TISSUE-SPECIFIC GENE EXPRESSION AND PROMOTER CHARACTERIZATION IN TRITICALE

CAROLYN RENEE PENNIKET

B. Sc. Biochemistry, University of Lethbridge, 2002

A Thesis

Submitted to the School of Graduate Studies

of the University of Lethbridge

in Partial Fulfillment of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY

Department of Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

ABSTRACT

Triticale (x *Triticosecale* Whitm.) is a cereal with favorable agronomic traits for a Canadian bioproduction platform crop. Appropriate tissue sampling times were determined and gene expression profiles were evaluated in five triticale seed tissues and eleven vegetative tissues using the Affymetrix Wheat GeneChip®. Genes that were expressed, not expressed, tissue-specific, tissue-enriched and developmentally regulated were identified. The percentage of probe sets on the wheat GeneChip with gene ontology annotations was improved from less than 3% to over 76% using homologous sequence identification and annotation transfer. This information was used to determine functions and processes over-represented within the identified gene lists and provide biological meaning to the results. Expression of candidate genes was further evaluated using qRT-PCR, RNA *in situ* hybridization and promoter characterization. This study has provided a comprehensive triticale gene expression atlas; knowledge regarding triticale development, gene function, expression and regulation; and tools enabling further triticale research and development.

ACKNOWLEDGMENTS

It is with immense gratitude that I acknowledge my thesis supervisor Dr. André Laroche. He has been a supervisor, mentor, coach, supporter and friend throughout this process. I can't thank him enough for all of his time and effort guiding me through my research and writing and making my PhD experience enjoyable. I will always respect his patience and the calm and gentle manner with which he approaches the encouragement of others, including myself, to challenge themselves and achieve success. I am also grateful to the guidance shown to me by my thesis co-supervisor Dr. Brent Selinger who provided critique, advice and positive encouragement when I needed it most. I would like to thank my committee members Dr. James Thomas and Dr. Hans-Joachim Wieden for their time, efforts and advice in meetings and the review of my thesis. It is greatly appreciated.

There are countless colleagues and individuals at Agriculture and Agri-Food Canada and the University of Lethbridge who have provided assistance in some form or other at one time or another. Thank-you to the School of Graduate Studies for their assistance throughout my program. Thank-you to the many individuals in the Biology department at the UofL and thank-you to so many staff members at AAFC-LRC for all their assistance and contributions. While I cannot name them all here I appreciate the contributions of each and every one. Great research is a team effort and I had a great team.

It is with gratitude that I acknoweledge my colleagues at AAFC-ECORC, Dr. Laurian Robert and Frances Tran for their valuable collaboration and for providing the data for the reproductive tissues. Conversations regarding data analysis and review of my results were beneficial and greatly appreciated.

I consider it an honor to have worked alongside Michele Frick for so many years. I have learned so much from Michele regarding technical expertise, lab management, research planning and biology in general that I feel my success is due in large part to the knowledge she has shared with me. She has also been a good friend and supporter throughout my graduate studies and her kindness is truly appreciated.

With deepest gratitude I thank Eric Amundsen for his assistance and support throughout my graduate work. I cannot thank him enough for the patience, care and support he has shown me. I cherish that you are proud of me.

My parents have played such an important role in forming the individual I am today that a large part of the credit for my success must go to them. They have taught me through example not to be afraid to take risks in life or in pursuing a career. I have watched them take risks in their own careers and watched them find success and happiness. They have also taught me not to give up when things get tough and to have pride in my achievements. I have learned so much through my education but it is these life lessons that are most important. I also feel that my parents have lived vicariously through me at times throughout my PhD, so to my parents - we did it!

Finally I would like to thank all of my friends and family for their unlimited support. I have had so many expressions of support throughout my graduate studies from so many friends and relatives both close and distant that I am truly grateful to have such a network behind me. An honorable mention goes to my Uncle Arlan for giving me a start in research and for also challenging me to take risks and to have a sense of humor.

I hope I have made my grandmother proud.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

All supplementary tables are available in electronic format only and are provided in the accompanying CD or digital drive.

LIST OF FIGURES

LIST OF ABBREVIATIONS

˚C	degree Celsius
µg	microgram
µl	microlitre
µM	micromolar
µmole	micro mole
A	Absent
A260/280	Absorbance at 260 nm over 280 nm
ABA	Abscisic acid
ABRE	ABA-responsive element
ABRE	Abscisic acid responsive element
AGP	ADP-glucose pyrophosphorylase
ANOVA	Analysis of variance
anti-DIG AP	anti-Digoxygenin antibody fragments linked to Alkaline Phosphatase
attB	Site-specific attachment sites
B2G	Blast2GO
BAC	Bacterial artificial chromosome
BCP	Bicellular stage of pollen development
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BP	Biological process
bZIP	Basic leucine zipper
C	Constitutive expression
CAB	Chlorophyll a/b binding protein

CaMV	Cauliflower mosaic virus
CC	Cellular component
cDNA	Complementary DNA
cds	coding sequence
CEL	Probe cell intensity
cm	centimetre
cRNA	Complementary RNA
Ct	Threshold cycle
CV	Coefficient of variation
cv.	Cultivar
DAG	Directed acyclical graph
dbEST	EST database
DDGS	Distiller's dry grains and soluble
DIG	Digoxygenin
DNA	Deoxyribonucleic acid
DPA	Days post-anthesis
DRE	Dehydration-responsive element
DRG	Developmentally regulated gene
ds cDNA	Double stranded cDNA
DTT	Dithiothreitol
EC	Evidence codes
ECORC	Eastern Cereal and Oilseeds Research Center
ELP	Expression-level polymorphism
Em	Early-methonine-labelled
ESP	Endosperm-specificity palindrome

EST	Expressed sequence tag
e-value	Expectation value
FDR	False discovery rate
G/R	Ratio of green to red
GA	Gibberellic acid
Galβ(1,3)GalNAc	β-galactose-1,3-N-acetylgalactosamine
gbss1	Granule-bound starch synthase 1
GFP	Green fluorescent protein
GO	Gene Ontology
GS	Glutamine synthetase
h	hour
HCL	Hierarchical clustering ()
HMG2	Tomato 3-hydroxy-3-methylglutaryl coenzyme A gene 2
IDA	Inferred from direct assay
IEA	Inferred from electronic annotation
ISS	Inferred from sequence or structural similarity
kb	kilo base pairs
LB	Luria Broth
LCM	Laser capture micro-dissection
LMW	Low molecular weight
$\log_2$	Logarithmic scale to base 2
Lox1	Lipoxygenase 1
m	metre
M	Marginal
MeJA	Methyl jasmonate

MeV	Multiple Experiment Viewer
MF	Molecular function
min	minute
ml	millilitre
mM	millimolar
MPG	Mature pollen grain stage of development
mRNA	Messenger RNA
MS	Murashige and Skoog nutrient mix
N_2	Nitrogen
NBT/BCIP	Nitro blue tetrazolium chloride/5-Bromo-4-chloro-3-indolyl phosphate
NGS	Next generation sequencing
nm	nanometre
nM	nanomolar
nr	Non redundant protein database
P	Present
PCA	Principal component analysis
PCR	Polymerase chain reaction
PHA	Polyhydroxyalkanoate
pmol	picomole
PPAP	Plant Proteome Annotation Program
p-value	Probability value
qRT-PCR	Quantitative real time-PCR
rbcS	Small subunit of ribulose-1,5-bisphosphate carboxylase
RCA	Reviewed computational analysis
RMA	Robust multiple-array average

RNA	Ribonucleic acid
rpm	rotations per minute
S	Expression specific to a particular tissue or tissue group
s	second
SAG	Senescence associated gene
SC	Expression specific to and constitutively expressed in a group of tissues
sec	seconds
SEF4	Soybean Embryo Factor 4
SEM	Scanning electron microscopy
SFP	Single feature polymorphism
SOTA	Self organizing tree algorithm
ssRNA	Single stranded RNA
TAIR	The Arabidopsis Information Resource
TCP	Tricellular stage of pollen development
TET	Tetrad stage of pollen development
TritFLcDNA	Triticale (x *Triticosecale* Wittm.) full length cDNA
TSS	Transcription start site
T-triticain-γ	Triticale homologue of wheat triticain γ
T-WGA	Triticale homologue of wheat germ agglutinin
T-ω-SEC	Triticale homologue of rye ω-secalin
U	unit
UCDNA	University of Calgary, University Core DNA Services
UNM	Uninucleate stage of pollen development
UV	Ultra violet
v/v	volume per volume

WGA	Wheat germ agglutinin (WGA)
WGA-A	WGA isolectin A
WGA-B	WGA isolectin B
WGA-D	WGA isolectin D
wheat GeneChip	Affymetrix GeneChip® Wheat Genome Array
WUN-motif	Wounding responsive element
Z	Zadoks'

EXECUTIVE SUMMARY

Triticale (x *Triticosecale* Whitm.) is a cereal with agronomic traits that make it an ideal candidate for a bioproduction platform crop in Canada. To succeed as a bioproduction platform, genetic tools and knowledge concerning triticale development, gene function, and gene expression timing, location, level and regulation are required. The goals of this study were to provide a pool of knowledge regarding triticale development, gene function, gene expression and regulation and to provide some tools for transgene expression in triticale to enable it as a bioproduction platform crop.

The linear-linear segmented regression temporal profile of leaf senescence was determined for flag and third leaves by measuring relative changes in chlorophyll content under controlled growth conditions (see Chapter 2). This profile was used to determine sampling times for tissue collection in gene expression studies.

The Affymetrix GeneChip® Wheat Genome Array was the largest available resource for profiling gene expression in triticale at the time of this experiment. While this array contains probes for a majority of triticale transcripts, less than 3% of the probe-sets had gene ontology annotations making functional classification of genes identified using the array difficult. The probe set target and consensus sequences were used for a series of homology searches against public databases and assembled cDNA libraries to identify functional information for probe sets on the array (see Chapter 3). By combining the results from these methods the percentage of probe sets on the wheat array with functional gene ontology annotations was improved to over 76%.

Using the Wheat Genome array, gene expression profiles were determined in triticale seed tissues (embryo, endosperm, crease, pericarp and epiderm) and triticale

vegetative tissues (leaf tissue at 5 developmental stages, stem tissue at 4 developmental stages, and early root and coleoptile tissues) and compared to expression in reproductive tissues (see Chapter 4). Results of the gene expression analysis identified genes on the wheat array that are expressed in triticale, that are not expressed in triticale, that are specifically expressed in each of the seed and vegetative tissues, genes whose expression was enriched in each tissue, and genes with expression that was developmentally regulated in stem and leaf tissues. The functional annotation information was used to determine molecular function and biological process categories that are over-represented within the identified gene lists allowing the identification of tissue-specific transcription factors and the biological processes and molecular functions involved in developmental processes such as leaf senescence. This study has provided the most comprehensive gene expression atlas in triticale to date.

A method for *in situ* hybridization of RNA transcripts in triticale seeds was developed and used for further evaluation of the expression patterns of candidate seed-specific transcripts identified by the gene expression analysis (see Chapter 5). Transcript expression patterns were evaluated at the cellular level and throughout an increased number of developmental stages in triticale seeds. An embryo-specific wheat germ agglutinin transcript was localized to the root epidermis and root cap cells within the embryo and an embryo-specific glutamine synthetase was localised to the procambial cells within the embryo and scutellum. The expression of a crease-specific protein kinase was restricted to the vascular cells in the crease at later stages of seed development. A putative crease-specific ubiquitin conjugating enzyme was found to be expressed in regions of the endosperm near the time of tissue sampling for the microarray analysis.

Promoter sequences for three triticale gene homologues of genes previously identified as tissue-specific in other cereals were sequenced and partially characterized to identify regulatory elements responsible for their expression patterns in triticale (see Chapter 6). *Cis*-acting sequence elements in the promoters were electronically identified using PlantCARE and the locations of putative seed-specific elements were used to design promoter deletion fragments. Promoter deletion fragments were used in transient expression studies in leaf and seed tissues to determine the effects of the putative elements on transgene expression. Different patterns of expression were identified using promoter deletion fragments of different lengths. Fragments were identified that were capable of expression in both seed and leaf tissues, throughout the seed but not in leaf, in only the embryo, embryo and endosperm and in endosperm and seed coat. Several promoter elements suspected to be involved in the regulation of the triticale homologues of wheat germ agglutinin, tritcain-γ and ω-secalin based on the transient expression results were also identified for further evaluation. The promoters identified are novel triticale sequences not encompassed by existing patents.

The temporal profile of senescence allowed the determination of appropriate tissue sampling times to capture molecular events important to the developmental process senescence (Chapter 2). Sampling of various tissues and developmental stages enabled the creation of a gene expression atlas for triticale (Chapter 4). Improvement of the GO information for the wheat GeneChip (Chapter 3) allowed functional analysis and biological interpretation of the gene expression results in Chapter 4. *In situ* hybridization of RNA transcripts in triticale seeds (chapter 5) allowed the further evaluation of the expression patterns of candidate seed-specific transcripts identified by the gene

expression analysis and improved the spatial resolution of the transcript expression. Characterization of promoters from candidate seed-specific transcripts using genome walking, promoter deletion fragment generation and transient expression analysis (Chapter 6) has allowed the identification of putative promoter elements involved in seed-specific expression and promoter fragments that may be suitable candidates for stable transgene expression.

The research presented provides a comprehensive expression atlas of triticale and the most complete GO annotation set for the wheat GeneGhip to date, resources that will greatly enhance future research efforts in triticale, wheat, other cereals and even non-cereal plants. The temporal profile of senescence, *in situ* hybridization and transient expression methods are techniques that could be applied to an increased number of tissues, transcripts and promoters in triticale and other plants. These methods have demonstrated the ability to improve upon the knowledge of gene expression and regulation for candidate genes identified in microarray studies using tissue-type rather than cell type samples. The promoter analysis has provided a number of sequences to direct transgene expression for research or bioproduction purposes in triticale or other cereals without infringing on existing patents. Together this research has vastly improved the volume of existing knowledge regarding triticale development, gene function, gene expression and regulation and has provided resources and tools for future research and applications in triticale, cereals and plants.

CHAPTER 1 The knowledge required to support bioindustrial triticale

1.1 LITERATURE REVIEW

1.1.1 Plants and the bio-industry

A steady decline in profit margins for farmers over the last twenty years (Statistics Canada, December 19, 2007) has led to the demand for new and innovative initiatives in agriculture that will add value to current operations in both western Canada and worldwide. Increasing profit margins for producers can parallel the societal demands associated with renewable resources. A regulation requiring five percent average renewable content in gasoline by December 2010 was adopted by the Government of Canada in their Renewable Fuels Strategy (Environment Canada, September 16, 2012), and the ecoENERGY for Biofuels program which supports the production of renewable alternatives to gasoline and diesel and encourages the development of a competitive domestic industry for renewable fuels will run from April 1, 2008 to March 31, 2017 (Natural Resources Canada, September 16, 2012). The bio-industry is emerging as an opportunity to meet these demands that will allow farmers to increase the value of their operation by using traditional crop plants to produce high value-added products.

Plants naturally produce a variety of compounds that can be used for commercial and industrial purposes, the major classes being proteins, carbohydrates and lipids (Goddijn and Pen, 1995). In addition to the compounds naturally produced in plants, there is an indefinite potential to increase plant value through genetic engineering. High yields of valuable products can be produced economically within different renewable

plant systems by altering plant metabolism, over-expressing plant metabolites or expressing novel heterologous compounds.

The production of a variety of compounds such as platform chemicals, bio-plastics, enzymes, prebiotics, emulsifiers, ethanol, adhesives and bio-diesel can be achieved with plant systems. There are several advantages to using plants to produce valuable products or their precursors. Extensive infrastructure already exists for large scale production of plant biomass for the production of desired bio-products. Another primary advantage of plant production is the lower production costs relative to other more expensive bacterial or animal transgenic systems (Ma et al., 2003; Goddijn and Pen, 1995; Takaiwa, 2005). Using crop species has the advantage that the agronomics are well known, they generally are not invasive and they do not usually have wild or weedy relatives in Canada (Macdonald, 2004). Recombinant proteins produced in plants can be stored in plant storage organs such as seeds, leaves or stems which can then be easily harvested for the extraction of these proteins. Proteins can also be highly and stably accumulated in a tissue-specific manner when expressed under the control of appropriate promoters (Takaiwa, 2005). Under tissue-directed expression, some products such as oleochemicals may be harvested directly from seeds reducing the need for expensive refining methods from fossil fuels. The stable integration of transgenes for the production of heterologous proteins in plants has the advantage that identical offspring can be produced by self-fertilization, resulting in stable inheritance and a renewable source for the desired bio-product (Goddijn and Pen, 1995).

1.1.2 Plant potential for bio-industry

1.1.2.1 Yield and seed size

Achieving higher yielding crops is a major goal of agricultural research. One of the most basic ways to increase the value of a crop is to increase its yield. Increasing yield is desirable for both traditional crop plants and genetically engineered crops used to produce novel compounds. Crop or plant total yield is the sum of all shoot and root tissues. In many crops only a portion of the plant is edible or harvested and in these crops, increasing the crop yield index (the ratio of harvested to total yield) is as important if not more important than increasing the total yield (Gong et al., 2005). In cereals, this involves the allocation and partitioning of photosynthates to the grain, the predominant sink tissue, thereby increasing the yield and crop productivity (Taiz and Zeiger, 2002). The cereal grain is a dry one seeded fruit known as a caryopsis, in which the wall of the seed becomes fused to the seed coat during its development (Esau, 1977). The caryopsis is the most important source of carbohydrates in the human diet making cereals the most agriculturally and economically important crops. An understanding of the processes involved in the development of the caryopsis is therefore of great significance to agriculture and world economics (Drea et al., 2005). Understanding the development of the wheat grain involves understanding the transport of photosynthates to the sink tissues. Achieving a maximum grain to straw ratio (maximum crop yield index) results from efficient usage of nutrients and resources, and minimal waste left in unused straw. An understanding of the processes involved in assimilate partitioning helps plant breeders select and develop varieties with improved transport to the edible or usable portions of the plant (Taiz and Zeiger, 2002). Many plant species have an excess of photosynthetic

capacity (Smidansky et al., 2003). Altering the demand for photosynthates in sink tissues can take advantage of this excess to increase yields.

The size of the seed is an important attribute if products are to be extracted from seed tissues. Endosperm cell proliferation during early seed development is a vital component of determining final seed size (Garcia et al., 2005). Promoters of genes active in early seed development have been evaluated for the modification of seed size (Tiwari et al., 2006).

1.1.2.2 Starch biosynthesis

Starch is one of the most important plant products to man and is composed of two major components, amylose and amylopectin, that assemble to form starch granules (Burrell, 2003). The major sources of starch are cereal crops, including wheat, rice and maize, and the root crop, potato. Starches have a wide range of industrial applications in addition to their importance to the human diet. These diverse applications include additives in cement, drilling, paper industry, biodegradable plastics, textiles and pharmaceutical compounds (Burrell, 2003). Some lines of maize and rice contain waxy starches that are composed almost entirely of amylopectin with little or no amylose and have economically valuable functional qualities (Slade et al., 2005). Potential uses for waxy starches include the food, paper and adhesives industries (Burrell, 2003). Plant transformation of non waxy cereal crops such as wheat or triticale to produce waxy starches has the potential to increase the value of these crops for industrial uses. Indeed, the starch biosynthesis pathways may provide the most opportunity for altering seed content and improving crop quality and/or value through genetic engineering.

The cereal grain consists of the embryo and the carbohydrate-rich endosperm surrounded by the fused remnants of the integuments and the inner layer of the ovary (the pericarp, Esau, 1977). It is the endosperm that contains the nutrients required for germination. These nutrients are the storage compounds accumulated during seed development. In the endosperm, starch biosynthesis pathways convert sucrose transported from photosynthetic tissues into starch, the main storage compound. In cereals, the endosperm consists mainly of starch (60%-70%) followed by storage proteins (approximately 15%, van Dongen et al., 2004). The starch stored in the endosperm has many commercial uses, either in its natural or modified forms (Goddijn and Pen, 1995). For example, the main conventional feedstocks for the fermentation industry are molasses, sugar and starch derivatives (Goddijn and Pen, 1995). High levels of enzymes used in the breakdown of sucrose (such as sucrose synthase) in the storage tissue (endosperm) during the grain filling period, increase the amount of photosynthates accumulated in the grain (Yang et al., 2004). Altering the processes involved in the conversion of sucrose to starch can increase the yield of starch or allow the accumulation of novel carbohydrates and consequently increase the crop value. Increasing the endosperm activity of ADP-glucose pyrophosphorylase (AGP), a key enzyme in starch biosynthesis, has been shown to increase plant yield in wheat and rice (Smidansky et al., 2003). This was accomplished by expressing a modified maize AGP large subunit sequence (*Sh2r6hs*) under the control of an endosperm-specific promoter. Seed weight per plant was increased 31% and 23% in wheat and rice respectively (Smidansky et al., 2003). The activity of other carbohydrate metabolism enzymes has been altered in dicot

species as well, demonstrating that yield can be manipulated in plants by modification of key carbohydrate metabolism enzymes (Smidansky et al., 2003).

1.1.2.3 Heterologous proteins

Plants are attractive systems for the production of heterologous proteins. They can produce large amounts of specific proteins at a relatively low cost when compared to alternative systems (Ma et al., 2003). The expression of novel proteins relies on the stable integration of the transgene in the plant genome. This results in the inheritance of the trait and ultimately a renewable source of the valued protein. The expression of heterologous proteins in plants has been explored as a direct delivery system for bioactive peptides, vaccines and heterologous plant proteins that improve nutritional quality and human health (Ma et al., 2003; Takaiwa, 2005; Ko et al., 2008; Gomord et al., 2010).

1.1.2.4 Bioplastics

More than 265 million tonnes of plastics are produced worldwide every year and are utilized in almost every industry (PlasticsEurope, 2011). Disposal of petroleum based plastics has become a large problem and biodegradable plastics are emerging as one solution. Biodegradable plastics are made from polyhydroxyalkanoate (PHA) polymers naturally synthesized in many bacterial species and accumulated at high levels during fermentation (Goddijn and Pen, 1995). Plants can be engineered to produce these polymers from acetyl-CoA in a manner similar to fatty acid and lipid biosynthesis by diverting acetyl-CoA to PHA production (Reddy et al., 2003). Starch-producing crops have an advantage over oil seed crops in terms of yield, but the diversion of acetyl-CoA

towards PHA synthesis is more complex making manipulation more difficult (Reddy et al., 2003). More research on the metabolic pathways involved is required before PHAs can be produced in starch plants for production of biodegradable plastics. Production of PHAs on an agronomic scale would allow synthesis at a level that would be more competitive with petroleum based plastics and cheaper to produce in comparison to fermentation (Reddy et al., 2003).

While the biosynthesis of PHAs for bioplastic production may be difficult in the cereals, numerous other plant proteins have also been used for the production of bioplastics including wheat gluten. Protein-based biomaterials are advantageous in terms of biodegradability and are among the polymers with fast rates of degradation (Domenek et al., 2004). Wheat gluten is plasticized with the addition of glycerol in conjunction with heat treatments (Domenek et al., 2004). An added advantage to using wheat gluten is the ability to generate bioplastics with differing properties dependent on the processing without diminishing its biodegradability (Domenek et al., 2004; Jerez et al., 2007).

The production of these industrial or commercial products will benefit from research focused on plant growth and development, in particular seed development. Promoters of genes expressed at high levels within the seed such as storage proteins are of particular interest for use in genetic engineering. Also, promoters of genes of unknown function, whose expression is limited to, or enhanced in the seed, have potential for use in the expression of economic traits in transgenic plants (Girke et al., 2000).

1.1.3 Triticale: a potential solution to the demand for a bio-industry in Canada

Hexaploid triticale (X *Triticosecale* Wittm.), is a man made plant that resulted from a cross between diploid rye (*Secale cereale*, 2n=14) and tetraploid durum wheat (*Triticum turgidum*, 2n=28). The resulting hybrid contains one set of rye chromosomes (R) and two sets of wheat chromosomes (AB) to form the hexaploid (2n=42) genome AABBRR (Kuleung et al., 2004). Octoploid triticales (2n=56=AABBDDRR), derived from hexaploid wheat varieties (*Triticum aestivum*, 2n=42=AABBDD), were the first triticales to be produced. However, for the most part, octoploid cultivars were not successful and the majority of triticale grown worldwide today is hexaploid varieties (Ammar et al., 2004).

Triticale has favorable qualities of both wheat and rye. It is a hardy, stress tolerant, disease resistant and high yielding plant that is ideally suited for use as a bio-industrial platform in Canadian environments. Triticale is a highly productive crop for both grain and plant biomass. Spring triticale is 5 to 20 percent higher yielding than Canadian spring wheat and winter triticale varieties are 10 to 20 percent higher yielding with similar winter hardiness to Canadian winter wheats (Salmon, 2004). Triticale has increased disease resistance to most of the economically important cereal crop diseases (Horlein and Valentine, 1995) and has few insect pests (Salmon, 2004). Triticale is adaptable to a variety of environments and outperforms wheat on marginal lands (Giunta et al., 1999; Rubio et al., 2004). The adaptability to these environmental conditions is ideal for growth in many Canadian environments. It also allows the potential for triticale to be grown on sub-optimal farmland, thus increasing the value of land used for the production of high value-added products and limiting the removal of optimal farmland

from crop production for human food consumption. This in turn will help to limit public opposition to the use of plants for purposes other than food production. Triticale in Canada is currently grown for forage, as an alternative to barley silage and to a lesser extent as feed grain for livestock (Salmon, 2004). Triticale is not currently grown in Canada for human consumption, an advantage in terms of a crop platform for genetic engineering of novel bio-products.

Triticale plants and grains are visually distinguishable from other cereal crops, which has the advantage of limiting the potential for seed mixing at grain elevators and intercrop mixing in the field. Differentiation between transgenic crops and non-transgenic crops is extremely important for economic success and public acceptance, as demonstrated by losses incurred by the StarLink™ corn fiasco (Schmitz et al., 2005). Triticale is predominantly self-fertile and does not hybridise with rye, wheat or other cereals under natural conditions (Hills et al., 2007), which also limits the potential of releasing transgenic material into the environment. Color traits such as the blue aleurone trait in wheat could also be included in triticale as tools for the detection of transgenic varieties and to monitor gene flow to other cultivars (Hanson et al., 2005; Matus-Cadiz et al., 2007). Enhancing color differences between transgenic and non-transgenic varieties even has the potential for detection of contaminated grain using visual or automated imaging techniques thus limiting the entry of transgenic varieties into food chains or nontransgenic grain supplies.

The above properties make triticale an attractive model for bio-industrial technology and value-added farming of non-food products. Triticale represents less than 0.3% of the cultivated cropland worldwide and is used mainly as an animal feed source

(FAO, 2010). Thus, using triticale for purposes other than human consumption, that are more profitable, will have minimal impacts on human food markets. The lack of use for human consumption and limited current production also contribute to the risk limiting factors and the economic potential when evaluating the use of triticale as a bio-industrial platform.

Triticale also has the advantage of being physiologically similar in many ways to a number of cereal crops (e.g., wheat and barley) that are widely grown and harvested in Canada. This is important for the harvest and extraction of plant products. Products targeted to the grain can be easily harvested using existing farm equipment and separated using existing milling technologies. Remaining plant biomass can also be harvested using existing farm equipment for downstream production. This limits the need for new capital equipment for current producers of wheat, barley, oats and rye, thus reducing the input costs and increasing the economic benefits of growing triticale as a high-value added crop.

Economic analyses have demonstrated that wheat can compete with conventional fermentation feedstocks, such as starch and sugar, in terms of profitability and that a two fold increase in world-wheat reserves could lead to the production of many chemical products via bio-processing without reducing wheat stocks for human consumption (Koutinas et al., 2004). The same could be achieved with triticale but with the added advantage of not having to increase wheat reserves and without threatening the food supply through genetic engineering of wheat. Triticale can be provided to the fermentation industry as a feedstock at a reduced cost compared to wheat and corn but is higher yielding than wheat allowing profit margins for growers that are equivalent to

corn, currently the most cost efficient option for ethanol and fermentation chemical production in North America (Bergstra, 2006). Ethanol production from triticale also has the added advantage of higher value of distiller's dry grains and soluble (DDGS) byproduct as a feed source which partially offsets the lower processing costs for ethanol production from corn (Bergstra, 2006).

1.1.4 The need for tissue-specific promoters

For economic reasons, the bio-industry is concerned with localization, diversity and stability of bio-products, fractionation potential, yield, market and consumer acceptance. Regulatory and government agencies are concerned with environmental and health impacts of new bio-products. To achieve maximal extraction efficiencies and yields and minimize bio-safety risks, it is necessary to characterize and select novel developmental- and tissue-specific promoters to target molecules to the specific tissues intended for industrial end-uses.

The requirement for promoter novelty is due to the presence of patents on many existing available promoters and the need to be free of these cost prohibitive patents in order to produce a commercially available product in plant systems. Many of the tissue-specific promoters characterized from plants are patented, and while it is possible to license existing technology, the added expense limits their use as a means to increase agricultural profit margins (Sechley and Schroeder, 2002; Dunwell, 2005).

Plant genetic engineering has become a very controversial topic and despite the potential for socioeconomic and environmental benefits, the safety for both the environment and the public has been brought into question (Potenza et al., 2004).

Concerns include possible antigenic effects of transgene products, transfer of transgenes to related wild or cultivated species through pollen drift, and toxic or unintentional effects on non-target organisms (Potenza et al., 2004). In order to minimize the potential risks of genetically engineered plants, the appropriate level of gene expression must be established in both a spatial and temporal manner within the plant. Careful selection of promoters to target transgene expression can ensure the necessary level of spatial and temporal control of gene expression.

Most of the characterized tissue-specific promoters in the databases are currently from dicot species (Potenza et al., 2004). Using endogenous promoters identified and characterized from triticale, or other closely related monocot species, such as wheat and rye, will be more acceptable than promoters from unrelated dicot species in terms of public acceptance and approval by regulatory agencies. Endogenous promoters will also likely improve functionality in comparison to using previously characterized promoters from dicot species.

Much of the control of gene expression is regulated at the level of transcription (Baginsky et al., 2010). Therefore, strong promoters conferring high levels of gene expression are often required for biotechnology purposes. Directed high levels of tissue and developmental stage-specific gene expression requires several factors, including the appropriate promoters, enhancers and leader sequences, optimization of codon usage, the absence of mRNA-destabilizing sequences and the presence of the necessary transcription factors in the targeted tissue/cells (Goddijn and Pen, 1995). Although there is a large fraction of conserved sequences among eukaryotes, especially in genes coding for core cellular functions, regulatory sequences can be quite diverse (Clarke et al.,

2003). For example, among the promoters of the rice glutelin multigene family, there is little sequence similarity other than a few conserved motifs involved in endosperm-specific expression (Yoshihara and Takaiwa, 1996). Even well characterized heterologous plant promoters may show different patterns of expression when transformed into a different species, or even a different cultivar of the same species (Potenza et al., 2004).

Promoters from the highly characterized gene family encoding the ribulose-1,5-bisphosphate carboxylase small subunit (*rbcS*) show a light inducible and tissue-specific expression pattern that is highly conserved when transformed among various dicot species (Schaffner and Sheen, 1991). However, *rbcS* promoters show a more variable pattern of expression in monocots, even when a monocot *rbcS* promoter from rice (a C3 plant) was transformed into maize (a C4 plant), another monocot (Nomura et al., 2000). Despite being one of the best-characterized plant tissue-specific promoters, variable activity of this promoter in heterologous systems highlights the need to characterize novel tissue-specific homologous promoters (Potenza et al., 2004).

A lack of homogeneity among cereal genes currently known to be expressed in the grains from plants of different species but of similar function, may require that novel endogenous promoters be used for transgene expression in cereal grains. Comparison of wheat genes expressed during grain development to the rice genome showed the largest classes of genes with no rice homologue were those expressed in the endosperm (Drea et al., 2005). Even the predominant storage proteins in cereals, which are the gliadins (prolamins) in wheat and glutelins (globulins) in rice, have a significant level of variation in promoter control of gene expression (Shewry and Halford, 2002). Only two of six

prolamin promoters from barley, oat, rice and wheat maintained their endosperm specificity when transiently expressed in both barley and wheat (Vickers et al., 2006). Purolindoline gene promoters in wheat direct expression exclusively in the starchy endosperm cells and not the aleurone cells but lose specificity in other cereals, despite the fact that over 1000 bp of promoter sequences have been used for transformation (Wiley et al., 2007). Even for cereal endosperms in which many promoters have been identified and widely studied, few of these promoters produce suitable amounts of recombinant proteins and in the spatial and temporal manner desired (Lamacchia et al., 2001).

Some promoters thought to be tissue-specific are expressed in other tissues. For example, promoters of various seed protein genes are expressed not only in developing seeds, but also in the pollen of dicot species (Zakharov et al., 2004). The expression in pollen has implications for the acceptability of the use of such promoters. If the transgene expressed has negative impacts on pollen development it could result in pollen sterility and would be unsuitable for biotechnological production in seeds. On the other hand, if the transgene is expressed in pollen and does not cause sterility, it may impact the ability to contain the products of biotechnology, a condition that is certain to be required in plant molecular farming (Macdonald, 2004).

Constitutive promoters are often used for over-expression of a gene of interest to determine functional activity of the encoded protein. These promoters often drive high levels of expression in all plant tissues and would therefore seem favorable for the production of a valuable bio-product in large volumes. There are, however, several drawbacks to using constitutive promoters. Some of the most commonly used constitutive promoters, such as the cauliflower mosaic virus (CaMV) 35S promoter,

originate from plant virus sources (Odell et al., 1985). These may be publicly unacceptable due to the perception of risk to human health from sequences derived from infective species, even if the species themselves do not pose a threat to human health as in the case of a plant virus. Constitutive expression of a transgene means that the expressed protein will be in parts of the plant that may be left in the environment even after harvest (i.e. root tissue). This may be unacceptable to public and regulatory agencies due to the potential perceived environmental threats (Canadian Biotechnology Advisory Committee, 2002). Constitutive over-expression of a transgene where and/or when it is not normally expressed using endogenous plant promoters from actin and ubiquitin genes has also been shown in some cases to have detrimental effects on plant development (Potenza et al., 2004). This may be counterproductive to reaching the goal of producing large volumes of a valuable bio-product. This again demonstrates the critical importance of using tissue-specific promoters to direct expression and maintain optimum plant performance.

Ultimately, it is important that the promoter used to drive expression of the desired bio-product is stably transformed. Transient expression often does not result in the same level of specificity as stable expression (Fujiwara and Beachy, 1994; Kluth et al., 2002). This may be due to; 1) incorrect recognition of *cis*-acting elements, 2) lack of correct post-transcriptional controls in heterologous species, 3) lack of chromosomal integration and resultant control mechanisms, 4) incorrect methylation patterns, or 5) vector sequence interference (Vickers et al., 2006).

1.1.5 Promoter structure and complexity

A promoter is a regulatory region of a gene located upstream of the coding sequence. Promoters contain specific sequences recognized by transcription factor proteins to signal transcription initiation (Griffiths et al., 1996). They generally consist of a core promoter approximately 40 base pairs (bp) upstream (-40) of the transcription start site and a set of transcription control sequences called promoter proximal elements that may extend up to 200 bp upstream (Nikolov and Burley, 1997). As well as the core and upstream promoter region required for initiation of transcription, there may be distal enhancer elements found even further from the transcription start site in either direction. Some elements are located infrequently up to 2 kilo base pairs (kb) upstream of the transcription start site and these DNA sequence motifs may also be involved in the regulation of transcription (Nikolov and Burley, 1997). These motifs are referred to as *cis*-elements, regulatory elements, enhancer elements or enhancers (Potenza et al., 2004). The type, number, position and combination of these *cis*-elements can regulate gene expression in a tissue-specific, stress- or pathogen-related, substrate-related, light- or developmentally-regulated manner (Potenza et al., 2004). These *cis*-elements can be involved with activation or silencing of gene expression under associated conditions. The activation or silencing of the specific *cis*-elements is dependent on the presence, absence or activity of their associated binding proteins (Potenza et al., 2004).

Eukaryotic RNA polymerase II cannot recognize target promoters directly. It is thought that transcriptional activators bind to proximal or distal elements, causing chromatin remodeling, thus allowing the recognition of the TATA element (Nikolov and Burley, 1997). The interaction of transcription factors with specific *cis*-elements

regulates the rates and profiles of transcription initiation and elongation (Abebe et al., 2006). The interaction of various transcription factors, transcriptional activators and co-activators with distal, proximal and core promoter elements ensures strict control of transcription and may make the use of heterologous promoters more problematic as they may not satisfy the interaction requirements in the targeted tissue. While some transcription factors are expressed constitutively, others are expressed in a tissue-, developmental- or environmental-specific manner (Chen et al., 2002). Developmentally regulated or tissue-specific promoter activity is therefore dependent on the pattern of expression of endogenous trans-acting factors in a transformed plant.

1.1.6 Promoters and promoter elements conferring tissue-specific expression patterns

The pattern and level of gene expression is determined by the presence, number and combination of *cis*-acting elements in the promoter region and the presence or absence of factors that interact with these elements (Dynan, 1989). Several promoters and promoter elements have been identified for targeting transgene expression in a tissue-specific/enhanced, cell type-specific, organelle-specific and developmentally regulated manner or in response to environmental conditions (reviewed by Potenza et al., 2004).

