Westgate *et al.*, 1996). In all studies the Ψ_w of the floral organ was reported to remain unchanged or only fall marginally, as compared to the Ψ_w of the flag leaf. Saini (1997) proposed that the relative immunity of the inflorescence to dehydration is likely to be related to it being protected by two or more leaf sheaths.

2.4 Translocation of complex carbohydrates to developing pollen

Long distance transport of assimilates from leaves into reproductive tissues such as the tapetum is driven by differences in osmotic potentials. Thus, cleavage of sucrose at the site of phloem uploading and metabolism of the cleaved products are rate controlling steps in the delivery of carbohydrates to the reproductive tissues (Eschrich 1980).

2.4.1 The role of cell wall invertase in carbohydrate delivery to pollen

Plant invertases (β -D-fructofuranosidase EC 3.2.1.26) constitute a family of enzymes that hydrolyse sucrose into glucose and fructose. In plants three types of invertases have been purified, including; cell wall, vacuolar and cytoplasmic. Invertases have been found to have a role in a number of plant physiological process ranging gene expression to long-distance nutrient allocation and are involved in regulating carbohydrate partitioning, developmental processes, hormone responses and biotic and abiotic interactions (Roitsch and Gonzalez 2004; Tymowska-Lalanne and Kreis 1998). However, for the purpose of this literature review carbohydrate partitioning and related issues of sugar delivery to developing pollen are the focus.

Sucrose is the major form of carbohydrate transported from photosynthetically active tissues (i.e. primarily leaves) to non photosynthetic tissues such as

flowers, fruit, seed, and roots (Ruan *et al.*, 2010). To utilize sucrose as an energy source requires the molecule to be cleaved at the α 1- β 2-glycosidic bond. Cleavage can be achieved by two different enzymes invertase (β - fructofuranosidase) or glucosyltransferase sucrose synthase (Figure 2.12 from Koch 2004).

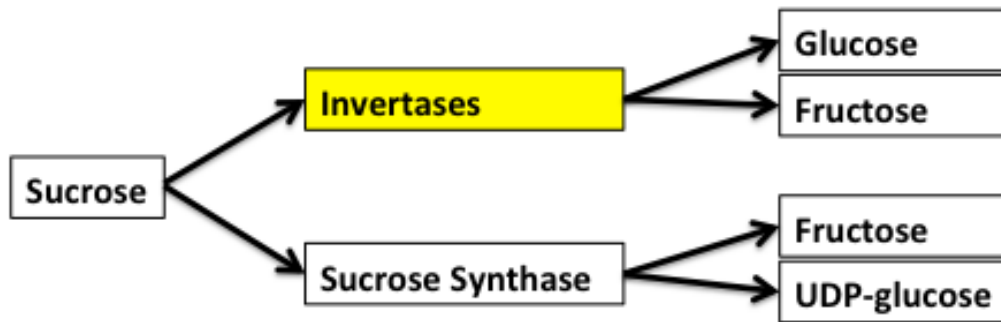


Figure 2.12 Sucrose cleavage enzymes in developing anthers (Koch 2004).

Invertases catalyse the irreversible hydrolyses of sucrose to glucose and fructose while glucosyltransferase sucrose synthase performs reversible cleavage of sucrose to UDP-glucose and fructose (Roitsch and Gonzalez 2004). Developing microspores are symplastically isolated thus must receive carbohydrates via apoplastic transport and CW-INV are critical in this transport process (Roitsch 1999; Sturm 1999) (Figure 2.13).

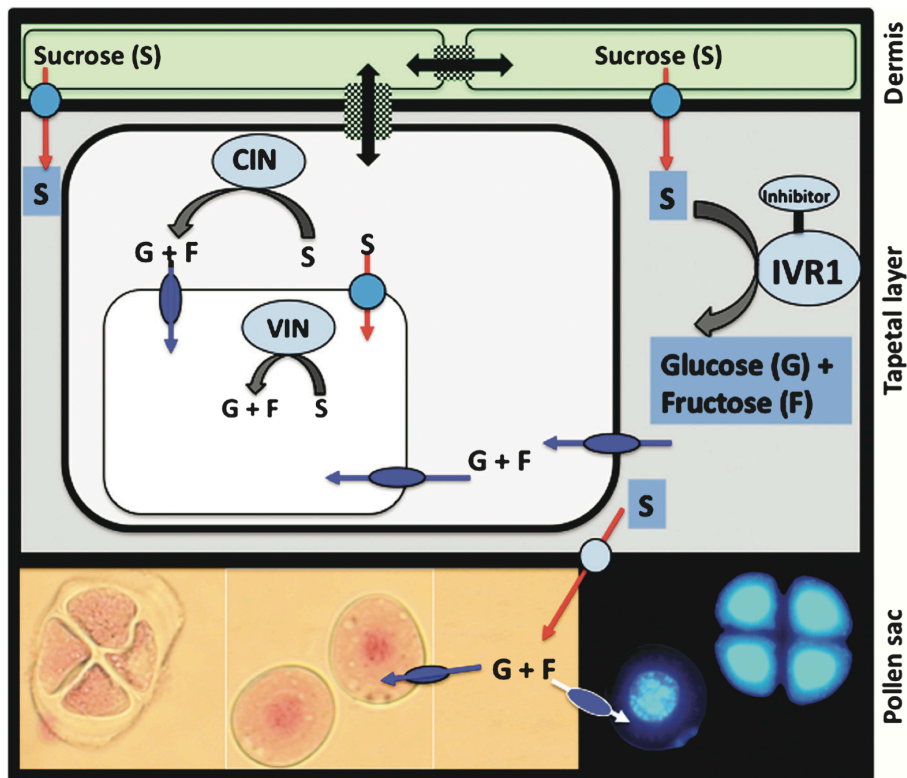


Figure 2.13 A Schematic Diagram representing the flow of carbon nutrients to developing pollen in the pollen sac of anthers. The diagram illustrates a section of an anther and the flow of sugars to the symplastically isolated developing pollen. In the dermis, symplastic (double-headed arrow) and active transport of sucrose (red arrow) occurs and within the tapetal layer the three broad classes of invertases that can break sucrose down to glucose and fructose, if they are expressed, are shown. CIN is the cytoplasmic invertase, VIN is the vacuolar invertase and IVR1 is the cell wall invertase. In the pollen sac layer the developing pollen (tetrad and uni-nuclear stages) from wheat are shown as acetocarmine stained cells (lower left side of panel) and DAPI-stained fluorescent images (lower right side of panel). The network involved in maintaining the sugars at an optimum level includes the interactions of a protein inhibitor with the cell wall invertase to moderate the breakdown of sucrose to glucose and fructose (Webster *et al.*, 2012).

The three different invertase isoforms (cell wall, vascular, and cytoplasmic) are categorised depending on the solubility, subcellular localization, and pH optima. The existence of different invertase isoforms is likely to provide greater flexibility in the control of sucrose metabolism, translocation, or storage under different internal and external stimuli and at various developmental stages of tissue or at the whole plant level (Akhunov *et al.*, 2003). Vacuolar and cell wall invertases share numerous enzymatic properties, e.g. they cleave sucrose most efficiently between pH 4.5-5.0 and attack the disaccharide complex from the fructose residue (Sturm 1999).

CW-INV and V-INV are synthesized as pre-proteins, with a long leader sequence that is cleaved off during transport and protein maturation (Figure 2.14 from Tymowska-Lalanne and Kreis 1998). The leader sequence consists of two segments: a signal peptide and a remaining leader sequence. The N terminal domains of CW-INV and V-INV have up to 100 amino acid residues comprising of a signal peptide and a N-terminal peptide that may be related to folding, targeting, and the regulation of activity (Sturm 1999).

Both CW-INV and V-INV contain a motif N-DPN-G/A box close to their N-terminus in the mature proteins (Figure 2.14 from Tymowska-Lalanne and Kreis 1998). However, CW-INV differs from V-INV in its valine cysteine domain, where the former is characterized with WECPD, where as the latter becomes WECVD (Ji *et al.*, 2005). The other structural point of difference between CW-INV and V-INV is the presence of an additional C-terminal extension in V-INV (Tymowska-Lalanne and Kreis 1998). The genomic organization and intron-exon structure is moderately conserved between monocots and dicots, which some notable exceptions (Tymowska-Lalanne and Kreis 1998).

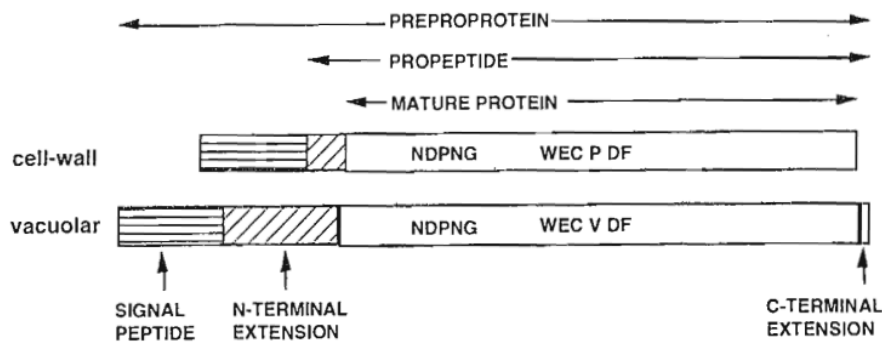


Figure 2.14 Schematic representation of cell-wall bound and vacuolar pre-protein invertases. The pre-protein sequence includes the signal peptide, N- and C-terminal extensions and the mature protein. The peptide sequences NDPNG and WEC P/VDF, respectively, the β -fructosidase motif and the catalytic site (Tymowska-Lalanne and Kreis 1998).

2.4.2 Cell wall invertases

Cell wall invertases are central to phloem uploading in some but not all sucrose importing structures. Cell wall invertases contribute most prominently to the development of pollen, which is symplastically isolated in the developing anthers

(Roitsch *et al.*, 2003). The essential function of Inv-CWs in carbohydrate supply to pollen is evident from antisense repression of Inv-CW Nin88 in tobacco, which is specifically expressed in the tapetum and pollen of tobacco. (Goetz *et al.*, 2001) (Le Roy *et al.*, 2013). Similarly, cold and/or ABA-induced male sterility in rice involves a disruption of sugar transport in anthers, caused by repression of cell wall invertase expression (Oliver *et al.*, 2007).

2.4.3 Cell wall invertase three dimensional structure

CW-INV belong to the family 32 of glycoside hydrolases (GH32; www.cazy.org).

These enzymes use sucrose as the preferential donor and transfer the fructosyl moiety to a variety of acceptors, including (sucrose hydrolysis), sucrose, and fructan (levelan or inulin) (Le Roy *et al.*, 2007). Phylogenetic analysis of the GH32 family in a number of plant species shows (based on protein sequence homology) that the 3-D structure is likely highly conserved, consisting of N-terminal five-bladed propeller domain, followed by a C-terminal domain formed by two β -sheets. The active sites (residues N116, N143, and N299 in Figure 2.15 from Verhaest *et al.*, 2006) where complex formation between invertases and sucrose occurs, are located on the β -propeller domain. The presence of these highly conserved active sites underpin the catalytic efficacy for CW-INV to hydrolyse the glycosidic bond in sucrose, thus cleaving in into substrates glucose and fructose (Verhaest *et al.*, 2006).

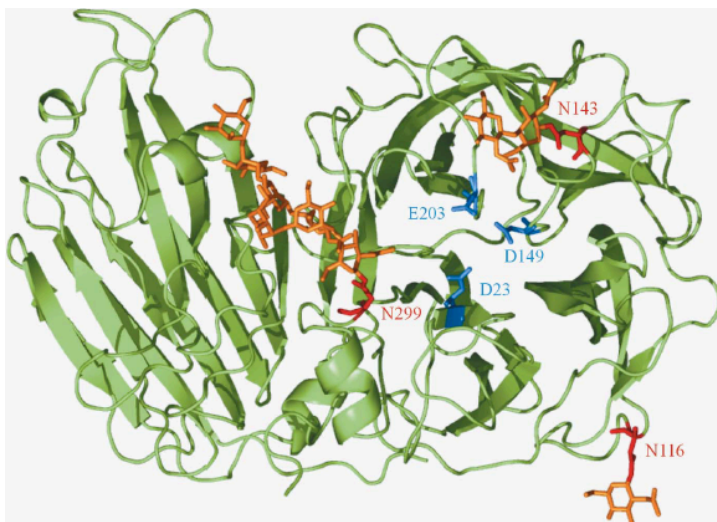


Figure 2.15 Three dimension protein structure of *Arabidopsis thaliana* cell wall invertase (PDB accession: 2AC1). Based on protein modelling three putative glycosylation sites have been determined where cell wall invertase binds to for a complex with sucrose (residues N116, N143, and N299) (Verhaest *et al.*, 2006).

2.5 Factors regulating the expression of cell wall invertases

Invertases play a critical role in plant metabolism and development and thus their expression must be tightly regulated to ensure expression at the appropriate developmental stage and tissue. Regulation of invertases at a transcription level has been widely studied in a range of crop species and has been found to be influenced by a range of signals including; low oxygen, plant hormones, sugars, and environmental and pathogenic cues (Koch 1996; Long *et al.*, 2002; Proels and Roitsch 2009). In studies examining the selective down regulation of cell wall invertase isoforms in what under water deficit condition the authors Koonjul *et al.*, (2005) reported that gene regulation was highly specific to individual isoforms both spatially and temporally. Water deficit applied specifically the period just prior to and during pollen meiosis showed specificity of down regulation for the cell wall invertase gene *IVR1* but not *IVR3* (also a cell wall invertase) or *IVR5* (a vacuolar invertase). These results appear consistent with previous conclusions drawn by Sturm (1999) and Godt and Roitsch (1997). Recent studies of cell wall invertase expression profiles in the developing seed in cotton also showed temporal and spatial specific profiles, however the lack of a reference genome in cotton precluded the authors from discerning isoform specific patterns despite the probable existence of gene homologs due to the polyploid nature of the cotton genome (Wang and Ruan 2012).

Koonjul *et al.*, (2005) concluded that *IVR1* gene expression regulation was also most likely to occur at a transcription level. However, there appears to be conflicting evidence in the literature on this issue with regards to whether regulation is most strongly influenced at a transcription or post transcription level. Greiner *et al.*, (2000) reported that transcriptional regulation of CW-INV was unlikely alone to provide the rapid inhibition required for silencing of enzyme activity. In a review of invertase inhibitors in plants Rausch and Greiner

(2004) concluded that regulation of *IVR1* expression appeared to involve both transcriptional and post transcription mechanisms. At a post transcription level regulation is thought to be attributed to: (a) a shift in the pH above or below the optimal activity threshold for the gene (i.e. 4-5), (b) developmentally regulated proteolytic degradation, and (c) the formation of complexes between CW-INV and proteinaceous invertase inhibitors (Jitsuahara *et al.*, 2002).

Jin *et al.*, (2009) reported that the inhibition CW-INV under stress conditions could have important evolutionary survival implications where by reduction in enzymatic activity under constrained environmental conditions could ensure that each developing microspore would receive only a small amount of carbohydrate sufficient to support fertilization. This prioritization of assimilate allocation would maximize the number of viable microspores capable of undertaking fertilization, and thus increasing the chance for the survival of the species. However, in the cultivation of modern wheat this genetic risk management strategy is not favourable in due the negative implications it has of seed size and seed number.

2.5.1 The role of abscisic acid in reproductive tissues sensing water deficit

The first step in switching on a specific response to water deficit is the perception of the signal by specific receptors (Barnabas *et al.*, 2008). There is a strong link between role of ABA in; plant response to abiotic stress, plant metabolic processes, and development (Hey *et al.*, 2010) (Figure 2.16). ABA has long been recognised as a 'stress hormone' due to its role in controlling plant water dynamics through regulation of stomatal conductance and carbon dioxide uptake (Kim *et al.*, 2010), which it turn induces expression of stress related genes (Shinozaki and Yamaguchi-Shinozaki 2007). In wheat and rice the causal link between abscisic acid accumulation under water deficit conditions in both source and sink tissue and the incidence of male sterility has been investigated in a number of studies (Ji *et al.*, 2011; Oliver *et al.*, 2007; Westgate *et al.*, 1996). Oliver *et al.*, (2007) also reported a similar link exists between ABA production in response to cold stress and pollen sterility. Endogenous ABA application has

been shown to cause expression of a number of genes that are involved in both drought and cold stress (Thomoshow 1999).

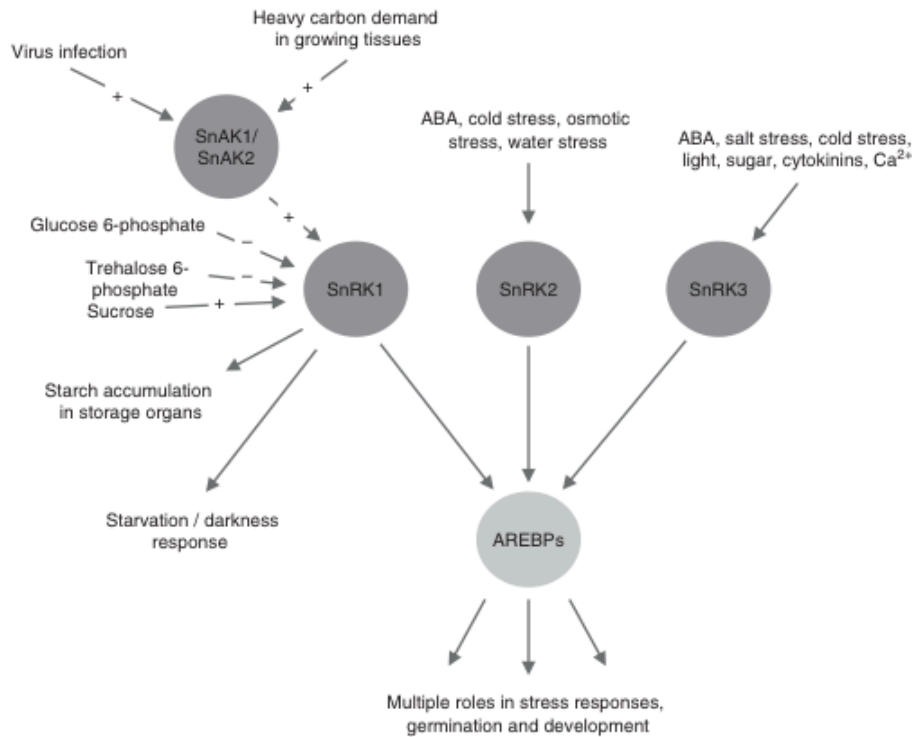


Figure 2.16 ABA response element binding proteins (AREBPs) as hubs at the interface between metabolic and stress signaling networks, potentially targeted for phosphorylation by all three SNF1-related protein kinases (SnRK1 to 3) and with multiple roles in stress responses, germination and development (Hey *et al.*, 2010).

An important finding by Ji *et al.*, (2011) established a clear link between ABA accumulation in rice anthers under cold-stress the repression of the rice CW-INV gene OSINV4. These results gave further strength to previous findings that ABA was strongly involved in sugar signalling (Arenas-Huertero *et al.*, 2000; Arroyo *et al.*, 2003; Rook *et al.*, 2001). The overall conclusion drawn in the Ji *et al.*, (2011) study was that ABA accumulation in the young microspore stage of pollen development in rice resulted in pollen sterility attributed to the down regulation of the CW-INV gene OSINV4, and thus resulted in a reduction of 'sink' strength in the developing anther. A finding from this study with important implications for potential selection of frost and/or drought tolerant rice varieties was that amongst the varieties tested there were clear varietal differences in sensitivity to ABA accumulation in anthers as measured by the degree of pollen

sterility that resulted. The authors reported this difference in varietal susceptibility to ABA may be a result of regulatory mechanisms of ABA catabolism that are tuned differently in tolerant lines.

Despite the clear link being established between cause and effect in these studies they were not able to elucidate the molecular basis of the overlap between ABA and sugar signalling pathways. Potential explanations of this cross-talk between ABA and sugar in the stress-sugar signalling network have been proposed by (Hey *et al.*, 2010) (Figure 2.17); (1) sugar signalling could be directly mediated by ABA; (2) ABA could modulate sugar signalling by priming tissue to respond to sugars; and (3) convergence and cross talk between the two pathways through transcription factors. It has been suggested that SnRK1 (sucrose non-fermenting-1-related protein Kinase-1) may interact closely with ABA in the sugar mediated drought response. Although not yet proven, protein kinases may act on ABA response element binding factors (AREBP's) through phosphorylation (Hey *et al.*, 2010). Evidence to this claim is apparent in findings of from (Umezawa *et al.*, 2004), which found that a SnRK2-type protein kinase in *Arabidopsis* found to control stress responsive gene expression and improve drought tolerance when over-expressed. It is also important to point out that ABA-independent regulatory pathways influencing drought-inducible gene expression also exist ((Yamaguchi-Shinozaki and Shinozaki 2005).

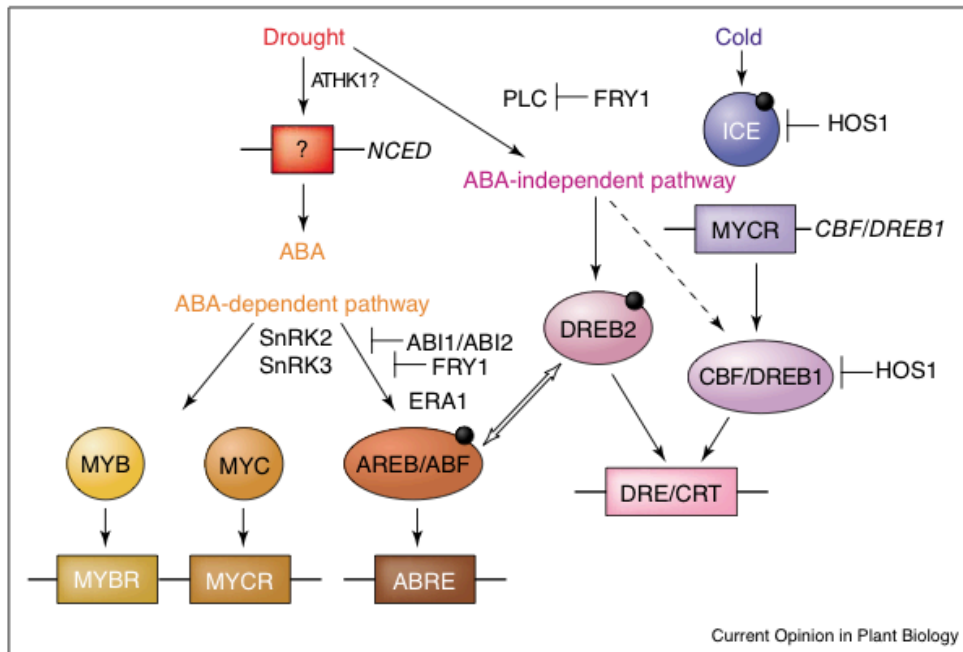


Figure 2.17 Regulatory network of gene expression in response to drought and cold stresses: specificity and crosstalk of gene networks. Cis-acting elements that are involved in stress-responsive transcription are shown in boxes. Transcription factors that control stress-inducible gene expression are shown in circles or ovals. Small shaded circles indicate the modification of transcription factors in response to stress signals for their activation, such as phosphorylation. The upper part of the Figure shows transcription cascades that are involved in rapid and emergency responses to drought and cold stress, such as those involving ICE, DREB2 or NCED. Lower parts of the Figure show transcription cascades that are involved in slow and adaptive processes in stress responses, such as those involving AREB/ABF, MYB, MYC and CBF/DREB1. A two-component histidine kinase 1 (ATHK1) is thought to function as an osmosensor, and phospholipase C (PLC) is thought to function upstream of the DREB2 system (Shinozaki and Yamaguchi-Shinozaki 2000). HOS1 functions as a negative regulator of ICE and CBF/DREB1. FRY1 functions as a negative regulator of drought, cold and ABA responses (Xiong and Ishitani 2006). SnRK2- and SnRK3- like protein kinases are involved in ABA signaling (Guo *et al.*, 2002; Mustilli *et al.*, 2002; Yoshida *et al.*, 2002). ABI1/ABI2 functions as a negative regulator for ABA signaling. A dashed arrow indicates that CBF4/DREB1D are involved in gene expression in response to drought stress (Haake *et al.*, 2002). The open double-headed arrow suggests crosstalk between DREB2 and AREB/ABF that is based on DRE/CRT's acting as a coupling element for ABRE. ERA1, ENHANCED RESPONSE TO ABA1; MYBR, MYB recognition site; MYCR, MYC recognition site (Shinozaki *et al.*, 2003).

2.6 Comparative genomics

Using *Arabidopsis* as a model species for characterising drought inducible gene during microarray studies it was discovered that drought-inducible genes fall into two categories; (1) proteins that function in a drought tolerance capacity, including chaperones, late embryogenesis abundant (LEA proteins), osmotin, antifreeze proteins, mRNA-binding proteins, key enzymes for osmolyte biosynthesis, water channel proteins, sugars, proline transporters, detoxification

enzymes, and proteases; (2) proteins involved in regulating signal transduction and stress –response gene expression, including various transcription factors, protein kinases, phosphatases, and phospholipid metabolism enzymes (Shinozaki *et al.*, 2003). Many of these stress-inducible transcription factors have also been reported in rice, where the same system of classification (i.e. regulatory or functional) (Rabbani *et al.*, 2003). Comparative analysis between the between rice and Arabidopsis for their respective groups of identified drought inducible genes has highlighted that among the 73 genes identified, 51 have also been reported in Arabidopsis to have similar function(Shinozaki and Yamaguchi-Shinozaki 2007). Given the degree of conservation in the number of genes identified and the similarity in functionality these results will likely proved beneficial for similar comparisons between rice and wheat.

2.7 Breeding for reproductive stage drought tolerance

The selection pressures during the evolutionary development of the Poaceae ensure the early landraces of wheat had the genetic potential for survival and reproduction even under sub-optimal growing conditions. However, the intensive selective breeding strategies, which have led to the development of our modern day wheats, have focused on yield under optimal conditions rather than survival under abiotic and biotic stress conditions. (Barnabas *et al.*, 2008).

Breeding for tolerance to drought stress specifically for reproductive stage growth in wheat has had limited success for a variety reasons, most notably the size and polyploid nature of the wheat genome, the quantitative nature of drought tolerance as a trait, and the increasing variability in climatic conditions as a result climate change (Collins *et al.*, 2008; Fleury *et al.*, 2010). At a technical level there is also evidence to suggest lack of accurate and efficient high throughput germplasm screening methods for evaluating sensitivity and/or tolerance to drought has also contributed to the slow progress on advancing this trait in new varieties (Dolferus *et al.*, 2013; Ji *et al.*, 2010). The commercial imperatives driving private wheat breeding programs which favour selection of germplasm for high yield under optimal growing conditions also hampers efforts to introgress genes which confer yield stability to drought and other abiotic

stress conditions (Cattivelli *et al.*, 2008). Recent advances in genomics platforms and technologies to identify key genes underlying drought tolerance QLT's and molecular markers to track genes through breeding populations there after are in detail in the following section.

2.7.1 Plant genetic improvement

This text was taken from Diepeveen D Webster H and Appels R (see Diepeveen *et al.*, (in press)). Key words and Figure 1. from the original publication have been placed in Appendix I to conserve space in this chapter.

The integration of biotechnology concepts and technologies are now accepted as part of an overall strategy in plant breeding that needs to consider water conservation, soil fertility, resistance of domesticated plants and animals to pests and diseases, the production of quality grain from resilient, productive, and adapted varieties. Solutions involve on the one hand innovation and improved management of resources, best practices, so that scale of effort is proportionate to need and opportunity. Complementing this is the need to capture biotechnological breakthroughs that can radically change the phenotypes of domesticated plants and animals for maximum performance in specified environments. Speed in delivering new genotypes for planting is crucial and marker assisted selection applications provide more informative diagnostic probes for identifying favourable combinations of alleles and genome segments in new germplasm. The multidisciplinary approaches that now characterize crop improvement enable a holistic systems approach to the study of expression and regulation of genes (and gene networks) at various levels of the plant (i.e. organism, tissue, organ and cellular level) and developmental stages, in order to identify key cues in the plant environment which influence the final phenotype of the plant.

Initiatives to generate reference genome sequences for rice, barley and maize have provided the basis for an increasing number of structural/functional studies cataloguing and characterizing major genes underlying important agronomic traits. Completion of the reference sequence in rice heralded the

beginning of a new era (the post genome era) in which a suite of complementary 'omics' disciplines (i.e. transcriptomics, proteomics, metabolomics) can now exploit and leverage the knowledge of the full genome sequence to discern at greater resolution than ever before, the complex regulatory mechanisms determining expression and function. For wheat, the genome sequence representation in relation to high resolution genetic maps is still in progress and will be based on a chromosome-by-chromosome analysis (Feuillet *et al.*, 2012).

As crop researchers transition into the 'post genome' era, the expanded knowledge base conferred by having sequenced reference genomes will facilitate improved efficiency for delivering crop genetic gains by way of more targeted research that is better equipped to accurately identify, evaluate and quantify: (1) the genetic potential held in crop genomes (and or sub species/population) which is currently under-utilized, in as yet uncharacterized gene(s) and/or unidentified alleles, that confer to plants the ability to thrive in rapidly changing environmental conditions caused by climate change (and the emerging biotic and abiotic stresses accompany these changes), and (2) the relative level of plasticity of particular crop genomes to evolve and thrive in response to changing environmental stresses, and the extent to which favourable heritable mutations and/or epigenetic changes may influence this adaptive evolution.

Parallel and ongoing research for the advancement of genomics platforms, sequencing technologies, and molecular marker development will continue to be essential in providing the molecular tools to capture and introgress existing and evolving allelic diversity for new variety development. Wide crossing and genetic transformation programs are currently the main sources of new genetic variation. In the case of an emerging threat to wheat in the form of the stem rust Ug99, for example, a basic strategy is to build a combination of genes into a new source of resistant phenotypes which may include unknown additive genes to achieve long-term resistance (Singh *et al.*, 2006). This process can be accelerated by new insights into the molecular basis for disease resistance. Similarly, increased efficiencies in photosynthesis and water use, as well as the incorporation of nutritional attributes, also require more fundamental changes

to the genomes of mainstream cereals in order to provide grain for future generations. An important variable in dealing with a polyploid crop such as wheat is that it shows homoeologous gene silencing (Bottley and Koebner 2008; Mochida *et al.*, 2003; Zhang *et al.*, 2008(A)). Bottley and Koebner (2008) identified homoeolog non-expression for 15 single-copy genes across a panel of 16 wheat varieties and found that for the expression profiles of eight genes only two varieties shared the same pattern of silencing. Homoeologous gene silencing can be tissue specific and is most likely a significant factor in generating phenotypic variation.

Plant breeding in the post-genome sequencing era is developing capacity to utilize the very large databases of molecular markers linked to key agronomic traits. Genome wide association studies (GWAS) in rice, maize and barley have provided high density single nucleotide polymorphisms (SNPs) coverage of genome regions controlling components of yield. For wheat the high density SNP technology is also in place (Cavanagh *et al.*, 2013). The genome sequences available for cereals and the parallel functional (mutation and transformation) studies in rice and *Arabidopsis*, have provided examples of successful “walking” to a gene of interest. The challenge for plant genetic improvement resides in combining genetic gain and improvement in management practises for increasing yield, with the targeting of food crops to particular food products. Changing environments in soil fertility and climate need to be considered as classical G x E (genotype by environment) interactions, but in the context of changes in field and trait-management variables. Although historically improvements have been achieved through plant breeding based on crosses between parents that show attributes of interest, more recent developments have, as noted above, considered targeted introductions of genes through transformation technologies.

The releases of contemporary varieties now generally describe the integration of molecular markers (the outputs from genomics) into the selection technologies utilized for both the selection of parents as well as the selection of progeny from crosses to carry forward in the breeding scheme (Miedaner and Korzun 2012;

UnSang *et al.*, 2011). Although the use of molecular markers provides fast results it is critical that the linkage between the molecular marker and the trait of interest is reliable. In the post-genome sequence era the availability of molecular markers tends to be no longer limiting and this has turned the focus on improving the accuracy of assessing the phenotypes that are important for success of the breeding program.

2.7.2 The phenotype and environment

A phenotype of central importance in crop breeding is yield stability, both in terms of total yield (underpinned by grain number and size) and consistency of quality attributes specific to end use and classification for processing. This is particularly important as climate change introduces conditions not previously experienced. Wheat grain yield has many components including whole plant attributes for optimal establishment and maturity/height features that in turn confer adaptive traits better enabling the plant to be suited to the environment in which it is grown. The supply of sugars to immature pollen in the developing head, and thereafter maturing grain during grain fill, has been identified as a critical plant metabolic phase determining grain yield and size, a result which has been consistently reported in studies on tomato (Ruan 2012), cereals (Dolferus *et al.*, 2013), and tobacco/Arabidopsis (Le Roy *et al.*, 2013). Abiotic stresses (particularly water deficiency and frost) affect the developing spike and maturing grain due to the down regulation of key carbohydrate mobilization and cleavage genes (cell wall invertases and fructosyl transferases and exohydrolases). Tolerance at the genetic level requires expression of these genes under stress conditions to be maintained so that the developing pollen and grain receive sufficient energy supplies for development to reach completion unimpeded.

Although components of yield can be readily defined, variation in the measured phenotype remains a significant factor in breeding programs. Breeding design must be sufficiently flexible to allow changes in crossing decisions to be made when instances of negative interactions between genotypes and environments arise. Designing trials and activities in a systematic way allows optimisation of

information to identify the genotypic basis of phenotypic differences in germplasm and the selection of lines for use in the crosses. The process by which these activities are structured requires the identification of parents for crosses, allocation into a design, data collection, and then the analysis of data from derived progeny. The science of trial design and the optimisation of crossing/mating information is extensive (Kempton 1982; Singh and Hinkelmann 1998; Smith *et al.*, 2006; Williams *et al.*, 2011) and can be simulated using tools such as QU-Gene (www.uq.edu.au/lcafs/qugene).

Important to information collection is the concept that the breeding process involves the generation of genetic variation, followed by the selection of elite germplasm, and the characterisation of elite lines for desired traits. This process has the potential for errors to be transferred into consecutive stages of the breeding process. Developing trial designs such as “chained trial design” that minimise type-II errors (false-positives and true-negatives) are becoming important considerations for breeding. The chained designs cover multiple testing of a plot (or grain sample) by incorporating within and between replication in each step tests to allow variance estimates. The statistical methods utilized are based on mixed-model restricted likelihood (REML).

With increasing amounts and complexity of information being generated from breeding operations, breeders need to collate this information into a format for decision making. Decision making methodologies has been well researched (Csáki 1985; Dams and Hunt 1977; Saaty 2005) but less often applied to plant breeding (Timmermann 2006). The concept of a data matrix formed by taking germplasm lines and matching each line with the results for the plant breeder to interpret is widely used. Decision matrices use a weighing or prioritising process that allows the combining of multiple results into scores to enable data reduction and easier decision making. Statistical approaches which calculate estimated breeding values (EBVs) for a trait associated with each line are designed to reduce environmental and/or extraneous factors and thus greatly enhance the decision matrix process. Diepeveen (2011) demonstrated this process for grain quality traits in a wheat breeding program.

The incorporation of marker/genomic information into the decision matrix process distorts the decision matrix since inputs of this type comprise data points that are highly correlated with each other when they are genetically closely linked to each other (see Figure 1 and the associated box). The incorporation of correlated data structures within a multivariate REML analyses is one tool that can be applied for data reduction of genomic information and the subsequent estimation of EBV's for decision making.

2.7.3 Comparative phenomics

A major effort to capture plant phenotype descriptions in a standardized way is being undertaken through efforts such as the Plant Ontology project in Gramene (Ni *et al.*, 2009) and Crop Ontology (Shrestha *et al.*, 2012), based on the premise that essential features of traits can be identified across a wide range of species. For example the Zadoks scale is widely used to describe stages in plant development and it has been incorporated into the Biologische Bundesanstalt Bundessortenamt and Chemical industry (BBCH) scale (Meier 2001). The BBCH scale provides the basis for describing equivalent stages in a range of cereals which can then be linked through a website such as Gramene (Ni *et al.*, 2009). A particular advantage for cross-referencing phenotypes of diverse but related species is that advances made in well characterised species can be more easily applied to complex species such as wheat, thus facilitating meaningful comparisons.

Plant phenotypes including aspects of maturity such as earliness *per se*, vernalization responses (Vrn), photoperiod response (Ppd) and height phenotypes (Rht genes), efficient plant establishment based on root growth, early vigour in vegetative growth, cold tolerance in some environments, tillering and leaf expansion constitute important features of the overall plant that have been especially well characterized at the gene-level through comparative genomics and phenomics (Feuillet *et al.*, 2012). Abiotic stress tolerance in breeding is generally environment-specific and can range from being simple to complex at a genetic level. The study of boron and aluminium tolerance

(Delhaize *et al.*, 2012; Feuillet *et al.*, 2012), and water deficit stress (preceding section) in crops such as wheat have benefited from analyses of these traits in a range of related plant species.

Barrero *et al.*, (2011) provided a detailed analysis of the potential value gained from utilizing comparative genomics as a selection tool in the preliminary stages of the candidate gene identification pipeline, in which gene structure and function information accumulated on well studied species (particularly rice) can be transferred to complex polyploid species such as wheat. Using the grain weight QTL called QTgw.ipk-7D QTL as an example, this trait mapped to the most telomeric bin (7DS4-0.61-1.00) in the physical map of wheat chromosome arm 7DS. Barrero *et al.*, (2011) estimated that a total of 173 annotated wheat EST loci were located in the wheat chromosome bin 7DS4- 0.61-1.00 (www.wheat.pw.usda.gov/cgi-bin/westql/map_locus.cgi). These wheat ESTs were then assigned positions on the rice genome and those ESTs uniquely mapping to rice chromosome 6, which is the syntenic region to wheat chromosome 7DS, were identified for further analysis. This process identified 23 wheat ESTs aligned to rice chromosome 6 from position 369,919 to 25,106,896 bp. To further narrow this 25 Mbp region, Barrero *et al.*, (2011) screened the rice genome for overlapping grain weight QTLs and identified a rice qrGW-6 QTL flanked by the G200 and C235 molecular markers that were located at 50.1 and 53.5 cM on the rice chromosome 6, respectively. The genetic markers overlapping this region from the Rice Genome Annotation database (www.rice.plantbiology.msu.edu) resulted in 11 rice genetic markers that mapped specifically to rice chromosome 6 from 6,978,639 to 9,153,520 bp. In this way it was possible for Barrero *et al.*, (2011) to reduce a 25 Mbp region for analysis down to a 2.1 Mbp region, in order to search for possible grain weight candidate genes. These candidates include genes in a diverse class of E3 ubiquitin ligases involved in protein degradation. The GW2 gene for grain weight is in this class of proteins. Also included in this region was an E2FE transcription factor which negatively regulates cell division.

2.7.4 Changing agendas to meet new challenges and conclusions

Future trends in plant genetic improvement will tend to focus on utilizing molecular biological techniques to consolidate pyramiding gene-networks for new purposes/end-products, combined with yield gains. While new phenotypes such as bioethanol production from specialized pastures and biofactories for speciality products will continue to emerge, the basic concepts for tackling the new challenges will tend to follow the approaches outlined in this article. Transformation technologies have significant contributions to make in defining the functions of gene networks and providing a potential basis for new varieties when suitable government legislation/policy can be developed. The genetic engineering of crop plants to produce bio-factories in controlled environments is more likely to receive increasing attention in the future for producing high value molecules such as vaccines (Huang *et al.*, 2010; Phoolcharoen *et al.*, 2011).

The genomic resources of grain crops can now contribute to refining molecular marker development and mapping strategies for increasing the efficiency of both the breeding processes and the utilization of diverse germplasm resources. Harnessing the genetic potential of novel allelic combinations in diverse germplasm resources will require breakthroughs in disrupting linkage blocks and/or altering recombination rates along the length of the chromosome (Higgins 2012; Saintenac *et al.*, 2011; Yousafzai *et al.*, 2010).

The main aim of researchers in the 'post-genome era' is to exploit data from crop genome sequences, using complimentary 'omics' tools, to continue delivering genetic gains at a rate that meets increasing global demand for grain commodities and provides yield stability to withstand climate change stresses. The success in obtaining reference genome sequences for broad acre crops is providing new high throughput technologies for assaying variation in combinations of genes that was not previously possible. In particular it is providing the basis for improved understanding of mechanisms regulating the epigenome (post-replication modification of chromatin) which holds promise for delivering crop genetic gains via strategically targeting new adaptive alleles.

Chapter 3: General materials and methods

3.1 Plant materials and growth in the glasshouse

Plant material was grown out to maturity in the Murdoch University PC2 glasshouse facilities four times over three consecutive years. The purpose of growing plant material varied, including: (1) phenotyping the Westonia x Kauz double haploid (WxK DH) population (225 lines) and the respective parents, Westonia and Kauz (April-October 2010), (2) seed multiplication of the individual lines in the WxK DH population (October 2010-April 2011), and (3) two reproductive stage water deficit (RSWD) stress experiments for gene expression analysis on a subset of WxK DH lines and the respective parents (May-November 2011 and June-December 2012).

Uniform protocols were conducted in preparation for, and management of, plant growth intervals between 2010-2013, for parameters relating to: pre-germination seed treatment, preparation of the plant growth medium (soil), and control of environmental conditions in the PC2 glasshouse facilities for the duration of the plant growth intervals. The specifications of these generic protocols were:

(1) pre-germination seed treatment for fourteen days to promote seed imbibement and subsequent coleoptile emergence, the conditions of which involved light exclusion, chilling at 4⁰C, and soaking in water. At the end of the fourteen day treatment germinated seed was sown into individual pots housed in the PC2 glasshouse facilities (containing a uniformly prepared sandy loam soil growth medium) to a depth of 2cm.

(2) uniform preparation of soil such that the final homogenous mix contained equally dispersed constituents, including; two parts loam, one part river sand and 3g of basal fertiliser (mgKg⁻¹): 1220N, 358P, 819K). Upon homogenisation the soil mix was transferred into individual free draining pots and steam pasteurized.

(3) control of temperature and relative humidity in the PC2 glasshouse facilities within a threshold range. The average upper and lower threshold values for

temperature and relative humidity respectively were 15°C (night) -25°C (day) and 80%(night)-60%(day). The PC2 glasshouse facilities lacked the necessary infrastructure to control day length. Hence, seasonal differences in planting dates for the various plant growth intervals (and their associated day lengths) resulted in variation for this parameter. The control of irrigation scheduling was carried out using an irrigation mat system, where by individual free draining pots were watered three times daily to maintain average soil volumetric water content of 60% by way of capillary action.

3.1.1 Selection of population structure and variety cross

A double haploid (DH) wheat population was chosen on the basis that all progeny in this type of cross are true breeding (i.e. homozygous at every locus) which enables more accurate evaluation of genotype to phenotype associations, and thus identification of specific loci controlling the quantitative trait of interest (Semagn *et al.*, 2010). Given effects from individual loci controlling the trait of interest in this study, RSWD stress tolerance, are often small and highly influenced by environmental factors, accurate dissection of RSWD stress tolerance QTL's requires precise phenotype data. Hence, the homozygosity of DH lines and the variation in phenotypes often expressed within this type of population would also confer greater accuracy in the RSWD stress QTL analysis process through robust phenotypic data collection.

The DH population chosen was from a cross between parents Westonia and Kauz, the selection of which was based on the fact both parents are well suited to the low rainfall minimum fertility soil conditions of the South West Australian wheat growing region. Westonia and Kauz are both classified as drought tolerant, an adaptability which is underpinned by the capacity to maintain high concentrations of stem carbohydrate (relative to the total dry weight of the plants) during flowering (Zhang *et al.*, 2009(A)). The experimentally determined stem carbohydrate content level for Westonia and Kauz is $40 \pm 5\%$ and $\approx 40\%$ respectively (Conocono 2002; Rajaram *et al.*, 2002).

Westonia (pedigree: CO-1190-203/84-W-127-501) is a variety that was developed specifically for Western Australian growing conditions and is classified as an Australian premium white (APW) spring wheat and has an erect growth habit, medium height, and early maturity (Australia 2013). Kauz (pedigree: Jupateco F 73/Bluejay/UresT81) is a variety developed by the International Maize and Wheat Improvement Centre (CIMMYT, EI Batan, Mexico) (Butler *et al.*, 2005) was bred for adaptability for growth in localities where low to medium rainfall conditions predominate (CIMMYT 2013). Seed stocks for the Westonia x Kauz double haploid (WxK DH) population were provided by the Department of Agriculture and Food, Western Australia (DAFWA).

3.2 Characterizing the WxK DH population

Genotypic and phenotypic characterisation was carried out on 225 lines of the WxK DH population and the respective parents, Westonia and Kauz (Zhang *et al.*, 2009(A)). Phenotype data was collected for: (i) evaluation of variation amongst the WxK DH lines in phenology and morphology, (ii) benchmarking the growth stage at which pollen meiosis occurs. Genotypic characterization of the WxK DH's was also undertaken to evaluate the pattern of allelic inheritance from the DH parents.

3.2.1 Phenotyping – phenology and morphology

The unit of measurement for phenological observations was the number of growth days (days after sowing, DAS) a DH line took to reach key growth stages. DAS data was collected for: auricle distance at 5cm (AD5), boot swollen (BS), awns first visible (AFV), full head emergence (FHE), and anthesis commenced (AC). DAS growth stages were also cross referenced against the Zadoks decimal growth scale for cereals (Zadoks *et al.*, 1974), and assigned the appropriate scale value: AD5= Z43, BS=Z45, AFV= Z49, FHE=Z59, AC= Z60. According to the DAS interval length between benchmarked growth stages individual WxK DH lines and the DH parents were assigned to maturity groups (Table 3.1). Categorical groupings were also made for morphological parameters, including: plant height (PH) at maturity (Z91), leaf waxiness (cuticle thickness on vegetative growth e.g. flag leaf and stem), and spike characteristics (Table 3.2).

Table 3.1 Phenological categories of the WxK DH population for days taken to reach key growth stages.

Maturity class	Days after sowing (DAS) taken to reach benchmarked growth stages				
	Auricle distance 5cm (AD5)	Boot swollen (BS)	Awns first visible (AFV)	Full head emerged (FHE)	Anthesis commenced (AC)
Early	42-60	44-61	44-62	45-63	45-64
Mid	61-80	62-81	63-82	64-83	65-84
Late	81-100	82-101	83-102	84-103	85-104
Extra late	101-140	102-141	103-142	104-143	105-145

Table 3.2 Morphological categories of the WxK DH population

Plant height	Spike characteristics	Leaf waxiness
Super dwarf (35-60cm)	Spikelet number per row (7 or 9)	Cuticle absent
Dwarf (61-80cm)	Floret number per spikelet (3 or 4)	Cuticle sparse
Tall (81-100cm)	-	Cuticle medium
Super tall (101-144cm)	-	Cuticle dense

Field based phenotyping was also undertaken on the 225 WxK DH lines and the respective parents (Westonia and Kauz) at the DAFWA Katanning research station ((33.7° S, 117.6° E) where the material was being grown out to maturity for seed multiplication in 2010 (June-November). Phenotypic measurements observed included plant height and leaf waxiness, and categorised as per outlined in Table 3.2.

3.2.2 Phenotyping - cytology

Cytological evaluations were undertaken for two purposes: (1) as part of the experimental design process, or (2) evaluating morphological changes in reproductive structures during glasshouse experiments, in response to plus and minus RSWD stress conditions. Cytology undertaken for experimental design was necessary due to the fact pollen development during stem elongation is a physiological process that cannot be visualised, and its temporal onset shows strong genetic variation that is variety specific. For each of the experimental lines selected it was therefore necessary to determine an auricle distance length at which pollen development (specifically meiosis) could be assumed to be occurring, thus enabling it to be used as a visual marker (See Appendix II, Figure 8.2 for an example using WxK DH line D02-105).

Cytological evaluation of pollen development was undertaken at 2cm increments of auricle distance extension, ranging from an auricle distance 2cm (AD2;40) up until visualization of awns first emerging from the leaf sheath (AFV; Z59).

Developing spikes from the primary tiller of plants were cut from the stem and dissected out from the surrounding leaf sheaf. Spike samples were fixed in Carnoys fixative (ethanol: glacial acetic acid, 3:1) at room temperature for four hours and stored at -20°C using either: (i) 3% aceto carmine, and (ii) DAPI (DAPI-4',6' diaminino-2-phenylindole-2HCl; product number D9524; Sigma Australia).

Anthers (all three per floret) were excised from the fixed spikes using a high resolution dissecting microscope at x40 magnification from the primary floret (F-1) of spikelets located in the mid region of the spike (rachis position S-3 to S-5) and placed on a microscope slide. Development of anthers and ovum within the spike is asynchronous (Sibony and Pinthus 1987), where by spikelets in the mid-section of the rachis (S-4 to S-6) develop earlier than the upper (S-7 to S-9) and lower spikelets (S-1 to S-3) and the lateral florets (F-1 and F-2) developing faster than medial florets (F-3, F-4, and F-5). Because of this asynchronous pattern of development the same 'representative' spikelets and florets had to be sampled in all heads. See Figure 3.1 below for explanation of spikelet and floret positional naming scheme. 7µl of DAPI solution (1ug/ml DAPI in a buffer with 50% glycerol and 10mmol/L citrate, pH 4.5) was pipetted onto the slide so as to engulf the anther and a cover slip was placed on the sample. To enable optimal penetration of the DAPI into the anthers microspore mother cells (MMCs) a tissue squash was performed by gently tapping the cover slip so as to release the encapsulated MMC's into the surrounding DAPI solution. The prepared slides were left to incubate for 10 minutes in the dark at room temperature and then examined.

The tissue staining protocol utilizing aceto-carmine was congruent to that followed for DAPI until completion of the anther squash step, where upon the aceto-carmine stained samples were gently heated over a low flame for 3 seconds to promote evaporation of excess ethanol. Observation of stained MMC's

was carried out using an Olympus BX51 Photomicroscope with a Panasonic Colour CCTV camera attached (Model: WV-CL830/G) at 40x magnification, utilizing fluorescence (Brightfield) for the DAPI stained samples and differential interference contrast (Normanski) for those stained with aceto carmine. Evaluation of pollen cell division stage was undertaken using published pollen cell division stage images generated from studies conducted by Zhang and Wilson (2009) as a reference to benchmark against. Cell division stage was recorded relative to the auricle distance stage at which the sample was harvested and temporal variations in onset of cell division stage between the experimental lines selected was also recorded.

Mature pollen was collected from the spike of the primary tiller of plants at the onset of anthesis, by excising anthers from florets located in the middle region of the spike with forceps. Anthers were transferred to 2ml microcentrifuge tubes, preserved in 500 μ l of 70% ethanol (per sample), and stored at 4^oC. Prior to staining the anther-ethanol suspension was agitated briefly by hand inversion to dislodge mature pollen from the anther sac, where upon 5 μ l of pollen-ethanol suspension was pipetted onto a microscope slide, to which 12 μ l of potassium iodide iodine stain solution (KI/I₂; 2.5gKI, 250mg I₂, 125ml H₂O) was finally added.

Observations and photography of potassium iodide iodine stained pollen was conducted using the same microscope and camera apparatus as was used for pollen cell division staging (i.e. DAPI and aceto carmine). Observations were recorded per transect of the slide until 100 pollen grains had been observed. The viability status of pollen was assessed on the basis of the extent of starch deposition per pollen grain (\geq 75% fully fertile (optimal for fertilization), 50-74% moderate fertile (sub-optimal for fertilization), 25-49% partial fertility (critically low significantly impeding fertilization), 0-25% complete sterility (fertilization completely impeded)).

General observations of morphological changes during anther development were recorded for plants in both the water deficit and control treatments during the

glasshouse RSWD stress experiments. Morphological changes were observed on fresh non-fixed tissue under 40x magnification, using a high resolution dissecting microscope (dissection protocol as per describe above). Changes in anther length and tissue colour were noted at the key auricle distance growth stages (AD5, AD10, AD15), full head emerged (FHE), and anthesis commenced (AC). Statistical testing was performed to evaluate correlations between these variables and, phenotypic growth stage and pollen cell division cycle.

3.2.3 Pollen sterility

Pollen sterility, as an inverse measure of grain set (i.e. the presence of a formed grain in a floret), was observed for all florets on the primary spike of mature wheat plants and was calculated according to Equation 3.1.

$$\text{Sterility (\%)} = \left[\frac{\text{number of florets per spike} - \text{number of grains per spike}}{\text{Number of florets per spike}} \right] \times 100$$

Equation 3.1 Calculating pollen sterility as a measure of grain set.

When examining each floret for the absence/presence of grain both its vertical and horizontal coordinates within the spike were recorded (i.e. on the spike and within the spikelet respectively) (Figure 3.1). Categorisation of grain presence/absence on the basis of floret position was necessary for downstream statistical testing to probe for potential significant differences in spatial distribution (vertical and horizontal) of grain set within the spike, in addition to the standard total grain set per spike measurement.

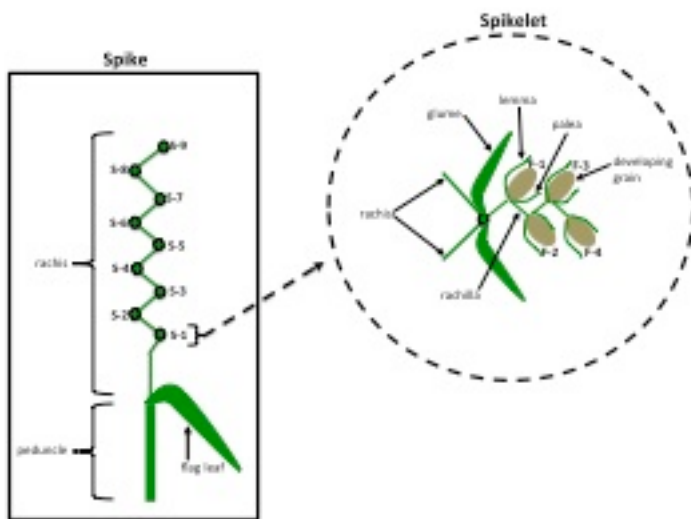


Figure 3.1 Naming scheme for reproductive organs of spike inflorescence. Spikelet position is preceded by the letter S and numbered acropetally along the rachis. Spikelet vertical position was subject to further classification namely: Lower spike = S-1 and S-2; Middle spike = S-3, S-4, and S-5; and Upper spike = S-7 and S-9. Floret position within the spikelet is denoted by the letter F and ordered acropetally along the rachillas where; F-1= primary floret, F-2 = secondary floret, F-3= tertiary floret, and F-4= quaternary floret (modified from (Wilhelm and McMaster 1996)).

3.2.4 Genotyping and selection of WxK lines for reproductive stage water deficit stress testing

A subset of lines in the WxK DH population was selected for inclusion in glasshouse RSWD stress experiment, the selections of which were made on the basis of data accumulated from extensive molecular and phenotype characterization of the WxK DH lines. The primary objective in selecting individual DH lines was to choose contrasting genotypes that would enable evaluation of genotype-phenotype associations. Selection of lines also had to take into account that the WxK DH population has large variation in plant height and flowering time ($\cong 100\text{cm}$ and $\cong 100\text{day}$ respectively) (Table 3.2). It was therefore necessary to reduce potential confounding effects in plant response to +RSWD stress and -RSWD stress (control), attributable to large differences in these genetic variables, by selecting only plants belonging to a single categorical grouping for plant height and flowering time.

Selection of WxK lines for glasshouse experiments involved a four stage process (Figure 3.2); **Stage 1**- assigning WxK lines to categorical groups based on number

of DAS taken to reach the AC (anthesis commenced) growth stage and plant height of the lines at maturity (Table 3.2). Lines belonging to the Early, Mid, and Extra late 'DAS to AC' group were excluded, as were the lines in the PH (plant height) group 'Dwarf', 'Super dwarf', and 'Super tall'. Lines retained for further rounds of selection belonged to the 'Mid-AC and Tall-PH' group. **Stage two** – the genotype data from the marker saturated ((199 SSR's, 91 829 SNP's) WxK DH genetic map (Zhang *et al.*, 2013) was mined to identify WxK DH lines showing either a strong parental genotypic likeness (i.e. $\geq 60\%$ of alleles inherited from parent one and $\leq 40\%$ from parent two) or lack thereof (i.e. % alleles inherited from parent one \cong % alleles inherited from parent two). WxK DH lines showing alleles inherited in equivalent numbers from both parents were ruled out for further selection. **Stage 3** – involved identification of lines belonging specifically to the Mid-maturity DAS to growth stage auricle distance 5cm in length (AD5) and subsequent exclusion of all lines falling into other maturity classes for this growth stage. The maturity categories (i.e. Early, Mid, Late, Extra late) for the AD5cm growth stage had an average 24 day threshold range and distribution within this interval followed a normal curve (data not shown). Subsequent selection also had to take into account that within this group two lines should only be selected if the difference for DAS to AD5 was no greater than seven days (Table 3.3). **Stage 4** – Of the remaining lines eligible for selection one Westonia-like (i.e. $\geq 60\%$ of alleles were inherited from Westonia) and one Kauz-like (i.e. $\geq 60\%$ of alleles were inherited from Kauz) DH was selected.

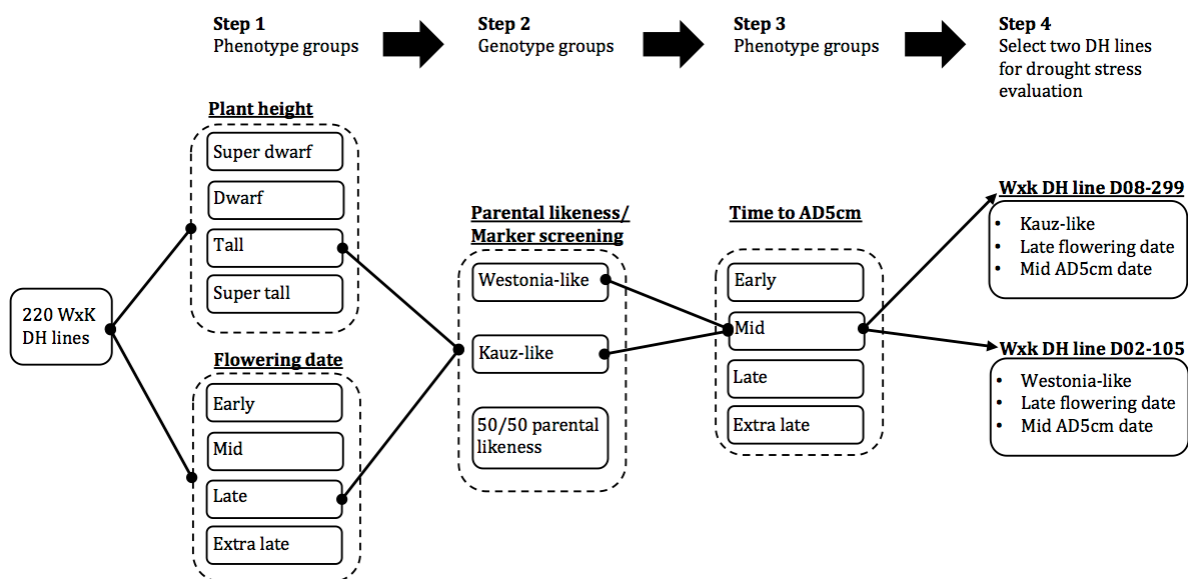


Figure 3.2 Steps involved in the selection of WxK DH lines for inclusion in the RSWD stress experiments. Change peduncle to auricle distance

Table 3.3 Summary of categorical grouping data for the WxK lines selected for RSWD stress experiment testing.

	DAS to auricle distance 5cm (AD5)	DAS to anthesis commenced (AC)	Plant height at maturity (cm)	Allelic inheritance pattern
D02-105	55	74	83	Kauz predominant (70%)
D08-299	48	63	86	Westonia predominant (65%)

3.3 Experimental design

The experiment was designed to enable gene expression analysis in response to RSWD stress. The experiment design was 4x4 factorial (Table 3.4) with three biological replicates.

Table 3.4 Factors and levels in the RSWD stress glasshouse experiments

Factor		Level	
(A)	Treatment	(i)	-RSWD stress (control)
		(ii)	+RSWD stress
(B)	Variety	(i)	Westonia
		(ii)	Kauz
		(iii)	WxK DH line D08-299
		(iv)	WxK DH line D02-105
(C)	Growth stage *	(i)	Auricle distance 5cm
		(ii)	Auricle distance 10cm
		(iii)	Auricle distance 15cm
		(iv)	Full head emerged
(D)	Tissue type	(i)	Spike
		(ii)	Leaf

* Using an aceto-carmin stain method it was found that pollen meiosis most often occurred between auricle distance 5-8cm in all four lines tested.

3.3.1 Water deficit treatment

The method of water deficit (drought) treatment used exclusion of water and was attained by placement of potted plants in vermiculite (a medium that increases the rate of soil dehydration) filled plastic tubs. The drought treatment was designed to impose water deficit stress on plants, such that maximum stress was incurred during reproductive stage development, specifically pollen meiosis. A leaf water potential of -10 to -14 Bar is equivalent to mild to moderate water deficit stress in the field and was used as the benchmark stress level necessary to be achieved by the onset of the pollen meiosis growth stage (experimentally determined in the current work to be equivalence to auricle distance length 5cm (AD5)). It was necessary to perform pre-experiment testing to determine the number of days it would take for leaf water potential of well watered plants, once removed from the irrigation mats and placed in the vermiculite medium, to drop to -10 to -14Bar. These results then enabled the calibration of the number of growth days (prior to auricle distance 5cm) the plants would need to have the RSWD stress imposed (\approx 4-5 days). The duration of the RSWD stress treatment was designed to last ten consecutive days so that its culmination on average occurred 1-3days prior to anthesis.

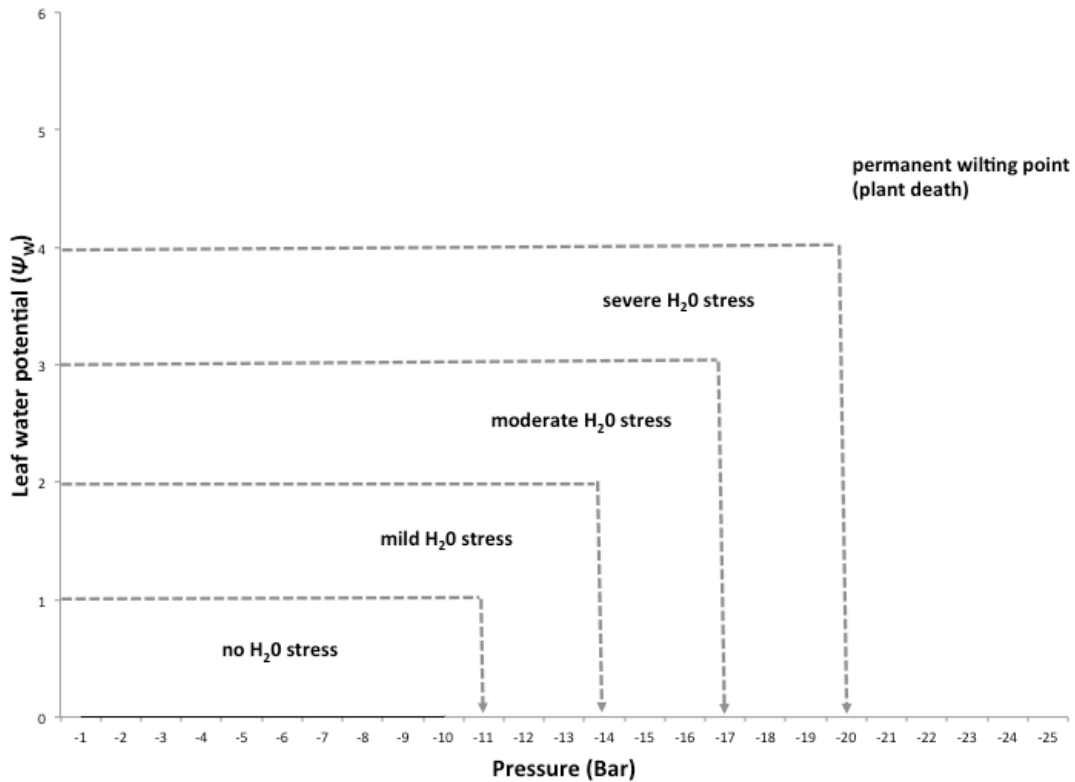


Figure 3.3 The relationship between leaf water potential and water stress intensity (Meron *et al.*, 1987).