Confining transgene expression to specific tissues through targeted expression is more likely to be acceptable to both the public and regulatory agencies (Canadian Biotechnology Advisory Committee, 2002). Tissue-specific promoters have been identified in plants for fruit, seed/grain, tubers/roots, flowers, pistils, anther/pollen and leaf/green tissue (reviewed in Potenza et al., 2004). The following literature review will

focus on promoters specific to the seed/grain and leaf/green tissue with an emphasis on promoters and promoter elements in monocot species.

Localizing expression of novel products in seed tissues would be desirable as it allows the bio-product to be easily harvested, handled, and stored using existing farm infrastructure thus making production more economically feasible for farmers. The seed also remains in the environment for a relatively short period of time when harvested, reducing the potential environmental impact of bio-product production. Seed-specific transgene expression has been used to improve seed nutrient and milling qualities and for the production of industrial or pharmaceutical compounds (Burrell, 2003; Takaiwa, 2005).

Cereal seed storage proteins and starch are produced at high levels during seed development and exclusively within the seed (Shewry et al., 1994), making the promoters of these genes ideal candidates for transgene expression and biotechnology purposes. Several of these promoters have been well studied. Study of the promoters of storage protein genes such as the hordein (barley), gliadin and glutenin (wheat), glutelin (rice) and secalin (rye) has led to the identification of several seed-specific elements (Kluth et al., 2002). Well known seed-specific *cis*-acting elements include the RY repeat (Ezcurra et al., 1999), the GCN4 element (Muller and Knudsen, 1993), the prolamin-box (PROL) (Wu et al., 2000) and the AACA motif (Takaiwa et al., 1996). The GCN4 element is also referred to as the prolamin box or endosperm box and consists of two-well conserved motifs, the -300 core sequence or the E motif (TGHAAARK; H=A/C/T, R=A/G; K=G/T) and the GCN4 element or N motif (TGASTCA; S=G/C, Lamacchia et al., 2001; Vickers et al., 2006). This element is recognized by several basic leucine zipper (bZIP)

transcription factors (Vickers et al., 2006). Another seed-specific element is the C-box (AGCT core), an element recognized by the rice transcription activator-1 expressed in rice endosperm cells and is thought to be a generic activator of rice seed genes (Vickers et al., 2006). The AACA motif is common to glutelin genes and is implicated in high levels of seed-specific expression in rice (Takaiwa et al., 1996).

Some elements found in monocot promoters were first characterized in dicot species. The RY element (CATGCA[TG]) is found in almost all legume seed storage protein promoters, in approximately 50% of cereal seed storage protein promoters, and in many non-storage protein gene promoters (Vickers et al., 2006). Positioning of the RY element at 70-100 bp upstream of the transcription initiation site is important for embryonic storage tissue-specificity (Vickers et al., 2006). The CACA box is found in many monocot and dicot seed storage gene promoters and is involved in tissue-specificity of a *Brassica napus napA* napin seed storage protein gene promoter (Ericson et al., 1991; Ellerström et al., 1996). It also has both positive and negative regulatory effects on the bean β-phaseolin promoter (Vickers et al., 2006). The Soybean Embryo Factor 4 (SEF4) element is found in soybean β-phaseolin promoters and is bound by the endosperm nuclear factor SEF4 and contributes to enhancement of transcription (Lessard et al., 1993).

Investigation of the promoter region of a rice storage protein gene, *GluB-1*, demonstrated that combinatorial interactions between *cis*-elements may play an important role in promoter activity and tissue-specific gene expression (Wu et al., 2000). The minimal promoter of *GluB-1* (197 bp upstream of the transcription start), able to confer endosperm-specific expression contains GCN4, PROL, AACA and ACGT core-

containing motifs (Wu et al., 2000). By varying combinations of these elements Wu et al. (2000) concluded that GCN4 is a core element required for endosperm-specific expression and that it may require two or more partners to form a functional unit. The GCN4 motif, PROL and AACA are often associated and together confer endosperm-specific enhancement (Wu et al., 2000). GCN4 was the only element demonstrated to be essential for endosperm-specific transcription until the identification of an endosperm-specificity palindrome (ESP) element (ACATGTCATCATGT) in the oat globulin promoter (*AsGlo1,* Vickers et al., 2006). This ESP element was shown to be required for endosperm-specificity and contributes to high levels of expression (Vickers et al., 2006). Interrupted palindromic sequences can be recognized by transcription factors that bind to their *cis*-elements as dimers (Vickers et al., 2006).

The *cis*-acting elements in the leader sequence of barley lipoxygenase 1 (*Lox1*) gene, expressed during grain development, are required for embryo-specific expression (Rouster et al., 1998). The region containing the TATA box and downstream sequences (75 bp upstream of the transcriptional start site) conferred only 10% of the full promoter activity but the upstream region (an additional 54 bp) containing an embryo-specific element and a general enhancer element increases promoter activity (Rouster et al., 1998).

A promoter fragment from the granule-bound starch synthase 1 (*gbss1*) gene in wheat drives transgene expression restricted to the endosperm and pericarp of the developing grain (Kluth et al., 2002). This fragment was also found to contain many of the above listed elements required for seed- and endosperm-specific expression.

The mature cereal grain consists of the starchy endosperm, the embryo (commercially known as the germ) and the seed coat (known as the bran). The seed coat and embryo can be removed from the starchy endosperm during milling processes. Therefore, there is the potential to drive separate expression of multiple transgenes in the different seed tissues. Seed coat expressed transgenes can be used as visual markers to distinguish transgenic varieties or for the expression of an extractable product. The differentiation of transgenic seeds has proven to be important and valuable in the prevention of costly mistakes such as the StarLink™ corn fiasco (Schmitz et al., 2005). Seed coat tissues can also play a role in protection of the interior developing tissues from pathogens. The promoter of the barley *Lem2* gene, inducible by salicylic acid, was evaluated for the purposes of targeting antifungal gene expression in wheat and barley tissues involved in the primary route of fungal infection (Abebe et al., 2006). In transgenic plants, the *Lem2* promoter directed expression in the seed protective organs: lemma, palea and glumes, and in the epicarp or outer seed coat tissue layer (Abebe et al., 2006). *Lem2* is not normally expressed in epicarp, and the transgene expression is thought to be a result of a failure to repress expression through methylation (Abebe et al., 2006). This type of aberrant expression is useful for the expression of antifungal agents in protective tissues. Other instances of beneficial expression patterns as a result of methylation loss in promoter regions may be exploited in the future for production of bio-products; however, the stability in methylation over subsequent generations would have to be determined on a case by case basis.

Leaf tissue has the potential to be transformed for improved nutrient quality for forage feed material or the production of industrial commercial or pharmaceutical

compounds. Several well-characterized genes are expressed specifically in chlorophyll-containing leaf tissue. Many of these genes are light inducible such as members of the multi-gene family encoding the small subunit of ribulose-1,5-bisphosphate carboxylase (*rbcS*, Nomura et al., 2000). Promoters of the *rbcS* genes contain both positive and negative regulatory elements that confer light-inducible and tissue-specific expression (Nomura et al., 2000). Examination of the *rbcS* promoter sequence has shown that *cis*-acting elements confer organ-specific and light-inducible regulation but that the elements for organ-specificity may vary among species (Nomura et al., 2000). The I box (GATAAGGCGCGCC) and part of the monocot *rbcS* consensus sequence (GGCCACT) are motifs involved in light regulation (Nomura et al., 2000) and identification of these sequences may be useful in characterizing leaf-specific promoters. Chlorophyll a/b binding (CAB) protein genes are also highly expressed in green tissue and respond to light and diurnal or circadian rhythms (Churin et al., 2003). *CAB1* gene expression in wheat is restricted to chlorophyll containing cells (Churin et al., 2003), another feature that makes it an attractive candidate; transgenes would not be expressed in root tissue remaining in the environment following plant harvest. Four protein binding regions have been mapped in the wheat *CAB1* gene promoter. Three of these (CAB1-A, CAB1-B and CAB1-C) coincide with leaf-specific expression and binding of Myb-like transcription factors to these is important for maximal expression (Churin et al., 2003). CAB and RbcS proteins are highly abundant in green tissue making their gene promoters attractive choices for transgene expression; however, they are both members of gene families and the expression levels could be the sum of many gene contributions (Potenza et al., 2004). One would have to characterize the promoter from each gene family member in order to

select one conferring the desired levels of expression. Selecting a strong promoter from a single-gene family in which there can be confidence in transgene expression levels is more ideal.

Tissue-specific promoters in many cases would be more accurately described by the term, tissue-enhanced. In many cases, the expression of genes driven by these promoters is not entirely restricted to a particular tissue. Expression may occur in other tissues at minimal levels or due to induction by certain conditions as in the case of the *Lox1* gene promoter, which contains an element that confers expression in leaf tissue in response to methyl jasmonate (MeJA) and wounding (Rouster et al., 1998). Transient expression studies of seed-specific promoters also rarely result in the expected level of tissue-specificity, which may be a result of the unusual properties of the endosperm (cell size, starch content, triploidy, or high genome copy number; Russell and Fromm, 1997). Regulation by metabolite levels may also be important for appropriate promoter activity as demonstrated by the increased expression of the maize *waxy* gene promoter in a low starch background (Russell and Fromm, 1997).

In studies on the effects of water stresses on various tissues in rice it was found that the genes induced in a specific tissue in response to stress had promoters that were not enriched in the common stress response elements: the dehydration-responsive element (DRE) and the ABA-responsive element (ABRE, Zhou et al., 2007). This may indicate that the tissue-specific gene expression, even in response to global stress, may be due to tissue-specific promoter elements mediating the response.

Other *cis*-elements responsible for tissue-specific expression may be identified through sequence comparisons. The availability of more sequence data should prove to make this possible and should aid in selecting sequences for further functional analysis.

1.1.7 Promoters and promoter elements conferring developmentally- or temporally-specific expression patterns

Directed gene expression is important for the production of novel traits in plants not only in a tissue-specific manner, but also in a temporal- or developmental-specific manner. Several studies have focused on finding genes expressed at a particular developmental stage in seeds (Girke et al., 2000; Tiwari et al., 2006) and in other plant organs (Abebe et al., 2006). Studies of gene expression related to *Arabidopsis* seeds have often focused on profiling gene expression in a developmental context rather than spatially within the seed due to difficulties associated with the small seed size. Cereal seeds, for the most part are larger, and spatial expression patterns are therefore easier to observe. Temporal expression patterns are also important in determining the utility of a promoter for transgene expression. For instance, the endosperm undergoes an initial proliferative phase of development, followed by the grain filling or storage phase and finally a maturation phase (Berger, 2003). The number of cells formed early in endosperm development during the proliferative phase is critical to the sink strength of the seed and its final size (Tiwari et al., 2006). Promoters of genes expressed in endosperm tissue during this phase have been identified in *Arabidopsis thaliana* and similar strategies may be used to alter seed size in other plant species (Tiwari et al., 2006).

Senescence is the last step in plant development and the timing of this event in relation to grain filling is critical to ensuring the maximum amount of photosynthates are reallocated from leaf tissues to the grains. During this process drastic physiological changes occur in the vegetative tissues and there is an enlargement of most cellular components accompanied by a reduction in chloroplast size, a disorganization of the cytoplasm and decreases in chlorophyll and soluble protein contents (Masclaux et al., 2000). The loss of photosynthetic activity can limit crop yield; therefore an improved understanding of the process is economically and agriculturally important. Significant changes in gene expression occur during leaf senescence as most leaf genes are down-regulated while a set of genes, senescence associated genes (SAGs), are up-regulated (Guo et al., 2004). As well as being part of the normal developmental program of plants, senescence can also be induced by biotic and abiotic stresses, adding to the complexity of its regulation (He et al., 2002). Characterization of promoters of genes expressed at the onset of senescence can help to understand these processes. Analysis of genes expressed during senescence can allow the identification of regulatory sequences associated with senescence and provide insight into the regulatory mechanisms involved.

1.1.8 Global analysis of spatial and temporal gene expression

Many gene expression studies in plants have focused on the response to various biotic or abiotic conditions. More recently, researchers have begun to focus their attention on studies that are more global in scope, that is, studies that survey the gene expression at the organ, tissue and cellular levels (Girke et al., 2000; Wellmer et al., 2004; Wilson et al., 2004; Ma et al., 2005).

Cellular state and timing of development are controlled by gene expression but can be altered by environmental factors. To fully understand the effects that a particular environmental factor, such as disease or temperature, may have on gene expression, it is necessary to be able to distinguish between changes in gene expression due to environmental variables from normal developmental changes in gene expression (Altenbach and Kothari, 2004). By global comparison of transcript accumulation among various cell or tissue types, it can be determined which changes in gene expression reflect changes in cellular state. Global expression analysis methods allow the survey of most, if not all, genes simultaneously rather than individually.

Higher plants possess a relatively simple developmental process, with only three nonreproductive organ systems (leaves, stems and roots) and fewer than 25 major tissue and cell types (Esau, 1977) and therefore provide a good model for defining organ- and tissue-specific genome expression patterns during development. Technological advances and the release of an abundance of sequence data have allowed global profiling of gene expression. Several plants species have been fully sequenced in recent years including the model plant, *A. thaliana* (The Arabidopsis Genome Initiative, 2000) and more recently two sub-species of rice, the model cereal plant (Goff et al., 2002; Yu et al., 2002). *Brachypodium* is also emerging as a new model for temperate grass species, such as wheat and barley, and the genome sequence was recently released (The International Brachypodium Initiative, 2010). Studies of gene expression in triticale will depend upon resources available from these model plant species for which there is abundant information. Triticale studies will also largely depend on information resources from its most closely related species, including rye, and especially wheat, for which there is a

large amount of sequence data currently available. The wheat genome is very large at 16,000 million base pairs (Mb); 140x, 40x and 5x that of the Arabidopsis, rice and the human genomes respectively. In addition to an extremely large genome, common bread wheat (*Triticum aestivum*) is a polyploid species, containing six sets of seven chromosomes (2n=6x=42, AABBDD). The large genome size of wheat is a result of polyploidization as well as amplification of transposable elements and duplication of chromosomal segments. Approximately 90% of the wheat genome consists of repeated sequences and 70% of known transposable elements (Li et al., 2004). The genomic duplications result in many large gene families, with homologues either nearby in chromosomal location or dispersed on multiple chromosomes. The genome size, sequence duplications and large amounts of repeated sequence, make the assembly of genomic sequences extremely difficult in comparison to organisms with smaller simpler genomes. For these reasons, a complete reference sequence of the wheat genome may not be completed for some time. Despite the lack of a completely sequenced genome, wheat has one of the largest collections of expressed sequence tags (EST) sequence data publicly available at well over one million EST's (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). This abundance in sequence data has resulted in the emergence of methods for the analysis of gene expression that are more global in scope. In this context, global studies of wheat can be thought to refer to the analysis of a large number of genes simultaneously as there is no true global gene expression analysis method yet available for all wheat genes. Triticale also lacks a complete genome sequence; triticale research relies heavily on wheat sequence and information resources.

In order to identify uniquely expressed genes, specific cells or tissues are isolated with the assumption that cell/tissue-specific genes ought to confer cell/tissue-specific phenotypes (Galbraith and Birnbaum, 2006). Microarrays are powerful tools for the global analysis of gene expression within specific cell types or tissues and can measure the individual transcript level of tens of thousands of genes simultaneously (Girke et al., 2000; Alba et al., 2004; Wilson et al., 2004; Ma et al., 2005). However, the spatial resolution is limited to the level of separation of tissues or cell types prior to RNA sample preparation. The large amount of RNA required for microarray experiments often limits the spatial resolution to complex organs such as leaves or grains (Leader, 2005). Global analysis of cell type-specific gene expression can be achieved by purifying cell types using fluorescence-activated sorting, laser-capture manipulation, and micro sampling, or by the comparison of whole plant samples to those with ablation, mis-expression or mutation in a particular cell type (reviewed by Galbraith and Birnbaum, 2006). For the purposes of identifying promoters for tissue-targeted transgene expression, the spatial resolution of gene expression analysis need not be at the cellular level. Tissue-specific resolution should be adequate for the identification of candidate genes for further functional analysis of promoters. Once candidate genes are chosen from global analysis of tissue-specific gene expression, spatial resolution can be refined within tissues using *in situ* hybridization (Drea et al., 2005) or reporter gene expression studies (Lamacchia et al., 2001; Abebe et al., 2006). In order to identify novel tissue-specific plant promoters, tissue-specific gene expression needs to be examined. The simplest means of examining expression specificity of tissues is to physically separate them prior to global analysis of gene expression (Galbraith and Birnbaum, 2006).

Microarrays have become a popular and efficient method for analysis of global gene expression through analyzing the accumulation of RNA transcript in tissues of interest. Gene expression microarrays are either cDNA-based arrays or probe-based arrays. cDNA amplicon arrays are less reproducible and less specific than microarrays produced using oligonucleotides as probes. Oligo-arrays avoid cross-hybridization problems associated with shared sequence motifs (Galbraith and Birnbaum, 2006).

High throughput sequencing methods, known as next generation sequencing (NGS), have recently increased in frequency and popularity due to the rapid increase in technological advances and decrease in cost per sequence (Bräutigam and Gowik, 2010). These sequencing methods have emerged as the preferred method for global transcriptional analysis due to the fact that they are not limited to the survey of a set of genes such as would be found on a microarray, but rather can determine relative quantities of all sequences present in a sample. While NGS has the ability to provide quantitative data for more transcripts in a sample compared to microarrays, the analysis of the data is more complex and requires bioinformatics tools that are not yet as streamlined as those that have been developed for microarray analysis, in particular for organisms without a reference genome. The length of sequence reads from NGS methods is shorter than that of traditional Sanger sequencing, meaning that assembly of sequencing reads into contigs to create a transcriptome sequence database must be done prior to quantification of expression in species without a reference genome (Bräutigam and Gowik, 2010). NGS data from various triticale and rye tissue samples was used to generate *de novo* transcriptome assemblies, and distinguish between wheat and rye sequences in the triticale transcriptome (Xu et al., 2011). This reference will be sure to

prove useful in future transcriptional analysis studies in triticale. While both microarray and NGS analysis of gene expression have their advantages and limitations, the results from each are certain to advance the knowledge and understanding of regulation of plant gene expression.

A large proportion of plant genes (60-77%) have been estimated not to exhibit strong tissue-specific expression (Girke et al., 2000). However, global cell-profiling studies have indicated that there are a significant number of genes that do appear to be cell-specific or highly localized to a few cell types (Birnbaum et al., 2003; Galbraith and Birnbaum, 2006). In seed *versus* leaf hybridizations of a spotted array of a developing *Arabidopsis* seed cDNA library, 30% of the cDNAs were seed enhanced and 12% seed-specific with similar values for seed *versus* root hybridizations (Girke et al., 2000). In a comparison of pollen to vegetative tissues of *Arabidopsis*, 469 genes were enriched and 162 genes were specifically expressed in pollen (5.7% and 2.0% of genes surveyed respectively, Becker et al., 2003). In another analysis of the male gametophyte-expressed transcriptome, 1,355 (9.7%) transcripts were specifically expressed in the male gametophyte throughout its development (Honys and Twell, 2004). Analysis of the Arabidopsis root cell types revealed between 200-400 uniquely expressed transcripts, about 2-4% of transcripts detected in the root (Birnbaum et al., 2003). Comparison of expression in 13 Arabidopsis organs, found a smaller number of genes unique to each organ but found that organ systems have distinct transcriptional signatures (Schmid et al., 2005). This might suggest that there is less organ-specificity than cell-specificity. On the other hand, the larger number of organ tissues surveyed may be the cause for the reduction in unique transcripts. In either case, the results highlight the need for

functional analysis of the expression of candidate genes chosen from microarray analysis to fully determine their level of specificity. In another study, gene expression in 17 different Arabidopsis organs was compared using a whole genome oligo array and found that only 16% of the total genes were expressed in all examined organs (Ma et al., 2005). Between 0.1% and 3.2% of genes were specifically expressed in germinating seed, silique, pistil, stamen, petal, sepal, stem, leaf, hypocotyl or root (Ma et al., 2005). Of 207 genes differentially expressed in the developing wheat caryopsis, more than half showed enrichment in a single cell type, and eighty-nine showed specific expression in a single cell type by *in situ* hybridization (Drea et al., 2005). The greatest portion of enriched and specific genes were found in endosperm tissues (Drea et al., 2005). All of these studies support the conclusion that there is a significant level of organ-, tissue-, and cell-specific expression and that expression patterns are defining characteristics of organ, tissue and cellular development.

Global analysis of gene expression has revealed that plants also have temporally distinct expression profiles at different stages in development. A temporal analysis of gene expression in developing grain tissues revealed that almost a quarter of the genes surveyed showed significant differential expression when compared to pre-anthesis wheat spikes (Leader, 2005). Studies of senescing leaves in Arabidopsis using enhancer trap lines have estimated that as much as 10% of the entire genome is expressed in senescing leaves (He et al., 2001). Analysis of an Arabidopsis leaf senescence transcriptome revealed more than 130 transcriptional regulators expressed in senescent leaves (Guo et al., 2004). These transcriptional regulators may give some insight into the promoter regulatory sequences of SAGs.

Global gene expression analysis methods are also often used to survey gene expression changes in response to stresses. Even in response to a global stress (that is stress at the level of the organism, and not at the cellular level), changes in gene expression can be in a cell-, tissue-, or organ-specific manner. Changes in gene expression in Arabidopsis in response to salt, osmotic and cold stress were significantly different between roots and leaves with only minor overlap (Kreps et al., 2002). Using the whole genome rice array to analyze gene expression changes in response to drought and high salinity stresses in shoot, flag leaf and panicles, it was found that changes in gene expression were more similar in the same organ in response to different stresses than between organs in response to the same stress (Zhou et al., 2007). Only a limited number of stress responsive genes were shared between organs suggesting that stress responses are independently regulated within organs. It is also interesting to note that organ-specific gene expression in response to stress did not exhibit enrichment of common stress related regulatory elements, DRE and ABRE motifs (Zhou et al., 2007). Organ-specific stress responses may be mediated by organ-specific regulatory elements. However, the statement that responses to environmental cues are mediated by organ-specific regulatory elements is probably more accurate and encompassing.

Plants perceive a range of environmental cues to regulate their development accordingly. Light is one of the most important of these cues and can have profound effects on plant growth and development (Leyser and Day, 2003). Results of organ-specific gene expression in response to different light conditions found that only a small set of light-regulated genes were shared among organ types and that among organs some genes even showed opposite light-regulation patterns (Ma et al., 2005).

Genes that are involved in the same processes are assumed to be co-expressed therefore identifying spatial and temporal gene expression patterns can help to indicate the function of genes of unknown function that are co-expressed with well characterized genes or transcription factors (Leader, 2005). Global analysis of gene expression cannot only help to identify novel tissue- or cell-specific promoters, but also non-coding sequences of similarly regulated genes can be mined for conserved sequences that will help to identify candidate regulatory *cis*-acting elements (Potenza et al., 2004).

Global studies of genes regulated in a tissue-specific or temporal manner provides a more comprehensive understanding of the transcriptional responses within each tissue and provides a starting point for further elucidating the role of individual genes involved in tissue differentiation, which may be of great value in crop engineering. It also aids in the identification of temporal- and tissue-specific promoters and responsible *cis*-elements within them that are important both for basic study and crop engineering applications.

1.1.9 Summary

Plants can be used successfully to produce a number of industrial, commercial and pharmaceutical bioproducts. Using plants creates a renewable resource for bioproducts and has the potential to increase profit margins for the Canadian agricultural sector. Proteins, carbohydrates and lipids naturally produced in plants can be utilized for commercial and industrial purposes. Alternatively, modification of plant biochemical pathways to produce new compounds can be achieved. The production of heterologous compounds can also be achieved through directed transgene expression.

Hexaploid triticale, a cross between diploid rye and tetraploid durum wheat, has favourable agronomic qualities that make it an ideal candidate for a Canadian and potentially even a worldwide bioindustrial platform. It is high yielding, stress tolerant, adaptable and disease resistant and can be grown and harvested using current Canadian agronomic practices. To make this bioindustrial platform a success in terms of economic benefits and social acceptance, novel endogenous tissue-specific promoters will be required for the expression of transgenes used for bioproduct production. Endogenous promoters will have more functionality than promoters from other unrelated species such as dicots and tissue-specific promoters will allow directed expression of transgenes to the harvestable crop fractions.

Promoters can contain a number of motifs or elements that determine the temporal and spatial expression pattern of a gene. Characterization of these elements and identification of which elements are required for temporal- or tissue-specific expression will be critical to isolating promoters for transgene expression purposes. Many seed-specific elements have already been characterized in previously identified promoters and the presence of these elements in triticale promoters will aid in their characterization.

The identification of novel temporal- and tissue-specific promoters will first require global analysis of gene expression in triticale to identify genes expressed in a temporally and spatially-specific manner. By first identifying genes expressed in a particular tissue and/or at a particular developmental stage, the promoters of these genes can be characterized and evaluated for their potential in controlling expression of transgenes for production of novel bioproducts. The objectives of this study were to identify novel temporal- and tissue-specific promoters in triticale. To achieve these

objectives I first needed to determine appropriate tissue sampling times that would capture gene expression associated with important tissues and developmental processes. While most tissue sampling times can be determined by comparison to growth curves, the timing of leaf senescence is highly dependent on the chosen growth conditions. Therefore, the temporal profile of senescence under the growth conditions used had to be determined and this is addressed in Chapter 2. Once sampling times are determined, the gene expression in the various tissues must be analyzed and compared to determine transcripts that are expressed in a tissue-specific or developmentally-enriched mannter. This was addressed in Chapter 4 using a microarray approach to measure gene expression using the wheat GeneChip. The goals of the expression analysis were to identify tissue-specific, tissue-enriched and developmentally regulated transcripts in triticale seed and vegetative tissues. Determining the function of candidate transcripts and processed that are important to particular tissues or developmental stages requires adequate functional annotation information. To ensure adequate functional annotation information was available for analysis of the microarray results, the improvement of the functional annotation of the wheat GeneChip was addressed and the improved annotation information (Chapter 3) was used to perform functional analysis and allow biological interpretation of the expression results in Chapter 4. Due to limitations in the number of samples used for the microarray expression analysis, an improvement in the spatial resolution of the expression of candidate transcripts was sought and addressed using *in situ* hybridization of RNA probes to triticale seed tissues in Chapter 5. Finally the promoters of candidate tissue-specific transcripts were characterized using sequence

analysis and transient expression of promoter deletion fragments in Chapter 6 to identify novel triticale tissue-specific promoters.

1.2 LITERATURE CITED

Abebe, T., Skadsen, R., Patel, M., and Kaeppler, H. (2006). The *Lem2* gene promoter of barley directs cell- and development-specific expression of *gfp* in transgenic plants. Plant Biotechnology Journal **4,** 35-44.

Alba, R., Fei, Z., Payton, P., Liu, Y., Moore, S.L., Debbie, P., Cohn, J., D'Ascenzo, M., Gordon, J.S., Rose, J.K.C., Martin, G., Tanksley, S.D., Bouzayen, M., Jahn, M.M., and Giovannoni, J. (2004). ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. The Plant Journal **39,** 697-714.

Altenbach, S.B., and Kothari, K.M. (2004). Transcript profiles of genes expressed in endosperm tissue are altered by high temperature during wheat grain development. Journal of Cereal Science **40,** 115-126.

Ammar, K., Mergoum, M., and Rajaram, S. (2004). The history and evolution of triticale. In Triticale improvement and production. FAO plant production and protection paper 179, M. Mergoum and H. Gomez-Macpherson, eds (Rome: Food and Agriculture Organization of the United Nations), pp. 1-9.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408,** 796-815.

Baginsky, S., Hennig, L., Zimmermann, P., and Gruissem, W. (2010). Gene expression analysis, proteomics, and network discovery. Plant Physiology **152,** 402-410.

Becker, J.D., Boavida, L.C., Carneiro, J., Haury, M., and Feijo, J.A. (2003). Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiology **133,** 713-725.

Berger, F. (2003). Endosperm: the crossroad of seed development. Current Opinion in Plant Biology **6,** 42-50.

Bergstra, R. (2006). Manufacturing competitiveness of triticale for industrial fermentation (MTN Consulting Associates). Pp 2.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science **302,** 1956-1960.

Bräutigam, A., and Gowik, U. (2010). What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biology **12,** 831-841.

Burrell, M.M. (2003). Starch: the need for improved quality or quantity-an overview. Journal of Experimental Botony **54,** 451-456.

Canadian Biotechnology Advisory Committee (2002). Improving the regulation of genetically modified foods and other novel foods in Canada. Report to the Government of Canada Biotechnology Ministerial Coordinating Committee. (Ottawa ON: Canadian Biotechnology Advisory Committee). pp 62.

Chen, W., Provart, N.J., Glazebrook, J., Katagiri, F., and Chang, H.S. (2002). Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell **14,** 559.

Churin, Y., Adam, E., Kozma-Bognar, L., Nagy, F., and Börner, T. (2003). Characterization of two Myb-like transcription factors binding to CAB promoters in wheat and barley. Plant Molecular Biology **52,** 447-462.

Clarke, B., Lambrecht, M., and Rhee, S. (2003). Arabidopsis genomic information for interpreting wheat EST sequences. Functional & Integrative Genomics **3,** 33-38.

Domenek, S., Feuilloley, P., Gratraud, J., Morel, M.-H., and Guilbert, S. (2004). Biodegradability of wheat gluten based bioplastics. Chemosphere **54,** 551-559.

Drea, S., Leader, D.J., Arnold, B.C., Shaw, P., Dolan, L., and Doonan, J.H. (2005). Systematic spatial analysis of gene expression during wheat caryopsis development. Plant Cell **17,** 2172-2185.

Dunwell, J.M. (2005). Review: intellectual property aspects of plant transformation. Plant Biotechnology Journal **3,** 371-384.

Dynan, W.S. (1989). Modularity in promoters and enhancers. Cell **58,** 1-4.

Ellerström, M., Stålberg, K., Ezcurra, I., and Rask, L. (1996). Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. Plant Molecular Biology **32,** 1019-1027.

Environment Canada. Accessed September 16, 2012. http://www.ec.gc.ca/default.asp?lang=En& n=714D9AAE-1&news=836027A7-252D-461F-A539-9CC10159D0E4,

Ericson, M.L., Muren, E., Gustavsson, H.-O., Josefsson, L.-G., and Rask, L. (1991). Analysis of the promoter region of napin genes from *Brassica napus* demonstrates binding of nuclear protein in vitro to a conserved sequence motif. European Journal of Biochemistry **197,** 741-746.

Esau, K. (1977). Anatomy of Seed Plants. (New York: Wiley).

Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K., and Rask, L. (1999). Interaction between composite elements in the napA promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. Plant Molecular Biology **40,** 699-709.

FAO. (2010). Statistical DataBases. Agriculture. http://faostat.fao.org/site/339/ default.aspx

Fujiwara, T., and Beachy, R.N. (1994). Tissue-specific and temporal regulation of a β-conglycinin gene: roles of the RY repeat and other cis-acting elements. Plant Molecular Biology **24,** 261-272.

Galbraith, D.W., and Birnbaum, K. (2006). Global studies of cell type-specific gene expression in plants. Annual Review of Plant Biology **57,** 451-475.

Garcia, D., FitzGerald, J.N., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed sized in Arabidopsis. Plant Cell **17,** 52-60.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C., and Ohlrogge, J. (2000). Microarray analysis of developing Arabidopsis seeds. Plant Physiology **124,** 1570-1581.

Giunta, F., Motzo, R., and Deidda, M. (1999). Grain yield analysis of a triticale (× *Triticosecale* Wittmack) collection grown in a Mediterranean environment. Field Crops Research **63,** 199-210.

Goddijn, O.J.M., and Pen, J. (1995). Plants as bioreactors. Trends in Biotechnology **13,** 379-387.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.-l., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002). A draft sequence of the rice genome (*Oryza sativa L. ssp. japonica*). Science **296,** 92-100.

Gomord, V., Fitchette, A.-C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., and Faye, L. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. Plant Biotechnology Journal **8,** 564-587.

Gong, Y.H., Zhang, J., Gao, J.F., Lu, J.Y., and Wang, J.R. (2005). Slow export of photoassimilate from stay-green leaves during late grain-filling stage in hybrid winter wheat (*Triticum aestivum* L.). Journal of Agronomy and Crop Science **191,** 292-299.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C., and Gelbart, W.M. (1996). An introduction to genetic analysis. (New York: W.H.Freeman and Company).

Guo, Y., Cai, Z., and Gan, S. (2004). Transcriptome of Arabidopsis leaf senescence. Plant, Cell & Environment **27,** 521-549.

Hanson, B.D., Mallory-Smith, C.A., Shafii, B., Thill, D.C., and Zemetra, R.S. (2005). Pollen-mediated gene flow from blue aleurone wheat to other wheat cultivars. Crop Science **45,** 1610-1617.

He, Y., Fukushige, H., Hildebrand, D.F., and Gan, S. (2002). Evidence supporting a role of jasmonic acid in Arabidopsis leaf senescence. Plant Physiology **128,** 876.

He, Y., Tang, W., Swain, J.D., Green, A.L., Jack, T.P., and Gan, S. (2001). Networking senescence-regulating pathways by using Arabidopsis enhancer trap lines. Plant Physiology **126,** 707.

Hills, M.J., Hall, L.M., Messenger, D.F., Graf, R.J., and Beres, B.L. (2007). Evaluation of crossability between triticale (x*Triticosecale* Wittmack) and common wheat, durum wheat and rye. Environmental Biosafety Research **6,** 249-257.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biology **5,** 1-13.

Horlein, A., and Valentine, J. (1995). Triticale (× *Triticosecale*). In Cereals and psuedocereals, Williams. J.T., ed (London: Chapman and Hall), pp. 187-221.

The International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463,** 763-768.

Jerez, A., Partal, P., Martínez, I., Gallegos, C., and Guerrero, A. (2007). Protein-based bioplastics: effect of thermo-mechanical processing. Rheologica Acta **46,** 711-720.

Kluth, A., Sprunck, S., Becker, D., Lörz, H., and Lütticke, S. (2002). 5′ deletion of a *gbss1* promoter region from wheat leads to changes in tissue and developmental specificities. Plant Molecular Biology **49,** 665-678.

Ko, K., Ahn, M.H., Song, M., Choo, Y.K., Kim, H.S., Ko, K., and Joung, H. (2008). Glyco-engineering of biotherapeutic proteins in plants. Molecular Cells **25,** 494-503.

Koutinas, A.A., Wang, R., and Webb, C. (2004). Evaluation of wheat as generic feedstock for chemical production. Industrial Crops and Products **20,** 75-88.

Kreps, J.A., Wu, Y., Chang, H.-S., Zhu, T., Wang, X., and Harper, J.F. (2002). Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. Plant Physiology **130,** 2129-2141.

Kuleung, C., Baenziger, P.S., and Dweikat, I. (2004). Transferability of SSR markers among wheat, rye, and triticale. Theoretical and Applied Genetics **108,** 1147-1150.

Lamacchia, C., Shewry, P.R., Di Fonzo, N., Forsyth, J.L., Harris, N., Lazzeri, P.A., Napier, J.A., Halford, N.G., and Barcelo, P. (2001). Endosperm-specific activity of a storage protein gene promoter in transgenic wheat seed. Journal of Experimental Botany **52,** 243-250.

Leader, D.J. (2005). Transcriptional analysis and functional genomics in wheat. Journal of Cereal Science **41,** 149-163.

Lessard, P.A., Allen, R.D., Fujiwara, T., and Beachy, R.N. (1993). Upstream regulatory sequences from two β-conglycinin genes. Plant Molecular Biology **22,** 873-885.

Leyser, O., and Day, S. (2003). Mechanisms in Plant Development (Malden, MA: Blackwell Publishing Ltd.).

Li, W., Zhang, P., Fellers, J.P., Friebe, B., and Gill, B.S. (2004). Sequence composition, organization, and evolution of the core Triticeae genome. The Plant Journal **40,** 500-511.

Ma, L., Sun, N., Liu, X., Jiao, Y., Zhao, H., and Deng, X.W. (2005). Organ-specific expression of Arabidopsis genome during development. Plant Physiology **138,** 80-91.

Ma, J.K.-C., Drake, P.M.W., and Christou, P. (2003). The production of recombinant pharmaceutical proteins in plants. Nature Reviews Genetics **4,** 794-805.

Macdonald, P. (2004). The science and regulation of plant molecular farming in Canada. In The 8th international symposium on the biosafety of genetically modified organisms (Montpellier, France), pp. 187-189.

Masclaux, C., Valadier, M.H., Brugière, N., Moro-Gaudry, J.F., and Hirel, B. (2000). Characterization of the sink/source transition in tobacco (*Nicotiana tabacum* L.) shoots in relation to nitrogen management and leaf senescence. Planta **211,** 510-518.

Matus-Cadiz, M.A., Hucl, P., and Dupuis, B. (2007). Pollen-mediated gene flow in wheat at the commercial scale. Crop Science **47,** 573-579.

Muller, M., and Knudsen, S. (1993). The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. The Plant Journal **4,** 343-355.

Natural Resources Canada. Accessed September 16, 2012. http://oee.nrcan.gc.ca/transportation/alternative-fuels/programs/10163

Nikolov, D.B., and Burley, S.K. (1997). RNA polymerase II transcription initiation: A structural view. Procedings of the National Academy of Sciences of the United States of America **94,** 15-22.

Nomura, M., Katayama, K., Nishimura, A., Ishida, Y., Ohta, S., Komari, T., Miyao-Tokutomi, M., Tajima, S., and Matsuoka, M. (2000). The promoter of *rbcS* in a C3 plant (rice) directs organ-specific, light-dependent expression in a C4 plant (maize), but does not confer bundle sheath cell-specific expression. Plant Molecular Biology **44,** 99-106.

Odell, J.T., Nagy, F., and Chua, N.-H. (1985). Identification of sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature **313,** 810-812.

PlasticsEurope. (2011). Plastics - the Facts 2011, An analysis of European plastics production, demand and recovery for 2010 (PlasticsEurope).

Potenza, C., Aleman, L., and Sengupta-Gopalan, C. (2004). Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. In Vitro Cellular & Developmental Biology - Plant **40,** 1-22.

Reddy, C.S.K., Ghai, R., Rashmi, and Kalia, V.C. (2003). Polyhydroxyalkanoates: an overview. Bioresource Technology **87,** 137-146.

Rouster, J., van Mechelen, J., and Cameron-Mills, V. (1998). The untranslated leader sequence of the barley *lipoxygenase 1* (*Lox1*) gene confers embryo-specific expression. The Plant Journal **15,** 435-440.

Rubio, S., Jouve, N., and González, J.M. (2004). Biolistic transfer of the gene *uidA* and its expression in haploid embryo-like structures of Triticale (× Triticosecale Wittmack). Plant Cell, Tissue and Organ Culture **77,** 203-209.

Russell, D.A., and Fromm, M.E. (1997). Tissue-specific expression in transgenic maize of four endosperm promoters from maize and rice. Transgenic Research **6,** 157-168.

Salmon, D.F. (2004). Production of triticale on the Canadian Prairies. In Triticale improvement and production. FAO plant production and protection paper 179, M. Mergoum and H. Gomez-Macpherson, eds (Rome: Food and Agriculture Organization of the United Nations), pp. 99-102.

Schaffner, A.R., and Sheen, J. (1991). Maize *rbcS* promoter activity depends on sequence elements not found in dicot *rbcS* promoters. The Plant Cell **3,** 997-1012.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. Nature Genetics **37,** 501-506.

Schmitz, T.G., Schmitz, A., and Moss, C.B. (2005). The economic impact of StarLink corn. Agribusiness **21,** 391-407.

Sechley, K.A., and Schroeder, H. (2002). Intellectual property protection of plant biotechnology inventions. Trends in Biotechnology **20,** 456-461.

Shewry, P.R., and Halford, N.G. (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. Journal of Experimental Botony **53,** 947-958.

Shewry, P.R., Tatham, A.S., Halford, N.G., Barker, J.H.A., Hannappel, U., Gallois, P., Thomas, M., and Kreis, M. (1994). Opportunities for manipulating the seed protein composition of wheat and barley in order to improve quality. Transgenic Research **3,** 3-12.

Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N., and Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology **23,** 75-81.

Smidansky, E., Martin, J., Hannah, C., Fischer, A., and Giroux, M. (2003). Seed yield and plant biomass increases in rice are conferred by deregulation of endosperm ADP-glucose pyrophosphorylase. Planta **216,** 656-664.

Statistics Canada. Accessed December 19, 2007. Table 002-0022 - Farm product price index, annual (index, 1997=100), CANSIM (database), Using E-STAT (distributor). http://estat.statcan.ca/cgi-win/cnsmcgi.exe?Lang=E&ESTATFile=EStat\English\CII_1_E.htm&RootDir=ESTAT/

Taiz, L., and Zeiger, E. (2002). Plant Physiology. (Sunderland, Massachusetts: Sinauer Associates, Inc.).

Takaiwa, F. (2005). Health-promoting transgenic rice suppressing life-related disease and type-I allergy. In Rice is life: scientific perspectives for the 21st century Proceedings of the World Rice Research Conference held in Tsukuba, Japan, 4-7 November 2004. pp 102-104.

Takaiwa, F., Yamanouchi, U., Yoshihara, T., Washida, H., Tanabe, F., Kato, A., and Yamada, K. (1996). Characterization of common *cis*-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. Plant Molecular Biology **30,** 1207-1221.

Tiwari, S., Spielman, M., Day, R.C., and Scott, R.J. (2006). Proliferative phase endosperm promoters from *Arabidopsis thaliana*. Plant Biotechnology Journal **4,** 393-407.

van Dongen, J.T., Roeb, G.W., Dautzenberg, M., Froehlich, A., Vigeolas, H., Minchin, P.E.H., and Geigenberger, P. (2004). Phloem import and storage metabolism are highly coordinated by the low oxygen concentrations within developing wheat seeds. Plant Physiology **135,** 1809-1821.

Vickers, C., Xue, G., and Gresshoff, P. (2006). A novel *cis*-acting element, ESP, contributes to high-level endosperm-specific expression in an oat globulin promoter. Plant Molecular Biology **62,** 195-214.

Wellmer, F., Riechmann, J.L., Alves-Ferreira, M., and Meyerowitz, E.M. (2004). Genome-wide analysis of spatial gene expression in Arabidopsis flowers. Plant Cell **16,** 1314-1326.

Wiley, P., Tosi, P., Evrard, A., Lovegrove, A., Jones, H., and Shewry, P. (2007). Promoter analysis and immunolocalisation show that puroindoline genes are exclusively expressed in starchy endosperm cells of wheat grain. Plant Molecular Biology **64,** 125-136.

Wilson, I.D., Barker, G.L.A., Beswick, R.W., Shepherd, S.K., Lu, C., Coghill, J.A., Edwards, D., Owen, P., Lyons, R., Parker, J.S., Lenton, J.R., Holdsworth, M.J., Shewry, P.R., and Edwards, K.J. (2004). A transcriptomics resource for wheat functional genomics. Plant Biotechnology Journal **2,** 495-506.

Wu, C.-Y., Washida, H., Onodera, Y., Harada, K., and Takaiwa, F. (2000). Quantitative nature of the Prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal *cis*-element requirements for endosperm-specific gene expression. The Plant Journal **23,** 415-421.

Xu, Y., Badea, C., Tran, F., Frick, M., Schneiderman, D., Robert, L., Harris, L., Thomas, D., Tinker, N., Gaudet, D., and Laroche, A. (2011). Next-Gen sequencing of the transcriptome of triticale. Plant Genetic Resources **9,** 181-184.

Yang, J., Zhang, J., Wang, Z., Xu, G., and Zhu, Q. (2004). Activities of key enzymes in sucrose-to-starch conversion in wheat grains subjected to water deficit during grain filling. Plant Physiology **135,** 1621-1629.

Yoshihara, T., and Takaiwa, F. (1996). *cis*-Regulatory elements responsible for quantitative regulation of the rice seed storage protein glutelin *GluA-3* gene. Plant & Cell Physiology **37,** 107-111.

Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., and Yang, H. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science **296,** 79-92.

Zakharov, A., Giersberg, M., Hosein, F., Melzer, M., Muntz, K., and Saalbach, I. (2004). Seed-specific promoters direct gene expression in non-seed tissue. Journal of Experimental Botony **55,** 1463-1471.

Zhou, J., Wang, X., Jiao, Y., Qin, Y., Liu, X., He, K., Chen, C., Ma, L., Wang, J., Xiong, L., Zhang, Q., Fan, L., and Deng, X. (2007). Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. Plant Molecular Biology **63,** 591-608.

CHAPTER 2 The temporal profile of leaf senescence in triticale

2.1 INTRODUCTION

Senescence is the final stage in leaf development and is defined as the gradual coordinated deterioration of its functions with age as nutrients are remobilized from vegetative tissues for storage in sink tissues such as tubers, fruit or grains (Hafsi et al., 2000; Madeira and de Varennes, 2005; Gregersen and Holm, 2007). Leaf senescence progresses in an age-dependent manner but complex interactions of developmental and environmental signals influence the timing, extent and rate of senescence (Gan and Amasino, 1997; Hafsi et al., 2000; Buchanan-Wollaston et al., 2003). The deterioration of leaf tissue during senescence is not a random process but is highly regulated as leaf cells undergo changes in cell structure, metabolism, and gene expression ensuring a highly controlled degradation of cellular organelles and their proteins, nucleic acids, carbohydrates and lipids (Lee et al., 2004; Gregersen and Holm, 2007; Lim et al., 2007). The vascular system is preserved allowing the efficient translocation of assimilates until late stages of leaf senescence (Kichey et al., 2005). Leaf senescence is a slower process than other types of programmed cell death, which are generally acute and rapid, ensuring the efficient remobilization of nutrients from senescing leaves to developing sink tissues (Lee et al., 2004; Lim et al., 2007). Senescence begins shortly after full expansion of the leaf and involves a decrease in photosynthetic activity and an increase in catabolic activity (Madeira and de Varennes, 2005; Gregersen et al., 2008). Visibly, leaves change color due to chlorophyll loss, reduced water content, organelle degradation and finally membrane deterioration (Hafsi et al., 2000; Balazadeh et al., 2008).

In cereals, the process of senescence is of major agricultural importance because it occurs during the grain filling period and is related to nutrient use efficiency and yield (Gregersen et al., 2008). The leaf is the major photosynthetic organ and the timing of translocation of nutrients from the vegetative leaf tissue to developing grains impacts yield, making this process critical to the production of grains and its understanding is pertinent to the improvement of agricultural yields (Hafsi et al., 2000). The improvement of nutrient mobilization from the vegetative parts of the plant to the grain has been one of the major objectives of plant breeding (Gregersen and Holm, 2007). Early senescence may be yield-limiting in crop plants by reducing the growth phase, and therefore the total amount of assimilates that can be remobilized to the grain (Lim et al., 2007). On the other hand, delayed senescence, can result in delayed remobilization and unused assimilates remaining in vegetative tissues at harvest, resulting in a lower harvest index (i.e., the grain weight as a percentage of total biomass at the time of harvest, Gong et al., 2005). Delayed senescence is a trait of 'stay-green' varieties that have prolonged periods of photosynthesis and the potential for higher yields. However, these 'stay-green' varieties have been shown to have a lower harvest index due to delayed remobilization of photosynthates (Gong et al., 2005; Gregersen et al., 2008). Despite the potential to increase the total yield due to prolonged photosynthesis, this trait may be unfavourable in climates where the growing season is limited and yield potential must be met during a short harvest season.

In addition to timing, rates of senescence are also important. Rapid senescence might be beneficial for adjusting leaf area to water availability in regions with frequent drought or heat stress, while slow senescence seems to be characteristic of cultivars bred

for high yield in less stress prone environments (Saulescu et al., 2001). Selecting for or genetically manipulating the senescence pattern to suit a given environment has the potential to positively impact crop yields. Studying leaf senescence to understand the biological processes involved will provide the knowledge needed to identify targets for improvement and allow proper selection in breeding programs.

Understanding the molecular mechanisms involved in senescence has been the focus of many studies in the last decade, particularly in the cereal crops, due to the improved yield potential. Analysis of senescence associated gene (SAG) expression contributes to the understanding of the processes involved and identifies targets for selection and improvement. Of late, genome wide analysis using cDNA library construction and/or microarrays has been employed in *Arabidopsis* (Buchanan-Wollaston et al., 2003; Guo et al., 2004; Lin and Wu, 2004; Buchanan-Wollaston et al., 2005) and *Populus* (Andersson et al., 2004). More recently, researchers have also begun to examine the transcriptome in response to senescence using microarrays for cereal crops, including wheat (Gregersen and Holm, 2007) and barley (Parrott et al., 2007; Jukanti et al., 2008). Microarray analysis can identify genes whose transcription is induced (up regulated) and repressed (down regulated) during senescence and will provide a more detailed picture of the transcriptome involved in leaf senescence. It is important to carefully sample tissue to be used in microarray experiments as it can have a significant impact on the results (Zhang et al., 2004). Determining the appropriate sampling times requires determining the temporal profile of senescence.

While the senescence of a crop is often obvious in the field, documenting the process is difficult because of the need for frequent sampling during periods of rapid

change (Adamsen et al., 1999). There are several well-established methods for quantitatively measuring leaf senescence symptoms including chlorophyll content, photochemical efficiency, senescence associated enzyme activities, change in protein levels, membrane leakage, and gene expression (Lim et al., 2007). Methods such as gene expression assays, cell content assays or any other method that requires a tissue sample or causes wounding are only able to reveal information regarding the time of sampling and do not reflect the progression of senescence. Non-invasive methods are key to developing a temporal profile of the progression of senescence over the life cycle of the plant and/or specific leaves and several methods have been developed for evaluation of senescence at the leaf and crop level (Adamsen et al., 1999; Hafsi et al., 2000; Yamamoto et al., 2002; Madeira and de Varennes, 2005).

The obvious visible symptom of leaf senescence is yellowing, or loss of greenness, resulting from the degradation of chlorophyll in the chloroplasts (Thompson et al., 1998). Degradation of the chloroplasts is the earliest structural change during senescence, which leads to the degradation of chlorophyll due to the disruption of the thylakoid membranes (Lee et al., 2004; Lim et al., 2007). Therefore, greenness is a useful parameter for documenting the progression of senescence in leaves and several methods have been developed for its assessment at the field canopy and leaf levels.

Visual observation of greenness is a simple, efficient and cost effective method for assessing the progression of senescence. The survivorship curve assay is based on visual examination to determine the time at which half of the leaf turns yellow (Hafsi et al., 2000). While efficient, visual assessment alone is inaccurate due to the subjective nature of human visual observations. This approach has been found to overestimate

senescence, not capture early events in the senescence process and is limited in its sensitivity compared to other methods (Hafsi et al., 2000).

Chlorophyll extraction and measurement by spectrophotometry is another common method for quantitatively assessing greenness in leaf tissue that has been employed in several studies of senescence (Oh et al., 1997; Lu et al., 2001; Woo et al., 2001; Yamamoto et al., 2002). Aside from being a laborious and time consuming method, it is destructive and therefore does not allow the study of the progression of senescence within a chosen plant or leaf over a period of time.

Portable hand held radiometers can be used to measure spectral reflectance, that can be used to calculate a number of vegetative indices related to photosynthetic activity and has been used to evaluate the senescence of crop canopies (Adamsen et al., 1999). This method measures reflected light without making contact and many plants are averaged in a single measurement (Adamsen et al., 1999). This is useful for evaluating crop canopies in a field but cannot separate the contributions from different plants or leaves. In order to analyze leaf senescence, measurements should be taken on a single leaf basis along with age information. Individual leaves of a plant have different ages and consequently may undergo senescence at different time points. For example, the flag leaf of triticale is the last leaf to emerge and develop and is consequently younger than other leaves.

Numerical image analysis has been used to determine the ratio of green to red (G/R) in images of both leaves (Hafsi et al., 2000) and canopy (Adamsen et al., 1999) in wheat using a digital camera. G/R has higher day to day variation than other methods (Adamsen et al., 1999). Lighting quality, angle and scale are all important parameters

that can be difficult to control when using digital images to assess the greenness and consequently senescence of leaf tissue.

An alternative, non-destructive method for measuring greenness is with chlorophyll meters such as the SPAD-502 (Minolta Canada Inc., Mississauga, Ontario). For studies of senescence where the chlorophyll content must be measured on the same leaves over time, this non-destructive method is useful (Yamamoto et al., 2002). Chlorophyll meters have been used successfully to evaluate leaf greenness in rice, maize, wheat, and potato (Yamamoto et al., 2002; Madeira and de Varennes, 2005). The SPAD-502 measures the amount of chlorophyll in the leaf, which is related to leaf greenness, by transmitting light through a leaf at wavelengths of 650 and 940 nm. The 650 nm light corresponds to peak chlorophyll attenuation of red light, while the 940 nm signal is not absorbed by chlorophyll. The value obtained represents the ratio between transmittances of these wavelengths through the leaf tissue, and consequently it tends to normalize for variables not related to chlorophyll concentration such as leaf thickness and cuticle reflectance (Adamsen et al., 1999). The sensor head of the SPAD-502 meter is clamped directly onto the leaf to make a measurement and care must therefore be taken to avoid any physical damage to leaves over repeated measurements. The SPAD-502 chlorophyll meter is a simple, quick, rapid and non-destructive method for estimating chlorophyll content in leaves and surveying the relative differences in chlorophyll content throughout leaf senescence (Yamamoto et al., 2002).

It is hypothesized that with a detailed temporal profile of senescence under controlled growth conditions, the optimal sampling times for gene expression studies could be identified. In this experiment, a temporal profile of senescence in triticale

leaves was defined under controlled growth conditions using the non-destructive method of measuring greenness with the SPAD-502 chlorophyll meter. This profile will be used as a reference for tissue sampling in future gene expression experiments aimed at determining the molecular characteristics of senescence in triticale. These experiments will help to elucidate the genes and mechanisms involved in age-dependent leaf senescence of triticale and identify targets for selection in crop improvement breeding programs.

2.2 MATERIALS AND METHODS

2.2.1 Plants and growth conditions

The triticale (x *Triticosecale* Wittm.) cultivar used in this study was AC Certa. Three seeds were planted equally spaced, 2 cm deep into 1 gallon pots of Cornell mix, a steam-pasteurized (121˚C for 8h) mixture of loam, sand, and peat (3:1:1, by volume) (Boodley and Sheldrake Jr., 1977). Seedlings were thinned to 2 per pot after 1 week growth. All plants were grown in growth cabinets (Conviron, Winnipeg, Canada) in order to maintain disease and stress free environments. Growth conditions were 16 hour photoperiods at 18˚C and 8 hour dark periods at 15˚C. Light intensities of 350 µmole m^{-2} s^{-1} of photosynthetic photon flux density were provided by cool white fluorescent tubes and incandescent bulbs.

2.2.2 Measurement of relative chlorophyll content

Plants were checked daily and heads were tagged at the date of anthesis. Chlorophyll measurements were taken on a near daily basis following anthesis using a Minolta SPAD-502 Chlorophyll Meter (Minolta Canada Inc., Mississauga, Ontario). The value determined by the chlorophyll meter provides an indication of the relative amount of chlorophyll present in plant leaves. The measurements were made 10 times at distances of 2-3 cm from the stalk spanning the width of the leaf and the average of the 10 measurements for each leaf was recorded. Measurements were taken for flag leaves (last leaf to emerge) and third leaves (third from top of stem) from the date of anthesis until the chlorophyll meter readings were less than 10. Readings were also disregarded

when leaves became dry and brittle because previous studies have shown anomalous performance of the meter under those conditions (Adamsen et al., 1999).

2.2.3 Statistical analysis and determination of senescence

Analysis of relative chlorophyll readings was performed using SAS, version 9.1 software (SAS Institute, Cary, NC). The NLIN procedure using the Marquardt method was used to fit a linear-linear segmented regression to the data from each leaf of each plant measured to determine the number of days post-anthesis (DPA) that relative chlorophyll content begins to decline. Linear-linear segmented regressions were fitted to the data for each leaf measured. Plants that had data sufficient in quality and quantity to meet the criteria to determine the convergence points, were used to determine the mean number of DPA when relative chlorophyll content declines. Convergence criteria included a change in error associated with the NLIN of < 0.000001 with 50 iterations of the procedure or a failure to find two best-fit linear equations.

2.3 RESULTS

Plants reared in growth chambers are in an environment relatively free of stress and disease, an advantage of growth chambers over field experiments. No visible damage was observed in leaf areas subjected to chlorophyll measurement with the SPAD-502 meter as a result of the measurements. The trend observed in relative chlorophyll content determined by the SPAD-502 readings is a plateau followed by a steady decline (Figure 2.1). This trend can be summarized by fitting a linear-linear segmented regression to determine the point of initiation of decline in chlorophyll content for each leaf. Of the 96 flag leaves and 96 third leaves measured, 92 flag leaves and 80 third leaves had sufficient data to meet the criteria for fitting a linear-linear segmented regression and finding a convergence point for the two segments. The mean number of DPA for which the segments converged was 27.4 DPA (± 1.4, $p > 0.05$) and 21.5 DPA (± 1.7, $p > 0.05$) for flag and third leaves respectively (Table 2.1). Flag leaves reached a stage of measurable senescence an average of 6 days later than third leaves. Visibly, leaves showed greater amounts of yellowing after the convergence point, which correlated with the decline in relative chlorophyll content measured by the SPAD-502 chlorophyll meter (Figure 2.2).

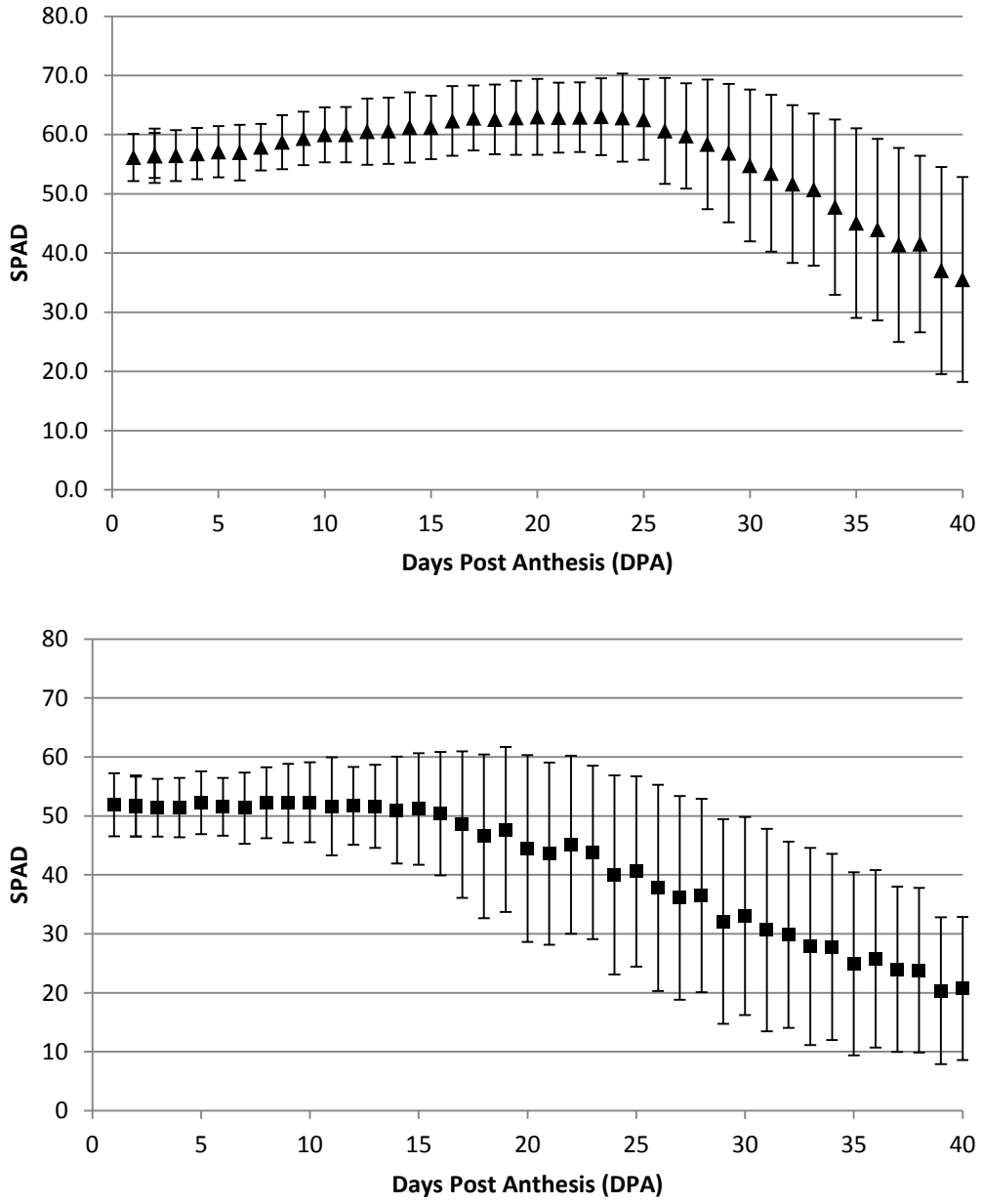


Figure 2.1 Temporal changes in relative chlorophyll content after anthesis in triticale (AC Certa) leaves. ▲ flag leaves, ■ third leaves. Values are means (flag leaf n = 92, third leaves n = 80) of SPAD-502 readings. Error bars are +/-1 standard deviation.

Table 2.1 Convergence points in triticale leaf relative chlorophyll content. Convergence points were determined as the number of days post-anthesis (DPA) that leaf relative chlorophyll content began to decline.

Leaf	Number of observations (leaves)	Convergence Point Mean DPA	Standard Deviation
Flag Leaf	92	27.4	± 1.4
Third Leaf	80	21.5	± 1.7

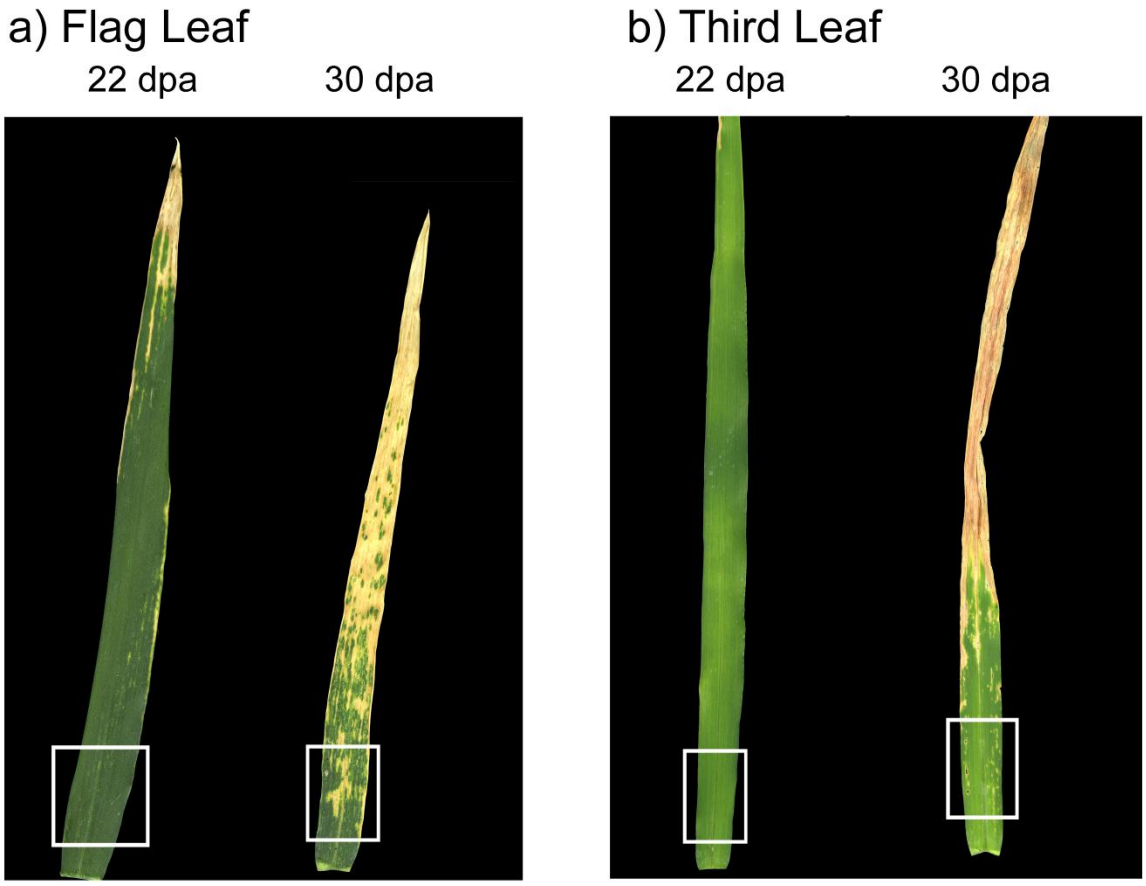


Figure 2.2 Representative leaf tissues before and after relative chlorophyll content decline. Areas of leaves measured (white boxes) for relative chlorophyll content with the SPAD-502 chlorophyll meter are shown before (22 dpa) and after (30 dpa) relative chlorophyll content decline. a) Flag leaf at 22 days post-anthesis (dpa) and 30 dpa, b) third leaf at 22 dpa and 30 dpa

2.4 DISCUSSION

The SPAD-502 chlorophyll meter was found to be a useful method for determining the profile of senescence in triticale. The pattern observed in relative change in chlorophyll was similar to that reported in a wheat study (Adamsen et al., 1999). Flag and third leaves had similar patterns of senescence although the pattern of the third leaf was shifted compared to the flag leaf by approximately 6 days earlier. Leaf primordial initiation rate is approximately 0.35 primordia per day for spring wheat grown under controlled environmental conditions (Heyne, 1987). If triticale has a similar primordia initiation rate to wheat, this would correspond to the overall difference in developmental timing of senescence and the 6 days difference would correspond to the two leaf difference in development. If this is the case, the senescence pattern observed using the SPAD-502 chlorophyll meter could potentially be extended to other leaves by shifting the pattern accordingly. The rate of primordial initiation in triticale under the chosen growth conditions would have to be examined to confirm this.

Comparing leaves from different plants using the raw SPAD values is not useful since there is a large amount of variation in values of individual leaves at anthesis (see error bars in Figure 2.1). SPAD values for flag and third leaves at anthesis varied from 44 to 63 and from 41 to 67, respectively. It is more useful to compare the pattern of senescence or the relative change in SPAD values, which tend to follow similar trends among plants.

Using the linear-linear segmented regression identifies the point at which the relative chlorophyll content begins to decline for each leaf in number of DPA. Since

chlorophyll degradation is the first visible symptom of senescence (Buchanan-Wollaston et al., 2003; Lee et al., 2004), this time point will be a useful reference for sampling tissue for future molecular studies of senescence in triticale. Comparing the molecular status of samples taken before this time point with samples taken after this time point should help to reveal some of the mechanisms involved in leaf senescence in triticale. It is important to note that referencing the results from this study for the purpose of studying senescence should only be done in studies using the same growth conditions in the same cultivar. Changes to the growth conditions or cultivar may cause changes to the rate, extent or pattern of senescence (Zhao et al., 2007).

This study has demonstrated that measuring chlorophyll content in leaves during the post-anthesis developmental stage results in similar patterns to previous cereal crop studies. This study surveyed chlorophyll content throughout post-anthesis leaf development to ensure detection of changes occurring due to senescence. The results have allowed the development of a temporal profile of chlorophyll content that can be used as a reference for senescence in triticale grown under the specified conditions. By sampling before and after the point in development at which chlorophyll content begins to decline as determined by this study, relative changes in gene expression can be used to identify triticale SAGs.

2.5 LITERATURE CITED

Adamsen, F.J., Pinter, P.J., Barnes, E.M., LaMorte, R.L., Wall, G.W., Leavitt, S.W., and Kimball, B.A. (1999). Measuring wheat senescence with a digital camera. Crop Science **39,** 719-724.

Andersson, A., Keskitalo, J., Sjodin, A., Bhalerao, R., Sterky, F., Wissel, K., Tandre, K., Aspeborg, H., Moyle, R., Ohmiya, Y., Bhalerao, R., Brunner, A., Gustafsson, P., Karlsson, J., Lundeberg, J., Nilsson, O., Sandberg, G., Strauss, S., Sundberg, B., Uhlen, M., Jansson, S., and Nilsson, P. (2004). A transcriptional timetable of autumn senescence. Genome Biology **5,** R24.

Balazadeh, S., Parlitz, S., Mueller-Roeber, B., and Meyer, R.C. (2008). Natural developmental variations in leaf and plant senescence in *Arabidopsis thaliana*. Plant Biology **10,** 136-147.

Boodley, J.W., and Sheldrake Jr., R. (1977). Cornell peat-lite mixes for commercial plant growing. New York State College Agricultural Life Science Information Bulletin **43,** 8.

Buchanan-Wollaston, V., Earl, S., Harrison, E., Mathas, E., Navabpour, S., Page, T., and Pink, D. (2003). The molecular analysis of leaf senescence - a genomics approach. Plant Biotechnology Journal **1,** 3-22.

Buchanan-Wollaston, V., Page, T., Harrison, E., Breeze, E., Lim, P.O., Nam, H.G., Lin, J.-F., Wu, S.-H., Swidzinski, J., Ishizaki, K., and Leaver, C.J. (2005). Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. The Plant Journal **42,** 567-585.

Gan, S., and Amasino, R.M. (1997). Making sense of senescence (molecular genetic regulation and manipulation of leaf senescence). Plant Physiology **113,** 313-319.

Gong, Y.H., Zhang, J., Gao, J.F., Lu, J.Y., and Wang, J.R. (2005). Slow export of photoassimilate from stay-green leaves during late grain-filling stage in hybrid winter wheat (*Triticum aestivum* L.). Journal of Agronomy and Crop Science **191,** 292-299.

Gregersen, P.L., and Holm, P.B. (2007). Transcriptome analysis of senescence in the flag leaf of wheat (*Triticum aestivum* L.). Plant Biotechnology Journal **5,** 192-206.

Gregersen, P.L., Holm, P.B., and Krupinska, K. (2008). Leaf senescence and nutrient remobilisation in barley and wheat. Plant Biology **10,** 37-49.

Guo, Y., Cai, Z., and Gan, S. (2004). Transcriptome of Arabidopsis leaf senescence. Plant, Cell & Environment **27,** 521-549.

Hafsi, M., Mechmeche, W., Bouamama, L., Djekoune, A., Zaharieva, M., and Monneveux, P. (2000). Flag leaf senescence, as evaluated by numerical image analysis, and its relationship with yield under drought in durum wheat. Journal of Agronomy and Crop Science **185,** 275-280.

Heyne, E.G. (1987). Wheat and Wheat Improvement. (Madison, Wisconsin, USA: American Society of Agronomy, Inc.).

Jukanti, A.K., Heidlebaugh, N.M., Parrott, D.L., Fischer, I.A., McInnerney, K., and Fischer, A.M. (2008). Comparative transcriptome profiling of near-isogenic barley (*Hordeum vulgare*) lines differing in the allelic state of a major grain protein content locus identifies genes with possible roles in leaf senescence and nitrogen reallocation. New Phytologist **177,** 333-349.

Kichey, T., Le Gouis, J., Sangwan, B., Hirel, B., and Dubois, F. (2005). Changes in the cellular and subcellular localization of glutamine synthetase and glutamate dehydrogenase during flag leaf senescence in wheat (*Triticum aestivum* L.). Plant and Cell Physiology **46,** 964-974.

Lee, R.-H., Lin, M.-C., and Chen, S.-C. (2004). A novel alkaline α-galactosidase gene is involved in rice leaf senescence. Plant Molecular Biology **55,** 281-295.

Lim, P.O., Kim, H.J., and Gil Nam, H. (2007). Leaf senescence. Annual Review of Plant Biology **58,** 115-136.

Lin, J.-F., and Wu, S.-H. (2004). Molecular events in senescing Arabidopsis leaves. The Plant Journal **39,** 612-628.

Lu, C., Lu, Q., Zhang, J., and Kuang, T. (2001). Characterization of photosynthetic pigment composition, photosystem II photochemistry and thermal energy dissipation during leaf senescence of wheat plants grown in the field. Journal of Experimental Botany **52,** 1805-1810.

Madeira, A.C., and de Varennes, A. (2005). Use of chlorophyll meter to assess the effect of nitrogen on sweet pepper development and growth. Journal of Plant Nutrition **28,** 1133-1144.

Oh, S.A., Park, J.H., Lee, G.I., Paek, K.H., Park, S.K., and Nam, H.G. (1997). Identification of three genetic loci controlling leaf senescence in *Arabidopsis thaliana*. The Plant Journal **12,** 527-535.

Parrott. David, L., McInnerney, K., Feller, U., and Fischer, A., M. (2007). Steam-girdling of barley (*Hordeum vulgare*) leaves leads to carbohydrate accumulation and accelerated leaf senescence, facilitating transcriptomic analysis of senescence-associated genes. New Phytologist **176,** 56-69.

Saulescu, N.N., Ittu, G., and Mustatea, P. (2001). Dark induced senescence as a tool in breeding wheat for optimum senescence pattern. Euphytica **119,** 205-209.

Thompson, J.E., Froese, C.D., Madey, E., Smith, M.D., and Hong, Y. (1998). Lipid metabolism during plant senescence. Progress in Lipid Research **37,** 119-141.

Woo, H.R., Chung, K.M., Park, J.H., Oh, S.A., and Ahn, T. (2001). ORE9, an F-box protein that regulates leaf senescence in Arabidopsis. The Plant Cell **13,** 1779-1790.

Yamamoto, A., Nakamura, T., Adu-Gyamfi, J.J., and Saigusa, M. (2002). Relationship between chlorophyll content in leaves of sorghum and pigeonpea determined by extraction method and by chlorophyll meter (SPAD-502). Journal of Plant Nutrition **25,** 2295-2301.

Zhang, W., Shmulevich, I., and Astola, J. (2004). Microarray quality control. (John Wiley & Sons, Inc.).

Zhao, H., Dai, T., Jing, Q., Jiang, D., and Cao, W. (2007). Leaf senescence and grain filling affected by post-anthesis high temperatures in two different wheat cultivars. Plant Growth Regulation **51,** 149-158.