3.3.2 Soil volumetric water content (SVMC)

The soil volumetric water content was measured daily between 7.00-8.00am in all experimental pots to ensure:

- (1) the rate of soil moisture loss in the drought pots was consistent with pre-determined daily loss rates ($\approx 6\%$) which would facilitate the necessary drop in leaf water potential after four days under +RSWD stress conditions (i.e. water withheld and pots submerged in vermiculate) which was also timed to coincide with the onset of pollen meiosis (growth stage AD5). See below section 'Plant water status' for threshold plant water potential values conferring water stress.
- (2) the SVMC of the -RSWD stress pots was maintained within a threshold range 21-24% (consistent with optimal plant available water levels for plant root systems growing in free draining pots) and mitigating effects from water stress or water logging.

Measurement of soil volumetric water content was undertaken using a HydroSense Soil Water measurement system (CD620 display unit and CS620

water content reflectometer) fitted with 12cm probes (accuracy +/-3% volumetric water content with conductivity <2dSm⁻¹). Campbell Scientific Australia (CSA) Pty. Ltd., Thurlington Central QLD Australia. Measurements were taken by fully inserting the probe rods into the potted soil and engaging the HydroSense Soil Water measurement system in the read mode. The HydroSense Soil Water measurement system is sensitive to changes on dielectric permittivity and the probe had a resolution >0.1%. The calibration coefficients used by the probes operating system to transform the measurement signal to SVMC were experimentally determined by the manufacturers using Equation 3.2, and performed on range of agricultural soil types, including sandy loam which was the soil medium used in the current work.

$$\theta_v = \frac{V_{\text{water}}}{V_{\text{soil}}} = \frac{[M_{\text{water}}/\rho_{\text{water}}]}{[M_{\text{soil}}/\rho_{\text{soil}}]} = \frac{\theta_g * \rho_{\text{soil}}}{\rho_{\text{water}}}$$

Where:

- θ_v = volumetric water content (cm³/cm³)
- θ_g = gravimetric water content (g/cm³)
- V_{water} = volume of water (m³)
- V_{soil} = volume of soil (m³)
- ρ_{water} = density of water (g/cm³)
- ρ_{soil} = density of soil (g/cm³)
- M_{water} = mass of water (g)
- M_{soil} = mass of soil (g)

Equation 3-2 Soil volumetric water content (θ_v) is the volume of liquid water per volume of soil with volume being the ratio of mass to density (ρ) (Fredlund and Xing 1994).

3.3.3 Plant water status Ψ_w

Leaf water potential (Ψ_w) was used as a measure of plant hydration, measurement for which were taken on both the -RSWD stress and +RSWD stress plants using a pressure chamber (Model 600) fitted with a bladed grass compression gland (PMS Instrument Company, Oregon, USA). The components which determine leaf water potential are described in Equation 3.3, where Ψ_w is always negative because Ψ_s and Ψ_M are negative and Ψ_P does not offset them, and as such the more negative a value of the greater the extent of water deficit stress in plant tissues.

$$\Psi_W = \Psi_S + \Psi_P + \Psi_M$$

Where:

- Ψ_W = water potential
- Ψ_S = solute potential
- Ψ_P = pressure potential
- Ψ_M = matrix potential

Equation 3-3 Components of leaf water potential (Blum 2011).

As soil moisture is depleted through evapotranspiration, soil water potential is reduced. Leaf water potential in turn must further reduce to create the necessary gradient differential between the soil, roots and the leaves which through the mechanism of capillary action works to draw water into the roots and up the stem following the pressure gradient from high to low. As such the more negative a value of Ψ_W , the greater the extent of water deficit stress in plant tissues.

Ψ_W measurements were taken during the RSWD stress experiment from the second youngest leaf between 7.00-8.00am at the following intervals during the fourteen day RSWD stress treatment; start (day 0), middle (day 7), end (day 14), and 7 days post treatment end. Leaves were carefully harvested to ensure a clean edged transverse cut in the medial portion of the leaf. Immediately following collection each leaf was placed in the pressure chamber with the cut end being threaded through the pressure gland so as to leave 0.5-1cm of tissue exposed on the upper side of the pressure gland end ensuring the upper and lower surfaces of the leaf were flush against the pressure gland (i.e. not folds in the leaf). The pressure gland was then tightened around the leaf to create an air seal and the chamber pressurised slowly (8-10 bars minute). The cut margin of the leaf was observed for expulsion of sap, which is the approximate visual indicator of when the pressure in the chamber is equal to the leaf pressure. The number of Bar/Mpa at which the sap expulsion occurred was recorded and cross referenced against experimentally determined correlations between leaf water potential (Ψ_W) values and plant water status (Table 3.5).

Table 3.5 Level of leaf RSWD stress gauged by measuring leaf water potential (Meron *et al.*, 1987).

Pressure bomb chamber pressure at point of leaf sap expulsion (Bar)	Level of plant water stress
> -10	No stress
≤ -10 to > -12	Mild stress
≤ -12 to > -14	Moderate stress
≤ -14 to > -16	High stress
≤ -16	Severe stress (permanent wilting point)

3.3.4 Relative humidity and ambient temperature

It was necessary to monitor relative humidity (RH) and ambient temperature (AT) during the two PC2 glasshouse experimental growth periods. RH and AT in addition to leaf area index (LAI) are the major determinants of crop evapotranspiration (ET) (Allen *et al.*, 1998). Although the PC2 glasshouse facilities were set to keep these two factors constant in the event of a system failure during the RSWD treatment any changes in RH and ambient temperature would drastically influence the rate of onset and intensity of the RSWD stress treatment. A reduction in temperature or increase in relative humidity would decrease net evapotranspiration losses while an increase in temperature and a decrease in relative humidity would increase the net evapotranspiration rate

Monitoring of RH and AT was undertaken using a Decagon Em50G data collection system fitted with an EHT temperature /relative humidity sensor housed in a radiation shield (ICT International Pty. Ltd. Armidale, NSW, Australia). Data stored in the data logger was downloaded and configured in the manufacturers specified compatible software ECH₂O utility (See Appendix II for Figures 8.3, 8.4, 8.5, 8.6, and 8.7 for a summary of RH and AT data collected in the two years of RSWD stress experiments)

3.3.5 Sampling schedule and tissue collection

Plant material was harvested from developing wheat heads and leaves for evaluation at six consecutive growth stages during the experimental (Table 3.6). The purpose of collection varied and at specific harvest times related to either more of the following; gene expression analysis, cytological observation, plant water potential, or grain set measurement. Collection and fixation methods for

cytological samples were described in section 'Phenotyping-cytology'. Samples collected for gene expression were snap frozen in liquid N upon collection in the glasshouse and subsequently transferred to -80°C for long term storage. Samples collected for plant water potential and grain set did not require fixation or freezing and were processed for analysis immediately upon collection.

Table 3.6 Harvesting schedule for the 2012 RSWD stress experiment

Growth stage	Tissue sampled	Number of samples per treatment group (WxK DH D08-299 Control, WxK DH D08-299 Drought, WxK DH D02-105 Control, WxK D02-105 Drought)				
		Gene expression (semi quantitative and RNASeq)	Cytology (anther/pollen development)	Cytology (pollen starch content)	Grain set	Plant water potential
Auricle distance 5cm (AD5)	Leaf Spike	3 3	- 3	- -	- -	3 1
Auricle distance 10cm (AD10)	Spike	3	3	-	-	-
Auricle distance 15cm (AD15)	Leaf Spike	3 3	- 3	-	-	3 -
Full head emerged (FHE)	Leaf Spike	3 3	- 3	-	-	3 -
Anthesis commenced (AC)	Leaf Spike	-	3	3	-	3 -
Maturity	Spike	-	-	-	3	-

3.4 Molecular Protocols

3.4.1 Extraction of genomic DNA from leaf and developing head tissue

Total nucleic acid was extracted from 2-4g of leaf or 0.2-2g of head tissue.

Tissues were immersion in liquid nitrogen housed in a porcelain mortar and pestle and pulverised to a powder consistency and subsequently transferred to a 1.5ml microcentrifuge tube. Samples were homogenized by adding 600µl DNA extraction buffer, 600µl of phenol/chloroform/iso-amyl alcohol (25:24:1), briefly hand mixed by inversion and then centrifuged for 10 minutes at 1400rpm. Upon centrifugation the upper aqueous phase was transferred into a clean 1.5ml microcentrifuge tube and the lower phase discarded. Another 600µl of phenol/chloroform/iso-amyl alcohol (25:24:1) was added followed by hand mix inversion, placed on ice for two minutes, and centrifuged for five minutes at 1400rpm. Transfer of the upper aqueous phase to a new tube and discard of the lower phase was repeated. 60µl of 3M sodium acetate 3M (pH4.8) and 600µl of

100% ethanol was added, mixed briefly by hand inversion, left to precipitate at room temperature for two minutes, and centrifuged for five minutes at 1400rpm. The supernatant was discarded and the formed pellet retained at the base of the tube. An ethanol wash was performed twice by adding 1ml of 70% ethanol, centrifuging for two minutes and discard of the supernatant. The purified pellet was left to dry, where after it was resuspended in 100µl of sterile H₂O. The extracted genomic DNA was quantified using a nanodrop ND-1000 spectrophotometer (Thermo Fisher, Scoresby, Vic., Australia).

3.4.2 Extraction of RNA from leaf and developing head tissue

RNA was extracted from 2-4g of leaf or 0.2-2g of head tissue. To enable high throughput tissue disruption a Quiagen Tissue Lysis system was used. Frozen tissues (stored at -80°C) in 2ml microcentrifuge tubes (Eppendorf South Pacific Pty. Ltd, North Ryde, NSW, Australia) were secured in 2 adaptor casings (each holding 24 2ml microcentrifuge tubes) and placed in a bath of liquid nitrogen to prevent thawing. Three stainless steel beads (5mm diameter) were placed in the tubes, the tubes capped and further secured by the adaptor casing lid. The adapter casings were then placed on a Quiagen Tissue Lyser machine for 2 minutes at 30Hz. The samples were disrupted to a fine powder consistency. There after RNA extraction followed the standard Trizol Quaigen RNeasy Plant Mini Kit protocol. Following extraction RNA was eluted in 30µl of DNase free H₂O, quantified on a nanodrop ND-1000 spectrophotometer (Thermo Fisher, Scoresby, Vic., Australia) and stored at -80°C.

3.4.3 Polymerase chain reaction (PCR)

PCR's were performed using the Taq DNA Polymerase kit from Fisher Biotec Australia (Wembley, WA). PCR reaction mixtures following the manufacturer specifications for a 20µl reaction volume were: 1x Buffer (670mM Tris-HCL pH 2.8, 166mM [NH₄]₂SO₄, 4.5% Triton X-100, 2mg/ml gelatin), 25mM MgCl₂, 200µM dNTP, 2.5 units (U) Taq DNA polymerase, 6µl 1% cresol red, 10pmol upstream and downstream primers with 200ng of template DNA. PCR reactions were performed in a 0.2ml microfuge tube in a DNA Thermo Cycler model (Model: GeneAmp PCR System 2700, Applied Biosystem, Mulgrave, Vic.,

Australia). Depending on the T_m values of the primers used the standard thermal cycling profile used was: incubation (3 minutes at 94°C), followed by 35 cycles of denaturation (30 seconds 94°C), primer annealing (30 seconds 57°C), extension (1 minute 72°C). Additional extension was performed (10 minutes 72°C). Deviations from this standard protocol are described in detail where relevant in subsequent chapters.

3.4.4 Reverse transcription (RT),

Reverse transcription (RT) was performed with Applied Biosystems (AB) high-capacity cDNA reverse transcription kit 200 reactions (Mulgrave, Vic., Australia). Reverse transcriptase was used to transcribe RNAs into cDNA. The reaction mixture was: 10x buffer, 25x dNTP (100mM), 1 unit (U) of RNase inhibitor, 2 units of 10x RT random hexamers or downstream primers, 1 unit of MultiScribe™ Reverse Transcriptase. After adding 2 µg of total nucleic acid RNase-free water was added to bring the final reaction volume up to 10 µl. Thermo cycler conditions were as follows: hold 1- 25°C for 10 minutes, hold 2- 37°C for 60 minutes, hold 3- 85°C for 5 minutes, hold 4- 4°C for infinity.

3.4.5 Primer design for RT, PCR, and DNA sequencing

All oligonucleotide primers used in this study were synthesized by GeneWorks Pty Ltd (Thebarton, South Australia). Primers used for cloning and sequencing of DNA were designed using the Primer3 ver. 0.4.0 (<http://primer3.sourceforge.net/>) using the reference wheat genome sequence (Chinese spring ctg 506) to enable the entire sequence length of the gene of interest to be spanned by a five overlapping fragments. The positions of individual primers were chosen such that intron-exon boundaries would be spanned. Touch down PCR was used for cloning and sequencing reactions, and the specific thermocycler conditions for specific primers in discussed in a later chapter.

Semi quantitative gene expression evaluation was undertaken using three sets of primers, two from previously published gene expression studies ((Ji *et al.*, 2010; Koonjul *et al.*, 2005), the third set being designed on previously unpublished

isoforms sequences, findings which arose subsequent to this current work (Table 3.7). The latter primer set was also designed using the Primer3 ver. 0.4.0 (<http://primer3.sourceforge.net/>).

Table 3.7 Primers used for cell wall invertase (*IVR1*) gene expression analysis

Origin of primer		Primer sequence 5'-3'	Fragment length (bp)
(Ji <i>et al.</i> , 2010)	Forward	ACGAGGGTCTACCCGAGGAA	140
	Reverse	AACAAGCGTCTGGGGCAGTA	
(Koonjul <i>et al.</i> , 2005)	Forward	CATGAGGGGGATCGCGGCGTTGTA	726
	Reverse	ACCCTTGACGGCCTTGTTGCTGAC	
H Webster thesis	Forward	TAC CCG GAG CTC CAG TCG	395
	Reverse	ACC CGG CCA ACC TCT CCG A	

3.4.6 Agarose gel electrophoresis

Agarose gels were prepared in dye cast moulds and electrophoresis was carried out using a BioRad (Gladesville, NSW, Australia) Sub-Cell GT agarose gel system. Gels were prepared for a range of agarose percentages (1-4%) of the total volume of the gel solution (dissolved in 1xTAE buffer) and stained with cybersafe prior to solidification. PCR products and a single lane of a 1KbDNA marker ladder (Life Technologies, Inc., Melbourne, Australia) were loaded into the cast wells of the agarose gels (immersed in 1xTAE buffer) and electrophoresed at 65V for 75minutes for optimal band separation. Evaluation and image capture of separated bands on gels was undertaken with UV illumination using a BioRad Gel Documentation System. Post image capture evaluation of separated bands was performed using the freeware software (<http://www.gelanalyzer.com>).

3.4.7 Cloning of PCR products

To clone DNA, purified PCR products were ligated to pGEM-T Easy (Promega Corp, Annandale NSW) cloning vectors, competent *E.coli* was transformed with ligated products and recombinant plasmid DNA purified from cell cultures. Purified PCR products were ligated to pGEM-T Easy cloning vectors in a 1:3 insert:vector ratio (Equation 3.4).

$$\text{ng of insert} = \frac{\text{ng of vector} \times \text{kb of insert}}{\text{kb size of vector}} \times \text{insert:vector molar ratio (1:3)}$$

Equation 3.4 Calculating the insert vector ratio (Promega Corp. pGEM-T Easy Vector System technical manual (2011).

Ligation reactions were prepared to a total reaction volume of 10µl and incubated at 72°C for 30minutes (Table 3.8).

Table 3.8 Volumes of individual constituents in ligation reactions.

Reaction component	Standard reaction (µl)	+ve control (µl)	-ve control (µl)
2x T4 DNA ligase buffer	5	5	5
pGEM-T Easy Vector (54ng/µl)	1	1	1
DNA insert	X	-	-
T4 DNA ligase	1	1	1
H ₂ O	(added to make total reaction volume of 10µl)	(added to make total reaction volume of 10µl)	(added to make total reaction volume of 10µl)
Control DNA (2ng/µl)	-	2	-

3.4.8 Restriction endonuclease digestion

Restriction digestion of plasmids from transformed bacteria with Eco RI (Promega Corp, Annandale NSW) was necessary to check for the following: (i) presence of restriction sites, (ii) the size of the fragments in cloning vectors. Reactions (Table 3.9) were incubated at 37°C for 1hour.

Table 3.9 Restriction endonuclease digestion reaction components

Reaction component	Standard reaction (µl)	+ve control (µl)
EcoRI (10U/µl)	0.5	-
10x Buffer	2	2
DNA (total of 1ug)	X	10
H ₂ O	(added to make total reaction volume of 20µl)	8

3.4.9 Transformation of competent E.coli cells

Competent cells were prepared from commercial stocks sourced from Promega Corp, using the modified rubidium chloride method outlined in manufacturers users guide (Promega’s Protocols and Applications guide 2011). Stocks of cells stored at -80°C were streaked on LB medium plates and incubated at 37°C overnight. 10ml of LM medium was inoculated with a single colony from a plate and incubated at 37°C overnight with shaking (225 strokes/min). 10% of the incubated culture was used to inoculate 100ml of LB in a flask. The flask solution was incubated at 37°C with shaking (225 strokes/min) until OD₆₀₀ levels reached

0.45-0.6. There after cells were chilled on ice for 10minutes, centrifuged for 5 minutes at 4,500g at 4⁰C for collection. Cells were resuspended in chilled titration buffer I (100mM RbCl, 10mM CaCl₂, 50mMCl₂.4H₂O, 30mM KOAc, and 15% glycerol (pH5.8), equivalent to 40% of the culture used. The solution was chilled on ice for 5 minutes, and centrifuged for 5 minutes 4,500g at 4⁰C thereafter. Collected cells were resuspended with gentle swirling in 4 ml of chilled titration buffer II (10mM MOPS, 75mM CaCl₂, 10mMRbCl, and 15% glycerol (pH6.5) and chilled on ice for 10 minutes. Resuspended cells were aliquoted into 100µl volumes in 1.5µl microcentrifuge tubes, snap frozen, and preserved at -80⁰C.

3.4.10 Heat shock transformation and bacterial growth

An aliquot of frozen competent cells was thawed on ice. Upon thawing the ligation mixture added to the competent cells in a 1.5ml microcentrifuge tube and the solution was mixed gently by flicking, then placed on ice for 30 minutes. The solution was heat-shocked in a 42⁰C water bath for 45seconds and returned to incubate on ice for 5 minutes. 900µl of LB broth was added to the solution and subsequently shaken at 225 strokes/min for 1.5 hours at 37⁰C.

3.4.11 Selection for transformants

Identification of transformants was undertaken by antibiotic selection and blue white screening. LB-ampicillin (100ugmL⁻¹ ampicillin) were prepared and 30minutes prior to streaking with the transformed bacterial culture the plates were applied with 20µl of 50mgmL⁻¹ X-Gal (5-bromo-4-chloro-3-indoly-B-D-galactopyranoside and 100µl of 100mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and warmed to 37⁰C . 200µl of the incubated transformation culture solution was streaked out on the plates, and incubated at 37⁰C for 24 hours. Upon completion of incubation the plates were evaluated for presence/absence and frequency of blue and white colonies. White colonies being those which most commonly contain inserts, while blue colonies do not.

3.4.12 Analysis of transformants by PCR

To check for the presence of inserts individual white colonies were collected and resuspended in 20 μ l of sterile H₂O and thereafter were subject to PCR analyses using the SP6 and T7 primers. PCR reaction were formulated to a total volume of 20 μ l and contained individual reaction components as follows: 4mM MgCl₂, 200mM dNTP's, 10pmol of primers, 1X Buffer, and 0.05U *Taq* DNA polymerase. Thermo cycler conditions were: incubation (3 minutes at 94^oC), followed by 25 cycles of denaturation (30 seconds 94^oC), primer annealing (30 seconds 55^oC), extension (1 minute 72^oC). Additional extension was performed (10minutes 72^oC). When visualised on a 2% agarose gel those PCR products that contained transformants with inserts showed the amplicon plus 140 bases of the plasmid vector, while in the insert free transformants did not.

3.4.13 Analysis of transformants by plasmid DNA mini-prep restriction digestion

Colonies found to contain inserts by PCR analysis were grown overnight in 10ml of LB broth containing 100 μ g mL⁻¹ of ampicillin and subjected to shaking at 225rpm. Plasmid DNA was purified from the bacterial cell cultures using a BioRad Aurum Plasmid Mini Kit (Gladesville, NSW, Australia). An aliquot of plasmid DNA was digested with EcoRI to confirm the presence of the insert in the pGEM-T Easy vector. The digest products were run on a 2% agarose gel to determine the size of the insert using the non recombinant pGEM-Teasy plasmid as a DNA marker.

3.4.14 DNA purification from agarose gels

PCR products were purified from low melting point agarose gels using the Promega Wizard SV Gel and PCR clean up System. After evaluation and image capture of separated bands on agarose gels bands were cut from the gel under UV illumination on a Gel Documentation System using a scalpel. The excised bands were transferred to individual 1.5ml microcentrifuge tubes. A volume of membrane binding solution (equivalent to the weight of the excised band in grams) was added to the gel fragments and vortexed for 10 minutes at 65^oC until dissolution of the solidified fragment. The dissolved solution was transferred to a filter column assembly, allowed to incubate at room temperature for one minute,

and centrifuged at 10,000g for 1 minute. Waste collected at the bottom of the assembly was discarded and then 700µl of membrane wash solution was added to the filter column. The filter column assembly was re-centrifuged for 1 minute and the flow through waste discarded. The membrane wash step was repeated with a 500µl of the solution. The filter system was then transferred to a new sterile 1.5ml microcentrifuge tube and the purified DNA held in the filters was eluted by adding 50µl of nuclease free H₂O, allowed to incubate for 2 minutes at room temperature, and finally centrifuged at 16,000g for 2 minutes.

3.4.15 Plasmid DNA purification

The BioRad Aurum Plasmid Mini Kit spin format (Gladesville, NSW, Australia) was used for purification of plasmid DNA. Methods followed the manufacturers protocol specifications. Upon culturing 1.5ml of the transformed bacterial suspension was transferred to a sterile 2ml microcentrifuge tube to which 250µl of re-suspension and lysis solution was added as well as 350µl of neutralizing solution. The solution was vortex briefly and centrifuged for 5 minutes at 10,000rpm to collect pellet cell debris. The clear supernatant was then transferred to a filter column assembly, centrifuged for 1 minute at 10,000rpm, and the collected waste was discarded. 750µl of wash solution was added to the filter assembly and the sample was re-centrifuged for 2 minutes to remove residual cell debris. The filter system was then transferred to a new sterile 1.5ml microcentrifuge tube and the purified DNA held in the filter was eluted by adding 50µl of nuclease free H₂O, allowed to incubate for 2 minutes at room temperature, and finally centrifuged at 16,000g for 2 minutes.

3.4.16 DNA sequencing

DNA sequencing was performed on both cloned inserts and PCR products from semi-quantitative gene expression analysis. Complete sequences for the cloned inserts were obtained by sequencing double-stranded templates. T7 and SP6 primers were used to sequence the clones in the pGEM-T easy vector. Direct sequencing on semi-quantitative PCR products was undertaken using the primers listed in Table 3.7. Sequences were generated on an ABI Prism XL 3730 sequencer (Applied Biosystems, Foster City, CA, USA) using BigDye3.1 (BDV3.1)

and were prepared as $\frac{1}{4}$ of the volume for full sequencing reactions recommended by ABI. The amount of template used in the sequencing reaction was 150-300ng and 1-3ng for plasmid DNA and purified PCR product reactions respectively, while all other reaction components for the two types of reactions remained constant; 2 μ l of Dye terminator mix, 3.2pmoles of primer, and H₂O to an amount that brought the final reaction volume up to 10 μ l. Reactions were performed in a DNA thermo cycler (Model: GeneAmp PCR System 2700, Applied Biosystem, Mulgrave, Vic., Australia) for 3 minutes at 94 °C, 50 cycles of 94 °C for 30 seconds, 57 °C for 30 seconds, and 72 °C for 1 minute. An additional 10 minutes extension was performed at 72°C.

After completion of PCR extension the reactions were purified. Purification involved the following steps: transfer of the PCR contents to a 0.5ml microcentrifuge tube, addition of 1 μ l of EDTA, 1 μ l of 3M NaOAc (pH 5.2), 25 μ l of 100% EtOH. The contents of the tube was mixed by pipetting and left to incubate at room temperature for 20 minutes, followed by centrifugation at 14,000g for 30 minutes. The supernatant was discarded and the remaining pellet at the bottom of the tube was washed with 125 μ l of EtOH and centrifuged for a further 5 minutes. The supernatant was then removed and the purified pellet was left to air dry at room temperature for 15 minutes to remove residual EtOH.

3.4.17 RNASeq

Details listed in Chapter 5 methods

Chapter 4: Genome-level identification of cell wall invertase genes in wheat for the study of drought tolerance

Publication contributions (Webster *et al.*, 2013):

Hollie Webster – experimental work and text preparation

Gabriel Keeble-Gagnère, Catherine Feuillet, and Frédéric Choulet– assembly of contig506 wheat chromosome 3B

Bernard Dell – text preparation

John Fosu-Nyarko – guidance on experimental work

Yasuhiko Mukai – in situ hybridization image

Paula Moolhuijzen – guidance for bioinformatics

Jizeng Jia and Xiuying Kong – *Aegilops tauschii* (wheat D genome progenitor) sequence

International Wheat Genome Sequencing Consortium – coordination of wheat genome sequencing

Rudolf Appels – Guidance on experimental design and text preparation

Hollie Webster^A, Gabriel Keeble-Gagnère ^B, Bernard Dell^A, John Fosu-Nyarko^C, Y Mukai^D, Paula Moolhuijzen^B, Matthew Bellgard^B, Jizeng Jia^E, Xiuying Kong^E, Catherine Feuillet^F, Frédéric Choulet^F, International Wheat Genome Sequencing Consortium^G, Rudi Appels^{B H}

^ASchool of Biological Sciences and Biotechnology, Murdoch University, Perth, Western Australia

^BCentre for Comparative Genomics, Murdoch University, Perth, Western Australia

^CState Agricultural Biotechnology Centre, Murdoch University, Perth, Western Australia

^DLaboratory of Plant Molecular Genetics, Division of Natural Science, Osaka Kyoiku University, Kashiwara, Osaka 582-8582, Japan

^EKey Laboratory of Crop Gene Resources and Germplasm Enhancement, Ministry of Agriculture, National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing China

^FINRA UMR 1095, Génétique Diversité et Ecophysiologie des Céréales, 63100 Clermont-Ferrand, France

^GInternational Wheat Genome Sequencing Consortium, Executive Director Kellye Eversole (www.wheatgenome.org)

^HCorresponding author. Email: rappels@ccg.murdoch.edu.au

4.1 Abstract

In wheat (*Triticum aestivum* L.) drought induced pollen sterility is a major contributor to grain yield loss and is caused by the down-regulation of the cell wall invertase gene (CW-INV) *IVR1*. The *IVR1* gene catalyzes the irreversible hydrolysis of sucrose to glucose and fructose, the essential energy substrates which support pollen development. Down-regulation of *IVR1* in response to drought is isoform specific and shows variation in temporal and tissue specific expression. *IVR1* is now prompting interest as a candidate gene for molecular marker development to screen wheat germplasm for improved drought tolerance. The aim of this study was to define the family of *IVR1* genes to enable: (1) individual isoforms to be assayed in gene expression studies, and (2) greater accuracy in *IVR1* mapping to the wheat genetic map and drought tolerance QTL analysis. Using a CW-INV specific motif as a probe, wheat genomics platforms were screened for the presence of unidentified *IVR1* isoforms. Wheat genomics platforms screened included the IWGSC wheat survey sequence, the wheat D genome donor sequence from *Aegilops tauschii*, and the CCG wheat chromosome 3B assembly: contig506. Chromosome specific sequences homologous to the query motif were isolated and characterised. Sequence annotation results showed five previously unidentified *IVR1* isoforms exist on multiple chromosome arms and on all three genomes (A, B, and D): *IVR1-3A*, *IVR1-4A*, *IVR1-5B*, *IVR1.2-3B*, *IVR1-D*. Including three previously characterised *IVR1* isoforms (*IVR1.1-1A*, *IVR1.2-1A* and *IVR1.1-3B*), the total number of isoform gene family members is eight. The *IVR1* isoforms contain two motifs common to CW-

INV's (NDPN and WECPDF) and a high degree of conservation in exon 4, suggesting conservation of functionality. Sequence divergence at a primary structure level in other regions of the gene was evident amongst the isoforms, which likely contributes to variation in gene regulation and expression in response to water deficit within this sub-family of *IVR1* isoforms in wheat.

4.2 Introduction

In wheat (*Triticum aestivum L.*) grain number per ear is the main determinant of yield and is significantly reduced by pollen sterility when water deficit is incurred during pollen meiosis. Expression of the cell wall invertase (CW-INV) gene *IVR1* in tapetal tissues that surround pollen during its formation is essential in maintaining pollen fertility. *IVR1* functions to catalyze the irreversible hydrolysis of sucrose to glucose and fructose, the essential energy substrates which support pollen development. *IVR1* is down-regulated by reproductive stage water deficit stress (RSWD), the effect of which leads to pollen being starved of its energy source, resulting in sterility, reduced grain number, and ultimately reduced yield (Koonjul *et al.*, 2005).

In view of this evidence, and the increasing frequency of drought events in Australia over the last decade, there has been growing demand for wheat pre-breeding research to elucidate the molecular determinants of *IVR1* gene regulation and expression under water deficit and non-limited conditions (Ji *et al.*, 2010; Powell *et al.*, 2012). Underpinning the interest for improved understanding of sugar metabolism during head development is the objective of breeding wheat varieties capable of maintaining pollen fertility, and thus grain set, under water deficit stress conditions. Achieving this objective would better enable the genetic potential of existing wheat germplasm, capable of high yields in non-limited water environments, to be fully exploited in regions where water is frequently limited during the period in the growing season which coincides with head development and grain fill (Dolferus *et al.*, 2011; Reynolds *et al.*, 2005).

Evidence supporting molecular approaches to breeding wheat for stable pollen fertility was seen in recent studies by (Ji *et al.*, 2010), where wheat varietal-

specific expression of *IVR1* under RSWD stress conditions was examined. Results from this study were consistent with similar studies by Koonjul *et al.*, (2005) and demonstrated a positive correlation between *IVR1* expression and pollen fertility. However, the degree of *IVR1* down-regulation in response to RSWD stress showed significant differences between varieties, in that certain varieties maintained gene expression levels under +RSWD stress conditions (and thus maintained pollen fertility), while others did not. This result is consistent with *IVR1* being a strong candidate gene for RSWD stress tolerance QTL analysis (Ji *et al.*, 2010).

Plant invertases (β -D-fructofuranosidase EC 3.2.1.26) constitute a family of enzymes that hydrolyse sucrose to glucose and fructose. In plants, three types of invertases have been identified, namely, cell wall (CW-INV), vacuolar (V-INV) and cytoplasmic (C-INV). Invertases have roles in a number of plant physiological processes related to long-distance nutrient allocation as well as regulating developmental processes, hormone responses and biotic and abiotic interactions (Roitsch and Gonzalez 2004; Tymowska-Lalanne and Kreis 1998). In order to provide a source of energy, sucrose is transported from leaves to the stem, reproductive structures and roots (Ruan *et al.*, 2010). The sucrose molecule is cleaved at the α 1- β 2-glycosidic bond, either by invertases or sucrose synthase (SST). Invertases catalyse the irreversible hydrolysis of sucrose to glucose and fructose while SST catalyses reversible cleavage of sucrose to UDP-glucose and fructose (Roitsch and Gonzalez 2004). Upon cleavage of sucrose, glucose and fructose are transported into the cells via hexose transporters located in the plasma membrane (Bush 1999).

The invertase gene family consists of many members within each of the three broad functional groups (i.e. CW-INV, V-INV, C-INV) (Tymowska-Lalanne and Kreis 1998). In the present study the gene nomenclature used defines *IVR1* isoforms as belonging to the cell wall invertase (CW-INV) gene sub-family in wheat. The term isoform is used in the context that it designates different proteins derived from related genes. A widely accepted explanation of multiple invertase isoform existence within each of the functional groups is that this

allows for greater flexibility in the control of sucrose metabolism, translocation and storage under different internal and external stimuli and at various developmental stages or tissues, or at the whole plant level (Akhunov *et al.*, 2003). (Koonjul *et al.*, 2005) found that individual *IVR1* and V-INV isoforms were expressed in a highly specific manner in response to water availability, with respect to both tissue type and developmental stage. Similar results were recently reported by Crismani *et al.*, (2011), who showed that, in rice, closely related *IVR1* isoforms within anther specific transcripts exhibited clear temporal variation under non-limiting water conditions (i.e. –RSWD stress).

Studies of invertase transcription regulation conducted in a number of different crop species have found that transcription is influenced by a range of signals including; low oxygen, plant hormones, sugars, and environmental and pathogenic cues (Koch 1996; Long *et al.*, 2002; Proels and Roitsch 2009). In studies specifically examining the role of sucrose in regulating CW-INV expression, varied results were reported; some CW-INV appear to have expression reduced by sucrose accumulation, while others isoforms are upregulated (Kaufman *et al.*, 1973; Sturm and Chrispeels 1990). Tymowska-Lalanne and Kreis (1998) suggest that variable CW-INV transcription under the control of sucrose signalling may be related to different isoforms being regulated in a highly specific manner, as a reflection of subtle but important differences in their function.

The aim of this paper was to define *IVR1* isoforms at the genome level in wheat, in order to provide a basis for greater accuracy in discriminating between the isoforms that are expressed in developing wheat heads under plus and minus RSWD stress conditions. Evidence from this study suggests that characterisation of the *IVR1* isoforms is essential to further gene expression studies as it will enable definitive assays of single isoforms. *IVR1* isoform specific assays will permit variation in spatial and temporal expression to be analysed in the context of improving understanding of *IVR1* gene regulation and functionality.

4.3 Materials and methods

4.3.1 Wheat BAC sequencing

The assembly of BACs from flow-sorted chromosome 3B DNA was as described by Paux *et al.*, (2008) and utilized a modified HICF SNaPshot protocol. In summary, five restriction endonucleases were used to digest DNA samples. Anchoring of the BAC contigs to a genetic map was carried out using established SSR markers as well as probes for insertion-based polymorphisms, using a mapping population derived from a cross between wheat varieties Chinese Spring x Renan and genetic lines carrying chromosomal deletions (Paux *et al.*, 2008). For contig506 (ctg506), the minimum tiling path was established using the FPC software at a high stringency ($1e^{-75}$) and refined by analyses using LTC software (Frenkel *et al.*, 2010). Sequencing was carried out at 5-8x coverage and the Solexa-Illumina sequencing was carried out at >300x coverage using 75 base pair reads. Additional 454 sequencing using 8 kb mate-pair libraries resolved ambiguities in the sequence assembly. The 480 kb portion of ctg506 (1.3 Mb) analysed in this paper has been submitted to Genbank (Genbank submission pending).

4.3.2 Identification of *IVR1* isoforms and sequence annotation

Access to the genome survey sequences for the chromosome arms of wheat before publication (<http://urgi.versailles.inra.fr/Species/Wheat/Sequence-Repository>) was provided by the International Wheat Genome Sequencing Consortium (IWGSC). To screen for the presence of previously unidentified *IVR1* isoforms a BLASTN query was carried out across the entire wheat survey sequence using two invertase specific motifs: (1) 5'CAGATCCAAATGGTACGTA'3, residues which span a β -Fructosidase motif (NDPN) and which are common to all invertase genes (i.e. CW-INV, V-INV, C-INV), and (2) 5'TGGGAATGCCCCGATTTC'3, residues which span a catalytic domain (WECPDF) and which are present only in CW-INV isoforms (Koonjul *et al.*, 2005; Tymowska-Lalanne and Kreis 1998). Determination of sequence-based *IVR1* categorisation required the presence of both the NDPN and WECPDF motifs. The BLASTN analysis was carried out across all available chromosome arm datasets and yielded subject sequence contigs identified by accession numbers, which were

then used to obtain the respective contigs from the specific chromosome arm on which they were located (Table 4.1). The contig sequences were imported into Geneious Pro (version 5.5.6, <http://www.geneious.com/>) for alignment and annotation using the *IVR1* full-length cDNA sequence, accession AF030420.1, as a reference.

Table 4.1 *IVR1* isoform accession identified in the IWGSC wheat survey sequence using the β -fructosidase (NDPN) and catalytic domain (WECPDF) motifs and isoform homology to the full length *IVR1* cDNA (AF030420.1). Note: tabulated percentage values for isoform homology to the reference sequence AF030420.1 were deduced after annotation of the isoform intron-exon boundaries and based on homology of coding region sequence.

IWGSC survey sequence contig ID	Chromosome arm location	Contig length (base pairs)	Proposed <i>IVR1</i> nomenclature	Query motif protein ID	Isoform homology to AF030420.1 (%)
10503511	3B	4630	<i>IVR1-3B</i>	WECPDF, NDPN	82.6
10595081	3B	11558	<i>IVR1-3B</i>	NDPN	84.4
529524	3A	2756	<i>IVR1-3A</i>	WECPDF, NDPN	82.7
7174266	4A	14238	<i>IVR1-4A</i>	WECPDF, NDPN	83.9
10924301	5B	25605	<i>IVR1-5B</i>	WECPDF, NDPN	92.9
4492773	5D	13924	<i>IVR1-5D</i>	WECPDF, NDPN	95.9

4.3.3 Plant material and DNA extraction

Wheat (*Triticum aestivum* L. cv Westonia and cv Kauz) seed stocks were provided by the Department of Agriculture and Food Western Australia (DAFWA). Seed stocks were pre-germinated at 4°C for 14 days and then planted into free draining pots with a uniformly prepared soil medium consisting of 2 parts loam, 1 part river sand, and 3 g of basal fertilizer (mg kg⁻¹: 1220 N, 368P, 819K), and grown in controlled glasshouse conditions with ambient temperature during the plant growth period of 9°C (night) and 18°C (day). Pots were watered three times daily using an irrigation mat system to maintain average soil volumetric water content of 60%. Leaf tissue samples were taken for DNA extraction at the early tillering phase (Zadoks growth scale stage Z20), frozen in liquid nitrogen, and DNA extraction was performed using a generic phenol-chloroform extraction method. The extracted DNA was quantified using a Nanodrop ND-1000 spectrophotometer.

4.3.4 *IVR1* sequencing and analysis

The genomic sequence ctg506 (Genbank submission pending), as defined in this manuscript, was located on chromosome 3B of Chinese Spring and contained

IVR1.1-3B. This isoform was used as the reference sequence to design primers (each spanning intron-exon junctions) to amplify a total of five overlapping fragments spanning the entire length of *IVR1.1-3B* (Table 4.2). Primers were designed with Primer3 version 0.4.0 (<http://primer3.sourceforge.net/>). Each PCR reaction was 20 µl, containing: 200 ng of template DNA, 10 pmol each of a forward and reverse primer pair specific to an individual genomic fragment, 200 µM dNTP, 2.5 mM MgCl₂, 1x reaction buffer, 6 µl of 1% cresol red, and 2.5 units of Taq polymerase (Fisher Biotech, Australia). Touch down PCR (TD-PCR) conditions were: incubation for 3 min at 94°C, followed by 35 cycles of 1 min denaturation (94°C), 30 seconds annealing (decreasing by 1°C per cycle until the primer specific annealing temperature was reached, as in Table 2), and 30 second extension (72°C). An additional ten minutes extension at 72°C was also performed.

The TD-PCR products were separated on a 2% agarose gel and DNA purified using a Wizard SV Gel and PCR Clean up system (Promega Corp, Annandale, Australia). The purified genomic DNA fragments were cloned by ligation to the pGEM-T Easy cloning vector (Promega Corp, Annandale, Australia). Both strands of the cloned *IVR1* isoform fragments were sequenced with the Sp6 and T7 promoter using BigDye3.1 (BDV3.1 plus dgtp) termination chemistry, and were run on an ABI Prism XL 3730 sequencer (Applied Biosystems, Foster City, Calif., USA).

Table 4.2 TD-PCR primer used for amplification of overlapping fragments spanning *IVR1.1-3B*

Primer pair (name)	Primer nucleotide sequence	Fragment length (base pairs)	TD-PCR annealing temperature (°C)
IVR1-1 Forward IVR1-4 Reverse	CTCTCTATCGTCGATCGAGCTA TGTAGTACATTGGCCCTGCATC	572	58-53
IVR1-4 Forward IVR1-6 Reverse	GATGCAGGGCCAATGTACTACA CTCCAGTGTCTGCACATATATC	577	58-53
IVR1-6 Forward IVR1-7 Reverse	GATATATGTGCAGACACTGGAG GTGTAAGCTAGTGACCTGGATC	1050	60-55
IVR1-7 Forward IVR1-9 Reverse	GATCCAGGTCCTAGCTTACAC ATATGTTACCTTGTTGGGTCA	513	60-55
IVR1-9 Forward IVR1-13 Reverse	TGACCCAACAAGGTAACATAT CTACTGCTCCAGCAGCTTGTTTCATC	570	58-53

4.4 Results

4.4.1 The genome level structure of wheat *IVR1* isoforms on chromosome 3B

Our sequence assembly of a BAC contig from chromosome 3B (ctg506) provided the genome sequence of two *IVR1* isoforms, *IVR1.1-3B* and *IVR1.2-3B*, in addition to the two previously characterised genes on chromosome 1A (Francki *et al.*, 2006). The two previously characterised *IVR1* isoforms, *IVR1.1-1A* (Genbank accession: HF571522.1) and *IVR1.2-1A* (Genbank accession: HF571523.1) were mapped to chromosome 1A using a genomic *IVR1* representative of AF030420.1 (cited in Koonjul *et al.*, (2005)).

The ctg506 is between the *Sr2* rust resistance and fusarium head blight resistance (*Fhb1*) loci, and includes the molecular marker STS104 (genetic map in Figure 4.1). The assembly initially comprised 12 BAC clones, with estimated insert sizes of 70 – 180 kb, compiled into a minimum tiling path. The linear sequence was established using a combination of Sanger and Solexa-Illumina sequencing technologies. The initial assembly had 21 contigs. Subsequent analyses of the FPC based assemblies and additional 454 sequencing of 8 kb mate pair libraries refined the assembly.

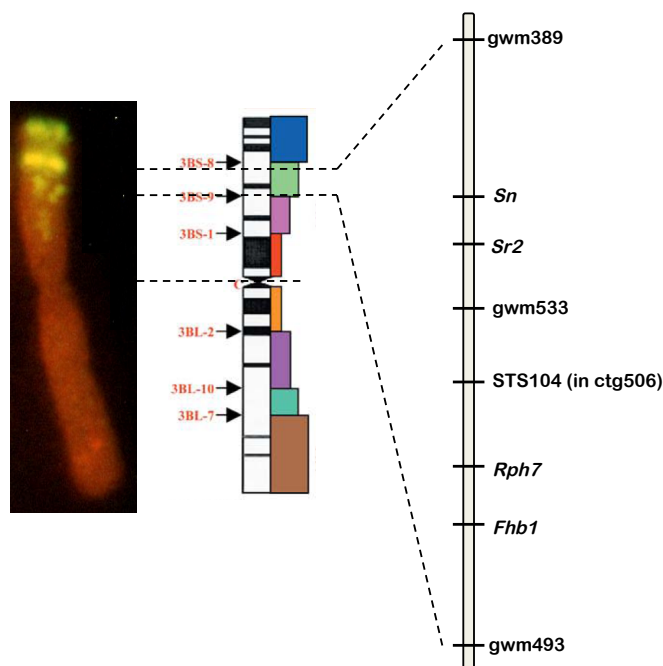


Figure 4.1 A summary molecular-genetic map of the region of chromosome 3B in which the genome segment characterised as ctg506 is located. The left panel indicated the *in situ* distribution of a repetitive sequence family found in ctg506, whereas the middle panel

indicated the deletion bin location (Sourdille *et al.*, 2004) of the genetic map shown in the right-most panel (modified from the 3B consensus map, <http://ccg.murdoch.edu.au> accessed 10 October 2011). The molecular marker STS104 defines the location of ctg506.

Figure 4.2 shows the details for 480 kb of ctg506 to indicate the overall structure of this genome sequence and the location of the *IVR1* isoforms. The highly repetitive region upstream from *IVR1.1-3B* is comprised of members of the Sc119.2 repetitive DNA family that are widely used as chromosome markers (left-hand panel Figure 4.1 (McIntyre *et al.*, 1990)) and are distributed in the ctg506 region of chromosome 3B at a relatively low representation compared to other sites in the genome (note the more intense band slightly distal to the region of interest). A SNP (BF293133) was found further downstream on the ctg506 genome segment shown in Figure 4. 2 and has been assigned a genetic location in chromosome 3B in the position expected for ctg506 (Zhang *et al.*, 2008(B)), providing additional confirmation of the chromosomal location of the sequences analysed.

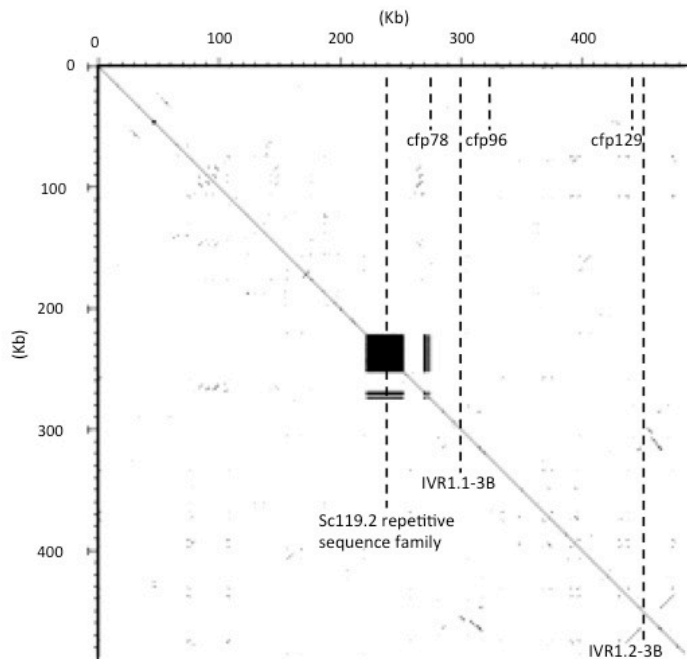


Figure 4.2 A dot plot of the 480-kb sequence defined within ctg506 against itself indicating that it is relatively free of major repeat sequences except for the cluster of sequences at 220-260kb. The repetitive sequence located at 220-260kb comprises variants of the 120pb, pSc119.2, sequence which is well known as a molecular marker (McIntyre *et al.*, 1990) and is particularly prominent in the terminal region of chromosome 3B (left panel in Figure 4.1).

Annotation of the 480 kb genome sequence was carried out using GenScan (<http://genes.mit.edu/GENSCAN.html>) as well as Augustus and EuGene (in TriAnnot <http://urgi.versailles.inra.fr/Species/Wheat/Triannot-Pipeline/>; (Leroy *et al.*, 2012)). The results consistently showed two *IVR1* isoforms at coordinates 300 kb and 454 kb in the genome sequence. In Figure 4.3 the primary structure for the *IVR1* isoforms is presented based on manual refinement of the annotations in the Geneious program, using AF030420.1 as a reference sequence to identify intron-exon boundaries. The *IVR1.1-3B* isoform has 7 exons and is clearly closely related to the reference cell wall invertase AF030420.1, including a characteristic small exon of 3 amino acids (DPN, part of the NDPN motif) and a conserved catalytic domain in exon 4 (WECPDF). The *IVR1.2-3B* isoform also has the DPN and WECPDF motifs but has a 9 exon structure and greater sequence divergence in other parts of the coding regions (as compared to the reference AF030420.1), particularly in exon 3.

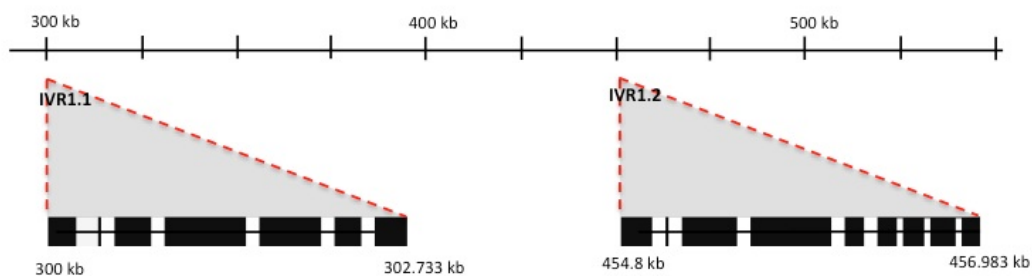


Figure 4.3 Detailed genome level arrangement of the *IVR1* isoforms located on ctg506 from chromosome 3B. A seven exon and nine exon structure is observed in *IVR1.1-3B* (left) and *IVR1.2-3B* (right), respectively, and divergence between the isoforms exists at an amino acid level (data not shown) particularly in exon 3. The black boxes indicate exons and ascending order from left to right corresponds to exon numbers 1 to 9.

4.4.2 The *IVR1* gene family in wheat

Sequence analysis of the cloned genomic DNA TD-PCR products amplified from exon 3 and exon 7, in both Westonia and Kauz, provided the initial evidence for the existence of additional *IVR1* isoforms in the wheat genome that had not yet been identified. The primers used to assay the fragments in question spanned exon-intron boundaries and thus were expected to have high specificity to assay *IVR1.1-3B* given they were designed using the Chinese Spring chromosome 3B sequence. Three homologous groupings were apparent in both the Westonia and Kauz sequences of exon 3 and 7, one group of which was homologous to *IVR1.1-*

3B (ctg506) but the remaining two did not show homology to any of the available annotated *IVR1* sequences for wheat, including: *IVR1.1-3B* (ctg506), *IVR1.1-1A* (Genbank: HF571524.1), *IVR1.2-1A* (Genbank: HF571525.1) and the full length *IVR1* cDNA AF030420.1. This evidence suggested at least two additional *IVR1* isoforms existed and that there was likely to be an *IVR1* sub-family in wheat.

The IWGSC survey sequences for the chromosome arms of wheat (<http://urgi.versailles.inra.fr/Species/Wheat/>) enabled three previously unidentified isoforms of the *IVR1* sub-family to be identified using the invertase specific motif NDPN, and the CW-INV specific motif WECPDF, as probes. Using these probes to screen the wheat D genome donor sequence from *Triticum tauschii* (Jia *et al.*, 2013) and the ctg506 chromosome 3B assembly, two additional uncharacterised *IVR1* isoforms were also identified. Figure 4.4 shows the isoform gene structure determined using the full-length *IVR1* cDNA AF030420.1 as a reference. The characterised and annotated *IVR1* isoforms show the conserved motifs DPN (exon 2) and WECPDF (exon 4). Conservation around the WECPDF motif spans the entire length of exon 4 (see protein sequence of exon 4 in the lower panel of Figure 4.4). Divergence in intron-exon structure is apparent, with five members for the *IVR1* family having 8 exons (*IVR1.1-1A*, *IVR1.2-1A*, *IVR1-4A*, *IVR1-5B*, *IVR1-D* genome), 2 members having 7 exons (*IVR1.1-3B*, *IVR1-3A*), and *IVR1.2-3B* comprising 9 exons.

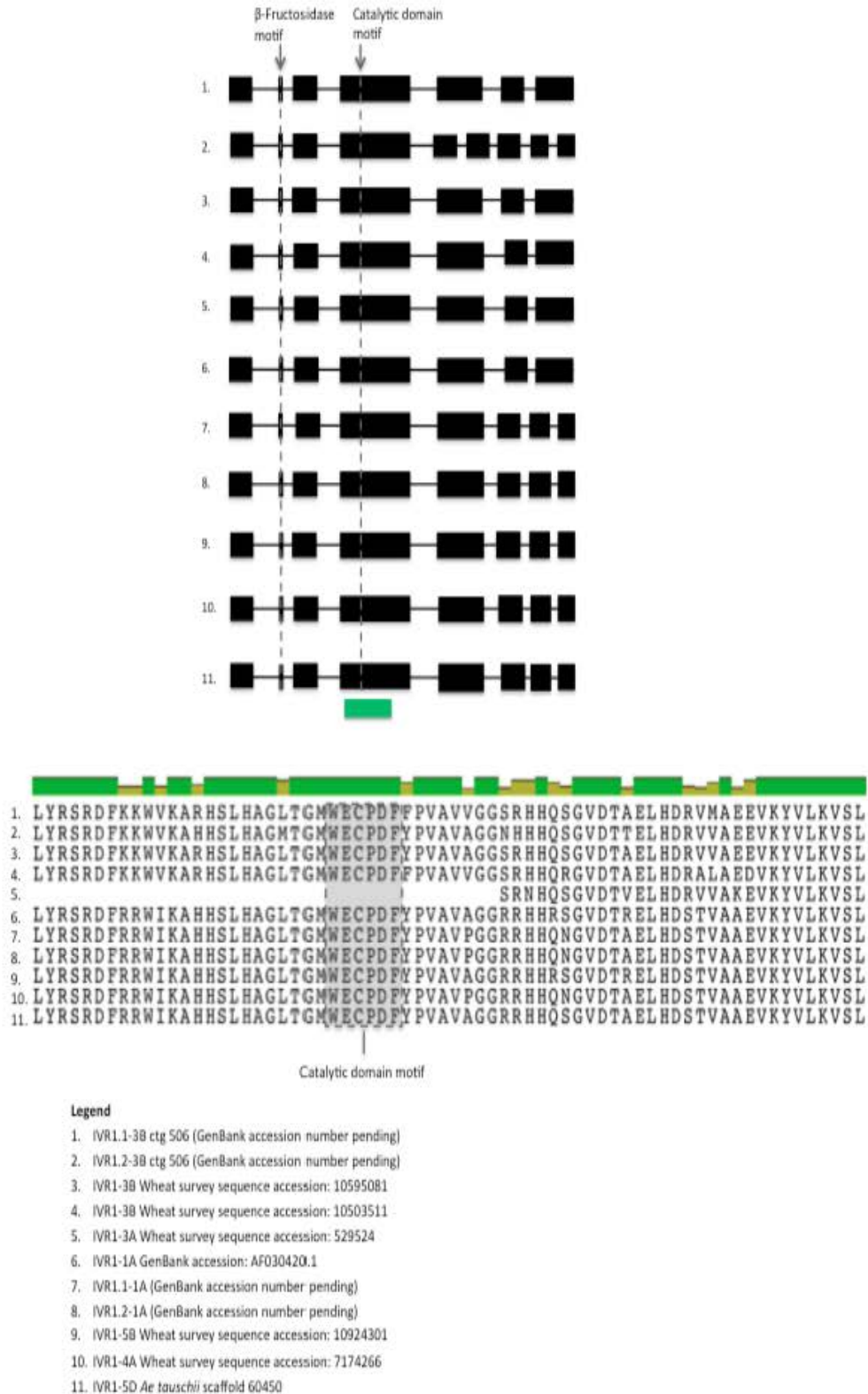


Figure 4.4 Detailed intron-exon structural arrangement of *IVR1* isoforms and a portion of the deduced amino acid sequence from exon 4. In the upper panel of the Figure exons are depicted by black boxes, with exon number ascending from left to right (corresponding to exons 1-9). Isoform intron-exon structure is highly conserved in the region extending from exons 1-5, as are the DPN and WECPD motifs. Divergence is apparent at the 3' end of the gene where the isoforms have either a total number of seven exons (sequence # 1, 3, 4, 5), eight exons (sequence # 7, 8, 9, 10, 11), or nine exons (sequence # 2). The bar

underneath exon 4 indicates the location of the sequences shown in the lower panel. The lower panel of the Figure shows the high degree of conservation that exists between the *IVR1* isoforms at an amino acid level, for the region spanning exon 4. In the protein alignment identical, conserved and analogous regions between the various *IVR1* isoforms are demarcated by the colour of the line seen above the alignment; where the dark green line represents homology and the light olive green line represents a high degree of conservation. Note; the IWGSC wheat survey sequence contig_529524, from which *IVR1-3A* was annotated, contained the amino acid sequence spanning the DPN motif in exon two (not shown) but was missing the portion of the sequence spanning the first half of exon 4 (where the WECPDF motif is usually located). The protein translation of the available portion of contig_529524 showed a high degree of homology with the annotated reference sequence AF030420.1, grounds which in parallel with the conservation of the DPN motif in contig_529524, indicate it is likely to be an *IVR1* isoform.

The evidence presented in Figure 4.4 was based on the screening of individual chromosome arm libraries, as distinct from counting sequence variants. These data suggest that at least eight *IVR1* isoforms exist in wheat, including: *IVR1.1-1A*, *IVR1.2-1A*, *IVR1-3A*, *IVR1.1-3B*, *IVR1.2-3B*, *IVR1-4A*, *IVR1-5B*, and *IVR1-D* genome. From these results we propose that using their chromosomal locations is a useful way to identify the isoforms. In Figure 4.5 the nucleotide sequence of exon 4 (containing the WECPDF motif) is shown and illustrates that *IVR1* isoforms are identifiable by their distributions of SNPs, enabling clear differentiation between the isoforms. We note that our estimate of the number of *IVR1* isoforms in wheat is a minimum estimate since the IWGSC database of survey sequences from wheat chromosomes was incomplete when screened.

```

1. ACGCCTCCAAGTCTTTTCGACCCGGCCAAGAAACGACGTGTGCTCTGGGGCTGGGCCAATGAGTCCGACACCGTCCCGACGACCTCCACAAGGGCTGG
2. ACGCCTCCAAGTCTTTC--GACCCGGCCAAGAA-CGCCGAGTGCTTTGGGGCTGGGCCAATGAGTCTGACACCGTTACCGACGACCCGCCACAAGAGCTGG
3. ACGCCTCCAAGTCTTTCGACCCGGCCAAGAAACGACGAGTGTCTCTGGGGCTGGGCCAATGAGTCTGACACCGTTACCGACGACCCGCCACAAGGGCTGG
4. ACGCCTCCAAGTCTTTCGACCCGGCCAAGAAACGACGAGTGTCTCTGGGGCTGGGCCAATGAGTCTGACACCGTTACCGACGACCCGCCACAAGGGCTGG
5. ACGCCTCCAAGTCTTTCGACCCGGCCAAGAAACGACGAGTGTCTCTGGGGCTGGGCCAATGAGTCTGACACCGTTACCGACGACCCGCCACAAGGGCTGG
6. ATGCATCGAAGTCTTCTATGACCCGGTCAAGAAAGCGCCGCTGCTCTGGGGCTGGGCCAATGAATCCGACACTGTCCCCGACGACCCGCAACAGGGTTGG
7. ATGCATCGAAGTCTTCTATGACCCGGTCAAGAAAGCGCCGCTGCTCTGGGGCTGGGCCAATGAATCCGACACTGTCCCCGACGACCCGCAACAGGGTTGG
8. ATGCATCGAAGTCTTCTATGACCCGGTCAAGAAAGCGCCGCTGCTCTGGGGCTGGGCCAATGAATCCGACACTGTCCCCGACGACCCGCAACAGGGTTGG
9. ATGCATCGAAGTCTTCTATGACCCGGTCAAGAAAGCGCCGCTGCTCTGGGGCTGGGCCAATGAATCCGACACTGTCCCCGACGACCCGCAACAGGGTTGG
10. ATGCATCGAAGTCTTCTTTCGACCCGGTCAAGAAAGCGCCGCTGCTCTGGGGCTGGGCCAATGAATCCGACACTGTCCCCGACGACCCGCAACAGGGTTGG
11. ATGCATCGAAGTCTTCTTTCGACCCGGTCAAGAAAGCGCCGCTGCTCTGGGGCTGGGCCAATGAATCCGACACTGTCCCCGACGACCCGCAACAGGGTTGG

```

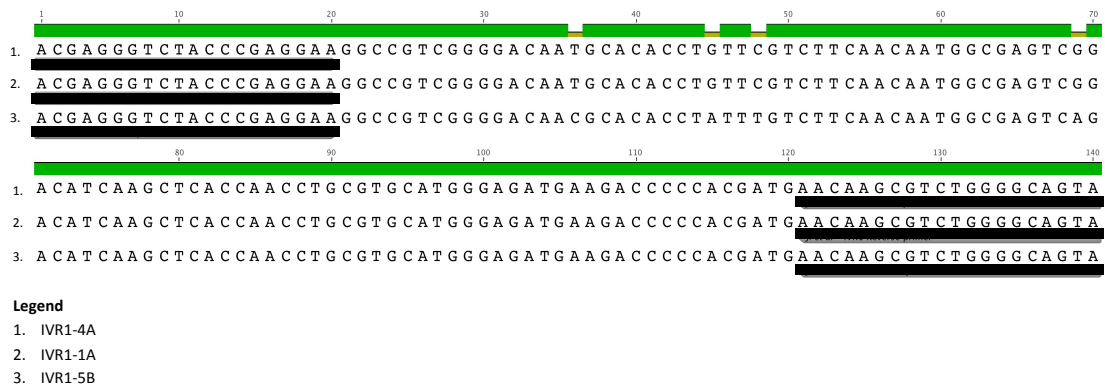
Legend

1. IVR1.1-3B ctg 506
2. IVR1.2-3B ctg 506
3. IVR1-3B Wheat survey sequence accession: 10595081
4. IVR1-3B Wheat survey sequence accession: 10503511
5. IVR1-3A Wheat survey sequence accession: 529524
6. IVR1-1A NCBI accession: AF030420.1
7. IVR1.1-1A GenBank submission ID:1527570
8. IVR1.2-1A GenBank submission ID pending
9. IVR1-5B Wheat survey sequence accession: 10924301
10. IVR1-4A Wheat survey sequence accession: 7174266
11. IVR1-5D *Ae tauschii* scaffold 60450

Figure 4.5 Comparison of wheat *IVR1* isoform nucleotide sequence in exon 4. In the multiple sequence alignment identical, conserved and analogous regions between the various *IVR1* isoforms are demarcated by the colour of the line seen above the alignment; where dark green represent homology, light olive represents a high degree of conservation and red represents low conservation.

Screening of the IWGSC survey sequence, the ctg506 chromosome 3B assembly, and the D genome sequence with PCR primers previously used to study the expression of *IVR1* in wheat (Ji *et al.*, 2010; Koonjul *et al.*, 2005), indicated that several *IVR1* isoforms on different chromosomes were assayed by these primers. As shown in Figure 4.6(a), the primers of Ji *et al.*, (2010) are predicted to assay *IVR1-1A*, *IVR1-4A* and *IVR1-5B*. The primers used by Koonjul *et al.*, (2005) would similarly assay *IVR1.1-3B* and *IVR1-5B* (Figure 4.6(b)).