CHAPTER 3 Annotation improvement of the Affymetrix GeneChip® Wheat Genome Array

3.1 INTRODUCTION

Microarray technology has been extensively used in the last decade for the examination of gene expression in a wide range of plant species (Rensink and Buell, 2005; Galbraith, 2006). Many of these microarrays have been designed for organisms without a complete genome sequence. The Affymetrix GeneChip® Wheat Genome Array (referred to hereafter as the wheat GeneChip) was designed primarily from Unigene clusters of expressed sequence tags (ESTs) from cDNA sequencing projects. Despite the fact that the wheat GeneChip was made commercially available in 2005 and has been widely used for gene expression studies within the wheat research community (Gupta et al., 2008), there has been little improvement to the annotation data for the array. Sufficient annotation information is critical to the biological interpretation of expression results. Controlled vocabularies facilitate the exchange and comparison of biological data. The Gene Ontology (GO, http://www.geneontology.org/) project (Ashburner et al., 2000) provides a controlled vocabulary of terms for describing gene product characteristics and gene annotation data. The GO is one of the most extensively used functional annotation vocabularies for gene and protein sequences, and has become the standard in the field of functional genomics. By using this vocabulary and applying these terms to sequence data, large data sets can easily be sorted, classified, and compared without first trying to determine if sequence descriptions are actually describing the same function or characteristic.

The annotations for the wheat GeneChip provided by Affymetrix® contain GO annotations for only 3.16% of the 61,115 probe sets on the array (not including control and reporter probe sets, annotation release August 11, 2010). While annotations other than GOs are available for the wheat GeneChip, such as the target description, these annotation formats do not utilize a controlled vocabulary and therefore do not easily allow classification of expression results into functional categories.

To enable classification of tissue-specific/enriched gene lists by GO categories and functional enrichment analysis, an improvement in GO annotation of probe sets on the wheat GeneChip is required. While manual annotation has the ability to guarantee a high level of accuracy (Jones et al., 2007), it is very time consuming and is most commonly done only on a final set of candidate probe sets or transcripts. This still prevents the ability to perform functional enrichment studies in which the functions of genes expressed in a particular tissue of interest or under a particular condition are compared to the plant or system as a whole. Another problem with manual annotation is the variability in factors determining the choice of annotation. It is often difficult to tell from manual annotations if a sequence similarity limit was used as a quality control mechanism, or if so, what this cutoff was. In many cases, the cutoff is not set and can vary from one transcript to another. Automated annotation methods have the advantage of applying the same criteria to a large number of sequences.

Using an automated approach to annotate the wheat GeneChip has the potential to improve the percentage of probe sets with GO annotations and due to the nature of these methods could be repeated over time, using the same criteria, as more sequence and annotation information in public databases becomes available for plant species. Other

groups have attempted to improve the GO annotation of the wheat GeneChip, and though improvement over the Affymetrix® annotations was achieved, the number of probe sets assigned GO annotations was still just over one third of the wheat GeneChip (Götz et al., 2011). This is still limited in terms of classification of gene lists and the confidence one would have in functional enrichment analysis results.

To improve the percentage of GO annotations available and address the issues with manual annotation, an automated annotation was performed for the wheat GeneChip. Several automated annotation programs have been created to assign functional annotations with varying characteristics and limitations in terms of high throughput sequence annotation (Götz et al., 2008). We used the Blast2GO (B2G, http://www.blast2go.org/home) application for functional annotation of the Affymetrix GeneChip® wheat genome array. The Blast2Go annotation suite is a bioinformatics tool for the automated functional annotation of sequence data based on sequence homology transfer and using the GO vocabulary (Conesa and Götz, 2008; Götz et al., 2008). The Blast2Go suite has high throughput capabilities, incorporates GO evidence codes (EC) to assign annotation quality and includes annotation statistics, visualization and data mining tools (Götz et al., 2008). Using the Basic Local Alignment Search Tool (BLAST, Altschul et al., 1990) homologous sequences are identified and functional information is transferred to the sequences represented on the wheat GeneChip. Using different query sequence information can return different blast results which can in turn result in variation in annotation information. The automated annotation was carried out using five separate methods of homologous sequence identification and B2G annotation to compare their ability to improve the GO annotation levels of the wheat GeneChip. The results of

each of these methods were also merged to provide the maximum number of GO annotations that can be used for GO classification of gene lists and functional enrichment analysis with a reasonable level of confidence. The steps for each of the methods are outlined within and can be repeated in the future as annotation information of sequence data in the public databases improves. The percentage of GO annotations available for the wheat GeneChip was increased from 3.16% to 76.52% using these methods.

3.2 MATERIALS AND METHODS

Annotations improvement was carried out for the 61,115 probe sets on the Affymetrix GeneChip® wheat genome array. Five separate methods of annotation improvement (referred to as Step 1 through Step 5 throughout) were tested and examined to determine the method that would provide the most comprehensive annotation set for the wheat GeneChip. All BLASTx steps were performed using the NCBI non redundant protein (nr) database (downloaded Feb 11, 2011). Blast results (in .xml format) were then imported into Blast2GO for the remaining procedures in the annotation processes.

3.2.1 Annotation step 1

The functional annotation of the wheat GeneChip are available in the B2G-FAR repository (http://bioinfo.cipf.es/b2gfar/affyannot2010, (Götz et al., 2011). The annotation data provided were obtained using the wheat GeneChip target sequences. Target sequences are the portion of the Consensus or Exemplar sequence from which the probe sequences were selected. Consensus/exemplar sequences are the sequences used at the time of design to represent the transcript that the GeneChip' probe set measures. A consensus sequence results from base-calling algorithms that align and combine sequence data into groups. An exemplar sequence is a representative cDNA sequence for each gene (http://www.affymetrix.com/estore/support/help/IVT_glossary/index.jsp, hereafter consensus/exemplar sequences are referred to collectively as consensus sequences). The annotation of the target sequences in the B2G-FAR repository are missing BLAST results for 29,189 of the 61,115 (47.76%) target sequences. To be sure that these probe sets could not be annotated using the same methods, the functional annotation was repeated in

our hands. BLASTx of the target sequences against the nr database was performed using an expectation value (*e*-value) cutoff of 1*e*-3 and 20 hits were retrieved per query. These blast parameters were permissive to ensure that a large amount of information was retrieved from the blast step which could then be filtered using more stringent conditions in later steps of the annotation process (Conesa and Götz, 2008). BLAST results were mapped and annotated using default B2G parameters which are an e-value filter of 1*e*-6, default gradual evidence code (EC) weights, a GO weight of 5 and an annotation threshold of 55. Therefore only sequences with a blast *e*-value lower than 1*e*-6 were considered in the annotation formula, the query-hit similarity value adjusted by the EC weight of the source GO term had to be at least 55 and the annotation of a parent node was possible if multiple child nodes existed that did not reach the annotation threshold. These values optimize the ratio between annotation coverage and annotation accuracy (Conesa and Götz, 2008).

3.2.2 Annotation step 2

BLASTx of the wheat GeneChip consensus sequences (61,115 total) against the nr database was performed using an e-value cutoff of 1*e*-3 and a maximum of 20 hits retrieved per query. Mapping and annotation were carried out in B2G using default parameters.

3.2.3 Annotation step 3

To retrieve homologous sequences that would extend the sequence length used for annotation, a BLASTn with the wheat consensus sequences was performed against a full

length triticale cDNA sequence reference set containing a total of 213,671 assembled elements from 454 sequence data (Xu et al., 2011). An *e*-value cutoff of 1*e*-5 was used for the BLASTn and a maximum of 20 hits were retrieved per query. The top hit for each query was subsequently used in a BLASTx of the nr database using an *e*-value cutoff of 1*e*-5 and 20 hits retrieved per query. Mapping and annotation were performed in B2G using default parameters.

3.2.4 Annotation step 4

Sequences for the representative public IDs (the accession number of the representative sequence on which the probe set is based) for the wheat GeneChip were retrieved from the public databases (GenBank, dbEST or RefSeq) in fasta format and used for annotation improvement. Only representative public ID sequences with a total sequence length less than 5 kb were used in order to avoid BAC clones and genomic DNA sequences that may contain multiple genes. The 8 representative public ID sequences greater than 5 kb (representing 32 probe sets) are shown in Table 3.1. An additional 35 representative public ID sequences representing 36 probe sets were also not included in the analysis since their records have been removed from the public databases (also shown in Table 3.1). The remaining representative public ID sequences (representing 61,047 probe sets) were used in a BLASTx against the nr database with an *e*-value cutoff of 1*e*-3. 20 hits were retrieved per representative public ID sequence. Mapping and annotation were carried out in B2G using default parameters.

Table 3.1 Probe sets excluded from annotation steps 4 and 5. Probe sets and their representative Public IDs not included in annotation improvement steps 4 and 5 due to a sequence length greater than 5 kb or removal of their records from public databases.

Probeset ID	Representative ID*	Molecule Type	Sequence Length (bp)	Representative ID Sequence Description**
TaAffx.143995.16.A1_at TaAffx.143995.16.S1_s_at TaAffx.143995.8.A1_at TaAffx.143995.14.S1_at TaAffx.143995.20.A1_at TaAffx.143995.7.S1_at TaAffx.143995.18.A1_at TaAffx.143995.7.A1_s_at TaAffx.143995.21.S1_s_at TaAffx.143995.8.S1_at TaAffx.143995.7.A1_at TaAffx.143995.22.S1_at TaAffx.143995.18.S1_at TaAffx.143995.20.S1_s_at TaAffx.143995.21.A1_at TaAffx.143995.14.A1_at TaAffx.143995.12.S1_at TaAffx.143995.22.A1_at TaAffx.143995.12.A1_at	AY485644	genomic DNA	438,828	Triticum monococcum phosphatidylserine decarboxylase, ZCCT2, ZCCT1, and SNF2P genes, complete cds; nucellin pseudogene, complete sequence; putative transposase, phosphatidylinositol phosphatidylcholine transfer protein sec14 cytosolic-like protein, and phytochrome P450-like protein genes, complete cds; and unknown genes.
TaAffx.143995.17.A1_at TaAffx.143995.17.S1_s_at	AY188331	BAC clone/ genomic DNA	133,625	Triticum monococcum DV92 chromosome 5AL BAC 231A16, complete sequence.
TaAffx.143995.23.S1_at TaAffx.143995.23.A1_at TaAffx.143995.13.A1_at TaAffx.143995.13.S1_at	AY188332	BAC clone/ genomic DNA	95,541	Triticum monococcum chromosome 5AL clone BAC 609E6, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
TaAffx.143995.19.A1_at TaAffx.143995.19.S1_at	AY188333	BAC clone/ genomic DNA	112,328	Triticum monococcum chromosome 5AL clone BAC 719C13, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
TaAffx.143995.6.S1_at TaAffx.143995.6.A1_at	AY491681	BAC clone/ genomic DNA	101,101	Triticum monococcum DV92 BAC 109N23, complete sequence.
Ta.187.1.S1_at	U10187	mRNA	7,360	Triticum aestivum cytosolic acetyl-CoA carboxylase mRNA, complete cds.
Ta.250.1.S1_at	AF258608	mRNA	5,346	Triticum aestivum starch synthase III mRNA, complete cds.
Ta.114.1.S1_at	AF029895	mRNA	6,993	Triticum aestivum acetyl-coenzyme A carboxylase mRNA, complete cds, nuclear gene encoding plastid protein.
Ta.951.1.A1_at	BQ280449	N/A	N/A	Record removed
Ta.13701.1.S1_at	BQ280458	N/A	N/A	Record removed
Ta.5480.2.S1_at	BQ280758	N/A	N/A	Record removed
Ta.26339.1.A1_at	BQ280779	N/A	N/A	Record removed
Ta.8558.1.S1_at	BQ280941	N/A	N/A	Record removed
Ta.561.2.S1_x_at	BQ280951	N/A	N/A	Record removed
Ta.28062.1.S1_s_at	BQ280971	N/A	N/A	Record removed
Ta.12872.1.A1_at	BQ280973	N/A	N/A	Record removed
TaAffx.131611.1.S1_at	BQ281041	N/A	N/A	Record removed
Ta.13706.1.S1_at	BQ281222	N/A	N/A	Record removed
Ta.12910.1.S1_at	BQ281259	N/A	N/A	Record removed
Ta.12912.1.S1_at	BQ281262	N/A	N/A	Record removed
Ta.4595.1.S1_at	BQ281281	N/A	N/A	Record removed
Ta.12939.1.S1_at	BQ281406	N/A	N/A	Record removed
Ta.12946.1.S1_at	BQ281443	N/A	N/A	Record removed
Ta.12949.2.A1_a_at	BQ281458	N/A	N/A	Record removed
Ta.7742.1.S1_at	BQ281483	N/A	N/A	Record removed
Ta.12911.1.S1_at	BQ281484	N/A	N/A	Record removed
Ta.29627.1.A1_at	BQ281547	N/A	N/A	Record removed
Ta.1677.1.S1_at	BQ281548	N/A	N/A	Record removed
Ta.8512.2.S1_x_at	BQ281568	N/A	N/A	Record removed
Ta.12974.1.S1_at	BQ281614	N/A	N/A	Record removed
Ta.22550.1.A1_at	BQ281632	N/A	N/A	Record removed
TaAffx.97428.1.S1_at	BQ281681	N/A	N/A	Record removed
Ta.12992.1.S1_at	BQ281717	N/A	N/A	Record removed
TaAffx.4392.1.A1_at	BQ281739	N/A	N/A	Record removed
Ta.12887.1.S1_at	BQ281752	N/A	N/A	Record removed
Ta.12999.1.S1_at	BQ281770	N/A	N/A	Record removed
Ta.13004.1.S1_at	BQ281790	N/A	N/A	Record removed
Ta.13007.1.S1_at	BQ281799	N/A	N/A	Record removed
Ta.1618.1.S1_s_at	BQ281801	N/A	N/A	Record removed
Ta.7964.1.A1_at	BQ281876	N/A	N/A	Record removed
Ta.1660.1.A1_at	BQ282004	N/A	N/A	Record removed
Ta.24712.2.S1_at Ta.24712.2.S1_x_at	BQ282055	N/A	N/A	Record removed
Ta.3292.1.S1_a_at	BQ282081	N/A	N/A	Record removed

*Representative Public IDs are GenBank (NCBI) accession numbers

**Representative Public ID sequence descriptions are the definition for the accession in GenBank.

3.2.5 Annotation step 5

In another effort to retrieve homologous sequences with increased sequence length for retrieval of additional annotation information, a BLASTn of the representative public ID sequences against the NCBI nucleotide database with an *e*-value cutoff of 1*e*-5 and a maximum of 5 hits per query was performed. The top hit for each query was retrieved as a fasta formatted sequence and subsequently used in a BLASTx against the nr database with an *e*-value cutoff of 1*e*-5 and 20 hits retrieved per query. Mapping and annotation were carried out in B2G using default parameters.

3.2.6 Annotation of sequences not successfully annotated using an annotation threshold of 55

To retrieve a higher amount of annotation information, sequences with GO associations from mapping but unable to be annotated at a threshold of 55 were selected and annotation was performed using a reduced threshold of 45 as recommended by the developers of B2G (International Course in Automated Functional Annotation and Data Mining, Orlando, Florida, USA, November 16-18, 2010).

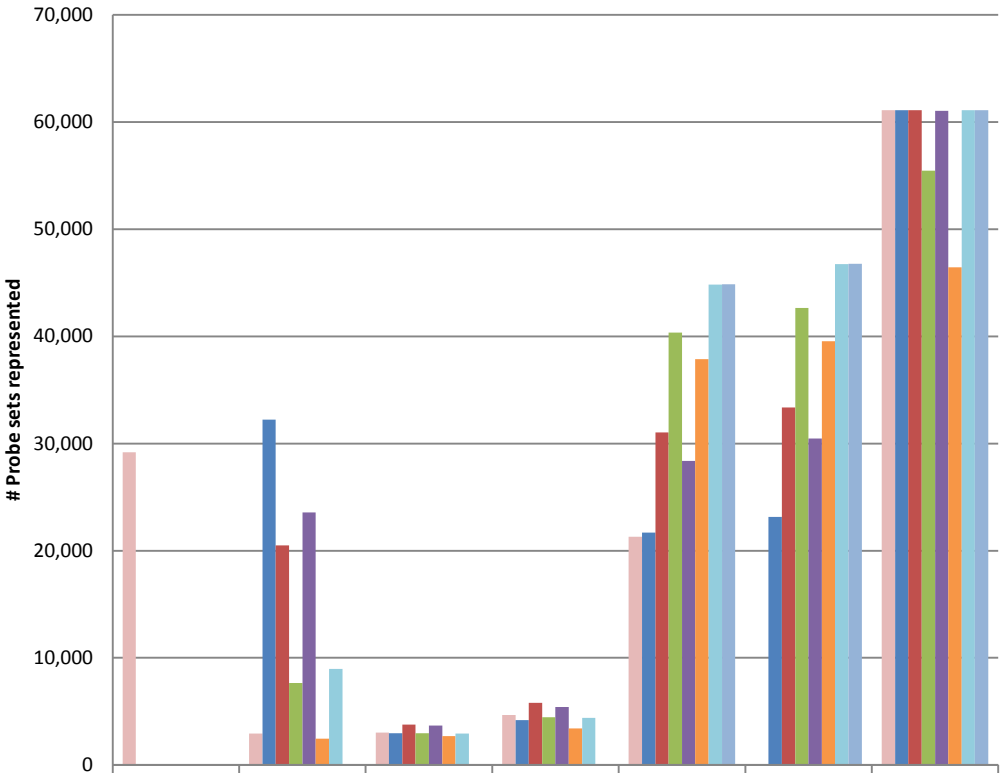
3.2.7 Annotation merger

In all steps the original probe set ID was maintained in the sequence identifier to allow the merging of the annotation results. From steps 1-5, annotations were exported from B2G (as .annot files) and then merged in a separate project file in B2G. Annotations produced at a threshold of 55 from steps 1-5 were merged together followed by merging the annotations produced at a threshold of 45 to these to provide the highest

percentage of GO annotations for the wheat GeneChip. Finally this set of annotations was merged with the GO annotations provided by Affymetrix® for the wheat GeneChip (annotation release 31, August 11, 2010).

3.3 RESULTS

The annotations for the wheat GeneChip provided by Affymetrix® contain gene ontology (GO) annotations for only 3.16% of the 61,115 wheat probe sets (annotation release August 11, 2010). Each of the five annotation methods used here were able to improve upon this percentage by over 34%. The resulting data distribution from all annotation steps is shown in Figure 3.1. All annotations from steps 1-5 at annotation thresholds 55 and 45 are provided in supplementary Table S3.1. This table also contains the target descriptions, representative public IDs and GO annotations provided in the Affymetrix GeneChip® wheat genome array annotation file (release Aug 11, 2010) and the results of the annotation merger and their conversion to the GOslim vocabulary (generic). From this table users can retrieve annotation information for any candidate probe set from the wheat GeneChip. Annotations from all five annotation methods are available in the table for comparison along with the blast and mapping results information from each step. Annotations are provided as GO terms as well as GO ID numbers and the combined annotations from steps 1-5 are also provided separated into the three GO domains; biological process, molecular function and cellular component. The annotations have also been converted to GOslim (generic) to allow classification using broader categories. The final annotation set from the merger of steps 1-5 at a threshold of 55 with those at a threshold of 45 and the GO annotations from Affymetrix® are also provided in Supplementary Table S3.2 in a format that can be imported into B2G for data mining and functional enrichment analysis. These resources will allow users of the wheat GeneChip to quickly assign function to candidate genes identified in expression studies and functionally classify gene lists. By using the provided annotation



	A) Without Blast	B) No Blast Hits	C) No Mapping	D) No Annot	E) Annot 55	F) Annot 55/45	G) Total
B2G-Far	29,189	2,939	3,013	4,660	21,314		61,115
Step 1	0	32,233	2,976	4,203	21,703	23,154	61,115
Step 2	0	20,495	3,777	5,810	31,033	33,377	61,115
Step 3	0	7,667	2,974	4,462	40,340	42,662	55,443
Step 4	0	23,582	3,670	5,409	28,386	30,465	61,047
Step 5	0	2,454	2,695	3,411	37,867	39,530	46,427
Steps 1-5 Merged	0	8,975	2,926	4,391	44,823	46,747	61,115
Steps 1-5 Merged with Affymetrix					44,843	46,765	61,115

Figure 3.1 Automated annotation results distribution for B2G-Far, steps 1-5 and Affymetrix® annotations. No Mapping values correspond to the number of probe sets that had Blastx hits but no GO associations to those Blastx hits. No Annot values correspond to the number of probe sets that had GO associations to the Blastx hits retrieved during mapping but that did not pass the annotation threshold of 55. Blank cells could not be calculated because this information is not available from Affymetrix® or B2G-Far.

set, multiple gene lists can be compared to identify functional enrichment that may be the cause of phenotypic differences. It is our hope that these annotation resources will expedite the understanding of biological meaning within expression results.

3.3.1 Step 1; repeat of the B2G-Far wheat GeneChip annotation

The results of the wheat GeneChip Blast2Go-Far annotation effort are found on the B2G-Far website (http://bioinfo.cipf.es/b2gfar/arrayannot:wheatx, March 21, 2011, Götz et al., 2011). From the efforts of B2G-Far, 34.93% of the sequences on the wheat chip were successfully annotated. This is a significant increase from the GO annotations provided by the Affymetrix® annotation release but still left over 65% of the probe sets without a GO annotation, making functional comparisons among gene groups problematic. The B2G-Far annotation resulted in 47.7% of probe set target sequences with no BLAST result (29,189 probe sets, Figure 3.1 A). In our repeat of the BLASTx and annotation steps, a blast result of 'no hit' against sequences in the nr database was obtained for the majority of sequences without a BLAST result in the B2G-Far repository. Our repetition of the annotation resulted in all 61,115 probe sets on the wheat GeneChip having a BLAST result (including BLAST results of 'no hit' for 32,233 probe set target sequences, Figure 3.1 B). In our study the number of probe sets without mapping or annotations was slightly reduced (in comparison to the annotation performed by B2G-Far), and the number of probe sets with functional annotations was slightly increased by 0.58% (353 more probe sets) to a total of 21,703 probe sets having a functional annotation (Figure 3.1, E). These slight differences are likely due to differences in the public database (nr) during the time between annotations performed by

B2G (Aug, 6, 2010) and this study (Feb. 11, 2011). Reducing the annotation threshold to 45 for un-annotated probe sets resulted in an additional 1,451 probe sets (2.37%) with functional annotations (Figure 3.1, F).

3.3.2 Step 2; annotation using the wheat GeneChip consensus sequences.

The sequence length distribution of the Affymetrix® wheat chip consensus sequences is shown in Figure 3.2. On average, the wheat consensus sequence lengths are twice as long as the target sequences, which resulted in 19.2% fewer sequences without BLAST hits (Figure 3.1, B). Of the 61,115 sequences queried against the nr database, 20,495 sequences (33.54%) had no sequence alignments (singletons, Figure 3.1, B) while 40,620 sequences (66.46%) had at least one BLAST hit (Figure 3.1, C+D+E). Of these, 36,843 (60.28%, Figure 3.1, D+E) of the query sequences had at least one blast hit that was associated with one or more GO terms that were retrieved during mapping and 3,777 sequences had no GO associations (Figure 3.1, C). Annotation at the default threshold of 55 resulted in 31,033 annotated sequences (50.78% of the sequences, Figure 3.1, E) while 5,810 mapped sequences (9.51%) had no GO annotation (Figure 3.1, D). Reducing the annotation threshold to 45 for un-annotated sequences resulted in an additional 2,344 probe sets (3.84%, Figure 3.1, F-E) with functional annotations.

3.3.3 Step 3; annotation using homologous sequence search to increase sequence length

The wheat consensus sequences were used to query an in-house database of triticale (x *Triticosecale* Wittm.) full length cDNA (TritFLcDNA) sequences for

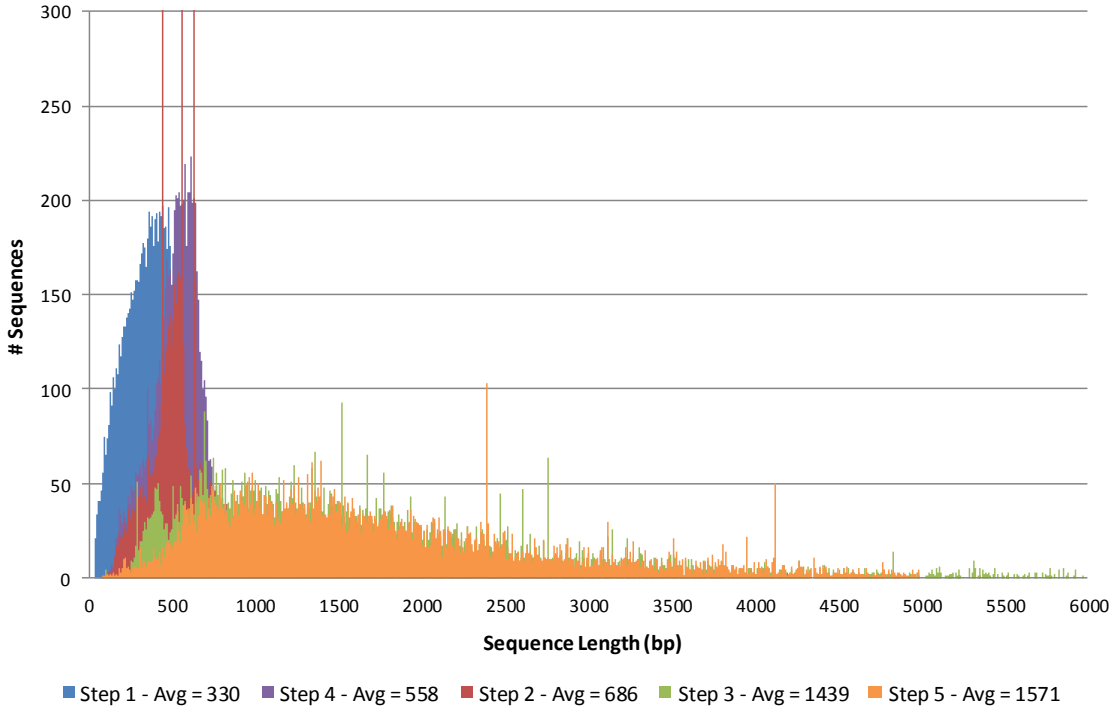


Figure 3.2 Sequence length distribution for sequences used for annotation improvement in steps 1-5. Sequences greater than 600 bp, and frequencies (# sequences) greater than 300 are not included in the image for visualization purposes.

homologues with a sequence length greater than the query consensus sequence. Triticale is a hybrid species that contains the A and B genomes from wheat and the R genome from rye and therefore should contain many wheat homologues. A total of 55,443 wheat consensus sequences had homology to a TritFLcDNA sequence when compared using BLASTn with an *e*-value cutoff of 1*e*-5. The TritFLcDNA sequences have an average length of 1,439 bp, which is considerably longer than the average length of the wheat consensus sequences (686 bp, See Figure 3.2). The longer TritFLcDNA sequences were then used to retrieve annotations using BLASTx and B2G. Of the 55,443 sequences, 47,776 sequences (Figure 3.1, C+D+E) had a BLASTx hit when compared to the nr database and 44,802 sequences (Figure 3.1, D+E) had at least one BLAST hit associated with one or more GO terms. The use of longer sequences was more advantageous for retrieval of additional GO associations and resulted in the highest percentage of annotated sequences of the 5 annotation strategies. Reducing the annotation threshold to 45 for un-annotated sequences resulted in an additional 2,322 probe sets (3.80%) with functional annotations (Figure 3.1, F-E).

3.3.4 Step 4; using representative sequences to retrieve annotations

Another strategy for using longer sequences to gain an increase in annotation data was to use the representative public ID sequences for each probe set on the wheat GeneChip. These sequences are generally longer than the wheat target sequences but are not as long as the wheat consensus sequences (See Figure 3.2). A BLASTx of these sequences was performed against the nr database and the results used to retrieve annotation data using Blast2GO. Eight of the representative public ID sequences

(relating to 32 probe sets) were over 5 kb and were omitted from the annotation process to avoid transference of annotations from adjacent genes. These sequences included genomic DNA fragments (primarily BAC clones) and a few mRNA sequences over 5 kb (Table 3.1). Another 35 representative ID sequences (relating to 36 probe sets) have been removed from the public database since the creation of the wheat chip and were not included in the annotation step. Of the remaining 61,047 sequences queried, 37,465 sequences (Figure 3.1, C+D+E) had at least one BLASTx hit and 33,795 were associated with at least one GO term (Figure 3.1, D+E). 28,386 probe sets (46.45% of the probe sets on the wheat chip, Figure 3.1, E) were assigned a functional annotation using the representative public ID sequences and B2G default annotation threshold of 55. Reducing the annotation threshold to 45 for un-annotated sequences resulted in an additional 2079 probe sets (3.40%) with functional annotations (Figure 3.1, F-E).

3.3.5 Step 5; using representative sequences to find homologues with an increased length for annotation

Many of the representative public ID sequences are ESTs and therefore still have a limited sequence length. In fact, 22,275 (36.45%) probe sets have representative public ID sequences that have a length less than 500 bp. To improve annotation success by using sequences with increased length, homologues to the representative public ID sequences were retrieved by performing a BLASTn against NCBI nucleotide database . A total of 46,427 representative public ID sequences had homologues in the nucleotide database that were longer than the query sequence but less than 5 kb. The distribution of these sequence lengths are shown in Figure 3.2. These sequences were used for functional

annotation of the wheat GeneChip using Blast2Go. I identified that 43,973 of the homologous sequences (Figure. 3.1, C+D+E) had at least one BLASTx hit when compared to the NCBI nr database and 41,278 could be associated with at least one GO (Figure 3.1, D+E). Following annotation, 37,867 probe sets (61.96% of probe sets on the wheat GeneChip, Figure 3.1, E) were assigned a functional annotation using the homologues of the representative IDs and the default annotation threshold of 55. Reducing the annotation threshold to 45 for un-annotated sequences resulted in an additional 1,663 probe sets (2.72%) with functional annotations (Figure 3.1, F-E).

3.3.6 The majority of Blast hits met the required parameters

Figure 3.3 shows the BLAST results *e*-value distribution (A) and sequence similarity distribution (B). In the steps that employed multiple BLASTs (steps 3 and 5) the *e*-value cutoff was reduced in the BLASTn step in order to increase the quality of the BLASTn hit and to ensure that sequences with a reasonable level of homology were being used to provide annotations and to avoid an excess of erroneous *cis*-annotation that might occur with less stringent parameters. In all steps the majority of BLAST hits had an *e*-value that was well below the cutoff. The sequence similarity distribution graphs (Figure 3.3 B) show that for all annotation methods employed, the majority of sequences had blast similarity values of 50-60% or higher. This distribution suggests that using the annotation threshold of 55 was adequate in most cases since the B2G annotation rule is largely dependent on sequence similarity (Conesa and Götz, 2008).

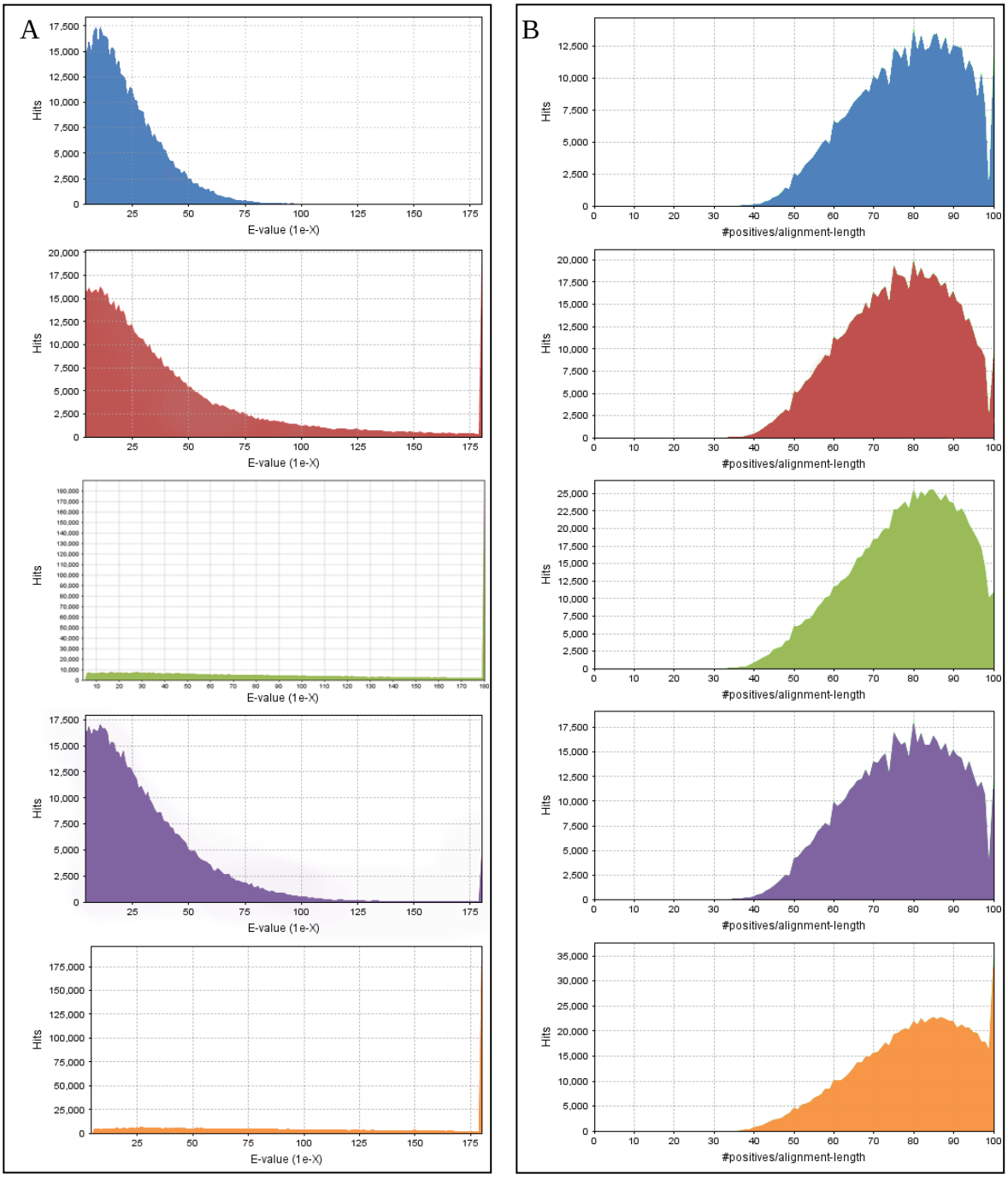


Figure 3.3 Blast results statistics for steps 1-5 (top to bottom respectively). A) *e*-value distribution for blast hits. B) sequence similarity distribution of blast hits. Note the differences in scale of y axes.

3.3.7 Increasing sequence length through homologue identification improves annotation percentages

One of the biggest factors in increasing the percentage of probe sets on the wheat chip with a functional annotation was the use of homologous sequences with an increased sequence length. The two annotation methods which employed the use of homologous sequences identified through BLASTn searches (steps 3 and 5) produced the highest percentages of probe sets on the wheat chip with functional annotations (Figure 3.1). The correlation between average sequence length and the percentage of the probe sets that were annotated in each step is shown in Figure 3.4. The relationship between average sequence length and percentage of probe sets annotated is positive and has a near linear ($R^2 = 0.989$) phase between average sequence lengths of 300 to 1,600 bp. The ideal situation for annotation would be to have the complete coding sequences for all transcripts represented by the probe sets on the wheat GeneChip.

3.3.8 Annotation methods reveal sequence conservation among plants and *Poaceae*

The species distribution of blast hits for all five blast methods employed is shown in Figure 3.5. For all five methods the primary 5 blast hit species were consistent. These were *Oryza sativa* (rice), *Zea mays* (corn), *Sorghum bicolor* (sorghum), *Vitis vinifera* (common grape vine) and *Arabidopsis thaliana*. The top three blast hit species are model plants belonging to the same family as wheat (*Poaceae*). Over 85% of all blast hits in all steps were to known plant species. This blast hit distribution provides confidence to the annotation methods as it was expected that the majority of sequence homology would be to plant species. The majority of blast hits are to plant species that are used as model

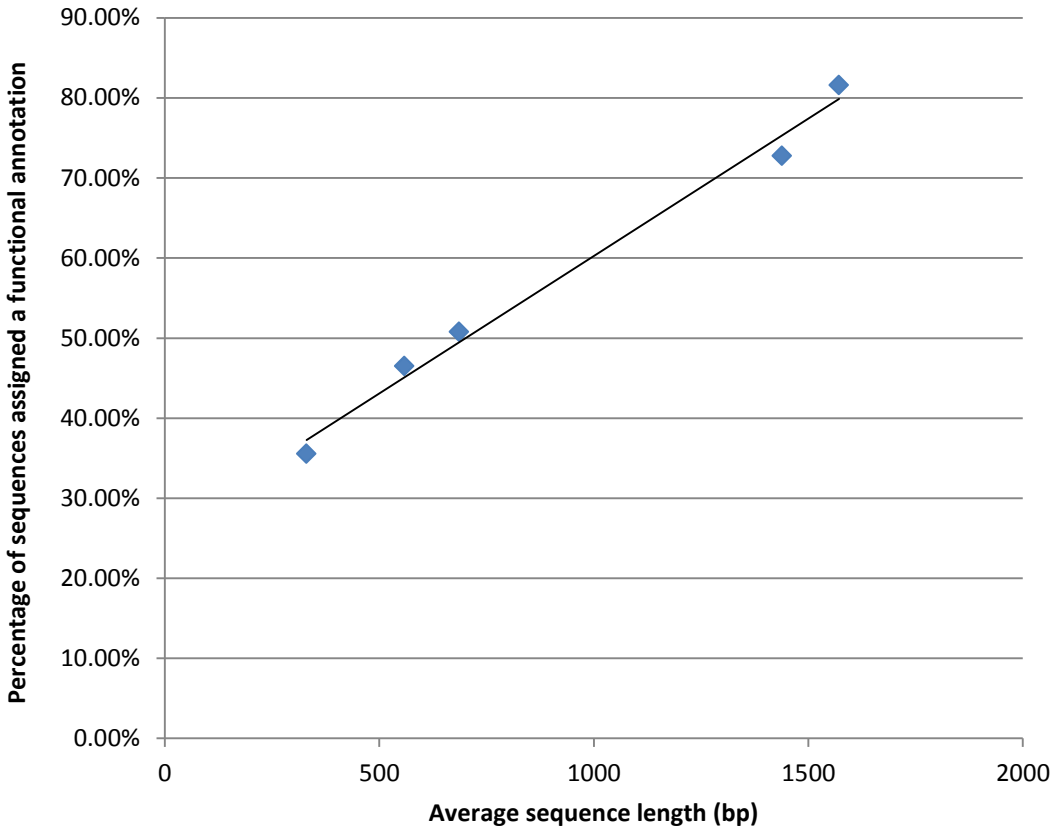


Figure 3.4 Correlation between average sequence length (bp) and the percentage of sequences that can be assigned functional annotations using Blast2Go.

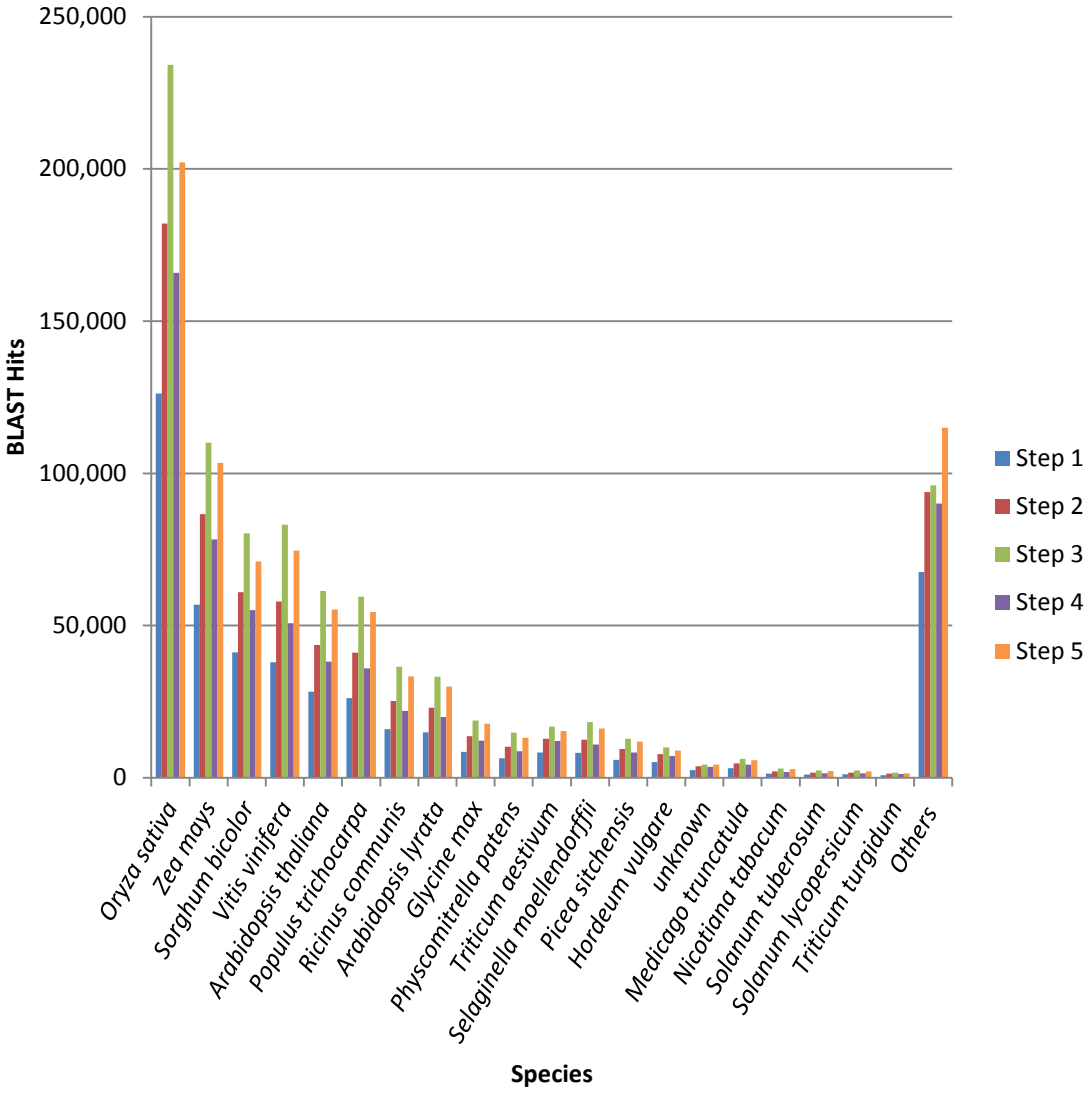


Figure 3.5 Species distribution for blast hits for annotation steps 1-5.

organisms, which emphasizes the disparity in annotated protein sequence data between model and non-model plant species in the nr database. This trend also highlights the importance of the annotation and characterization of sequences from model species for the application of knowledge to non-model species such as wheat. When looking at the species from which only the top blast hits are derived, the distribution favors cereal species as the predominant sources for the best blast hits (Figure 3.6). *Oryza sativa*, *Sorghum bicolor*, *Zea mays*, *Triticum aestivum* and *Hordeum vulgare* are the primary species to which sequence homology was found. The fact that these top blast hits are primarily from the grasses (*Poaceae*) provides an increased confidence in the sequence description assigned during the annotation processes and is indicative of the high level of sequence and functional conservation shown to be among species in the grass family (Campbell et al., 2007). While the majority of blast hits were to model plant species, the model plant *Brachypodium distatichyon* was only minimally represented (less than 0.3% of top blast hits for each step, included in others, Figure 3.6). The majority of Brachypodium protein sequences added to the nr database were added after the date of the nr database download (Feb 11, 2011) used in this study (25,292 out of 25,821 protein entries with a modification date after Feb 11, 2011). There are continually new submissions to public databases and it should be noted that repeating the procedures described within has the potential to further improve annotations as newly annotated sequence information becomes available.

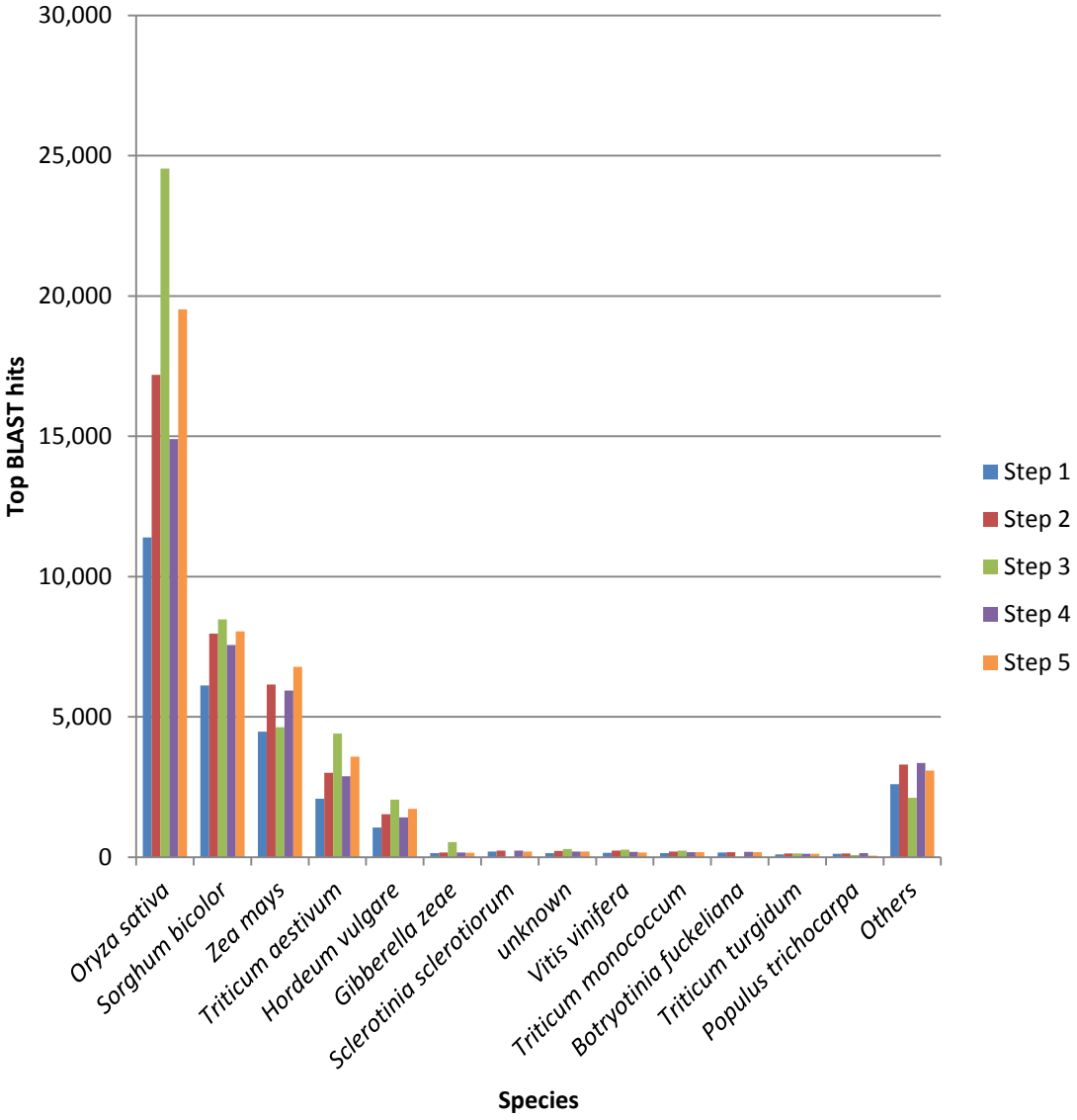


Figure 3.6 Species distribution for top blast hits generated for annotation steps 1-5.

3.3.9 The majority of annotations are from electronic annotation resources

For all annotation methods the majority of GOs were retrieved from the UniProtKB, TAIR and GR_protein databases (Figure 3.7). The majority of annotations (over 85%) in steps 1 to 5 were retrieved from the UniProt Knowledgebase (UniProtKB, http://www.uniprot.org/help/uniprotkb). The proportions of annotations that come from this database highlight the importance of having a central hub for the collection of functional data. The UniProtKB database consists of a manually reviewed section (UniProtKB/Swiss-Prot) and an un-reviewed computationally analyzed section (UniProtKB/TrEMBL, Schneider et al., 2009). The reviewed UniProtKB/Swiss-Prot section accounts for only 3.3% of all sequences in the UniProtKB database (Release 2011_06 of May 31st, 2011). Over 87% of annotations in all of steps 1-5 were inferred from electronic annotation (IEA). The EC distribution highlights the importance of including electronic annotations in the annotation method as they are the primary source of retrieved GOs (Figure 3.8). After IEA, inferred from reviewed computational analysis (RCA), inferred from direct assay (IDA) and inferred from sequence or structural similarity (ISS) were the next most common ECs respectively (Figure 3.8). Not including the IEA category of annotation resources would drastically reduce the percentage of probe sets on the wheat GeneChip that would be assigned functional annotations.

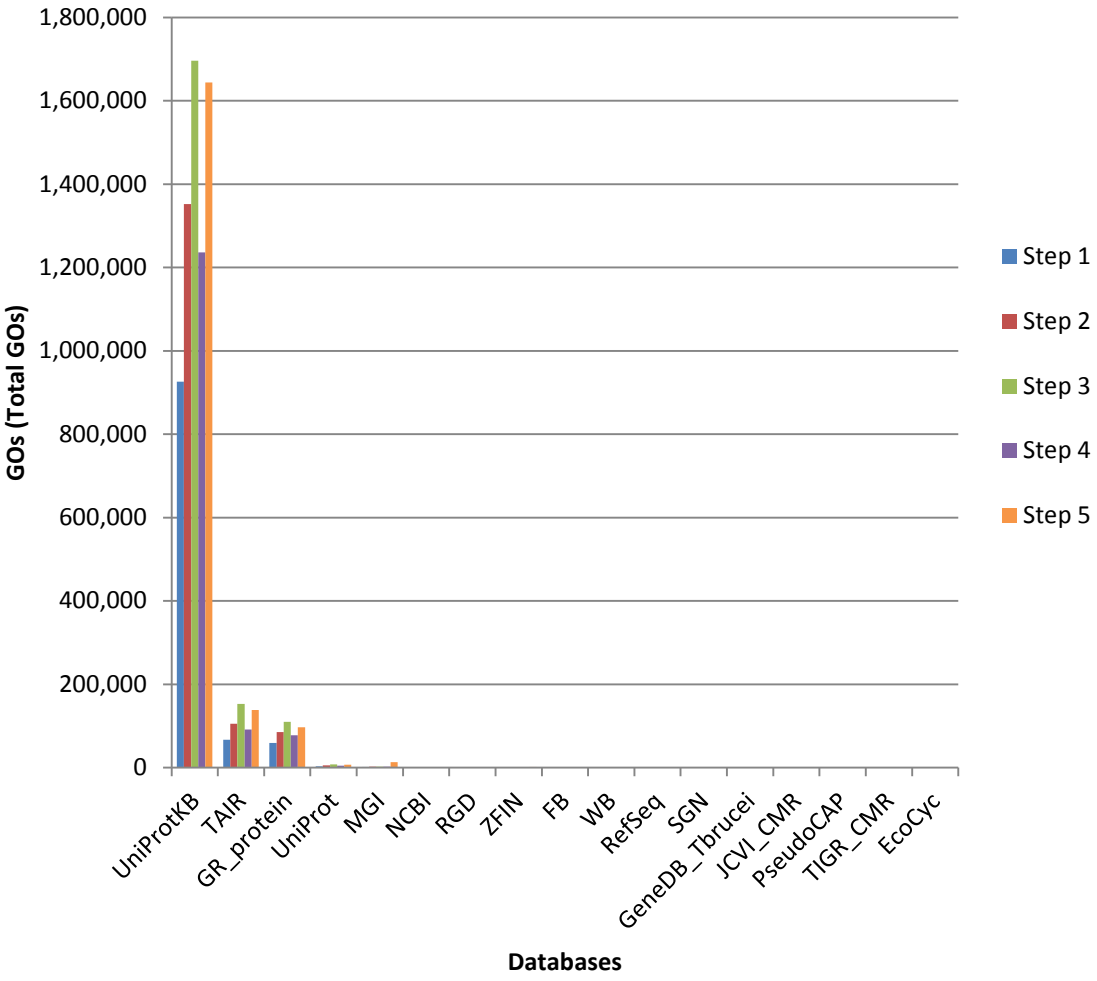


Figure 3.7 Mapping database sources for annotation steps 1-5. Database source sites are as follows; UniProtKB (http://www.uniprot.org/), TAIR (http://www.arabidopsis.org/), GR_Protein (http://www.gramene.org), UniProt (http://www.uniprot.org/), MGI (http://www.informatics.jax.org/function.shtml), NCBI (http://www.ncbi.nlm.nih.gov/), RGD (http://rgd.mcw.edu/), ZFIN (http://zfin.org/cgi-bin/webdriver?MIval=aa-ZDB_home.apg), FB (http://flybase.org/), WB (http://wormbase.sanger.ac.uk/), RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/), SGN (http://solgenomics.net/), GeneDB_Tbrucei (http://www.genedb.org/Homepage), JCVI_CMR (http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi), PseudoCAP (http://www.pseudomonas.com/), TIGR_CMR (http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi), EcoCyc (http://ecocyc.org/).

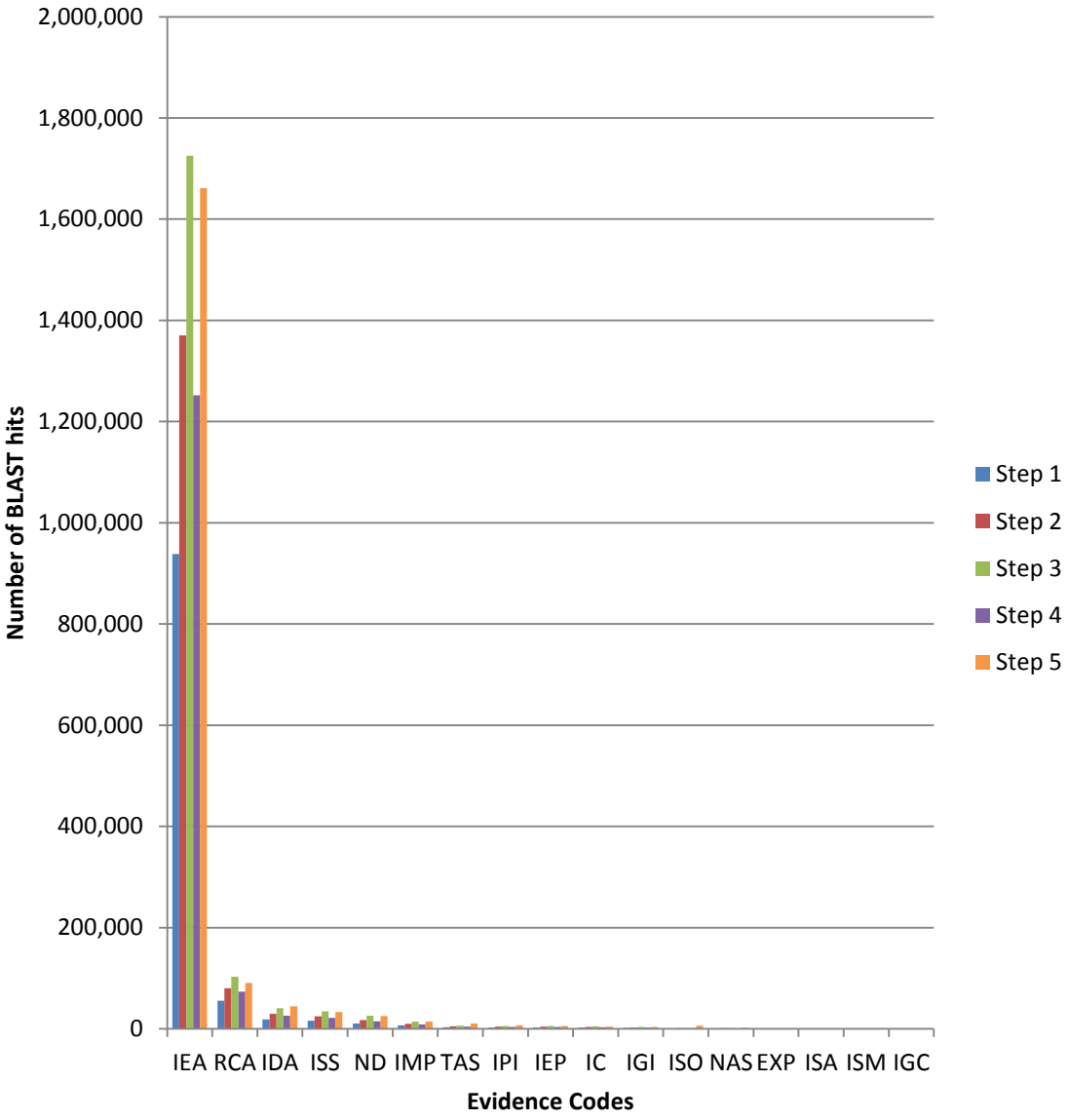


Figure 3.8 Evidence Code distribution of gene ontologies retrieved during mapping in the annotation procedures performed in steps 1-5. IEA; inferred from electronic annotation (0.7), RCA; inferred from reviewed computational analysis (0.9), IDA; inferred from direct assay (1.0), ISS; inferred from sequence or structural similarity (0.8), ND; no biological data available (0.8), IMP;inferred from mutant phenotype (1.0), TAS; traceable author statement (0.9), IPI; inferred from physical interaction (1.0), IEP; inferred from expression pattern (1.0), IC; inferred by curator (0.9), IGI; inferred from genetic interaction (1.0), ISO; inferred from sequence orthology (0.8), NAS; non-traceable author statement (0.8), EXP; inferred from experiment (1.0), ISA; inferred from sequence alignment (0.8), ISM; inferred from sequence model (0.8), IGC; inferred from genomic context (0.7). Numbers in brackets are the evidence code weights.

3.3.10 There is little variance in depth of annotation among annotation methods used

The goal of any annotation process is to retrieve the most information possible about the sequence in question. In terms of GO, this means finding an appropriate GO association at the deepest possible level within the GO hierarchy. The relationships between GOs in the database are organized in a hierarchy and can be represented by a directed acyclical graph (DAG, The Gene Ontology Consortium, 2001). The hierarchy is organized with one of the three root nodes (biological process - BP, molecular function - MF and cellular component - CC) at the top of the graph and each node below is a child of the parent node above it. The terms get more specialized with each node going down the chart. The GO distribution for each of the three GO domains from steps 1-5 of the wheat GeneChip annotation are shown in Figure 3.9. While all steps have similar patterns of GO level distribution, the total number of annotations retrieved by each method differs, in part due to the differing number of starting sequences. To ensure that one method was not better suited to retrieving annotations at a deeper GO level which would provide more detailed annotation information, the average depth of GO annotations was examined (Figure 3.10). The calculation of the average depth of GO level took into account the total number of GOs retrieved by each method and therefore should be independent from the starting number of sequences used. There was no significant difference (much less than 1 standard deviation) in the depth of GOs retrieved by each annotation method and therefore the five annotation methods used were considered equal in terms of retrieving GOs at the deepest possible level. Without a difference in the depth of annotations

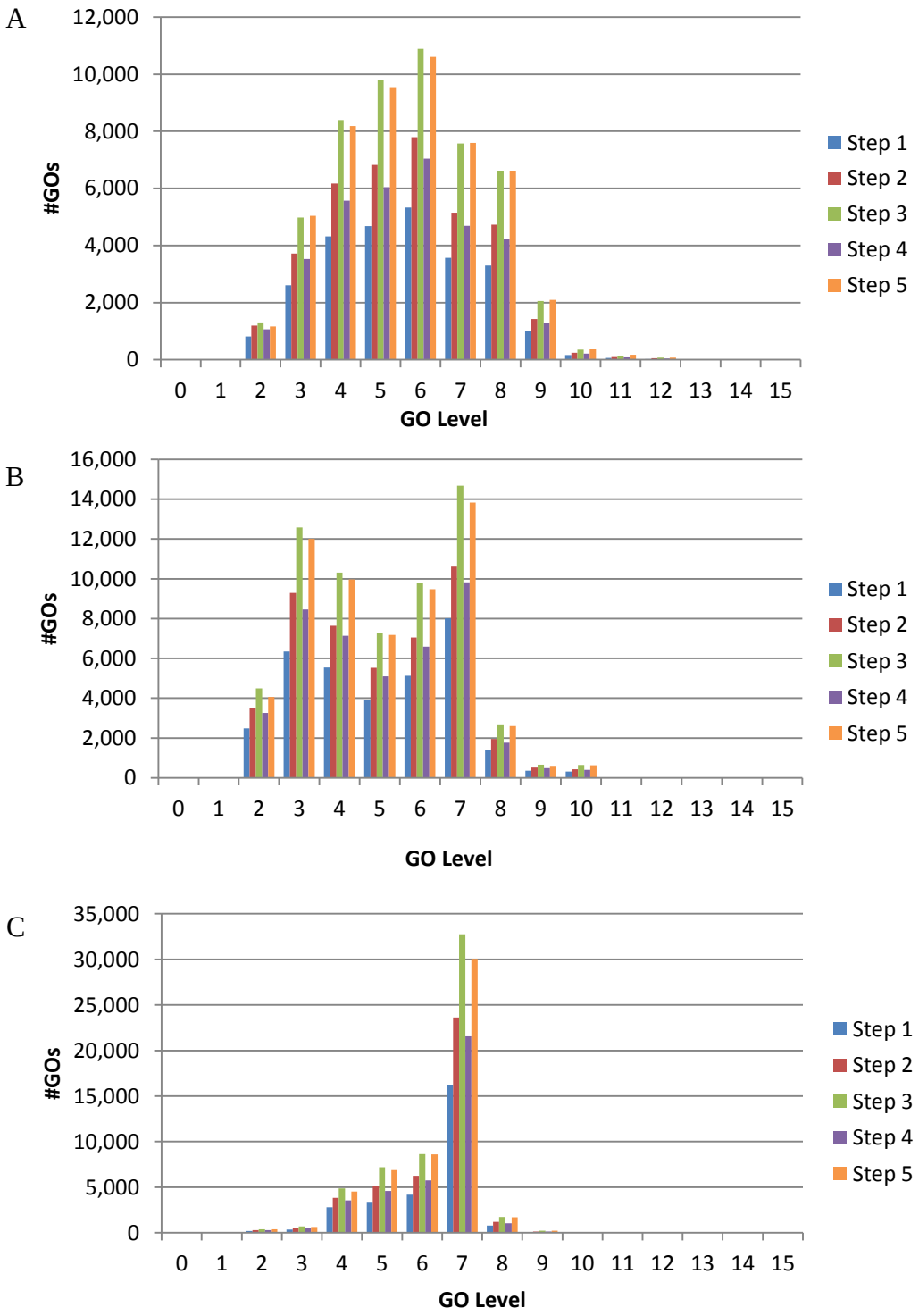
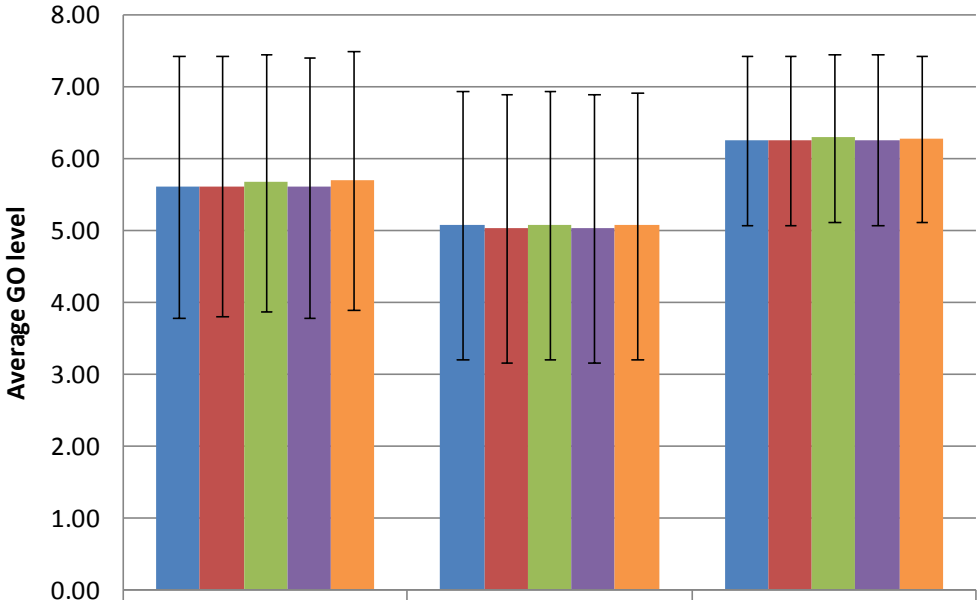


Figure 3.9 GO level distribution for GO annotations applied to probe sets in steps 1-5. A) Biological Process GO level distribution, B) Molecular Function GO level distribution, C) Cellular Component GO level distribution.



	BP	MF	CC
Step 1	5.61	5.08	6.26
Step 2	5.61	5.03	6.26
Step 3	5.67	5.07	6.29
Step 4	5.60	5.04	6.26
Step 5	5.69	5.07	6.28

Figure 3.10 Average GO level for each of the GO domains for annotation processes Steps 1-5. BP; biological process, MF; molecular function, CC; cellular component. Error bars = +/- one standard deviation.

retrieved it is therefore the method that retrieved annotations for the greatest number of probe sets that is considered the preferred method.

3.3.11 Step 3 provided the most annotations for the wheat GeneChip

Of the annotation strategies applied (steps 1-5), step 3 produced the highest number of functionally annotated probe sets for the wheat GeneChip (Fig. 3.1). The use of the full length cDNA sequences from a cereal plant related to wheat (the origins of the A and B genomes in triticale are from wheat) was extremely beneficial to finding protein sequences with annotations that could be transferred to the probe sets on the wheat GeneChip. This also indicates that assembled full length sequence data from wheat genome sequencing projects should be extremely useful to the annotation of wheat resources like the wheat GeneChip when they become available. Following step 3, step 5 produced the next highest number of functionally annotated probe sets for the wheat GeneChip. Step 5 also utilized a two-step blast process in which the first blast was a search for homologous nucleotide sequence with an increased sequence length followed by searching for homologous protein sequences with functional information. Based on our results, maximizing the length of transcribed sequence information is the most important factor in retrieving functional annotation information. A positive relationship between sequence length and ability to annotate has been observed previously using several datasets (Conesa and Götz, 2008) and is confirmed by my results.

3.3.12 Using multiple annotation strategies is beneficial to the annotation of a data set

The annotation strategies were compared to determine if the probe sets that they were able to annotate were similar to or unique from other annotation methods. The overlap and intersection of the annotation methods is presented in Figure 3.11. It was found than no two methods had 100% overlap in terms of annotating the same probe sets (Figure 3.11A). Each annotation strategy (including the Affymetrix® annotations) provided functional annotations for probe sets which could not be annotated by any other method (Figure 3.11B). Therefore to provide the most annotations possible to the wheat GeneChip the annotations from steps 1-5 were merged together. The merged annotations were then combined with the Affymetrix® annotations to provide the highest possible number of annotated probe sets on the GeneChip (44,843 probe sets). This annotation set is provided in Supplementary Table S3.2 in Blast2Go compatible format (.annot) which can be imported into Blast2Go. This file allows users to perform data mining on the annotation set, to easily search for probe sets belonging to specific functional categories, for classification of groups of probe sets and for enrichment analysis within the Blast2Go software to determine functional enrichment present in gene lists that result from expression studies using the wheat GeneChip.

3.3.13 Reducing the annotation threshold increased annotations by over 3%

By reducing the annotation threshold from 55 to 45 for sequences that were not annotated at a threshold of 55, the percent of the probe sets that were successfully annotated was increased by up to 3.84% for each step. The total combined annotations following the

A

	Affy.		Step 1		Step 2		Step 3		Step 4		Step 5	
Affy.	1,930		22,431	20,501	31,231	29,301	40,379	38,449	28,704	26,774	37,950	36,020
			728	1,202	198	1,732	39	1,891	318	1,612	83	1,847
Step 1	22,431	728	21,703		31,511	9,808	41,779	20,076	30,194	8,491	40,144	18,441
	20,501	1,202			478	21,225	1,439	20,264	1,808	19,895	2,277	19,426
Step 2	31,231	198	31,511	478	31,033		42,192	11,159	32,004	971	41,214	10,181
	29,301	1,732	9,808	21,225			1,852	29,181	3,618	27,415	3,347	27,686
Step 3	40,379	39	41,779	1,439	42,192	1,852	40,340		42,208	1,868	**43,609**	3,269
	38,449	1,891	20,076	20,264	11,159	29,181			13,822	26,518	5,742	34,598
Step 4	28,704	318	30,194	1,808	32,004	3,618	42,208	13,822	28,386		40,524	12,138
	26,774	1,612	8,491	19,895	971	27,415	1,868	26,518			2,657	25,729
Step 5	37,950	83	40,144	2,277	41,214	3,347	**43,609**	5,742	40,524	2,657	37,867	
	36,020	1,847	18,441	19,426	10,181	27,686	3,269	34,598	12,138	25,729		

B

	Affy.	Step 1	Step 2	Step 3	Step 4	Step 5
42,601						
44,738						
41,610						
44,744						
44,742						
44,823						
44,843						

Figure 3.11 Annotation method overlap comparison. A) Pairwise comparison of annotation methods. The annotation methods are listed in the vertical and horizontal headers. The dark green is the total number of probe sets assigned a functional annotation by a single method under the default parameters. The light green is the probe sets that were annotated by both intersecting methods (by the method listed in the column and the row). Salmon colored boxes are probe sets that were annotated by one method but not by the intersecting method, with the box in the upper right corners listing probe sets annotated by the method listed in the column header and the box in the lower left listing probe sets annotated by the method listed in the row header. Yellow boxes list the number of probe sets annotated when the annotations of the intersecting methods are merged. B) Multiple annotation results merger comparison. The yellow boxes represent the methods from which the results were merged with the total number of probe sets assigned functional annotation by these merged methods listed to the left.

merger of the annotations from all 5 methods increased by 3.15% (Fig. 3.1). This strategy results in a balance between maximum annotation information provided and a high degree of confidence in the majority of assigned annotations. While the confidence in the annotations assigned at a threshold of 45 may be lower than those assigned at a threshold of 55 there is still some functional information about the probe set sequence provided. With a decreasing annotation threshold it is increasingly prudent to manually curate the annotations of any candidate genes (Jones et al., 2007; Götz et al., 2008). The annotations for a sub-set of probe sets on the wheat GeneChip were manually curated (a set of 3,852 probe sets representing triticale reproductive tissue-specific transcripts, courtesy of Laurian Robert, data not shown) and it was determined that the annotations provided by my results could not be improved upon using manual annotation methods while staying within the specified thresholds (i.e., e-value filter of 1*e*-3 or lower). This validates my methods of automated annotation for wheat GeneChip annotation improvement.

3.3.14 Annotation levels on the wheat GeneChip have been improved by 73.36%

Following the annotation improvement methods applied above, we have increased the percentage of probe sets on the wheat GeneChip with a GO annotation from 3.16% to 76.52% (61,115 probe sets total). For the three branches of the gene ontology, this corresponds to 60.97% of probe sets having a molecular function (MF) annotation (an increase from 2.87%), 52.34% of probe sets having a biological process (BP) annotation (an increase from 2.42%) and 62.99% of probe sets having a cellular component (CC) annotation (an increase from 1.38%). These are also significant increases over the

annotations provided by B2GOFar which annotated 34.93% of the probe sets on the GeneChip (23.94% with BP, 27.20% with MF and 27.38% with CC). The top 50 most frequently annotated GO terms for the wheat GeneChip for each of the three GO domains following the merger of steps 1-5 with the annotations provided by Affymetrix® are shown in Figures 3.12 to 3.14.

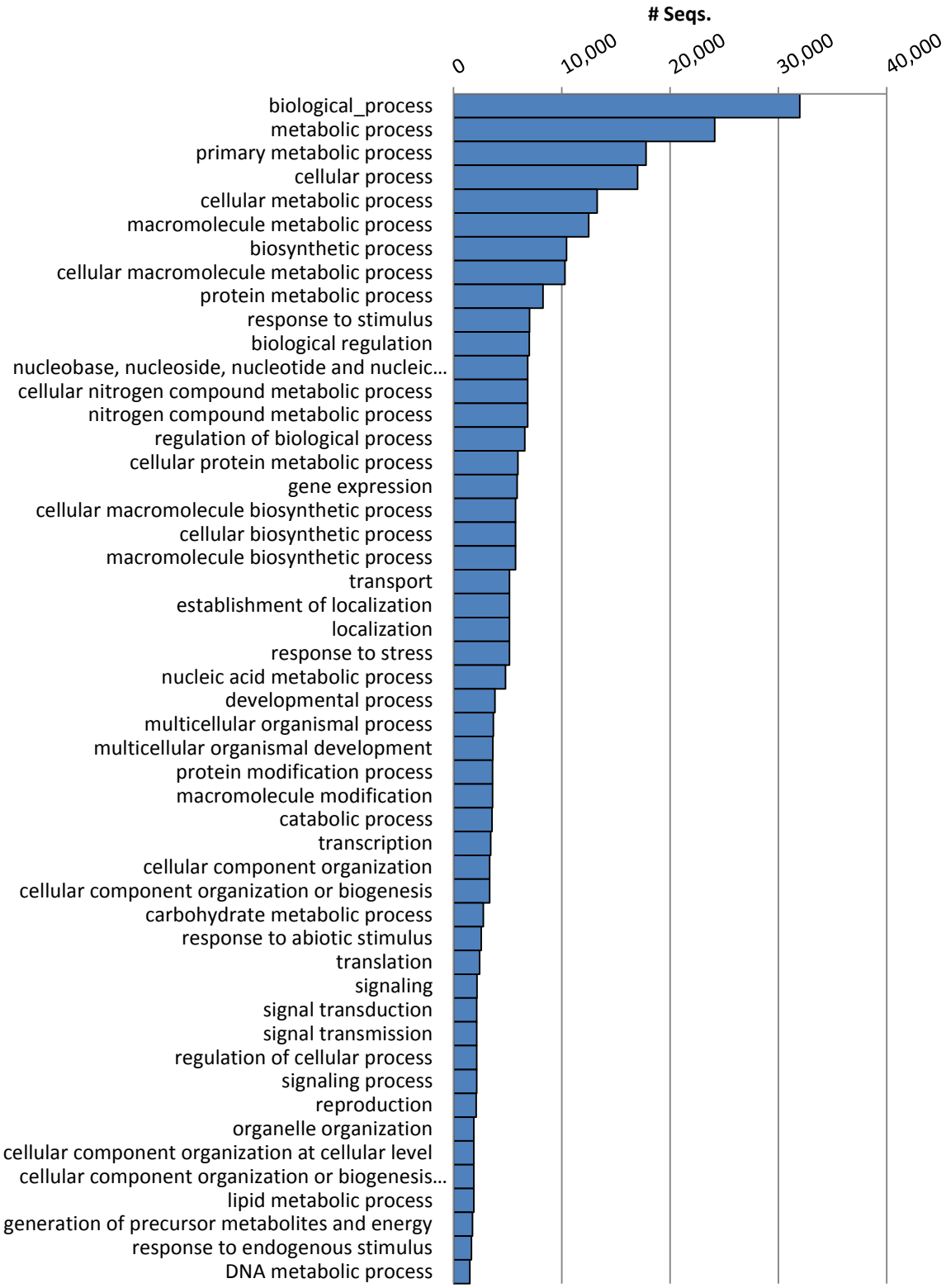


Figure 3.12 Sequence distribution of the top 50 most frequent biological process GO annotations associated with the wheat GeneChip. For some categories the entire term could not fit in the graph and part of the term is replaced with "..".