(a)



(b)

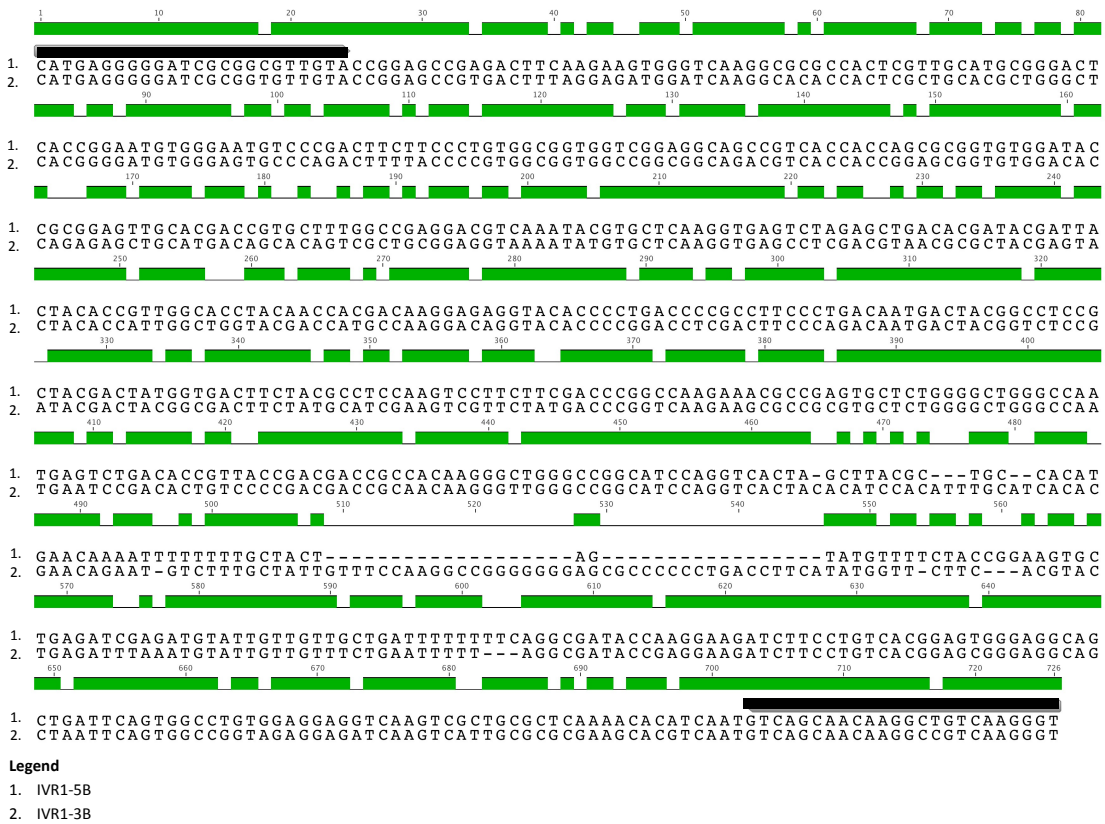


Figure 4.6 (a) Comparisons of nucleotide sequence of *IVR1* PCR products in exon 7, assayed by primers utilized by Ji *et al.*, (2010) for observing *IVR1* gene expression under RSWD stress conditions. The Ji *et al.*, (2010) primer sequences are demarcated in the alignment by a solid black line under the nucleic acid sequence. The forward primer is positioned in the alignment at co-ordinates 1 to 20 (5'-3') and the reverse primer is positioned at the co-ordinates 121-140 (5'-3'). The degree of homology amongst the *IVR1* isoforms is demarcated by the colour of the line seen above the alignment, where dark green represents homology and light olive green represents a high degree of conservation. Note: the Ji *et al.*, (2010) primer sequences is homologous in the *IVR1* isoforms on chromosome 4A, 1A and 5B and therefore lacks specificity to assay a chromosome-specific isoform of the gene. (b) Comparison of nucleotide sequences of *IVR1* PCR products in exon 4, assayed by primers utilized by Koonjul *et al.*, (2005) for observing *IVR1* gene expression under RSWD stress conditions. The Koonjul *et al.*, (2005) primer sequences are demarcated in the alignment by a solid black line above the nucleic acid sequence. The forward primer is positioned in the alignment at co-ordinates 1-24 (5'-3') and the reverse primer is positioned at 703-726 (5'-3'). The degree of homology amongst the *IVR1*

isoforms is demarcated by the colour of the line seen above the alignment, where dark green represents homology and light olive green represents a high degree of conservation. Note the Koonjul *et al.*, (2005) primer sequence is homologous in the *IVR1* isoforms on chromosomes 5B and 3B and therefore, lacks specificity to assay a chromosome-specific isoform of the gene.

4.4.3 Comparative genomics of cell wall invertases in wheat and rice

Within the rice genome sequence eight cell wall invertase isoforms has been identified (Cho *et al.*, 2005; Sturm 1999). The extensive phenotypic analyses associated with genes and regions of rice provided a reference for studies in wheat. Phylogenetic analysis indicated that among the cell wall invertase isoforms in rice, the LOC_Os04g33720 sequence on chromosome 4 was most closely related to the AF030420.1 sequence. The LOC_Os04g33720 gene is preferentially transcribed in anther tissues although there is expression at lower levels (i.e. 30% or less) in other tissues (Sharma *et al.*, 2012). It was of interest to align the 5' upstream regions from the LOC_Os04g33720 gene of rice (-507 to 0) with available upstream regions from the wheat *IVR1* isoforms, *IVR1.1-3B* (-1069 to -562) and *IVR1-D* genome (-1777 to -1356), to identify conserved motifs.

The analysis shown in Figure 4.7 provides evidence that motifs conserved between rice and wheat are located upstream of *IVR1.1-3B*, indicating that these may be significant for anther specific expression of *IVR1* isoforms. Similar results were obtained for *IVR1.2-3B* and *IVR1-D* genome upstream regions aligned with the respective rice region.

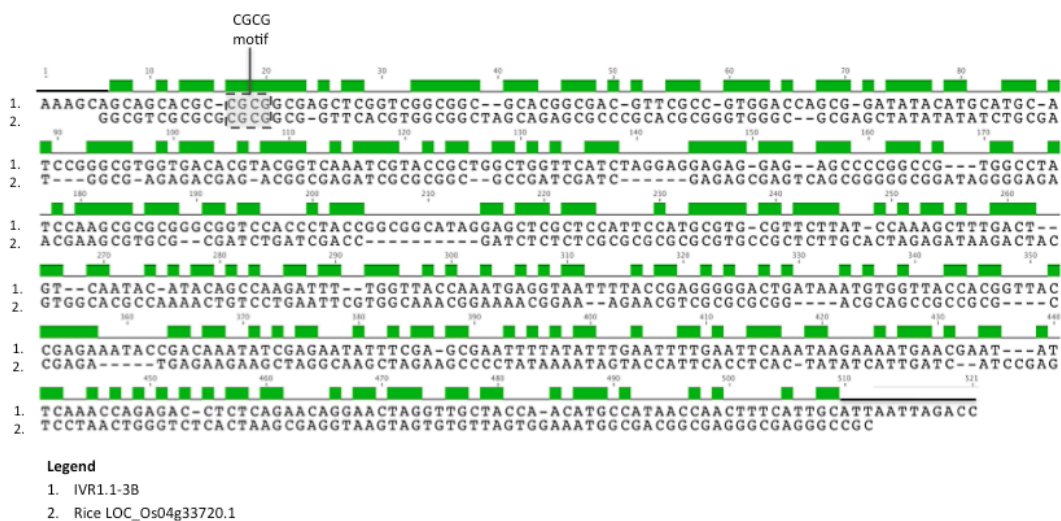


Figure 4.7 Comparison of amino acid sequence in the 5' region upstream of *IVR1.1-3B* in wheat and the most closely related CW-INV in rice LOC_Os04g33720.1 (OsCIN3) (Cho *et al.*, 2005), located on chromosome 4. Scrutiny of identical regions within the alignment

(demarcated by a green line above the homologous regions) show a conserved motif (i.e. the CGCG box highlighted in grey) located in the alignment from residues 17-20. An important pollen-specific promoter motif (AGAAA) is also located upstream from the wheat *IVR1.1-3B* isoform from residues 355-359 (ProFITS site: <http://bioinfo.cau.edu.cn/ProFITS/>, accessed 27 April 2012)

4.5 Discussion

The data in the present paper have further defined the *IVR1* sub-family in wheat at the genome level. Complexity at this level was suggested by Koonjul *et al.*, (2005) and the genome-level analysis of *IVR1* has indicated that the primers used to assay transcripts in water deficit experiments (Ji *et al.*, 2010; Koonjul *et al.*, 2005) actually assay several isoforms from different chromosomal locations. The significance of these findings is that transcription studies need to define the chromosomal location of the isoform being assayed so that suitable molecular markers can be transferred to wheat breeding.

As in cotton (Rapp *et al.*, 2010; Udall *et al.*, 2007), wheat shows homologous gene silencing (Bottley and Koebner 2008; Mochida *et al.*, 2003; Zhang *et al.*, 2008(A)). In an analysis of a panel of 16 wheat varieties, Bottley and Koebner (2008) identified homoeolog non-expression for 15 single-copy genes. The detailed expression profiles of eight genes indicated that only two varieties of the 16 shared the same pattern of silencing and thus the homoeologous silencing observed in wheat is likely to be significant in generating phenotypic variation. In this context it is evident that in relating variation in expression of *IVR1* isoforms in wheat to water deficit stress tolerance, it is important to define the chromosomal origin of the *IVR* isoform being studied in transcriptional analyses.

There is growing evidence to suggest that sucrose conversion by cell wall invertases in the apoplastic space is one of the key rate limiting steps in carbohydrate metabolism in reproductive structures and supplying photosynthetic assimilates for storage (Eschrich 1980). The basis of this dependency is that developing pollen microspores are symplastically isolated and thus, must receive carbohydrates via apoplastic transport in which *IVR1* isoforms play a critical role (Roitsch 1999; Sturm 1999). Impeded apoplastic transport can be attributed to the down regulation of *IVR1* in anthers, which

results in the pollen being starved of the energy substrate, as there is no alternate pathway for the delivery of sucrose. The breakthrough in establishing the central role of CW-INV's in phloem uploading of sugars was in the study of the Nin88 CW-INV gene in tobacco (Goetz *et al.*, 2001), which was specifically expressed in the tapetum and pollen. In their study Goetz *et al.*, (2001) demonstrated that tissue specific suppression of the Nin-88 gene, under the control of its own promoter sequence, resulted in defects in the early stages of pollen development and thus sterile pollen.

In studies examining the selective down-regulation of *IVR1* isoforms under plus and minus RSWD stress conditions in wheat, Koonjul *et al.*, (2005) reported that gene expression was highly specific to individual isoforms both spatially and temporally. Similar conclusions were drawn by (Godt and Roitsch 1997; Sturm 1999). In the Koonjul *et al.*, (2005) study, RSWD stress applied specifically during the period just prior to and following pollen meiosis showed immediate and significant down-regulation of one *IVR1* isoform (*IVR1S*) but no decline in another (*IVR1L*). Conversely, in the control plants (i.e. -RSWD stress) the *IVR1S* isoform also declined over the duration of pollen development while *IVR1L* increased. According to the results shown in Figure 4.6(b), which highlighted the PCR products assayed by the Koonjul *et al.*, (2005) *IVR1* primers, we propose that the two *IVR1* isoforms assayed in this gene expression analysis were located on chromosome 3 (*IVR1.1-3B*) and chromosome 5 (*IVR1-5B*). However, determination of which isoform was located on chromosome 3 and 5 was not possible without access to isoform specific PCR product sequences.

The Koonjul *et al.*, (2005) invertase gene expression analysis also highlighted that two genes (*IVR3* and *IVR5*), closely related to the *IVR1* family, did not show decline in expression under drought stress. The *IVR3* and *IVR5* genes were reported as being members of the CW-INV and V-INV families, respectively. In the work related to the current study, detailed examination of the *IVR3* (Genbank accession: AF030421) and *IVR5* (Genbank accession: AF069309) gene sequences confirmed that *IVR5* belongs to the V-INV family but that *IVR3* is not a CW-INV or a V-INV isoform. Our results showed that the *IVR3* sequence is: (i) highly

divergent from the *IVR1* family both at a primary sequence level and exon structure, (ii) located on chromosome 6A in wheat (IWGSC Survey Sequence contig-4429358), and (iii) a member of the Fructan 1-exohydrolase gene family, with its closest relative being the *1-FEHw1* (Genbank accession: AJ516025) gene in wheat. Similarly our analysis of putative wheat cell wall invertases genes characterized in studies by Ma *et al.* (2012) showed the genes in question (TaCwi's) are not cell wall invertases and also do not belong to the other two invertase family sub groups (vacuolar and cytoplasmic). In wheat the TaCwi genes are also most closely related to the FEH family.

The studies in the present paper provide greater resolution in defining *IVR1* isoforms for expression studies. The comparison in Figure 4.7 showed that the 5'-upstream region for *IVR1.1-3B* (ctg506: 298676kb to 300000kb) had significant similarity to an equivalent region for LOC_Os04g33720 (ch4: 20411920kb to 20412390kb). The shared motifs included the CGCG box and CGACG motif. The CGCG box has been characterized in Arabidopsis (Yang and Poovaiah 2002) and often occurs as multiple elements as is the case for region shown in Figure 4.7. The CGACG motif has been studied in the cis-control of amylase (*Amy3D*) expression in rice (Hwang *et al.*, 1998). Within the *IVR1.1-3B* region the ACTCAT (related to osmoregulation), AGAAA (within a pollen specific promoter) and CATGTG (related to drought response) motifs can be identified (ProFITS site: <http://bioinfo.cau.edu.cn/ProFITS/>) and their presence is consistent with *IVR1* expression being significant in maintaining a supply of hexose sugars to the developing pollen.

Gene duplication is widely recognized as a platform from which gene families can evolve through mechanisms of chromosomal remodelling, locus deletion or point mutation (Wang *et al.*, 2010). The data in this manuscript further support this hypothesis, with regards to the evolution of complex gene families in wheat (derived from a common ancestral gene), by providing evidence of the existence of a multi member *IVR1* sub-family, which has a shared functionality in hydrolysing hexoses in reproductive structures. We propose the sequence divergence that exists amongst the eight *IVR1* isoforms, located on four separate

chromosome arms of the wheat genome, is likely to have causal links to subtle differences in isoform functionality and regulatory mechanisms, particularly in response to changing water availability conditions. It is plausible to suggest that sequence variation in the *IVR1* sub-family provides wheat the flexibility to maintain carbohydrate supply to its developing reproductive structures, thus ensuring viable pollen development even in water deficit stress conditions. Gene expression studies remain the foundation methodology by which the relationship between the sequence variation of the *IVR1* isoforms and their corresponding functional differences can be better understood. The identification of the chromosomal origins of the *IVR1* isoforms undertaken in this study will now enable more accurate transcriptional analysis and improved capacity to relate variation in *IVR1* isoform expression to water deficit stress tolerance in wheat.

Chapter 5: Investigating a Westonia x Kauz mapping population for double haploid lines suitable for studying the effects of reproductive stage water deficit

Chapter contributions:

Generation of ICIM mapping outputs reported in this Chapter involved direct interactions with Jiankang Wang (CIMMYT/Chinese Academy of Agricultural Science, Beijing). Hollie Webster attended the ICIM workshop run by Dr Wang in February 2013 (UWA).

The statistical analysis reported in this chapter also involved a major collaboration with Dean Diepeveen (DAFWA, Department of Agriculture and Food, Western Australia).

The 90K chip data was generated by Matthew Hayden (DEPI, Department of Environment and Primary Industries, LaTrobe University Melbourne).

The analyses reported in this Chapter required a major collaboration with a colleague at Murdoch University, Gabriel Keeble-Gagnere, in the form of handling and storage of the large datasets as well as running analyses to answer questions posed in the biological context.

5.1 Abstract

The work in this chapter characterised a doubled haploid population of 225 lines derived from a cross between Westonia and Kauz (WxK DH), two lines known for contrasting water deficit stress response mechanisms. Phenotype and genome level characterization of the WxK DH population was conducted for genetic map construction and QTL detection of traits underpinning response to reproductive stage water deficit (RSWD) stress. QTL analysis using a WxK DH 90K based SNP chip genetic map (consisting of 9740 markers) showed 15 significant (>LOD 3) marker-trait associations existed for a range of drought

escape traits relating to the number of growth days taken to reach key growth stages during the course of spike reproductive development. The drought escape QTL were found to be highly correlated with each other, and each trait co-located at the same position on chromosomes 5A, 5B, and 5D. The high level of correlation between the traits showed direct selection for just one of traits would maximize the effects on the plant reproductive stage phenotype. Amongst the QTL detected *QtAD5-5B*, which was mapped to chromosome 5B (290-292cM) between markers *wsnp_Ku_c10434_17255840* and *wmc075b*, was found to have the strongest influence on the number of growth days taken to reach the most critically sensitive growth stage of spike development, that being the end of pollen meiosis. The additive effect of *QtAD5-5BL* was found to correspond to a shift of 10 days in the onset of this growth stage, the direction of which in each WxK DH line depended on the parental origin of the alleles inherited for the QTL genotype. The results showed days taken to reach the completion of pollen meiosis could be manipulated by selecting for an easy to assess phenotypic allele variant, so as prevent the occurrence of the development stage coinciding with a known time in the growing season that frequently incurs water deficit stress. Structural and functional characterization of genes located within the *QtAD5-5B* interval (none of which have been previously identified) showed a number of genes have hypothetical function relating to key pollen development processes and/or abiotic stress response pathways. Collectively these results show the suitability of the germplasm and marker suite used to identify two highly contrasting WxK DH lines for inclusion in high-resolution transcriptome studies (reported in Chapter 6).

5.2 Introduction

Based on findings from a number of studies (Fischer 2007; Miralles and Slafer 1995; Reynolds *et al.*, 2005; Shearman *et al.*, 2005) there is now strong evidence to show optimizing reproductive traits in wheat for tolerance to abiotic stress should be elevated to a high priority breeding target. This conclusion was drawn in a review of studies focussed on evaluating reproductive fertility in cereals, the traits contributing to yield potential, and the relative susceptibility of flowering to abiotic stress, particularly drought (Reynolds *et al.*, 2009(A)). In this review it

was suggested that breeding wheat for optimized fertility would be a necessary combative measure to ensure yields increase concomitantly with increasing demands for wheat as a food source. These increases must be achieved in spite of increasingly climatic variability that is now prevailing throughout many of the world's major wheat growing regions.

Given that traits underpinning wheat response to RSWD stress are quantitatively inherited, as are most drought response traits (Tuberosa and Salvi 2006), the discovery of major QTL continues to hold promise for improvement of these traits through marker assisted selection (MAS). The importance for selecting for this trait in rice has resulted in the release of several new rice varieties for growth in India, Bangladesh, and Nepal, specifically for rainfed crop environments that commonly experience water deficit stress mid to late in the growth season when reproductive development is occurring (Kumar *et al.*, 2012).

It is well understood that flowering in plants is the key adaptive trait that confers evolutionary fitness, through ensuring flowering occurs at optimal times for pollination, thus increasing the chance of successful reproduction. A suite of flowering and related reproductive development traits underpin grain yield outcomes in the field, among them being the structure and formation of reproductive organs, panicle morphology, pollen meiosis, pollination, fertilization and seed development (Reynolds *et al.*, 2009(B)). Research into reproductive fitness to date has focussed mostly on the effect of abiotic stress on post-anthesis development (i.e. grain filling/size)(Sinclair and Jamieson 2006; Yang and Zhang 2006). Amongst the reproductive development phases the young microspore stage of pollen development has been shown in wheat and rice to be the most susceptible to water deficit stress (Ji *et al.*, 2010; Saini *et al.*, 1984). RSWD stress induced pollen sterility and the resulting loss of grain set is of particular importance in a West Australian wheat breeding context, as it is in many breeding programs throughout the world that develop wheat cultivars for Mediterranean climates, because of the the high frequency with which water

deficit occurs concomitantly with reproductive development in these climatic regions (Maccaferri *et al.*, 2011).

Studies by Ji *et al.*,(2010) showed that genetic variability also exist for RSWD stress tolerance in wheat and theoretically such germplasm could be used in marker assisted backcrossing programs to breed for this trait. Amongst the germplasm tested in the Ji *et al.*, (2010) study, two lines Halberd (and Australian developed cultivar) and SYN604 (a synthetic wheat developed by CIMMYT), exhibited comparatively higher levels relative to the other lines tested, for resistance to water deficit imposed at the young microspore stage of pollen development. Clearly studies such as this, which identify germplasm showing contrasting sensitivity to RSWD stress, are the necessary first step towards introgressing the trait into commercial germplasm (Dolferus *et al.*, 2011). Achieving this outcome however requires additional steps including identifying large effect QTL driving the trait, demonstrating the QTL are stable across genetic background and environment, and finally introgressing favourable allele variants into elite germplasm in marker assisted backcross programs (Xu 2010). While the discovery of locally adapted germplasm for Australian growing conditions exhibiting tolerance to this developmental phase-specific stress is a promising start there remains a significant gap in understanding the molecular and genetic basis of plant response to RSWD stress. This knowledge gap is the primary constraint to developing new varieties possessing a stable RSWD response ideotype.

A review by Fleury *et al.*,(2010) highlighted that the challenges hampering QTL detection for RSWD stress traits in wheat are similar to those also encountered for QTL studies in most drought related traits per se. In this review Fleury *et al.*,(2010) scrutinized the body of literature relating to drought tolerance QTL discovery in wheat and concluded that a lack of genetic and genomics resources were the primary cause of the relatively slow progress. This conclusion is consistent with a number of reviews summarizing recent advances in molecular approaches for genetic improvement broadly across the *Triticeae* tribe (Borevitz and Chory 2004; Maccaferri *et al.*, 2006; Tuberosa *et al.*, 2002).

Collectively these reviews highlight that while there have been significant advancements in wheat genomics technologies in recent years these advancements are considerably less than what has been achieved in rice (Dolferus *et al.*, 2013). Rice, for example, has a full genome reference and a superior SNP marker platform (as compared to wheat), which can provide resolution to the gene level (McNally *et al.*, 2006). In studies by Schnurbusch *et al.*, (2007) and Alfres *et al.*, (2009) homologous genes in wheat were mapped via synteny and showed that rice remains a useful model species to dissect the genetic basis of drought tolerance traits in wheat. In the context of this thesis rice was the optimal species for benchmarking outputs from this thesis, based on the fact a number of candidate genes of interest to this work, particularly cell wall invertase (*IVR1*) and UDP-glucose-6-dehydrogenase, have been comprehensively characterised and mapped in rice (Crismani *et al.*, 2011; Nguyen *et al.*, 2004; Oliver *et al.*, 2005; Wang *et al.*, 2010).

In addition to the lack of genomics resources Fleury *et al.*, (2010) also highlighted inadequate experimental design as the other major impediment to accurate QTL detection for drought traits in wheat. These constraints relate to poorly defined targets for: growth stages and/or physiological processes of critical drought stress vulnerability, type of drought stress and choice of treatment to mimic its effect, selection of mapping population (population size and variability in phenology), and marker saturation in genetic maps. In terms of understanding the inherent complexities of accurate QTL detection, particularly for RSWD stress tolerance, a considerable amount can be learned from the findings and methods of the numerous studies published on this topic in rice.

Mishra *et al.*, (2013) reviewed all published rice RSWD stress QTL, many of which were claimed by the respective authors to be QTL of major effect on the QTL phenotype. The reported QTL included: chromosome 1, *qDTY1.1* (Vikram *et al.*, 2011) and *QTLgys1.1* (Babu *et al.*, 2003; Kumar *et al.*, 2007), chromosome 3, *qDTY3.1* (Venuprasad *et al.*, 2009; Vikram *et al.*, 2011), chromosome 4 *QTLgys4.1*

(Babu *et al.*, 2003), and chromosome 12 *qDTY_{12.1}* (Bernier *et al.*, 2007; Mishra *et al.*, 2013). Mishra *et al.*, (2013) concluded that most of the reported QTL were in fact only minor and that inaccurate estimation of the association between the trait and markers had occurred and was a result of poor experimental design. These claims were based on the fact many of the reported QTL were not stable across genetic background or environment and thus the findings were likely confounded by large interactions between: genotype x environment, QTL x environment, and QTL x germplasm genetic background.

Similar experimental design flaws are also apparent in the published studies evaluating tolerance to RSWD stress in wheat and are the likely reason that, to date, no QTL for tolerance to RSWD have been identified. A QTL for spike fertility in response to abiotic stress was identified on the long arm of wheat chromosome 1B (Demotes-Mainard *et al.*, 1996). This trait was mapped in response to variation to solar radiation (shading), a form of stress which was reported by the authors as being an equivalent stress to the suite of stresses belonging to the abiotic category, which includes: water deficit, thermal stress (heat and frost), salinity, and nutrient deficiency/toxicity. Based on this premise Demotes-Mainard *et al.*, (1996) inferred alleles underpinning this chromosome 1B spike fertility QTL would likely also be important to RSWD stress response.

The approach to use alternative (i.e. non water deficit) screening methods to evaluate wheat RSWD stress response has also been utilized in other studies, despite the methodology being unproven (Dolferus *et al.*, 2011). Examples include shading stress methods used by (Ji *et al.*, 2011) and salinity stress methods used by (Munns *et al.*, 2010). Comparative analysis of gene expression data generated from studies evaluating a range of abiotic stress types showed that some responses networks are common to a range of abiotic stress types, particularly expression of transcription factors in the early stress detection phase (Rahaie *et al.*, 2013). However, a recent review by Bosco de Oliveira *et al.*, (2013) highlighted increasing evidence to show that while some response pathways and/or mechanisms are common to all abiotic stresses many show a number which are distinctly unique. This finding was demonstrated in both

wheat (Peng *et al.*, 2009) and barley (Ahmed *et al.*, 2013) with respect to the different gene networks and response mechanisms observed between water deficit stress and salinity stress.

Based on these findings it is then plausible to suggest gene expression and QTL analysis studies attempting to evaluate water deficit stress but which do so by imposing 'alternative' abiotic stresses (i.e. to mimic water deficit stress), reduce their power to detect true associations between the trait and the gene(s) and/or allele(s) driving it. For QTL studies confounded phenotype data reduces the accuracy with which: (1) true associations between markers/alleles and traits is detected, (2) significance level is estimated for the detected QTL (Xu 2010). Essentially these errors would be introduced due to the fact that 'off-target' alleles (i.e. non water deficit stress related) would be expressed in response to the 'mimic' water deficit treatment.

Accurate detection of genomic regions associated with response to RSWD stress requires confounding influences from environmental effects, such as variability in stress onset, duration, and intensity, to be excluded (Tuberosa and Salvi 2006). Because RSWD stress in wheat has been found to be at critical levels of sensitivity during the 10-15 days prior to anthesis (Ji *et al.*, 2010) glasshouse conditions provide the necessary infrastructure to minimize environmental variability, and in addition the capacity to apply water deficit stress for the correct growth stage, duration, and intensity. Final consideration should also be given to how to apply water deficit stress to individual lines in mapping population, many of which are known to exhibit a large range in flowering time (Langridge *et al.*, 2006). This issue however, can be dealt with either by developing tailored methods to impose water deficit stress on plants within a mapping population specific to their individual growth stages or by selecting a subset of lines which have very similar phenology and as such can have water stress imposed on upon them at the same time.

Understanding of the genetic determinants underpinning RSWD stress response in wheat is essential, to ensure a more targeted and successful approach to breed

new cultivars for this this stage specific trait is achieved. One such approach explored in this thesis is the construction of a genetic map for a biparental population Westonia x Kauz (WxK), of parents with contrasting RSWD stress tolerance phenotypes. The WxK DH genetic map was constructed in this work to enable QTL detection and identification of allelic variants in these regions with a causal link to variability observed in the trait. The outputs of this chapter provided the biological material to carry out high resolution spike transcriptome analysis for the network of genes involved in carbohydrate metabolism.

5.3 Materials and methods

5.3.1 Marker analysis and WxK DH genetic map construction

The WxK DH genetic map constructed by (Zhang *et al.*, 2013) using the wheat 9k chip was used as an anchor to construct a high resolution 90k chip map with the inclusive composite interval mapping ICIM software described by (Li *et al.*, 2007). The 9K chip genetic map consisting of 199 microsatellite (SSR), 2378 single nucleotide polymorphism markers, was saturated with a further 7162 SNP markers using ICIM (Li *et al.*, 2010) to provide a map which included traits for QTL analysis. The 90K chip genotyping arrays comprised 91,829 SNP, of which 261 were Infinium I (two probes per SNP) and 91,568 were Infinium II (one probe per SNP). Of the 91, 829 SNP, 81, 587 produced functional assays, which represented an 89% conversion rate (Cavanagh *et al.*, 2013; Wang *et al.*, in press).

Methods relevant to this chapter including plant material and growth in the glasshouse, characterization of the WxK DH population, experimental design and molecular protocols are outlined in detail in Chapter 3 (General Material and Methods).

5.3.2 ICIM QTL analysis

QTL analysis utilized the inclusive composite interval mapping (ICIM) approach, which was applied using the ICIM QTL mapping software package (Li *et al.*, 2007). QTL screens followed a stepwise regression of simultaneous consideration of all marker information (<http://www.isbreeding.net/>). The

interval for QTL analysis was 1.0 cM, which utilized a stepwise regression probability of 0.01. A QTL with a minimum LOD \geq 3.0, and a contribution rate of 10 % was defined as significant in the WxK genetic map.

5.4 Results

5.4.1 A genetic resource for defining doubled haploid lines for detailed analysis of response to reproductive stage water deficit stress

A total of 9740 molecular markers comprising in part the wheat 90K SNP chip (Wang *et al.*, in press) were used to assay 225 doubled haploid lines from the WxK DH population. The 90K SNP assay on the WxK DH lines was carried out by Dr Matthew Hayden (Victoria DPI). The assay provided the basis for constructing a high density WxK DH genetic map based an anchor map from a lower density WxK DH genetic map published by Zhang *et al.*, (2013) constructed using the wheat 9K SNP chip.

The 90K chip markers were integrated into the WxK DH 9K chip genetic map without removing markers that were co-segregating. Map construction was carried out in collaboration with Dr Jiankang Wang (CIMMYT/CAAS, Beijing) using the inclusive composite interval mapping software package (Li *et al.*, 2007). A whole genome alignment of the final iteration of the WxK DH 90K chip based genetic map against the MAGIC 9K SNP genetic map (Figure 5.1 presents chromosome 5B from the alignment) showed minimal crossover in marker order between the two maps. This result confirmed there is a high level of consensus in marker ordering between the respective maps and validated the suitability of the WxK DH 90K chip genetic map for further studies. As expected, a high degree of marker alignment was evident between the WxK DH 90K chip based map and 9K chip based genetic map (alignment not shown).

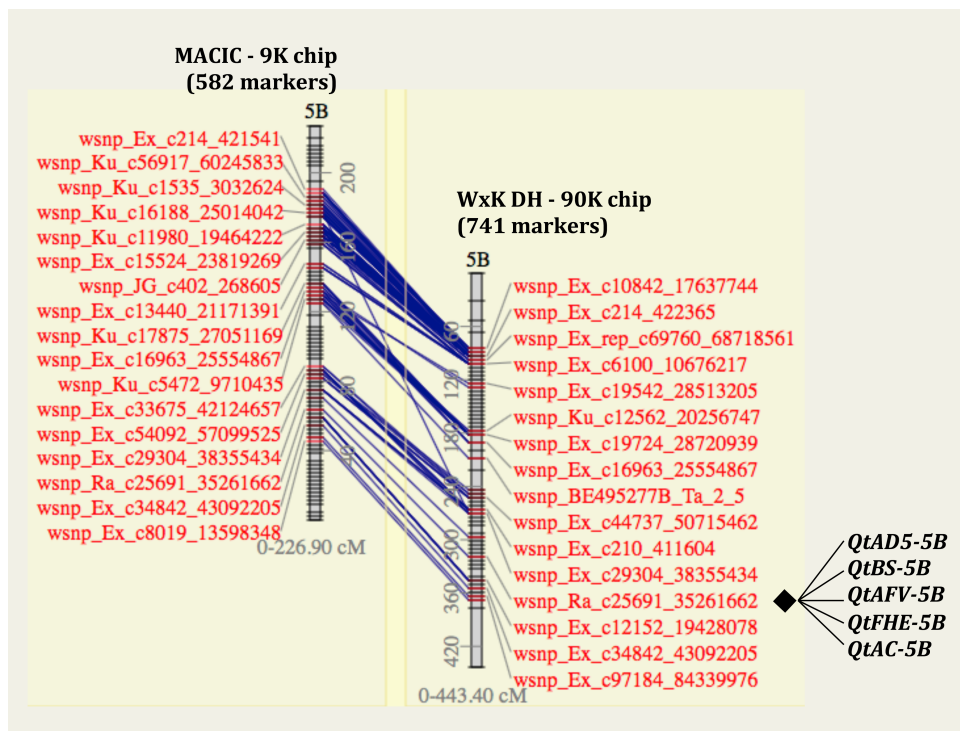


Figure 5.1 The WxK DH 90k chip based genetic map (right) aligned to the wheat composite SNP genetic map (left). The wheat composite SNP map is based on the MAGIC wheat 9k chip map construction by Cavanagh *et al.*, (2013). Markers names highlighted in red have correspondence between the two maps. The incidence of a small number of markers showing disagreement (denoted by crossover of blue lines in alignment) is most likely due to repetitive DNA sequences. The traits mapped as QTL are: *QtAD5-5B* growth days after sowing taken to reach the penultimate internode auricle distance 5cm growth stage, *QtBS-5B* growth days to boost swollen, *QtAFV-5B* growth days after sowing taken to reach the awns first visible growth stage, *QtFHE-5B* growth days after sowing taken to reach the full head (spike) emerged growth stage, and *QtAC-5B* growth days after sowing taken to reach the anthesis commenced growth stage. The presentation of the genetic maps in CMap (version 0.16) does not, for the sake of clarity, present all the molecular markers in a single image. The full details of the maps shown, at a higher levels of magnification can be obtained from the CMap location at the website, <http://ccg.murdoch.edu.au/cmap/ccg-live>. The CMap ID's for the maps shown in the figure are: left - Wheat2 SNP map 2013, right - Westonia x Kauz9740.

The high-density map provided the basis for associating the phenotypes, assessed in the glasshouse and the field (Katanning 2010) with specific chromosome regions. The traits measured included number of growth days to reach key growth stages including: penultimate internode auricle distance at 5 cm in length (AD5), boot swollen (BS), awns first visible (AFV), full head emerged (FHE), and anthesis commenced (AC). Additionally, phenotype data were also collected for plant height (seedling and at maturity), and leaf waxy/glaucousness (for example see Appendix III, Figure 8.8). The AD5 phenotype was of particular interest because findings published in a study by Ji *et al.*, (2011) indicated this was a useful phenotypic marker corresponding to the

end of pollen meiosis, a stage of plant development reported to be critically sensitive to RSWD stress. Validation of AD5 as an indicator of pollen development is shown in Figure 5.2, for one of the lines analysed in this study (DH D02-105). The Figure also presents a summary of the method undertaken to generate a temporal profile of auricle distance and thereafter link it to pollen cell division stage, rate of soil drying upon withholding water from pots, and extent of tissue dehydration. The analytical methods underpinning Figure 5.2 were first undertaken during the experimental design and refinement phase of the project on the WxK DH parents (Westonia and Kauz) and were subsequently repeated after selection of the two DH lines chosen for experimentation in the two consecutive years of RSWD stress experiments in 2011 and 2012.

In the lower portion of Figure 5.2 cytological images are shown for acetocarmine stained pollen dissected from anthers incrementally at four developmental stages extending throughout the course of auricle distance elongation of the penultimate internode. Consistent with the findings of Ji *et al.*, (2011) these images confirm that at AD5 pollen microspore mother cells (MMC's) are undergoing meiosis II. Variation of two to three days was observed for the exact number of growth day after sowing (DAS) at which pollen meiosis II was completed but all samples evaluated from the AD10 growth stage had reached formation of the tetrad of haploid microspores. These cytological observations formed the basis for the design of RSWD stress experiments in that they enabled the identification of two fundamental phenotypic targets (and corresponding DAS), including: (1) the stage at which water should be withheld from the plants, and (2) the stage at which after having water withheld (under controlled atmospheric conditions) plants should have reached moderate levels of water stress (-11 to -14 Bar, see Chapter 3 General Materials and methods for explanation of determination of water stress grades and corresponding Bar values). Gene expression data accumulated from the RSWD stress experiments is presented in Chapter 6.

In the upper portion of Figure 5.2 the line graph shows that on average the soil volumetric water content (SVWC) in both the plus and minus RSWD stress

designated pots at the start of the stress treatment (day zero) was between 21.5% and 23.5% in the 2011 and 2012 experiments respectively. By day three of the RSWD stress treatment, when developing spikes were undergoing pollen meiosis II (as visualised by the AD5 growth stage), the SVWC had dropped to between 4.5% (2011) and 11.5% (2012). No significant difference in the equivalent temporal SVWC comparison in the –RSWD stress pots was observed.

Leaf turgor (LT) of the flag leaf was used as the measure of tissue hydration but because this was a destructive measurement and because significantly reducing the photosynthetic capacity of the plants (by sampling too many leaves) would have introduced confounding effects on the RSWD stress experiment, LT was not measured at daily intervals throughout the RSWD stress treatment interval, but rather once at RSWD stress treatment commencement and once upon completion. Hence, the linear trend between LT at RSWD stress treatment start (day zero) and end (day ten) was used to extrapolate the level of tissue hydration in the -RSWD stress plants for the duration of the experiment in between the respective sampling points. Figure 5.2 shows that at day zero of the RSWD stress treatment there was no significant differences in LT between -RSWD and +RSWD stressed plants in 2011 or 2012. For the two years of data the range in average LT at the start of the treatment was 5.5-7 Bar. This range is consistent with tissue hydration levels conferring adequate water status to support all normal plant function (See Figure 3.3 in Chapter 3). At the onset of AD5 (day 2.5-5 days under RSWD stress) the extrapolated level of water status for the +RSWD stressed plants in 2011 and 2012 respectively can be deduced from Figure 5.2 as being in the range from -10 to -14 Bar and -10 to -13.5 Bar, threshold ranges of which span the mild (-10 to -12Bar) to moderate (-12 to -14Bar) water stress categories. The extrapolated LT levels for the –RSWD stress plants (2011 and 2012) at the same time point show no significant difference relative to their day zero levels.

Collectively the results presented in Figure 5.2 show that for the germplasm evaluated in this study: (1) AD5 can be used with a high degree of certainty as a visual marker for pollen meiosis II, (2) to achieve levels of soil dehydration that

confer moderate water deficit stress by the AD5 growth stage water must be withheld from the potted soil at AD0. Key aspects of this experimental design included accounting for variation in cultivar maturity, atmospheric conditions in glasshouse facilities, and hydraulic properties of the potted soil medium used.

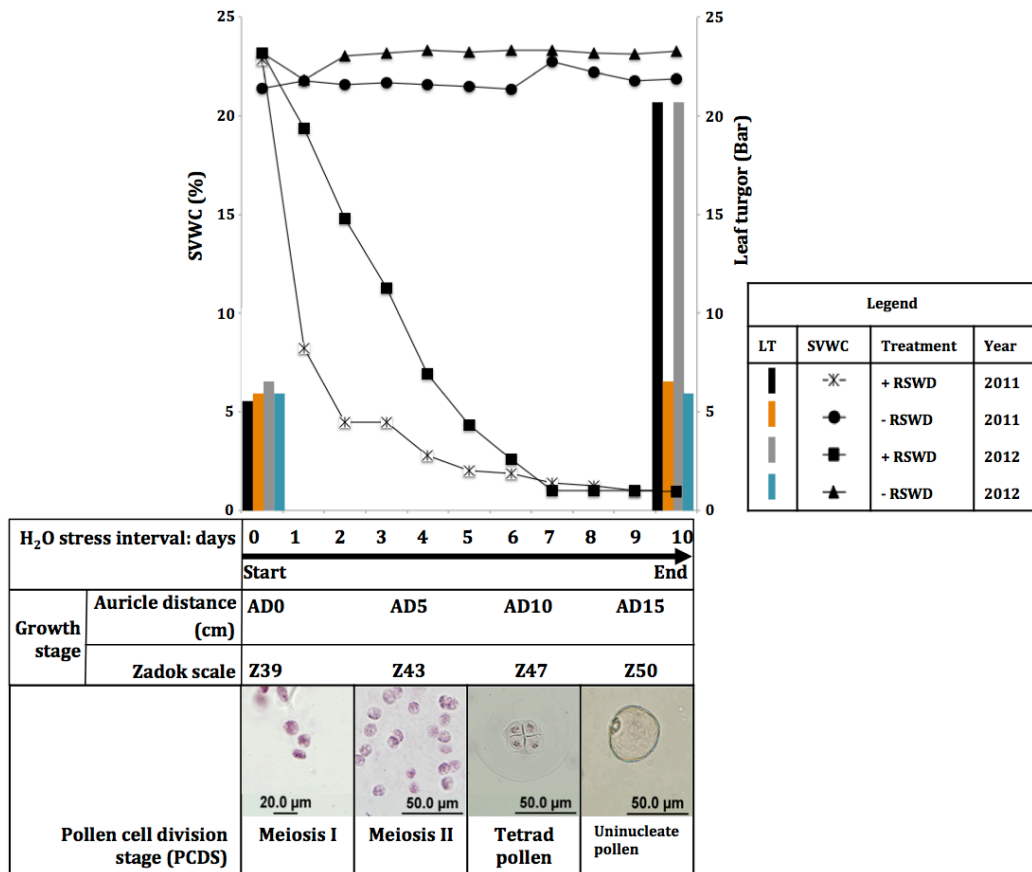


Figure 5.2 Data from the 2011 and 2012 glasshouse experiments summarizing the benchmarking of penultimate internode auricle distance (AD) phenotypes against pollen cell division stage (PCDS), rate of soil drying down in response to withholding water during reproductive stage development (SVWC), and the level of leaf turgor (LT) conferred by contrasting watering regimes (+/- RSWD stress). The central portion of the Figure presents auricle distance (ADcm) of the penultimate internode as a scale of reproductive stage development in the spike (for each AD stage the equivalent Zadoks classification is also listed directly below the AD value). Positioning of the AD/Zadoks scale in the centre of the Figures provides the means to link SVWC and LT data (both presented in the upper portion of the Figure) with PCDS data presented in the lower portion of the Figure. Results comprising the Figure were obtained from evaluating samples collected over the course of the ten day RSWD stress interval in the two successive years of experiments. The results shown relate to WxK DH line D02-105 and are representative of similar analyses for Westonia, Kauz and WxK DH line D08-299 (data not shown). The inset to the right of the Figure shows the colour and shape scheme used differential results comprising the graph in the upper portion of the Figure, as they relate to the specific response variable measured (LT and SVWC), watering regime imposed (+/- RSWD stress), and year of sample collection (2011 and 2012). On the graph average SVWC (%) data is plotted as a trend line against the left side y-axis and average LT (Bar) data as a histogram against the right side y-axis. Both SVWC and LT share the same x-axis, which corresponds to the daily increments comprising the ten day RSWD stress interval that commenced at AD0cm and concluded at AD15cm. The cytological images of developing pollen represent the most consistently observed PCDS at auricle distances ranging from

AD0cm to AD15 (Z39-Z50). The linked AD-PCDS results are aligned to the x-axis of the upper portion of the Figure to show the temporal profile of development from AD0cm (at RSWD stress interval day zero) to AD15cm (at RSWD stress interval day ten). Cytological data show that the AD5cm marker provides the phenotypic target at which mild to moderate RSWD stress must be achieved and the temporal trend line for rate of soil drying down upon withholding water show the number days taken to reach the target soil LT and SVWC is 2.5-5 days.

Phenotype data collected in the glasshouse and the field (Katanning) in 2010 was subjected to principal component analysis (PCA) (Figure 5.3) to identify possible biases in the distribution of measured phenotypic traits in the population. No particular bias was evident and over 50% of phenotypic variation was explained by two dimensions in the PCA, those being 36% and 18% attributable to and PC1 and PC2 respectively. Results from analysis of variance (ANOVA) on the phenotype data (Table 5.1) confirmed there was significant variation between the lines comprising the WxK DH population with respect to morphological traits observed in the field (DAFWA Katanning research station) and phenological traits observed in the glasshouse (Murdoch University). Plant height at maturity (MH) measured in the glasshouse was the exception being found to show no significant difference at a statistical level in the ANOVA despite being identified in the PCA as one of the major traits underlying phenological variability in the first dimension of the PCA.

In Figure 5.3 each of the 225 WxK DH lines represents one data point, the individual values for which are based on a linear combination of each of the phenotyped traits. The loading values (data not shown) are the linear regression co-efficient of each of the traits for a given dimension in the PCA. The larger the co-efficient for a trait the more important that trait is in explaining the variation in that particular dimension. The stepwise progression in calculating each new dimension in the PCA involved the same formula as the one prior except that variation explained by the previous PC dimension was removed. Figure 5.3 shows PC1 is predominantly affected by height related traits (i.e. height at fourth leaf stage, tillering, and anthesis) and collectively these positively correlated traits accounted for the greatest level of variation in the phenotyped population at 36%. PC2 shows phenological staging attributes (i.e. number of growth day after sowing to reach AD5, BS, AFV, FHE and AC) are also strongly positively

correlated and collectively these account for another 18% of the phenotypic variation.

The above results are consistent with trends in double haploid populations for significant variation in plant height and phenology, the latter often being 30 days or more (Reynolds *et al.*, 2009(B)). In the WxK DH population plant height ranged from 25-120cm and growth days taken to reach anthesis-commenced was between 45-120 DAS. Because different phenological stages show different sensitivity to environment factors the large variation in phenology of the WxK DH population means that the stage of interest to this study, pollen development (particularly pollen meiosis), could potentially have been subject to significant GxE interaction had consideration not been given to robust experimental design, specifically DH line selection, timing of RSWD stress, and control of environmental conditions in the glasshouse.

Because this study was concerned with identifying and measuring phenotype QTL associated with RSWD stress it was important to try and prevent putative QTL interacting with the growth environment and genetic background (particularly genes of major effect) of the WxK DH population. To ensure environmental variation did not impede the statistical detection of minor gene effects contributing to possible RSWD stress QTL the water deficit studies were conducted in glasshouse conditions and repeated as such over two successive years. Similarly steps were also taken to ensure the major effect genes contributing to plant height (*Rht*) and phenology (*Ppd* and *Vrn*) did not confound detection of minor gene contributing to the RSWD stress response trait.

Initially a subset of 29 WxK DH lines belonging to a defined phenological maturity and height category were identified as potential candidates for the RSWD stress experiments. This suite of lines was then characterized for their genotype in order maximize genetic differences within the defined phenotypic group for plant height and phenology. Figure 5.3 is annotated to show two WxK DH lines identified as belonging to the target phenotypic group, their strong

phenotypic similarity being evidenced by their close proximity in both the PC1 and PC2 dimension of the principal component analysis plot.

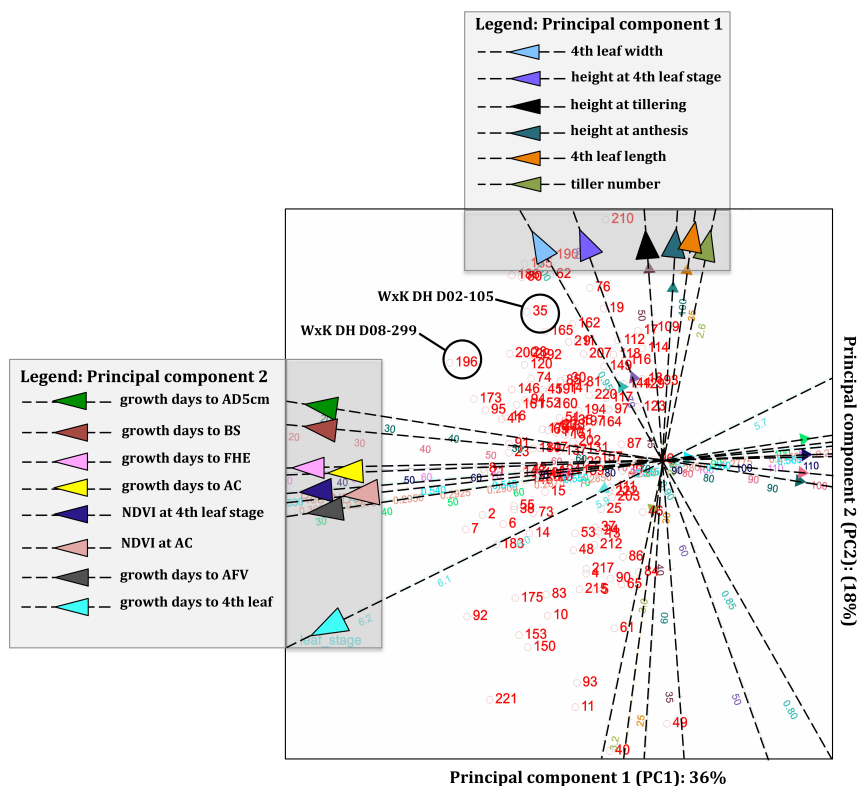


Figure 5.3 Principal component (PC) biplot to evaluate how all traits phenotyped in the WxK DH population relate. The phenotype data point for each WxK DH line evaluated is represented by a red numerical value. The first two principal components of the analysis, PC1 and PC2, are plotted on the x-axis and y-axis respectively and the percentage of variability explained by the respective PC's is listed at the end of the axis titles. The horizontal and vertical position of each WxK DH data point provides a measure of where the data points are situated in the spectrum of phenotypes for the traits associated with PC1 and PC2. Broken black lines overlaid with a coloured triangle represent each of the phenotyped traits correlating in PC1 and PC2 and the angle of these lines relative to each other shows how closely they relate. A legend is presented for PC1 on the upper x-axis and for PC2 on the left y-axis, both of which provide a colour scheme to delineate each of the traits that strongly correlate in the given PC dimension. The data points encircled in a black line represent two WxK DH lines, D08-299 and D02-105, identified during preliminary phenotyping as belonging to the target plant height and phenological maturity group. Based on the strong phenotypic relatedness of these two WxK DH lines they were deemed strong candidates for downstream RSWD stress response gene expression analysis.

The details of the phenotypic traits of interest are provided in Table 5.1. The mapping of these traits as QTL to the WxK DH genetic map, indicated 18 were linked to at least one of the phenotyped traits (Table 5.2). Markers *w SNP_Ku_c10434_17255840* and *wmc075b* were found to co-segregate with traits AD5, BS, AFV, FHE, and AC on chromosome 5B. These observations are consistent with those reported by Zhang *et al.*, (2013). The co-location of QTLs

QtAD5, *QtBS*, *QtAFV*, *QtFHE* and *QtAC* on chromosome 5A, 5B and 5D suggested the traits are strongly correlated and that water deficit stress imposed in the growth stage interval between AD5 and AD10 period of head development would maximize the effects on the plant reproductive stage phenotype.

Table 5.1 Statistical evaluation of phenotyped traits in the WxK DH population. Location: site at which the material phenotyped was grown and phenotype data collected. Trait name: trait on which data was collected, Sample size: number of DH lines on which phenotype data was collected for a particular trait, Mean: mean of the phenotypic trait, Variance: variance of the phenotypic trait, Standard error: standard deviation of the phenotypic trait, P-value: measure of significant difference. Trait name abbreviations: AD5 -Auricle distance 5cm, BS – boot swollen, AFV – awns first visible, FHE – full head emerged, AC –anthesis commenced, MH – mature height.

Location	Trait type	Trait Name	Sample Size	Mean	Variance	Standard error	P-value
Field (Katanning)	Morphological	Seedling vigour	216	9.5	5	2.24	0
		Seedling height	218	2.26	0.54	0.74	0
		Seedling leaf waxiness	212	2.15	0.72	0.85	0
Glasshouse	Phenological	AD5	225	72.51	677.34	26.03	0
		BS	225	76.01	630.74	25.11	0
		AFV	225	82.03	596.38	24.42	0
		FHE	225	87.88	586.02	24.21	0
		AC	225	93.86	561.62	23.70	0
	Morphological	MH	225	80.21	367.73	19.18	0.32

Table 5.2 Summary of QTL identified from inclusive composite interval mapping (ICIM). Trait name: trait on which data was collected, Chromosome: chromosome on which the QTL was mapped to, Position: scanning position in cM on the chromosome (see also Figure 5.1), Left marker: name of the left-side marker of the identified QTL, Right marker: name of the right-side marker of the identified QTL, LOD: LOD score caused of the QTL, PVE (%): phenotypic variation explained by QTL at the current scanning position, ADD: estimated additive effect of QTL at the current scanning position. Trait name abbreviations: SH - seedling height, SCW - Seedling cuticle waxiness, AD5 -Auricle distance 5cm, BS - boot swollen, AFV - awns first visible, FHE - full head emerged, AC -anthesis commenced, MH - mature height.

Trait	QTL	Chromosome	Position (cM)	Left marker	Right marker	LOD	PVE (%)	ADD
SH	<i>QtSH-G-1D</i>	1D	131	CAP8_c1305_148	Kukri_c33670_506	3.18	4.79	-0.16
	<i>QtSH-G-4B</i>	4B	84	Tdurum_contig64772_417	2Rht1	3.76	5.67	0.18
	<i>QtSH-G-4D</i>	4D	236	tplb0027d07_1388	2Rht2	10.96	24.57	-0.36
	<i>QtSH-G-5D</i>	5D	157	Excalibur_c34793_1260	gwm292b	4.13	6.18	-0.18
SCW	<i>QtSCW-K-3A</i>	3A	193	wmc264	WMC428	20.95	34.78	-0.50
AD	<i>QtAD5-5A</i>	5A	313	wsnp_Ex_c5598_9854423	wsnp_Ex_c24577_33826666	24.41	23.41	-12.64
	<i>QtAD5-5B</i>	5B	292	wsnp_Ku_c10434_17255840	wmc075b	18.25	16.40	-10.52
	<i>QtAD5-5D</i>	5D	157	Excalibur_c34793_1260	gwm292b	20.53	18.87	11.32
BS	<i>QtBS-5A</i>	5A	313	wsnp_Ex_c5598_9854423	wsnp_Ex_c24577_33826666	22.79	22.18	-11.87
	<i>QtBS-5B</i>	5B	292	wsnp_Ku_c10434_17255840	wmc075b	17.31	16.23	-10.09
	<i>QtBS-5D</i>	5D	157	Excalibur_c34793_1260	gwm292b	19.45	18.60	10.84
AFV	<i>QtAFV-5A</i>	5A	313	wsnp_Ex_c5598_9854423	wsnp_Ex_c24577_33826666	22.54	22.16	-11.54
	<i>QtAFV-5B</i>	5B	292	wsnp_Ku_c10434_17255840	wmc075b	16.68	15.44	-9.58
	<i>QtAFV-5D</i>	5D	157	Excalibur_c34793_1260	gwm292b	20.25	19.41	10.77
FHE	<i>QtFHE-5A</i>	5A	313	wsnp_Ex_c5598_9854423	wsnp_Ex_c24577_33826666	22.86	21.88	-11.37
	<i>QtFHE-5B</i>	5B	292	wsnp_Ku_c10434_17255840	wmc075b	17.57	15.93	-9.64
	<i>QtFHE-5D</i>	5D	157	Excalibur_c34793_1260	gwm292b	21.48	20.24	10.90
AC	<i>QtAC-5A</i>	5A	314	VPA	wsnp_Ex_c214_421541	21.04	20.93	-10.86
	<i>QtAC-5B</i>	5B	292	wsnp_Ku_c10434_17255840	wmc075b	14.64	13.82	-8.79
	<i>QtAC-5D</i>	5D	157	Excalibur_c34793_1260	gwm292b	19.86	19.76	10.55
MH	<i>QtMH-4A</i>	4A	39	cfD002b]	wsnp_JD_c1796_2496653	2.77	1.89	-2.64
	<i>QtMH-4A</i>	4B	99	RAC875_rep_c117526_220	wsnp_Ra_c4418_8012732	38.53	38.57	11.94
	<i>QtMH-4B</i>	4D	239	tplb0027d07_1388	2Rht2	34.84	36.64	-11.58
	<i>QtMH-4D</i>	5B	292	wsnp_Ku_c10434_17255840	wmc075b	5.41	3.80	-3.73
	<i>QtMH-5B</i>	5D	159	VrnD1	RFL_Contig1091_1538	3.49	2.40	2.99

5.4.2 Selection of two double haploid lines from the Westonia x Kauz double haploid population for water deficit stress studies

In order to select two WxK DH lines for further study, steps were taken to ensure the major effect genes contributing to plant height (*Rht*) and phenology (*Ppd* and *Vrn*) did not confound detection of minor gene contributing to the RSWD stress response trait. A set of 29 lines identified with the following phenotype: anthesis commenced (AC) 65 – 84 days after planting, height 81 – 100 cm. Among the 29 lines, 8 were selected with a time to AD5 within 7 days of each other. Figure 5.3 is annotated to show two WxK DH lines identified as belonging to the target phenotypic group, their strong phenotypic similarity being evidenced by their close proximity in both the PC1 and PC2 dimension of the principal component analysis. To maximize genetic differences within this confined phenological group, the two lines indicated had maximum differences in terms of marker alleles. The line D02-105 showed 70% Kauz alleles and 30% Westonia alleles while D08-299 showed 65% Westonia alleles and 35% Kauz alleles. Genotypic selection of D02-105 and D08-299 using the 90K SNP chip was validated by genotype clustering results produced from principal component analysis (PCA Figure 5.4).

The PCA of marker segregation produced a ranking of correlation for genotype 'likeness' between the 225 WxK DH lines. In Figure 5.4 the pattern of marker segregation (i.e. alleles of markers present in each DH line) showed two distinct groups of allelic inheritance from the DH parents, those being either Westonia-like or Kauz-like in the distinct clustering of data points along the PC1 dimension (x-axis) between the ranges -1.6 to -0.5 (lower left cluster) and -0.3 to 0.7 (upper right cluster). The small out group of clustered data points in the lower right corner of Figure 5.4 was the result of marker information consistently missing from a set of 12 WxK DH lines in the 90K SNP chip analysis. When observed in the PC2 dimension (y-axis) the two WxK DH subfamilies show a more continual spread in genotypic variation, as opposed to clustering into further sub-families. The spread of data in the PC2 dimension is consistent with random segregation of markers.

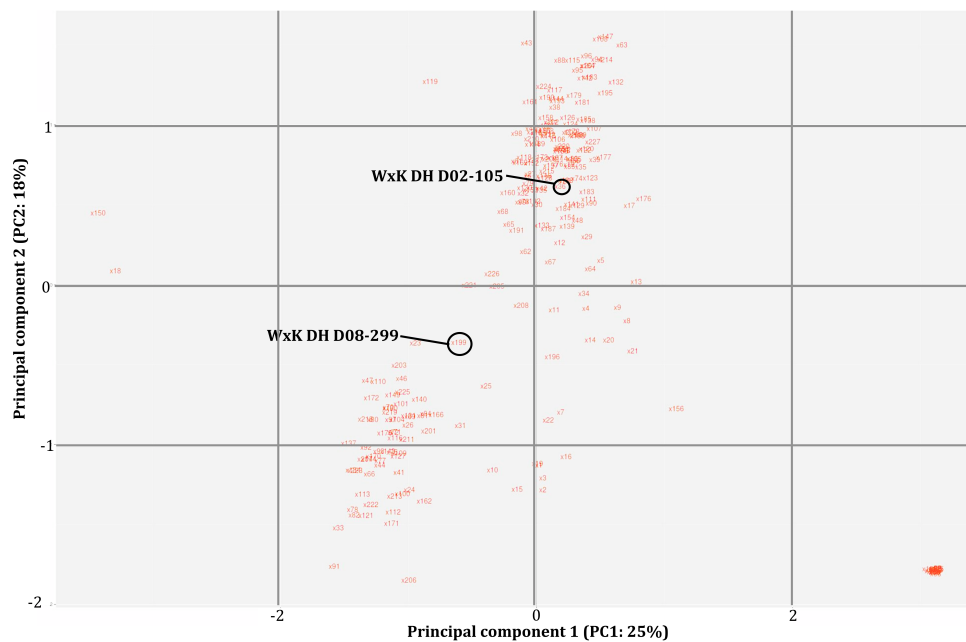


Figure 5.4 Principal component (PC) biplot to evaluate the distribution of alleles assayed using the 90K chip within the WxK population. The marker assay data point for each WxK DH line evaluated is represented by a red numerical value. The first two principal components of the analysis, PC1 and PC2, are plotted on the x-axis and y-axis respectively and the percentage of variability explained by the respective PC's is listed at the end of the axis titles. The horizontal and vertical position of each WxK DH data point provides a measure of where the data point is situated in the spectrum of genotypes associated with PC1 and PC2. The data points encircled in a black line represent two WxK DH lines (D08-299 and D02-105) that were identified during preliminary phenotyping as being strong candidates for downstream RSWD stress response gene expression analysis. D08-299 and D02-105 are positioned in two separate genotype clusters: D02-105 at 0.18 (PC2) and 0.62 (PC1), and D08-299 at -0.61 (PC1) and -0.35 (PC2). Positioning of the two lines in separate cluster shows they have contrasting genotypes, thus validating their selection for water deficit stress response gene expression analysis.

To validate the marker based analysis for line selection, a trait-based approach was also used in the selection of D02-105 and D08-299. It was expected selection for contrasting RSWD response phenotypes would change the allelic frequencies of segregating plus or minus alleles at possible RSWD response QTL, thus affecting the pattern of QTL inheritance, the magnitude of the QTL and its direction. Although variation for water deficit stress tolerance is continuous in mapping populations based on the traits quantitative nature, by ruling out intermediate phenotypes the extreme phenotypes can be distinguished. The frequency of the plus alleles and minus alleles was then expected to increase and decrease accordingly in the high and low contrasting RSWD stress response phenotypes. Grain set in response to RSWD stress (an inverse phenotypic marker for RSWD stress induced pollen sterility) was used as the target

phenotype to measure RSWD stress response and as such phenotypic contrasts were selected from the extremities of the traits response scale.

The 2011 RSWD stress experiment was a fore-runner to the 2012 experiment and preliminary grain set characterization from the former showed there was a significant difference between D02-105 and D08-299 in their grain set response to RSWD stress (Figure 5.5(A)). Analysis of variance on the combined 2011 and 2012 grain set data sets (See Appendix III, Table 8.1) showed a significant main effect on grain set response from treatment (i.e. +/- RSWD stress) and variety, there was also an interaction between these main effects. In the 2011 experiment (Figure 5.5(A)) the trend for reduced grain set in +RSWD stress plants was observed in all four lines tested but in D08-299 the extent of this reduction was not statistically significant. Conversely the extent of grain set reduction in D02-105 +RSWD stress plants relative to the -RSWD stress plants was highly significant, with the average dropping from 68.5% in the -RSWD stress group down to 21.7% in the +RSWD group. Assuming the average grain set per spike under control conditions (-RSWD stress) reflects the true genetic potential for grain set in each of the lines tested the extent of grain set reduction in response to RSWD stress equates to a yield loss of 32% for Westonia, 59% for Kauz, 68% for D02-105, and 5% for D08-299. Yield loss was calculated by taking the sum of the difference between grain set under the +RSWD stress and -RSWD treatments, dividing it by the -RSWD stress grain set value, and multiplying by one hundred to achieve a percentage.

In the 2012 experiment (Figure 5.5(B)) the trend for reduced grain set in the +RSWD stressed plants relative to the -RSWD stress plants was again observed in both D02-105 and D08-299 but in this data set the reduction was significant in both lines, where as in 2011 only D02-105 showed a statistically significant response. The DH parental lines Westonia and Kauz were not included in the 2012 experiment. The 2012 data showed that D02-105 grain set dropped from 66% in the -RSWD stress plants down to 47% in the +RSWD stress plants, while for D08-299 grain set was 63% and 52% in the +RSWD stressed and -RSWD

plants respectively. Based on the 2012 results RSWD stress induced yield losses were calculated as being 29% for D02-105 and 17% for D08-299.

Despite some variability in the grain set response between the two years of data (i.e. for magnitude in grain set reduction between treatments within the same line and for differences between lines for the same treatment type) the overall trends in the two consecutive experiments for lines D02-105 and D08-299 were the same. These trends being that D02-105 consistently showed a significant grain set reduction in response to +RSWD stress and that the extent of this reduction was also consistently found to be significantly more than the response observed in D08-299. On the basis of these results D02-105 was deemed critically sensitive to RSWD stress while D08-299 was deemed to have only a mild to moderate sensitivity. In the 2011 experiment (Figure 5.5(A)) both DH parental lines also exhibited a significant reduction in grain set in response to +RSWD stress relative to the -RSWD stress treatment, however the magnitude of +RSWD stress induced grain set reduction in Kauz was significantly higher than Westonia. This result were consistent with the findings for the two DH lines tested (D08-299 and D02-105) and shows that there is also a significant contrasting grain set in response to RSWD stress phenotype between the two DH parents Westonia and Kauz.

When considering the 2011 and 2012 grain set results in the context of the marker assay data (i.e. Westonia versus Kauz alleles) for the two DH lines, which showed D08-299 is Westonia like and D02-105 is Kauz like, the collective results suggest the contrasting segregation pattern of parental alleles in the two DH lines may also be driving the grain set in response to RSWD stress phenotype. Although grain set data was only collected for two DH lines in the WxK DH population it is plausible to suggest, based on the D02-105 and D08-299 data, that all DH lines in the WxK population would exhibit the same grain set RSWD stress response phenotype that is exhibited in the parent to which they show the greatest allelic likeness. These results also show that the trait based approach used in the thesis to identify two DH lines with contrasting RSWD stress

response phenotypes for downstream analysis of gene expression response to water deficit was sound.