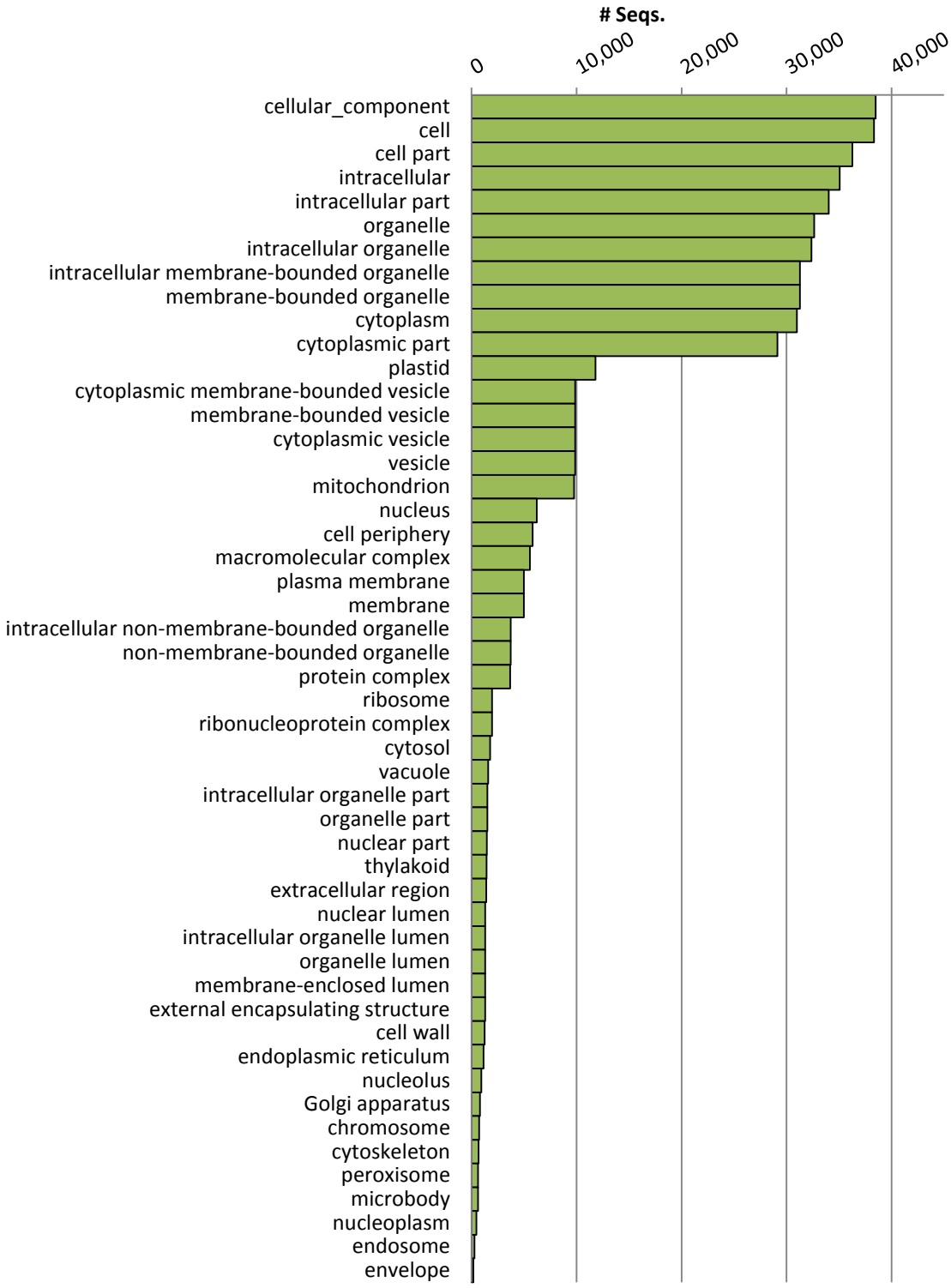


Figure 3.13 Sequence distribution of the top 50 most frequent cellular component GO annotations associated with the wheat GeneChip.

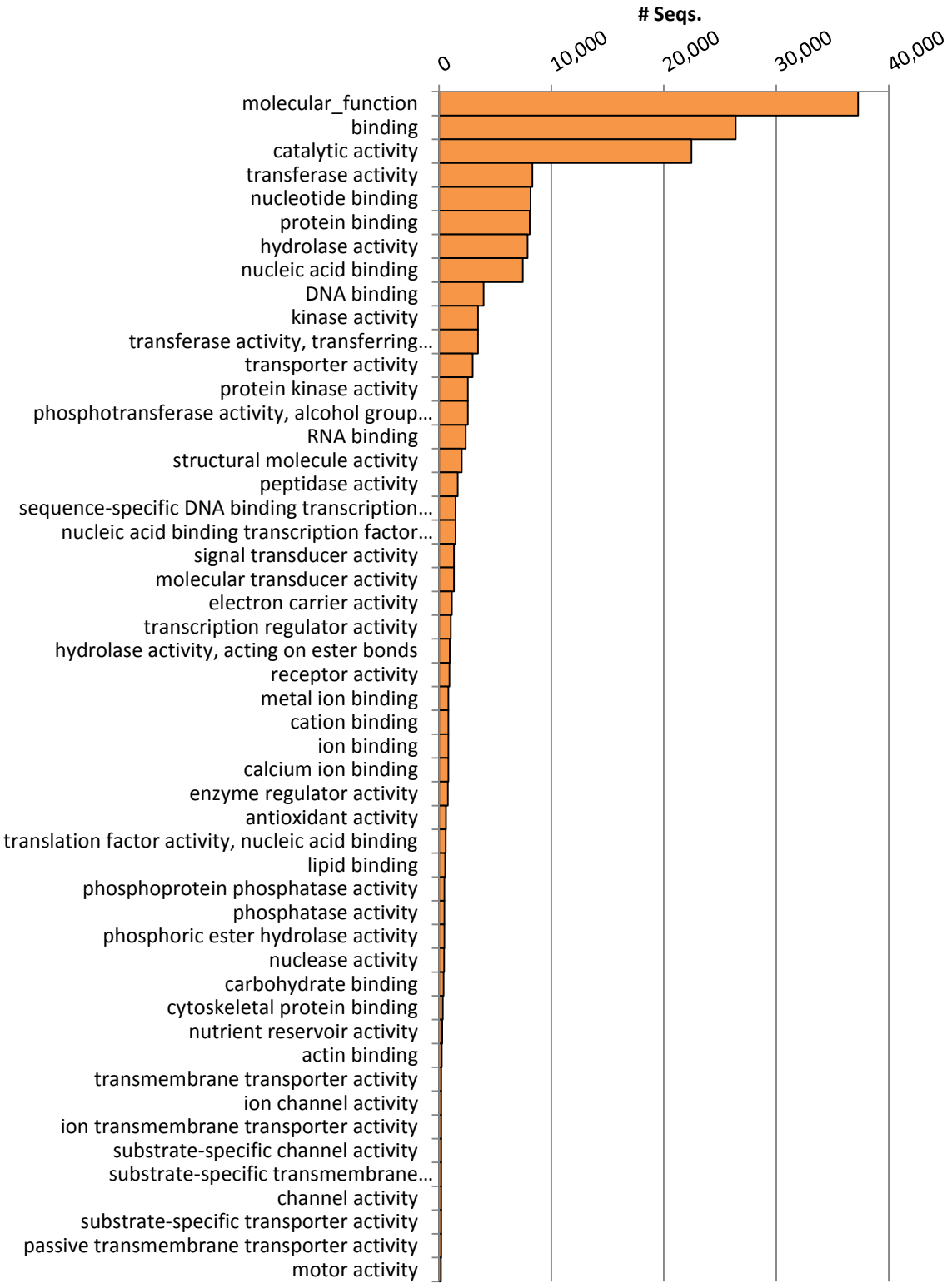


Figure 3.14 Sequence distribution of the top 50 most frequent molecular function GO annotations associated with the wheat GeneChip. For some categories the entire term could not fit in the graph and part of the term is replaced with "..".

3.4 DISCUSSION

We have generated functional annotations for 76.49% of the probe sets on the Affymetrix GeneChip® Wheat Genome Array in the standardized vocabulary of the Gene Ontology (GO) using sequence homology searches and the automated annotation tool Blast2Go. When combined with the annotations from Affymetrix® (resulting in 76.52%), these results provide annotations for the highest number of probe sets on the wheat GeneChip to date to my knowledge.

Using a two-step homology search for annotations, in which the first step was to identify a full length cDNA sequence with homology to the probe set consensus sequence, produced the highest number of annotated probe sets of all the strategies utilized (see Figure 3.1, Step 3 E and F). High throughput sequencing and assemblies of the sequencing results are being generated at an increasing rate due to the recent advances in sequencing technology and their cost reduction (Mardis, 2008; Bräutigam and Gowik, 2010). As sequencing efforts for non-model plants continue to generate assemblies that are added to the public databases it is reasonable to expect an improvement in annotation for other sequence resources such as the wheat GeneChip.

In our hands utilization of full length assembled cDNA sequences from triticale produced the highest proportion of annotated probe sets. If more full length cDNA sequences for wheat and other *Poacea* species are added to the databases, then the strategy applied in step 5 should increase the number of probe sets on the wheat GeneChip that can be annotated. Sequences that will remain un-annotated may be those sequences that are unique to wheat and those for which there is no protein function information for homologous sequences in other species. The sequencing efforts of the

wheat genome are ongoing (The International Wheat Genome Sequencing Consortium, http://www.wheatgenome.org/) and an accurate annotation of the genome will be vital to its usefulness to the plant research community. The accuracy of the annotations will likely undergo improvement overtime as more sequence annotations are manually curated and as more experimental evidence of gene function becomes available (Jones et al., 2007).

The majority of annotations were transferred from sequences with annotations that were inferred from an electronic annotation method (evidence code IEA, Figure 3.8) followed by annotations from reviewed computational analysis (evidence code RCA, Figure 3.8) and from the UniProtKB database (Figure 3.7). The majority of UniProtKB records are from the TrEMBL part of the database which consists of computer annotated translation of coding sequences in public nucleic acid databases (Schneider et al., 2004). The ongoing growth and improvement in this database requires that researchers continue to submit their nucleotide sequences to public databases. The proportion of unreviewed sequences in the UniProtKB database, as well as the predominance of the annotations with the IEA EC, highlights the need for the genomics community to continue to curate and verify annotations through experimentation.

The Plant Proteome Annotation Program (PPAP) has focused on the manual annotation of plant-specific proteins or protein families with an emphasis on the two fully sequenced model plants *Arabidopsis thaliana* and *Oryza sativa* (Schneider et al., 2009). While wheat (*Triticum aestivum*) is one of the ten most highly represented plant species in the UniProtKB/swiss-Prot database it only accounts for less than 3% of the manually annotated plant proteins (October 2008 release). Therefore there is a large dependence

on finding sequence similarity to proteins from other species with manual annotations in order to assign a functional annotation to wheat sequences. As the sequencing efforts in non-model plant species such as wheat continue, manual annotation efforts will be required to confirm these functional assignments. Our blast results (Figures 3.5 and 3.6) indicate that currently, the annotation of rice (*Oryza sativa*) proteins is particularly important to the assignment of functional annotations to sequences from other monocot species such as wheat.

Several sequence and genome information databases incorporate GO annotations including several plant specific databases such as The Arabidopsis Information Resource (TAIR, Berardini et al., 2004), Gramene (Liang et al., 2008) and UniProtKB (Schneider et al., 2004) to name a few. Accessing these annotation data and application of the data to sequences to be annotated can be cumbersome if the researcher has to access each database individually. Blast2Go accesses several public databases to retrieve all applicable GO information for the sequences in question (Conesa et al., 2005). The automation of the GO retrieval allows rapid functional annotation of a large sequence data set. This automation also enables easily repeating an annotation procedure at future dates as more submissions are made to the protein databases (for example, the addition of a large number of *Brachypodium distachyon* submissions to the nr database after Feb 11, 2011) allowing continual improvement of an annotation set. The functional annotation of the sequences from which microarrays are designed is critical to the biological interpretation of expression data. With functional annotation information, one cannot only assess which sequences are expressed in an experimental condition, but which functions are present in these expressed sequences and whether these functions are over-

or under-represented in comparison to the plant as a whole or to another experimental condition. Using Blast2Go we were able to provide functional annotations for over 76% of the probe sets on the wheat GeneChip, the largest amount to date to our knowledge. We are confident that the level of functional annotation is now sufficient to perform functional classification and enrichment analysis of expression data from experiments utilizing the wheat GeneChip. As well, additional gene functions may be predicted through expression pattern similarity to sequences with assigned functions.

The increase in functional annotation information is presented in a format that is easily searched, classified and utilized for functional enrichment analysis. This will provide biological meaning to our survey of gene expression in various tissues throughout development in triticale, in which we used the wheat GeneChip. We can now determine which functions and biological processes are enriched in each tissue and may therefore play a role in the development and differentiation of these various tissues throughout development. By comparing these genes with functional assignments to genes of unknown function that share similar expression patterns, we can assign putative functional annotations and identify additional candidate genes that may also have important roles in tissue differentiation. Utilization of the annotation data set for these purposes is presented in chapter 4.

3.5 LITERATURE CITED

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology **215**, 403-410.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics **25**, 25-29.

Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L.A., Yoon, J., Doyle, A., Lander, G., Moseyko, N., Yoo, D., Xu, I., Zoeckler, B., Montoya, M., Miller, N., Weems, D., and Rhee, S.Y. (2004). Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiology **135**, 745-755.

Bräutigam, A., and Gowik, U. (2010). What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biology **12**, 831-841.

Campbell, M.A., Zhu, W., Jiang, N., Lin, H., Ouyang, S., Childs, K.L., Haas, B.J., Hamilton, J.P., and Buell, C.R. (2007). Identification and characterization of lineage-specific genes within the *Poaceae*. Plant Physiology **145**, 1311-1322.

Conesa, A., and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics **2008**, 1-12.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**, 3674-3676.

Galbraith, D.W. (2006). DNA microarray analyses in higher plants. OMICS - a Journal of Integrative Biology **10**, 455.

The Gene Ontology Consortium. (2001). Creating the Gene Ontology resource: Design and implementation. Genome Research **11**, 1425-1433.

Götz, S., Arnold, R., Sebastián-León, P., Martín-Rodríguez, S., Tischler, P., Jehl, M.-A., Dopazo, J., Rattei, T., and Conesa, A. (2011). B2G-FAR, a species-centered GO annotation repository. Bioinformatics **27**, 919-924.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research **36**, 3420-3435.

Gupta, P.K., Mir, R.R., Mohan, A., and Kumar, J. (2008). Wheat genomics: present status and future prospects. International Journal of Plant Genomics **2008**, 1-36.

Jones, C., Brown, A., and Baumann, U. (2007). Estimating the annotation error rate of curated GO database sequence annotations. BMC Bioinformatics **8**, 170.

Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E.S., Casstevens, T., Hurwitz, B., McCouch, S., Ni, J., Pujar, A., Ravenscroft, D., Ren, L., Spooner, W., Tecle, I., Thomason, J., Tung, C.-w., Wei, X., Yap, I., Youens-Clark, K., Ware, D., and Stein, L. (2008). Gramene: a growing plant comparative genomics resource. Nucleic Acids Research **36**, D947-D953.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics **9**, 387-402.

Rensink, W.A., and Buell, C.R. (2005). Microarray expression profiling resources for plant genomics. Trends in Plant Science **10**, 603-609.

Schneider, M., Tognolli, M., and Bairoch, A. (2004). The Swiss-Prot protein knowledgebase and ExPASy: providing the plant community with high quality proteomic data and tools. Plant Physiology and Biochemistry **42**, 1013-1021.

Schneider, M., Lane, L., Boutet, E., Lieberherr, D., Tognolli, M., Bougueleret, L., and Bairoch, A. (2009). The UniProtKB/Swiss-Prot knowledgebase and its plant proteome annotation program. Journal of Proteomics **72**, 567-573.

Xu, Y., Badea, C., Tran, F., Frick, M., Schneiderman, D., Robert, L., Harris, L., Thomas, D., Tinker, N., Gaudet, D., and Laroche, A. (2011). Next-Gen sequencing of the transcriptome of triticale. Plant Genetic Resources: Characterization and Utilization **9**, 181-184.

CHAPTER 4 Global analysis of gene expression in hexaploid triticale

4.1 INTRODUCTION

An increase in global population, coupled with changes in the global environmental status, has caused an increase in the global demand for food. At the same time there has been an increase in demand for industrial and commercial products that are agriculturally produced from renewable resources. Cereals are the main agricultural products in Canada and can be used to supply the nutritional demands of the human population as well as a host of other industrial and commercial products. While supplying global consumer needs, improved cereal production also invigorates the Canadian agricultural economy by making improvements in yield, efficiency and value that are highly sought after in the industry. Triticale (X *Triticosecale* Wittm.) is an excellent cereal candidate for improvement to meet the demand for production of industrial bio-products in Canada. Triticale is a man-made species with favorable agronomic traits including high yield, low input requirements, adaptability to many Canadian climates and environments and resistance to most of the Canadian crop diseases (Horlein and Valentine, 1995; Giunta et al., 1999; Salmon, 2004). These traits make triticale an ideal candidate cereal crop for value-added bio-production in Canada

It has been shown that improved agricultural practices and breeding techniques can achieve increased yields and resistance to environmental factors (Repellin et al., 2001) but the production of novel industrial bio-products in cereals will require a more rapid biotechnology approach than conventional breeding and agronomics. Genetic transformation is a means of diversifying the range of products produced in cereals.

Transgenes can be used to alter the relative amount or type of protein or starch in the cereal grain or plant tissue (Goddijn and Pen, 1995). These transgene products can be used to make new or improved processed products. In order to achieve this improvement or alteration of cereal characteristics, promoters driving high levels of transgene expression will be required.

Regulatory and economic constraints require that transgene expression is localized, stable, and expressed at relatively high levels (Canadian Biotechnology Advisory Committee, 2002; Takaiwa, 2004). To achieve maximum extraction efficiencies and yields, and minimize bio-safety risks, it is necessary to select developmental- and tissue-specific promoters to target transgene expression. Identification and characterization of novel promoters free of infringing patents will also increase the return on investment for producers. The identification of these specific promoters will require the transcriptional profiling of triticale to identify tissue- and developmental stage-specific genes. Once genes with these specific expression patterns are identified, their upstream non-coding promoter sequences can be evaluated for their potential use in transgene expression.

Gene expression studies in plants often focus on the transcriptional response to various biotic or abiotic conditions. More recently, researchers have begun to focus attention on global transcriptional profiling studies that survey gene expression at the organ, tissue and cellular level (Girke et al., 2000; Wellmer et al., 2004; Wilson et al., 2004; Ma et al., 2005a; Druka et al., 2006; Benedito et al., 2008; Wan et al., 2008; Le et al., 2010). Global profiling of transcript accumulation among various cell or tissue types can help to determine which changes in gene expression reflect changes in cellular state.

Identification of genes regulated in a spatial or temporal manner provides a more comprehensive understanding of the transcriptional responses within each tissue (Altenbach and Kothari, 2004). This in turn provides a starting point for further elucidating the role of individual genes involved in tissue differentiation which may be of great value in crop engineering. It also aids in the identification of temporal- and tissue-specific promoters, and responsible *cis*-elements within them, that are important both for improving our understanding of gene regulation and plant development and for crop engineering applications.

Microarray profiling of gene expression has become a popular tool for the identification of plant genes responsive to a number of environmental factors (Kreps et al., 2002; Rabbani et al., 2003; Yamakawa et al., 2007; Zhou et al., 2007; Laudencia-Chingcuanco et al., 2011). Tissue-specific differences in gene expression in response to environmental conditions or stress treatments have highlighted the need to understand the spatial differences in gene expression that exist under optimal conditions prior to determining the effects of stress (Ma et al., 2005a; Zhou et al., 2007). A lesser and more recent application of microarray expression analysis has been the global profiling of gene expression patterns among organ, tissue and even cell types as well as throughout development (Girke et al., 2000; Gregersen et al., 2005; Ma et al., 2005b; Galbraith and Birnbaum, 2006; Le et al., 2007; Kawashima et al., 2009; Le et al., 2010). Much of the developmental profiling using microarrays has focused on the model plants Arabidopsis (*Arabidopsis thaliana*, Ma et al., 2005a; Schmid et al., 2005; Le et al., 2010) and rice (*Oryza sativa*, Ma et al., 2005b; Wang et al., 2010) and has covered embryogenesis (Girke et al., 2000; Kondou et al., 2006), male gametophyte development (Honys and

Twell, 2004; Lu et al., 2006), seed development (Girke et al., 2000; Becerra et al., 2006; Furutani et al., 2006; Kondou et al., 2006; Le et al., 2007; Le et al., 2010), and leaf senescence (Andersson et al., 2004; Guo et al., 2004; Lin and Wu, 2004).

Hexaploid triticale contains the A and B genomes from wheat (*Triticum durum*) and the R genome from rye (*Secale cereale*). Wheat and rye resources can therefore be used for functional genomics studies in triticale. Due to its size, the complete genome of wheat is not fully sequenced like that of Arabidopsis or rice; however, there are over one million wheat expressed sequence tag (EST) sequences publicly available, more than for most organisms and more than 138 times that available for rye (dbEST release 120701, October 8th, 2012, http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). This large amount of wheat EST data was made available in recent years and has allowed the creation of several wheat microarray platforms including the commercially available GeneChip® Wheat Genome Array (Affymetrix, Inc. Santa Clara, CA, USA). The GeneChip® Wheat Genome Array contains 61,127 probe sets representing 55,052 transcripts (including 12 control probe sets representing 3 transcripts) making it a useful and comprehensive tool for global gene expression studies.

The ability to profile the expression of such large numbers of genes at the same time makes microarrays an attractive tool for cereal researchers. The large number of genes surveyed across a small number of samples (a typical scenario for microarray experiments) also poses the potential of falsely identifying changes in gene expression. For example in analysis of a 61,115 probe set array using a Student's t-test with a probability (P) value cutoff of $P \leq 0.05$ it would be expected that, by chance, 5% of the probe sets on the array would be identified as having significant differential expression

and could be false positives. This phenomenon is known as the multiple testing problem (Quackenbush, 2005). There is also the potential for errors to be introduced at each step in the micorarray process including sample preparation, slide scanning, image processing, and data normalization and analysis and these errors may affect what is identified as having a significant change in expression. Alternate techniques therefore, are recommended (and often required for publication) to verify microarray results. Quantitative real time-PCR (qRT-PCR) is one commonly used method to verify the observed patterns of expression in the sample set under investigation.

Expression analysis methods including microarray and qRT-PCR have the ability to identify lists of genes with particular expression patterns that may be of interest for further investigation. The biological meaning from such expression studies, however, lies in the functions of genes that share similar expression patterns. Therefore it is the functional annotation of the identified genes that can provide biological meaning and without sufficient functional annotation, interpretation of the expression results is difficult. Controlled vocabularies such as the Gene Ontology (GO, http://www.geneontology.org/) provide terms for describing gene product characteristics and gene annotation data and facilitate the interpretation and comparison of expression study results (The Gene Ontology Consortium et al., 2000). By using this vocabulary and applying these terms to sequence data, large data sets such as microarrays can easily be sorted, classified, and compared. Meaning can be extracted from the expression results by examining functional enrichment within identified gene lists. Over or under-representation of a particular biological pathway may indicate its importance in the

regulation of the biological process, or the differentiation and development of a tissue under study.

We hypothesize that global transcript profiling in triticale will enable the identification of spatial differences in expression between tissue types under optimal growth conditions and will identify candidate genes for characterization of promoters that will be suitable for transgene expression. In addition to identifying transcripts suitable for promoter characterization, the objectives of this experiment also set out to provide a reference data set of gene expression in triticale that could be used by other triticale and cereal researchers as a gene expression atlas and a reference and comparative data set for the investigation of future hypothesis-driven research. It is also expected that the exploitation of functional information for transcripts with differential expression patterns will allow the identification of biological pathways and regulatory networks controlling plant tissue development and differentiation.

To identify triticale genes expressed in a developmental- and tissue-specific manner, microarray profiling of gene expression was conducted on early root and shoot tissue, leaf tissue at five developmental stages (seedling, tillering, early booting, early senescence and late senescence), stem tissue at four developmental stages (early stem development, tillering, early booting and early senescence) and five seed tissues (embryo, endosperm, crease, pericarp and epiderm). The gene expression profiling was performed using the Affymetrix GeneChip® Wheat Genome Array. This work was combined with the microarray profiling of gene expression in three reproductive tissues (anther, stigma and ovaries) at four different stages of development (tetrad and uninucleate microspore stage and bicellular and tricellular pollen stage) and mature pollen by colleagues at the

Eastern Cereal and Oilseeds Research Center (ECORC, AAFC), for a complete profile of gene expression in triticale.

Probe sets identified as representing genes with tissue-specific, tissue-enriched and developmentally regulated expression patterns were subsequently used for functional enrichment analysis using the GO annotations provided for the wheat array in Chapter 3. This functional enrichment analysis was used to identify biological processes and molecular functions that may be indicative of cellular state and help to identify pathways that may be important to the development and differentiation of the tissues studied. The expression of a selection of genes identified by microarray as tissue-specific, tissue-enriched or developmentally regulated were also verified using qRT-PCR. The data will also provide a basis for comparison for future studies by highlighting the spatial and temporal differences in gene expression in the absence of stress or other treatments.

4.2 MATERIALS AND METHODS

4.2.1 Plants and growth conditions

The triticale (X *Triticosecale* Wittm.) cultivars used in this study were AC Certa and AC Alta. Plant material was produced by Agriculture and Agri-Food Canada in Swift Current, Saskatchewan as part of their annual seed increase. For leaf, seed and stem tissues, AC Certa was used and three seeds were planted equally spaced, 2 cm deep into several 1 gallon pots of Cornell mix, a steam-pasteurized (121˚C for 8 h) mixture of loam, sand, and peat (3:1:1, by volume, Boodley and Sheldrake Jr., 1977). Seedlings were thinned to 2 per pot after 1 week growth. All plants were grown in growth cabinets (Conviron, Winnipeg, Canada) in order to maintain a disease and stress free environment. Growth conditions were 16 h photoperiods at 18˚C and 8 h dark periods at 15˚C. Light intensities of 350 µmole m^{-2} s^{-1} of photosynthetic photon flux density were provided by cool white fluorescent tubes and incandescent bulbs. For reproductive, root and coleoptile tissues, AC Alta was used and plants were grown and sampled under similar conditions at ECORC.

4.2.2 Plant tissue collection

Leaf and stem tissues were removed from plants using RNase-free scalpel blades. The collected tissues were packaged in tin-foil, immediately frozen in liquid nitrogen (N_2) and stored at -80°C until further processing. Seeds were dissected under a stereomicroscope on an RNase-free surface using RNase-free scalpel blades to separate seed tissues. Instruments were cleaned with Absolve™ (Perkin Elmer Life and

Analytical Sciences, Cat. No. 6NE9711) then rinsed with RNase-free water between dissections to eliminate RNA degradation by exogenous RNases. Samples were collected at growth stages detailed in Table 4.1 for a total of 29 combinations of different tissues and developmental stages.

Seedling, tillering and early boot leaf stages (Figure 4.1A) were determined by comparison to Feekes and Zadoks' growth scales for wheat (Large, 1954; Zadoks et al., 1974) while sampling times for early and late senescence were determined by a previous experiment (Chapter 2). Early senescence leaf tissue was collected at 22 days post-anthesis (DPA) and late senescence tissue was collected at 30 DPA. Plants were discarded after collection of a particular tissue. The material extracted from a group of plants with the same planting date is considered a biological replicate. Tissue was collected from 3 replicates with different planting dates.

Stem tissues stages were determined by comparison to the Zadoks' growth scale for wheat (Zadoks et al., 1974). Internode stem samples were collected from above the crown to the second node (internodes 2 and 3). The nodes were not collected and the leaf sheath was completely removed. Stem sections were taken to determine physiological differences associated with sampling times (Figure 4.1, B)

Seed tissues were collected at the soft dough stage (Zadoks' growth scale, 85; Figure 4.2a). In order to prepare enough RNA for the planned experiments, seed tissue from several plants was pooled. Approximately 10 to 30 grains were dissected per spike from approximately 1 to 3 spikes per plant. The material extracted from a group of plants with the same planting date is considered a biological replicate. Tissue was collected from 3 replicates with different planting dates. Approximately 25 to 30 plants were used

Table 4.1 Triticale tissues collected for RNA isolation for use in expression analysis.

Sample Name	Cultivar	Zadoks growth scale	Developmental Stage	Tissue Collected	Section of tissue collected	Replicates
Root	AC Alta	7	Emergence	Root tissue 7 days after sowing	Entire root	3
Coleoptile	AC Alta	7	Emergence	Shoot tissue 7 days after sowing	Entire shoot	3
Seedling	AC Certa	13	Seedling	Leaf tissue	entire leaf	3
Tillering	AC Certa	23	Tillering	Leaf tissue	entire leaf	3
Flag Leaf	AC Certa	49	Early Boot	Flag leaf tissue	entire leaf	5
22DPA	AC Certa	75+	Early Senescence	Flag leaf tissue	2-3 cm section 2 cm from stalk	3
30DPA	AC Certa	85+	Late Senescence	Flag leaf tissue	2-3 cm section 2 cm from stalk	3
Stem21	AC Certa	21	Initial tillering	Stem tissue	internode stem tissue between crown and second node	3
Stem45	AC Certa	45	Early Boot	Stem tissue	internode stem tissue between crown and second node	3
Stem59	AC Certa	59	Ear clear of leaf sheath	Stem tissue	internode stem tissue between crown and second node	3
Stem75	AC Certa	75	Early Senescence	Stem tissue	internode stem tissue between crown and second node	3
Embryo	AC Certa	85	Soft dough stage of seed development	Embryo	embryo with attached scutellum	3
Endosperm	AC Certa	85	Soft dough stage of seed development	Endosperm	mid-section of endosperm	3
Crease	AC Certa	85	Soft dough stage of seed development	Crease	vascular bundle from ventral groove with most of seed coat tissues removed	3
Pericarp	AC Certa	85	Soft dough stage of seed development	Pericarp	large mid-section of pericarp seed layer	2
Epiderm	AC Certa	85	Soft dough stage of seed development	Epiderm	large mid-section of epiderm seed layer	2
Anther tetrad	AC Alta	41	tetrad microspore stage	Anthers	entire anther	3
Ovary tetrad	AC Alta	41	tetrad microspore stage	Ovaries	Ovary removed from stigma	3
Stigma tetrad	AC Alta	41	tetrad microspore stage	Stigmas	stigma removed from ovary	3
Anther UNP	AC Alta	49	Uninucleate microspore stage	Anthers	entire anther	3
Ovary UNP	AC Alta	49	Uninucleate microspore stage	Ovaries	Ovary removed from stigma	3
Stigma UNP	AC Alta	49	Uninucleate microspore stage	Stigmas	stigma removed from ovary	3
Anther BNP	AC Alta	55-57	Bicellular pollen stage	Anthers	entire anther	3
Ovary BNP	AC Alta	55-57	Bicellular pollen stage	Ovaries	Ovary removed from stigma	3
Stigma BNP	AC Alta	55-57	Bicellular pollen stage	Stigmas	stigma removed from ovary	3
Anther TNP	AC Alta	59	Tricellular pollen stage	Anthers	entire anther	3
Ovary TNP	AC Alta	59	Tricellular pollen stage	Ovaries	Ovary removed from stigma	3
Stigma TNP	AC Alta	59	Tricellular pollen stage	Stigmas	stigma removed from ovary	3
Pollen	AC Alta	60	Mature Pollen	Pollen grains	Isolated polen grains	3

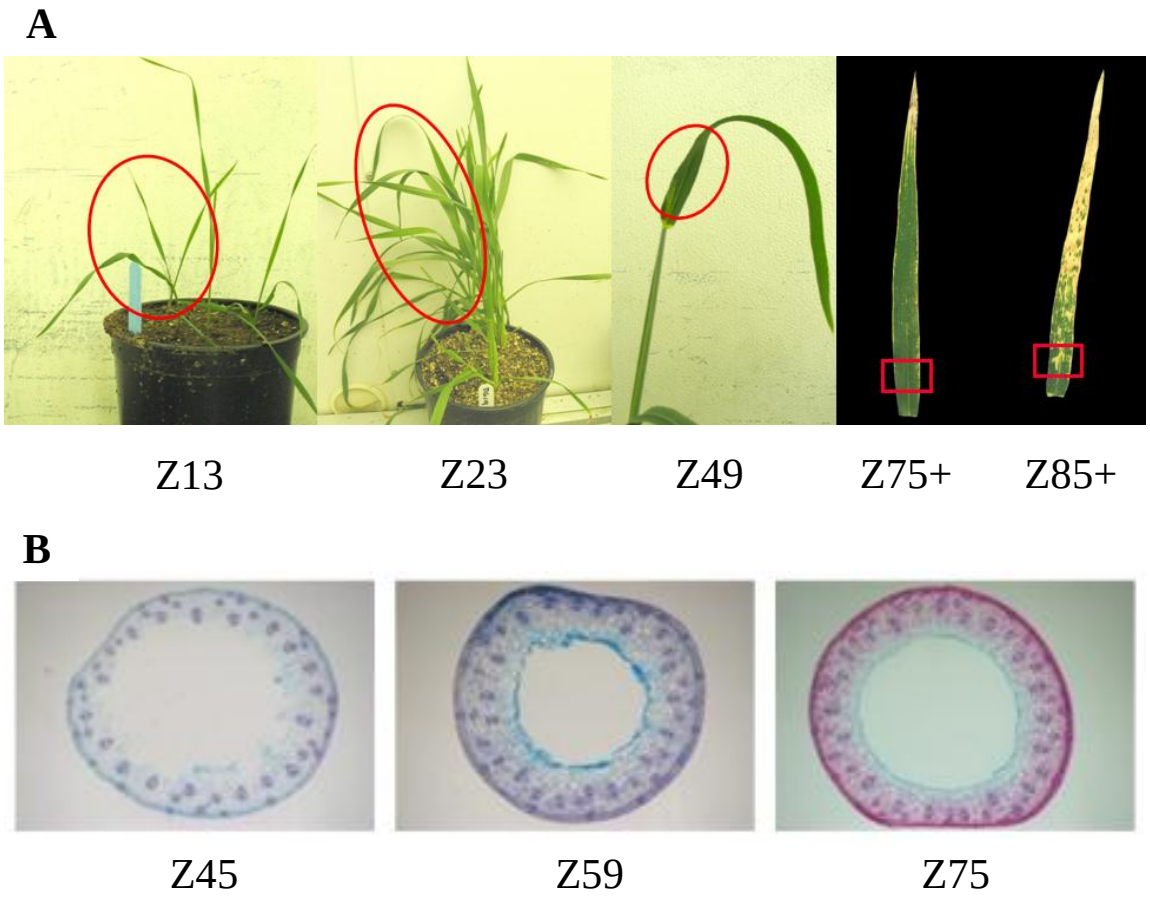


Figure 4.1 Representative stem and leaf tissues sampled for microarray analysis. A) Leaf tissues sampled for isolation of RNA for use in expression analysis. Leaf tissues were sampled at the developmental stages indicated below according to the Zadoks' (Z) growth scale. Red circles/squares indicate areas of leaf tissues sampled. From left to right images are of leaf tissue at the seedling, tillering, flag leaf, early senescence (22 DPA) and late senescence (30 DPA) stages of development. B) Stem tissues sampled for isolation of RNA for expression analysis. Stem tissues were sampled at the developmental stages indicated below the images according to the Zadoks' growth scale. From left to right images are of stem tissue at the early booting, ear clear of leaf sheath and early senescence stages of development. Stem tissues were sectioned with a with a vibrating blade microtome (Leica Microsystems) and stained with safranin O (red, stains lignin) and astra blue (blue, stains hemicelluloses) for imaging on a dissecting microscope (Leica Microsystems). Stem tissue at the initial tillering stage of development (Z21) is not shown due to the tenderness of the tissue at this stage preventing sectioning and subsequent staining and imaging. Stem tissue images are courtesy of Michele Frick (Lethbridge Research Centre, Agriculture and Agri-Food Canada, Lethbridge).