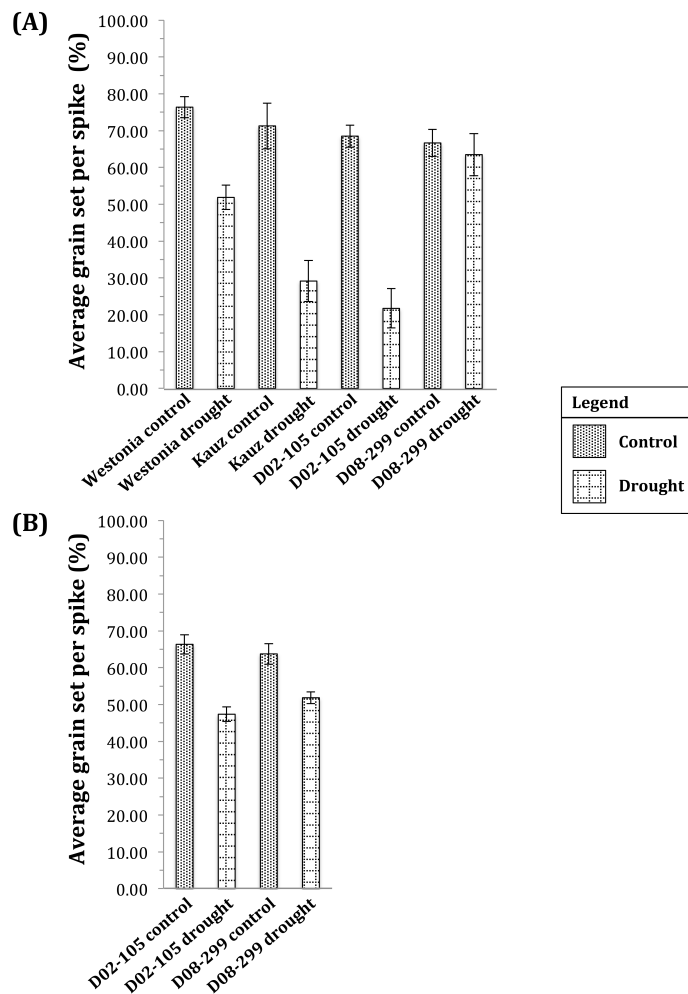


Figure 5.5 (A) 2011 RSWD stress experiment average grain set (per spike) data, according to water treatment (+/- RSWD stress) during reproductive development for lines: Westonia, Kauz, D02-1-5, and D08-299, (B) 2012 RSWD stress experiment average grain set (per spike) data, according to water treatment (+/- RSWD stress) during reproductive development for lines: D02-1-5 and D08-299. Columns in the histograms plotted along the x-axis show data (average of three replicates) for individual treatment groups according to variety/DH line tested and watering treatment imposed. Data are presented as columns with standard error bars and contrasting fill patterns to demarcate data accumulated on plus and minus RSWD stressed plants and the legend to the right of the figure provides the visual explanation of the column fill patterns. Data for the DH parents Westonia and Kauz are not presented in figure 5.5(B) because the lines were not included in the experiment.

Since the maturation time of D02-105 and D08-299 were not identical it was important to confirm that the differences in response to RSWD stress was not due to a trivial explanation such as differences in environment at the time of imposing the stress. Details are provided in Appendix II (Figure8.3) and suggest there were no consistent differences between the temperature and humidity

across the 2011 and 2012 experiments that could account for the differences in response to RSWD stress. Details in chapter 3 (section 3.3.2 and section 3.3.3) outlining how the method for imposing the water deficit stress at the equivalent developmental stage in both lines (despite differences in maturation time) show the differences in response to RSWD stress was not related to confounding effects from the water stress being imposed at different developmental stages.

As part of the characterisation of the DH lines selected for the RSWD stress studies the three drought escape QTL (*QtAD5-5A*, *QtAD5-5B*, *QtAD5-5D*) contributing to the number of growth days taken to reach the AD5 trait were evaluated for genotypic differences (Table 5.3). Table 5.3 shows the pattern of QTL inheritance for *QtAD5-5A* (from chromosome 5A) is the same in both D02-105 and D08-299, as denoted by the QQ genotype that represents the Westonia parent. A non-contrasting *QtAD5-5A* genotype (i.e. QQ for both genotypes) means that the copy of the QTL could not be the causal factor for variation seen in the AD5 trait in D08-299 and D02-15, i.e. as was observed in the earlier phenotyping studies in the glasshouse in 2010 studies (results discussed in more detail below).

Table 5.3 also shows that for the *QtAD5-5B* and *QtAD5-5D* AD5cm QTLs (located on chromosome 5B and 5D respectively) D02-105 inherited the Kauz genotype (qq) and D08-299 inherited the Westonia genotype (QQ). Importantly *QtAD5-5BL* and *QtAD5-DL* were differentiated with regards to their relative contribution to the AD5 trait. For the *QtAD5-5B* genotype, the additive effect is negative and thus causes a shift towards a faster time (i.e. growth days after sowing taken to reach) to AD5 for D08-299 (QQ QTL genotype) but a slower time to AD5 for D02-105 (qq QTL genotype). Appendix III (Equation 8.1) provides the details for the calculation of genotype specific *QtAD5-5B* and *QtAD5-D* additive effects for the respective QTL genotypes. Conversely for *QtAD5-5D* the genotype specific additive effect is positive and causes a shift towards a slower time to AD5 for D08-299 (QQ QTL genotype) and a faster time to AD5 for D02-105 (qq QTL genotype).

The preliminary phenotyping of the WxK DH population and the parents showed that the number of growth days after sowing taken to reach AD5 was contrasting between Westonia and Kauz, with the former being faster (56 DAS), and the latter being slower (63 DAS). A contrast of similar extent was also observed between D08-299 and D02-105, the respective values for which were 48DAS and 55DAS. The evidence that D08-299 and D02-105 had contrasting QTL genotypes added further weight to the selection of these lines for inclusion in the RSWD stress studies. Marked variation in the direction of *QtAD5-5B* in particular provided the opportunity to discern if causal links could be identified between the expression of alleles underpinning the QTL and the endpoint phenotype of interest in the RSWD stress response trait, grain set. Studies using RNASeq technology to further investigate the response to RSWD stress in D02-105 and D08-299 are presented in Chapter 6.

Table 5.3 ICIM Bayesian classification of genotypes of additive (and dominant) QTL identified from inclusive composite interval mapping. The results presented are the posterior probability and predicted genotype for the two QTLs identified for AD5cm, and the predicted genotypic value for the WxK DH lines evaluated, D02-105 and D08-299. Trait name: trait on which data was collected, WxK DH line: respective line in the WxK DH population the QTL genotype data relates to, Phenotype: phenotypic value (measure as number of days after sowing (DAS) when trait was reached. Three AD5cm QTL were identified and the QTL genotype data for each is presented in blocks of three columns underlying column headers titled: *QtAD5-5AL*, *QtAD5-5BL*, and *QtAD5-5DL*. For each defined QTL there are three columns of associated data: (i) P(QQ): Posterior probability the QTL genotype is QQ, (ii) P(qq): Posterior probability the QTL genotype is qq, and (iii) Genotype: QTL genotype as determined by the posterior probabilities. The scale for QTL posterior probabilities is used as a measure of probable genotypic inheritance and the scale for which is 0-1, where '1' equals 100% inheritance ranging through to '0' which equals 0% inheritance.

Trait name	WxK DH line	Phenotype	<i>QtAD5-5AL</i>			<i>QtAD5-5BL</i>			<i>QtAD5-5DL</i>		
			(i) P(QQ)	(ii) P(qq)	(iii) Genotype	(i) P(QQ)	(ii) P(qq)	(iii) Genotype	(i) P(QQ)	(ii) P(qq)	(iii) Genotype
AD5	D02-105	55	1.00	0.00	QQ	0.00	1.00	qq	0.00	1.00	qq
AD5	D08-299	48	1.00	0.00	QQ	1.00	0.00	QQ	0.60	0.40	QQ

5.4.3 Investigation of the chromosomal location of the drought escape *QtAD5* QTL and cell wall invertase (*IVR1*) genes

Extensive studies were carried out to identify sequence polymorphisms between *Westonia* and *Kauz* for the different isoforms of the *IVR1* genes defined in Chapter 4. The interest in the chromosomal location of the *IVR1* genes was based on their reported contribution to plant response to RSWD stress (Ji *et al.*, 2011). Although the comprehensive sequence analysis in Chapter 4 led to the discovery of the isoforms defined by chromosome arm location, the analysis was unable to identify polymorphisms that could be used to locate the *IVR1* genes within a high resolution genetic map. The following results summarize a target approach taken for mapping *IVR1-5BL* on chromosome 5BL only, as an example of the resolution that is achievable. The remaining members of the cell wall invertase (CW-INV) family in wheat were not investigated at this level.

The particular interest in chromosome 5BL was based on the QTL analyses presented in Table 5.2 and Table 5.3. The process of evaluating preliminary gene expression and QTL mapping results in a stepwise fashion led to the decision to target *IVR1-5BL* for mapping, where by:

(1) initially *IVR1-5BL* and *IVR1-5DL* were judged to have equal likelihood of potentially being genes of major effect contributing the RSWD stress response phenotype, given both isoform were found to be co-locating on chromosomes on which “drought escape” *QtAD5* QTL’s had been mapped (Table 5.3). No *QtAD5* QTL were identified on the chromosome arms on which CW-INV’s *IVR1.1-3B*, *IVR1.2-3B*, and *IVR1-4A* were mapped. Lack of *IVR1* isoform co-location on a chromosome arm on which a *QtAD5* was situated was taken by the author as evidence the gene(s) were less likely to be contributing to the RSWD stress response phenotype and as such these genes were relegated to lower mapping priority

(2) the QTL genotype for both *QtAD5-5BL* and *QtAD5-5DL* QTL was found to be contrasting in the two DH lines selected for downstream gene expression analysis (Table 5.3). Moreover the ICIM predicted additive effect of the contrasting *QtAD5-5BL* genotypes was found to most closely match the observed

phenotype of the trait in D02-105 and D08-299, specifically matching the direction of the additive effect. This result suggested *QtAD5-5BL* is the major effect QTL driving the trait (and not *QtAD5-5DL*) and that allele variants underlying the QTL interval relate to factors of the contrasting RWSW stress response phenotype observed between DH lines D08-299 and D02-105. The *IVR1-5BL* was thus the cell wall invertase (CW-INV) isoform for mapping, based on its potential contribution to the *QtAD5-5BL* trait

(3) preliminary gene expression analysis results (detailed in Chapter 6) showed of the *IVR1* isoforms *IVR1-5BL* was the strongest candidate gene potentially contributing to RWSW stress response, and thus *IVR1-5BL* became the target for genetic mapping preferentially over *IVR1-5DL*.

To date there is no full genome reference sequence for the majority of chromosome of interest to the current work, including 5A, 5B, 5D or 4A. Sequencing activities for these chromosomes is ongoing with completion not scheduled until 2016 (pers comm. Kellye Eversole, Executive Director International Wheat Genome Sequencing Consortium 2013). It should be noted also that while *IVR1.1-3B* and *IVR1.2-3B* were not mapped in the WxK DH 90K chip based genetic map they were mapped to the wheat chromosome 3B consensus map on a scaffold comprising ctg506 (Webster *et al.*, 2012; Keeble-Gagnère *et al.*, unpublished).

The limitations described above also had effect downstream for selection of gene targets for mapping, namely effect on the accuracy with which *IVR1-5BL* and *QtAD5-5BL* could be mapped relative to each other, and also affected the extent of loci characterisation in and around the interval for the latter. In the absence of a full genome reference many of the mapping results presented below were generated from and/or validated by using: (1) the Wheat genome survey sequence (GSS) (IWGSC unpublished), which is contiguous partial genome reference, and (2) the MIPS gene model reference (IWGSC and Klaus Mayer unpublished) which is an annotation of the former.

Analysis of the wheat 90K chip assay data (Wang *et al.*, in press) on a diversity panel of 12500 cultivars showed that a number of SNP were locating on the contigs which span the genomic regions where the respective *IVR1* genes are located (Table 5.4 data provided by Dr. Matthew Hayden). Preliminary analysis showed that fifteen out of the sixteen *IVR1-5BL* contig located markers were monomorphic and could not be used for mapping due to their co-location on the A, B and D genomes. One polymorphic marker (marker 72622) was found to map to the contig on which *IVR1-5BL* is located but the polymorphism was not evident in the assay of the MAGIC 8 way cross. Five markers were found to co-locate on the *IVR1-5DL* contig, none of which were found to be polymorphic (data not shown, Matt Hayden pers. comm. Victorian Agrobiosciences Centre La Trobe, October 2013).

Marker 72622 was found to be segregating in at least one of the populations in the 90K chip diversity panel assay. Theoretically the genetic map of a population in which the polymorphism was detected could have been used as an alternative anchor map for mapping *IVR1-5BL*. However, the diversity panel minor allele frequency for marker 72622 was only 30 in 12500 and thus the chance of finding a population with the segregating SNP were very low (pers comm. Matt Hayden Victorian Agrobiosciences Centre La Trobe, October 2013).

At present the analysis of the 90k chip marker assay (Wang *et al.*, in press) on the diversity panel is being refined to integrate additional markers (pers comm. Dr Matthew Hayden Victorian Agrobiosciences Centre La Trobe, October 2013) and it is hoped some of these may show polymorphisms for the *IVR1* contigs into the genetic map of the Opata Synthetic cross (data not shown). Subject to public release of this information the suitability of these markers to fine map the *IVR1* genes in the WxK DH genetic map could be investigated.

Table 5.4 Summary of wheat 90K chip SNP markers located on *IVR1* contigs from the IWGSC Wheat GSS. Cultivar specific 90K chip assay numbers denoting the data set from which the tabulated data was extracted were: Westonia: 66.063.02, Kauz: 138.12.877

<i>IVR1</i> isoform contig ID	90K chip marker ID	Assigning polymorphism to genome (A, B, D)
<i>IVR1.1-3BL</i> ctg_1053511	No marker founds	NA
<i>IVR1.2-3BL</i> ctg_10595081	12665	Monomorphic (A, B, and D)
	12666	Monomorphic (A, B, and D)
	49394	Monomorphic (A, B, and D)
	57156	Monomorphic (A, B, and D)
	71735	Monomorphic (A, B, and D)
<i>IVR1-4AL</i> ctg_7174266	23541	Monomorphic (A, B, and D)
	39640	Monomorphic (A, B, and D)
<i>IVR1-5BL</i> ctg_10924301	23541	Monomorphic (A, B, and D)
	39640	Monomorphic (A, B, and D)
	57156	Monomorphic (A, B, and D)
	72622	Polymorphic (B genome) in the 90K chip diversity panel lines but not in the MAGIC 8 way cross
<i>IVR1-5DL</i> ctg_10924301	22750 monomorphic	Monomorphic (A, B, and D)
	23541 monomorphic	Monomorphic (A, B, and D)
	39640 monomorphic	Monomorphic (A, B, and D)
	41822 m monomorphic	Monomorphic (A, B, and D)
	57156 monomorphic	Monomorphic (A, B, and D)

In view of the lack of polymorphism required for mapping *IVR1-5BL* the alternative strategy of synteny based physical mapping to rice was undertaken. Rice was selected as the syntenic anchor from which to map the *IVR1-5BL* in the wheat genome for the following reasons: (1) the family of cell wall invertase in rice have been extensively studied and their genomic locations mapped, (2) wheat and rice are members of the *Poaceae* family and have been found to shares a high proportion of genes and gene order (Alfares *et al.*, 2009; Schnurbusch *et al.*, 2007), and (3) phylogenetic analysis showed that, of known CW-INV characterised from numerous plant species, the family of CW-INV's in rice are the closest relatives (data not shown). Syntenic mapping of *IVR1-5BL* in wheat was based on *CIN3* in rice as the most closely related homologue (detailed below).

The open source J. Craig Venter Institute (JCVI) Rice-Wheat Syntenic Mapping platform (<http://jcv.org/wheat/index.php>), which integrates a number of wheat and rice genome resources (including >500,000 ESTs, BAC clones, and > 5,000 bin mapped ESTs), was utilized to carry out the syntenic mapping of the *IVR1*

genes in wheat. The family of CW-INV's in rice consists of eight members, of which the copy of the gene on chromosome four of rice (*CIN3*; Genebank accession AY578160, Rice Genome Annotation ID: LOC_Os04g33720) has the highest degree of homology to all cell wall invertases in wheat (homology; *IVR1.1-3B* 81%, *IVR1.2-3B* 60%, *IVR1-4AL* 78%, *IVR1-5BL* 82%, *IVR1-5DL* 82%). The specific chromosome location of *CIN3* in rice was obtained from the Rice Genome Annotation platform (Figure 5.6).

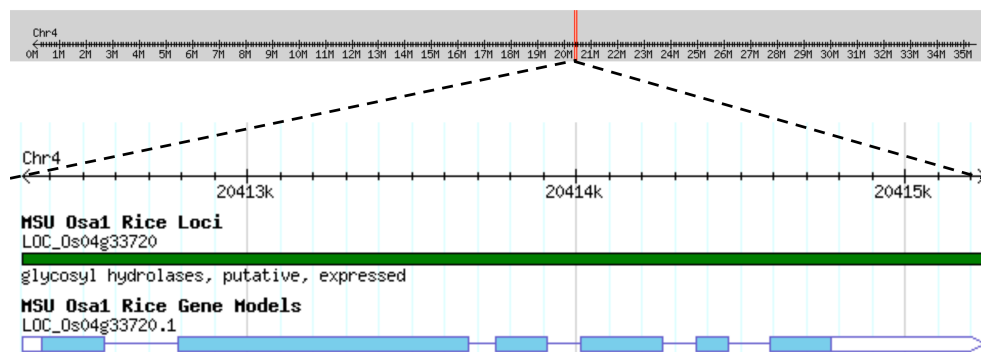


Figure 5.6 Rice Genome Annotation Project (version 1.7.0) genome browser view of rice cell wall invertase *CIN3* located on chromosome 4 (20411162-20416162bp)

The suite of wheat EST markers comprising the JCVI platform have been mapped to their syntenic location in the rice genome by identifying the homologue in the complete set of published rice pseudo molecules. Utilization of this data set enabled the identification of the approximate genomic localisation of homologous genes between the wheat and rice genomes and the broader syntenic blocks to which they belong. In the current work the wheat ESTs linked to rice pseudo molecule reference set were filtered to identify marker sequences specifically shared between rice chromosome 4 and all 21 chromosomes comprising the wheat genome (Figure 5.7(A)). Figure 7(A) shows the syntenic relationship between rice and wheat is much more complex than previously thought, in that homology exists at a micro-colinearity level as opposed to a whole chromosome level (Liu *et al.*, 2006). Using rice chromosome 4 as an example, Figure 5.7(A) highlights that the syntenic relationship of rice to wheat is defined by a vast number of small linkage blocks that co-locate throughout the entire wheat genome. Analysis of marker data used to construct Figure 5.7(A) showed that rice chromosome 4 contains 1755 wheat EST marker

sequences, which represent a suite of genes originating from all 21 chromosomes in wheat. Of the 1755 shared markers it was found that a subset of 33 wheat EST homologues were conserved between rice chromosome 4 and wheat chromosome 5B.

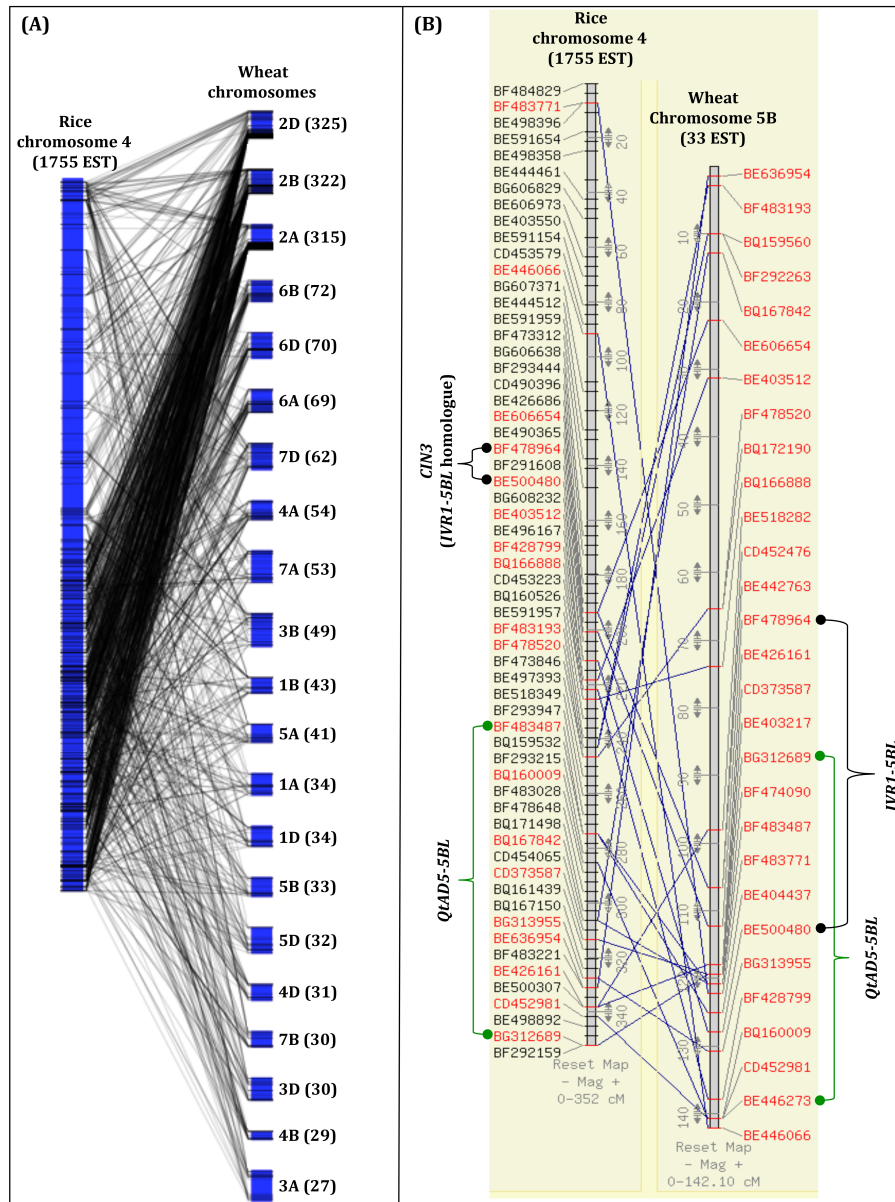


Figure 5.7 (A) A genetic map alignment showing the syntenic relationship between rice chromosome 4 (left side) and the respective chromosome specific linkage blocks in the wheat genome (right side) (J. Craig Venter Institute Wheat Genome Database 2013, <http://jcv.org/wheat/Figureview/Rice-chr4-vs-Wheat.php>), **(B)** Mapping of *IVR1-5BL* in the wheat chromosome 5BL WxK DH 90K chip based genetic map, based on syntenicity to the rice chromosome 4 genetic map. Wheat chromosome 5BL is positioned on the right and rice chromosome 4 is positioned on the left. EST markers names highlighted in red have correspondence between the two chromosomes, while marker names highlighted in black in the rice chromosome 4 map are EST which share homology to wheat chromosome other than chromosome 5B. Coloured brackets denote the broad mapping position of *IVR1-5BL* (black brackets) and *QtAD5-5BL* (green brackets). EST markers to which the ends of the brackets are aligned show the flanking marker positions for the genetic features to which

they relate. The presentation of the genetic maps using (Version 0.16) as the alignment viewer does not, for the sake of clarity, present all the molecular markers in a single image. The full details of the maps shown, at a higher levels of magnification can be obtained from the CMap location at the website, <http://ccg.murdoch.edu.au/cmap/ccg-live/>. The CMap ID's for the maps shown in the figure (B) are: left – Rice_Wheat5B, right – Wheat Rice4.

The location (marker bin and locus co-ordinates) of the 33 wheat-rice shared markers on the rice 4-wheat 5BL chromosome segment was imported into CMap (<http://ccg.murdoch.edu.au/cmap/ccg-live/>), a marker name-based correspondence program designed for carryout comparative analysis of genetic maps. The CMap alignment (Figure 5.7(B)) co-ordinates for the genomic location of *CIN3* in rice (chromosome 4: 20411162-2041612) were used to identify two markers BF478964 and BE500480, of which the interval between them defined the genome region on chromosome 5BL in wheat where *IVR1-5BL* is located. The locations of the region BF478964 and BE500480 define in rice and wheat respectively are 199.11-211.27cM and 112.21-127.49cM. Similarly Figure 5.7(B) also shows that the markers BE404437 and BQ16009 define the approximate location of *QtAD5-5BL* in wheat and rice.

With the broad syntenic location of *IVR1-5BL* and *QtAD5-5BL* mapped to chromosome 5 of wheat it was also of interest to evaluate the graphical genotypes of D02-105 and D08-299 for chromosome 5B, with respect to distribution of alleles at loci along the chromosome, either derived from Westonia or Kauz (Figure 5.8). While results from the 90K chip marker assay of the WxK DH population (discussed in detail in Chapter 5.4.2) showed D08-299 at a whole genome level contains 65% of alleles derived from Westonia and 35% from Kauz, the chromosome 5B specific proportion are 96% Westonia and 4% Kauz. The same chromosome 5B trend for allele segregation favouring Westonia was also evident for D02-105 but not to the same extent. At the whole genome level D02-105 showed 30% of its alleles are derived from Westonia and 70% from Kauz but for chromosome 5B the proportions shift to 62% and 38% derived from Westonia and Kauz respectively.

Figure 5.8 presents a visual distribution of ordered loci on chromosome 5B (based on the ICIM derived WxK DH 90K chip genetic map outputs) for D02-105 and D08-299, in which each marker-linked loci has been colour coded to represent the parental origin of the allele for a given loci. This result shows that for D08-299 chromosome 5B exists as a linkage block of almost exclusively Westonia alleles, separated only by a small linkage block of Kauz alleles at loci positioned at the very distal portion of the long arm. Chromosome 5B of D02-105 shows Westonia alleles occupy loci in a linkage block which spans the entire short arm and extends to approximately the medial region of the long arm. The distal portion of chromosome 5B of D02-105 is entirely comprised of Kauz alleles

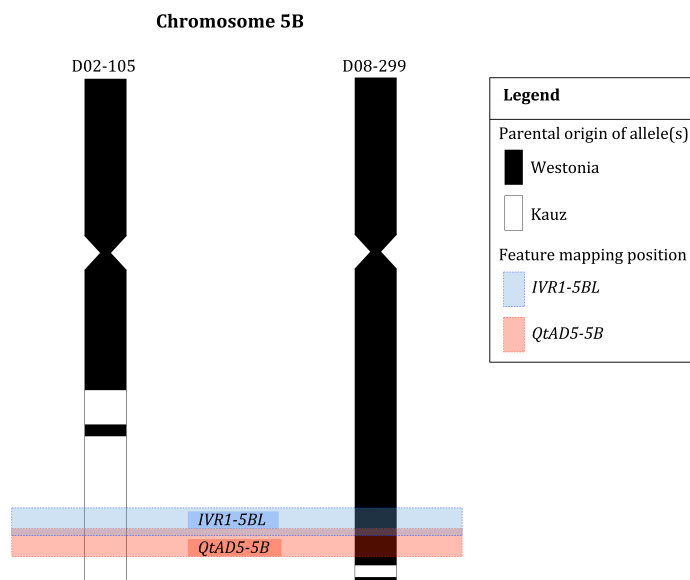


Figure 5.8 The graphical genotype of chromosome 5BL for WxK DH line D02-105 (left) and D08-299 (right). The chromosomes are colour coded to show the parental origin of alleles occupying loci along the chromosome and the broad region in which the syntenically mapped features *IVR1-5BL* and *QtAD5-5BL* are located. The legend inset shows the colour code scheme used to differential the origin of parental alleles and the mapping features.

The graphical genotype results presented in Figure 5.8 taken together with the syntenic mapping of *IVR1-5BL* and *QtAD5-5BL* (Figure 5.7) show the features are located in linkage blocks, which have contrasting alleles in D02-105 and D08-299. In D02-105 the linkage block is Kauz in origin and for D08-299 it is Westonia. Collectively these results show that both *IVR1-5BL* and *QtAD5-5BL* could be contributing to the contrasting phenotype seen between D02-105 and D08-299 for the grain set in response to RSWD stress. In view of this result it

was then also of interest to characterise the loci within and near to the *QtAD5-5BL* interval and with the view to evaluating their gene expression.

Because there are no assembled genomic scaffolds for chromosome 5B it was not possible to definitively characterize all genes underlying *QtAD5-5BL*. Availability of other genomic resources including the wheat GSS (IWGSC in press) and the MIPS gene annotation reference (IWGSC and Klaus Mayer unpublished) were utilized in the current work as an interim strategy for this purpose. The sequences for 13 markers situated within a 14cM region that spanned the *QtAD5-5BL* interval (identified from the WxK DH 90Kchip based genetic map) were used as probes to identify wheat GSS contigs located in the region of interest and to identify MIPS gene models corresponding to the contigs in question (Table 5.5). The mapping distance between flanking markers of *QtAD5-5BL* itself was 1.8cM but the no markers were mapped within this interval. The suite of 13 markers was refined to eliminate markers that were found to co-located on the same contig.

Using the MIPS annotation reference a total of 22 predicted genes were identified from 6 marker-linked wheat GSS contigs. A further 6 gene predictions were called from the sequences comprising four wheat GSS contigs which were not annotated on in the MIPS gene model reference set. Non-MIPs defined gene predictions presented in Table 5.5 were generated via contig submission to the URGI hosted gene annotation platform, Triannot (<http://wheat-urgi.versailles.inra.fr/Tools/Triannot-Pipeline> version 3.8). Each MIPS gene prediction is characterised by a high confidence class score (HSC), a ranking based on homology to fl-CDNAs for wheat, barley, Brachypodium, and rice, the purpose of which being to distinguish between highly reliable gene predictions and putative pseudo genes. Five of the MIPS gene predictions were classed as HCS_1 (>70% coverage to a reference gene), two as HCS_3 (<70>50% coverage to a reference gene), twelve as USL_1 (no reference gene match), and one as LCS (match to reference gene but with low identity).

Predicted protein sequences for the MIPS and Triannot defined gene models were used to blast the NCBI protein database for functional classification. While all genes predictions presented in Table 5.5 were of interest to pursue in downstream gene expression analysis the putative functional characterisation results highlighted four genes as being strong candidates for potential involvement in pollen development and/or water deficit stress response processes in developing spike tissue, those being: (1) transcription factors (Ta5b1Loc028528 and Ta5b1Loc037737), (2) UDP-glucose 6-dehydrogenase (Ta5b1Loc037731), and (3) strictosidine synthase (non MIPS defined but based on LOC_Os03g53950 from rice).

Table 5.5 Summary of genes underlying the chromosome 5BL drought escape QTL *QtAD5-5BL*. Gene prediction information for 27 genes is arranged in the six columns to the left of the table, including: Marker - 90K chip marker mapped in the WxK DH genetic map, Position - genetic location (cM) of the 90K chip marker on the WxK DH genetic map of chromosome 5B, Contig ID - Wheat GSS contig ID for the contig comprising the marker sequence, MIPS gene model - MIPS gene model identification code denoting the gene model annotated from the contig, HCS class - confidence score of the prediction call on the MIPS gene mode (where HCS_1 = >70% coverage to a reference gene, HCS_3 = <70%>50% coverage to a reference gene, LCS-1 = match to a reference gene but with low identity, USL_1 = no reference gene match). The four columns to the right of the table relate to the functional characterization of the gene models and summarise information relating to: Species top hit - the species from which the protein of highest homology was found with respect to the protein sequence of the MIPS gene model, Homology - degree of homology (%) between the MIPS gene model (alignment subject sequence) and the top hit query sequence, Function - putative function based on the predicted function of the top hit query, Accession - identification code denoting the defined protein sequence or predicted protein of the top hit query. * Denotes entries in the MIPS gene model column for which no gene model was available. In instances when a wheat GSS contig was found to contain a gene based SNP 90K chip marker that was mapped in the WxK DH chromosome 5B genetic map but had no MIPS defined gene model the contig comprising the marker sequence was obtained and annotated using the URGI Triannot gene annotation platform (<http://wheat-urgi.versailles.inra.fr/Tools/Triannot-Pipeline>). For these entries functional characterisation was undertaken via protein homology search using the Triannot defined predicted protein sequence as the alignment subject.

Marker	Position	Contig ID	MIPS gene model	MIPs locus	HCS class	Functional characterization via comparative analysis			
						Species top hit	Homology	Function	Accession
IAAV5992	283.1	5BL-10752063	>Ta5blLoc001259.1	Ta5blLoc001259	HCS_1	<i>Aegilops tauschii</i>	97%	Protein kinase dsk1 protein	EMT01423.1
			>Ta5blLoc001260.1	Ta5blLoc001260	USL_1	<i>Oryza sativa</i>	40%	Putative retrotransposon protein	LOC_Os04g40260.1
			>Ta5blLoc001261.1	Ta5blLoc001261	HCS_3	<i>Zea mays</i>	60%	Hypothetical ZEAMMB73_918040 protein	AFW78571.1
Excalibur_c874_1479	284.21	5BL-10903425	>Ta5blLoc028520.1	Ta5blLoc028520	LCS_1	<i>Sorghum bicolor</i>	58%	Hypothetical Sorbidraft protein	XP_002452318.1
			>Ta5blLoc028521.1	Ta5blLoc028521	USL_1	<i>Triticum urartu</i>	58%	Hypothetical TRIUR3_07023 protein	EMS59394.1
			>Ta5blLoc028522.1	Ta5blLoc028522	HCS_1	<i>Hordeum vulgare</i>	98%	Predicted stem specific protein	BAJ94066.1
			>Ta5blLoc028523.1	Ta5blLoc028523	USL_1	-	-	Uncharacterised	-
			>Ta5blLoc028524.1	Ta5blLoc028524	USL_1	<i>Oryza sativa</i>	43%	Putative retrotransposon protein	LOC_Os07g40510.1
			>Ta5blLoc028525.1	Ta5blLoc028525	USL_1	<i>Oryza sativa</i>	33%	Putative protein (uncharacterised)	LOC_Os12g07560.1
			>Ta5blLoc028526.1	Ta5blLoc028526	USL_1	<i>Oryza sativa</i>	65%	Putative NLI interacting factor-like	LOC_Os01g43870.6

			>Ta5blLoc028527.1	Ta5blLoc028527	USL_1	<i>Oryza sativa</i>	40%	phosphatase protein Myosin head family protein	LOC_Os05g46030.1
			>Ta5blLoc028528.1	Ta5blLoc028528	HCS_1	<i>Triticum urartu</i>	96%	Transcription factor 3C polypeptide 3	EMS62723.1
wsnp_Ku_c16188_25014042	285.86	5BL-2266331	not in MIPS*	-	-	<i>Hordeum vulgare</i>	95%	Uncharacterised protein	AK362557
			not in MIPS*	-	-	<i>Brachypodium distachyon</i>	82%	Putative acyl-CoA-binding domain-containing protein	100834733
			not in MIPS*	-	-	<i>Hordeum vulgare</i>	91%	Putative tRNA synthetase protein	AK361994
wmc075b	292.58	5BL-10805310	not in MIPS*	-	-	<i>Oryza sativa</i>	80%	G-beta repeat domain containing protein	LOC_Os04g50660
BS00011514_51	293.6	5BL_2865077	not in MIPS*	-	-	<i>Oryza sativa</i>	81%	Putative stritiosidine synthase protein	LOC_Os03g53950
wsnp_RFL_Contig2809_2587619	293.6	5BL-7524439	>Ta5dlLoc001591.1	Ta5dlLoc001591	HCS_1	<i>Aegilops tauschii</i>	100%	Transcriptional adapter ADA2	EMT00694.1
IACX5702	296.37	5BL-10926070	>Ta5blLoc037731.4	Ta5blLoc037731	HCS_1	<i>Aegilops tauschii</i>	96%	UDP-glucose 6-dehydrogenase protein	EMT27332.1
			>Ta5blLoc037732.1	Ta5blLoc037732	USL_1	<i>Oryza sativa</i>	36%	Eukaryotic aspartyl protease domain containing protein	LOC_Os11g08100.1
			>Ta5blLoc037733.1	Ta5blLoc037733	HCS_3	<i>Aegilops tauschii</i>	75%	Nudix hydrolase 14 chloroplastic protein	EMT28421.1
			>Ta5blLoc037734.1	Ta5blLoc037734	USL_1	<i>Oryza sativa</i>	26%	Putative thaumatin protein	LOC_Os12g43390.1
			>Ta5blLoc037735.1	Ta5blLoc037735	USL_1	-	-	Uncharacterised	-
			>Ta5blLoc037736.1	Ta5blLoc037736	USL_1	<i>Oryza sativa</i>	39%	Putative meiotic coiled-coil 7 protein	LOC_Os09g10850.1
			>Ta5blLoc037737.4	Ta5blLoc037737	HCS_1	<i>Aegilops tauschii</i>	95%	Putative WRKY transcription factor 57	EMT21988.1
			>Ta5blLoc037738.1	Ta5blLoc037738	USL_1	<i>Oryza sativa</i>	47%	Putative retrotransposon protein	LOC_Os04g19510
VrnB1	297.87	5BL-10800239	not in MIPS	-	-	<i>Triticum aestivum</i>	98%	Vernalisation protein	HQ130483.2

5.5 Discussion

The studies in this Chapter have provided a detailed molecular genetic map for wheat, based on 225 segregating double haploid (DH) lines from a cross between the varieties Westonia (from the breeding program in Western Australia) and Kauz (a CIMMYT release). The purpose for the detailed analysis was to provide a basis for choosing two WxK DH lines for further analysis in RSWD stress experiments. The genetic map presented in this chapter used the earlier 9K SNP WxK DH map (developed by Zhang *et al.*, (2013)) as an anchor, and incorporated additional markers from the 90K SNP chip (Wang *et al.*, in press). The map was developed using ICIM software and included a total of 9494 polymorphic markers (199 SSR and gene markers, 9295 SNPs). Wang *et al.*, (in press) also generated an SNP based map for Westonia x Kauz (based on WxK DH DNA sample derived from the current work) and its alignment to the map produced in this chapter showed a high degree of consensus. Construction of the molecular genetic map also enabled identification of the genetic location of a key phenotypic marker of interest to this thesis, AD5, the developmental stage at which the penultimate internode is at 5cm in length. AD5 was used as a visual cue to estimate when pollen meiosis in the developing spike was complete. This non-destructive phenotypic marker defined the time at which RSWD stress was applied for the transcriptome studies in Chapter 6.

Given no QTL for RSWD stress response or equivalent traits have been mapped in wheat to date a priority objective of this work was to conduct QTL analysis for the trait using the WxK DH 90K chip map. Evaluation of the literature and the URGI wheat QTL repository (<https://urgi.versailles.inra.fr/GnpMap/mapping/qtl/queryQtlSelect.do?progressInd=1&speciesGroupId=2>) confirmed no such QTL have been published. The repository results did show however that multiple QTL, of a related nature for number of days taken to reach flowering, were mapped in the genetic map for the wheat cultivar Renan, and were found to co-locate on chromosomes 1A, 2B, 2D, and 5A (Map ID: RER_040618, no LOD scores listed). Importantly no QTL for onset flowering and/or drought tolerance/escape were listed in the URGI repository or in the wider literature for wheat chromosome 5B.

QTL analysis results from the current work showed that 15 significant ($>LOD\ 3$) marker-trait associations exist for a range of drought escape traits relating to the number of growth days taken to reach key growth stages during the course of spike reproductive development. The drought escape QTL including *QtAD5*, *QtBS*, *QtAFV*, *QtFHE* and *QtAC* were highly correlated with each other, with each trait co-locating at the same position on chromosomes on 5A, 5B, and 5D. The co-location of the defined QTL on the group 5 chromosomes suggests the traits are correlated with potentially single causal factors resulting in pleiotropic effects. Data from similar studies in wheat by Worland and Law (1986) and Hart *et al.*, (1993), evaluating the relative contribution of genes contributing to traits including plant height and flowering time, showed that the many pairs of independent discrete mutations are closely linked to major effect genes *Rht8* and *Ppd1*, which affect plant height and flowering time respectively. In both cases the authors concluded that closely linked genes, rather than single genes, account for the occurrence of correlated traits. Analysis to determine which genes underlying the *QtAD5-5BL* interval exhibit these linkages is presented in Chapter 6. The high level of correlation between the reproductive development traits mapped in this study also shows direct selection for even one of them could potentially maximize the effects on the plant reproductive stage phenotype.

Amongst the QTL detected *QtAD5-5B*, which was mapped to chromosome 5B (290-292cM) between markers *w SNP_Ku_c10434_17255840* and *wmc075b*, was found to have the strongest influence on the number of growth days taken to reach the most critically sensitive growth stage of spike development, that being the end of pollen meiosis. The discovery of *QtAD5-5B* QTL as having the strongest effect of the number of growth days required to reach AD5 is also consistent with findings that amongst Triticeae species the group 5 chromosome have the highest density of QTL and major loci controlling plant adaptation to environment (Cattivelli *et al.*, 2002). The additive effect of *QtAD5-5BL* was found to correspond to a shift of 10 days in the onset of this growth stage, the direction (+/- days) of which in each WxK DH line depended on the parental origin of the alleles inherited for the QTL genotype. These results show the onset of pollen

meiosis (i.e. AD5) could potentially be manipulated (by selecting for a particular major effect *QtAD5-5BL* allele variant) to occur either before or after a time in the growing season that commonly encounters water deficit.

Structural and functional characterization of genes located within the *QtAD5-5B* interval (none of which have been previously identified), showed a number of the genes have hypothetical functions relating to key pollen development processes and/or abiotic stress response pathways. The mapping distance of the QTL interval for *QtAD5-5B* is 2cM, which is well below the average 10-30cM as determined by a review of QTL detection methods in bi-parental populations (Cavanagh *et al.*, 2008). Taken together the high level of significance estimated for *QtAD5-5B* (i.e. >LOD 3) and its narrow interval can be taken as evidence showing the alleles underpinning the QTL had a true association with the QTL phenotype. Four genes within a 5cM region spanning *QtAD5-5B* were deemed particularly strong candidates for gene expression analysis in downstream transcriptome studies including: two transcription factors, based on the known role of this class of gene in early stress detection (Rahaie *et al.*, 2013), UDP-glucose, based on its known involvement in sugar metabolism and energy supply to developing pollen (Roitsch and Gonzalez 2004), and strictosidine synthase a gene shown to be tissue specific in anthers, upregulated by stress, and thought to be involved in programmed cell death of tapetal tissues surrounding pollen during development (Ageez *et al.*, 2005).

Another important output from this Chapter was the identification of two WxK DH lines, D02-105 and D08-299, for more detailed studies on tolerance to RSWD stress (Chapter 6). The two lines were similar in maturity time and height but were markedly different, with respect to inheritance pattern of alleles from the parents Westonia and Kauz. As expected variation was evident between the two lines for response to RSWD stress as measured by grain set at maturity. This finding was consistent in two independent glasshouse experiments that were repeated over successive years (2011 and 2012) (Figure 5.5). However, the variation observed was more marked in the 2011 experiment, where by D02-105 showed a highly significant reduction in grain set in response to +RSWD

stress at 68% in 2011 but only a 29% reduction in 2012. No significant difference was observed in the measured response for D08-299 for 2011 compared to 2012.

Although the two successive experiments were conducted in the same controlled glasshouse facility, 2011 data recorded for the hourly temperature and relative humidity showed a +4°C temperature spike and 20% decrease in relative humidity had occurred across the duration of the final five days of the ten day RSWD stress treatment for the D02-105 plants (See Appendix II, Figure 8.3). This spike in ambient conditions in the glasshouse did not affect the rate of RSWD stress onset in the D08-299 plants in 2011 because the treatment, imposed to start at the onset of the AD5 growth stage in D08-299 which was seven days ahead of the AD5 stage in D02-105, had already been completed prior to the onset of the temperature spike, by which time the D08-299 plants had already been re-watered. Collectively these results suggest that D02-105 is clearly more sensitive in response to RSWD stress, as compared to D08-299, and that this higher degree of sensitivity is likely exhibited across the spectrum of water deficit stress from mild range stress to high. Since D02-105 is clearly Kauz-like, especially when a specific chromosome such as 5B is considered (Figure 5.8), the Kauz alleles confer a more sensitive response to environmental stress compared to Westonia alleles. One basis for this may be that Westonia senses the stress and responds quickly, while Kauz is not able to sense the stress and hence is not able to respond in a timely way, as argued by Zhang *et al.*, (2009(A)) for the mobilisation of stem carbohydrates, a trait which they studied in response to +/- RSWD stress in the WxD DH population.

A final objective in constructing the detailed 90K chip based WxK genetic map was to identify the genetic location of cell wall invertase (*IVR1*) isoforms (identified in Chapter 4), in order to test claims made by Ji *et al.*, (2011) that *IVR1* plays an important role in plant response to RSWD stress. Due to an absence of polymorphisms between the *IVR1* isoforms it was not possible to use this approach to map the genes in the WxK DH genetic map. However an estimate of the genetic location of the *IVR1* isoform on chromosome 5BL (*IVR1-5BL*) was

determined using the syntenic location of the closest cell wall invertase homologue in rice *CIN3*, situated on rice chromosome 4. The data indicated a well defined homology exists between the rice and wheat genome in the respective chromosomal regions where *CIN3* and *IVR1-5BL* are located. The evidence included: (1) the existence of homologs for both UDP-glucose -6-dehydrogenase and *CIN3/IVR1-5BL* on wheat chromosome 5BL and rice chromosome 4, and (2) positioning of the two genes within relative close proximity to each other in an equivalently defined chromosomal region in the respective species.

The existence of homology between small chromosomal regions spanning the location of *IVR1-5BL* and *CIN3* in wheat and rice respectively is consistent with previous studies that have shown conservation of gene content and order does exist between the two genomes, but at a complex microlinearity level (Liu *et al.*, 2006). This phenomena has been found particularly true for key loci controlling abiotic stress tolerance in wheat and rice but does not hold true for certain biotic stress genes and QTL, many of which are of more recent in evolutionary origin and as such are not usually located in syntenic linkage blocks (Langridge *et al.*, 2006). The high likelihood of conservation of microlinearity between wheat chromosome 5B and rice chromosome 4 for the alleles underpinning the syntenically mapped *QtAD5-5BL* QTL will also help characterise (in following projects to this thesis) the portion of the gene space which was not able to be characterised on account of incomplete coverage of that area by the IWGSC wheat GSS.

It was of interest that this region on chromosome 5BL also included the vernalisation locus, *Vrn1* (Galiba *et al.*, 1995; Snape *et al.*, 1997) and the frost tolerance loci *Fr-1* and *Fr-2* (reviewed in Pasquariello *et al.*, (2014) because these loci are also thought to be relevant in water deficit stress response. In studies by Zhang *et al.*, (2014) *Vrn-B1* was found to be polymorphic in the WxK DH population and co-segregated with the markers defining the boundaries of the *QtAD5-5BL* QTL region. Westonia was found to have the *Vrn-B1a* allele and Kauz had the *vrnB1* allele. In addition a RSWD stress response QTL on chromosome 4

of rice *QTLgys4.1* (Babu *et al.*, 2003) is situated upstream of the mapped position of rice cell wall invertase CIN3 (presented in Figure 5.7(B)). Using the map position of flanking markers associated with *QTLgys4.1* (markers RG939 and RG476 from (Babu *et al.*, 2003) the approximate mapping position of these markers in the rice chromosome 4 genetic map in Figure 5.7(B) was found to define a region where the QTL is situated at Markers BE446066 and BF478965. Using the syntenically linked EST in Figure 5.7B the mapping position of markers BE446066 and BF478965 on chromosome 5B of wheat shows the spanning region between the two markers overlaps with the mapped position of *QtAD5-5BL* and *IVR1-5BL*. Importantly the *QTLgys4.1* was detected under the same water deficit stress methodology as utilized in this thesis, namely withholding of water at the rice equivalent spike developmental stage from Z39-Z50.

The work in Chapter 5 defined some doubled haploid lines at the genome level and showed the suitability of the germplasm and marker suite used to construct a genetic map from which to identify two highly contrasting DH lines for inclusion in high-resolution transcriptome studies. Most notably the results revealed a region on chromosome 5BL with contrasting alleles underpinning the major effect drought escape QTL *QtAD5-5BL*. The suite of genes characterised in this region provides a list of targets to investigate with regards to discerning the genetic basis of variation in the *QtAD5-5BL* phenotype (growth days taken to reach pollen meiosis completion) and potential causal links to variation observed in the phenotype for grain set response to RSWD stress. Tracking the transcript expression of these candidates in the RNASeq data accumulated from the two years of RSWD stress experiments forms the basis of content presented in chapter 6.

Chapter 6: Genotype specific gene expression in the developing spike of double haploid wheat lines under water deficit stress

Chapter contributions:

The RNASeq data reported in this Chapter were generated in DEPI Victoria (LaTrobe University Melbourne). Hollie Webster worked with Matthew Hayden and Josquin Tibbits in DEPI for 2-3 weeks in 2012 and again in 2013.

The analyses reported in this Chapter required a major bioinformatics collaboration with a colleague at Murdoch University, Gabriel Keeble-Gagnere, in the form of handling and storage of the large datasets as well as running analyses to answer questions posed in the biological context.

6.1 Abstract

In this study transcriptional profiling was undertaken on the whole spike to identify genes responsive to reproductive stage water deficit (RSWD) stress with a particular focus on the expression profiles of the carbohydrate metabolism network transcripts. Despite the clear limitations of working from an incomplete wheat genome reference and limited time available to process the particularly large RNASeq data set generated in this project, the results presented in this chapter make a significant contribution to knowledge about the molecular determinants of pollen fertility in wheat. Against the backdrop of recent breakthrough findings on carbohydrate metabolism in wheat leaves and stems (Xue *et al.*, 2013) and potato tubers (Ferreira and Sonnewald 2012) the RNASeq data in this thesis was leveraged to construct the first network wide roadmap for carbohydrate metabolism in developing spikes.

In doing so, a number of previously held assumptions about the reported major effect gene, *IVR1*, in the pollen fertility trait and role in the carbohydrate metabolism network were challenged. New insights from these results shows the pollen fertility should be considered quantitative and that optimal balancing of assimilate partitioning between energy storage, utilization, and structural

development is the main determinant of the relative fitness of pollen at maturity and its capacity to successfully fertilize ovum. Based on enzyme transcript expression analysis amongst the pathways comprising carbohydrate metabolism, starch synthesis did not predominate over other pathways. Conversely enzymes involved in the lipid metabolism pathway were highly represented along with hemicellulose production pathways, showing that in the germplasm we tested lipids serve as an important source of stored energy (in addition to starch) for pollen germination.

The findings highlighted a significant asynchrony between pollen development and: (1) UDP-glucose interconversion and partitioning into specific end fate, and (2) the timing and duration of tapetal programmed cell death. In particular, molecular cues regulating temporal expression of transcripts in the carbohydrate synthesis network signals may be derived from the tapetum relative to its developmental staging.

The findings formed the basis of discussions of future work with a particular focus on testing how allelic variants underpinning genotypic differences in tapetal programmed cell death could be targeted via marker assisted selection for improvement of the carbohydrate metabolism in developing spikes under water deficit stress trait.

6.2 Introduction

The transcriptome analysis reported in this Chapter was based on the extensive characterization of doubled haploid (DH) lines from the Westonia x Kauz cross (WxK) and the selection of two lines, D02-105 and D08-299, as reported in Chapter 5. The rationale for examining the transcriptome from developing heads under plus and minus RSWD stress conditions was to identify a suite of genes coding for proteins that could be utilized as biomarkers for identifying genotypes of wheat that are better adapted to tolerating water deficit stress early in head development. A focus for this work was to evaluate expression profiles using the RNASeq platform for key transcripts involved in spike development, in particular in transcripts associated with the drought escape QTL *QtAD5-5BL* and

transcripts comprising the carbohydrate metabolism network in developing anthers.

Transcriptional profiling studies in wheat have recently proven highly effective in identifying clusters of genes sharing expression patterns in response to a single factor (e.g. drought, tissue type, or growth stage) (Deveshwar *et al.*, 2011). Reviews by Mortazavil *et al.*, (2008) and Dillies *et al.*, (2013) on the capabilities of the RNASeq platform for such purposes have highlighted that the platform is more robust than the earlier hybridization based microarray platforms. The advantage of RNASeq technology is that it provides a larger range of transcriptome coverage and has significantly improved capability for the detection of low expression transcripts, as compared to microarrays.

Despite the apparent superior capabilities of RNASeq, the accuracy with which statistically and biologically relevant changes in transcript expression between two or more treatment groups can be identified via this platform is intrinsically linked to the method used for data normalization and analysis (Esnaola *et al.*, 2013). The task of RNASeq data analysis is particularly challenging in instances when data is derived from polyploid species such as wheat due to the assembly of short reads for diverse alleles, splice variants and gene family members (Hirsch *et al.*, 2014). In particular the accurate detection of shared temporal expression profiles amongst differentially expressed genes is severely constrained in RNASeq data when derived from highly replicated multi factorial experiments (Martin *et al.*, 2013). Collectively these two factors significantly increase the amount of variability in data that must be accounted for in normalization algorithms.

There also appears no clear agreement in the literature with respect to how normalization algorithms should be constructed so as not to mask meaningful variation that is linked to different genes having different levels of expression. This issue is particularly problematic for low expression transcripts, which have been shown to be far from normally distributed even after transformation (McCarthy *et al.*, 2012). An important example relevant to this thesis, for a class

of gene potentially vulnerable to this masking effect, is transcription factors (TF's). Despite not being comprehensively studied in plants, recent human and animal TF studies have shown that TF mRNA is significantly less stable than other classes of genes (Biggin 2011), and that expression levels span four orders of magnitude ranging from very low to very high abundance (Maglietta *et al.*, 2012; Vonlanthen *et al.*, 2014). TF's are known to play an important role in regulating the expression of genes underpinning drought response traits in plants through a cascade of reactions within the complex gene networks driving the trait(s). Hence, even small changes in abundance of naturally low expression TF's that act upstream in response pathways theoretically can have an important biological significance downstream in the pathway, by way of amplification effects along the cascade of reactions they set off (Tran *et al.*, 2013).

The principle underpinning RNASeq studies is that expression levels of a given RNA is measured as a unit comprising the number of sequenced fragments that map to a gene, correlates to abundance (Rapaport *et al.*, 2013). However, because genes vary in length the number of reads sequenced and mapped to a transcript per se is clearly biased by gene length and as such cannot be used as an absolute measure of transcript expression level. In addition to gene length, RNASeq generated data is vulnerable to a number of complex technical and biological biases including transcriptome size, sequencing depth, sample library size, and G-C content (Dillies *et al.*, 2013).

A wide range of software packages are available for RNASeq data analysis including Cuffdiff, edgeR, DESeq, PoissonSeq, baySeq and limma. Broadly speaking all of these software programs involve processing the data through the following analytical steps: normalization of counts, fitting the data to a statistical model, estimation of expression level, and testing for differential expression (Dillies *et al.*, 2013). These programs rely on modeling the transcript count data using either a poisson or negative binomial distribution. Downstream normalization processes and algorithms used for statistical analysis however vary widely between programs and are reportedly the causal factor of significant variability in results obtained from various methods when tested on a common

dataset (McGettigan 2013). In a recent benchmarking study of the commonly used RNASeq analysis software packages Rapaport *et al.*, (2013) found that the inherent differences in the algorithms used in each program significantly affect the specificity and sensitivity of the analysis pipeline for detection of differentially expressed transcripts.

Consistent with the review of literature described above on limitations to transcriptome studies using RNASeq, it was also found in this thesis that apparently 'standard' protocols for RNASeq data normalization and statistical testing introduced significant bias into the results. At the global data set level biases did not limit the process of statistical detection of significant experimental treatment effects. However, at the single transcript and gene network level, bias effects appeared to be magnified and as such the results obtained lacked the accuracy for detection of differences in expression profiles. In the case of the latter a manual approach for identification of definitive differences, was adopted although this method was not applicable for quantifying differences at a larger level.

Based on the above comments, the term differential expression (DE) in this chapter is used in two different contexts. In the global transcriptome analysis sections the term DE is applied in a statistical significance context. All other transcript expression results sections relating to pairwise comparisons between single transcripts for candidate genes families of interest (e.g. cell wall invertase *IVR1*'s) or subsets of transcripts associated with a genomic feature (i.e. the drought escape QTL *QtAD5-5BL*) or gene network (i.e. carbohydrate metabolism), use the term DE in the context that expression profiles were found to be measurably different.

Despite the clear challenges of RNASeq data analysis, the RNASeq platform is a powerful genomics resource for conducting high throughput gene expression assays that allow detailed characterization of phenotypes at the molecular level (Esnaola *et al.*, 2013). In particular the results and discussion in this Chapter highlight how, through the use of this platform, cell wall invertase (CW-INV's)

expression has for the first time been studied in concert with the whole suite of genes involved in carbohydrate metabolism in developing anthers. Through the use of the well characterized WxK DH population, the results presented make a significant contribution to understanding of the role cell wall invertases play in spike development and also provide insights into the key genes driving the pollen fertility under changing levels of plant available water.

6.3 Materials and methods

6.3.1 Samples submitted for RNASeq transcriptome analysis

Methods relevant to this chapter including plant material and growth in the glasshouse, experimental design, molecular protocols including RNA extraction and preparation of RNASeq cDNA libraries are outlined in detail in Chapter 3 (General Material and Methods).

The 2011 experiment was an exploratory experiment and indicated that, in our germplasm, cytological analysis of spikes (Chapter 5, Figure 5.2) from the time points of AD3, AD5 and AD6 did not provide significant resolution of the different stages of head development. Based on the cytological analysis, a single time point of AD3-AD5-AD6 was generated for comparison to AD5 in the 2012 experiment. The time points in the 2011 experiment (AD3, AD5, AD6 and AD8) each comprised pooled biological triplicates. The AD8 time point from 2011 was used for verifying trends observed in 2012 data, which comprised assay time points of AD5, AD10, AD15, and FHE.

6.3.2 Raw transcriptome data

Raw transcriptome data generated from RNASeq on tissue collected from the two years of RSWD stress experiments comprised: (2011) 1,827,043,502 (1.8 billion) raw Illumina reads (168GB compressed), and (2012) 4,539,707,778 (4.5 billion) raw Illumina reads (438GB compressed).

As noted above, biological replicates were pooled in the 2011 experiment for RNASeq analysis because it was an exploratory experiment to confirm the strategy for the 2012 experiment. In order to carry out the detailed analysis for

differential gene expression (DGE), it proved to be an advantage to keep the biological replicates separate and this was carried out for the 2012 experiment. In the 2012 experiment the biological stages analysed (AD5, AD10, AD15 and FHE) were more clearly differentiated.

6.3.3 Quality control of short reads

Reads were filtered using an in-house Python script for the quality control (QC) pipeline prompted the following actions: (1) discard reads with > 50% of bases with less than Q20, (2) trim poor-quality (< Q20) bases at the 3' end of the read, (3) discard reads with > 5% N bases, and (4) check for overlaps between the two pairs in a read and joining the reads if such an overlap was detected. Step (4) required the first 10 overlapping bases to match perfectly, and within the remainder of the overlap only a 1 mismatch was allowed. The resulting outputs combined overlapping reads and were saved as single-end reads. Upon completion of QC (steps 1-4) a total of 1,260,097,797 and 3,411,259,802 reads remained in the 2011 and 2012 datasets, respectively.

6.3.4 Preparation of alignment references for transcript assembly construction

The MIPS annotation of the IWGSC Wheat GSS (version: survey_sequence_gene_models_MIPS_feb2013.zip, released February 2013, available at: https://urgi.versailles.inra.fr/download/iwgs/Gene_models/) was used as the alignment reference for transcript assembly construction. Only those scaffolds from the wheat GSS containing MIPS annotations were used to construct the reference. In total, 507,286 GSS scaffolds were used, with a combined length of 2,313,894,040 bp. A new version of the MIPS annotation of the IWGSC Wheat GSS was released in February 2014, but was not available for the analysis carried out in this thesis.

6.3.5 Alignment of short reads to genome reference and generating counts

Filtered short reads were aligned to the MIPS wheat GSS reference using Tophat 2.0.9 (default parameters). The paired-end and single-end reads were aligned separately, with the junctions from the paired-end alignments passed to the single-end alignments with the -j option, as described in the Tophat manual

(<http://tophat.cbcb.umd.edu/manual.shtml>). The resulting .bam files were then individually name-sorted with Samtools and read counts calculated using the htseq-count command in HTSeq with the following options: --mode=intersection-nonempty --stranded=no -t exon -i gene_id. The counts for paired- and single-end reads were then combined into a single count table for each replicate. For the 2011 and 2012 datasets 1,046,827,663 (1.0 billion, 83% of total) and 2,205,604,403 (2.2 billion, 65% of total) reads respectively were mapped to the MIPS reference transcripts.

6.3.6 Identifying outliers

For each experimental sampling time point a counts matrix was constructed and loaded into the statistical software R (version 3.0.0). Multi-dimensional scaling (MDS) plots were produced using the edgeR package (plotMDS command). Outliers from the data were incrementally removed upon discussion of the iterative rounds of MDS, until the +/- RSWD stress data separated in one of the two dimensions.

6.3.7 PCA analysis

Preliminary estimates of the variation within the datasets were obtained using principal component analysis (PCA) (separately for 2011 and 2012). Transcripts with ≥ 100 counts (summed across all samples) were used in the analysis. Counts were subsequently log-transformed to reduce noise. PCA was conducted in R using the function: prcomp.

6.3.8 Total expressed transcripts and assigning of GO terms

A transcript was defined as expressed in an individual sample if, on average, ≥ 2 reads per replicate aligned to it. To evaluate transcript expression according to GO term categorisation 61,290 probes from the Affymetrix wheat array were Blast against the MIPS gene models with a cut-off of 75% coverage and 95% sequence identity. GO term evaluation was conducted on the 2012 data but not the 2011 data set because the level of replication was more extensive in the 2012 experiment. Of the 212,759 expressed transcripts in the 2012 data set, a total of 38,078 were assigned to gene models. GO terms for each probe were obtained

through PlexDB (<http://www.plexdb.org/plex.php?database=Wheat>) and assigned to corresponding transcripts. A SPARQL script (G Keeble-Gagnere and J Nystrom) was written to obtain 2nd-level GO terms from the list defined through PlexDB. For each expressed transcript, 2nd-level GO terms were assigned and total counts generated.

6.3.9 Comparison between expressed transcripts in 2011 and 2012 experiments

To assess reproducibility of expression between the two experiments, the set of expressed transcripts for the pooled AD3-5-6 time point in 2011, and the AD5 time point in 2012 were compared. Overall, a high degree of conservation was found between the two sets of expressed transcripts. In general fewer transcripts were expressed in 2012, but this may have been due to increased batch effects/noise in the 2012 data.

6.3.10 Identification of stable expression transcripts

Trimmed means of M (TMM) normalised counts were obtained from edgeR and used to define a set of “stable” transcripts. For each transcript defined by the MIPS Wheat GSS annotation, counts were ordered and the top and bottom 10 removed from the analysis. The mean, median, and variance as percentage of the median were calculated for the remaining counts. The top 1000 stable genes were analysed for both 2011 and 2012 experiments.

6.3.11 Global data set analysis for statistical calculation of differential transcript expression

Detection of transcript differential expression in the global RNASeq data set was calculated with edgeR. The following code was constructed by Gabriel Keeble-Gagnere for this purpose.

```
library(edgeR)

data = read.table("counts.raw.txt", header=T, row.names=1, com="")
col_ordering = c(1,2,3,5,7,8,9,10,11)
rnaseqMatrix = data[,col_ordering]
rnaseqMatrix = round(rnaseqMatrix)
```

```

rnaseqMatrix = rnaseqMatrix[rowSums(rnaseqMatrix)>=10,]
conditions = factor(c(rep("C", 4), rep("D", 5)))
write.table(rnaseqMatrix, file='counts.raw.filtered.txt', sep=' ', quote=F, row.names=T)

exp_study = DGEList(counts=rnaseqMatrix, group=conditions)
exp_study = calcNormFactors(exp_study)
pdf("mds.no-c3-d0.pdf")
plotMDS(exp_study)
dev.off()
counts.per.m = cpm(exp_study, normalized.lib.size=TRUE)
write.table(counts.per.m, file='counts.per.m', sep=' ', quote=F, row.names=T)
exp_study = estimateCommonDisp(exp_study)
exp_study = estimateTagwiseDisp(exp_study)
et = exactTest(exp_study)
tTags = topTags(et,n=NULL)
write.table(tTags[tTags$table$PValue <= 0.05,], file='counts.raw.txt.C_vs_D.edgeR.DE_results',
sep=' ', quote=F, row.names=T)

```

6.3.12 Calculation of differential transcript expression (pairwise comparisons between single transcripts)

During our preliminary RNASeq mapping activities a number of different data normalization and statistical testing for differential expression approaches were tested (data not shown). Comparison of results obtained from the various approaches showed a significant amount of variation, firstly in the data retained for mapping (after quality control (QC)), and secondly in the list of transcripts identified as being differentially expressed. Presentation of comparative analysis of methods for transcript assembly construction and detection of differential expression were beyond the scope of this thesis. Criteria for method selection of preliminary RNASeq pipeline protocols, relating to data QC and transcript assembly construction approach, followed standard protocols and was based on validation by principal component analysis (PCA) and multidimensional scaling (MDS).