Figure 4.2 Dissection of triticale seeds for the separation of seed tissues. Seeds at soft dough stage (a, Z85) were dissected by the following steps: embryos were removed by gently pushing them with the side of the scalpel blade through a diagonal incision made in the seed coat tissues (b-d), the ends of the grain were removed (e), a longitudinal section to one side of the crease (ventral groove) was made (f) to allow quick separation and removal of seed coat tissues, the translucent epiderm layer was removed and cut away near the crease (g), the endosperm was gently pushed away from the photosynthetic pericarp layer and vascularized crease tissues (h-i) and finally the crease was excised from the pericarp tissue by longitudinal sectioning (j). Tissues were immediately placed in liquid N_2 in an RNase-free plastic 50 ml tube set in an insulated container also filled with liquid N_2 (k).

for each replicate. Grains were removed from spikes and dissected as quickly as possible (approximately 1 minute per grain) to minimize induced changes in gene expression. Seeds were dissected according to the steps outlined in Figure 4.2. As each tissue segment was separated from the other tissues it was immediately frozen in liquid N_2 in RNase-free 50 ml tubes. This allowed the preservation of tissues on the bench top for a period of time that allowed several dissections before liquid N_2 levels needed to be refreshed. All frozen tissues were stored at -80°C until further processing. Instruments were cleaned with Absolve™ then rinsed with RNase-free water between dissections to eliminate RNA degradation by exogenous RNases.

Reproductive tissue stages of development were determined using anther squashes and DAPI staining (Figure 4.3). Once the pollen developmental stage was established the tissues were manually dissected under a stereomicroscope and immediately frozen in liquid N_2.

4.2.3 RNA extractions

Total RNA was isolated from crease, epiderm and all leaf and stem tissues using the RNeasy® Plant Mini Kit (QIAGEN, Cat. No. 74904) with on-column digestion of DNA during RNA purification using the RNase-Free DNase Set (QIAGEN, Cat. No. 79254). Total RNA isolated from stem tissues was further purified using the RNA cleanup protocol in the RNeasy® Mini Kit (QIAGEN, Cat. No. 74104) to remove carbohydrates. Total RNA was purified from 100 mg of each tissue and eluted in 50 µl of nuclease-free water and stored at -80°C.

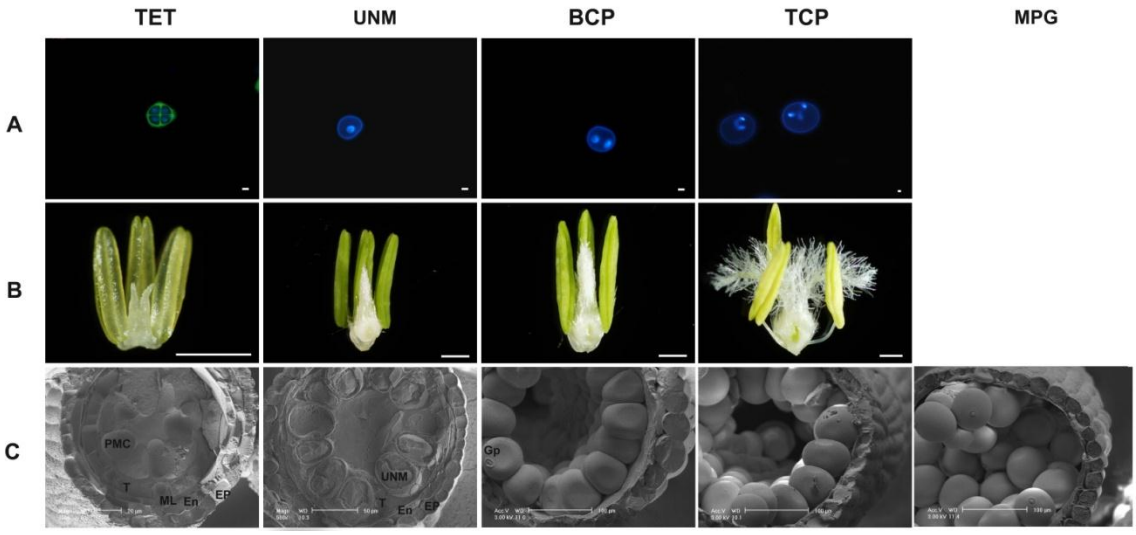


Figure 4.3 Reproductive tissues used for expression analysis. A) DAPI stained anther squashes to establish landmark stages of pollen development; TET, tetrad, UNM, uninucleate, BCP, bicellular and TCP, tricellular stages, MPG, mature pollen grains. B) images of attached anther, ovary and stigma tissues at each of the stages sampled. C) SEM on anther cross sections to show the pollen development at each of the stages sampled since developing pollen could not be imaged with light microscopy. Image courtesy of Frances Tran (ECORC, AAFC , Ottawa).

The presence of large amounts of carbohydrates in the embryo, endosperm and pericarp tissue made the RNA extraction using the above method difficult. Therefore, to increase the yield and purity, RNA was isolated from these three tissues using the method described by Singh et al. (2003). Extracting seed tissues in buffer (50 mM Tris-HCl pH 9, 150 mM NaCl, 1% sarcosyl, 20 mM ethylene diamine tetracetic acid [EDTA], and 5 mM dithiothreitol [DTT]) ensures maximum RNA solubility in the aqueous phase and the removal of most interfering polysaccharides and insoluble materials. RNA was then purified from the aqueous phase using a guanidine hydrochloride buffer system (Singh et al., 2003). Total RNA was isolated from 100 mg of each tissue using this method then dissolved in 100 µl of nuclease-free water and immediately further purified using the RNeasy® Mini Kit (QIAGEN, Cat. No. 74104) following the RNA cleanup protocol in the kit manual with on-column digestion of DNA using the RNase-Free DNase Set (QIAGEN, Cat. No. 79254). Purified RNA was eluted in 50 µl of nuclease-free water then stored at -80°C.

Dilutions of 1:50 were made for quantification of RNA samples using UV spectrophotometry at 260 nm. Based on concentrations, RNA samples were concentrated under vacuum to approximately 1 µg/µl or greater in volumes no less than 10 µl. The 10 µl RNA samples were shipped to the McGill University and Genome Quebec Innovation Centre on dry ice. Quality and quantity of RNA samples were verified by microfluidics-based automated electrophoresis using an Agilent 2100 bioanalyzer (Agilent Technologies).

Tissue collection and RNA isolation from root, coleoptiles and reproductive tissues were performed by colleagues at ECORC using similar protocols to the leaf RNA extraction protocol described above.

4.2.4 cDNA synthesis, cRNA synthesis and hybridization to Affymetrix GeneChip® Wheat Genome Arrays

Synthesis of double stranded cDNA (ds cDNA) and biotin labeled cRNA from all RNA samples was performed at the McGill University and Genome Quebec Innovation Centre. Quantity and quality of RNA and cRNA samples were verified by microfluidics-based automated electrophoresis using an Agilent 2100 Bioanalyzer prior to proceeding with the sample. Hybridization of the labeled cRNA samples to Affymetrix GeneChip® Wheat Genome Arrays and subsequent staining and scanning of the arrays were also performed at the Genome Quebec Innovation Centre. These methods were performed according to their Affymetrix protocol (https://genomequebec.mcgill.ca/nanuq Administration/download/affymetrixProtocol_en.pdf, June, 2004).

No triticale microarrays are commercially available. Triticale contains two wheat genomes (A and B genomes) and one genome from rye (R genome). Of the publicly available rye EST sequences, 99.12% had Blast hits to wheat publicly available sequences and 98.9% had a Blast hit with an expect (e) value of ≤ 1.00 *e*-7 (Blastn, July 27, 2007). Based on these sequence homology parameters, it is expected that the majority of rye sequences can be detected with microarrays designed for wheat sequence hybridization. Therefore, Affymetrix GeneChip® Wheat Genome Array, the largest

commercially available wheat array at the time of this study, was chosen for expression analysis of triticale RNA.

4.2.5 Affymetrix GeneChip® Wheat Genome Array data analysis

In total 87 Affymetrix GeneChip® Wheat Genome Arrays were hybridized with the cRNA from vegetative, seed and reproductive tissue samples described above. The raw probe cell intensity (CEL) data from these arrays were normalized to provide probe set summarizations using the Affymetrix® Expression console™ software. All 87 CEL files were first normalized using the Affymetrix® MAS 5.0 algorithm (Affymetrix, 2002, http://www.affymetrix.com/support/technical/whitepapers/sadd_whitepaper%.pdf) with default parameters (background estimation by weighted average of the lowest 2% of the feature intensities, and linear scaling of the feature level intensity values using the trimmed mean) to provide detection calls (P, A or M for present, absent and marginal respectively). Following MAS 5.0 normalization, the robust multiple-array average (RMA) algorithm (Irizarry et al., 2003) was performed to provide expression value summaries for each probe set on the arrays. Control and reporter probe sets were eliminated from the analysis leaving a total of 61,115 probe sets that were examined for expression analysis. The resulting intensity values for each probe set on each array range between 1.40 and 14.65 on a logarithmic scale to base 2 ($\log_2$) . Three biological replicates were analyzed for most samples with the exception of pericarp and epiderm tissues for which only two out of the three biological replicates collected were analyzed due to poor quality of the cRNA for the third replicates of these samples, and flag leaf for which five biological replicates were collected and analyzed.

4.2.5.1 Replicate chip analysis

Chip-to-chip correlations of overall RMA expression data were analyzed using Pearson correlation coefficients in Microsoft Excel® and visualized using within-class comparison scatter plots in FlexArray (data not shown , Blazejczyk et al., 2007) in order to identify if there were any poorly correlated replicates that needed to be removed from further analysis.

4.2.5.2 Identification of gene expression patterns

4.2.5.2.1 Identification of tissue-specific expression

MAS 5.0 detection calls of biological replicates were combined to provide consensus detection calls which were used to determine presence and absence of expression. Two methods of determining tissue-specific expression were applied to the data, a permissive method of majority consensus (>50%) among detections calls and a stringent method of absolute consensus (100%) among detection calls. Because of the possibility of falsely identifying tissue-specific expression using a permissive strategy and the possibility of not identifying all potentially tissue-specific expression using the stringent strategy, both strategies were used. For the permissive strategy, the probe set consensus detection calls had to have a majority of present (P) or absent (A) detection calls for the transcript to be considered present and absent respectively in a tissue of interest, and all other consensus detection calls were considered insufficient to determine tissue-specific expression. This means that for all sample types there had to be a minimum of two biological replicates (except for flag leaf which had to have 3 out of 5 biological replicates) that had detection calls of P for presence of expression, or A for

absence of expression of the transcript represented by a probe set (for pericarp and epiderm, >50% requires that both replicates have either P or A detection calls). For the stringent protocol, probe sets with absolute agreement in detection calls among all biological replicates were used to determine expression. Consensus detection call of 100% P for all biological replicates was considered present and representing an expressed transcript while a consensus call of 100% A for all biological replicates was considered absent and representing a transcript not expressed in the particular tissues/developmental stage. All other combinations of detection calls between biological replicates for a given probe set were considered insufficient to determine tissue-specific expression by the stringent method and were not used for comparative analysis. For example, for a sample with three biological replicates a consensus detection call of PPP would be considered present while AAA would be considered absent and PPA would be considered insufficient to determine tissue-specific expression. For both the permissive and the stringent protocols, filtering criteria were applied to the data by only selecting those probe sets that had a present or absent consensus call for further comparative analysis. Tissue-specific expression was determined for both the stringent and permissive methods by filtering for probe sets that had present consensus detection calls in the tissue of interest and absent consensus detection calls in all other tissues. Presence of expression (P) in a group of tissues was determined by the majority P detection call (permissive) or absolute P call (stringent) in any tissue within the group of tissues. Absence of expression (A) was determined by the majority A detection calls (permissive) or absolute A calls (stringent) in all tissues within the group. Constitutive expression (C) was determined by majority P detection calls (permissive) or absolute P detection calls

(stringent) in all tissues within the group. Specific expression constitutively expressed across a group of tissues (SC) was determined by majority P detection calls (permissive) or absolute P detection calls (stringent) in all tissues within the group and majority A detection calls (permissive) or absolute A detection calls (stringent) in all tissues not within the tissue group of interest. The probe sets that were identified using both the stringent and permissive strategies are considered to be the probe sets that represent tissue-specific expression with the highest level of confidence.

4.2.5.2.2 Identification of tissue-enriched expression

Tissue-enriched expression was considered to be greater than or equal to two-fold higher expression in a tissue of interest in comparison to all other tissues. To calculate tissue-specific enrichment of expression, MAS 5.0 consensus detection calls and RMA normalized expression values across all 29 tissue types (87 arrays) were used. Following RMA normalization an analysis of variance (ANOVA) was performed across the 29 different tissues types and a false discovery rate (FDR, Benjamini and Hochberg, 1995) was applied. Only probe sets with a significant difference in expression (adjusted p value ≤ 0.05) across tissue types and those with a majority P consensus detection call in the tissue of interest were considered for further evaluation. Following normalization and ANOVA, enriched expression was determined as greater than or equal to two-fold enrichment in the tissue of interest by pair wise comparisons between the tissue of interest and all other tissues by filtering for a difference in mean RMA expression values greater than or equal to two fold (a difference $\geq$ 1on a $\log_2$ scale).

Evaluation of four different strategies to determine expression enrichment was performed to optimize the probe sets identified as representing tissue-enriched expression (Figure 4.4). The first strategy used only the mean normalized RMA expression values to calculate fold change. For the remaining three strategies, the MAS5.0 detection calls were used to convert mean RMA normalized expression values to zero in the case of absence of expression. This conversion of RMA values to zero allows the detection of enriched probe sets when mean expression levels are similar in two tissues due to the compression of low expression values on a $\log_2$ scale, but are determined to be present in the tissue of interest and absent in other tissues by MAS5.0 detection calls which is often the case in genes expressed at low levels. The second strategy used mean RMA normalized expression values to calculate fold change following the conversion of these values to zero when the corresponding consensus detection call had a majority of absent (>50% A) detection calls. The third strategy was similar to the second strategy except that mean RMA normalized values were converted to zero when the corresponding consensus detection call did not have a majority (<50% P) of present detection calls. The fourth strategy was similar to the third strategy except that mean RMA normalized values were converted to zero when the corresponding consensus detection call did not have an absolute consensus (≠100% P) of present detection calls.

Evaluation of the four strategies revealed that the probe sets identified as enriched using only the first and/or second method (List 1 and 2 in Figure 4.4) were also found not to have a majority present consensus detection call in the tissue of interest and were therefore discounted as likely false positives. To evaluate the probability that the probe sets identified using strategies 3 and 4 were true positives, the probe sets were ordered

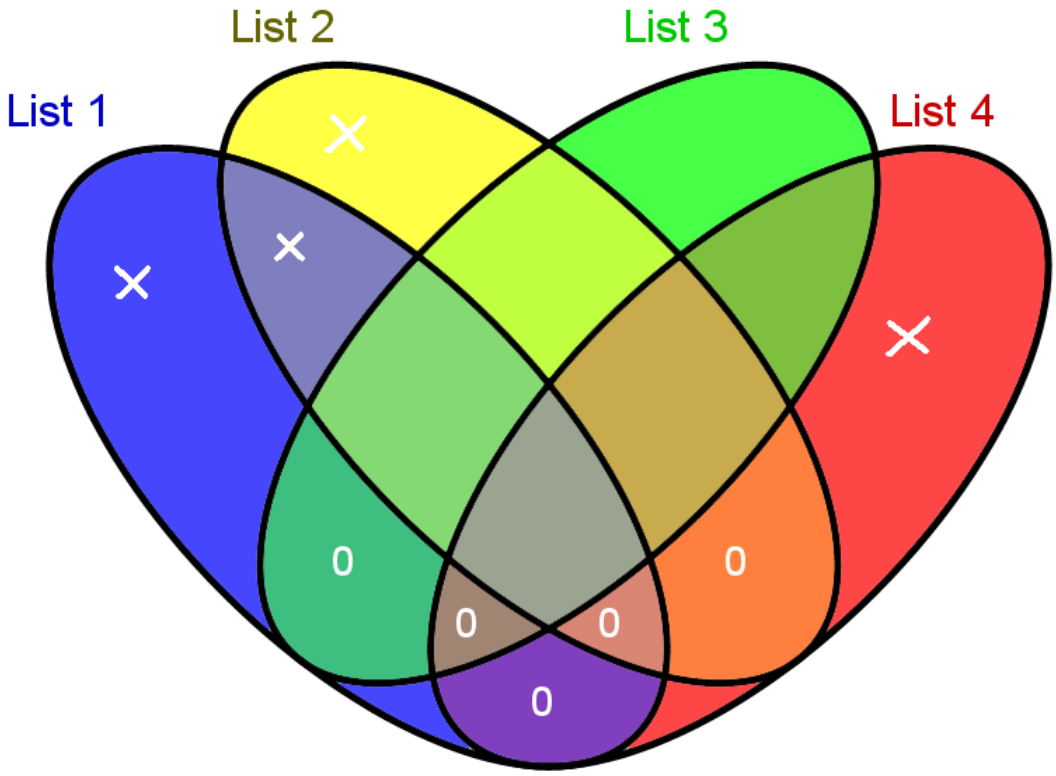


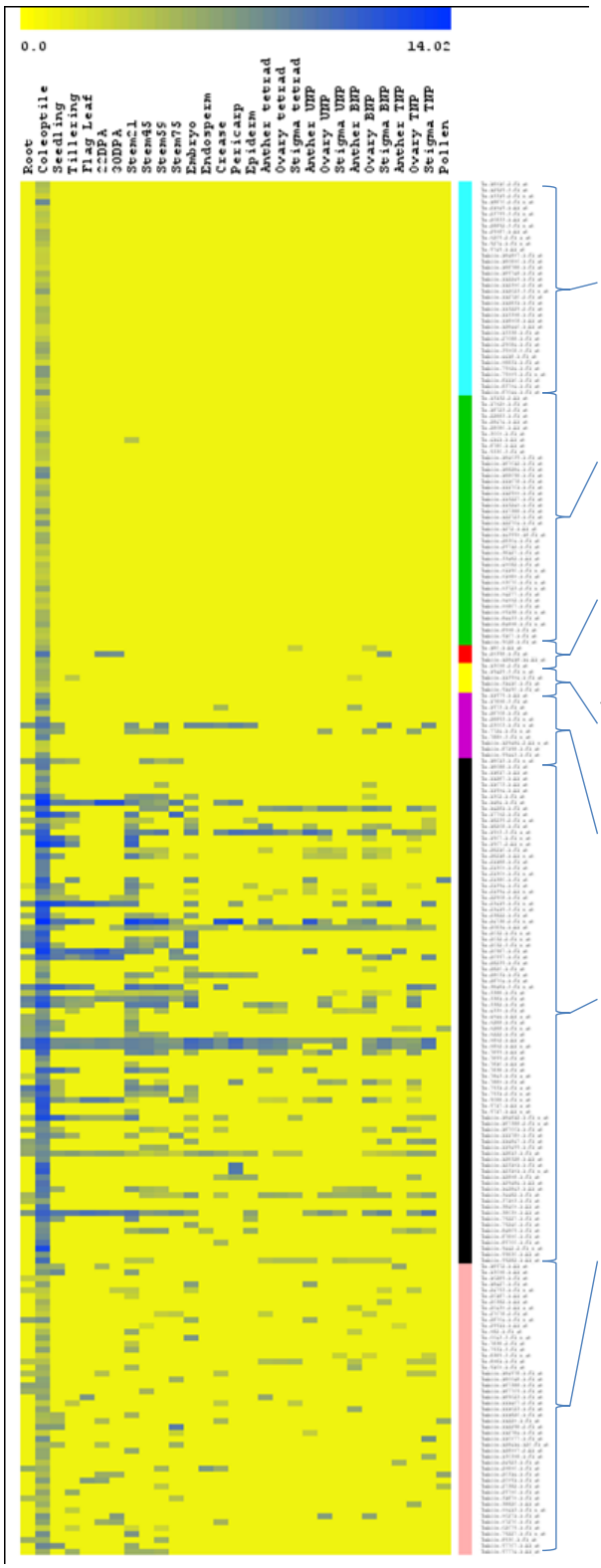
Figure 4.4 Venn diagram of the strategies evaluated for determining probe sets representative of tissue-enriched expression. List 1 probe sets with tissue-enriched expression when using the normalized mean RMA expression values, List 2 probe sets with tissue-enriched expression when mean RMA expression values were converted to zero when corresponding consensus detection calls had a majority of absent detection calls (>50%, A), List 3) probe sets with tissue-enriched expression when mean RMA expression values were converted to zero when corresponding consensus detection calls did not have a majority of present detections calls (< 50% P), List 4 probe sets with tissue-enriched expression when mean RMA expression values were converted to zero when corresponding consensus detection calls did not have absolute consensus of present detection calls (≠ 100% P). A 0 indicates that there was no overlap in the probe sets identified as tissue-enriched by the indicated strategies in any seed or vegetative tissue. Probe sets identified in categories marked with an X were disregarded as false positives. False positives in list 1 and 2 were due to there not being a majority present detection call in the tissue of interest and false positives in list 4 were due to presence of expression in tissues other than the tissue of interest at a level close to that of the expression in the tissue of interest (less than 2 fold difference of expression).

according to their overlap among the four strategies and their expression patterns were then visualized using expression heat maps. Figure 4.5 shows the heat map of coleoptile-enriched probe sets as a representative example. The ordered heat map revealed that probe sets that were identified using strategy 4 and not strategy 3 (Pink, Figure 4.5) were found to have majority consensus detection calls of present in tissues other than the tissue of interest and that expression was less than 2-fold lower and in many cases higher in at least one tissue compared to the tissue of interest before the conversion to zero. These probe sets were therefore discounted as likely false positives and removed from further analysis. The evaluation of the four strategies therefore left only the probe sets identified using strategy 3 (RMA values converted to zero when consensus detection call is non majority P) for further analysis of tissue-enriched expression. The probe sets identified as tissue-enriched using all four strategies (Black, Figure 4.5) would be considered to be tissue-enriched with the highest level of confidence.

4.2.5.2.2.1 Clustering of tissue-enriched genes

The Multiple Experiment Viewer (MeV) package from the TM4 software suite (Saeed et al., 2003) was used for hierarchical clustering of tissue-enriched genes in order to group genes according to their similarity of expression patterns. Hierarchical clustering (HCL) was performed on the mean RMA expression values of the probe sets following the conversion of expression values to zero when corresponding consensus detection call did not have a majority of present detection calls (<50% P).

Figure 4.5 Heat map of the probe sets identified as having coleoptile tissue-enriched expression using the four methods described in Figure 2. Each row represents the probe set expression values across the tissues studied and each column represents one of the tissue types studied. The horizontal color scale at the top of the heat map represents the expression values with yellow representing no expression and dark blue representing maximum expression. The mean expression values for each probe set in a particular tissue are represented by the intensity of color. The vertical color bar to the right of the heat map represents the probe sets identified using the different methods. Mean RMA expression values following the conversion of these values to 0 when the majority of detection calls were not P (< 50% present) were used to construct the heat map.



Light Blue = enriched probe sets identified in list 3 only

Green = enriched probe sets identified in list 2 and 3

Red = enriched probe sets identified in lists 1, 2 and 3

Yellow = enriched probe sets identified in list 3 and 4

Magenta = enriched probe sets identified in list 2, 3 and 4

Black = enriched probe sets identified in list 1, 2, 3, and 4
*Highest confidence

Pink = enriched probe sets identified in list 4 only

4.2.5.2.3 Identification of developmentally regulated gene expression

In order to identify genes whose expression levels change throughout development, expression patterns in tissues types that were sampled at varying stages of development were analyzed. Among the vegetative tissues, both stem and leaf tissues were sampled at multiple stages of development and were used for identification of developmentally regulated genes (DRGs). Following RMA normalization across all 29 tissues types (87 arrays) samples were grouped according to developmental stage then ANOVA and FDR (Benjamini and Hochberg, 1995) were performed on RMA normalized expression values only for the tissues sampled at multiple stages of development using the FlexArray software package (Blazejczyk et al., 2007).

4.2.5.2.3.1 Leaf developmentally regulated gene expression

ANOVA and FDR were performed across the 5 leaf sample types from seedling (Zadoks' 13) to late senescence (Zadoks' 85). Only probe sets with significantly differing expression values (adjusted p value ≤ 0.05) across leaf stages were considered for further analysis. Evaluation of four strategies to determine developmental regulation was performed to optimize the probe sets identified as DRGs. These strategies were the same as those evaluated for tissue-enriched expression analysis and as with the tissue-enriched expression analysis, only the third strategy, where mean RMA values are converted to zero in the case of a non-majority present consensus detection call, proved to be a valid method for the same reasons as listed for the tissue-enriched analysis. Following conversion of absent calls to zero, probe sets with a minimum two-fold difference in mean expression values (≥ 1 on the $\log_2$ scale) across all 5 leaf tissues were

considered to be developmentally regulated (leaf DRGs). This minimum difference was determined by comparison of the maximum to the minimum mean RMA expression values among the 5 leaf tissues.

4.2.5.2.3.2 Stem developmentally regulated gene expression

ANOVA and FDR were performed across the 4 stem sample types from Zadoks' 21 to Zadoks' 75. Only probe sets with significantly differing expression values (adjusted p-value ≤ 0.05) across stem tissues were considered for further analysis. Following normalization and filtering for adjusted p-values, the same steps were performed as for leaf DRG identification, including conversion of absent expression values to zero and filtering for probe sets with a minimum 2 fold difference in mean expression across all 4 stem stages to identify stem DRGs.

4.2.5.2.3.3 Clustering of developmentally regulated genes

Following the steps to identify leaf and stem DRG expression, clustering was performed on the probe sets to identify various patterns of developmentally regulated expression using the self organizing tree algorithm (SOTA, Herrero et al., 2001). This divisive clustering method is more appropriate when it is predicted that there are several diverse patterns of expression within the list of genes to be clustered. This was predicted for leaf and stem DRGs in contrast to the tissue-enriched genes which all share a similar pattern of expression and therefore used the aggregative method of hierarchical clustering. SOTA clustering was performed on the mean RMA expression values of the probe sets following the conversion of absent expression values to zero using the

Multiple Experiment Viewer (MeV) package from the TM4 software suite of software (Saeed et al., 2003).

4.2.6 Classification of gene lists

Lists of probe sets representing expressed, tissue-specific, tissue-enriched, and developmentally regulated genes were grouped into categories using GO annotations provided for the wheat array in Chapter 3. Lists of probe sets along with their associated GO annotations were imported into Blast2Go for categorization according to molecular function, biological process and cellular component.

4.2.7 Functional enrichment analysis

Blast2Go was used for functional enrichment analysis of probe set lists identified from the microarray analysis. GO annotations used for functional enrichment are the final set of annotations for the wheat array described in Chapter 3. The list of probe sets that were determined to be expressed in any tissue (majority P consensus detection call in any tissue, 41,918 probe sets), were used as the reference set as they represent expression in triticale under the conditions studied. Probe sets determined not to be expressed in any of the tissues studied were excluded from the reference set to avoid the comparison to wheat genes that may not be present in triticale and that are only expressed under conditions not encountered by this study (i.e., biotic and abiotic stress conditions). GO's present in the lists of probe sets were compared to GO's present in the reference set using a two-tailed Fisher's Exact Test (Fisher, 1922) to determine over and under representation. Enrichment results were filtered for those with a p-value ≤ 0.05. Only

the most specific terms for each tissue are presented to avoid the redundancy of parent and child terms, and only those with a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of probe sets are presented to show the terms most likely to be important to tissue differentiation and development.

4.2.8 qRT-PCR

Based on the microarray analysis, candidate genes were chosen for verification of the microarray results using qRT-PCR. Candidates were chosen based on their enrichment of expression in leaf and seed tissues or developmental regulation in leaf tissues as these were the samples that were available for qRT-PCR. Only two biological replicates of each leaf and seed tissue were used for qRT-PCR verification due to space limitations in the qRT-PCR plates. Amplification of target transcripts in each biological replicate was tested in technical triplicates according to standard qRT-PCR protocol.

For qRT-PCR, first-strand cDNA was synthesized using a 1 µg aliquot of the total RNA samples described above, oligo$(dT)_{18}$ (25 ng μL^{-1}), dNTPs (0.5 mM), RNaseOUT (2 units μL^{-1}, Invitrogen, Cat no. 10777-019), SuperScriptIII (10 units μL^{-1}, Invitrogen, Cat no. 18080-044), first strand buffer (Invitrogen) and 10 mM DTT in a 20 µL reaction incubated at 42°C for 50 minutes. Two cDNA synthesis reactions were performed for each RNA sample then pooled, diluted 1 in 2 for a total volume of 80 µL and stored at -20°C until use in qRT-PCR.

Primers for qRT-PCR were designed using Primer3 (Rozen and Skaletsky, 2000) with lengths near 21 bp, GC content near 50%, melting temperatures near 60°C and

targeting amplicons between 140 and 160 bp. The primers designed for qRT-PCR are listed and described in Table 4.2. Presumptive housekeeping genes (Table 4.3) were selected by their expression stability in previous experiments (actin, EF1a, contig1, contig5, data not shown) or in the present microarray experiment (*Ald*). Suitability of housekeeping genes for use as internal controls was evaluated using the method of Vandesompele et al. (2002), with a calculated internal control gene stability measure for each reference gene of ≤ 1.11. The geometric mean (0.95) of these five genes was used as an internal reference.

All primers were evaluated for their ability to amplify the targeted transcript with specificity by PCR using 5 µl of a 1:20 diluted aliquot of a pooled sample of cDNA prepared above with 0.5 µM forward primer, 0.5 µM reverse primer, 200 µM dNTPs, and 2.5 units of QIAGEN *Taq* DNA Polymerase (QIAGEN, Cat no. 201205) in a 25 µL reaction. Amplified products were purified using QIAquick® Gel Extraction Kit according to the kit manual (QIAGEN, Cat. No. 28706) and quantified using the PicoGreen® dsDNA Quantitation Kit (Roche, Cat no. P-7589).

For qRT-PCR, a 5 µL aliquot of 1:20 diluted cDNA template was used along with 0.3 µM gene specific primers, 12.5 nM fluorescein and 1X QuantiTect® SYBR® Green Master Mix (QIAGEN, Cat. no. 204143) in 20 µL reactions in an iCycler iQ® Real-Time Detection System (Bio-Rad Laboratories, Inc.). Reactions were also run simultaneously with purified and quantified PCR products in order to calculate relative copy numbers.

Table 4.2 Quantitative real-time PCR primer sequences for selected genes

Probe set ID	Representative Public ID	Microarray Expression Pattern	Forward Primer 5' to 3' / Reverse Primer 5' to 3'
TaAffx.99992.1.S1_at	CA593620	Tillering-specific	CCTTTCGTCGACTTCCTCAAG
			ACTTCTTGGTCACCGACGACT
Ta.27399.3.S1_at	CA718775	Embryo-enriched	GACCCATGATCAATCTGGAAG
			TATTGTCCAACCCAACAAAGG
Ta.2798.1.S1_s_at	BJ290016	Embryo-enriched	AGACTCGCAGGGAGCAGAT
			GGACTTGGTCTTGAACTTGGA
Ta.5720.1.S1_at	BJ296692	Embryo-enriched	AGTGCTCAGCTATGCACGGTA
			TGCATGCGTGTATCTAGGTGA
TaAffx.69777.1.S1_at	J02961.1	Embryo-enriched	GCAGGGTTTGCACTAACAACTA
			CATTCTGCGAGAAGAGTGGAG
Ta.23142.5.S1_x_at	CD868513	Endosperm-enriched	GGGTATATTCTTGCAGCCACA
			ATCAGTAGCCACCAACACCAG
Ta.11127.2.S1_a_at	CA708523	Endosperm-enriched	CAAATGCAGACCATATGAGCA
			CTTGATCCAAGCAGGAATTGA
Ta.13160.2.S1_x_at	CA635509	Crease-enriched	CTGACGTCCAGAACCCATC
			CTCCTTGGTCTCCTCAACCTC
Ta.20064.1.S1_at	CA666265	Crease-specific	GCTCGTCTCCTTGCCATCTA
			CCATTAAGCATCCGAACATGA
Ta.30755.1.S1_at	CN007962	Pericarp enriched	CTAGCTGAGTGCGTGTTGTTG
			GTACGGGCAACACAAATTTCA
Ta.590.2.S1_x_at	CA601369	Leaf developmentally regulated. Down regulated throughout development	AGCCTCAAGGGGAGAGTTTG
			ACATGAGGAAGTAGCGCAGAG

Table 4.3 Quantitative real-time PCR primer sequences for putative housekeeping genes.

Housekeeping genes	Representative Public ID	Forward Primer 5' to 3' / Reverse Primer 5' to 3'	Annotation
Actin-For	DN551593	GGAAAAGTGCAGAGAGACACG / TACAGTGTCTGGATCGGTGGT	putative actin
Ta.304.1.S1_at	CK171609	CCCTGAGGTGATTGCTGAGTA / TTCTTGGTCTGGAGCTTGTTC	putative fructose 1-,6- biphosphate aldolase (*Ald* gene)
Contig1	CK172186	AACCTTCCTAATCGCCTGGT / CCATATGTTCCTGCCACTGA	putative hexokinase
Contig5	CK155621	CTGCAGTGCGTGCATATTTT / AACAAGAACGATGCCGAGTT	putative protein
Ef1a	M90077	GGTGATGCTGGCATAGTGAA / GATGACACCAACAGCCACAG	Wheat translation elongation factor 1 alpha-subunit (TEF1) mRNA

Optimal template dilutions were determined by quantification of a dilution series, prior to running the full experiments.

Analysis of qRT-PCR results was performed using the freely available qBase software (version 1.3.5, http://medgen.ugent.be/qbase/). The threshold cycle (Ct) values within the log-linear range were used to estimate relative transcript abundance using a modified $2^{-\Delta\Delta CT}$ method that takes multiple reference genes into account after having determined gene-specific amplification efficiencies as outlined by Hellemans et al. (2007).

4.3 RESULTS

4.3.1 Quality control

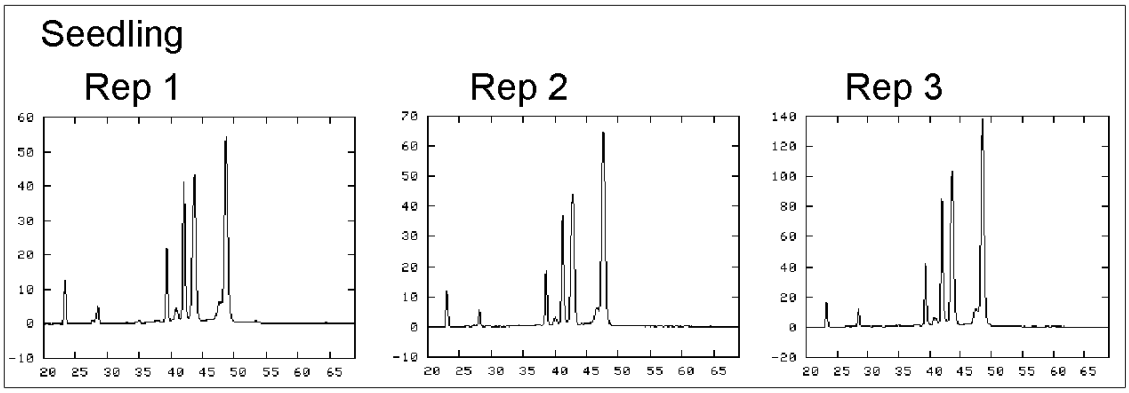
RNA was extracted from leaf tissues at five developmental stages and from five separate seed tissues. The quantity of RNA extracted from equal amounts of starting material varied among the tissues. Seed tissues in particular had higher variability among the amounts of total RNA extracted from 100 mg of tissue ranging from 2.1 µg from crease tissue to 178 µg from embryo tissue (Table 4.4). Among the leaf tissues, the late senescence samples had the least amount of total RNA as would be expected due to the degraded nature and reduced levels of ribosomal RNA in senescing tissue (Lim et al., 2005). RNA quality was also assessed using microfluidics-based automated electrophoresis (Agilent Bioanalyzer). Quantity and quality of RNA samples was sufficient to proceed with cRNA synthesis (Figure 4.6). The resulting cRNA quality was analyzed prior to hybridization to the GeneChips using the same automated electrophoresis. With the exception of 2 samples, the quality of cRNA was sufficient to proceed with hybridizations (Figure 4.7). Pericarp rep 3 and epiderm rep 3 samples showed large peaks to the left in the bioanalyzer charts, indicating large amounts of small cRNA fragments. The quality of pericarp rep 3 and epiderm rep 3 cRNA samples were therefore insufficient for further processing and hybridization to wheat GeneChips.