The introduction of bias in downstream steps in the RNASeq pipeline using the software packages, EdgeR and CuffDiff, was considerable and could not be reconciled. In particular the biases rendered process of calling differential

expression between two unique transcripts derived from the same sample (i.e. a sample on which the same experimental conditions were imposed) ambiguous. Bias was primarily due to distortion of the transcript expression signal and was underpinned by the normalization algorithms which were designed to: (1) nullify variation in transcript count numbers caused by differences in gene length, where by longer transcripts were detected as having higher expression signals by virtue of the fact they have a higher number of short read counts mapping to them, and (2) equilibrate transcript count numbers by normalizing each transcript assembly back against every transcript assembly in the whole data set.

In this Chapter we investigated new methodologies for standardizing each transcript assembly against a subset of pre-determined highly stable expression genes for which expression levels remain unaffected by treatment effects. In the case of transforming transcript count numbers by normalizing each transcript assembly back against every transcript assembly in the whole data set, it was found that the iterative normalization process caused the direction of the bias to change according to: (1) which transcript assembly's were normalized first to last and, (2) the order of the global data set to which each transcript assembly was normalized against. Hence the decision was taken not to normalize for these parameters in this thesis. The only parameter the data was normalized for was library size, a process which involved: (1) calculating the factor by which each sample library needed to be multiplied for its total number of reads to be equivalent to the largest library, (2) multiplying the number of counts for a transcript assembly according to conversion factor calculated for the sample library to which it belonged.

Because gene length was not normalized for differential expression analysis was only undertaken between two transcripts in the following circumstances: (1) when transcripts were homologs or isoforms of the same gene family and in either case where not significantly different in length, or (2) when transcripts were completely unrelated but were of a similar length. Examples for the latter included when it was of interest to: (a) compare the expression of two

transcripts that are involved in the same gene network or shared a common molecular function, or (b) benchmark change in level expression of a differentially expressed transcript against an equivalently sized stable expression transcript (i.e. candidate house keeping gene) from the core set of 1000 stable expression transcripts reported on in section 6.4.2 of this chapter.

The results section utilizing this normalization approach included: house keeping/stable expression genes (*GAPDH*'s), cell wall invertases (*IVR1*'s), QTL *QtAD5-5BL* interval genes, and gene belonging to the sucrose-starch metabolism network. Because this data was not analysed in a statistically results are not presented in a statistical but rather as pattern matching analysis results for discerning overall expression trends. Standard error bars calculated from average transcript expression across replicates are used to highlight measurably different levels of transcript expression.

6.3.13 Investigation of KEGG starch sucrose metabolism pathway

The 26 unique protein sequences in the rice starch sucrose metabolism pathway (Network pathway ID: osa00500) were downloaded from KEGG (http://www.genome.jp/kegg-bin/show_pathway?osa00500) and Blast-ed (blastp) against the MIPS wheat GSS annotation sequences. Genes in the KEGG pathway were assigned to MIPS transcripts based on top hit (top hit was defined as $\geq 75\%$ homology and $\geq 50\%$ coverage).

6.4 Results

6.4.1 Illumina sequencing and assemble construction using the MIPS wheat GSS reference

In 2011, 16 cDNA libraries were sequenced from four spike development stages (auricle distance AD3, AD5, AD6, AD8). In 2012, 96 cDNA libraries were also sequenced from four spike development stages but from a more expanded time frame of AD5, AD10, AD15, and FHE (full head emerged). As a result the cumulative total for transcriptome data capture for 2011 and 2012 was >1.2 Tbytes compressed, including 168Gbytes in 2011 and 438 Gbytes in 2012. Total number of reads generated was $>1.8 \times 10^9$ (2011) and $>4.5 \times 10^9$ (2012) of 50-

100bp in length. After quality control to filter out poor quality reads 1.2×10^9 (2011) and 3.4×10^9 (2012) million high quality reads were retained for mapping.

The proportion of transcripts from the 2011 data set that were also expressed in the 2012 data set was 92%. Reads retained for mapping equated to 66% and 75% of the 2011 and 2012 raw data sets respectively, and indicated the high quality of the short read data produced using the RNASeq platform.

Results generated from counts matrices (data not shown) for read assemblies of transcripts mapped to gene models showed that 1.0×10^9 and 1.2×10^9 reads were mapped to the MIPS wheat GSS reference in the 2011 and 2012 data set respectively. Using the value of total number of reads retained for mapping (i.e. post QC) this result shows the mapping efficiency was 83% for the 2011 data set and 65% for the 2012 data set. Figure 6.1 shows a histogram of total number of reads retained for mapping plotted against the total number of reads mapped to the MIPS genome reference for each experimental sample group. The result for the 2011 experiment (Figure 6.1(A)) shows that the mapping efficiency across all sample groups was highly stable (only 1% variation calculation not shown), while in 2012 (Figure 6.1(B)) it was less stable, exhibiting a variation between samples of 15%, calculation not shown). The variation in the 2011 experiment was much reduced because each of the AD3, AD5 and AD6 that were pooled to form AD3-5-6 were comprised of pooled biological replicates.

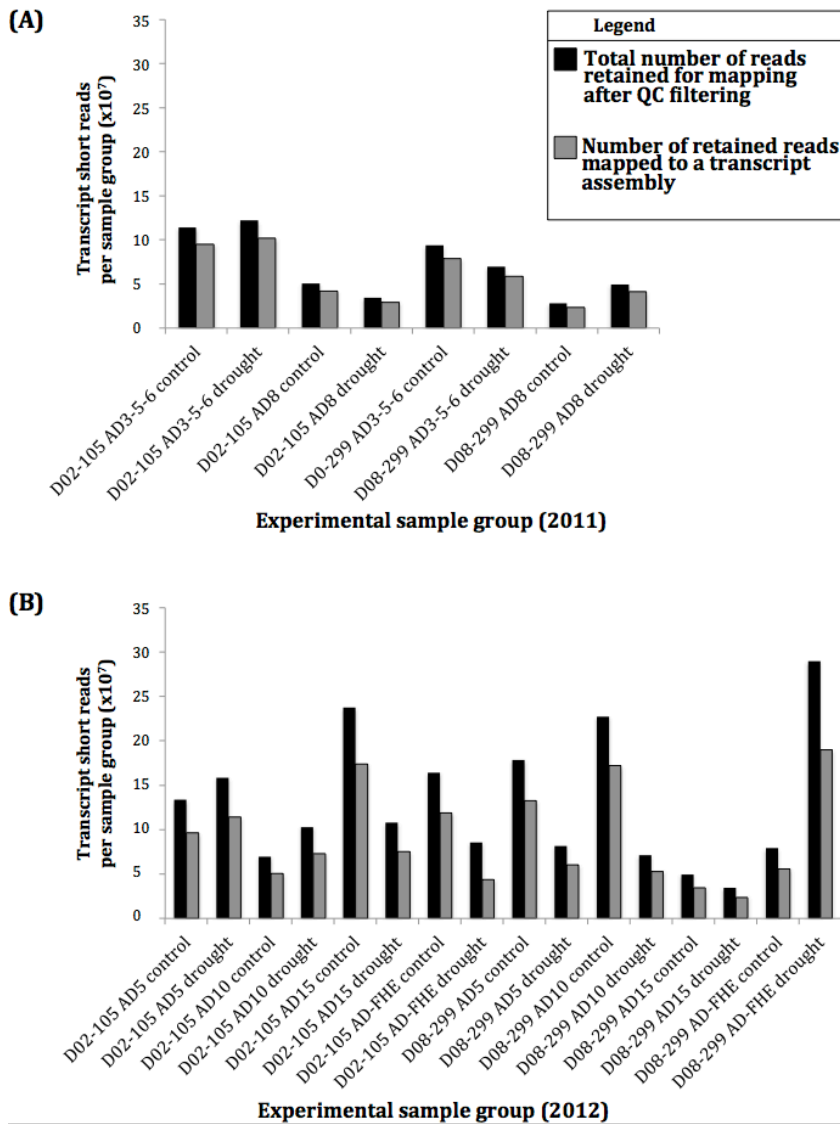


Figure 6.1 A histogram plotting total number of transcript short reads (after QC filtering) comprising the 2011 (A) and 2012 (B) RSWD stress experiment data sets, plotted against the total number of reads mapped to genome reference (MIPS wheat GSS annotation). Experimental sample groups are plotted along the x-axis and short reads- expressed transcripts (counts) are mapped on the y-axis. Each experimental sample group comprised two columns of data, which are colour coded to demarcation of values pertaining to total number of reads retained after QC (black columns) and total number of retained reads mapped to the genome reference (grey columns). Note that the AD stages 3, 5, and 6 samples from the 2011 experiment were pooled for the RNASeq summary in (A). Biological replicates in the 2012 experiment (B) were not pooled, as discussed in the methods.

Analysis of the mapped transcriptome data showed that at least 1.3×10^5 and 2.1×10^5 transcripts were expressed in the global data sets for 2011 and 2012 experiments respectively, and that 92% of transcripts expressed in the 2011 experiment were also expressed in the 2012 experiment. It should be noted however, values obtained for transcript global expression were expected to be an under representation of the true expression numbers for two reasons. Firstly,

the IWGSC wheat GSS is not a full genome reference. Secondly, the gene recognition algorithms used to construct the MIPS annotation are not yet robust enough to: (i) capture all features in the current version of the wheat GSS, (ii) filter out all pseudo features relating to repetitive sequences, and (iii) concatenate multiple features that should comprise one gene upon alternative splicing. Our analysis of the MIPS wheat GSS (version: survey_sequence_gene_models_MIPS_feb2013.zip, released February 2013, available at: https://urgi.versailles.inra.fr/download/iwgs/Gene_models/) showed the annotation currently comprises 9.6×10^6 predicted gene models, a value which will be refined as repetitive sequence features are removed.

Figure 6.2 shows the total number of mapped transcripts for 2011 and 2012 to the MIPS wheat GSS reference. In total the number of transcripts represented ranged from 8.3×10^4 to 11.4×10^4 in 2011 and from 8.9×10^4 to 13.7×10^4 in 2012. Proportionately these values show between 9-13% of the transcriptome was represented in the 2011 data set and between 9-14% in the 2012 data set. When considered in parallel with results for the total amount of reads mapped (Figure 6.1) the results presented in Figure 6.2 highlights that despite a degree of variability between the two data sets for the total amount of data mapped it appears to have had no significant effect on the proportion of the transcriptome being detected as being expressed. For example for the treatment group, D08-299 AD15 +RSWD stress (2012), the total number of reads mapped was 1.9×10^7 for the 2012 experiment and the proportion of the transcriptome detected as expressed was 11%. For the treatment group D02-105 AD3-5-6 (2011), 9.5×10^7 reads were mapped (representing almost a five times the amount of data) and yet the proportion of the transcriptome detected as expressed was also 11%. The proportion of the transcriptome that was detected as expressed in both 2011 and 2012 was much lower than had been anticipated and it is expected that this is due to the incomplete nature of the MIPS wheat GSS annotation.

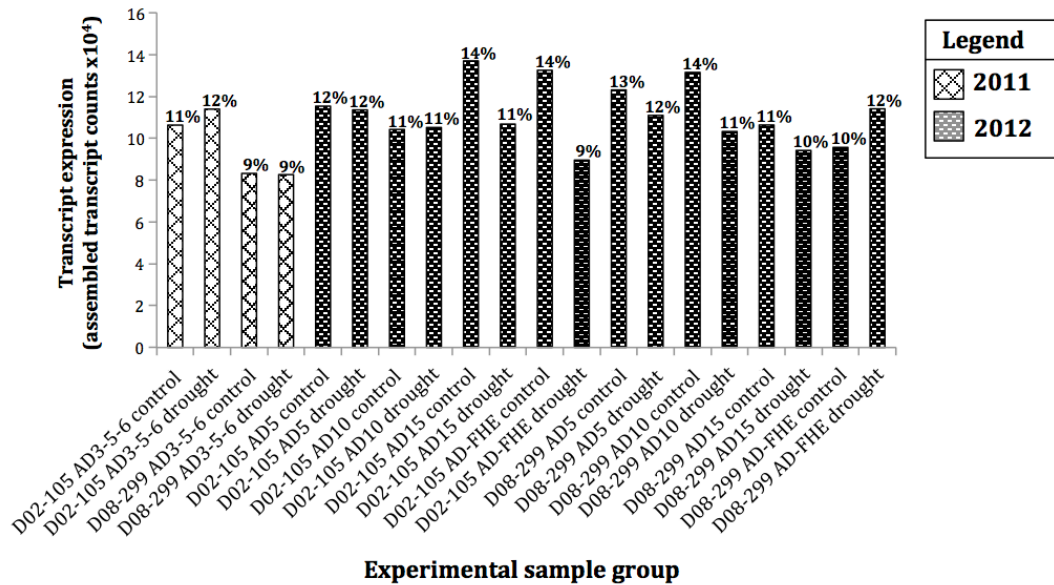


Figure 6.2 A histogram of the total number of transcripts expressed per experimental sample group in the 2011 and 2012 RSWD stress experiments. Columns representing data for each sample group are plotted along the x-axis and transcript counts are shown on the y-axis. A legend inset to the right of the histogram shows the colour scheme used to identify data arising from the 2011 and 2012 experiments. Percentages shown above each column convey the proportion of total transcripts expressed in experimental group as a proportion of the total number of annotated features in the MIPS wheat GSS.

In order to quantify the amount of variation within and between the 2011 and 2012 data sets, principal component analysis (PCA) was carried out (Figure 6.3). Within the 2011 data (Figure 6.3(A)) 68% of variance was contained in the first two principal components (PC1 and PC2) and for 2012 data (Figure 6.3(B)) the same PC levels accounted for 42% of the variance. The PCA for both the 2011 and 2012 data showed the replicates, in most cases were, clustering together in the respective years of data. The clustering of replicates highlights the variation within each experiment is predominantly due to differences between experimental treatment groups and was the result of main effects relating time experiment factors such as: treatment (+/-RSWD stress), time (growth stage), and variety (DH line/variety). Moreover, the consistent clustering of replicates in both years of data show the reproducibility of the experiments across years, and the soundness of the data obtained overall.

While variation on the whole was kept to a minimum, comparisons between Figure 6.4(A) and 6.4(B) show there is a degree of variation between the 2011 and 2012 experiments, as indicated by the increased scale of the PC1 and PC2 axes in 2012 PCA (Figure 6.4(B)). The major difference between the 2011 and

2012 datasets is the absence of time as a variable in the 2011 dataset as noted in Figure 6.1. The information in Table 6.1 provides a summary of the comparisons between the 2011 and 2012 data sets for identifying expressed transcripts common to both. In this table the time point comparison made was between the 2011 AD-3-5-6 data and the AD5 time point data in 2012 (the AD10, AD15, and FHE were not represented in the 2011 data). For D02-105 there was no significant difference detected between the two years for the number of transcripts expressed in the AD-3-5-6 as compared to AD5 in 2012, based on the very close values calculated in the respective years denoting the proportion the overlapping (i.e. commonly expressed) list of transcripts represent in the individual years of data. Conversely, a significant difference was detected for the equivalent comparison in the D08-299 data set, where an additional 42496 (-RSWD stress group) and 33576(+RSWD stress group) were expressed at the AD5 growth stage. Because the trend for increased numbers of transcripts expressed in the second experiment (2012) was consistent in both the -RSWD stress and +RSWD stress treatment groups in D08-299, the result suggests the 'year' effect was not due to variability in water deficit intensity.

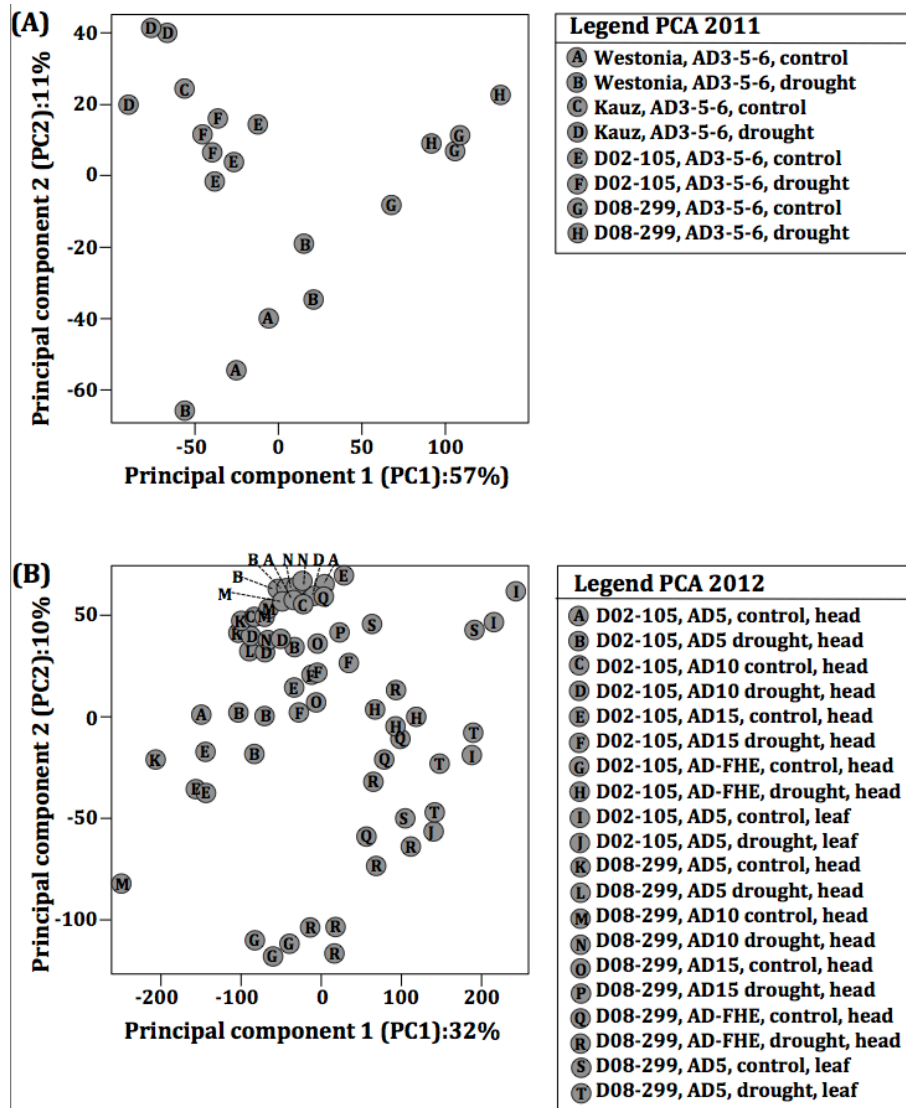


Figure 6.3 (A) Principal component analysis (PCA) of the mapped transcripts in the 2011 (A) and 2012 (B) data sets. In both plots the scale of variability for principal component 1 (PC1) is presented on the x-axis and the scale of variability for principal component 2 (PC2) is presented on the y-axis. Each data point in the plots represents one replicate of a particular experimental treatment group and is depicted by a grey circle with a letter in it. The letter assigned to individual data points corresponds to the treatment group and replicate for each data point and legend inset to the right of both plots ((A) and (B)) provides the full name for a treatment group according to letter.

Table 6.1 provides a summary of the comparisons between the 2011 and 2012 data sets for the identification of expressed transcripts common to both. As outlined previously the 2011 experiment data for the individual AD3, AD and AD6 were pooled, hence the only meaningful time point comparison to be made against the 2011 AD-3-5-6 data was from the AD5 time point data in 2012 (i.e. on account of AD10, AD15, and FHE not being represented in the 2011 data). For D02-105 there was no significant difference detected between the two years for

the number of transcripts expressed in the AD-3-5-6 as compared to AD5 in 2012 as evidence by the very close values calculated in the respective years denoting the proportion the overlapping (i.e. commonly expressed) list of transcripts represent in the individual years of data. Conversely, a significant difference was detected for the equivalent comparison in D08-299 data set, where by an additional 42496 (-RSWD stress group) and 33576 (+RSWD stress group) transcripts were expressed at the AD5 growth stage. Because the trend for increased number of transcripts expressed in the second experiment (2012) was consistent in both the plus and minus RSWD stress treatments for D08-299 the result suggests the ‘year’ effect was not due to variability in water deficit intensity.

Table 6.1 Summary of shared transcripts between the 2011 pooled AD3-5-6 data set and a subset of data comprising the AD5 growth stage from the 2012 data set.

Variety	Condition	2011 AD3-5-6 expressed transcripts ¹	2012 AD5 expressed transcripts ¹	Overlap	% of 2011	% of 2012
D02-105	-RSWD	106302	115346	94981	89.35	82.34
D02-105	+RSWD	113935	113563	99273	87.13	87.42
D08-299	-RSWD	83205	123098	80602	96.87	65.48
D08-299	+RSWD	82622	110930	77354	93.62	69.73

¹FDR cut-off of 0.02 was used.

6.4.2 Analysis of stable expression genes for use as house keeping genes

Because *GAPDH* was utilized as the primary housekeeping gene in the transcript expression analysis using semi-quantitative PCR (SQ-PCR) in the RSWD stress experiment samples (2011 and 2012) it was of interest to evaluate the stability of gene expression in the RNASeq data sets generated on the same samples. It was expected that cross-referencing *GAPDH* results between the two platforms would enable meaningful comparisons to be made. *GAPDH* was originally selected as a housekeeping gene based on published literature (Breen *et al.*, 2010(A)), which suggested that the gene maintained stable expression across all tissues, growth stages, and in all plant growth environments.

Preliminary SQ-PCR assays on samples from the 2011 and 2012 experiments suggested *GAPDH* was expressed in a stable manner in all samples evaluated (result discussed in more detail below). Quantification of *GAPDH* transcript

expression in these assays was via measurement of light intensity of PCR bands on agarose gels. Based on these preliminary results it was concluded that *GAPDH* met the criteria for use as a housekeeping gene. Conversely, equivalent evaluations of *GAPDH* transcript expression in the RNASeq data sets showed that in fact not one *GAPDH* gene, but multiple isoforms were being assayed in all samples collected across the two experiments. Subsequently the sequences for *GAPDH* primers used in the SQ-PCR assays (Table 6.2) were Blast'ed against the IWGSC wheat GSS. Blast results showed the primers assayed at least six isoforms located on five different chromosome arms including: 2DS, 2BS, 6AL, 6BL, 6DL.

Table 6.2 IWGSC GSS contigs on which GAPDH primer sequences are conserved (100% identity) and corresponding MIPS gene model ID's for the GAPDH isoforms annotated on the respective contigs.

Primer sequence	GSS contig	MIPS ID
>GAPDH_Forward CGAAGCCAGCAACCTATGAT	>6DL_3217060	Ta6DL-Loc003951
	>6BL_4402652	Ta6BL-Loc022066
	>6AL_5816732	Ta6AL-Loc014256
	>6AL_5749470	Ta6AL-Loc003193
	>2DS_5341461	Ta2DS-Loc017499
	>2BS-5189115	Ta2BS-Loc010950
>GAPDH_Reverse CGAAGCCAGCAACCTATGAT	>6DL_3217060	Ta6DL-Loc003951
	>6BL_4402652	Ta6BL-Loc022066
	>6AL_5816732	Ta6AL-Loc014256
	>6AL_5749470	Ta6AL-Loc003193
	>2DS_5341461	Ta2DS-Loc017499
	>2BS-5189115	Ta2BS-Loc010950

Statistical evaluation of transcript expression for the *GAPDH* isoforms summarized in Table 6.2 showed all six isoforms were differentially expressed (data not shown). Figure 6.4 provides an example of the variability seen in *GAPDH* isoform expression in the 2012 RNASeq data by highlighting the full head emerged (FHE) growth stage. The results are presented as a histogram with standard error bars. The error bars for the majority of pairwise comparisons do not overlap, thus providing evidence that *GAPDH* was differentially expressed in an isoform specific manner in response to both treatment type (+/- RSWD stress), and genotype (WxK DH line D02-105 versus D08-299). *GAPDH* transcript expression comparisons were also made between tissue type (leaf versus head), growth stage (AD3, AD5, AD6, AD10, and AD15) and variety

(Westonia, Kauz) (data not shown), and confirmed isoform specific differential expression was evident spatially, temporally and by genotype.

Figure 6.4 shows that of the *GAPDH* isoforms in the 2012 RNASeq data set the two isoforms located on chromosome 6A (MIPS wheat GSS gene model ID: Ta6a1Loc003193 and Ta6a1Loc014256) were expressed at almost non-detectable levels (<10 counts). The absence of expression of these two isoforms was consistent with results from the Westonia and Kauz data. The total combined expression data of *GAPDH* isoforms is plotted in Figure 6.4 (plotted as blue columns), and is consistent with the expression levels determined for *GAPDH* during earlier SQ-PCR assays (i.e. performed prior to RNASeq) at which time it was assumed the assayed products were for one *GAPDH* gene.

GAPDH expression in the SQ-PCR assays was based on quantification of PCR band light intensity. Statistical detection of differential expression via this method was not possible due to lack of sensitivity in the Gel doc UV light image capture apparatus used. However the results did provide an indication for the PCR product assayed by the *GAPDH* primers being: (1) stably expressed (at high levels) in all tissues assayed, and (2) consisting of a single band suggesting a single transcript product. In Figure 6.4 an image for an SQ-PCR assay of for *GAPDH* is positioned above its equivalent assay in the RNASeq data (i.e. both assays were performed on the same sample). This comparison highlights how the light intensity of what appears to be one *GAPDH* PCR product is in fact four PCR products (of the same approximate size) overlaying each other, thus making the image appear as one band of high illumination intensity. Direct sequencing of *GAPDH* transcripts expressed in the 2011 and 2012 SQ-PCR assays was undertaken (data not shown) and confirmed the same four isoforms were being preferentially assayed by both platforms. Taken collectively results in Figure 6.4 show: (1) *GAPDH* isoforms do not define a set of stable-expression-genes in wheat and as such are not a suitable for use as a house-keeping genes to normalize expression levels of test genes: and (2) when selecting house-keeping genes for gene expression studies in polyploid species (based on evidence from

studies on diploid species) further work must be taken to identify isoforms and/or homologs and to design primers specific to the isoform selected.

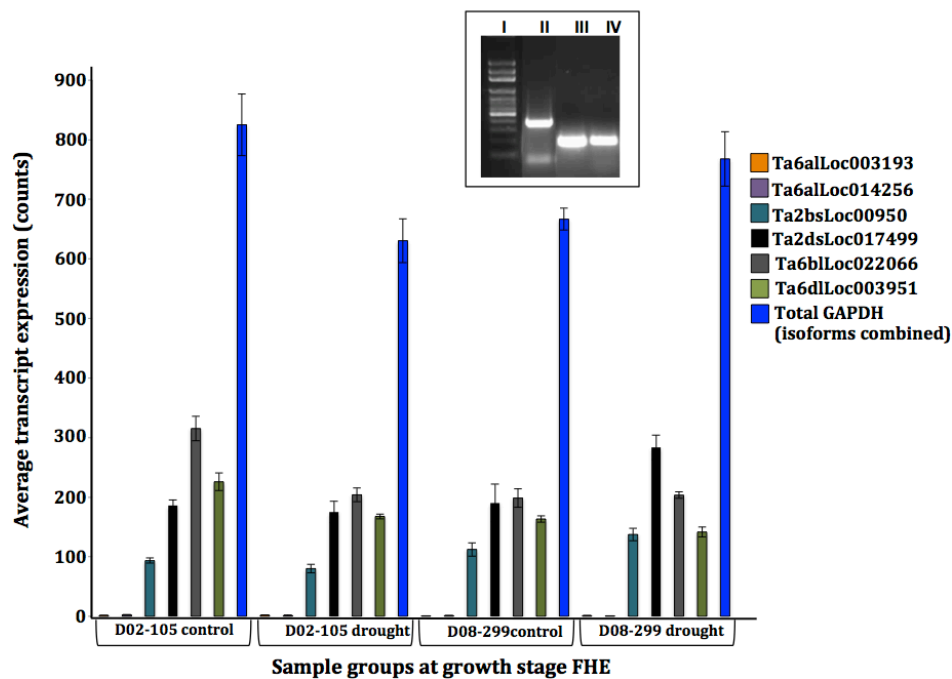


Figure 6.4 *GAPDH* isoform specific transcript expression at growth stage FHE (2012 RSWD stress experiment) in D08-299 and D02-105 under plus and minus RSWD conditions. The histogram present (for both DH lines and both experimental treatments tested) the individual expression of all six *GAPDH* isoforms identified in the IWGSC wheat GSS, relative to the cumulative (combined) total expression of all *GAPDH* isoforms. Inset (upper left corner) – an agarose gel image displaying the SQ-PCR band intensity of assayed products from D08-299 –RSWD stress samples at FHE. Gel bands are annotated (from left to right): (I) DNA ladder, (II) *IVR1* (multiple isoforms assayed), (III) *GAPDH* (multiple isoforms assayed), (IV) *EXPB11* (multiple isoforms assayed). The horizontal positioning of the gel image above the histogram allows for direct comparison between results obtained from the two contrasting gene expression analysis methods (SQ-PCR and RNASeq) on samples derived from the same tissue (D08-299 at FHE under –RSWD stress conditions).

In view of the fact that *GAPDH* genes did not meet the criteria for use as housekeeping genes, the RNASeq datasets were screened to identify a suite of stable expression transcripts. Expression stability was assessed by calculating the variance in expression thresholds for a given transcript relative to its median expression value. Analysis to determine the level of variation amongst the top 1000 most stable expression transcripts in the RNASeq data sets indicated the 2011 data set was more stable than the 2012 dataset. The 1000th most stable expression transcript in the 2011 data set exhibited 18% variance across all samples, while in the equivalently defined set of 2012 transcripts 47% variance was detected. The suite of 1000 stable expression transcripts was subject to GO

term classification (data not shown), the results from which showed that over 500 genes were associated with one of the following GO terms: molecular function – binding (GO:0005488), catalytic activity (GO:0003824), cellular component: cell (GO:0005623) and cell part (GO:0044464), processes: cellular processes (GO:0009987), metabolic processes (G):0008152).

Figure 6.5 (A) shows the variance detected in the expression of the *GAPDH* isoforms in the 2012 data set (2011 data not shown) ranged from 110-130%. This result shows that even the most stable expression *GAPDH* isoform, Ta2dsLoc017499 that has 63% variance around its median expression value, well in excess of that observed in the list of 1000 most stable expression transcripts. Variance calculated for the 2011 *GAPDH* isoform expression (data not shown) was much lower (ranging from 27-65%), but was still not within the defined suite of stable expression transcripts. It was concluded that while 2011 exhibited less transcript expression variance than 2012 at a global level (including *GAPDH* data) the result was likely skewed on account of the fact the 2011 data was derived from fewer samples than the 2012 dataset.

To enable identification of candidate housekeeping genes, the core list of 1000 stable expression genes identified in the 2012 data set was refined to 500. Figure 6.5 (B) show the expression variance for the 1st and 500th most stable expression transcripts. A transcript from chromosome 7, Ta7alLoc000499, was identified as having the most stable expression across all experimental samples, with a variance of 21%. The 500th most stable expressed transcript, Ta1asLoc008673, had expression variance of 43%. Comparing the 1st and 500th ranked stable expressed transcripts shows the difference in variance between the two is relatively low at 22%. While utilization of transcript expression variance thresholds to screen for expression stability has not been widely documented in RNASeq literature our results show it is an important and intuitive first step in the selection of a suite of candidate housekeeping genes for use as expression benchmarks in downstream comparative analysis of differential expression between two or more transcripts.

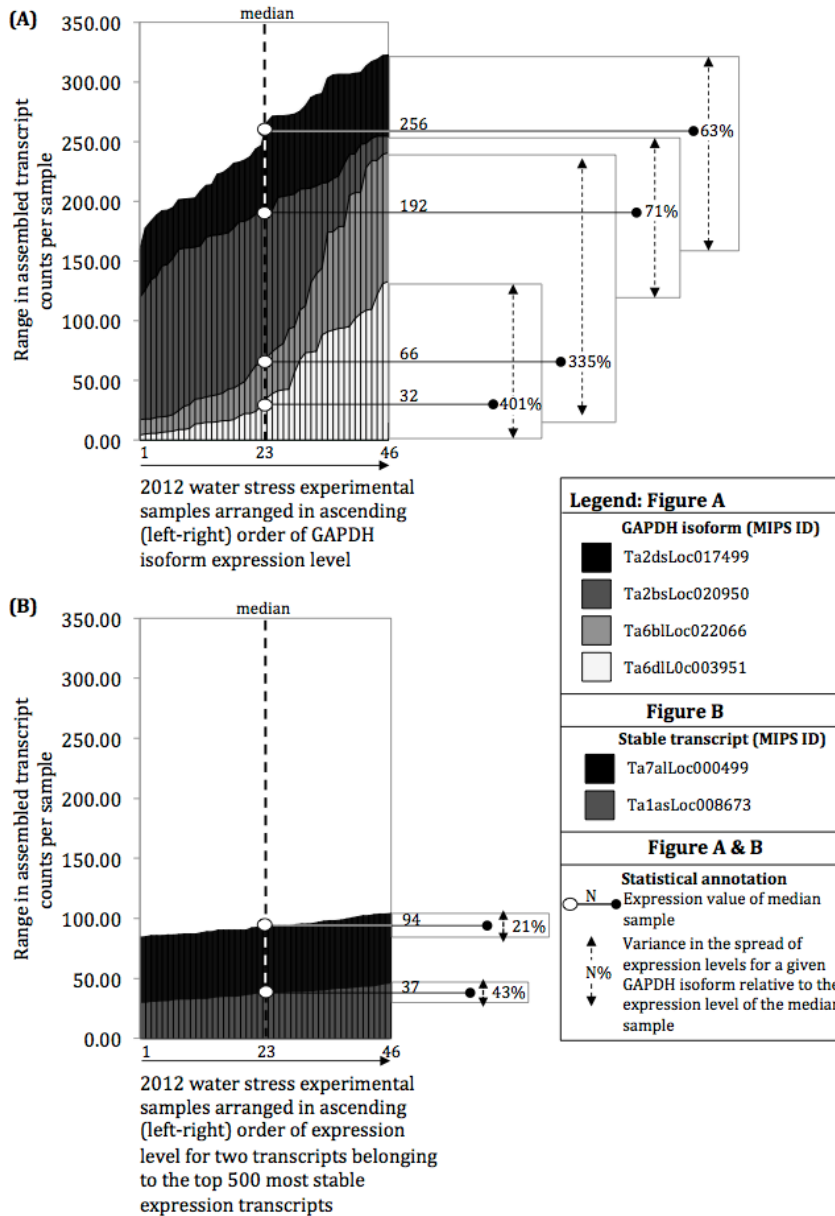


Figure 6.5 Stacked histograms showing variance in transcript expression, relative to the median expression value for each transcript, detected in samples from the 2012 RSWD stress experiment, for: (A) all GAPDH isoforms detected, and (B) transcripts ranked 1st and 500th in the list of stable expression transcripts. Expression level per sample for each transcript is plotted along the x-axis with data arranged left to right in order of ascending expression. The median expression value is denoted by a vertical broken black line in both histograms. The threshold range between the lowest and highest expression value for each transcript is annotated to the left of the histograms and corresponding to the median expression value and variances are listed there in. A legend inset to the right provides the visual explanation for colour coding of data for each transcript in the stacked histograms.

6.4.3 Global analysis for transcriptome wide changes in expression

Transcriptome wide analysis of the 2012 RSWD stress experiment RNASeq data showed a total of 14,627 transcripts were differentially expressed (FDR cut-off

0.02) in response to RSWD stress in at least one sample. Transcripts found to be differentially expressed represented 7% of the 1.2×10^5 transcripts mapped in the global data set. Because the genome reference used for mapping (i.e. MIPS annotation of the wheat GSS) does not at present account for all genes in the wheat genome it is expected some transcripts expressed in the data set that were not mapped may have been water deficit stress responsive in their expression. While refinement of the MIPS annotation would increase the overall number of transcripts mapped, and in turn increase the number of transcripts detected for differential response to RSWD stress, we expect our results for the overall proportion of the transcriptome that is water deficit stress responsive (7%) would not change significantly.

Transcripts found to be differentially expressed in response to RSWD stress were analysed for GO term classification (Table 6.3). This result presents, firstly, GO term classification (2nd level categories) of the global 2012 transcriptome data and the number of transcripts represented in each GO term category, and secondly the proportion of transcripts within each 2nd level category that were differentially expressed in response to RSWD stress. Information summarized in this table highlights two important findings:

(1) the cumulative total for transcripts representing 1st level GO terms including molecular function, biological process, and cellular component showed that each 1st level group had between 18-19% of the transcripts effected by RSWD stress. This results shows that while certain 2nd level GO terms associated with 1st level categories (i.e. molecular function, biological process, or cellular component) may be more responsive to RSWD stress each 1st level group is effected in equal measure, evidence which points to stress response being enacted at a whole of plant level,

(2) at the global data set level there is no clear correlation between the number of transcripts represented in a given 2nd level GO term category and the proportion of transcripts showing differential expression in response to RSWD stress in that particular category. This conclusion is highlighted by the fact some highly represented biological processes GO terms, such as those relating to localisation (which comprised 2046 transcripts), showed 15% of the group as

RSWD stress responsive. Conversely, of the 8815 transcripts comprising metabolic processes, from Table 6.3 it can be seen that 21% of this group were differentially expressed. Similarly the correlation was also absent in GO term categories represented by only very few transcripts.

In parallel to GO term analysis for identification of differentially expressed transcription factors (TF), the complete MIPS Wheat GSS annotations set was Blast'ed against the Wheat Transcription Factor database (http://planttfdb_v1.cbi.pku.edu.cn:9010/web/index.php?sp=tathat) to allow determination of the approximate number of TF's in the wheat genome. Blast results showed that 4009 of the MIPS Wheat GSS gene models had >75% identity to predicted TF's. Differential expression analysis of the global RSWD stress experiment RNASeq data subsequently showed that of the 4009 putative TF's 1602 were differentially expressed (equating to 39%) in response to the +RSWD stress treatment. The agreement between results obtained via GO term analysis and blasting the Wheat Transcription Factor database (http://planttfdb_v1.cbi.pku.edu.cn:9010/web/index.php?sp=ta), both of which showed in our data set TF's are among the most differentially expressed class of genes in response to RSWD stress, further validates consensus claims from the literature that TF's are one of the major drivers of plant response to water deficit stress.

Table 6.3 Gene ontology (GO) classification of transcripts differentially expressed in response to RSWD stress in the 2012 experiment. Each row of data comprises left to right details for: (A) classification according to 1st level GO term, (B) classification according to 2nd level Go term, (C) the number of transcripts that were assigned to a given 2nd level GO term. Two columns comprise the data for number of transcripts assigned to a given 2nd level GO term: Left – total number of transcripts, including non differentially expressed and differentially expressed, and Right – the number of differentially expressed transcripts only. Percentages listed in brackets behind the value for differentially expressed transcripts shows the portion which the value represent, with respect to the total number of transcripts assigned the given 2nd level GO term.

(A) Level 1 GO-term category	(B) Level 2 GO-term category	(C) Number of transcript assigned to 2 nd level GO-terms	
		Total number of transcripts (including differential and none differential expression)	Differentially expressed transcripts only
Biological process	biological adhesion	14	2 (14%)
	locomotion	23	3 (13%)
	positive regulation of biological process	23	7 (30%)
	rhythmic process	23	6 (26%)
	growth	38	4 (11%)
	negative regulation of biological process	63	9 (14%)
	immune system process	70	14 (20%)
	multi-organism process	168	25 (15%)
	reproductive process	257	41(16%)
	reproduction	302	52 (17%)
	developmental process	370	58 (16%)
	multicellular organismal process	407	65 (16%)
	biogenesis	954	205 (21%)
	response to stimulus	1727	355 (21%)
	regulation of biological process	1799	345 (19%)
	biological regulation	1896	357 (19%)
	localization	2046	315 (15%)
	cellular process	8120	1528 (19%)
metabolic process	8815	1821 (21%)	
	Total	27115	5212 (19%)
Molecular function	guanyl-nucleotide exchange factor activity	10	0 (0%)
	protein binding transcription factor activity	23	7 (30%)
	nutrient reservoir activity	51	7(14%)
	molecular transducer activity	153	28 (18%)
	enzyme regulator activity	179	30 (17%)
	antioxidant activity	194	57 (29%)
	nucleic acid binding transcription factors	305	57 (19%)
	receptor activity	326	60 (18%)
	signalling	609	112 (18%)
	electron carrier activity	791	138 (17%)
	transporter activity	965	150 (16%)
	structural molecule activity	1010	139 (14%)
	establishment of localization	2026	312 (15%)
macromolecular complex	2583	477 (18%)	

	single-organism process	5976	1245 (21%)
	catalytic activity	7590	1529 (20%)
	binding	8991	1692 (19%)
	Total	31782	6040 (19%)
Cellular component	side of membrane	7	2 (29%)
	cell junction	8	1 (13%)
	membrane-enclosed lumen	194	33 (17%)
	extracellular region	502	128 (25%)
	organelle part	1964	368 (19%)
	membrane part	2899	485 (17%)
	membrane	3628	630 (17%)
	organelle	6145	1148 (19%)
	cell	8131	1514 (19%)
	cell part	8131	1514 (19%)
	Total	31609	5823 (18%)

In addition to tracking changes in transcriptional profiles in response to RSWD stress an additional focus was to monitor temporal changes occurring throughout the course of reproductive stage development (AD5-FHE) and to monitor differences between genotypes. No significant spike transcriptome wide effect was detected for differential expression between genotypes (WxK DH lines D02-105 and D08-299) in the 2012 experiment. However, genotype was found to have a significant effect at the individual transcript level, where by a number of key transcripts showed genotype specific expression profiles. In a number of cases these transcripts were found to be involved in important processes underpinning the primary trait of interest in this thesis, tolerance to RSWD stress, and the expression results for these transcripts are presented in detail in the following sections. The effect of time (growth stage) was found to be highly significant at a whole spike transcriptome level. Particularly striking was the trend for increased expression over time with the highest number of transcript being expressed at FHE (data not shown).

While statistical analysis of the 2012 data showed there was no spike transcriptome-wide effect on transcript expression in response to genotype, results presented in Figure 6.6 suggest there was a clear interaction between time (growth stage) and genotype and also between time and treatment (+/- RSWD stress). Figure 6.6 comprises four genotype by treatment histograms showing pairwise comparison for the number of transcript differentially expressed between growth stages: AD5 vs. AD10, AD10 vs. AD15 and AD15 vs.

FHE. Each column of data represents the total number of transcripts differentially expressed between two growth. Transcripts comprising data for each column may have been expressed at an earlier growth stage in the comparison but not differentially expressed. Example: in histogram (A)(D02-105, +RSWD stress) in the pairwise comparison between AD5 vs. AD10 shows 2815 transcripts were differentially expressed in AD10 that were not differentially expressed in AD5 but they may still have been expressed at AD5 but not differentially.

Figure 6.6(A) shows that expression in D02-105 under control conditions (-RSWD stress) is relatively stable between growth stages ranging from AD5-AD15 but at onset of the FHE growth stage a whole suite of transcripts are differentially expressed that were either previously stable at AD15 or were not expressed at all. Of the 22,737 transcripts differentially expressed in a stage specific manner at the FHE stage, the proportion found to be up regulated as compared to down regulated was approximately equal. Conversely, Figure 6.6 (B) shows onset of differential expression in D02-105 +RSWD stress conditions is highly contrasting compared to the -RSWD stress conditions. This result highlights that D02-105, under +RSWD stress conditions, exhibits significant stage specific differential transcript expression at growth stages AD10, AD15, and FHE, with the highest proportion occurring at the AD15. Of the 12,647 transcripts found to be uniquely differentially expressed at AD15 the proportions found to be up regulated as compared to down regulated were also approximately equal.

Figure 6.6(C) shows that expression in D08-299 under control conditions(-RSWD stress) is relatively stable between growth stages ranging from AD5-AD10, with only a very small number of transcripts (206) showing up regulation at the AD10 stage. There after stage specific differential expression was detected at both the AD15 and FHE stage, the total values for which were 13,586 and 10,175 respectively. Comparisons between the up-regulated versus down-regulated proportions between AD15 and FHE show only marginal differences

were evident for up-regulation, while a significant drop occurred in the number of uniquely down regulated transcripts at FHE.

Similar to D02-105, D08-299 also showed highly contrasting results for stage specific differential expression in the +RSWD stress plants as compared to the -RSWD stress plants. In Figure 6.6(D) it is evident that stage specific differential expression occurs under +RSWD stress conditions most strongly at the AD10 and AD15 stage. The number of uniquely up-regulated genes between the two stages is not significantly different but the number found to be down regulated is significantly lower at AD15 at 1,155, as compared to AD10 that had 4,164.

These global analyses provide clear evidence of genotype specific temporal transcript profiles. Broadly speaking, both D02-105 and D08-299 follow a trend for increased differential expression at the mid to late stages of reproductive development (AD15-FHE). While both D02-105 and D08-299 show clear differences for stage specific differentially expressed transcripts in their water deficit stress responses, the comparison for stress response between the two lines shows striking differences. Particularly noteworthy is the difference seen between the two lines at the AD10 stage (data presented in the Figure 6.6 histograms in the AD5 vs. AD10 columns), which coincided with day 6-7 of the 10 day RSWD stress treatment imposed during the experiment.

At day 6-7 of the treatment in the 2012 experiment, plants assigned to the water deficit treatment group had reached moderate-severe levels of water deficit stress (refer to Figure 5.2 in chapter 5) and anthers in the developing spikes were undergoing the transition from pollen meiosis and the tetrad pollen stage (a time of development reported as being critically sensitive to water deficit stress Storme and Geelen (2013)). Given that the results in Figure 6.6 show D08-299 under +RSWD stress conditions exhibits a relatively high number (7031) of uniquely differentially expressed transcripts at the AD10 stage, as compared to D02-105 under +RSWD stress conditions (which showed only 2,808), there is sound evidence to claim D08-299 possesses molecular mechanisms which confer a much earlier stress detection and response. Moreover, taken in context with

results presented in Chapter 5 indicating that D08-299 was more tolerant to RSWD stress in both the 2011 and 2012 experiments compared to D02-105, it is plausible to suggest the capacity of D08-299 to fine-tune (differentially express) its transcript expression at the stage of critical sensitivity (pollen meiosis-tetrad pollen) is correlated to its superior level of RSWD stress tolerance.

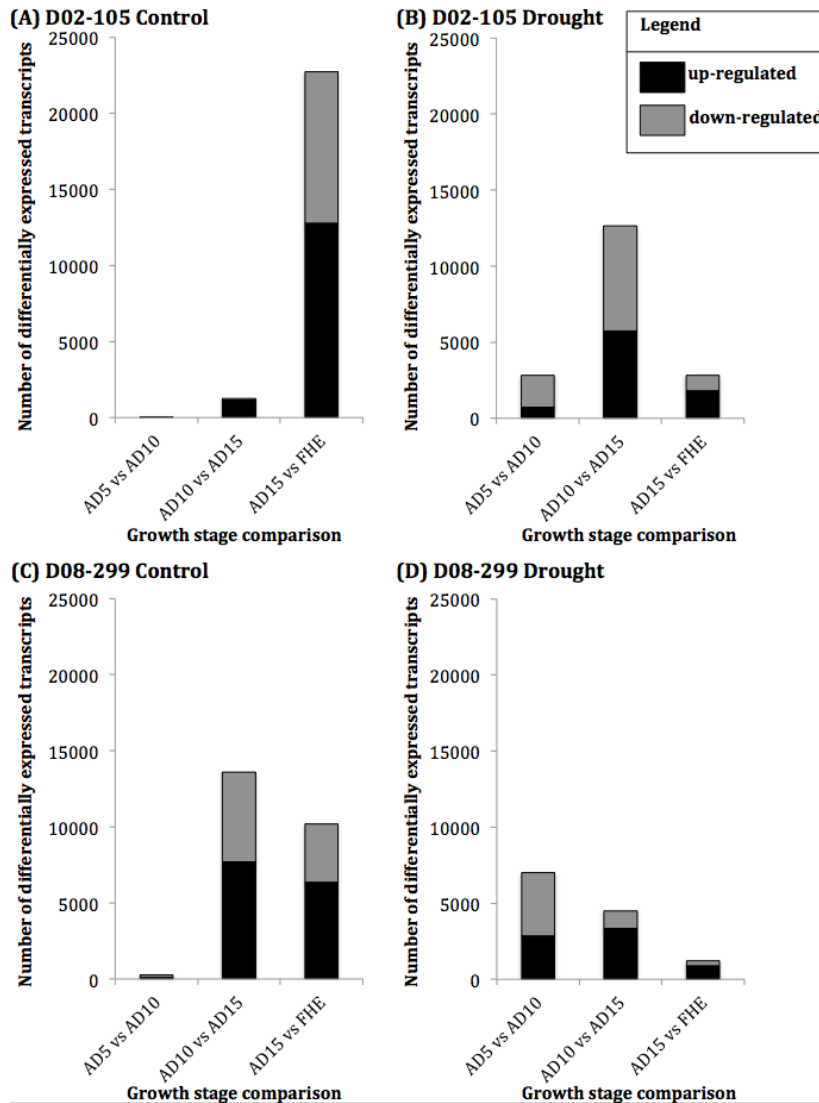


Figure 6.6 Pairwise comparison of differentially expressed transcripts for each growth stage sampled in the 2012 experiment. Only comparisons between sequentially order growth stages are presented and not between growth stages separated by one sampling interval.

6.4.4 *IVR1* transcript expression in developing spikes in response to plus and minus water stress during reproductive stage growth

Based on findings from Ji *et al.*, (2010), which showed *IVR1* plays an important function role in RSWD stress response in wheat, and the findings summarised in Chapter 5 of this thesis relating to the genetic mapping *IVR1-5BL* proximal to a

drought escape QTL (*QtAD5-5BL*) on chromosome 5BL, it was of interest to evaluate the expression profile of *IVR1* isoforms in the 2011 and 2012 RSWD stress experiments. Gene expression analysis of *IVR1* in the 2011 RSWD stress experiment utilized SQ-PCR as the assay platform and primers designed to have specificity for all five *IVR1* isoforms in the CW-INV family (*IVR1.1-3B*, *IVR1.2-3B*, *IVR1-4AL*, *IVR1-5BL*, and *IVR1-5DL*). Developing spikes were assayed at four AD elongation stages (AD3, AD5, AD6, and AD8), in four varieties/DH lines (Westonia, Kauz, D02-105 and D08-299) and in two differing water treatments (+/- RSWD stress). Despite considerable effort to optimize PCR thermocycler conditions over multiple rounds of PCR assays little or no *IVR1* expression was detected in the majority of the 96 samples collected in the 2011 experiment. The only sample group in which *IVR1* expression was detected was in the AD8 growth stage of WxK DH line D08-299. *IVR1* expression in the D08-299 samples was very low and while expression could be faintly detected, the PCR band illumination levels were still too low to be adequately captured as an photographic image using an agarose gel doc camera. Inability to perform image capture of the agarose gel bands also precluded the detection of possible differential expression in response to RSWD stress via agarose gel band light intensity analysis software.

The D08-299 AD8 *IVR1* PCR products were cut from the agarose gels for direct sequencing. Trace files obtained were used as probes to screen the IWGSC wheat GSS, and based on isoform specific SNP the *IVR1* isoform located on chromosome 5BL (*IVR1-5BL*) was identified as the transcript expressed in these samples. From these preliminary results a number of tentative conclusion were drawn: (1) the onset of *IVR1* expression is later in spike reproductive development than previously reported in studies by Ji *et al.*, (2010) and results from this thesis suggested that expression of the gene(s) may be closer to the tetrad pollen stage at AD10 (Z47) rather than pollen meiosis at AD5 (Z43) (See Figure 5.2 in Chapter 5), (2) the detection of *IVR1* expression in D08-299, but not in the other germplasm tested (i.e. Westonia, Kauz, and D02-105), is evidence of possible differential

expression between varieties/DH lines, either in response to RSWD stress or varietal differences in the temporal onset of gene expression, and (3) the detection of *IVR1-5BL* expression in D08-299 at AD8, but not *IVR1.1-3B*, *IVR1.2-3B*, *IVR1-4AL*, or *IVR1-5DL*, is evidence of possible isoform specific differential expression within a single variety, possibly also attributable to variation in temporal onset of gene expression.

Subsequent to the SQ-PCR performed on cDNA samples derived from RNA extracted from developing spike samples collected from the 2011 RSWD stress experiment, RNASeq was also performed on the same RNA samples. It was expected the deep sequencing capabilities of the RNASeq platform would likely detect even very low to minute transcript expression signals, such as the *IVR1*'s which may have been expressed, but not detected, in the SQ-PCR analysis of Westonia, Kauz, and D02-105 samples. Contrary to these expectations RNASeq performed on the 2011 RSWD stress experiment detected no transcript expression of any *IVR1* isoform family member in any of the 96 sample assayed, including D08-299 AD8 samples (i.e. the one sample group in which *IVR1-5BL* was detected via the SQ-PCR platform). These results provided the impetus to investigate what is the baseline sensitivity of the RNASeq platform to detect very low transcript expression

Based on the findings from the 2011 RSWD stress experiment, which highlighted the expression of *IVR1* is later than than previously thought, the sampling strategy for the 2012 RSWD stress experiment was refined. This process involved assaying developing heads using SQ-PCR for the expression of *IVR1*'s under control conditions only (i.e. -RSWD stress) from the AD3 growth stage through to two weeks post anthesis. Results from this refinement process (data not shown) indicated the most appropriate sampling strategy for the 2012 experiment should be to assay tissues incrementally at four time points from AD5 through to when the head is fully emerged from the sheath (Full head emerged FHE/Zadoks Z59).

An important difference between SQ-PCR results obtained from the refinement gene expression assays (conducted under –RSWD stress conditions only) and the 2011 RSWD stress experiment assays, was that in the refinement assays *IVR1* was found to be expressed as early as the AD5 growth stage in WxK DH line D02-105, while in the RSWD stress experiment assays it was found not to be expressed until AD8, but preferentially in D08-299 and not D02-105. To enable the refinement assays to be completed on the target growth stage tissues prior to the commencement of the 2012 experiment, new plants had to be germinated and grown ‘out of season’ over the 2011-2012 summer period. While glasshouse facilities used for all controlled studies in the current work have capabilities for control of ambient conditions (temperature and humidity) they do not have functionality for manipulation of light intensity or duration of light/dark interval.

The phenological development of plants sampled for the refinement assays was markedly faster (\approx ten days) than what was observed during the 2011 or 2012 RSWD stress experiments. The 2011 and 2012 experiments were planted in June and as such were subject to the light cues in keeping with the natural germination period of spring wheat grown in Australia. Although no light intensity or daylength data was recorded it is expected that daylength was a significant influencing factor on the rate of development of the plants grown for the *IVR1* gene expression refinement assays. The germination and seedling establishment under long day conditions most likely induced an increased rate of development in the plants in question, the likely causal factor for which being an earlier onset of gene expression in the major effect genes driving phenological development. Although likely confounded by duration and intensity of light interception factors the expression of *IVR1* at AD5 in D02-105 could not be ruled out as a result reflecting the natural expression onset of the gene in that particular DH line, and as such AD5 was selected as the starting point for tissue sampling in the 2012 RSWD stress experiment.

It should also be noted that while results obtained from the summer 2011/2012 refinement assays showed cessation of *IVR1* expression did not occur until

approximately seven days post anthesis, *IVR1* expression had dropped to almost non-detectable levels during the 2-3 days between full head emerged (FHE) and anthesis commenced (AC). The number of growth stages able to be sampled and processed for RNASeq was limited to four (i.e. to reduce cost of sequencing), hence the decision was made to target expression in developing spikes incrementally at the AD5, AD10, AD15, and FHE growth stages. The objective for selecting this sampling interval was that, based on the summer 2011/2012 refinement assay results, it would enable capture of the complete (or close to) expression profile of *IVR1* in both WxK DH lines D08-299 and D02-105.

Results from SQ-PCR analysis of the 2012 RSWD stress cDNA samples (Figure 6.7) showed *IVR1* expression was detected in both lines tested (D02-105 and D08-299), under both water treatments (+/- RSWD stress), at all growth stages sampled (AD5, AD10, AD15, and FHE), except for the AD5 stage of the -RSWD stress D08-299 treatment group. Overall, the levels of *IVR1* transcript expression, relative to *GAPDH* (the house keeping gene used), ranged from 10% to 95% of the expression level of *GAPDH*. Because primers used to assay *IVR1* expression were conserved across all *IVR1* isoforms it was not possible to discern which *IVR1* isoforms were being expressed in any given treatment group. Direct sequencing was undertaken to determine which *IVR1* isoforms were being expressed (results discussed in more detail in following paragraphs) but prior to this work being undertaken it was assumed expression levels presented in Figure 6.7 were cumulative totals of all *IVR1* isoforms being expressed in a given sample group.

Image capture and measurement of illumination intensity of SQ-PCR product bands on agarose gels was used to perform a statistical evaluation of differential expression between treatment groups for the 2012 RSWD stress experiment. An example is provided in Figure 6.7 for D02-105 at the AD5 sampling stage, where by an agarose gel image comprising the *IVR1* SQ-PCR bands assayed from one replicate of the +RSWD stress treatment group and one replicate of the -RSWD stress treatment group is positioned above the histogram columns for the same treatment groups respectively. This example shows clearly that

illumination intensity of the *IVR1* PCR band in the -RSWD stress sample is significantly higher than in +RSWD stress sample. No result equivalent to Figure 6.7 is presented for the 2011 RSWD stress experiment, due to non-detectable *IVR1* expression via image capture in the assays on these samples.

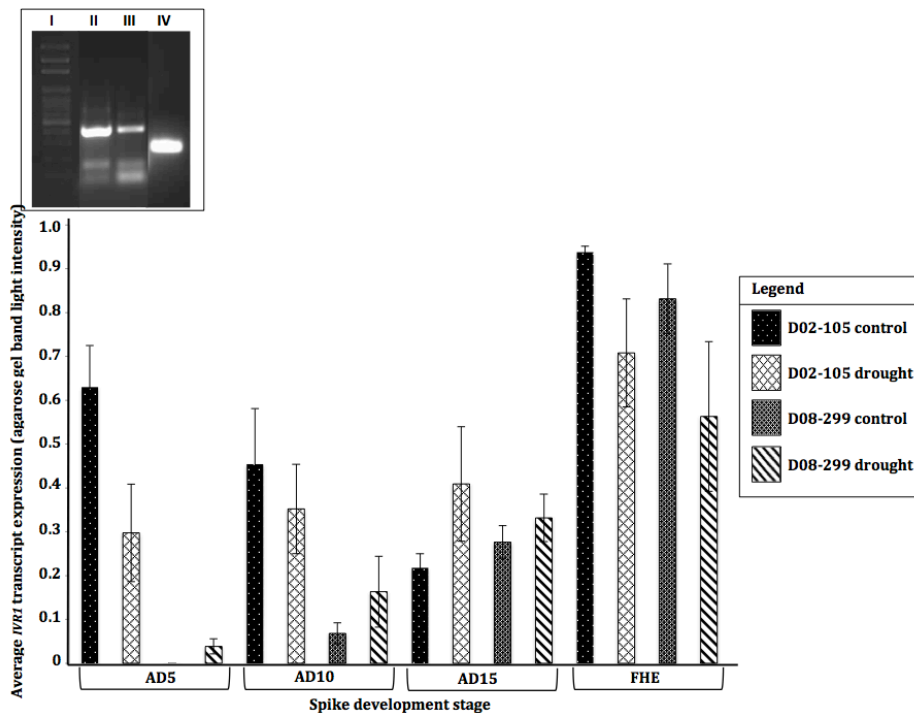


Figure 6.7 *IVR1* gene expression data (SQ-PCR) using primers designed to assay the whole suite of *IVR1* isoforms in the WxK DH cDNA samples from the 2012 RSWD stress experiment. Data was accumulated on developing spike tissue collected incrementally throughout the course of penultimate internode auricle distance (AD) extension, including AD5, AD10, and AD15, and culminating at full head emerged (FHE). Columns comprising the histogram are arranged in groups along the x-axis according to growth stage at sampling. Columns represent mean transcript expression level have been annotated with standard error bars for visualisation of significant differences between the varietal treatment group each column represents. The y-axis scale presents the unit of transcript expression measurement (agarose gel band illumination intensity), which uses a scale from 0-1, where zero is the complete absence of expression and 1 is maximal expression. The agarose gel band intensity of *GAPDH* (the house keeping gene) was used as the standard and thus the scale on the y-axis should be interpreted as *IVR1* expression as a proportion of *GAPDH*. Because primers used to assay *IVR1* had selectivity for all *IVR1* isoforms the PCR band light intensities reflect a cumulative quantification comprising all *IVR1* isoforms expressed in a given sample. A Figure legend inset right shows the colour scheme that defines the WxK DH line number and water treatment group (+/- RSWD stress) for each column. Positioned top left of the Figure is a representative SQ-PCR agarose gel image showing *IVR1* genes being assayed from the developing spike of WxK DH line D02-105 at AD5. Gel lane numbering corresponds to: (I) DNA ladder, (II) *IVR1* expressed in cDNA from D02-105 -RSWD stress, (III) *IVR1* expressed in cDNA from D02-105 +RSWD stress, and (IV) *GAPDH* expressed in cDNA from D02-105 -RSWD stress. The agarose gel image is positioned above the histogram so that PCR bands in lane II and III are directly above the column representing the mean gene expression for the treatment group to which they belong.

Analysis of variance (ANOVA) was performed on the 2012 RSWD stress experiment SQ-PCR data (Table 6.4). These results showed that there was no overall statistically significant effect on *IVR1* expression from experimental factors including: replicate (1, 2, and 3), variety (D08-299 versus D02-105), or water treatment (+/- RSWD stress). A significant effect on *IVR1* expression was detected from two experimental factors including: growth stage (AD5, AD10, AD15, and FHE) and Tissue (spike versus leaf). Table 6.4 also shows a significant overall effect on *IVR1* transcript expression caused by two factorial interactions, firstly between variety and growth stage, and secondly between growth stage and tissue type. No significant effect on *IVR1* transcript expression was detected for interactions between RSWD stress treatment type with variety, growth stage, or tissue.

Table 6.4 Analysis of variance (ANOVA) of *IVR1* band illumination intensity on agarose gels from SQ-PCR assays on samples collected in the 2012 RSWD stress experiment. Abbreviations listed in column headings denote the following: DF = degrees of freedom, SS = sums of squares, MS = mean sums of squares, F = F-statistic, P = P-value. Experimental factors evaluated for main effects and interaction effects on *IVR1* expression were: replicate, variety, growth stage, tissue type, and RSWD stress treatment type. Main effects from and interactions between experimental factors were deemed have a significant effect on *IVR1* expression when $P < 0.05$.

	DF	SS	MS	F	P
Rep	5	3965.5	793.1	1.77	0.14
Variety	1	813	813	1.82	0.18
Growth stage	3	25740.2	8580.1	19.2	0.001
Tissue type	1	2840.6	2840.6	6.35	0.015
water treatment type	1	390.7	390.7	0.87	0.36
Variety - Growth stage	3	5557.4	1852.5	4.14	0.011
Variety - Tissue type	1	6862.5	6862.5	15.3	0.001
Growth stage - Tissue type	0	0	*		
Variety - Treatment	1	145.1	145.1	0.32	0.57
Growth stage - Treatment	3	3444	1148	2.56	0.07
Tissue type - Treatment	1	599.3	599.3	1.34	0.25
Variety - Growth stage - Tissue type	0	0	*		
Variety - Growth stage - Treatment	3	569.7	189.9	0.42	0.74
Variety - Tissue type - Treatment	0	0	*		
Growth stage - Tissue type - Treatment	0	0	*		
Residual	48	21483.3	447.6		
Total	71	72411.2	1019.9		

Because ANOVA was limited to detection of overall effects driving differential expression of *IVR1* transcripts it was also of interest to undertake pairwise comparison for detection of differential expression between single treatment groups (e.g. D02-105 head AD5 -RSWD stress versus D02-105 head AD5

+RSWD). In Figure 6.7 standard error bars calculated for mean *IVR1* expression across three replicates (per experiment treatment group) allowed for pairwise comparisons to be made, many of which were found to be significantly different. Utilizing this approach *IVR1* transcript expression was identified as being differentially expressed in D02-105 in response to water stress treatment type (+/- RSWD stress) at growth stages AD5, AD15, and FHE. However, the direction of the effect was not consistent at all growth stages. The AD5 and AD15 showed *IVR1* transcript expression was down-regulated in the +RSWD stress samples (relative to the -RSWD stress samples) but at AD10 the effect was for *IVR1* transcript expression to be up-regulated in response to +RSWD stress samples (relative to the -RSWD stress samples).

Apparent in Figure 6.7 is a trend for *IVR1* transcript expression in D08-299 to be up-regulated in the +RSWD stress samples (relative to the -RSWD stress samples) from growth stages AD5-AD15, but overlap of standard error bars (at all three growth stages) shows the effect of water treatment type was not significant. In contrast the effect of water treatment on *IVR1* transcript expression was significant at the FHE growth stage and the +RSWD stress treatment caused expression to be down-regulated relative to the -RSWD stress condition.

While standard error bars of *IVR1* transcript expression in Figure 6.7 confirm the significant effect detected by ANOVA for the interaction between variety and growth stage, they also highlight as incorrect the ANOVA non-detection of a significant three-way interaction between variety, water treatment, and growth stage. Comparison of standard error bars between the two DH lines at each of the growth stage under the same water treatment type (e.g. D02-105 AD5 - RSWD versus D08-399 AD5 - RSWD) show:

- (1) the level of *IVR1* expression between the two lines is significantly different under +RSWD stress at the AD5 and AD10 but not AD15 or FHE, and
- (2) the difference under -RSWD stress is also significantly different at AD5 and AD10 but not AD15 or FHE.

Because primers used to assay *IVR1* transcript expression in the SQ-PCR assays were conserved across all *IVR1* isoforms it was of interest to undertake direct sequencing of the assayed products to individually identify isoforms that were expressed in the 2012 RSWD stress experiment samples. Figure 6.8 presents a multiple alignment of a portion of the transcript sequences assayed by the *IVR1* primers in the 2012 RSWD stress experiment samples. Based on *IVR1* isoform annotation results (presented in Chapter 4) isoform specific SNPs present in the individual trace files obtained from direct sequencing were able to be assigned to the *IVR1* isoform of origin.

Figure 6.8 is a representative alignment only (i.e. it does not present all trace files generated and as such does not present the full frequency of isoform specific SNPs), comprising all isoform specific SNPs represented in the collective trace file results for the 2012 RSWD stress experiment. Assignment of SNPs to isoforms of origin showed that amongst all samples evaluated *IVR1* isoforms expressed included *IVR1.1-3B*, *IVR1.2-3B*, *IVR1-4AL*, and *IVR1-5BL*. However the frequency of *IVR1-5BL* specific SNPs (data not shown) collectively across all samples was >75%, evidence from which it was concluded that *IVR1-5BL* was the dominant isoform being expressed in the developing spike.

The dominant expression of *IVR1-5BL* was also taken as evidence that different isoforms are not differentially expressed in developing spikes in response to type of water treatment. Preferential expression was detected however between tissue type, where by *IVR1.1-3B* and *IVR1.2-3B* were found to be the only *IVR1* isoforms expressed in flag leaf tissue at the AD5 stage and their expression in spike tissues was extremely low. The SNP results in Figure 6.8 also showed that the *IVR1* isoform located on chromosome 5DL (*IVR1-5DL*) was not expressed in any of the 2012 RSWD stress samples.

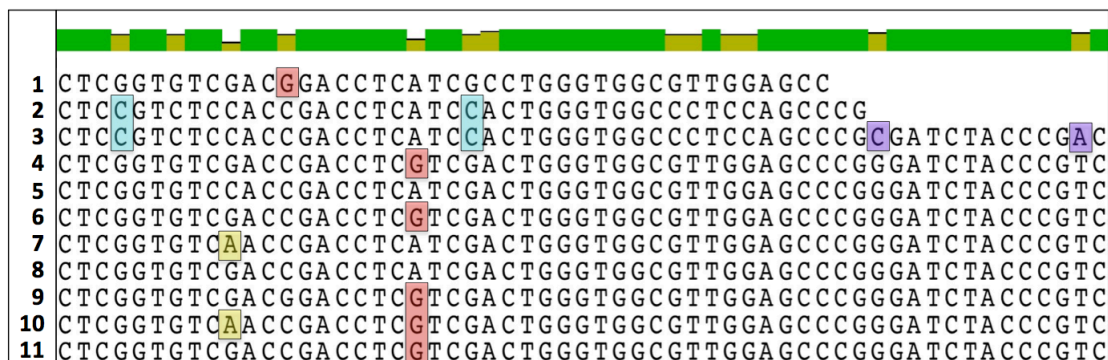


Figure 6.8 Multiple alignment of *IVR1* sequences obtained from direct sequencing of SQ-PCR assays on the 2012 RSWD stress experiment using primers designed to have selectivity for all members of the *IVR1* family. The alignment shows a portion of the region assayed for 11 individual SQ-PCR gel bands. SNPs in each sequence have been annotated to show the specific isoform from which it originates; aqua = *IVR1.1-3B*, purple = *IVR1.2-3B*, yellow = *IVR1-4AL*, red = *IVR1-5BL*. SNP specific to *IVR1-5D* are not present in the region of the transcript shown.

Collectively the SC-PCR-based *IVR1* transcript expression results presented in Figure 6.7, Table 6.4, and Figure 6.8 provided the first clear evidence in this work that *IVR1* expression in response to RSWD stress is strongly genotype specific. Moreover these results highlighted that *IVR1* expression occurs most strongly in the later stages of reproductive development at the AD15 and FHE stage (i.e. uninucleate microspore) as opposed to coinciding with pollen meiosis earlier in development (i.e. AD3 to AD5). Importantly the discovery of *IVR1.1-3B* and *IVR1.2-3B* expression in the flag leaf (discussed in following paragraphs) also provided evidence to question the accuracy of earlier claims by Oliver *et al.*, (2005) that *IVR1* genes are specifically expressed in anther tissue. The Oliver *et al.*, (2005) findings were from rice however and thus it is possible there are tissue expression specificity difference between wheat and rice.

Figure 6.9 shows a histogram of the pairwise comparison of *IVR1* expression from the SQ-PCR analysis of the 2012 RSWD stress experiment between samples from contrasting tissues at the AD5 growth stage (i.e. spike tissue versus flag leaf). Leaf tissue was sampled only at the AD5 growth stage. Mean expression values of *IVR1* transcripts (and associated standard error bars) plotted in Figure 6.8 are consistent with ANOVA results in Table 6.4, which indicated there was a significant main effect from tissue type on *IVR1* expression. Taken in context with the direct sequencing results from Figure 6.8 it can be concluded that data

for *IVR1* transcript expression in the flag leaf (presented Figure 6.9) comprises expression from both *IVR1.1-3B* and *IVR1.2-3B*. From Figure 6.9 it can also be seen that the expression level (i.e. of *IVR1.1-3B* and *IVR1.2-3B* combined) is not significantly effected by water treatment (i.e. +/- RSWD stress) in D02-105 but in D08-299 expression is upregulated by the +RSWD stress treatment.

Because direct sequencing results confirmed *IVR1-5BL* is the dominant *IVR1* isoform expressed in developing spike tissues it can also be concluded from Figure 6.9 that, at the AD5 stage, there are significant isoform specific differences between *IVR1* expression in the flag leaf (i.e. *IVR1.1-3B* and *IVR1.2-3B*) and the spike (*IVR1-5BL*). In D02-105 at the AD5 stage, *IVR1-5BL* was expressed in the spike at significantly higher levels than *IVR1.1-3B* and *IVR1.2-3B* in the flag leaf, under -RSWD stress conditions, but not under +RSWD stress conditions. The reverse effect was evident in D08-299, where by under -RSWD stress conditions *IVR1-5BL* was not expressed in the spike at AD5 at all, but *IVR1.1-3B* and *IVR1.2-3B* were expressed in the flag leaf. Under +RSWD stress conditions D08-299 expressed *IVR1-5BL* at very low levels which are significantly less than the combined level of *IVR1.1-3B* and *IVR1.2-3B* in the flag leaf.

As highlighted in section 6.4.3 (Global RNASeq analysis) spike transcriptome wide gene expression profiles of D08-299 provide evidence to suggest this DH line possesses the capacity for early detection of, and superior response to, RSWD stress. Results from Figure 6.9 also seem consistent with this conclusion, where by significant isoform and tissue specific upregulation of *IVR1.1-3B* and *IVR1.2-3B* in the flag leaf (i.e. the tissue thought to be the initial site of water deficit stress detection (Dorion *et al.*, 1996) may be part of the transcriptome wide effect seen in D08-299 for differential expression of key water deficit stress response transcripts.

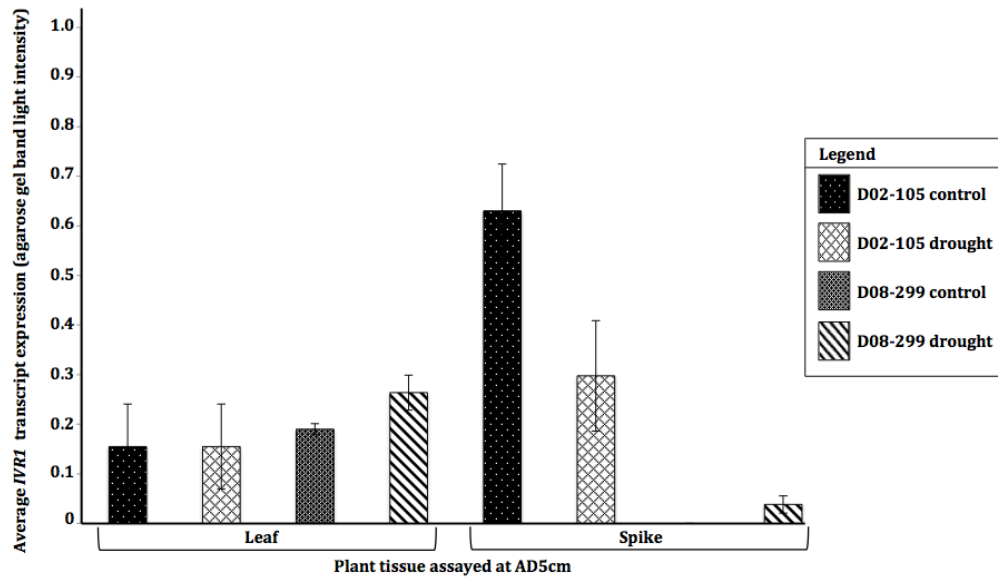


Figure 6.9 Tissue specific (spike and flag leaf) *IVR1* gene expression data (SQ-PCR) at early reproductive stage development (AD5) from the 2012 RSWD stress experiment. Primers used were designed to assay the whole suite of *IVR1* isoforms in wheat. Columns comprising the histogram are arranged in two groups along the x-axis according to tissue sample. Columns represent mean transcript expression level have been annotated with standard error bars for visualisation of significant differences between the varietal treatment group each column represents. The y-axis scale presents the unit of transcript expression measurement (agarose gel band illumination intensity), which uses a scale from 0-1, where zero is the complete absence of expression and 1 is maximal expression. The agarose gel band intensity of *GAPDH* (the house keeping gene) was used as the standard and thus the scale on the y-axis should be interpreted as *IVR1* expression as a proportion of *GAPDH*. Because primers used to assay *IVR1* had selectivity for all *IVR1* isoforms the PCR band light intensities reflect a cumulative quantification comprising all *IVR1* isoforms expressed in a given sample. A Figure legend inset right shows the column fill pattern that defines the WxK DH line number and water treatment group (+/- RSWD stress) for each column.

RNASeq mapping results of short read transcript data from the 2011 RSWD stress experiment were broadly consistent with SQ-PCR results generated on the same RNA samples. As mentioned in section 6.1.4.4, *IVR1-5BL* was the only *IVR1* isoform for which expression was detected via SQ-PCR in developing spike tissue in the 2011 RSWD stress experiment, and amongst all treatment groups comprising the experiment, it was only expressed in D08-299 at the AD8 growth stage in both the +RSWD stress and -RSWD stress samples. Despite the widely accepted deep sequencing capabilities of the RNASeq platform to detect even very low expression transcripts, *IVR1* transcript expression was not detected in any samples in the 2011 RSWD stress experiment, not even D08-299 at the AD8 growth stage (i.e. the sample in which *IVR1-5BL* had previously been detected by SQ-PCR).

RNASeq mapping results for *IVR1* in the 2012 RSWD stress experiment samples (Figure 6.10) differed from those obtained for 2011, where by the 2012 results showed that amongst developing spike samples four *IVR1* isoforms were expressed and detected at four growth stages during spike development. It is expected that variability in *IVR1* transcript expression between the two years of experimental data was likely due, in part, to a longer and later sampling interval in the 2012 (see Chapter 3, section 3.3.5 sampling schedule and tissue collection).

Comparison between *IVR1* transcript expression results generated on the 2012 RSWD stress experiment samples via the RNASeq platform (Figure 6.10) as compared to SQ-PCR (Figure 6.7 and 6.9) highlight some striking differences, which collectively point to a possible limitation of using RNASeq for detection of low expression transcripts and for the detection of differential expression between treatment groups on low expression transcripts. Two points of evidence provide grounds for this claim, firstly: via SQ-PCR methods expression of *IVR1.1-3B* and *IVR1.2-3B*, transcripts at the AD5 stage was detected in flag leaf samples from both DH lines tested and under both +RSWD and -RSWD stress treatments (Figure 6.9). RNASeq results detected no *IVR1* expression in the same flag leaf sample, including *IVR1.1-3B* and *IVR1.2-3B* transcripts, in either DH line under either water treatment type. Secondly, SQ-PCR results in Figure 6.7 showed that *IVR1* transcript expression at the AD10 and AD15 stage was detected at levels ranging from 10-50% of the expression of the house keeping gene (*GAPDH*). In contrast, while RNASeq did detect *IVR* expression at the same sampling stages, the range of expression levels detected were all extremely low, <9 counts across three replicates of data, values which were too low for detailed analysis.

Comparative analysis between Figure 6.7 and 6.10 provide strong grounds to question the overall sensitivity of the RNASeq platform for the detection of low expression transcripts. Within these limitations, there were some important overall trends conserved in the results obtained from the the two different

analysis platforms with regards to *IVR1* expression profiling in the 2012 RSWD stress experiment samples. Figure 6.7 and 6.10 show consistently that *IVR1* transcript expression levels at the FHE growth stage were significantly higher than the three growth stages sampled prior (i.e. AD5, AD10, and AD15). This finding is particularly important in that it provides further evidence to argue against the concept developed by Ji *et al.*, (2010) that *IVR1* has functional importance as a key gene underpinning pollen fertility, based on findings from their study that the gene is preferentially expressed at the most vulnerable stage of development, pollen meiosis (i.e. at the AD3-5 growth stage).

An important difference between Figure 6.7 and Figure 6.10 is the calling of differential expression at the FHE stage. In Figure 6.7, *IVR1* expression was identified, via SQ-PCR, as significantly down-regulated in response to +RSWD stress in both WxK DH lines tested. In Figure 6.10 the RNASeq data showed a trend in down-regulation of all three isoforms expressed (*IVR1-4AL*, *IVR1-5BL*, and *IVR1-5DL*) in response to +RSWD stress in D02-105 but not D08-299. Figure 6.10 also shows that while no significant down-regulation in expression occurred amongs the three *IVR1* isoforms detected in D08-299 at FHE, the isoform *IVR1-5BL* showed the least amount of variation between expression levels for +RSWD stress conditions as compared to -RSWD conditions.

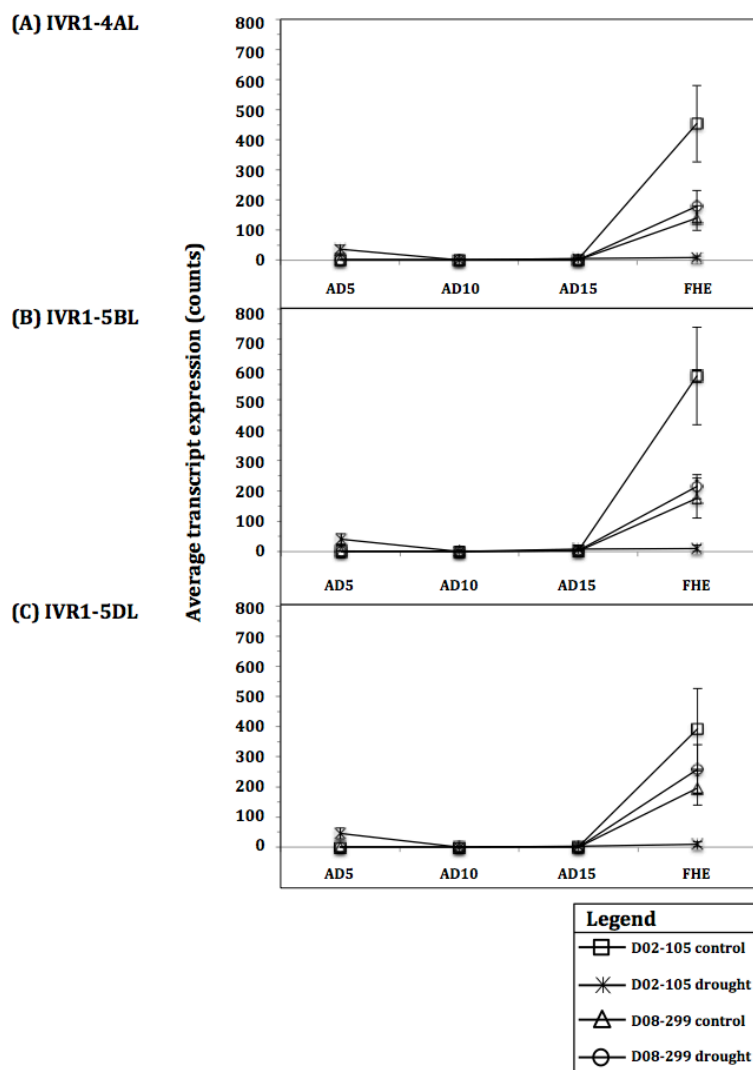


Figure 6.10 A line graph of *IVR1* isoform transcript expression across time (growth stage) during the course of reproductive development in spike tissue collected from the 2012 RSWD stress experiment for isoforms: (A) *IVR1-4AL*, (B) *IVR1-5BL*, and (C) *IVR1-5DL*. The unit of scale for mean transcript expression plotted against the y-axis is normalized counts. On the x-axis data is plotted left to right across growth stage including: AD3, AD5, AD10, and FHE. Differential expression between treatment groups can be visualised when standard error bars between the two data points in question do not overlap. A legend below right of the line graph provides the visual explanation of shaped which denote particular treatment groups for DH line number (D02-105 and D08-299) and water treatment group (+/- RSWD stress).

Similar to direct sequencing measures taken on the SQ-PCR assays of the 2012 RSWD stress experiment samples, isoform specific SNP were also used as landmarks to validate *IVR1* isoform specific mapping of transcript assemblies in the RNASeq data. Using *IVR1-5BL* as an example Figure 6.11 shows how isoform specific SNP in transcript assemblies were cross referenced against the particular MIPS Wheat GSS gene model used as a mapping reference for the transcript assembly (i.e. to confirm the same SNP was present in the mapping

reference sequence), hence confirming mapping of transcripts assemblies to the correct gene model.

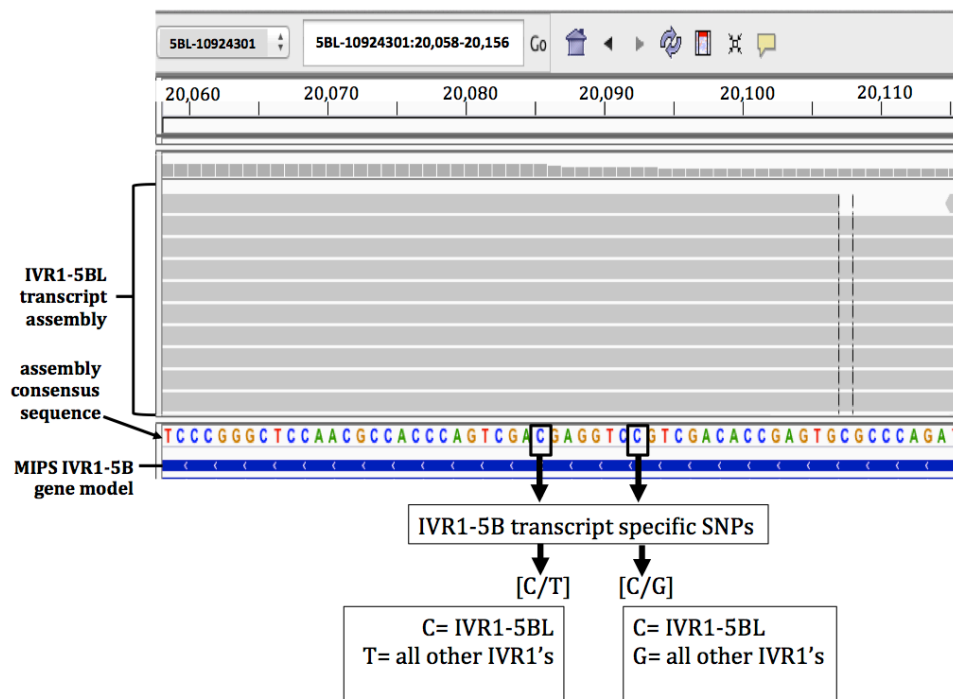


Figure 6.11 Multiple alignment using Integrative Genome Browser software (v 2.3, <http://www.broadinstitute.org/igv/>) of transcript short reads mapped to the MIPS Wheat GSS gene model for cell wall invertase *IVR1-5BL* from the 2012 RSWD stress RNASeq data set. Thick grey horizontal lines represent individual short reads mapped to the MIPS Wheat GSS gene model for *IVR1-5BL* (Ta5blLoc036480), which is represented at the bottom of the alignment as a thick blue line. Positioned in between the stack of *IVR1-5BL* transcript short reads and the MIPS gene model is the consensus sequence generated for the *IVR1-5BL* transcript assembly. In the consensus sequence two SNP unique to *IVR1-5BL* are denoted by enclosure in a black box. Collectively this Figure provides an example of how *IVR1* isoform specific SNP in individual short reads were used to ensure mapping of transcript assemblies was to the correct MIPS Wheat GSS gene model for a given *IVR1* isoforms.

6.4.5 Expression profiling of *QtAD5-5BL* associated transcripts in response to plus and minus reproductive stage water deficit stress

Transcript expression results presented in this section of chapter 6 are reported as overall trends for genotypic specific expression profiles. The term differential expression (DE) is used to denote identifiably ‘different’ transcript expression levels (i.e. in instances where standard error bars for two data points in a pairwise comparison do not overlap). Based on findings summarized in Chapter 5 (Figure 5.8) it was of interest to track the transcript expression profiles of genes located in the *QtAD5-5BL* region in D02-105 and D08-299 in response +/- RSWD stress conditions. Because *QtAD5-5BL* was found to be located in a linkage block

segregating in the two lines of interest it was hypothesized different expression patterns in response to +/- RSWD stress conditions between one or more of the allele variants could be contributing to the contrasting QTL phenotype and the grain set response to RSWD stress also observed between the two lines.

Figure 6.12 shows the expression profiles of seven of the twenty seven transcripts characterised from the *QtAD5-5BL* region (Chapter 5 Table 5.5) from the 2012 RSWD stress RNASeq data. It should be noted, that of the remaining twenty *QtAD5-5BL* associated transcripts not represented in Figure 6.12, a number of them were found to be expressed at some time points, under one or both water treatments. In all cases the expression level of this group of twenty transcripts was below thresholds for biological significance (i.e. <10 counts across three or more replicates). The known developmentally related gene, *VrnB1* was found to be in this category.

The functional characterization of the *QtAD5-5BL* (presented in Chapter 5) associated transcripts provided evidence to suggest that at least six of the seven transcripts might have functional importance to RSWD stress response processes. Among this group of transcripts there were three predicted transcription factors (Ta5bLoc028528, Ta5bLoc037737, and Ta5bLoc001591), a predicted protein kinase (Ta5bLoc001259), UDP-glucose-6-dehydrogenase (Ta5bLoc037731), and a predicted stritoidine synthase (Ta5bLoc038963). The functional importance to abiotic stress response processes in plants for the majority of these classes of genes have been previously reported, including: (1) transcription factors and their role in regulating expression at the transcriptional level (reviewed by Budak *et al.*, (2013)), (2) protein kinases and their role in abiotic stress response signal transduction and detection of drought induced ABA synthesis (reviewed by Tran *et al.*, (2013)), and (3) UDP-glucose-6-dehydrogenase for its role in modulating levels of UDP-glucose for maintenance of cell wall stability under dehydration stress (Hlavackova *et al.*, 2013). The predicted stritoidine synthase transcript (Ta5bLoc038963) belongs to a lesser studied class of anther specific genes, that have been shown to be upregulated in

response to abiotic stress and thought to be involved in pollen cell wall formation and degradation (Ageez *et al.*, 2005; Dobritsa *et al.*, 2009).

Both DH lines tested in the 2012 RSWD stress experiment, under both water treatments, showed the same pattern for transcript expression presence/absence among the group of *QtAD5-5BL* associated transcripts. However, Figure 6.12 shows clear genotypic differences exist between the expression profiles of the group of seven expressed transcripts, in response to +/- RSWD stress, extending from growth stage AD5 to FHE. The most striking contrast in expression profiles was apparent between D08-299 and D02-105 for the UDP-glucose 6-dehydrogenase (Ta5blLoc037731) transcript (Figure 6.12(G-H)). In D08-299 UDP-glucose 6-dehydrogenase expression under control conditions (-RSWD stress) was relatively stable and moderately expressed between growth stages AD5 and AD10, then significantly spiked at AD15, and at FHE dropped back to moderate expression levels (approximately equivalent to those observed between AD5-AD10). Under +RSWD stress conditions UDP-glucose 6-dehydrogenase expression in D08-299 was significantly upregulated (with respect to -RSWD stress expression levels) at growth stages AD5 and AD10, but significantly was down regulated at AD15 and FHE. From these results it can be concluded the overall effect of +RSWD stress on UDP-glucose 6-dehydrogenase expression in D08-299 is to cause a shift to an earlier expression peak. Despite the clear temporal shift no marked difference in abundance was apparent with respect to the level of expression observed in the low and high points of the expression profiles in the two treatments respectively.

The effect of +/- RSWD stress on UDP-glucose-6-dehydrogenase in D02-105 appeared not to cause the same temporal shift as observed in D08-299. In the case of D02-105 the primary effect +/- RSWD stress had on UDP-glucose 6-dehydrogenase expression in D02-105 was to alter transcript abundance (expression level) at growth stages AD10, and FHE. At both the AD10 and FHE growth stages expression of UDP-glucose 6-dehydrogenase was significantly down regulated by the +RSWD stress treatment, compared to -RSWD stress treatment.

When comparing the two DH lines to each other under the same water treatment conditions it can be seen from Figure 6.12(G) that there was a significant difference in abundance under the -RSWD stress conditions (Figure 6.12(G)) in all growth stages except the early development stage at AD5. The same is true for the +RSWD stress comparison, except the direction of the difference was reversed, in that the only non-significant difference was detected at the FHE stage. Overall the UDP-glucose 6-dehydrogenase expression results highlight some important trends, firstly that the Kauz 'like' UDP-glucose 6-dehydrogenase allele in D02-105 appears to be more water deficit stress susceptible to down regulation. Secondly, the transcript for the Westonia 'like' UDP-glucose 6-dehydrogenase allele in D08-299 appears unaffected at an abundance level by +RSWD stress but the allele also had a responsive expression mechanism to reduce the amount of time taken for peak expression to be achieved. Thirdly, UDP-glucose 6-dehydrogenase abundance level at peak expression in D08-299 under plus and minus RSWD stress conditions respectively was significantly higher than the peak expression levels in the equivalent treatments in D02-105. This significant contrast in abundance points to UDP-glucose 6-dehydrogenase potentially being a major effect allele driving the variation seen between the two DH lines with respect to the *QtAD5-5BL* phenotype.

The expression profile of the predicted stritosidine synthase transcript (Ta5blLoc038963) was also found to show clear within and between genotype differences in response to +/- RSWD stress (Figure 6.12(M-N)). Under -RSWD stress conditions Ta5blLoc038963 expression in D02-105 trends upwards from AD5 to FHE. Conversely in D08-299 Ta5blLoc038963 expression was stable between AD5-AD10, significantly increased (and then spiked) at AD15, and at FHE dropped back to moderate expression levels (approximately equivalent to those observed between AD5-AD10). Despite the apparent genotypic difference at a 'trend' level, crossover of standard error bars at time point AD5, AD10, and AD15 shows there was no significant difference in Ta5blLoc038963 expression between the two lines under -RSWD stress conditions, except at the FHE stage.

The +RSWD stress condition appears to cause a significant down regulation in Ta5blLoc038963 expression in D02-105 (relative to -RSWD stress treatment group) at the AD15 and FHE stage. For the equivalent plus versus minus RSWD stress comparison in D08-299 it can be seen that the stress had no significant effect on Ta5blLoc038963 expression at any time point. Interestingly comparison between the -RSWD stress expression profile for Ta5blLoc038963 and Ta5blLoc037731 (UDP-glucose 6-dehydrogenase) in D08-299 show the two genes share an almost identical expression trend, the only difference being the absolute expression level at each time point. No equivalent trend was observed in D02-105 data.

Expression evaluations of the three predicted transcription factors (Figure 6.12 (E-F), (I-J), and (K-L)) highlighted clear differences for expression response to +/- RSWD stress amongst this group. Broadly speaking in the D02-105 assays the expression profiles of each of the three transcripts, Ta5blLoc028528 (*3C polypeptide 3*), Ta5blLoc037737 (*WRKY 57*) and Ta5blLoc001591 (*ADA2*) under +RSWD stress conditions follows the same trend over time but not necessarily the same level of abundance. For example, under -RSWD stress conditions transcription factor Ta5blLoc028528 (*3C polypeptide 3*) (Figure 6.12 (E)) in D02-105 increased expression from AD5 to AD10 and then was consistently down regulated there after, with the lowest expression level in the profile being reached by FHE. The expression profile of Ta5blLoc028528 (*3C polypeptide 3*) in D02-105 under +RSWD stress conditions (Figure 6.12 (F)) showed the exact same temporal trend as the -RSWD stress assay (Figure 6.12 (E)) for D02-105 but the distinguishing feature being that absolute abundance levels were different. Overall the comparison between Figure 6.12 (E) and Figure 6.12 (F) shows that +RSWD stress conditions cause Ta5blLoc028528 (*3C polypeptide 3*) in D02-105 to be down regulated at all growth stages assayed but the only significant difference (via non overlap of standard error bars) was at the AD15 and FHE stages.

Comparison between Figure 6.12 (K) and (L) show the expression profile of Ta5blLoc001591 (Transcription factor *ADA2*) in D02-105 under +/- RSWD stress

was almost identical in direction and abundance between the two treatments but at FHE the effect of +RSWD stress caused a significant up regulation. Interesting there was also only a relatively small treatment effect on the expression of Ta5blLoc037737 (Transcription factor *WRKY 57*) (Figure 6.12 (I) and (J)) in D02-105 at the AD5 and AD10 growth stages, where by expression was moderately upregulated, but at the AD15 growth stage the effect of +RSWD stress seemed to cause a significant up regulation.

Evaluation of the expression profiles of the *QtAD5-5BL* associated TF's in D08-299 highlighted that transcriptional regulation of transcripts under the control of these TF's is likely modified in response to +RSWD stress to a much larger extent than in D02-105. In the case of Ta5blLoc001591 (*ADA2*) (Figure 6.12 (K) and (L)) expression was significantly upregulated in response to +RSWD stress at all four growth stages assayed. Moreover, the same significant up regulation effect was evident at the AD5, AD10, and FHE growth stage assays for Ta5blLoc037737 (*WRKY 57*) (Figure 6.12 (I) and (J)). However, no differential expression effect was detected between the plus and minus RSWD stress assays for Ta5blLoc028528 (3C polypeptide 3) at any growth stage (Figure 6.12 (E) and (F)).

Comparisons between the expression profiles of each of the *QtAD5-5BL* associated TF's in D02-105 and D08-299, under -RSWD stress conditions, shows that in most growth stage specific pairwise comparisons expression levels were not different, except in the case of Ta5blLoc037737 (*WRKY 57*) (Figure 6.12 (I)) and Ta5blLoc001591 (*ADA2*) (Figure 6.12 (K)) at the FHE stage. In both cases expression levels in D02-105 were significantly higher than D08-299. In contrast the between genotype comparisons for response to the +RSWD stress treatment showed differential expression was detected in Ta5blLoc028528 (3C polypeptide 3) at the AD5, AD10 and FHE growth stages. Genotypic variation for growth stage specific expression was also detected for Ta5blLoc001591 (*ADA2*) under +RSWD stress conditions at growth stages AD5, AD15, and FHE.

Because transcription factors are known to be a key class of gene underpinning drought response (via regulation of transcription) it was expected differential expression in the *QtAD5-5BL* associated TF's may have been detected in the +RSWD stress assays (for both DH lines tested) at the earlier growth stages assayed, including AD5 or AD10 stages (i.e. within 3-6 days of the RSWD stress being imposed, refer to Chapter 5 Figure 5.2 for explanation of growth stages spanning the +RSWD stress treatment). However, in the case of D02-105 this was clearly not the case, since differential expression in response to +RSWD stress was (in most cases) only detected at the FHE stage. These findings provided grounds to plausibly suggest D02-105 may lack the capacity for prompt detection of RSWD stress and/or prompt stress response. The evidence for D08-299 points to the opposite potentially being true in that early detection of RSWD stress at the AD5 stage may be a causal factor in the up regulation of TF's Ta5blLoc037737 (*WRKY 57*) and Ta5blLoc001591 (*ADA2*) in D08-299. In view of the fact cessation of the 10 day RSWD stress treatment coincided with the completion of the AD15 stage, the detection of differential expression in the *QtAD5-5BL* associated TF's in the D02-105 +RSWD stress treatment group at the FHE stage may actually be reflecting a response to re-watering and a subsequent return to favourable tissue hydration levels.

Whether or not differential expression in the *QtAD5-5BL* associated TF's in D08-299 initiate a cascade of beneficial expression responses in other key water deficit stress response transcripts remains unclear. However, this finding taken in context with the results from Chapter 5 indicate these loci may in fact contribute significantly to both the *QtAD5-5BL* escape QTL phenotype and the overarching grain set response to RSWD stress phenotype. The underlying premise to this claim being that the allelic variants of the loci in D08-299 (i.e. that were derived from the *Westonia* parent as opposed to the *Kauz* derived *QtAD5-5BL* alleles in D02-105), are driving the superior RSWD stress response in D08-299.

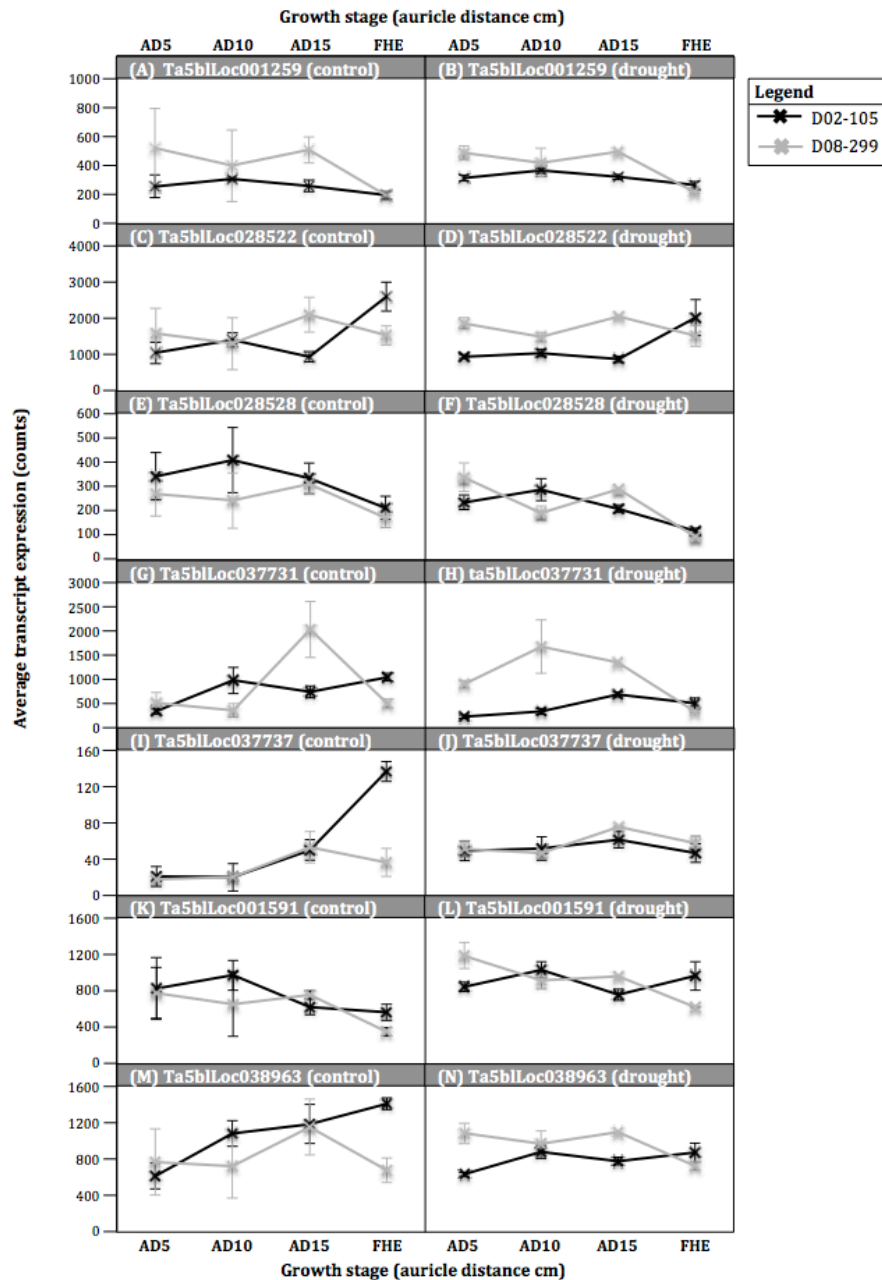


Figure 6.12 Expression profiling from the 2012 RSWD stress RNASeq assays of transcripts that are associated with the drought escape QTL, *QtAD5-5BL*. The Figure shows temporal expression profiles for seven of the twenty two *QtAD5-5BL* associated transcripts. Not all *QtAD5-5BL* associated transcripts were expressed and thus are not shown in the figure. The expression profiles are plotted as pairwise comparisons, showing each transcript expressed in D02-105 and D08-299, under the same water treatment type (+/- RSWD stress). A legend in-set upper right to the figure provides details of the colour scheme used in the line graphs to delineate data by genotype. The x-axis represents the temporal scale (growth stage (AD)) for expression and the y-axis represents the relative expression level (expression averaged across replicates). Pairwise comparisons for each transcript under control conditions (-RSWD stress) and drought conditions (+RSWD stress) are presented on the left and right side of each row of sub-Figures respectively. Correspondence between sub figures and MIPS defined transcript ID are: (A)-(B) Ta5blLoc001259, predicted protein kinase dsk1, (C)-(D) Ta5blLoc028522, predicted stem specific transcript, (E)-(F) Ta5blLoc028528, predicted transcription factor 'like' 3C polypeptide 3, (G)-(H) Ta5blLoc037731, predicted UDP-glucose 6-dehydrogenase, (I)-(J) Ta5blLoc037737, predicted transcription factor 'like' WRKY 57, (K)-(L) Ta5blLoc001591,

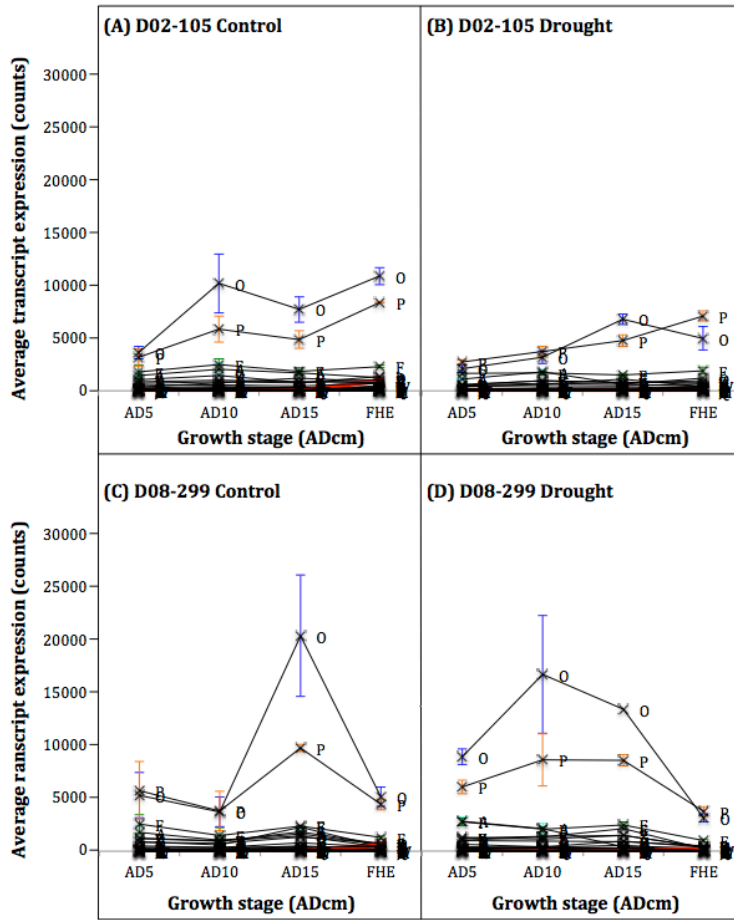
predicted transcription factor 'like' ADA2, (M)-(N) Ta5b1Loc038963, predicted stritosidine synthase protein transcript.

6.4.6 Gene expression analysis of the KEGG starch-sucrose metabolism network

Because reproductive development represent a high energy demand growth stage where developing spike require considerable amount of energy for viable gamete production and associated supportive tissues it was of interest to evaluate the gene expression at a network level for those genes involved in the synthesis of starch and sucrose in these structures. The primary concern being to identify genetic factors with a causal link to changes in carbohydrate synthesis and partitioning in developing anthers in response to RSWD stress. The overarching objective was to detect associations between transcript expression changes (at a temporal and/or abundance level) in this suite of genes, and variation in the phenotype of the grain set response RSWD stress trait.

Figure 6.13 shows the expression profiles from the 2012 RSWD stress RNASeq data set, accumulated on developing spikes for the two DH lines assayed under +/- RSWD stress, for genes involved in the carbohydrate metabolism network (based on the KEGG defined pathway in rice, Pathway ID: osa00052).

Complementing the genotypic specific assay presented in Figure 6.13 is a schematic representation of these enzymes (Figure 6.15, adapted from Ferreira *et al.*, (2012) and Xue *et al.*, (2013)) comprising the individual pathways of carbohydrate metabolism within developing anthers (tapetum and pollen).



Line graph annotation	KEGG starch-sucrose metabolism network ID	KEGG functional categorization	Rice accession	Wheat homolog MIPS gene model ID
—*— A	2.7.7.27	ADP glucose pyrophosphorylas	Os01g0633100	Ta1alLoc015962
—*— B	3.2.1.28	alpha,alpha-trehalase	Os10g0521000	Ta1alLoc016545
—*— C	2.4.1.21	starch synthase	Os02g0744700	Ta1dlLoc003427
—*— D	5.3.1.9	glucose phosphate isomerase	Os03g0776000	Ta1dsLoc010693
—*— E	3.2.1.37	beta-xylosidase	Os04g0640700	Ta2alLoc017176
—*— F	2.4.1.18	amylose isomerase	Os02g0528200	Ta2alLoc028150
—*— G	3.2.1.20	alpha-glucosidase	Os06g0675700	Ta2blLoc029972
—*— H	2.4.1.25	4-alpha-glucanotransferase	Os07g0627000	Ta2bsLoc000443
—*— I	3.2.1.2	beta-amylase	Os03g0141200	Ta2bsLoc010282
—*— J	3.2.1.21	beta-glucosidase	Os01g0508000	Ta2bsLoc020647
—*— K	3.1.1.11	pectinesterase	Os01g0788400	Ta3bLoc002340
—*— L	3.2.1.26	cell wall invertase (<i>IVR1-5BL</i>)	AY578160	Ta5blLoc036480
—*— M	2.7.1.1	hexokinase	Os01g0722700	Ta3bLoc033806
—*— N	2.4.1.1	starch phosphorylase	Os01g0851700	Ta3dlLoc022408
—*— O	1.1.1.22	UDP-glucose 6-dehydrogenase	Os04g48490	Ta5blLoc037731
—*— P	5.4.2.2	phosphoglucomutase	Os03g0712700	Ta4bsLoc021735
—*— Q	2.7.1.4	fructokinase	Os01g0894300	Ta5blLoc009695
—*— R	5.1.3.6	UDP glucuronic epimerase	Os02g0791500	Ta5blLoc019747
—*— S	2.4.1.15	trehalose 6-phosphate synthase	Os01g0730300	Ta5bsLoc007231
—*— T	2.4.1.13	sucrose synthase	Os02g0831500	Ta6alLoc002131
—*— U	3.2.1.1	alpha-amylase	Os02g0765600	Ta6alLoc004799
—*— V	3.2.1.4	cellulase	Os02g0778600	Ta6alLoc016436
—*— W	2.7.7.9	UDPG pyrophosphorylase	Os02g0117700	Ta6asLoc002715
—*— X	2.4.1.14	sucrose-phosphate synthase	Os02g0184400	Ta6bsLoc007096
—*— Y	3.2.1.-	isoamylase	Os08g0520900	Ta7asLoc002713
—*— Z	2.4.1.43	polygalacturonate 4-alpha-galacturonosyltransferase	Os02g0498700	Ta7asLoc014442

Figure 6.13 Temporal expression analysis of the suite of twenty six genes comprising the starch-sucrose metabolism network, identified in the wheat spike transcriptome. Expression profiles are from the 2012 RSWD stress RNASeq data set and probes used to identify the suite of transcripts are from the KEGG defined starch-sucrose metabolism network (*Oryza sativa japonica* (Japanese rice), RefSeq: 4324001, Pathway: osa00052, <http://www.genome.jp/kegg/pathway.html>). Temporal cluster profiles ranging from growth stage AD5 to FHE for suite of twenty six transcripts are presented for each of treatment group: (A) D02-105 under -RSWD stress, (B) D02-105 under +RSWD stress, (C) D08-299 under -RSWD stress, (D) D08-299 under +RSWD stress. The x-axis represents the temporal scale (growth stage (ADcm)) for expression and the y-axis represents the relative expression level (expression averaged across replicates). Temporal expression data are plotted as a connected line and each time point assayed for a given transcript is annotated with a letter corresponding to that specific transcript. A legend presented below the expression profiles provides information for the correspondence between: (i) the letter assigned to the transcript and the probe details defined by KEGG (including enzyme ID number, functional categorization, rice probe accession), and (ii) the MIPS ID for the wheat homolog to the rice probe. The expression profile of the two highest expressed transcripts, UDP-glucose 6-dehydrogenase and phosphoglucomutase, are annotated with standard error bars for each assay time point to allow for discernment of differential expression.

Figure 6.13 highlights that two enzymes, UDP-glucose 6-dehydrogenase and phosphoglucomutase were consistently the highest expressed transcripts (by an order of magnitude) amongst the suite of genes belonging to the carbohydrate metabolism network. UDP-glucose 6-dehydrogenase catalyses the conversion of UDP-glucose into UDP-glucuronate and partitions the assimilate into pathways for hemicellulose production or fatty acid synthesis in the pentose phosphate pathway (PPP). Phosphoglucomutase catalyses the conversion of D-glucose-1-P (the first intermediate from the conversion of UDP-glucose by UDP-glucose pyrophosphorylase in the PPP specific pathway) into D-glucose-6-P. Interestingly our data showed the chromosome 5B UDP-glucose 6-dehydrogenase homologue (Ta5bL037731) associated with *QtAD5-5BL*, was consistently the dominant form of the gene expressed in the developing spike tissues in both lines tested and under both water treatment types.

It is also apparent from Figure 6.13 that the expression profiles of the remainder of the transcripts in the network appear to cluster within a threshold of abundance across growth stages AD5 to FHE. This result was also consistent in both DH lines tested, and under both water treatment types. Due to time limitations in this thesis, relating to resolving data normalization and statistical analysis biases, it was not possible to discern meaningful differences resulting from treatment effects within the expression profiles of this cluster of

transcripts. However, we note the relative abundance of a number of these transcripts (at some time points), particularly some of those enzymes active in the starch synthesis pathway and pentose phosphate (PPP) pathways, were in excess of levels recognised as biologically high (i.e. >1000 counts). Hence, we expect treatment effects causing differential expression within this cluster of transcripts will be detectable upon refinement of the data analysis methods that will be undertaken following on from this thesis.

While Figure 6.13 shows the dominant expression of UDP-glucose 6-dehydrogenase and phosphoglucomutase transcripts amongst the suite of carbohydrate metabolism genes was apparent in both DH lines tested, under both +/- RSWD stress treatment, and at almost all growth stages assayed, there were striking within and between genotypic differences in abundance of these transcripts, in response to growth stage and water treatment type. Within and between genotype differences in the expression of UDP-glucose 6-dehydrogenase were comprehensively discussed in the text accompanying Figure 6.12 (G-H). However, the context of this discussion related to the possible role the gene plays in contributing to the drought escape *QtAD5-5BL* phenotype. In the current results section it is important to interpret UDP-glucose 6-dehydrogenase data in the context of carbohydrate gene network co-expression analysis.

According to the available pathways for partitioning of UDP-glucose into specific end use in developing spikes (outlined in Figure 6.15(A)) the expression profile for UDP-glucose 6-dehydrogenase (in Figure 6.13) highlights that the strongest demand for sucrose in these tissues is driven by the need for production of hemicellulose or redirection into the PPP (via hexokinase) for fatty acid synthesis. Based on extrapolation from UDP-glucose 6-dehydrogenase expression profiles under -RSWD stress conditions (Figure 6.13 (A) (C)) it is plausible to suggest that both DH lines tested may have two points of peak demand for UDP-glucose to be partitioned into pathways for hemicellulose production and/or fatty acid synthesis in the PPP. Our assay interval captured only one peak expression point for UDP-glucose 6-dehydrogenase in D08-299

(under the -RSWD stress conditions) at the AD15 growth stage (Figure 6.13 (C)), which coincides with the uninucleate microspore stage. However, the downward trend for expression between AD5 and AD10 suggests D08-299 may have had an earlier expression peak also, most likely between the AD0 and AD5 stages, the growth stages of which extend across pollen cellular division stages meiosis I and meiosis II. In D02-105 the first peak expression of UDP-glucose 6-dehydrogenase under -RSWD stress conditions occurs at AD10 (uninucleate pollen) and the second we extrapolate either at, or soon after, the FHE (mature pollen) growth stage, which was the final growth stage in the assay interval.

As discussed previously UDP-glucose 6-dehydrogenase expression in D08-299 appears unaffected at an abundance level by +RSWD stress but the allele also had a responsive expression mechanism to reduce the amount of time taken for peak expression to be achieved, shifting it from AD15 to AD10. From Figure 6.13(D) it is not possible to discern if the temporal shift in UDP-glucose 6-dehydrogenase expression in D08-299 under +RSWD stress conditions was accompanied by biologically meaningful changes in the other carbohydrate metabolism transcripts (all of which remained relatively low in comparison). However, what can be deduced is that the preferential partitioning of UDP-glucose, by UDP-glucose 6-dehydrogenase, into pathways for hemicellulose and fatty acid synthesis shows these two sucrose end uses remain the priority targets of carbohydrate metabolism in the developing spikes of D08-299.

Contrastingly there was no evidence of such an early stress detection response via modification of UDP-glucose 6-dehydrogenase expression under +RSWD stress conditions in D02-105 (Figure 6.13(B)). Expression of UDP-glucose 6-dehydrogenase in +RSWD stress conditions in D02-105 was significantly down regulated at all time points (relative to the control -RSWD stress treatment) except AD15. This result indicates levels of UDP-glucose partitioning into both hemicellulose and fatty acid synthesis (via the PPP) would have been significantly below maintenance level requirements for both normal tapetal and pollen development.

From Figure 6.13 it can be seen that for both lines tested, and under both water treatment types, the expression profile of phosphoglucomutase (the second most abundantly expressed carbohydrate synthesis transcript), follows almost the same profile of UDP-glucose 6-dehydrogenase in all four experimental factorial groups. In view of the clear correlation between expression profiles of these two transcripts the majority of conclusions outlined above, regarding the factors driving UDP-glucose 6-dehydrogenase expression and the implications of changes in expression profiles have on carbohydrate metabolism in the developing spikes, also apply to phosphoglucomutase. The main distinction between the expression profiles of phosphoglucomutase and UDP-glucose 6-dehydrogenase in Figure 6.13 (A), (C) and (D) appears to be that peak expression of phosphoglucomutase, while still very high relative to the cluster of low expression carbohydrate metabolism transcripts, is still significantly lower than UDP-glucose 6-dehydrogenase. The only exception to this trend is in the case of the D02-105 +RSWD stress group, where non overlap of standard error bars indicated that the average expression of phosphoglucomutase across most time points was higher than UDP-glucose 6-dehydrogenase.

Highly correlated co-expression of enzymes involved in the synthesis and conversion of UDP-glucose have also been documented in *Arabidopsis* (reviewed by Kleczkowski *et al.*, (2011)). Based on evidence from this thesis we believe a plausible explanation of this correlation between significantly high abundance transcripts is that it points to a delicate balancing of partitioning of assimilates to end uses which are a priority for successful reproductive development in the spike. Those end uses being fatty acid synthesis, in the case phosphoglucomutase mediated partitioning of UDP-glucose into the PPP, and hemicellulose and/or fatty acid synthesis via catalysis from UDP-glucose 6-dehydrogenase. Given both of these enzymes are closely linked to UDP-glucose catalysis, either directly in the UDP-glucose 6-dehydrogenase or once removed in the case of phosphoglucomutase, their dominance in the assay of the carbohydrate pathway analysis further validates that UDP-glucose represents a critical form of assimilate that can readily be partitioned to priority metabolic end uses and or allows flexibly interchanged between pathways.

The data in Figure 6.13 shows that the expression profile of CW-INV *IVR1-5BL* clusters with the numerous transcripts expressed at low to mid range levels. Due to the very low and in most cases absent expression of *IVR1-5BL* for ease of interpretation the expression profile for the enzyme has been highlighted as a red line. Interpreting *IVR1-5BL* expression in context with the suite of carbohydrate metabolism transcripts provides for the first time the opportunity to dissect its relative contribution to the synthesis of carbohydrates in developing anthers and the extent to which this contribution underpins fertile development of pollen in contrasting levels of water availability.

As outlined in the text accompanying *IVR1* expression results in Figure 6.10 the expression of *IVR1-5BL* was, in the case of AD10 and AD15 completely absent in both lines tested, present at very low levels at AD5 in D02-105, and expressed at peak levels at the FHE stage. It is important to reiterate that even the peak FHE *IVR1-5BL* expression levels were very low compared to the peak levels achieved in the majority of transcripts comprising the carbohydrate metabolism network (data not shown). The implication of these findings are covered comprehensively in the following discussion.

6.5 Discussion

The data in this Chapter demonstrate that spike transcriptome assays generated via the RNASeq platform allowed detailed characterization of RSWD stress response phenotypes at the molecular level, in a subset of WxK DH lines. During the course of analyzing the very large data set (1.5 Tbytes) a number of apparently "standard" analytical technologies had to be questioned and this reduced the time available, in this thesis, to complete all the biological associations that were possible.

6.5.1 Global analysis of transcriptome data

Based on the version of the MIPS gene model set used as a mapping reference in this study our data show that the spike transcriptome represents at least 14% of transcripts comprising the global wheat transcriptome. Based on our

calculations the MIPS reference set comprises 963,695 high confidence score predicted gene model loci, but upon further curation to remove redundancy caused by repetitive elements and alternative splicing, and improvement of gene detection algorithms to identify genes missed, this number will change (IWGSC in press). We note also final accurate estimations for the size of both the spike and global wheat transcriptome will only be possible once a full wheat genome sequence has been achieved, as opposed to fragmented IWGSC genome survey sequence on which the MIPS gene model predictions are currently based.

Although there was variation in the mapping efficiency between the two years of data (83% in 2011 and 65% in 2012) the transcriptome coverage was approximately equal (14%), a result which shows: (1) the larger amount of data accumulated in 2012 did not confer greater transcriptome coverage, and (2) the depth of sequencing in both experiments was sufficient to capture a high proportion of the spike transcriptome. While there is some anecdotal evidence from RNASeq benchmarking studies to suggest detection of low expression genes can be improved with increased sequencing depth, sensitivity is not improved for already highly expressed genes (Dillies *et al.*, 2013). Theoretically, increased sequencing depth in the current study may have increased the resolution to detect stronger expression signals for some low expression genes (e.g. *IVR1*'s and low expression TF's). However, in a review of RNASeq studies by Rapaport *et al.*, (2013) it was shown the decision to increase sequencing depth is a trade off, where by the proportion of false positives detected in low expression genes significantly increases with sequencing depth.

In the same review (Rapaport *et al.*, 2013) it was also shown that the frequency of these false positive calls is approximately equal amongst all the commonly used open source software packages for statistical detection of differential expression in RNASeq data. Importantly findings from this study showed that when utilizing EdgeR for statistical analysis, increased sequencing depth actually reduces the sensitivity to detect differential expression in the top half of most abundantly expressed transcripts. These findings validate two important aspects of the approach taken in the current study, namely not to increase the

level of sequencing depth and not to use EdgeR for statistical detection of differential expression beyond a global expression context. In two recent studies (McGettigan 2013; Zheng *et al.*, 2011) was concluded that when bias detection and correction in RNASeq data normalization cannot be resolved then the minimum approach to only correct for variation in library size between samples is an acceptable default approach. These claims are consistent with the approach taken in this thesis, where in the case of detailed gene-to-gene detection of differential expression amongst a gene network, QTL, or family of genes (i.e. *IVR1*'s) analysis was carried out using data that had only been scaled to equilibrate for library size.

An important research output from this Chapter is that RNASeq did not prove to be as sensitive as SQ-PCR for the detection of *IVR1* isoforms. This result highlights that while RNASeq is a valuable first pass transcriptome wide assay to identify a suite of candidate genes, there after a validation step for analysis of single transcript expression profiles is necessary through the use of gene specific primers in qRT-PCR or SQ-PCR assays. A challenge to carrying out meaningful validation is that, as yet, there remains no clear 'gold standard' for estimation of transcript abundance in RNASeq data (Roberts *et al.*, 2011). Hence, drawing meaningful comparisons between results obtained via RNASeq and PCR is difficult.

Two important findings from the current work highlight probable causes for the lack on congruency in the estimation of transcript abundance between RNASeq and PCR: (1) apparent standard RNASeq data normalization techniques involving iterative rounds of normalization per transcript assembly back against every transcript assembly in the global data set lead to major bias in the final estimation of transcript abundance, and (2) expression data for apparent standard house keeping genes (such as *GAPDH*) used as expression benchmarks in wheat PCR assays can be confounded with expression from multiple isoforms, which results in falsely high benchmarks being set for 'standard' expression. Our results for analysis of stable expression genes in the 2012 data set highlighted that *GAPDH* isoforms were not stably expressed and had a range in variance of

expression abundance between samples of between 110-130%. This result is consistent with conclusions drawn in a review by Tenea *et al.*, (2011) which highlighted results from numerous studies showing that many of previously accepted stable expression genes (e.g. actins, beta-globin, 18S rRNA) are in fact insufficiently stable across time, tissue, and treatment type.

Results from this thesis for the defined set of 500 most stable expression genes variance ranged between 21-43%. Based on these findings we would argue defining a core set of stable expression transcripts in RNASeq studies, with the lowest amount of variance in abundance between samples would greatly increase the overall accuracy with which differential expression is detected and quantified in RNASeq studies. We propose these transcripts should be used both as core set for normalizing the global data and as candidates for selection of house keeping genes. These conclusions are also consistent with those drawn in a study by Glusman *et al.*, (2013) in which it was found that normalizing transcript counts by a defined scaling factor (based on a subset of stable expression transcripts), greatly reduced bias and thus reduced that rate at which real differences in expression levels were masked by incorrect adjustment.

GO term analysis of the of differentially expressed transcripts from the 2012 data set yielded results particularly useful for mining the data set for follow up work to this thesis. Characterisation of transcripts in this way provided the opportunity to identify key functional classes of genes showing critical sensitivity to water deficit response. These results will enable whole suites of candidate stress response transcripts, belonging to a defined functional group, to be tracked in parallel for expression pattern matching analysis. Future work will have a strong focus to follow up on the suite of putative transcription factors (TF) found to be differentially expressed in the global data set, the impetus for which being this class of gene is know to have an important role in regulating transcript expression in drought response gene networks (Tran *et al.*, 2013). The results for GO term molecular characterization showed that while only a small number of transcripts were classified as TF's, proportionately the number differentially expressed was the highest (30%) amongst all classes of molecular

functional groups represented. In conjunction with the GO term analysis the full set of probes for all defined TF's in wheat from the Wheat Transcription Factor Database (http://planttfdb_v1.cbi.pku.edu.cn:9010/web/index.php?sp=ta) was also used to screen the MIPS Wheat GSS gene model set. These results (data not shown) highlighted that, 1602 TF's are represented in the global transcriptome (i.e. MIPS Wheat GSS), and 1356 in the spike transcriptome (i.e. 2012 RNASeq data set), and of the latter 185 were found to be differentially expressed in response to the +RSWD stress treatment.

A group of TF's of particular interest to this thesis, based on their putative function in determining the spike fertility trait, have been defined by a number of studies (Deveshwar *et al.*, 2011; Parish and Li 2010). This defined set TF's will provide useful targets in the preliminary follow up research to this thesis to identify trait specific (i.e. spike fertility) TF's that were differentially expressed in the 2012 spike transcriptome RNASeq data. In addition to this suite of previously defined spike fertility related TF's, the three TF's mapped within the interval of the drought escape QTL, *QtAD5-5BL* (in chapter 5) are also of particular interest for planned follow up research to this thesis. This group of TF's included: Ta5blLoc028528 (*3C polypeptide 3*), Ta5blLoc037737 (*WRKY 57*) and Ta5blLoc001591 (*ADA2*). In view of the fact clear genotypic difference were detected for the temporal expression of between the two WxK DH lines tested in this thesis under both +/- RSWD for all three TF's in question it would be of particular interest to test against claims made by Tran *et al.*, (2013) regarding the relative contribution TF's make to the overall drought response. The premise of these claims being that small changes in abundance of TF's, even naturally low expression TF's, and particularly TF's that act upstream in stimulus response pathways, can have large flow on effect downstream in the pathway by way of setting of a cascade of reactions within gene networks. Moreover, based on the fact the that *QtAD5-5BL* was mapped to a linkage block segregating between the D08-299 and D02-105 (Chapter 5) there are strong evidential grounds on which to pursue possible causal effects between the allelic variants of the TF's associated with the QTL and the contrasting phenotypes observed in

the two DH lines studies for the *QtAD5-5BL* phenotype and the grain set in response to water deficit stress phenotype.

6.5.2 Cell wall invertase *IVR1* and its relative contribution to starch accumulation in pollen and the pollen fertility trait per se

The phenomenon of abiotic stress induced pollen sterility, resulting from altered carbohydrate synthesis in developing spikes, has been widely documented in cereals (reviewed by Storme and Geelen (2013) and Dolferus *et al.*, (2013)). Accumulation of carbohydrates (via synthesis from sucrose) in both developing microspores and the tapetum is an essential component of pollen formation (Ruan *et al.*, 2010). In the case of the former to provide the energy to support cell expansion during maturation and as an energy reserve to support pollen tube growth during fertilization. Release of enzymes and carbohydrates from the tapetum (the supportive and nutritive tissue surrounding pollen) upon programmed cell death (PCD) of the tapetum also provides essential components for pollen cell wall formation (particularly exine and intine structures) (Zhang and Yang 2014).

While some studies including Dorian *et al.*, (1996); Koonjul *et al.*, (2005); and Ji *et al.*, (2010) have focussed on starch accumulation and the relative conversion of sucrose into this end fate in developing spikes as the key determinant of pollen fertility, it has been shown via starch mutant studies in *Arabidopsis* that starch accumulation per se is not necessarily a prerequisite for fertility (Niewiadowski *et al.*, 2005). In this study it was shown, both in the presence and absence of starch, that lipid bodies also constitute a significant proportion of the energy reserves used by pollen to support its growth and germination. Moreover, it is known that phloem unloading of sucrose in the anther and conversion into hexoses in the tapetum and pollen thereafter, has a number of fates in addition to starch accumulation, all of which collectively determine pollen fertility (Pacini 1996).

Figure 6.14 (formulated based on findings from a review by Pacini (1996)) shows the distribution of carbohydrate by end fate in mature pollen, including:

(1) energy production in the mitochondria, (2) energy conservation in the form of starch and lipids in amyloplast and vesicles respectively, (3) and conversion into structural components in the cell wall, including hemicellulose cellulose in the cell wall specifically, and tapetal debris resulting from PCD, which is subsequently incorporated into the pollen wall to form part of the exine. In view of this evidence it is defensible to suggest that pollen fertility should be considered a quantitative trait that is actually underpinned by a complex network of genes that partition sucrose to various end fates. Furthermore the degree to which sucrose is partitioned into a specific end fate likely reflects the relative importance of that end fate in conferring structural and metabolic integrity of mature pollen for successful fertilization of the ovum.

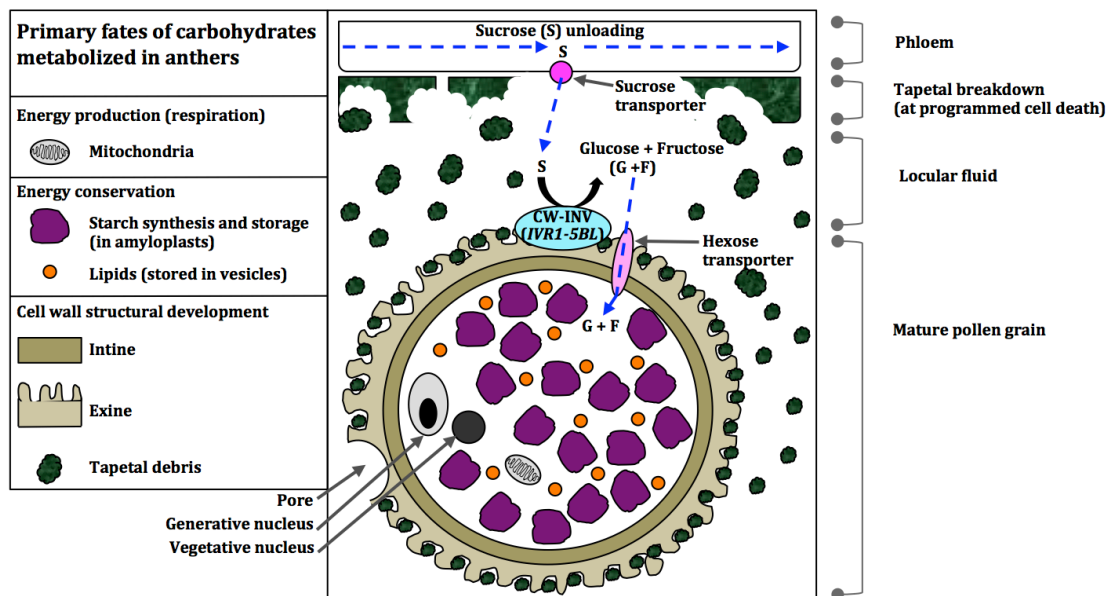


Figure 6.14 A schematic view representing the end fate (by cellular location in mature pollen at the stage of late tapetal breakdown) of carbohydrates synthesized from sucrose (diagram constructed based on findings review by Pancini (1996)). A legend inset to the left provides the visual explanation of the structural components of the pollen and tapetum where carbohydrates have been assimilated for energy construction, energy conservation, or structural development of the pollen cell wall.

An important output from this thesis is that it has highlighted that starch is only one of several factors contributing to pollen fertility. The questions raised by this observation are: (1) to what extent starch accumulation and the other competing end fates for sucrose in the anther contribute to the pollen fertility trait, (2) what are the key enzymes driving these competing sucrose partitioning pathways, and (3) how are these pathways altered by RSWD stress.

In particular two important concepts being challenged via results from this thesis are that *IVR1* is a major effect gene amongst the network of carbohydrate metabolisms genes, and as such is a major determinant of pollen fertility. Secondly, that *IVR1* expression onset is tightly linked to pollen meiosis, a pollen cellular division stage reportedly critically sensitive to water stress (Dolferus *et al.*, 2013; Dorion *et al.*, 1996; Koonjul *et al.*, 2005; Lalonde *et al.*, 1997), and a stage that was benchmarked to the AD5 growth stage in the subset of WxK DH lines tested in this thesis. The premise behind these claims being that hydrolysis of sucrose into glucose and fructose by *IVR1* is the predominant means by which developing pollen receive hexoses for conversion into starch and that this partitioning of assimilates to the pollen commences from the early microspore stage (Dorion *et al.*, 1996). Based on previous findings by Ji *et al.*, (2010) that *IVR1* was down regulated by water deficit stress, and that these expression changes were accompanied by decreased starch accumulation, it was concluded there was a direct and major effect causal relationship between the expression maintenance of *IVR1* and pollen fertility.

Results from this thesis showed that, while meaningful comparisons for *IVR1* expression abundance between SQ-PCR and RNASeq assay were constrained, there was overall concordance between results obtained via the two platforms, in that the highest level of expression was consistently detected at the FHE stage (when pollen has reached maturity), and not the early microspore stage (i.e. AD5). This finding was consistent in both genotypes tested irrespective of water treatment (+/- RSWD stress). *IVR1* expression was detected in D02-105 at the early development stage assayed (AD5) but at low levels compared to the FHE stage, while in D08-299 expression onset did not occur until mid development at the AD15 stage.

Moreover, the discovery of multiple isoforms in the *IVR1* family, and the undertaking of isoform specific assays using RNASeq, highlighted clear genotypic differences in the expression profile of *IVR1* isoforms at a temporal scale and in response to +/- RSWD stress. While *IVR1-5BL* was consistently found (via assays

from both SQ-PCR and RNASeq) to be the dominant isoform expressed in the two DH lines tested, the expression onset under -RSWD stress conditions was detected in D02-105 at AD5, but not until the AD15 growth stage in D08-299. Importantly *IVR1-5BL* expression onset and appeared unaffected in either line by water treatment type. However, *IVR1-5BL* transcript abundance was altered by water treatment type in D02-105 but not D08-299, with the former showing significant down regulation at all growth stages when the transcript was expressed.

In the Ji *et al.*, (2010) study it was also found that genotypic differences existed for the extent to which *IVR1* was down regulated by +RSWD stress in lines that had an expression onset early in spike development at the pollen meiosis stage (equivalent to the AD5 stage benchmarked in this thesis). An important distinction between results in that study and the present one is that our findings highlighted there are also genotypic differences at a temporal level for *IVR1* expression onset. Hence, we conclude our evidence points to the fact molecular cues regulating temporal expression of *IVR1* in the developing spike are not primarily determined by pollen cellular division stage.

Interestingly our cytological and grain set evaluations showed the +RSWD stress treatment did not significantly reduce pollen starch accumulation or grain set per spike in D08-299 but a significant reduction in both variables was detected in D02-105. In view of the fact starch accumulation in pollen has been documented in a number of species as commencing early in pollen development (wheat - Ji *et al.*, (2010); Rice – Ueda *et al.*, (2013); Arabidopsis – Liu *et al.*, (2013); Brachypodium – Sharma *et al.*, (2014)) our results for D08-299, showing that *IVR1* expression onset is not until mid development (AD15), highlight that in this specific DH line the demand for sucrose to be partitioned into starch synthesis must be met at least until AD15, via an alternative pathway not involving *IVR1*. As discussed in more detail in following sections the results for D02-105 also appear consistent with this claim, in that the low level of *IVR1* expression at early development (i.e. AD5) under -RSWD stress conditions would likely have been rate limiting in the cleavage of sucrose, i.e.

commensurate with the high demands at that the early pollen cellular division stages for hexoses required as a precursor for starch synthesis. While the results from D08-299 and D02-105 do validate *IVR1*'s previously claimed functional relatedness to starch accumulation in pollen, collectively they also provide compelling evidence to suggest *IVR1* is not a major effect gene in the developing spike starch synthesis pathway, nor is it a major effect genes in the overall pollen fertility phenotype.

Although analysis of *IVR1* expression in leaves during reproductive development was not a focus in this thesis results obtained from these assay validated an important recent finding by Xue *et al.*, (2013) in their transcriptome study of carbohydrate metabolism in wheat leaves and stems. Based on the fact CW-INV's perform irreversible hydrolysis of sucrose into fructose and glucose the authors concluded that in the leaf the role of the enzyme essentially is to lock carbon into end-use fates for structural development or energy utilization specifically in the leaf, i.e. as opposed to export to other tissues. In the context of the Xue *et al.*, (2013) model for the function of CW-INV's in the leaf, our results showing that the predominant form of the isoform expressed in leaves was *IVR1.1-3B* (but at levels below biological significance), highlight some interesting possibilities for the temporal expression of this isoform.

Although our assay interval did not include growth stages before reproductive development we predict *IVR1.1-3B* expression in the leaf would be at the highest levels during seedling establishment and/or early tillering when sucrose translocation to other tissues is low, i.e. on account of the strong demand for assimilate utilization in the leaf itself for tissue development and energy utilization. We also predict there would likely be no time in later development (especially not reproductive development) when *IVR1.1-3B* expression would be high in the leaf. The rationale for this claim is similar to conclusions drawn by Storme and Gleeson (2013), those being that later in development there is always a very strong demand from other tissues (e.g. the stem and/or developing spikes) which results in the majority of sucrose synthesized in the leaf being translocated.

6.5.3 UDP-glucose 6-dehydrogenase as a key enzyme in spike development

Preliminary results for the targeted approach to characterise and assay *QtAD5-5BL* interval genes provided strong evidence to show one gene in particular, UDP-glucose 6-dehydrogenase, may be having a large effect on the developmentally related time to AD5 (pollen meiosis) trait. Co-expression analysis from the 2012 RSWD stress experiment data for the suite of *QtAD5-5BL* associated transcripts showed that UDP-glucose 6-dehydrogenase expression had to most striking contrast between the two lines tested, both for response to +RSWD stress and temporal profile under the contrasting water treatment types.

These assays highlighted the effects of +RSWD stress on UDP-glucose 6-dehydrogenase on D08-299 caused: (1) significant up regulation (both at the AD5 and AD10 stage), and (2) the peak expression point to be shifted forward to the AD10 stage (i.e. relative to the peak AD15 expression observed in the -RSWD stressed conditions). In D02-105 the +RSWD stress treatment caused a shift in peak expression of UDP-glucose 6-dehydrogenase from AD10 to AD15, and also caused a significant down regulation in all time points assayed. Another important contrast between the two lines was that peak abundance of UDP-glucose 6-dehydrogenase in D08-299 was significantly higher than D02-105 in both water treatment types. However, overall the expression level of UDP-glucose 6-dehydrogenase in both lines under both water treatment types was significantly higher than all other *QtAD5-5BL* associated transcripts.

These result are broadly consistent with the findings from a recent study by Hlavackova *et al.*, (2013) in which the barley proteome was evaluated for expression changes in response to cellular dehydration caused by frost stress. In this study it was found that UDP-glucose 6-dehydrogenase was amongst a small number of proteins proteome wide, that when under the effect of frost induced cellular dehydration, either maintained a significantly high level of expression or were significantly upregulated. Based on the fact frost induced cellular dehydration leads to deleterious changes in cell wall structure Hlavackova *et al.*,

(2013) concluded UDP-glucose 6-dehydrogenase enables an important adaptive response to the stress, through its role in biosynthesis of cell wall components (i.e. hemicellulose). The rationale being, maintenance of UDP-glucose 6-dehydrogenase expression confers the ability to maintain cell wall integrity. In view of the fact water deficit stress also leads to cellular dehydration our results for the expression of UDP-glucose 6-dehydrogenase in D08-299 may provide evidence to show this particular DH line also shares a similar adaptive stress response to maintain cell wall integrity.

In particular the D08-299 result showing a shift in UDP-glucose 6-dehydrogenase expression to an earlier onset under water deficit condition may reflect that this DH has a greater capacity for early stress detection. Evidence to support this claim is illustrated by the fact the shift in expression of UDP-glucose 6-dehydrogenase from AD15 to AD10 corresponded approximately to a three day earlier expression onset (Refer to Chapter 5 Figure 5.2 for temporal staging of AD growth stages). Under stress conditions correspondence between the number of days under stress and auricle distance (AD) growth stage was that AD10 and AD15 occurred at six and nine days respectively during the ten day RSWD stress interval. By day six of the stress treatment interval soil volumetric water content had dropped to 2.5% (data not shown), which is in the range of severely limited plant available water. Hence, we conclude D08-299 underwent very early stress detection, likely within one to two days of the stress interval commencing, and that increased expression of UDP-glucose 6-dehydrogenase was part of this DH lines adaptive response. Based on the findings from the Hlavackova *et al.*, (2013) study the practical implications of this adaptive response would be that the major form of assimilate present in the tapetal cells, UDP-glucose, would be cleaved by UDP-glucose 6-dehydrogenase and partitioned into pathways for hemicellulose production to prevent major structural damage to cell wall integrity.

In view of the fact *QtAD5-5BL* (reported on in Chapter 5) was mapped to a linkage block which was segregating between the two DH lines we concluded the Kauz 'like' UDP-glucose 6-dehydrogenase allele in D02-105 appears to be more

water deficit stress susceptible to down regulation. Conversely, the transcript for the *Westonia* 'like' UDP-glucose 6-dehydrogenase allele in D08-299 is not susceptible to down regulation, and in addition has an adaptive expression response mechanism to reduce the amount of time taken for to reach peak expression. As presented in Chapter 5 calculations for the direction and magnitude of the additive effect for *QtAD5-5BL* in D08-299 (which possessed the *Westonia* QTL genotype) showed that time to AD5 (pollen meiosis) was shifted forward by ten days. The reverse was true for D02-105, which had the *Kauz* QTL genotype. If taken in context with the QTL genotype analysis the *QtAD5-5BL* assays provide strong grounds to suggest the allelic variants of UDP-glucose 6-dehydrogenase may have a causal link to the genetic variation in the time to AD5 (pollen meiosis) trait.

6.5.4 Key drivers of carbohydrate metabolism in developing anthers in relation to spike fertility under water deficit conditions

Although a wheat spike is a single structure it consists of different tissue types (e.g. tapetum, pollen, ovum, and vascular tissue), all which have their own requirements for carbon (Ferreira and Sonnewald 2012), and as such a spike should not be viewed as a single carbon sink. Hence, because in this study transcriptional profiling was undertaken on the whole spike, and not developing anthers or even pollen themselves, it was necessary when interpreting the expression data for the network of carbohydrate metabolism genes to bear in mind that the different tissues comprising the spike samples likely had different assimilate requirements and at different times. Furthermore different tissues have different roles and this affects how assimilates are utilized and to what degree (Ferreira and Sonnewald 2012). This point is particularly important with regards to how we interpreted the relative expression levels of enzymes from our data, in that abundance at the 'whole spike' level was likely cumulative and being driven by expression from different tissues.

Despite the limitation in this study associated with having assayed the whole spike transcriptome (i.e. as opposed to the anther transcriptome) the results generated to our knowledge are the first to report on the effect water deficit

stress has on the whole suite of genes involved in carbohydrate metabolism from early spike development through to head emergence from the sheath. These claims are based on the fact the majority of studies in wheat in this research space to date have focussed on conducting +/- RSWD stress assays on anthers and only a handful of candidate genes (Ji *et al.*, 2010; Koonjul *et al.*, 2005). Two recent studies employed a similar tissue specific network wide assay of carbohydrate metabolism as was carried out in this thesis (potato tuber tissue- Ferreira and Sonnewald (2012); wheat stems and leaves - Xue *et al.*, (2013)). In these studies results highlighted the power of undertaking this type of large scale co-expression analysis approach. In conjunction with the rice defined KEGG pathway for carbohydrate metabolism, findings from Ferreira and Sonnewald (2012) and Xue *et al.*, (2013) provided a template in this thesis to define a tissue specific carbohydrate metabolism pathway in developing spikes (Figure 6.15). To our knowledge this approach has not been previously reported. Importantly this template provides the means by which to: (1) contextualise two priority transcript candidates to this thesis (*IVR1* and UDP-glucose 6-dehydrogenase) with co-expression of the whole network of carbohydrate metabolism transcripts, and (2) define a roadmap for how assimilates are transported and converted from the phloem and into tapetal tissue and pollen.

Figure 6.15 (A) shows phloem unloading of sucrose (the only form of carbon which can be transported around the plant) in the anthers during early reproductive development when the tapetal cells remain intact (Ruan *et al.*, 2010), via two mechanisms: (1) symplastic channels (across plasmodesmata) directly into tapetal cells, and (2) apoplastically into the extracellular space between the tapetum and the pollen sac, via sucrose transporters. Amongst the literature there are conflicting reports regarding the specific location of sucrose hydrolysis by cell wall bound cell wall invertase (CW-INV), but we propose it occurs in the locular space, via cleavage by the dominant CW-INV isoform *IVR1-5BL*. Sucrose is cleaved by *IVR1-5BL* into hexoses (glucose and fructose), which are then exported into the pollen sac and taken up by developing microspores for conversion (and storage) into starch (Figure 6.15(B) and (C)) (Ueda *et al.*,

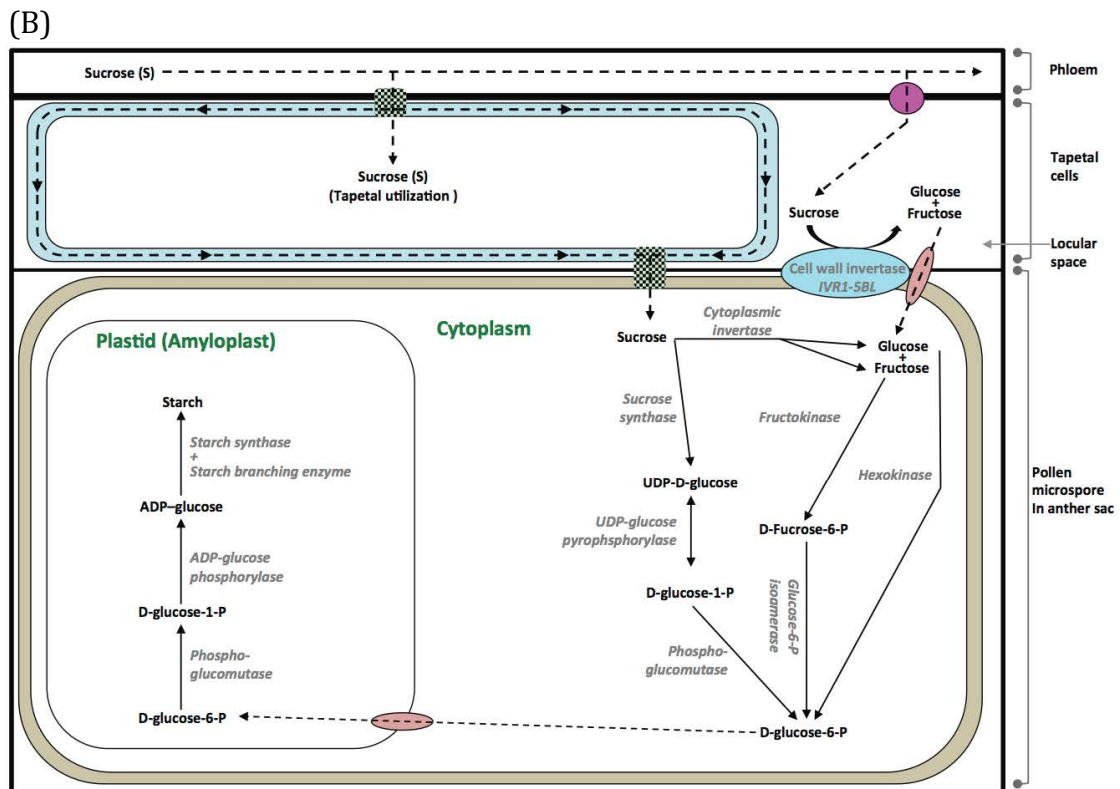
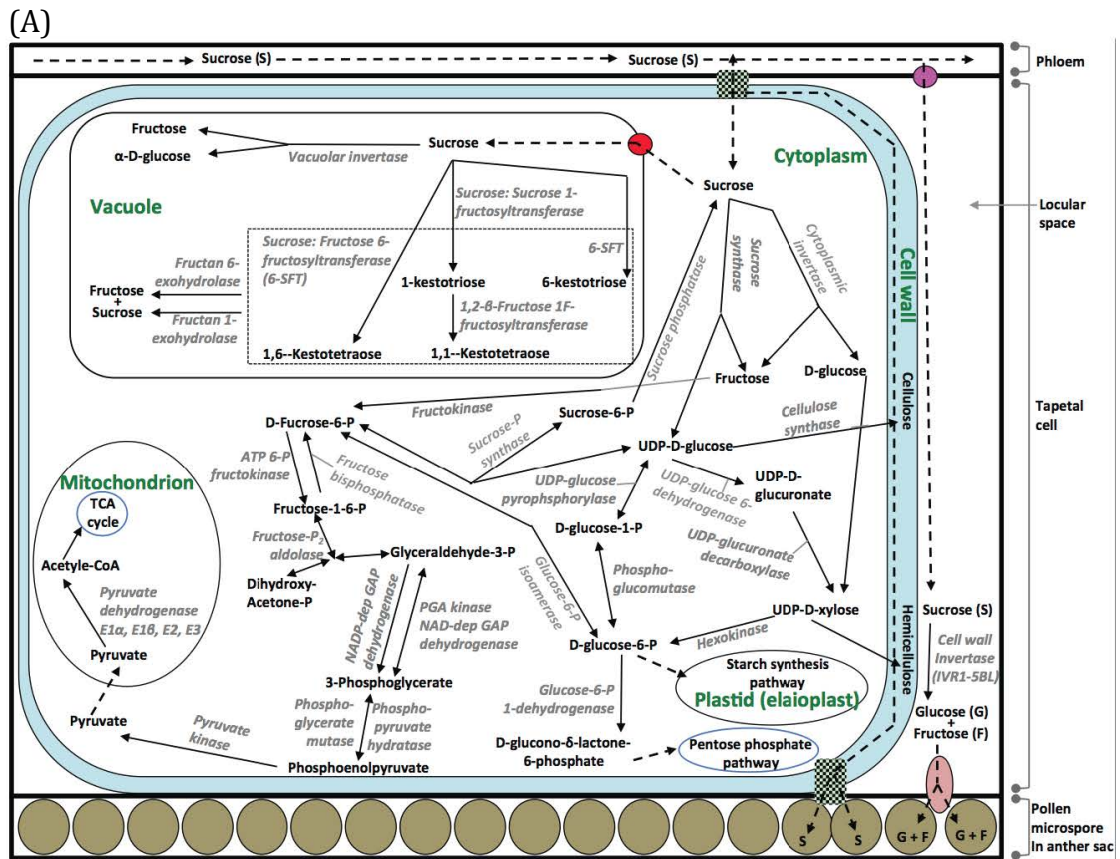
2013), or direct utilization for energy production and development of cellular structural components (i.e. particularly cellulose and hemicellulose comprising the cell wall) (Pacini 1996).

In the tapetum, sucrose may be hydrolysed in two sub cellular locations: (1) the vacuole, by either vacuolar invertase (V-INV) or Sucrose:Sucrose1-fructosyltransferase, and (2) the cytoplasm, by either sucrose synthase, which converts sucrose into fructose and UDP-glucose, or cytoplasmic invertase (C-INV), which converts sucrose into glucose or fructose (Ruan *et al.*, 2010). Based on these findings we proposed that UDP-glucose, and the enzymes that catalyse its conversion, represents a critical point of flexible interchange in the carbohydrate synthesis network for partitioning of photo assimilates for various end uses, including: respiration (TCA cycle), carbohydrate storage (starch synthesis), fatty acid synthesis (pentose phosphate pathway (PPP)), and structural development (cellulose and hemicellulose) (Ferreira and Sonnewald 2012). Hence, we propose monitoring expression of enzymes catalysing UDP-glucose reactions provides insight into the dominant pathways and fate(s) of sucrose once unloaded in the anther.

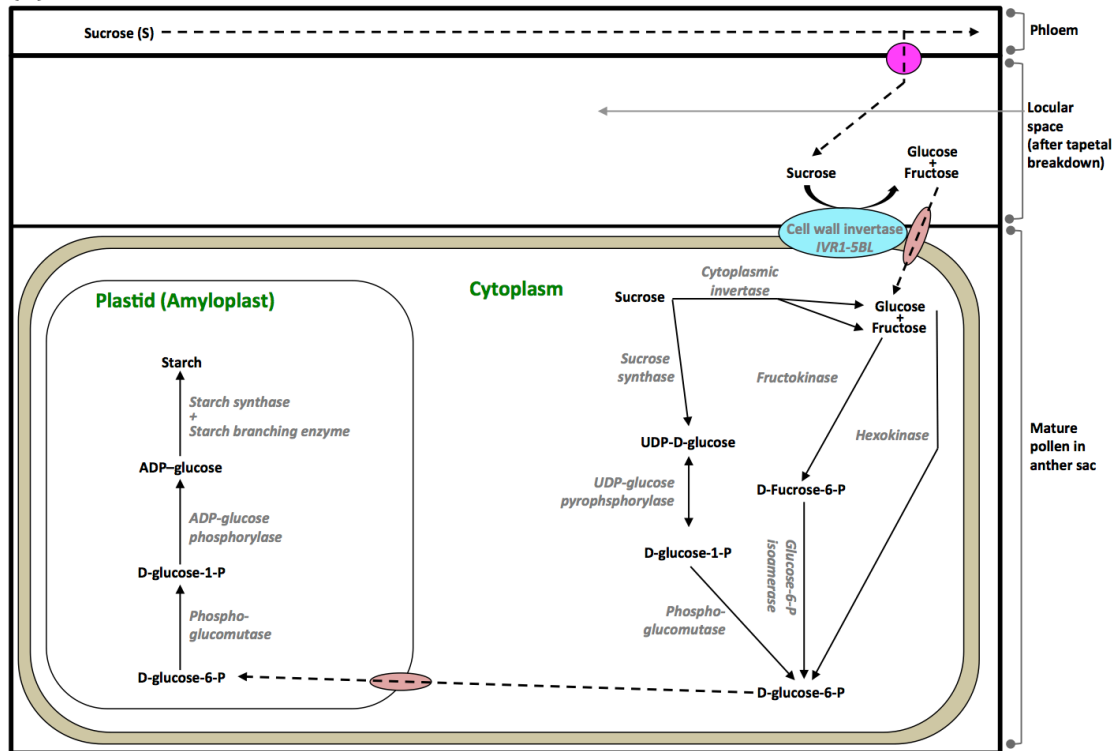
Figure 6.15 (A) shows that once symplastically transported into the cytoplasm of tapetal cells sucrose may be irreversibly hydrolysed by cytoplasmic and vacuolar invertases (C-INV and V-INV) (Roitsch and Gonzalez 2004). Due to there being no characterised gene model for C-INV's or V-INV's in the MIPS gene model reference our assay did not include these transcripts. However, we conclude theoretically these transcripts would likely have been at least moderately expressed during the AD5 to FHE assay interval. We postulate that instances when C-INV's or V-INV's could be highly expressed in the tapetum or pollen may be when there is a high demand for UDP-glucose to be partitioned into the mitochondria for respiration or into the pentose phosphate pathway (PPP). Furthermore, since UDP-glucose is also a precursor for hemicellulose production (Ferreira and Sonnewald 2012) theoretically biomass production in these tissues could be compromised if UDP-glucose was preferentially directed into the PPP and mitochondria (TCA for respiration) if an alternative precursor for the

hemicellulose synthesis pathway was not available. However, since glucose (synthesized by C-INV) represents such an alternative precursor (i.e. alternative to UDP-glucose) for hemicellulose synthesis both pathways can be active simultaneously.

Figure 6.15 (B) and 6.15(C) represent the two possible modes of sucrose transport in the pollen sac for uptake by developing microspores/pollen and conversion into starch. In Figure 6.15(B) the symplastic transport of sucrose to the developing pollen is shown (across plasmodesmata in the tapetum cell wall when it remains intact), and highlights how sucrose synthase or C-INV hydrolyse pollen sucrose respectively, into UDP-glucose or glucose and fructose (Ruan *et al.*, 2010). In Figure 6.15(C) the apoplastic sucrose pathway is represented and shows how after decomposition of the tapetum (upon completion of tapetal PCD), the apoplastic pathway, via CW-INV (*IVR1-5BL*) remains available but the symplastic pathway has disappeared (Parish and Li 2010). An important distinction between these two pathways is that hydrolysis of sucrose (intended for conversion into starch) by CW-INV (*IVR1-5BL*) is more energetically expensive and less efficient in the conversion of sucrose into hexoses than via the sucrose synthase route, with the former requiring an additional ATP unit (Ferreira and Sonnewald 2012).



(C)









Legend	
	Sucrose transporter (apoplastic pathway)
	Sucrose transporter (symplastic pathway)
	Plasmodesmata (symplastic pathway)
	Hexose transporter (apoplastic pathway)
	Assimilate movement
	Enzymatic reaction

Figure 6.15 Schematic representation of the major carbohydrate metabolism related enzymes and respective pathways in reproductive tissues of the developing spike (adapted from Ferreira *et al.*, (2012) and Xue *et al.*, (2013)), including pathways for: (A) phloem unloading of sucrose symplastically (directly into the tapetum) and appoplastically into the locular space. Sucrose conversion pathways represented in the tapetum (specific to cellular location) include the cytoplasm, mitochondria, vacuole, and plastid, (B) symplastic transport of sucrose from the tapetal cell wall (pre tapetal programmed cell death) into the developing microspore and appoplastic transport of sucrose (mediated by *IVR1-5BL*) from the locular space into the developing microspore, and (C) symplastic transport of sucrose (mediated by *IVR1-5BL*) from the locular space into the developing microspore/pollen after tapetal disintegration (i.e. post tapetal programmed cell death). NB: end fate of pathway represented in: (A) includes energy production (respiration), fatty acid synthesis, starch synthesis, and hemicellulose and cellulose production, and in (B) only the starch metabolism pathway extending across the cytoplasm into the plastid (amyloplast) is represented. A legend inset shows colour coded shapes and symbols denoting: (i) the method of assimilate transport (symplastic or appoplastic), (ii) the form of assimilate transported (sucrose or hexoses (glucose and fructose)), (iii) assimilate movement/transport, and (iv) enzymatic reactions synthesis and catalysis. The name of compounds synthesized and/or catalyzed in a reaction are written in black text and the name of enzymes synthesizing/catalyzing reactions are written in grey text.

Construction of the roadmap for carbohydrate metabolism in developing spikes (Figure 6.15) sheds new light on the determinants of carbohydrate metabolism in these tissues. Ferreira and Sonnewald (2012) concluded that, based on the fact sucrose synthase performs reversible hydrolysis of sucrose (as opposed to irreversible hydrolysis by CW-INV) and that the sucrose synthase mediated pathway to starch synthesis is less energetically expensive, sucrose synthase is the preferred sucrose hydrolysis pathway. If proven correct the theory would validate our earlier conclusion that in D08-299, since *IVR1-5BL* expression is not until AD15-FHE, the requirement for hexoses in pollen (generated from the hydrolysis of sucrose) for partitioning into the starch synthesis pathway must be met by sucrose synthase.

Our assays showed *IVR1-5BL* expression in D02-105 during early development (AD5 stage) was very low and there after was not found to be expressed again at significantly higher levels until the FHE stage. Hence we conclude sucrose transport into developing pollen in D02-105 prior to the FHE stage was likely also predominantly symplastic (via sucrose synthase), as represented in Figure 6.15(B). The wide body of research evaluating the level of starch accumulated in pollen upon maturity points to it being a cumulative total dependant of continual accumulation, i.e. starting from early development through to maturation (Liu *et*

al., 2013; Sharma *et al.*, 2014; Ueda *et al.*, 2013). Hence, the practical implication of our results showing that the *IVR1* hydrolysis of sucrose does not partition hexoses into the starch synthesis pathway at high levels until the FHE stage when tapetal breakdown has occurred (as represented in Figure 6.15(C)) likely reflects *IVR1* relatively small contribution to starch accumulation across the duration of the accumulation interval.

Importantly it has been shown that carbohydrate metabolism in the tapetum for the production of energy and cellular structural components also represent an important aspect of pollen fertility (Parish *et al.*, 2012). The tapetum not only acts as a conduit for symplastic flow of sucrose into the pollen (Figure 16.5 B), upon its breakdown during programmed cell death the tapetal debris also incorporates with the pollen to form essential components of the pollen cell wall (specifically the exine) (Figure 6.14). Hence, carbohydrate metabolism at levels sufficient to support optimal tapetal development are also very important to the overall pollen fertility trait (Storme and Geelen 2013). A number of studies (Reviewed in Parish *et al.*, (2012)) have shown abiotic stress induced disrupted timing of tapetal PCD significantly contributes to impaired development of pollen. In the study by Ji *et al.*, (2010) the onset of tapetal PCD (and subsequent apoplastic isolation of the pollen) in wheat was benchmarked to occur mid-late way through reproductive stage development (i.e. equivalent to the AD15 in the germplasm tested in this thesis). Contrastingly, we believe our results provide the first evidence to suggest onset and duration to completion of tapetal PCD is a trait showing genotypic variation.

Extrapolating on finding in this thesis and a number of other studies we believe it's plausible to suggest the expression of *IVR1-5BL* may be sound indicator for the timing of tapetal PCD onset. In view of the fact *IVR1-5BL* assays highlighted clear genotypic differences for both the onset of expression and differential expression in response to +RSWD stress we propose the use of *IVR1-5BL* as a signal for tapetal PCD. This represents a novel approach to track both natural variation (under non stress conditions) and stress induced changes in PCD, in a germplasm specific context. The evidence from other studies underpinning this

model includes: (1) starch starts to accumulate in pollen in the early developmental stages (i.e. pollen meiosis/AD5) and continues until just prior to pollen dehiscence (i.e. FHE) (Dolferus *et al.*, 2013), (2) prior to tapetal PCD both symplastic and apoplastic pathways for sucrose transport to the pollen for starch synthesis are theoretically available ((Ruan *et al.*, 2010), and (3) when both the symplastic (sucrose synthase mediated) and apoplastic (*IVR1* mediated) sucrose conversion into starch pathway are available the former is likely the preferred mode based on its increased efficiency of conversion and its less energetically expensive (Ferreira and Sonnewald 2012). Intuitively we believe these results collectively point to their being no metabolically justifiable point at which *IVR1* would be expressed for the partitioning of sucrose into starch synthesis in the pollen when the tapetal tissue remains intact and the symplastic sucrose synthase pathway is available.

Given that CW-INV (*IVR1*'s) are mediators specifically of the apoplastic form of sucrose transport, which is the only pathway available for sucrose transport to pollen after tapetal PCD we believe its plausible to suggest expression onset and relative abundance of *IVR1* expression is a measure of: (1) tapetal PCD onset, (2) the extent to which tapetal breakdown has advanced, and (3) timing of when tapetal PCD has completed and the tapetal debris is fully incorporated with the pollen exine.

Some studies report it is the disrupted timing of tapetal PCD (induced by abiotic stress) that contributes to impaired development of pollen (Parish and Li 2010). Somewhat contrasting to the Parish *et al.*, (2010) claims, our results provide evidence to suggest the timing of tapetal PCD in some germplasm is not altered by water deficit stress. Our assays showed genotypic specific temporal expression of *IVR1-5BL* was not altered in either of the WxK DH lines tested under normal or water deficit treatments. Using *IVR1* expression as an indicator for PCD onset and advancement to completion, our results show that in D02-105 PCD may commence at pollen meiosis (AD5) and extend through, albeit very slowly, to completion by pollen maturation at the FHE stage. The *IVR1* assays in D08-299 showed the onset and duration of PCD is likely much shorter, with

commencing at the AD15 stage and completion by the FHE stage. The results suggest that the onset of PCD is not affected by water deficit and the segregating allelic variant(s) causing the different onset of PCD between D08-299 and D02-105 are not under the control of drought response pathways.

Subject to further testing we believe this naturally occurring variation in the onset of PCD death may be one of the major underlying causes of genotypic variation between D08-299 and D02-105 in the RSWD stress pollen fertility and grain set response. If proven correct via cytological and metabolomic evaluations our model points to the timing and completion of PCD being a process less affected by water deficit and in turn less of a determinant of pollen fertility. Moreover we would argue it may in fact be that difference in the duration of connectivity between the tapetum and the pollen that alters the dynamics of how sucrose is partitioned into specific end fates and in the case of starch synthesis determines which starch synthesis pathway predominates. Taken in context with the assay results for the highly abundant UDP glucose-6-dehydrogenase and phosphoglucomutase enzymes these findings point to the major point of disruption potentially being the relative abundance of enzymes that synthesize or convert UDP-glucose. Given UDP-glucose represents a major point of flexible interconversion of substrate partitioning into various end fates in the carbohydrate metabolism network we believe its relative availability in both the tapetum and pollen is a probably large effect causal factor determining the pollen fertility trait.

Chapter 7: General discussion and conclusions

At the outset of this thesis the overall experimental approach was summarized in the thesis project pipeline in Chapter 1 (reproduced as Figure 7.1), with special reference to exploring the possibilities of discovering new aspects of drought (water deficit stress) tolerance in wheat. In this concluding chapter the achievements reported in Chapters 3, 4, 5 and 6 are summarized using the headings in Figure 7.1.

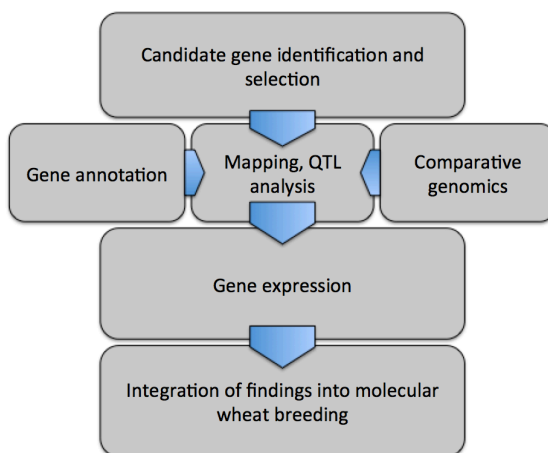


Figure 7.1 Thesis project pipeline

7.1 Candidate gene identification and selection

The literature, at the time of planning this thesis, was providing strong indications that cell wall invertase 1 (*IVR1*) was a key candidate gene in determining the tolerance of wheat to RSWD stress. During the course of identifying this gene in our reference wheat germplasm (225 doubled haploid lines from a Westonia x Kauz double haploid cross) it was evident that at least five isoforms of *IVR1* existed which thus complicated any detailed interpretations. These *IVR1* isoforms proved not to be polymorphic between Westonia and Kauz, nor in a wide range of other germplasm, and thus standard procedures for defining a genetic location could not be utilized for this purpose. The early release of the extensive IWGSC Wheat Genome Survey Sequences (IWGSC Wheat GSS) for each of the forty two chromosome arms provided the first confirmation of the existence of *IVR1* isoforms (Webster *et al.*, 2012).

The identification of the isoforms is a major research output from this thesis, on account of it removing significant ambiguity in interpreting the data related to this family of genes. The work in this thesis, including the analysis of the IWGSC Wheat GSS, ruled out that *IVR1* genes are located on chromosome 1A, as previously reported by Francki *et al.*, (2006). The *IVR1* isoforms that were definitively identified in the current work included *IVR1.1-3B*, *IVR1.2-3B*, *IVR1-4A*, *IVR1-5B* and *IVR1-5D* (Webster *et al.*, 2012), and their respective annotated sequences were submitted to GenBank.

In addition to the family of *IVR1* genes the extensive transcriptome analyses carried out in this thesis discovered some new candidate genes and these are summarized below in the gene expression section, in the context of the carbohydrate metabolism in developing spikes.

7.2 Gene annotation, Mapping/QTL analysis, and comparative genomics

The gene annotation carried out in this thesis focused mainly on the URGI_TriAnnot system (<https://urgi.versailles.inra.fr/triannot/>) as well as utilizing gene models from the Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu>). The main gene family analysed in detail was the *IVR1* family, the manual annotation of which provided robust gene models to enable purification and characterisation of the genes. Our analyses were confirmed by Associate Professor Wim Van den Ende (Katholic University, Leuven, Belgium), who is an expert in the field of structural and functional characterisation of sugar metabolizing enzymes. The Van den Ende group is currently utilizing the *IVR1* annotations (from this thesis) to test the model that some isoforms in the *IVR1* family are metabolically inactive but have gained functionality for regulating the expression of the metabolically active *IVR1*(s), by preventing the latter binding with the *IVR1* inhibitor (Le Roy *et al.*, 2013). Using the *IVR1* gene family as a test case, on going collaborations in this regard are providing outputs for wheat genomics groups to test aspects of gene evolution in wheat, epigenetic regulation, and genome re-arrangement.

The high density map for the Westonia x Kauz cross in this thesis (based on the 90K SNP analysis) is a major thesis output and was built on the Zhang *et al.*, (2013) 9K SNP map as an anchor. This work utilized the ICIM mapping software from Wang *et al.*, (2007) and engaged a number of colleagues in a collaborative process to resolve the details of the map. The phenotype measurements on the 225 WxK DH lines were carried out in triplicate and enabled the identification of a number of significant QTL, especially on chromosome 5A, 5B and 5D, and defined a target region for analysis of the underpinning genome. The traits mapping to the drought escape QTL (*QtAD5*), on chromosome arms 5A, 5B, and 5D, all related to spike development. It should be noted however that further testing is required to establish if these QTL are adaptive (i.e. expressed under certain environments/conditions only) or constitutive (expressed across all environments/conditions). Plans to conduct further studies beyond this thesis combining our suite of markers to screen a wide range of germplasm across multiple environments will facilitate field observations to identify if the QTL are adaptive or constitutive.

Transcripts associated with the respective QTL interval were of interest because allelic variation in this region(s) would be potentially relevant to responses to RSWD stress. The chromosome 5B QTL, *QtAD5-5B*, was studied in most detail. A major achievement from the QTL analysis was the identification of two WxK DH lines, D02-105 and D08-299, which differed in response to RSWD stress, a determination that was based on grain set per spike. The significant grain set reduction in D02-105 and non-significant reduction in D08-299 correlated with D02-105 and D08-299 having Kauz and Westonia QTL genotypes respectively. These two lines formed the basis of the large transcriptome study carried out in Chapter 6.

The comparative analysis carried out in this thesis focused on the *IVR1* gene on chromosome 5B and, based on microlinearity, could be assigned to a section of rice chromosome 4 which mapped near to the QTL for spike development (*QtAD5-5B*). A number of drought response and reproductive stage development

QTL have been defined in rice, one of which (*QTLgys4.1*) is located in the syntenic mapping position proximal to the *CIN3/IVR1* gene on the above rice chromosome 4 section, and this means that rice provides a valuable model for defining genes in wheat associated with response to RSWD stress. The families of cell wall invertase and UDP-glucose-6-dehydrogenase genes studied in this thesis have also been extensively studied in rice.

Screening of the URGI wheat QTL repository showed that no QTL have been defined on wheat chromosome 5B for traits relating to plant height and flowering time. Screening of the same repository also showed that to date no drought tolerance QTL have been defined on wheat chromosome 5B and thus the QTL identified in this thesis are new.

7.3 Gene expression

An additional major output from the thesis was that our evidence showed line-specific transcriptome-level responses to RSWD stress. Although a complete analysis of the RNASeq data was not possible, a significant finding was that amongst the set of annotated genes examined in detail, two genes (UDP-glucose-6-dehydrogenase and phosphoglucosyltransferase) that distinguished between the case study lines (D02-105 and D08-299), were predicted to be involved in the network of genes supplying sugars to tissues in the developing spike. Consistent with this finding, the *IVR1* expression analysis using SQ-PCR also distinguished between D02-105 and D08-299. These results were in agreement with the overall hypothesis proposed by Ji *et al.*, (2010), namely that supply of sugars to the early developing spike is particularly important in determining varietal response to RSWD stress. However, in contrast to the conclusions drawn by Ji *et al.*, (2010) our assays showed *IVR1* genes are not the major effect gene determining the proportional majority of sugar supplied to developing anthers in the spike.

While our results showed that the *IVR1* genes are not the main effect genes determining spike fertility these results did confirm the importance of *IVR1*'s as components of the suite of genes that define fundamental differences between

genotypes in their ability to respond to RSWD stress. The RNASeq analyses in Chapter 6 of this thesis indicated that there was also clear genotypic differences in the expression of transcripts associated with either the drought escape QTL, *QtAD5-5BL*, or the carbohydrate metabolism network. These findings highlighted a number of potential targets that may be contributing to the RSWD stress response trait. Particularly strong candidates were three predicted transcription factors (potentially involved in transcriptional regulation of drought response genes), a predicted stritosidine synthase gene (functionally related to tapetal break down), and UDP-glucose-6-dehydrogenase and phosphoglucomutase. UDP-glucose-6-dehydrogenase and phosphoglucomutase were the most abundantly expressed enzymes amongst the carbohydrate metabolism network.

Interestingly the effect of RSWD stress on the expression of UDP-glucose 6-dehydrogenase and phosphoglucomutase in D08-299 was to shift peak expression from late to early-mid development but no equivalent temporal shift was observed in D02-105, where the peak expression remained at late development. In addition to genotypic difference for temporal expression, RSWD stress caused significant down regulation in the expression level of these two genes in D02-105 (relative to the control -RSWD stress treatment group), but had no significant effect on the expression level of the genes in D08-299.

An unexpected but valuable discovery in this project was the identification of inherent flaws in apparent standard protocols for identifying and utilizing house keeping genes in wheat gene expression studies. In terms of analysing data outputs from RNASeq and SQ-PCR studies it is widely recognized that reference genes are required in order to normalize the levels of expression in a given study. Traditionally so-called housekeeping genes such as *GAPDH* have been used for this purpose, but this thesis has shown the existence of gene isoforms in wheat complicates analyses when using arbitrarily chosen reference genes. Particularly noteworthy is the contribution from the RNASeq analyses reported in Chapter 6, highlighting that an appropriate set of stable expression transcripts

can be identified and characterized in each individual experiment carried out, prior to normalization of data sets at the global scale.

7.4 Summary of thesis findings and integration into molecular breeding

The work in this thesis has indicated it is feasible to utilize the outputs in an applied context to define a suite of genes that are associated with various aspects of RSWD stress response in developing wheat spikes. Collectively our results show pollen fertility is a quantitative trait that is underpinned by a complex network of genes. While some of the important genes defined in this work may be specifically involved in the network driving pollen and tapetal carbohydrate metabolism, while others such as *VrnB1* are more likely involved in defining broader spike developmental, and may be involved in water deficit stress response and/or stress detection at the whole plant level.

Although this work has focussed on a developmental stage specific water deficit stress response, it is evident the thesis contributes to the wider body of information becoming available for wheat, much of which now indicates it should be feasible to define a haplotype for wheat using the allelic variation in genes associated with water deficit stress tolerance in different environments. The approach for using a suite of markers to define a complex phenotype was studied by McNeil *et al.*, (2009) for late maturing alfa amylase (LMA) and the authors argued for the use of a range of markers to provide a risk factor for the occurrence of the trait in a particular environment, rather than an absolute plus/minus score. This approach is equivalent to estimated breeding values developed for the breeding of domesticated animals. In view of the fact that RSWD stress tolerance is also a quantitative trait, this risk factor approach could be used to more effectively screen for the trait in early selection or backcrossing for sensitivity in particular environments.

In addition to utilizing the outputs of this thesis in an applied context for marker assisted selection of RSWD stress tolerance the outputs also provide valuable material for continuing basic research of the trait, i.e. further upstream in the

pre-breeding pipeline. In particular, Chapter 6 outlined a model to test the concept that variation observed between D02-105 and D08-299 for the tolerance to RSWD stress is due to possible genetic variability in the onset and duration of tapetal programmed cell death (PCD), and that these response-variables could be tracked using *IVR1* expression as a signal.

Our results suggest the duration of connectivity between the tapetum and the pollen may alter the dynamics of how sucrose is partitioned into specific end fates, and in the case of starch synthesis determine which starch synthesis pathway predominates. The underlying premise of this model is that there are inherent differences in both the efficiency of sucrose conversion to carbohydrate(s) and total energy cost of sucrose conversion to carbohydrate(s), via the apoplastic pathway versus the symplastic pathway (i.e. the pathways that predominate before and after tapetal PCD respectively).

Data in this thesis, and plant PCD studies more broadly (reviewed by van Doorn and Wouter (2011)), provide evidence that suggests manipulating the timing of tapetal PCD might be a particularly effective developmental process to target for the improvement of RSWD stress tolerance. The basis for this claim is that genes underpinning the trait appear to be most strongly regulated via pathways independent of both abiotic stress response and major plant height and flowering time genes (i.e. *Rht*, *PPD*, *Vrn*). Studies characterising different types of cell death in plants have shown that tapetal PCD is a form of autolytic PCD, which is associated with natural developmental processes, as opposed to the non-autolytic form, which is associated with hypersensitive response to stress and which is characteristic of necrotic like tissue damage (van Doorn and Woltering 2005).

Recent studies on tapetal PCD in rice and Arabidopsis (Chang *et al.*, 2014; Solis *et al.*, 2014; Zhang and Yang 2014) can now be used as a template to try and capture the full suite of tapetal PCD related genes in wheat. Using the large RNASeq data set derived in the current work the wheat spike transcriptome can be mined to identify allelic variants at tapetal PCD related loci, with a particular

focus on tracking allele segregation patterns in the WxK DH population. If through this work a significant causal relationship was established between allelic variants at one or a number of functionally related tapetal PCD loci it is expected these would be valuable additions to the proposed suite of biomarkers that could be used to quantify germplasm specific risk factors for RSWD stress induced pollen sterility.

Chapter 8:

8.1 Appendix I

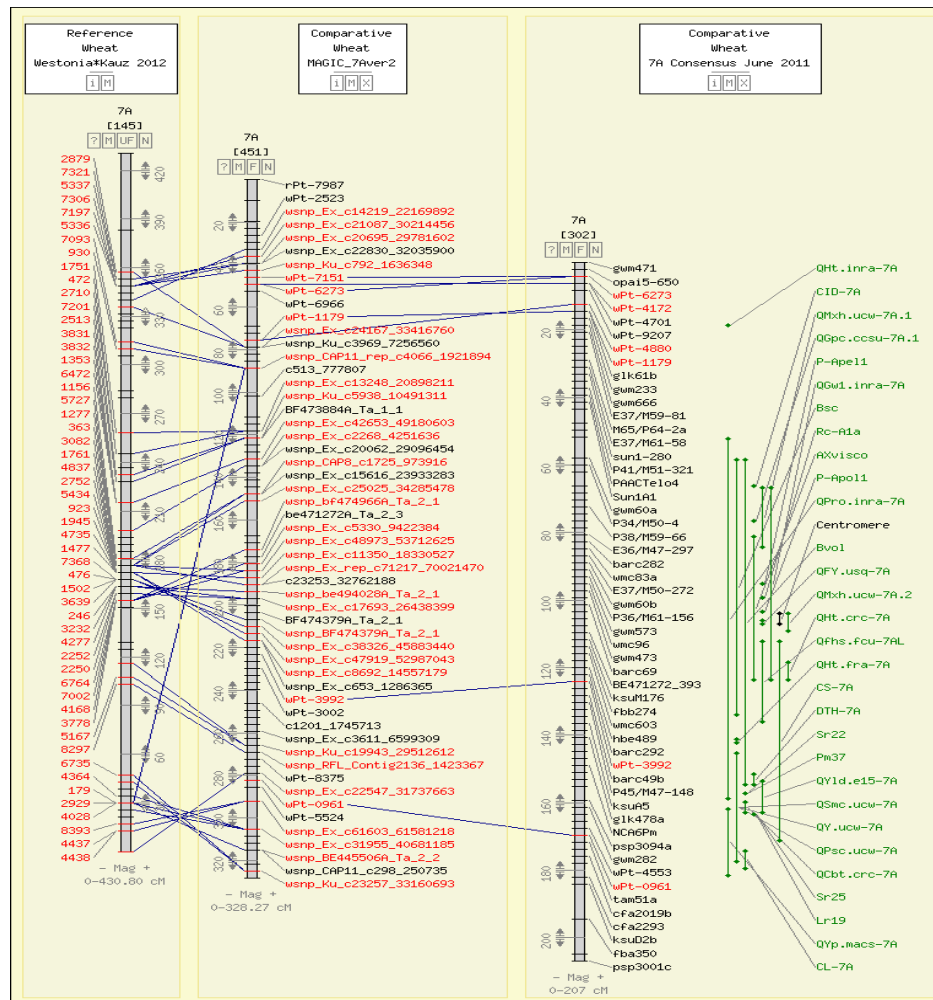


Figure 8.1 (From Diepeveen *et al.*, (in press)) The high resolution genetic maps shown in the figure derive from crosses between wheat lines Westonia x Kauz (far left panel) and a multi-parent cross called MAGIC (middle panel, based on Cavanagh *et al.*, (2013) and were produced by assaying many single nucleotide polymorphisms (SNPs) using the 9,000 SNP array made available by Illumina (Cavanagh *et al.*, (2013)). The genetic maps for chromosome 7A are shown as an example (for a published example for chromosome 2D of wheat Jia *et al.*, (2013)). The genetic map on the far right is a composite map and summarises the broad output from a literature focused on mapping QTL for phenotypes that are important in agriculture (green lines on the far right of the figure). The QTL usually occupy a region of the genetic map rather than a single point because of the variation inherent in measuring a phenotypic trait.

Additional information provided in Diepeveen *et al.*, (in press)

When crosses are made between two individuals, the genetic differences between the two individuals which can be defined at the level of DNA sequences,

will segregate among the progeny of the cross. The differences between individuals at the DNA levels are called polymorphisms and can be assayed as molecular markers using a range standard molecular technologies. The pattern of segregation of molecular markers relative to each other, among individual progeny, reflects their distribution along the long linear strand of DNA that makes up a chromosome so that molecular markers on different chromosomes will segregate at random from each other among the individual progeny from a cross. Molecular markers on the same chromosome will show non-random segregation from each other and this apparent linkage between the markers will depend on how far apart the molecular markers are from each other in the chromosome. When the molecular markers are far apart, the meiotic process of recombination has a greater probability of separating the molecular markers than when they are very close together (= closely linked) and thus the % recombination provides a measure of genetic distance. This distance is measured in units called centiMorgans (cM). When phenotypic characteristics are determined for each of the segregating lines derived from a cross, the variation in the traits measured can be statistically associated with a polymorphism in the molecular marker (= allele) between the parents used in the cross and since these associations generally account for proportions of variation in the phenotype, these associations are referred to as quantitative trait loci (QTL).

Molecular markers that are shared between the many crosses analyzed in the literature provide the basis for linking the maps to each other and, importantly, to link the QTL that have been mapped over an extensive period of time and in different environments, to the contemporary high resolution molecular genetic maps. The DNA sequences of the molecular markers also provide a direct link to the reference genome sequences that are being determined and allow the location, or projection, of QTL on to the DNA sequence for analysing the genes that underpin the QTL. The lines joining the maps indicate the shared location for a molecular marker and the Figure shows a generally good alignment between the maps although the detailed alignments of some regions indicate a reversal in the order of molecular markers in one map relate to the other. These

reversals in order may indicate genuine differences at the DNA level between wheat varieties due to an altered distribution of repetitive DNA sequences, or regions of the genetic map that require re-examination to check for errors. The maps illustrated are displayed using software called CMap (<http://ccg.murdoch.edu.au/cmap/ccg-live>)

8.2 Appendix II

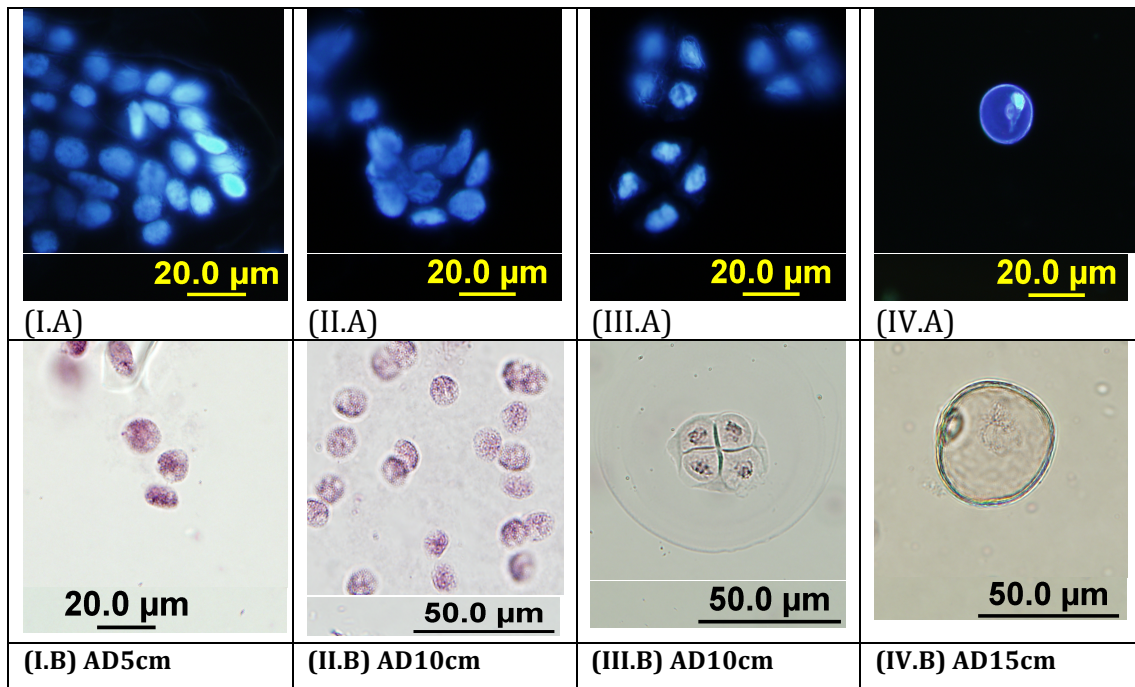


Figure 8.2 Cytological evaluation of timing of pollen development relative to auricle distance (cm) in WxK DH line D02-105, visualised by staining with aceto carmine (upper panel images I.A- IV.A) and DAPI (lower panel images I.B-IV.B). Data not shown for D08-299, Westonia, and Kauz. Image IA and IB (AD5cm) - pollen microspore mother cells (MMC's) undergoing meiosis I; Image II.A and II.B (AD10cm) - , pollen microspore mother cells (MMC's) undergoing meiosis II; Image III.A and III.B (AD10cm) - meiosis completed and tetrad of haploid microspores formed; Image IV.A and IV.B (AD15cm) - microspore disassociates from tetrad to form uninucleate microspore.

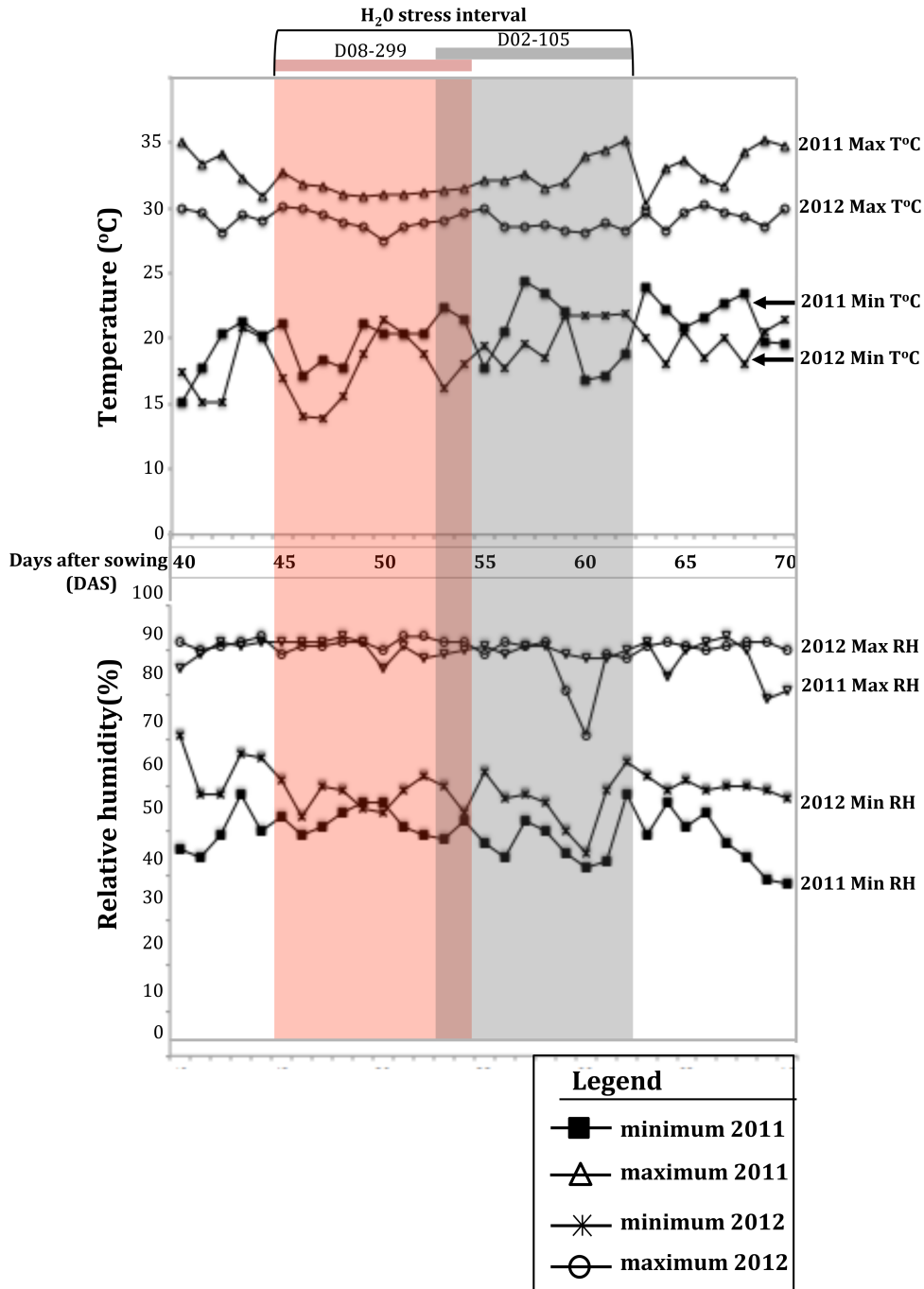


Figure 8.3 Daily threshold ranges for temperature (degrees Celsius) and relative humidity (%) for the duration of the experimental growth periods in the 2011 and 2012 WxK DH water stress experiments. Data was recorded in the glasshouse room of the Murdoch University PC2 controlled environment glasshouse facilities in which the experiment was conducted for both years of experimentation. Both relative humidity and temperature data are plotted as a line graph across time (days after sowing) on the x-axis. Temperature data is presented in the upper panel of the figure and relative humidity is presented in the lower panel. The y-axis shows the scale for which data is plotted against for temperature (upper panel) and relative humidity (lower panel). The lines of data representing upper and lower threshold values are annotated which shapes to enable differentiation (visual explanation for correspondence is provided in a legend inset). The RSWD stress interval for the two DH lines evaluated is annotated on the figure as red shading for D08-299 and grey shading for D02-105.

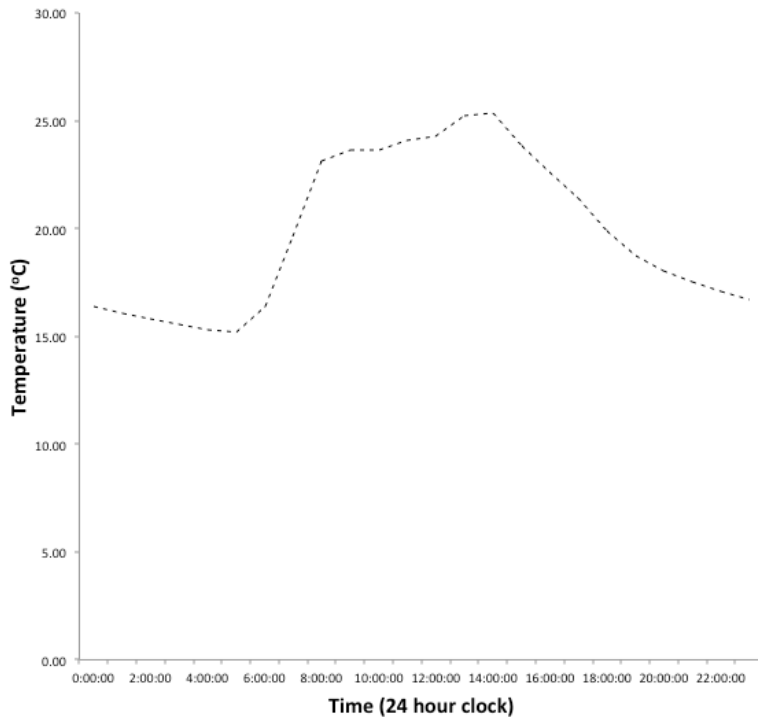


Figure 8.4 Hourly mean ambient temperature (°C) in the glasshouse calculated from data accumulated throughout duration of the plant growth period during the 2011 RSWD stress experiment.

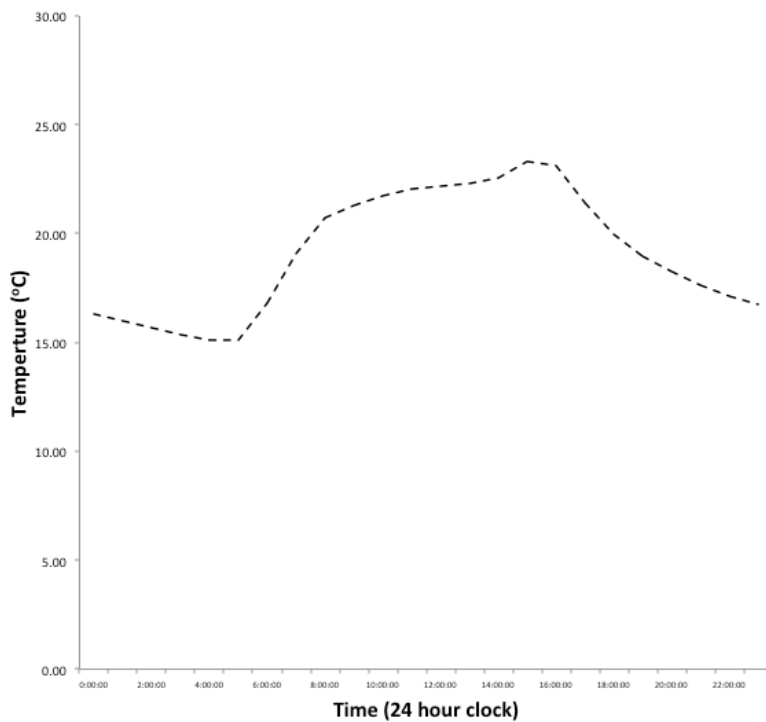


Figure 8.5 Hourly mean ambient temperature (°C) in the glasshouse calculated from data accumulated throughout duration of the plant growth period during the 2012 RSWD stress experiment.

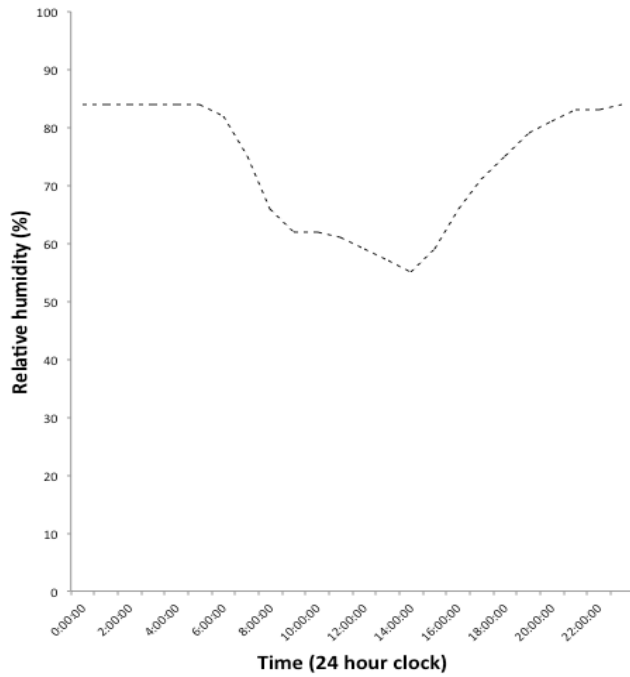


Figure 8.6 Hourly mean relative humidity (%) in the glasshouse calculated from data accumulated throughout duration of the plant growth period during the 2011 RSWD stress experiment.

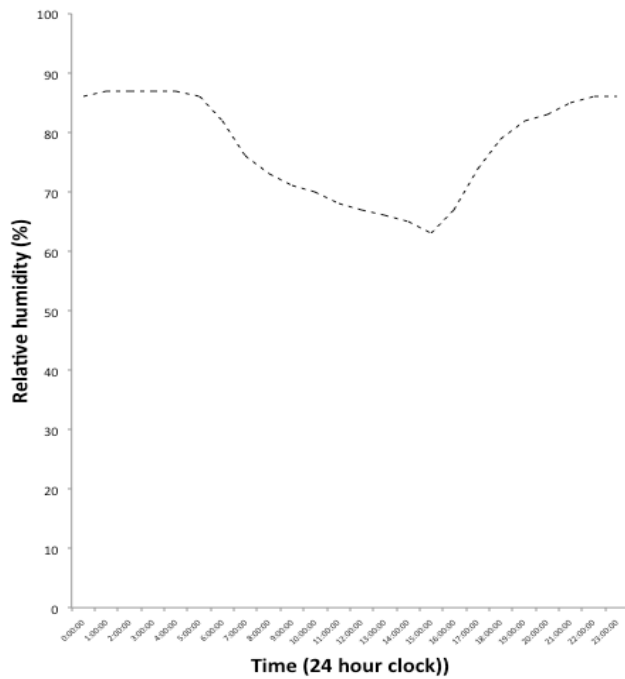


Figure 8.7 Hourly mean relative humidity (%) in the glasshouse calculated from data accumulated throughout duration of the plant growth period during the 2012 RSWD stress experiment.

8.3 Appendix III

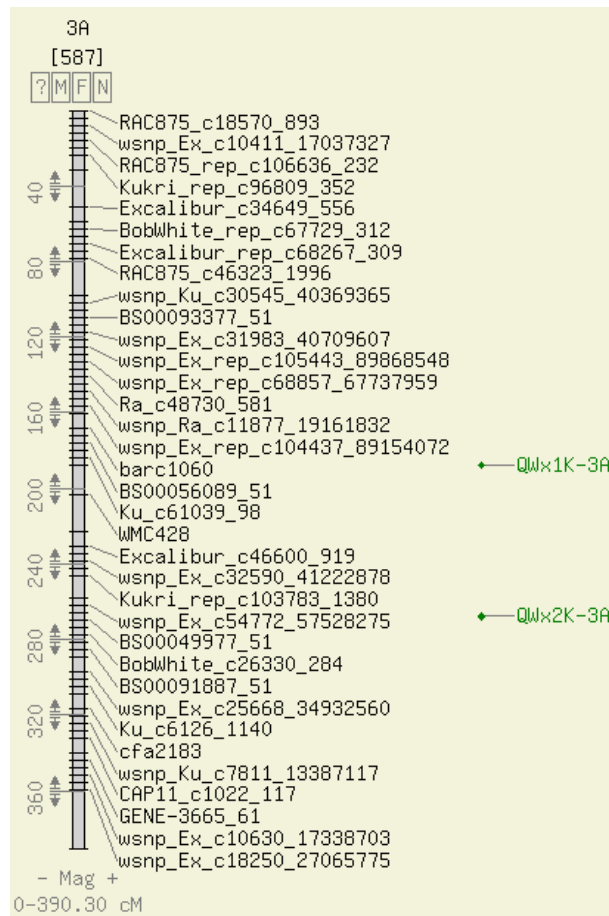


Figure 8.8 Genetic map of Westonia x Kauz DH genetic map wheat chromosome 3A (CMap ver0.16) annotated to show the position of QTL *QTLWxy-3AL*, a drought escape QTL for extent of waxy cuticle deposition on leaves and stems. The full details of the map shown, at a higher levels of magnification can be obtained from the CMap location at the website, <http://ccg.murdoch.edu.au/cmap/ccg-live>. The CMap ID for the map shown in the figure is: Westonia x Kauz9740.

Table 8.1 Statistical analysis of grain set (whole spike) data pooled from the 2011 and 2012 WxK DH RSWD stress experiments. Column heading are: DF - degrees of freedom, SS - Sums of squares, MS - mean sums of squares, P - P value.

	DF	SS	MS	P
Replicate	2	2.89	1.44	0.938
Year	1	26976.22	26976.22	<.001
Treatment	1	2006.43	2006.43	<.001
Variety	3	903.46	301.15	<.001
Year - Treatment	1	3090.56	3090.56	<.001
Year - variety	1	896.5	896.50	<.001
Treatment - variety	3	690.43	230.14	<.001
Year - Treatment - variety	1	1072.93	1072.93	<.001
Residual	46	1032.80	22.45	
Total	59	36672.21	621.56	

Additive effect* of QTL genotype on QTL phenotype

QQ = m + a

qq = m - a

Where:

- Q = Westonia allele
- q = Kauz allele
- m = mid parental value for the phenotyped trait **
- a = additive effect **

QtAD5-5BL

QQ genotype of D08-299
 = 59.5 + (-10.27)
 = 49.23

qq genotype in D02-105
 = 59.5 - (-10.27)
 = 69.77

QtAD5-5DL

QQ genotype of D08-299
 = 59.5 + (11.05)
 = 70.55

qq genotype in D02-105
 = 59.5 - (11.05)
 = 48.45

** additive effect unit of measure is days after sowing (DAS)
 ** mid parental and additive effect values are derived from Table 5.2*

Equation 8.1 Formula and calculations derived from the ICIM QTL analysis software used for determining the additive effect of the QTL genotype on the *QtAD5* trait.

References

- ABARE (2009) Australian Grains Report 2009.
- Abiko M, Akibayashi K, Sakata T, Kimura M, Kihara M, Itoh K, et al. Higashitani A (2005) High temperature induction of male sterility during barley (*Hordeum vulgare* L.) anther development is mediated by transcriptional inhibition. *Sexual Plant Reproduction* 18:91-100.
- ABS (2006) Australian Bureau of Statistics Year Book 2006. 1301.0.
- Acquaah G (2007) Principles of plant genetics and breeding. Blackwell Publishing Malden USA
- Ageez A, Kazama Y, Sugiyama R, Kawano S (2005) Male-fertility genes expressed in male flower buds of *Silene latifolia* include homologs of anther-specific genes. *Genes & genetic systems* 80:403-413.
- Ahmed IM, Cao F, Zhang M, Chen X, Zhang G, Wu F (2013) Difference in Yield and Physiological Features in Response to Drought and Salinity Combined Stress during Anthesis in Tibetan Wild and Cultivated Barleys. *PLoS one* 8:e77869.
- Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS, et al. Dvorak J (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res* 13:753-763.
- Alfares W, Bouguennec A, Balfourier F, Gay G, Berges H, Vautrin S, et al. Feuillet C (2009) Fine mapping and marker development for the crossability gene SKr on chromosome 5BS of hexaploid wheat (*Triticum aestivum* L.). *Genetics* 183:469-481.
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration guidelines for computing crop water requirements. Irrigation and drainage paper. Food and Agriculture Organization of the United Nations, pp. 1-5.
- Amasino R (2010) Seasonal and developmental timing of flowering. *The plant Journal* 61:1001-1013.
- Arenas-Huertero F, Arroyo A, Zhou L, Sheen J, Leon P (2000) Analysis of *Arabidopsis* glucose insensitive mutants, gin5 and gin6, reveals a central role of the plant hormone ABA in the regulation of plant vegetative development by sugar. *Genes and Development*:2085-2096.
- Arroyo A, Bossi F, Finkelstein R, R., Leon P (2003) Three genes that affect sugar sensing (Abscisic acid insensitive 4, Abscisic acid insensitive 5, and Constitutive triple response 1) are differentially regulated by glucose in *Arabidopsis*. *Plant Physiology* 133:231-242.
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208-218.
- Bioplatforms-Australia (2013) Variety and pedigree summary. <http://www.bioplatforms.com.au/special-initiatives/agriculture/wheat-datasets/wheat-sequencing/variety-sequencing/167-westonia>. Cited 14/10/2013
- Babu RC, Nguyen BD, Chamarek V, Shanmugasundaram P, Chezian P, Jeyaprakash P, et al. Sarkarung S (2003) Genetic analysis of drought resistance in rice by molecular markers. *Crop Science* 43:1457-1469.

- Barnabas B, Jager K, Feher A (2008) The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ* 31:11-38.
- Barrero RA, Bellgard M, Zhang X (2011) Diverse approaches to achieving grain yield in wheat. *Funct Integr Genomics* 11:37-48.
- Bennett MD, Rao MK, Smith JB, Bayliss MW (1973) Cell Development in the Anther, the Ovule, and the Young Seed of *Triticum aestivum* L. Var. Chinese Spring. *Philosophical Transactions of the Royal Society of London Series B* 266:39-81.
- Bernier Jrm, Kumar A, Ramaiah V, Spaner D, Atlin G (2007) A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. *Crop Science* 47:507-516.
- Biggin MD (2011) Animal transcriptome networks as highly connected, quantitative continua. *Dev Cell* 21:611-626.
- Blum A (2011) Plant Water Relations, Plant Stress and Plant Production Plant Breeding for Water-Limited Environments. Springer, pp. 11-52.
- Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. *Current Opinion in Plant Biology* 7:132-136.
- Bosco de Oliveira A, Mendes Alencar NA, Gomes-Filho E (2013) Comparison between the water and salt stress effects on plant growth and development. In: Akinci S (ed) Response of organisms to water stress. InTech, pp. 68-94.
- Bottley A, Koebner RM (2008) Variation for homoeologous gene silencing in hexaploid wheat. *Plant J* 56:297-302.
- Boyer JS (1982) Plant productivity and the environmental potential for increasing crop plant productivity, genotypic selection. *Science* 218:443-448.
- Breen J, Li D, Dunn DS, Bekes F, Kong X, Zhang J, et al. Ma W (2010(A)) Wheat beta-expansin (EXPB11) genes: identification of the expressed gene on chromosome 3BS carrying a pollen allergen domain. *BMC plant biology* 10:99.
- Breen J, Wicker T, Kong X, Zhang J, Ma W, Paux E, et al. Bellgard M (2010(B)) A highly conserved gene island of three genes on chromosome 3B of hexaploid wheat: diverse gene function and genomic structure maintained in a tightly linked block. *BMC plant biology* 10:98.
- Budak H, Kantar M, Yucebilgili Kurtoglu K (2013) Drought tolerance in modern and wild wheat. *The Scientific World Journal* 2013.
- Bush DR (1999) Sugar transporters in plant biology. *Curr Opin Plant Biol* 2:187-191.
- Butler JD, Byrne PF, Mohammadi V, Chapman PL, Haley SD (2005) Agronomic performance of the Rht alleles in a spring wheat population across a range of moisture levels. *Crop Science* 45:939-947.
- Cattivelli L, Baldi P, Crosatti C, Di Fonzo N, Faccioli P, Grossi M, et al. Stanca AM (2002) Chromosome regions and stress-related sequences involved in resistance to abiotic stress in Triticeae. *Plant Mol Biol* 48:649-665.
- Cattivelli L, Rizza F, Badeck F, W., Mazzucotelli E, Mastrangelo A, M., Francia E, et al. Stanca A, M. (2008) Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. *Field Crop Research* 105:1-14.

- Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol* 11:215-221.
- Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci U S A* 110:8057-8062.
- Chang F, Zhang Z, Jin Y, Ma H (2014) Cell Biological Analyses of Anther Morphogenesis and Pollen Viability in Arabidopsis and Rice Flower Development. Springer, pp. 203-216.
- Chaves MM, Flexas J, Pinheiro C (2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Ann Bot* 103:551-560.
- Cho JI, Lee SK, Ko S, Kim HK, Jun SH, Lee YH, et al. Jeon JS (2005) Molecular cloning and expression analysis of the cell-wall invertase gene family in rice (*Oryza sativa* L.). *Plant Cell Rep* 24:225-236.
- Choulet Fdr, Wicker T, Rustenholz C, Paux E, Salse Jr, Leroy P, et al. Gonthier C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *The Plant Cell Online* 22:1686-1701.
- CIMMYT (2013) Genetic Resources Information System for Wheat and Triticale (GRIS). <http://wheatpedigree.net>. Cited 07/10/2013
- Collins N, Tardieu F, Tuberosa R (2008) QTL approaches for improving crop performance under abiotic stress conditions: where do we stand. *Plant Physiology* 147:469-486.
- Conocono EA (2002) Improving yield of wheat experiencing post-anthesis water deficits through the use of shoot carbohydrate reserves University of Western Australia Perth.
- Crismani W, Kapoor S, Able JA (2011) Comparative Transcriptomics Reveals 129 Transcripts That Are Temporally Regulated during Anther Development and Meiotic Progression in Both Bread Wheat (*Triticum aestivum*) and Rice (*Oryza sativa*). *Int J Plant Genomics* 2011:931898.
- Csáki C (1985) Simulation and systems analysis in agriculture. Elsevier
- Dams T, Hunt KE (1977) Decision-making and agriculture: papers and reports, sixteenth International Conference of Agricultural Economists, held at Nairobi, Kenya, 26th July-4th August, 1976 16 International Conference of Agricultural Economists, Nairobi, Kenya (USA), 26 Jul-4 Aug 1976. University of Nebraska Press.
- Delhaize E, Ma JF, Ryan PR (2012) Transcriptional regulation of aluminium tolerance genes. *Trends Plant Sci* 17:341-348.
- Dembinska O, Lalonde S, Saini HS (1992) Evidence against the Regulation of Grain Set by Spikelet Abscisic Acid Levels in Water-Stressed Wheat. *Plant Physiol* 100:1599-1602.
- Demotes-Mainard S, Doussinault G, Meynard JM, Gate P (1996) Is it possible to diagnose a harvest problem of pollen sterility in wheat? *European Journal of Agronomy* 5:169-180.
- Devshwar P, Bovill WD, Sharma R, Able JA, Kapoor S (2011) Analysis of anther transcriptomes to identify genes contributing to meiosis and male gametophyte development in rice. *BMC plant biology* 11:78.

- Diepeveen D (2011) Data integration for decision making in wheat breedingMurdoch University.
- Diepeveen D, Webster H, Appels R (in press) Genomics: Plant genetic improvement. In: Leakey R (ed) Encyclopedia of Agricultural Science (2nd edition). Elsevier.
- Dillies M-As, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, et al.Estelle J (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in bioinformatics* 14:671-683.
- Dobritsa AA, Nishikawa S-I, Preuss D, Urbanczyk-Wochniak E, Sumner LW, Hammond A, et al.Swanson RJ (2009) LAP3, a novel plant protein required for pollen development, is essential for proper exine formation. *Sexual Plant Reproduction* 22:167-177.
- Dolferus R, Ji X, Richards RA (2011) Abiotic stress and control of grain number in cereals. *Plant Sci* 181:331-341.
- Dolferus R, Powell N, Xuemei J, Ravash R, Edlington J, Oliver S, et al.Shiran B (2013) The Physiology of Reproductive-Stage Abiotic Stress Tolerance in CerealsMolecular Stress Physiology of Plants. Springer, pp. 193-216.
- Dorion S, Lalonde S, Saini HS (1996) Induction of Male Sterility in Wheat by Meiotic-Stage Water Deficit Is Preceded by a Decline in Invertase Activity and Changes in Carbohydrate Metabolism in Anthers. *Plant Physiology* 111:137-145.
- Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316:1862-1866.
- Dvorak J, Luo MC, Yang ZL (1998) Genetic evidence of the orogin of *Triticum aestvum* L. . In: Damania AB, Valkoun G, Willcox G, Qualset CO (eds) Harlan Symposium. ICARDA Aleppo.
- Eschrich W (1980) Free space invertase, its possible role in phloem unloading. *Ber Dtsch Bot Ges* 93:363-378.
- Esnaola M, Puig P, Gonzalez D, Castelo R, Gonzalez JR (2013) A flexible count data model to fit the wide diversity of expression profiles arising from extensively replicated rna-seq experiments. *BMC Bioinformatics* 14:254.
- FAO (2010) FAOSTAT - Global wheat export value 2010.
- Ferreira SJ, Sonnewald U (2012) The mode of sucrose degradation in potato tubers determines the fate of assimilate utilization. *Frontiers in plant science* 3:23.
- Feuillet C, Stein N, Rossini L, Praud S, Mayer K, Schulman A, et al.Appels R (2012) Integrating cereal genomics to support innovation in the Triticeae. *Funct Integr Genomics* 12:573-583.
- Fischer RA (1973) The effect of water stress at various stages of development on yield processes in wheat. In: Slatyer RO (ed) *Plant response to climatic factors*. UNESCO Paris, pp. 233-241.
- Fischer RA (2007) Understanding the physiological basis of yeild potential in wheat. *Journal of Agricultural Science* 145:99-113.
- Fleury D, Jefferies S, Kuchel H, Langridge P (2010) Genetic and genomic tools to improve drought tolerance in wheat. *J Exp Bot* 61:3211-3222.
- Franchi GG, Bellani L, Nepi M, Pacini E (1996) types of carbohydrate reserves in pollen localization, systematic distribution and ecophysiological significance. *Flora* 191:143-159.

- Francki M, Appels R (2002) Wheat functional genomics and engineering crop improvement. *Genome Biol* 3:reviews1013.
- Francki MG, Walker E, Forster JW, Spangenberg G, Appels R (2006) Fructosyltransferase and invertase genes evolved by gene duplication and rearrangements: rice, perennial ryegrass, and wheat gene families. *Genome* 49:1081-1091.
- Fredlund DG, Xing A (1994) Equations for the soil-water characteristic curve. *Canadian Geotechnical Journal* 31:521-532.
- Frenkel Z, Paux E, Mester D, Feuillet C, Korol A (2010) LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. *BMC Bioinformatics* 11:584.
- Fu G-f, Song J, Xiong J, Li Y-r, Chen H-z, Le M-k, Tao L-x (2011) Changes of oxidative stress and soluble sugar in anthers involve in rice pollen abortion under drought stress. *Agricultural Sciences in China* 10:1016-1025.
- Galiba G, Quarrie S, Sutka J, Morgounov A, Snape J (1995) RFLP mapping of the vernalization (*Vrn1*) and frost resistance (*Fr1*) genes on chromosome 5A of wheat. *Theoretical and Applied Genetics* 90:1174-1179.
- Gill BS, Appels R, Botha-Oberholster AM, Buell CR, Bennetzen JL, Chalhoub B, et al. Sasaki T (2004) A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium. *Genetics* 168:1087-1096.
- Glusman G, Caballero J, Robinson M, Kutlu B, Hood L (2013) Optimal scaling of digital transcripts. *PloS one* 8:e77885.
- Gocal G, F, W., Sheldon C, Gubler F, Moritz T, Bagnall D, J., MacMillan C, P., et al. King R (2001) GAMB-like genes, flowering, and gibberellin signaling in *Arabidopsis*. *Plant Physiology* 127:1682-1693.
- Godt DE, Roitsch T (1997) Regulation and tissue-specific distribution of mRNAs for three extracellular invertase isoenzymes of tomato suggests an important function in establishing and maintaining sink metabolism. *Plant Physiol* 115:273-282.
- Goetz M, Godt DE, Guivarc'h A, Kahmann U, Chriqui D, Roitsch T (2001) Induction of male sterility in plants by metabolic engineering of the carbohydrate supply. *Proc Natl Acad Sci U S A* 98:6522-6527.
- Golberg R, B., Beals T, P., Sanders P, M. (1993) Anther development: Basic principles and practices. *The Plant Cell* 5:12.
- Goldberg RB, Sanders PM, Beals TP (1995) A novel cell-ablation strategy for studying plant development. *Philos Trans R Soc Lond B Biol Sci* 350:5-17.
- Gothandam K, Kim E, Chung Y (2007) Ultrastructural study of rice tapetum under low-temperature stress. *Journal of Plant Biology* 50:396-402.
- Greiner S, Koster U, Lauer K, Rosengranz H, Vogel R, Rausch T (2000) Plant invertase inhibitors: expression in cell culture and during development. *Plant Physiology* 27:807-814.
- Guo Y, Xiong L, Song CP, Gong D, Halfter U, Zhu JK (2002) A calcium sensor and its interacting protein kinase are global regulators of abscisic acid signaling in *Arabidopsis*. *Dev Cell* 3:233-244.
- Gupta AK, Kaur N (2005) Sugar signalling and gene expression in relation to carbohydrate metabolism under abiotic stresses in plants. *J Biosci* 30:761-776.

- Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. *Plant Physiol* 130:639-648.
- Hart G, Gale M, McIntosh R (1993) Linkage maps of *Triticum aestivum* (hexaploid wheat, $2n= 42$, genomes A, B & D) and *T. tauschii* ($2n= 14$, genome D) Progress in genome mapping of wheat and related species: proceedings of the 3rd Public Workshop of the International Triticeae Mapping Initiative, 22-26 September 1992, CIMMYT, Mexico. El Batañón, Mexico: International Maize and Wheat Improvement Center, [1992?
- Hawker JS, Jenner CF, Niemietz CM (1991) Sugar metabolism and compartmentation. *Australian Journal of Plant Physiology* 18:227-237.
- Helenius A, Tatu U, Marquardt T, Braakman I (1993) Protein folding in the endoplasmic reticulum *Cell biology and biotechnology*. Springer, pp. 125-136.
- Hey SJ, Byrne E, Halford NG (2010) The interface between metabolic and stress signalling. *Ann Bot* 105:197-203.
- Higgins JD (2012) Spatiotemporal asymmetry of recombination initiation and prophase I progression underlies the predominantly distal distribution of meiotic crossovers in Barley. *The Plant Cell*.
- Hirsch CD, Evans J, Buell CR, Hirsch CN (2014) Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. *Briefings in functional genomics* elt051.
- Hlavackova I, Vitamvas P, Santrucek J, Kosova K, Zelenkova S, Prasil IT, et al. Kodicek M (2013) Proteins involved in distinct phases of cold hardening process in frost resistant winter barley (*Hordeum vulgare* L.) cv Luxor. *International journal of molecular sciences* 14:8000-8024.
- Huang Z, Phoolcharoen W, Lai H, Piensook K, Cardineau G, Zeitlin L, et al. Chen Q (2010) High-level rapid production of full-size monoclonal antibodies in plants by a single-vector DNA replicon system. *Biotechnol Bioeng* 106:9-17.
- Hurtley SM, Helenius A (1989) Protein oligomerization in the endoplasmic reticulum. *Annual review of cell biology* 5:277-307.
- Huysmans S, El-Ghazaly G, Smets E (1998) Orbicules in Angiosperms: Morphology, Function, Distribution, and Relation with Tapetum Types. *Botanical Review* 64:240-272.
- Hwang YS, Karrer EE, Thomas BR, Chen L, Rodriguez RL (1998) Three cis-elements required for rice alpha-amylase Amy3D expression during sugar starvation. *Plant Mol Biol* 36:331-341.
- IPCC (2007) IPCC fourth assessment report-climate change 2007.
- IWGSC (in press) A chromosome-based draft sequence of the hexaploid bread wheat genome. *Science*.
- Jain M, Prasad PV, Boote KJ, Hartwell Jr AL, Chourey PS (2007) Effects of season-long high temperature growth conditions on sugar-to-starch metabolism in developing microspores of grain sorghum (*Sorghum bicolor* L. Moench). *Planta* 227:67-79.
- Ji X, Dong B, Shiran B, Talbot M, J., Edlington J, E., Huches T, et al. Dolferus R (2011) Control of ABA catabolism and ABA homeostasis is important for reproductive stage stress tolerance in cereals. *Experimental Botany* 55:15.

- Ji X, Shiran B, Wan J, Lewis DC, Jenkins CLD, Condon AG, et al. Dolferus R (2010) Importance of pre-anthesis anther sink strength for maintenance of grain number during reproductive stage water stress in wheat. *Plant, Cell & Environment* 33:926-942.
- Ji X, Van den Ende W, Van Laere A, Cheng S, Bennett J (2005) Structure, evolution, and expression of the two invertase gene families of rice. *J Mol Evol* 60:615-634.
- Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, et al. Mao L (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496:91-95.
- Jin Y, Ni DA, Ruan YL (2009) Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. *Plant Cell* 21:2072-2089.
- Jin Y, Yang H, Wei Z, Ma H, Ge X (2013) Rice Male Development under Drought Stress: Phenotypic Changes and Stage-dependent Transcriptomic Reprogramming. *Molecular plant*.
- Jitsuvara Y, Toyoda T, Itai T, Yamaguchi H (2002) Chaperone-like functions of high-mannose type and complex-type N-glycans and their molecular basis. *J Biochem* 132:803-811.
- Jung C, Muller AE (2009) Flowering time control and applications in plant breeding. *Trends in Plant Science* 14:563-573.
- Kaneko M, Inukai Y, Ueguchi-Tanaka M, Itoh H, Izawa T, Kobayashi Y, et al. Matsuoka M (2004) Loss-of-function mutations of the rice *GAMYB* gene impair alpha-amylase expression in aleurone and flower development. *Plant Cell* 16:33-44.
- Kapoor S, Kobayashi A, Takatsuji H (2002) Silencing of the tapetum-specific zinc finger gene *TAZ1* causes premature degeneration of tapetum and pollen abortion in petunia. *Plant Cell* 14:2353-2367.
- Kaufman PB, Ghosheh NS, Lacroix JD, Soni SL, Ikuma H (1973) Regulation of invertase levels in *Avena* stem segments by gibberellic Acid, sucrose, glucose, and fructose. *Plant Physiol* 52:221-228.
- Kempton R (1982) Adjustment for competition between varieties in plant breeding trials. *Journal of Agricultural Science* 98:599-611.
- Kilian B, Ozkan H, Pozzi C, Salamini S (2009) Domestication of the Triticeae in the fertile crescent. In: Feuillet C, Muehlbauer GJ (eds) *Genetics and genomics of the Triticeae*. Springer Science + Business Media.
- Kim SY, Hong CB, Lee I (2001) Heat shock stress causes stage-specific male sterility in *Arabidopsis thaliana*. *Journal of Plant Research* 114:301-307.
- Kim TH, Bohmer M, Hu H, Nishimura N, Schroeder JI (2010) Guard cell signal transduction network: advances in understanding abscisic acid, CO₂, and Ca²⁺ signaling. *Annu Rev Plant Biol* 61:561-591.
- Kleczkowski LA, Decker D, Wilczynska M (2011) UDP-sugar pyrophosphorylase: a new old mechanism for sugar activation. *Plant Physiology* 156:3-10.
- Koch K (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr Opin Plant Biol* 7:235-246.
- Koch KE (1996) Carbohydrate-Modulated Gene Expression in Plants. *Annu Rev Plant Physiol Plant Mol Biol* 47:509-540.
- Koonjul PK, Minhas JS, Nunes C, Sheoran IS, Saini HS (2005) Selective transcriptional down-regulation of anther invertases precedes the failure

- of pollen development in water-stressed wheat. *Journal of Experimental Botany* 56:179-190.
- Ku S, Yoon H, Suh HS, Chung Y-Y (2003) Male-sterility of thermosensitive genic male-sterile rice is associated with premature programmed cell death of the tapetum. *Planta* 217:559-565.
- Kumar A, Verulkar S, Mandal N, Variar M, Shukla V, Dwivedi J, et al. Mall A (2012) High-yielding, drought-tolerant, stable rice genotypes for the shallow rainfed lowland drought-prone ecosystem. *Field Crops Research* 133:37-47.
- Kumar R, Venuprasad R, Atlin G (2007) Genetic analysis of rainfed lowland rice drought tolerance under naturally-occurring stress in eastern India: heritability and QTL effects. *Field Crops Research* 103:42-52.
- Lalonde S, Beebe DU, Saini HS (1997) Early signs of disruption of wheat anther development associated with the induction of male sterility by meiotic-stage water deficit. *Sexual Plant Reproduction* 10:40-48.
- Langridge P, Paltridge N, Fincher G (2006) Functional genomics of abiotic stress tolerance in cereals. *Brief Funct Genomic Proteomic* 4:343-354.
- Le Roy K, Lammens W, Verhaest M, De Coninck B, Rabijns A, Van Laere A, Van den Ende W (2007) Unraveling the difference between invertases and fructan exohydrolases: a single amino acid (Asp-239) substitution transforms Arabidopsis cell wall invertase1 into a fructan 1-exohydrolase. *Plant Physiol* 145:616-625.
- Le Roy K, Vergauwen R, Struyf T, Yuan S, Lammens W, Matrai J, et al. Van den Ende W (2013) Understanding the role of defective invertases in plants: tobacco Nin88 fails to degrade sucrose. *Plant Physiol* 161:1670-1681.
- Leroy PI, Guilhot N, Sakai H, Bernard A, Choulet F, Theil S, et al. Feuillet C (2012) TriAnnot: a versatile and high performance pipeline for the automated annotation of plant genomes. *Frontiers in Plant Science* 3:1-14.
- Li H, Hearne S, Banziger M, Li Z, Wang J (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity* 105:257-267.
- Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. *Genetics* 175:361-374.
- Li N, Zhang DS, Liu HS, Yin CS, Li XX, Liang WQ, et al. Zhang DB (2006) The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. *Plant Cell* 18:2999-3014.
- Liu S, Zhang X, Pumphrey MO, Stack WS, Gill BS, Anderson JA (2006) Complex microlinearity among wheat, rice, and barley revealed by fine mapping of the genomic region harboring a major QTL for resistance to Fusarium head blight in wheat. *Funct Integr Genomics* 6:83-89.
- Liu Y, Cui S, Wu F, Yan S, Lin X, Du X, et al. Meng Z (2013) Functional conservation of MIKC*-Type MADS box genes in Arabidopsis and rice pollen maturation. *The Plant Cell Online* 25:1288-1303.
- Long JC, Zhao W, Rashotte AM, Muday GK, Huber SC (2002) Gravity-stimulated changes in auxin and invertase gene expression in maize pulvinal cells. *Plant Physiol* 128:591-602.
- Ma D, Yan J, He Z, Wu L, Xia X (2012) Characterization of a cell wall invertase gene TaCwi-A1 on common wheat chromosome 2A and development of functional markers. *Molecular Breeding* 29:43-52.

- Ma H (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Plant Biology* 56:393-434.
- Maccaferri M, Sanguineti MC, Demontis A, El-Ahmed A, Garcia del Moral L, Maalouf F, et al. Tuberosa R (2011) Association mapping in durum wheat grown across a broad range of water regimes. *J Exp Bot* 62:409-438.
- Maccaferri M, Sanguineti MC, Natoli V, Ortega JLA, Salem MB, Bort J, et al. Demontis A (2006) A panel of elite accessions of durum wheat (*Triticum durum* Desf.) suitable for association mapping studies. *Plant Genetic Resources: characterization and utilization* 4:79-85.
- Maglietta R, Liuzzi VC, Cattaneo E, Laczko E, Piepoli A, Panza A, et al. Buffoli F (2012) Molecular pathways undergoing dramatic transcriptomic changes during tumor development in the human colon. *BMC cancer* 12:608.
- Majewska-Sawka A, Rodriguez-Garcia MI, Jassen B (1993) Ultrastructural expression of cytoplasmic male sterility in sugar beet (*Beta vulgaris*). *Sexual Plant Reproduction* 6:22-32.
- Mamun EA, Alfred S, Cantrill LC, Overall RL, Sutton BG (2006) Effects of chilling on male gametophyte development in rice. *Cell Biol Int* 30:583-591.
- Mamun EA, Cantrill LC, Overall RL, Sutton BG (2005) Cellular organisation in meiotic and early post-meiotic rice anthers. *Cell Biology International* 29:903-913.
- Martin LB, Fei Z, Giovannoni JJ, Rose JK (2013) Catalyzing plant science research with RNA-seq. *Frontiers in plant science* 4.
- Mascarenhas J, P. (1990) Gene activity during pollen development. *Plant Physiology* 41:317-338.
- McCarthy JD, Chen Y, Smyth GK (2012) Differential expression analysis with respect to biological variation. *Nucleic Acids Res* 40:4288-4297.
- McCormick S (1993) Male Gametophyte Development. *Plant Cell* 5:1265-1275.
- McCormick S (2004) Control of male gametophyte development. *Plant Cell* 16 Suppl:S142-153.
- McGettigan PA (2013) Transcriptomics in the RNA-seq era. *Curr Opin Chem Biol* 17:4-11.
- McIntyre CL, Pereira S, Moran LB, Appels R (1990) New *Secale cereale* (rye) DNA derivatives for the detection of rye chromosome segments in wheat. *Genome* 33:635-640.
- McNally KL, Bruskiewich R, Mackill D, Buell CR, Leach JE, Leung H (2006) Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiology* 141:26-31.
- McNeil M, Diepeveen D, Wilson R, Barclay I, McLean R, Chalhoub B, Appels R (2009) Haplotype analyses in wheat for complex traits: tracking the chromosome 3B and 7B regions associated with late maturity alpha amylase (LMA) in breeding programs. *Crop and Pasture Science* 60:463-471.
- Meier U (2001) Growth stages of mono- and dicotyledonous plants. Monograph, BBCH. Federal Biological Research Centre for Agriculture and Forestry.
- Meron M, Grimes D, Phene C, Davis K (1987) Pressure chamber procedures for leaf water potential measurements of cotton. *Irrigation Science* 8:215-222.
- Miedaner T, Korzun V (2012) Marker-assisted selection for disease resistance in wheat and barley breeding. *Phytopathology* 102:560-566.

- Millar AA, Gubler F (2005) The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* 17:705-721.
- Miralles D, Slafer G (1995) Yield, biomass and yield components in dwarf, semi-dwarf and tall isogenic lines of spring wheat under recommended and late sowing dates. *Plant Breeding* 114:392-396.
- Mishra KK, Vikram P, Yadaw RB, Swamy BM, Dixit S, Cruz MTS, et al. Kumar A (2013) qDTY 12.1: a locus with a consistent effect on grain yield under drought in rice. *BMC genetics* 14:1-10.
- Mochida K, Yamazaki Y, Ogihara Y (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol Genet Genomics* 270:371-377.
- Moghaddam MRB, Van den Ende W (2012) Sugars and plant innate immunity. *Journal of Experimental Botany* 63:3989-3998.
- Morgante M, Salamini F (2003) From plant genomics to breeding practice. *Curr Opin Biotechnol* 14:214-219.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 5:621-628.
- Munns R, James RA, Sirault XR, Furbank RT, Jones HG (2010) New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. *Journal of Experimental Botany* 61:3499-3507.
- Mustilli AC, Merlot S, Vavasseur A, Fenzi F, Giraudat J (2002) Arabidopsis OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* 14:3089-3099.
- Naylor RL (2009) Managing food production systems for resilience. *Principles of Ecosystem Stewardship*. Springer, pp. 259-280.
- Nellemann C (2009) *The Environmental Food Crisis: The Environment's Role in Averting Future Food Crises: a UNEP Rapid Response Assessment*. UNEP/Earthprint
- Nesbitt M, Samuel D (1996) Hulled wheats. In: Padulosi S, Hammer K, Heller J (eds) *Proceedings of the 1st International Workshop on Hulled Wheats*. International Plant Genetics Research Institute Rome.
- Nguyen TT, Klueva N, Chamareck V, Aarti A, Magpantay G, Millena AC, et al. Nguyen HT (2004) Saturation mapping of QTL regions and identification of putative candidate genes for drought tolerance in rice. *Mol Genet Genomics* 272:35-46.
- Ni J, Pujar A, Youens-Clark K, Yap I, Jaiswal P, Teclé I, et al. McCouch S (2009) Gramene QTL database: development, content and applications. *Database* (Oxford) 2009:bap005.
- Niewiadomski P, Knappe S, Geimer S, Fischer K, Zschulz B, Unte US, et al. Schneider A (2005) The arabidopsis plastidic glucose 6-Phosphate/Phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. *The Plant Cell* 17:760-775.
- Nilson SE, Assmann SM (2010) Heterotrimeric G proteins regulate reproductive trait plasticity in response to water availability. *New Phytol* 185:734-746.

- Oliver SN, Dennis ES, Dolferus R (2007) ABA regulates apoplastic sugar transport and is a potential signal for cold-induced pollen sterility in rice. *Plant and Cell Physiology* 48:1319-1330.
- Oliver SN, Van Dongen JT, Alfred SC, Mamun EA, Zhao X, Saini HS, et al. Dolferus R (2005) Cold-induced repression of the rice anther-specific cell wall invertase gene OSINV4 is correlated with sucrose accumulation and pollen sterility. *Plant, Cell & Environment* 28:1534-1551.
- Oshino T, Abiko M, Saito R, Ichiishi E, Endo M, Kawagishi-Kobayashi M, Higashitani A (2007) Premature progression of anther early developmental programs accompanied by comprehensive alterations in transcription during high-temperature injury in barley plants. *Mol Genet Genomics* 278:31-42.
- Owen HA, Makaroff CA (1995) Ultrastructure of microsporogenesis and microgametogenesis in *Arabidopsis thaliana* (L.) Heynh. ecotype Wassilewskija (Brassicaceae). *Protoplasma* 185:7-21.
- Pacini E (1996) Types and meaning of pollen carbohydrate reserves. *Sexual Plant Reproduction* 9:362-366.
- Papini A, Mosti S, Brighigna L (1999) Programmed-cell-death events during tapetum development of angiosperms. *Protoplasma*:213-221.
- Papini E (1990) Tapetum and microspore function. In: Blackmore S, Knox RB (eds) *Microspores: Evolution and Ontogeny*. Academic Press London.
- Parish RW, Li SF (2010) Death of a tapetum: A programme of developmental altruism. *Plant Science* 178:73-89.
- Parish RW, Phan HA, Iacuone S, Li SF (2012) Tapetal development and abiotic stress: a centre of vulnerability. *Functional Plant Biology* 39:553-559.
- Pasquariello M, Barabaschi D, Himmelbach A, Steuernagel B, Ariyadasa R, Stein N, et al. Tagliafico E (2014) The barley Frost resistance-H2 locus. *Functional & integrative genomics*:1-16.
- Passioura J (2007) The drought environment: physical, biological and agricultural perspectives. *J Exp Bot* 58:113-117.
- Patrick JW (1997) Phloem unloading: Sieve Element Unloading and Post-Sieve Element Transport. *Annu Rev Plant Physiol Plant Mol Biol* 48:191-222.
- Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, et al. Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* 322:101-104.
- Peltonen-Sainio P, Kangas A, Salo Y, Jauhiainen L (2007) Grain number dominates grain weight in temperate cereals yield determination: evidence based on 30 years of multi-location trials. *Field Crop Research* 100:179-188.
- Peng Z, Wang M, Li F, Lv H, Li C, Xia G (2009) A proteomic study of the response to salinity and drought stress in an introgression strain of bread wheat. *Molecular & Cellular Proteomics* 8:2676-2686.
- Phoolcharoen W, Bhoo SH, Lai H, Ma J, Arntzen CJ, Chen Q, Mason HS (2011) Expression of an immunogenic Ebola immune complex in *Nicotiana benthamiana*. *Plant Biotechnol J* 9:807-816.
- Piepho HP (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. *Genetics* 156:2043-2050.

- Powell N, Ji X, Ravash R, Edlington J, Dolferus R (2012) Yield stability for cereals in a changing climate. *Functional Plant Biology* 39:539-552.
- Pressman E, Peet MM, Pharr DM (2002) The effect of heat stress on tomato pollen characteristics is associated with changes in carbohydrate concentration in the developing anthers. *Annals of Botany* 90:631-636.
- Proels RK, Roitsch T (2009) Extracellular invertase LIN6 of tomato: a pivotal enzyme for integration of metabolic, hormonal, and stress signals is regulated by a diurnal rhythm. *J Exp Bot* 60:1555-1567.
- Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, et al. Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiol* 133:1755-1767.
- Raghavan V (1988) Anther and pollen development in rice (*Oryza sativa*). *American Journal of Botany*:183-196.
- Rahaie M, Xue GP, Schenk PM (2013) The Role of Transcription Factors in Wheat Under Different Abiotic Stresses. *development*, 2(4), 59. In: Vahdati K, Leslie C (eds) *Abiotic Stress - Plant Responses and Applications in Agriculture*. InTech, pp. 368-385.
- Rajaram S, Borlaug NE, Ginkel M (2002) CIMMYT International wheat breeding. In: Curtis BC, Ramaram S, Gomez Macpherson H (eds) *Bread wheat improvement and production*. The Food and Agriculture Organization of the United Nations Rome Italy.
- Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, et al. Betel D (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome biology* 14:R95.
- Rapp RA, Haigler CH, Flagel L, Hovav RH, Udall JA, Wendel JF (2010) Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. *BMC Biol* 8:139.
- Rausch T, Greiner S (2004) Plant protein inhibitors of invertases. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* 1696:253-261.
- Reynolds M, Foulkes MJ, Slafer GA, Berr P, Parry MAJ, Snape JW, Angus WJ (2009(A)) Raising yeild potential in wheat. *Journal of Experimental Botany* 60:1899-1918.
- Reynolds M, Manes Y, Izanloo A, Langridge P (2009(B)) Phenotyping approaches for physiological breeding and gene discovery in wheat. *Annals of Applied Biology* 155:309-320.
- Reynolds M, Pellegrineschi A, Slovmann B (2005) Sink-limitation to yield and biomass: a summary of some investigations in spring wheat. *Annals of Applied Biology* 146:39-49.
- Reznickova SA, Willemse MTM (1980) Formation of pollen in the anther of *Lilium*. II. The function of the surrounding tissues in the formation of pollen and pollen wall. *Acta Botanica Neerlandica* 29:141-156.
- Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome biology* 12:R22.
- Roitsch T (1999) Source-sink regulation by sugar and stress. *Current Opinion in Plant Biology* 2:198-206.

- Roitsch T, Balibrea ME, Hofmann M, Proels R, Sinha AK (2003) Extracellular invertase: key metabolic enzyme and PR protein. *J Exp Bot* 54:513-524.
- Roitsch T, Bittner M, Godt DE (1995) Induction of apoplastic invertase of *Chenopodium rubrum* by D-glucose and a glucose analog and tissue-specific expression suggest a role in sink-source regulation. *Plant Physiol* 108:285-294.
- Roitsch T, Gonzalez MC (2004) Function and regulation of plant invertases: sweet sensations. *Trends Plant Sci* 9:606-613.
- Rolland F, Baena-Gonzalez E, Sheen J (2006) Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol* 57:675-709.
- Rook F, Card R, Munz G, Smith C, Bevan M, W. (2001) Impaired sucrose-induced mutants reveal the modulation of sugar-induced starch biosynthetic gene expression by abscisic acid signalling. *The plant Journal* 26:421-433.
- Rosenquist S (2007) Plant sugar signaling: regulation of starch and fructan metabolism. *Plant Biology and Forest Genetics*. Swedish University of Agricultural Sciences Uppsala.
- Ruan Y-L, Jin Y, Yang Y-J, Li G-J, Boyer JS (2010) Sugar Input, Metabolism, and Signaling Mediated by Invertase: Roles in Development, Yield Potential, and Response to Drought and Heat. *Molecular Plant* 3:942-955.
- Ruan YL (2012) Signaling role of sucrose metabolism in development. *Mol Plant* 5:763-765.
- Saaty TL (2005) Theory and applications of the analytic network process: decision making with benefits, opportunities, costs, and risks. RWS publications
- Saini H, Sedgley M, Aspinall D (1984) Development anatomy in wheat of male sterility induced by heat stress, water deficit or abscisic acid. *Functional Plant Biology* 11:243-253.
- Saini HS (1997) Effects of water stress on male gametophyte development in plants. *Sexual Plant Reproduction* 10:67-73.
- Saini HS, Aspinall D (1981) Effect of Water Deficit on Sporogenesis in Wheat (*Triticum aestivum* L.). *Annals of Botany* 48:623-633.
- Saini HS, Lalonde S (1998) Injuries to reproductive development under water stress, and their consequences for crop productivity. *Journal of Crop Production* 1:223-248.
- Saini HS, Westgate ME (1999) Reproductive development in grain crops during drought. *Advances in agronomy* 68:59-96.
- Saintenac C, Faure S, Remay A, Choulet F, Ravel C, Paux E, et al. Sourdille P (2011) Variation in crossover rates across a 3-Mb contig of bread wheat (*Triticum aestivum*) reveals the presence of a meiotic recombination hotspot. *Chromosoma* 120:185-198.
- Sairam R, Saxena D (2000) Oxidative stress and antioxidants in wheat genotypes: possible mechanism of water stress tolerance. *J Agron Crop Sci* 184:55-61.
- Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. *Nat Rev Genet* 3:429-441.

- Salter PJ, Goode JE (1967) Crop responses to water at different stages of growth (Research review number 2). Commonwealth Agricultural Bureaux Farnham Royal UK
- Schluepmann H, van Dijken A, Aghdasi M, Wobbes B, Paul M, Smeekens S (2004) Trehalose mediated growth inhibition of Arabidopsis seedlings is due to trehalose-6-phosphate accumulation. *Plant Physiol* 135:879-890.
- Schnurbusch T, Collins N, Eastwood R, Sutton T, Jefferies S, Langridge P (2007) Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the Bo1 boron toxicity tolerance locus of bread wheat. *Theor Appl Genet* 115:451-461.
- Semagn K, Bjornstad A, Xu Y (2010) The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology* 13.
- Sharma A, Singh MB, Bhalla PL (2014) Cytochemistry of pollen development in *Brachypodium distachyon*. *Plant Systematics and Evolution*:1-10.
- Sharma R, Agarwal P, Ray S, Deveshwar P, Sharma P, Sharma N, et al. Kapoor S (2012) Expression dynamics of metabolic and regulatory components across stages of panicle and seed development in indica rice. *Funct Integr Genomics*.
- Shearman V, Sylvester-Bradley R, Scott R, Foulkes M (2005) Physiological processes associated with wheat yield progress in the UK. *Crop Science* 45:175-185.
- Sheoran IS, Saini HS (1996) Drought-induced male sterility in rice: Changes in carbohydrate levels and enzyme activities associated with the inhibition of starch accumulation in pollen. *Sexual Plant Reproduction* 9:161-169.
- Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Plant Biology* 3:217-223.
- Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. *J Exp Bot* 58:221-227.
- Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. *Plant Biology* 6:410-417.
- Shrestha R, Matteis L, Skofic M, Portugal A, McLaren G, Hyman G, Arnaud E (2012) Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. *Front Physiol* 3:326.
- Sibony M, Pinthus MJ (1987) Floret initiation and development in spring wheat (*Triticum aestivum* L.). *Annals of Botany* 61:473-479.
- Sinclair T, Jamieson P (2006) Grain number, wheat yield, and bottling beer: an analysis. *Field Crops Research* 98:60-67.
- Singh M, Hinkelmann K (1998) Analysis of partial diallel crosses in incomplete blocks. *Biometrical journal* 40:165-181.
- Singh RP, Hodson DP, Jin Y, Huerta-Espino J, Kinyua MG, Wanyera R, et al. Ward RW (2006) Current status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen. *CAB Rev Perspect Agric Vet Sci Nutr Nat Resour* 1:13.
- Slafer GA, Garaus JL (2007) Physiological traits for improving wheat yield under a wide range of conditions. In: Spiertz JHJ, Struik PC, van Laar HH (eds)

- Scale and complexity in plant systems research: gene plant-crop relations. Springer, pp. 147-156.
- Smith AB, Lim P, Cullis BR (2006) The design and analysis of multi-phase plant breeding experiments. *The Journal of Agricultural Science* 144:393-409.
- Snape JW, Semikhodskii A, Fish L, Sarma RN, Quarrie S, A., Galiba G, Sutka J (1997) Mapping frost tolerance loci in wheat and comparative mapping with other cereals. *Acta Agic Hung* 45.
- Solis MT, Chakrabarti N, Corredor E, Cortes-Eslava J, Rodrigues-Serrano M, Biggiogera M, et al. Testillano PS (2014) Epigenetic Changes Accompany Developmental Programmed Cell Death in Tapetum Cells. *Plant and cell physiology* 55:16-29.
- Sorensen AM, Krober S, Unte US, Huijser P, Dekker K, Saedler H (2003) The Arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. *Plant J* 33:413-423.
- Sourdille P, Singh S, Cadalen T, Brown-Guedira GL, Gay G, Qi L, et al. Bernard M (2004) Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). *Funct Integr Genomics* 4:12-25.
- Storme N, Geelen D (2013) The impact of environmental stress on male reproductive development in plants, biological processes and molecular mechanisms. *Plant, Cell & Environment*.
- Sturm A (1999) Invertases. Primary structures, functions, and roles in plant development and sucrose partitioning. *Plant Physiol* 121:1-8.
- Sturm A, Chrispeels MJ (1990) cDNA cloning of carrot extracellular beta-fructosidase and its expression in response to wounding and bacterial infection. *Plant Cell* 2:1107-1119.
- Suwabe K, Suzuki G, Takahashi H, Shiono K, Endo M, Yano K, et al. Fujioka T (2008) Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. *Plant and cell physiology* 49:1407-1416.
- Suzuki K, Takeda H, Tsukaguchi T, Egawa Y (2001) Ultrastructural study on degeneration of tapetum in anther of snap bean (*Phaseolus vulgaris* L.) under heat stress. *Sexual Plant Reproduction* 13:293-299.
- Takegami MH, Yoshioka M, Tanaka I, Ito M (1981) Characteristics of isolated microsporocytes from liliaceous plants for studies of the meiotic cell cycle in vitro. *Plant Cell Physiol* 22:1-10.
- Taylor ML, Hudson PJ, Rigg JM, Strandquist JN, Schwartz Green J, Thiemann TC, Osborn JM (2012) Tapetum structure and ontogeny in *Victoria* (Nymphaeaceae). *Grana* 51:107-118.
- Tenea GN, Bota AP, Raposo FC, Maquet A (2011) Reference genes for gene expression studies in wheat flag leaves grown under different farming conditions. *BMC Research Notes* 4:373.
- Thomson MF (1999) Plant cold acclimation: Freezing Tolerance Genes and Regulatory Mechanisms. *Annu Rev Plant Physiol Plant Mol Biol* 50:571-599.
- Timmermann M (2006) The Breeder eye - theoretical aspects of the breeders decision-making. *Proceedings of the COST SUSVAR workshop on cereal crop diversity: implications for production and products*. ITAB, Paris, France, pp. 118-123.

- Toroser D, Plaut Z, Huber SC (2000) Regulation of a plant SNF1-related protein kinase by glucose-6-phosphate. *Plant Physiol* 123:403-412.
- Tran F, Penniket C, Patel RV, Provart NJ, Laroche A, Rowland O, Robert LS (2013) Developmental transcriptional profiling reveals key insights into Triticeae reproductive development. *The plant Journal* 74:971-988.
- Trevaskis B (2010) Goldacre Paper: The central role of the VERNALIZATION1 gene in the vernalization response of cereals. *Functional Plant Biology* 37:479-487.
- Tsuda M, Takami S (1992) Changes of water potential in rice panicle under increasing drought stress at various stages. *Plant Physiology* 62:41-46.
- Tuberosa R, Gill BS, Quarrie SA (2002) Cereal genomics: ushering in a brave new world. *Plant Mol Biol* 48:445-449.
- Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. *Trends Plant Sci* 11:405-412.
- Tymowska-Lalanne Z, Kreis M (1998) The plant invertases: Physiology, Biochemistry, and Molecular Biology. *Advances in Botanical Research* 28:71-117.
- Udall JA, Flagel LE, Cheung F, Woodward AW, Hovav R, Rapp RA, et al. Wendel JF (2007) Spotted cotton oligonucleotide microarrays for gene expression analysis. *BMC Genomics* 8:81.
- Ueda K, Yoshimura F, Miyao A, Hirochika H, Nonomura K-I, Wabiko H (2013) COLLAPSED ABNORMAL POLLEN1 gene encoding the arabinokinase-like protein is involved in pollen development in rice. *Plant Physiology* 162:858-871.
- Umezawa T, Yoshida R, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K (2004) SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 101:17306-17311.
- UN (2011) World Population Prospects, the 2010 Revision. <http://esa.un.org/unpd/wpp/index.htm>. Cited 25/10/2011
- UnSang Y, JunHyun C, YouChun S, HangWon K, JaeKeun S (2011) Marker assisted selection of brown planthopper resistance and development of multi-resistance to insect and diseases in rice (*Oryza sativa* L.). *Korean Journal of Breeding Science* 43:413-421.
- van Doorn WG (2011) Classes of programmed cell death in plants, compared to those in animals. *Journal of Experimental Botany* 62:4749-4761.
- van Doorn WG, Woltering EJ (2005) Many ways to exit? Cell death categories in plants. *Trends in Plant Science* 10:117-122.
- Varnier A-L, Mazeyrat-Gourbeyre F, Sangwan RS, Clement C (2005) Programmed cell death progressively models the development of anther sporophytic tissues from the tapetum and is triggered in pollen grains during maturation. *Journal of structural biology* 152:118-128.
- Venuprasad R, Dalid DO, Del Valle M, Zhao D, Espiritu M, Sta Cruz MT, et al. Atlin GN (2009) Identification and characterization of large effect quantitative trait loci for GY under drought stress in rice using bulk segregant analysis. *Theoretical Applied Genetics* 120:177-190.
- Verhaest M, Lammens W, Le Roy K, De Coninck B, De Ranter CJ, Van Laere A, et al. Rabijns A (2006) X-ray diffraction structure of a cell-wall invertase

- from *Arabidopsis thaliana*. *Acta Crystallogr D Biol Crystallogr* 62:1555-1563.
- Vikram P, Swamy BM, Dixit S, Ahmed HU, Cruz MTS, Singh AK, Kumar A (2011) qDTY1. 1, a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. *BMC genetics* 12:89.
- Vonlanthen J, Okoniewski MJ, Meningatti M, Cattaneo E, Pellegrini-Ochsner D, Haider R, et al. Marra G (2014) A comprehensive look at transcription factor gene expression changes in colorectal adenomas. *BMC cancer* 14:46.
- Wang A, Xia Q, Xie W, Datla R, Selvaraj G (2003) The classical Ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development. *Proc Natl Acad Sci U S A* 100:14487-14492.
- Wang E, Xu X, Zhang L, Zhang H, Lin L, Wang Q, et al. He Z (2010) Duplication and independent selection of cell-wall invertase genes GIF1 and OsCIN1 during rice evolution and domestication. *BMC Evol Biol* 10:108.
- Wang L, Ruan YL (2012) New insights into roles of cell wall invertase in early seed development revealed by comprehensive spatial and temporal expression patterns of GhCWIN1 in cotton. *Plant Physiol* 160:777-787.
- Wang S, Wong D, Forrest K, Allen A, Huang E, Chao S, et al. Akhunov E (in press) Polyploid wheat genomic diversity revealed by the high-density 90,000 SNP array. *Plant Biotechnol J*.
- Webster H, Keeble G, Dell B, Fosu-Nyarko J, Mukai Y, Moolhuijzen P, et al. Appels R (2012) Genome-level identification of cell wall invertase genes in wheat for the study of drought tolerance. *Functional Plant Biology* 39:569-579.
- Westgate ME, Passioura J, B., Munns R (1996) Water status and ABA content of floral organs in drought stressed wheat. *Australian Journal of Plant Physiology* 23:763-772.
- Wiese A, Elzinga N, Wobbes B, Smeekens S (2004) A conserved upstream open reading frame mediates sucrose-induced repression of translation. *Plant Cell* 16:1717-1729.
- Wilhelm WW, McMaster GS (1996) Spikelet and floret naming scheme for grasses with a spike inflorescence. *Crop Science* 36:1071-1073.
- William HM, Trethowan R, Crosby-Galvan EM (2007) Wheat breeding assisted by markers: CIMMYT's experience. *Euphytica* 157:307-319.
- Williams E, Piepho HP, Whitaker D (2011) Augmented p-rep designs. *Biom J* 53:19-27.
- Wilson ZA, Zhang DB (2009) From *Arabidopsis* to rice: pathways in pollen development. *Journal of Experimental Botany* 60:1479-1492.
- Worland AJ, Law CN (1986) Genetic analysis of chromosome 2D of wheat. I. The location of genes affecting height, daylight insensitivity, hybrid dwarfism and yellow rust resistance. *Zeitschrift fur Pflanzensuchtung* 96:331-345.
- Wu H, Yang M (2005) Reduction in vacuolar volume in the tapetal cells coincides with conclusion of the tetrad stage in *Arabidopsis thaliana*. *Sexual Plant Reproduction* 18:173-178.
- Xiong L, Ishitani M (2006) Stress signal transduction: components, pathways, and network integration. In: Rai AK, Takabe T (eds) *Abiotic Stress tolerance in plants*. Springer Dordrecht, pp. 3-29.
- Xu Y (2010) *Molecular plant breeding*. CABI Oxfordshire United Kingdom

- Xue G-P, Drenth J, Glassop D, Kooiker M, McIntyre CL (2013) Dissecting the molecular basis of the contribution of source strength to high fructan accumulation in wheat. *Plant Molecular Biology* 81:71-92.
- Xue GP, McIntyre CL, Jenkins CL, Glassop D, van Herwaarden AF, Shorter R (2008) Molecular dissection of variation in carbohydrate metabolism related to water-soluble carbohydrate accumulation in stems of wheat. *Plant Physiol* 146:441-454.
- Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci* 10:88-94.
- Yang J, Zhang J (2006) Grain filling of cereals under soil drying down. *New Phytol* 169:223-236.
- Yang T, Poovaiah BW (2002) A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. *J Biol Chem* 277:45049-45058.
- Yoshida R, Hobo T, Ichimura K, Mizoguchi T, Takahashi F, Aronso J, et al. Shinozaki K (2002) ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in Arabidopsis. *Plant Cell Physiol* 43:1473-1483.
- Yousafzai FK, Al-Kaff N, Moore G (2010) The molecular features of chromosome pairing at meiosis: the polyploid challenge using wheat as a reference. *Funct Integr Genomics* 10:147-156.
- Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stage of cereals. *Weed Research* 14:415-421.
- Zhang C, Guinel FC, Moffatt BA (2002) A comparative ultrastructural study of pollen development in Arabidopsis thaliana ecotype Columbia and male-sterile mutant apt1-3. *Protoplasma* 219:59-71.
- Zhang D, Yang L (2014) Specification of tapetum and microsporocyte cells within the anther. *Current Opinion in Plant Biology* 17:49-55.
- Zhang DB, Wilson ZA (2009) Stamen specification and anther development in rice. *Chinese Science Bulletin* 54:2342-2353.
- Zhang J, Dell B, Biddulph B, Drake-Brockman F, Walker E, Khan N, et al. Appels R (2013) Wild-type alleles of Rht-B1 and Rht-D1 as independent determinants of thousand-grain weight and kernel number per spike in wheat. *Molecular Breeding*:1-13.
- Zhang J, Dell B, Biddulph B, Khan N, Xu Y, Luo H, Appels R (2014) Vernalization gene combination to maximize grain yield in bread wheat (*Triticum aestivum* L.) in diverse environments. *Euphytica* 198:439-454.
- Zhang J, Dell B, Conocono E, Waters I, Setter T, Appels R (2009(A)) Water deficits in wheat: fructan exohydrolase (1-FEH) mRNA expression and relationship to soluble carbohydrate concentrations in two varieties. *New Phytol* 181:843-850.
- Zhang J, Huang S, Fosu-Nyarko J, Dell B, McNeil M, Waters I, et al. Appels R (2008(A)) The genome structure of the 1-FEH genes in wheat (*Triticum aestivum* L.): new markers to track stem carbohydrates and grain filling QTLs in breeding. *Molecular Breeding* 22:339-351.
- Zhang W, Chao S, Manthey F, Chicaiza O, Brevis JC, Echenique V, Dubcovsky J (2008(B)) QTL analysis of pasta quality using a composite microsatellite and SNP map of durum wheat. *Theor Appl Genet* 117:1361-1377.

- Zhang Y, Primavesi LF, Jhurrea D, Andralojc PJ, Mitchell RA, Powers SJ, et al. Paul MJ (2009(B)) Inhibition of SNF1-related protein kinase1 activity and regulation of metabolic pathways by trehalose-6-phosphate. *Plant Physiol* 149:1860-1871.
- Zheng W, Chung LM, Zhao H (2011) Bias detection and correction in RNA-Sequencing data. *BMC Bioinformatics* 12:290.
- Zhou Z, Tian W, Liu H, Cao L (2012) Food consumption trends in china Australian Government Department of Agriculture, Fisheries and Forestry.
- Zinselmeier C, Lauer MJ, Boyer JS (1995(a)) Reversing drought-induced losses in grain yield: sucrose maintains embryo growth in maize. *Crop Science* 35:1390-1400.
- Zohary D, Hopf M (2000) *Domestication of Plants in the old world*. Oxford University Press Oxford