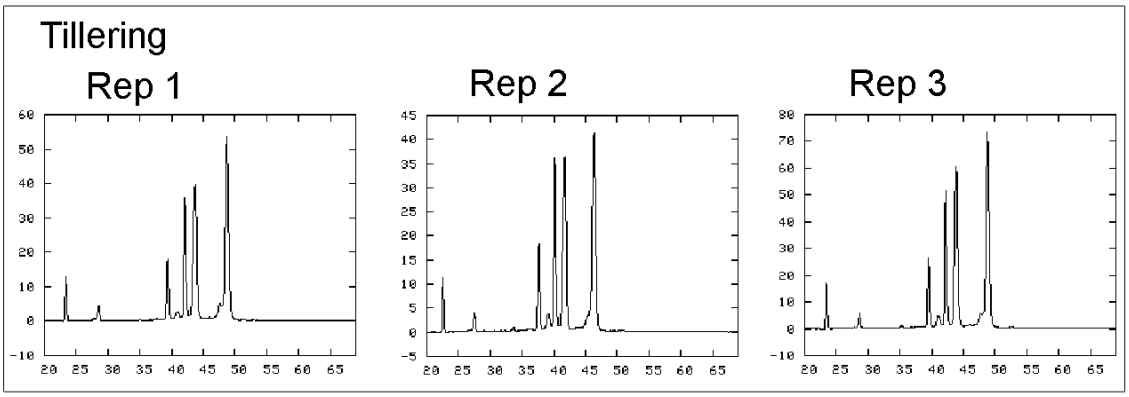
RNA and cRNA from the remainder of the tissue samples (root, coleoptiles, stem and reproductive tissues) were analyzed by colleagues at ECORC and quantity and quality were similarly assessed prior to hybridization to wheat GeneChips (data not shown).

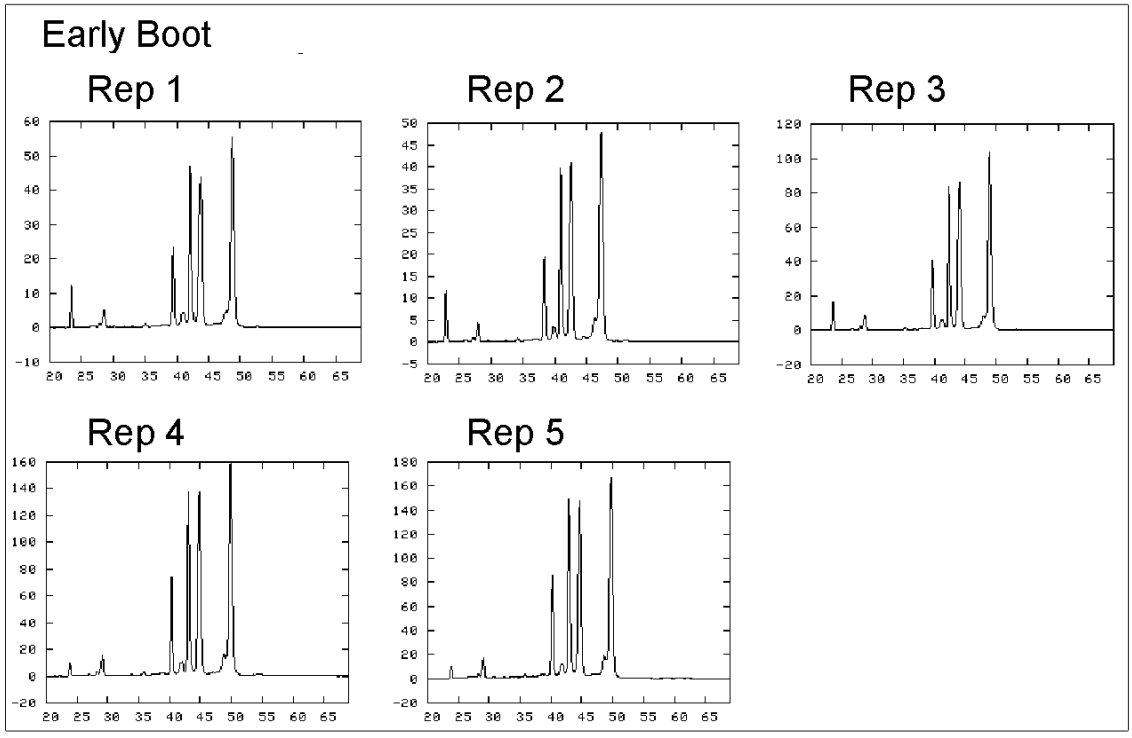
Table 4.4 RNA extraction yields from triticale tissues. RNA was extracted from 100 mg of tissue and quantified by UV spectrophotometry at 260 nm. The relative purity of the RNA samples was also determined by a ratio of absorbance at 260 nm over 280 nm ($A_{260/280}$).

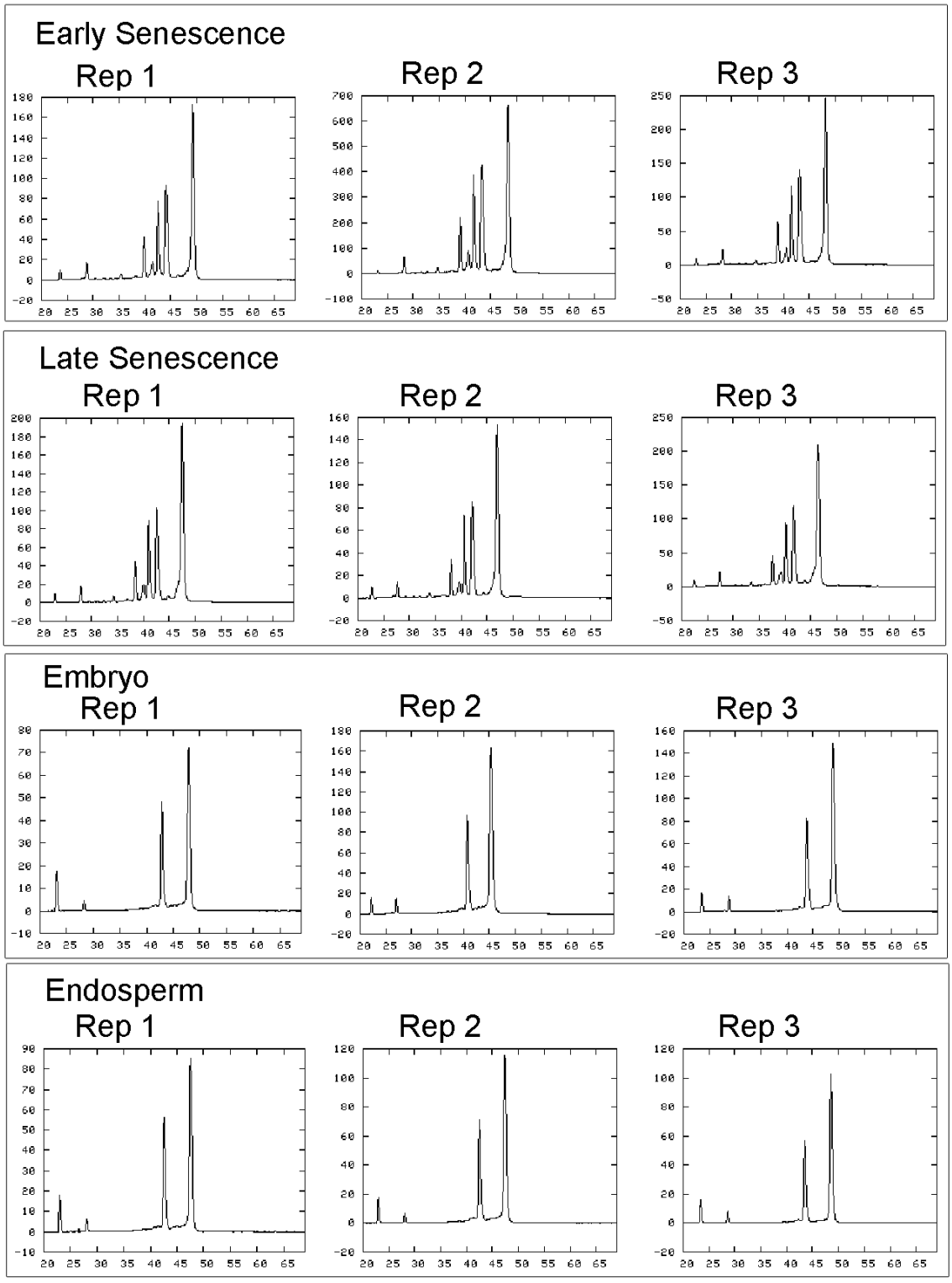
	Rep 1		Rep 2		Rep 3		Rep 4		Rep 5	
Tissue	RNA	$A_{260/280}$	RNA	$A_{260/280}$	RNA	$A_{260/280}$	RNA	$A_{260/280}$	RNA	$A_{260/280}$
Seedling	36.6 µg	2.011	35.7 µg	2.017	34 µg	2.000				
Tillering	47.7 µg	2.004	28.5 µg	2.012	16.2 µg	2.000				
Early Boot	33.4 µg	2.012	36.6 µg	2.011	30.8 µg	2.000	25.5 µg	2.003	29.1 µg	1.993
Early Senescence	15.7 µg	1.962	23.9 µg	1.992	14.4 µg	2.000				
Late Senescence	6.5 µg	1.970	8.5 µg	2.024	9.3 µg	1.979				
Embryo	22.6 µg	1.992	83.05 µg	1.997	177.8 µg	2.002				
Endosperm	16.9 µg	2.000	20.8 µg	1.993	81.1 µg	1.987				
Crease	10.3 µg	1.839	3 µg	1.875	2.1 µg	1.909				
Pericarp	28.2 µg	1.918	39.2 µg	1.960	52.6 µg	1.838				
Epiderm	8.1 µg	1.929	6.1 µg	1.906	2.7 µg	1.929				

Figure 4.6 RNA quality control. Charts produced from the Bioanalyzer (Agilent Technologies) automated electrophoresis analysis of RNA samples. The x-axis is time in seconds, y-axis is absorbance. Sharp peaks and flat baseline indicate good quality RNA. Peaks (from left to right) are; internal reference (23 seconds), 5S RNA (~28 seconds), degraded 23S RNA (green tissues only, between 36 to 40 seconds) 16S RNA (green tissues only, between 42 to 45 seconds),18S RNA (between 41 to 44 seconds), and 28S RNA (between 45 to 50 seconds).


Seedling
Rep 1
Rep 2
Rep 3


Tillering
Rep 1
Rep 2
Rep 3


Early Boot
Rep 1
Rep 2
Rep 3
Rep 4
Rep 5


Early Senescence
Rep 1
Rep 2
Rep 3
Late Senescence
Rep 1
Rep 2
Rep 3
Embryo
Rep 1
Rep 2
Rep 3
Endosperm
Rep 1
Rep 2
Rep 3

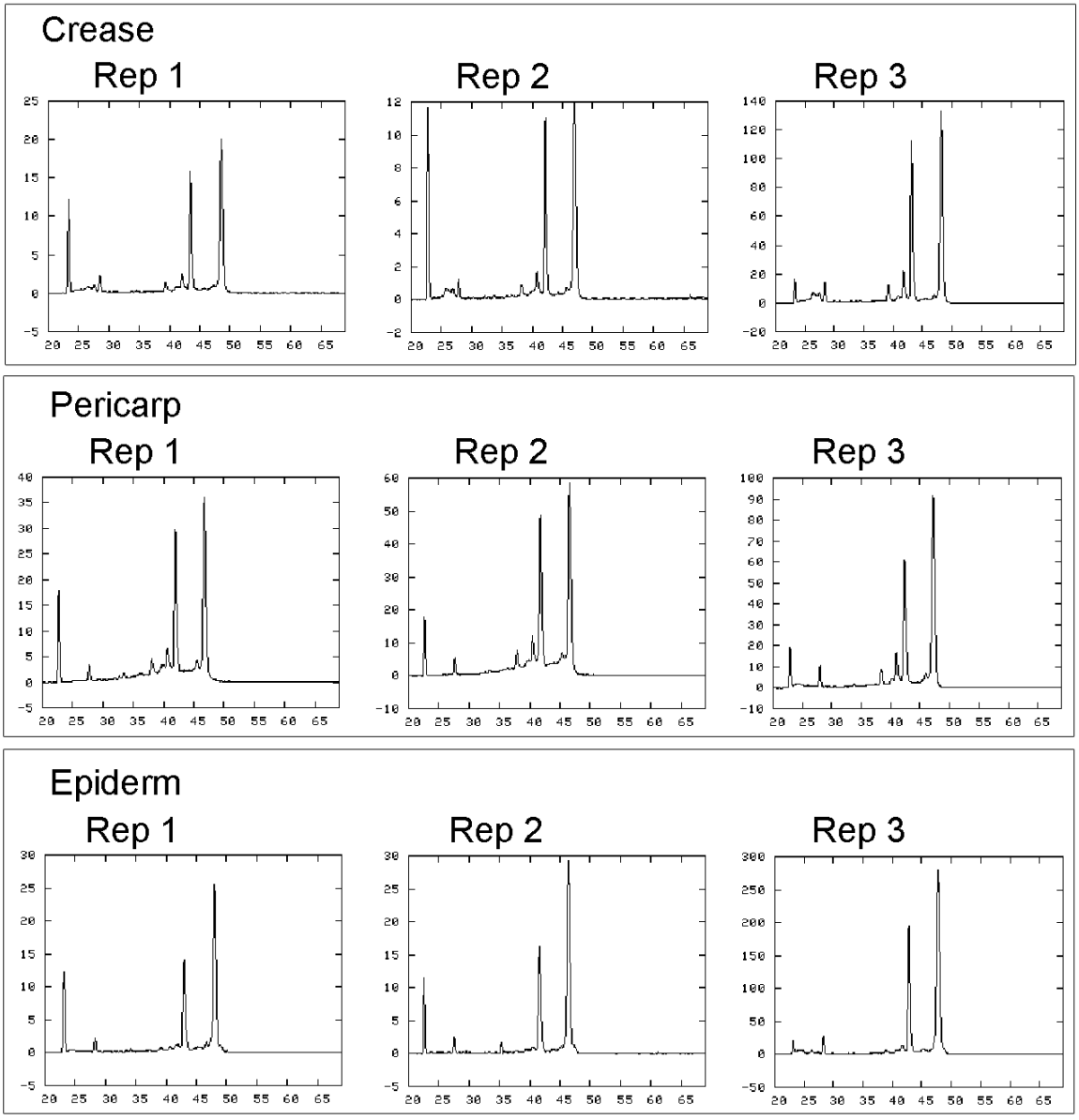
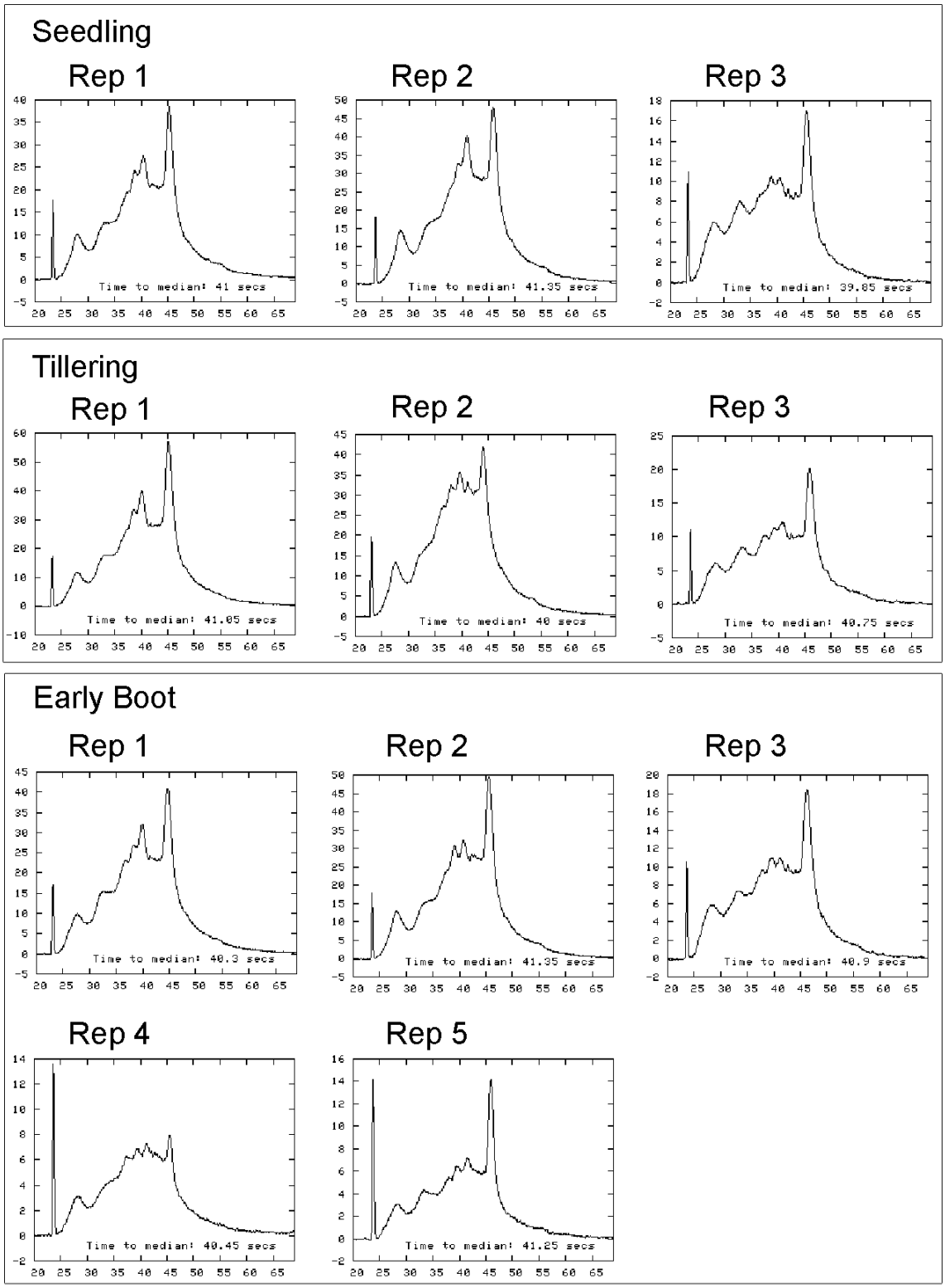
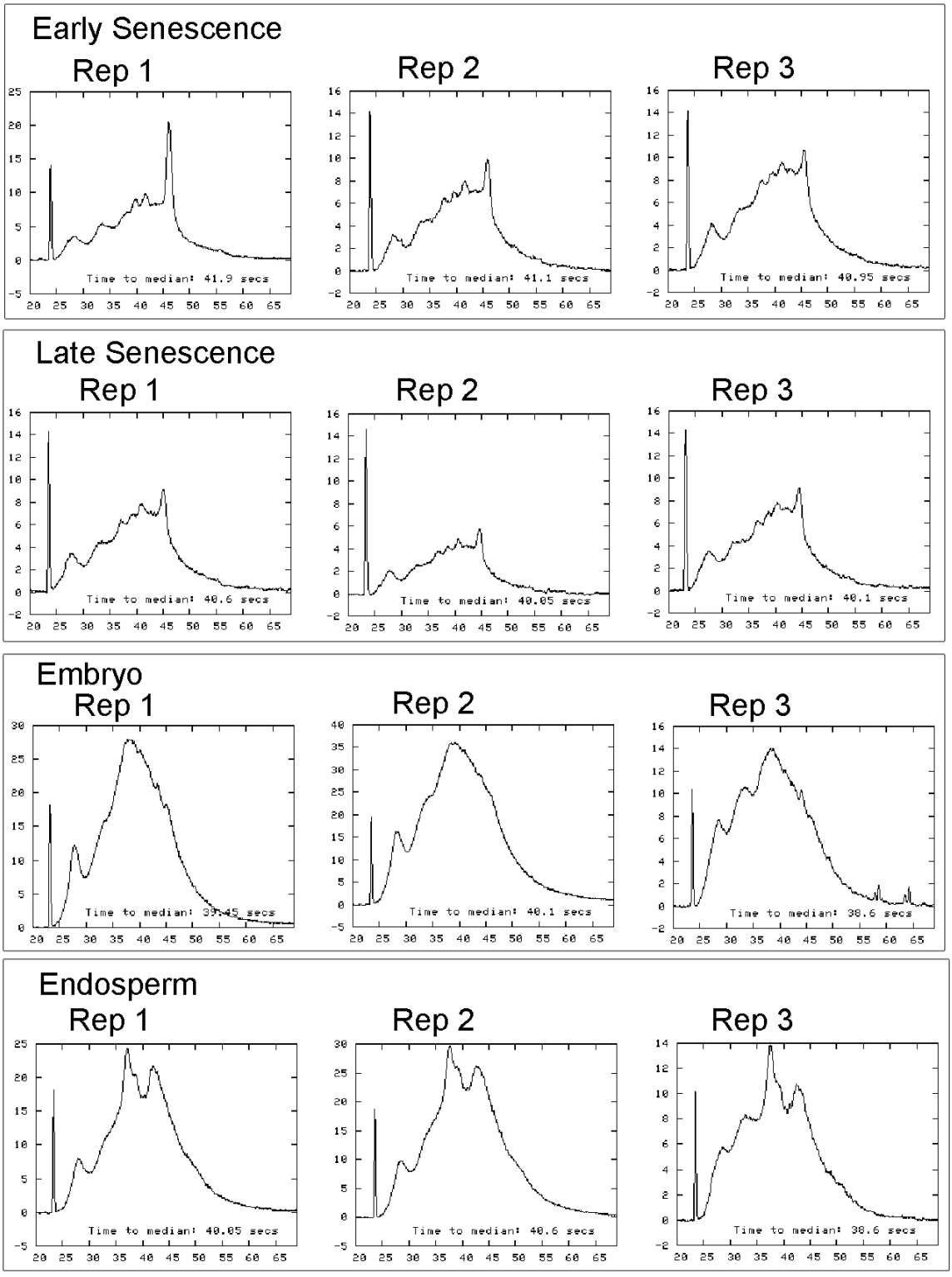
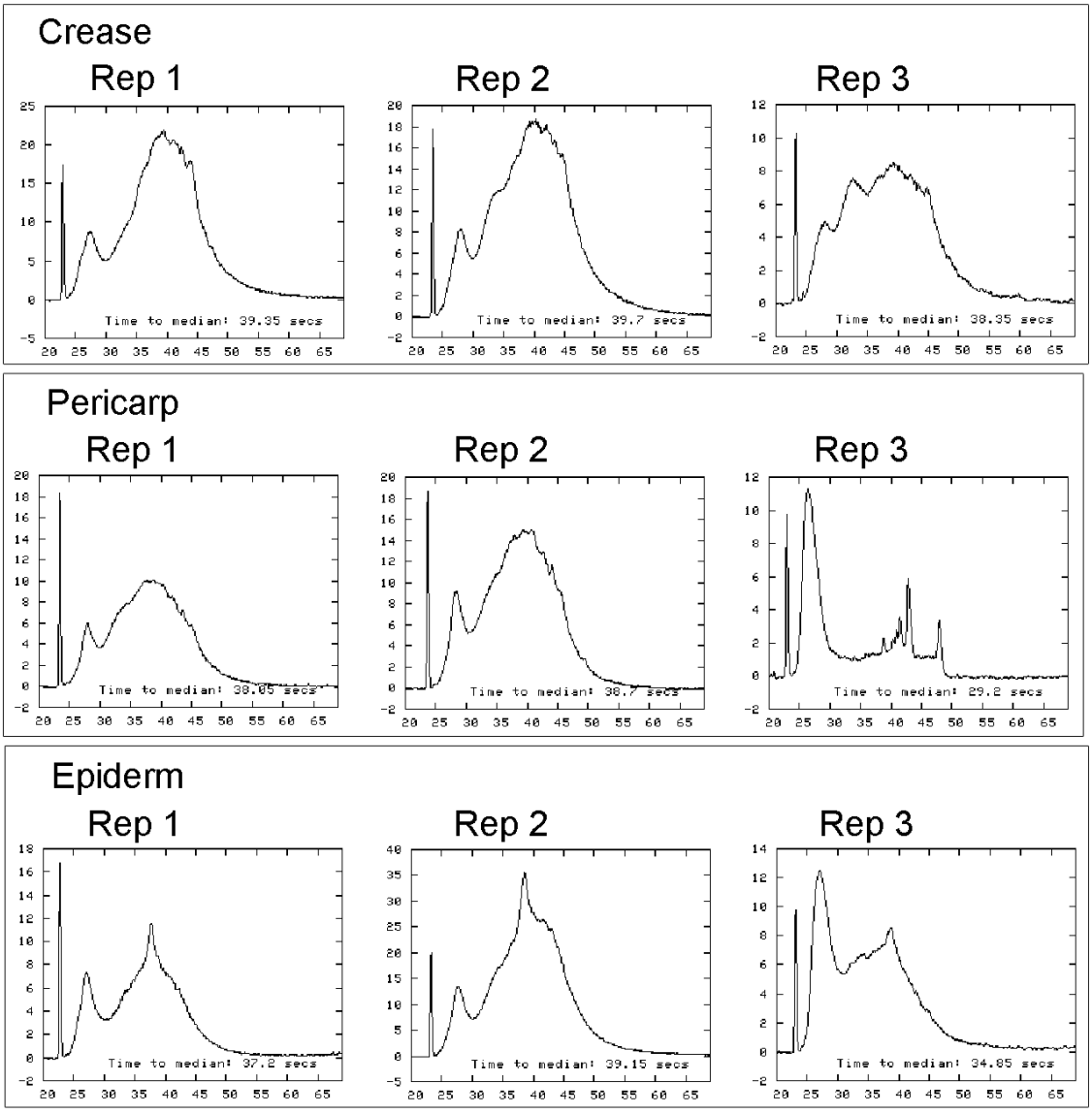

Crease
Rep 1
Rep 2
Rep 3
Pericarp
Rep 1
Rep 2
Rep 3
Epiderm
Rep 1
Rep 2
Rep 3

Figure 4.7 cRNA quality control. Charts produced from Bioanalyzer (Agilent Technologies) analysis of cRNA samples. The x-axis is time in seconds and the y-axis is absorbance. Sharp furthest left peak is an internal reference. Following internal reference peak (23 seconds) large peaks early in the charts indicate small fragments of cRNA and poor quality. Smaller peaks to the left and larger peaks to the right indicate large cRNA fragments and good quality cRNA.


Seedling
Rep 1
Time to median: 41 secs
Rep 2
Time to median: 41.35 secs
Rep 3
Time to median: 39.85 secs
Tillering
Rep 1
Time to median: 41.05 secs
Rep 2
Time to median: 40 secs
Rep 3
Time to median: 40.75 secs
Early Boot
Rep 1
Time to median: 40.3 secs
Rep 2
Time to median: 41.35 secs
Rep 3
Time to median: 40.9 secs
Rep 4
Time to median: 40.45 secs
Rep 5
Time to median: 41.25 secs


Early Senescence
Rep 1
Time to median: 41.9 secs
Rep 2
Time to median: 41.1 secs
Rep 3
Time to median: 40.95 secs
Late Senescence
Rep 1
Time to median: 40.6 secs
Rep 2
Time to median: 40.05 secs
Rep 3
Time to median: 40.1 secs
Embryo
Rep 1
Time to median: 39.45 secs
Rep 2
Time to median: 40.1 secs
Rep 3
Time to median: 38.6 secs
Endosperm
Rep 1
Time to median: 40.05 secs
Rep 2
Time to median: 40.6 secs
Rep 3
Time to median: 38.6 secs


Crease
Rep 1
Time to median: 39.35 secs
Rep 2
Time to median: 39.7 secs
Rep 3
Time to median: 38.35 secs
Pericarp
Rep 1
Time to median: 38.05 secs
Rep 2
Time to median: 38.7 secs
Rep 3
Time to median: 29.2 secs
Epiderm
Rep 1
Time to median: 37.2 secs
Rep 2
Time to median: 39.15 secs
Rep 3
Time to median: 34.85 secs

All hybridizations of samples to wheat GeneChips passed quality control standards (Supplementary Table S4.1). The RMA normalization method includes a background correction, normalization, log transformation and median polishing steps that result in values all above 1 (log_2) with no zero values (Irizarry et al., 2003). RMA normalized expression values on the chips range between 1.92 and 15.02 (and are in logarithmic scale (base 2), Supplementary Table S4.2, columns AE to DM).

To evaluate similarity between biological replicate samples, Pearson correlation coefficient values were calculated among replicates using RMA expression values across each array (Supplementary Table S4.3). Pearson correlation coefficients between biological replicates ranged from 0.946 (tillering tissue) to 0.996 (ovary TNP tissue) for 1 pair-wise comparison with a replicate of the same tissue. Among the vegetative and seed tissues, Pearson correlation coefficients ranged from 0.946 to 0.994 (root, coleoptile, stem21 and stem59 tissues). All Pearson correlations coefficients for biological replicates were close to one and therefore considered acceptable (Draghici et al., 2006) and all samples were kept for further analysis.

4.3.2 Sample comparison

To determine the similarity between sample types (tissue types) a principal component analysis (PCA) in FlexArray was performed on the mean RMA normalized expression values to separate the tissues according to expression variation (Figure 4.8). Anther TNP and pollen have a high degree of expression variation compared to all other samples but are still closer to other reproductive tissues than vegetative tissues in the first principal component which accounts for the largest percentage of variation. Anthers at

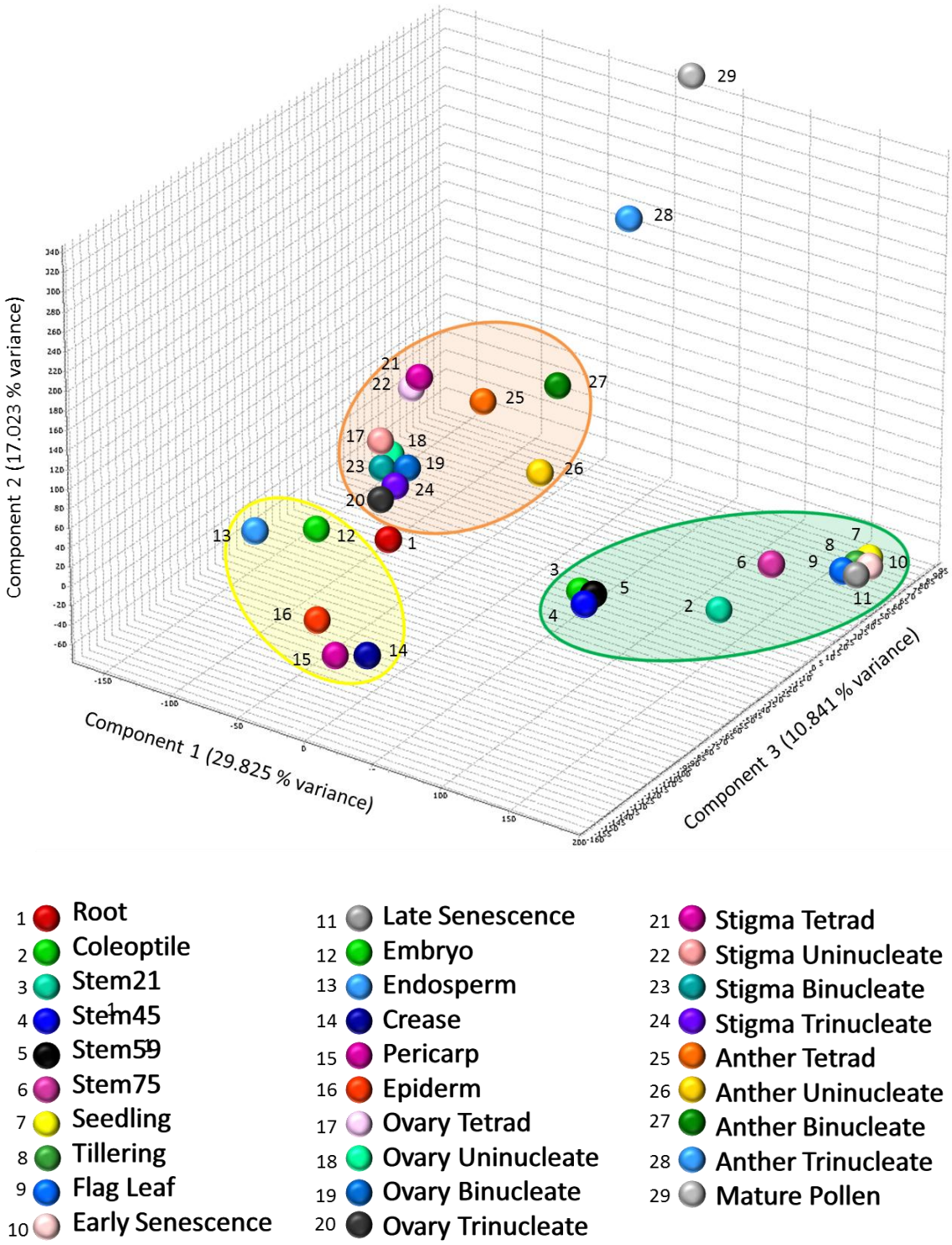


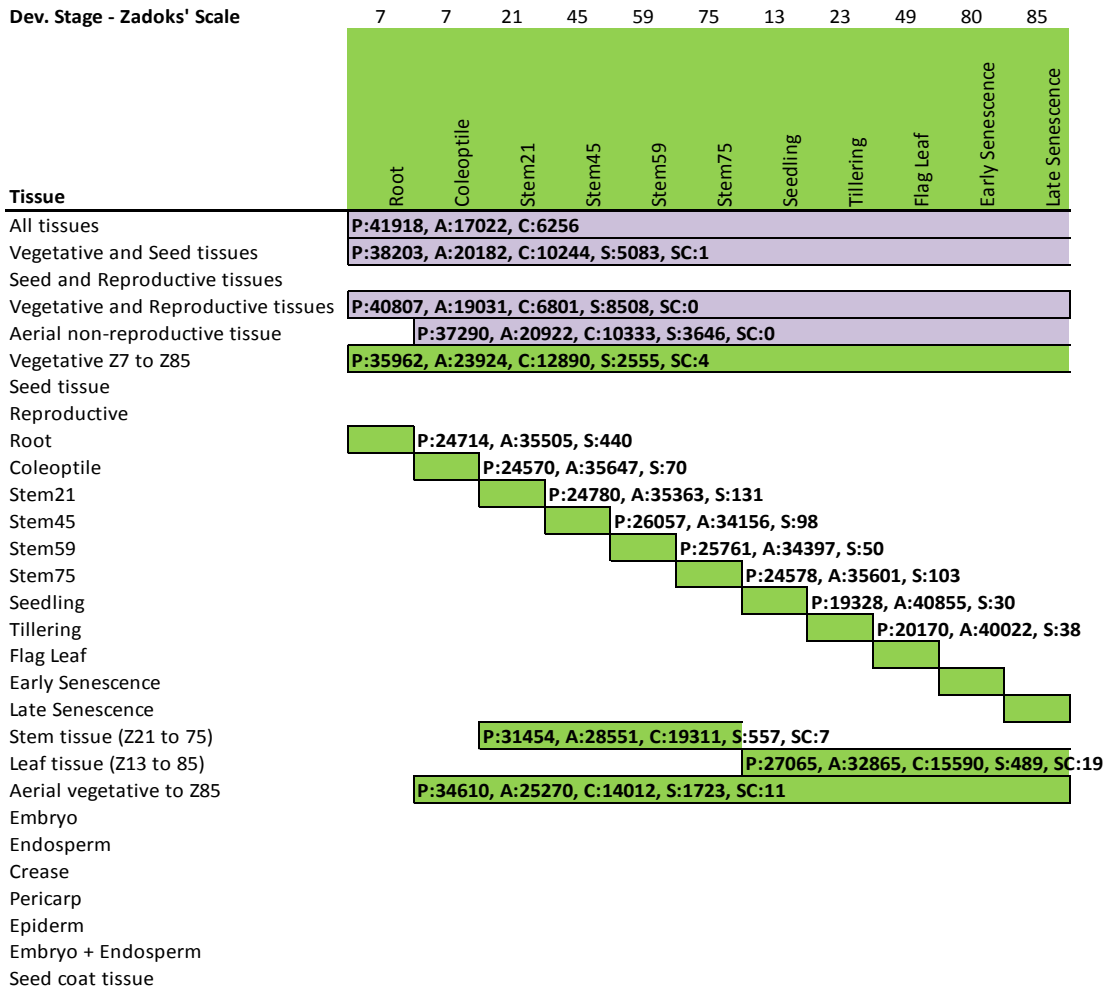
Figure 4.8 PCA plot of 29 RMA normalized sample means. Samples were plotted using PCA to show the relationships among tissues. Three main tissue clusters were observed, vegetative (green oval), seed (yellow oval) and reproductive (orange oval).

this stage of development consist of primarily pollen near maturity so most of the difference is likely due to the unique expression profile in pollen, a result that has also been observed in Arabidopsis expression profiling (Schmid et al., 2005). Root tissue also fell near seed tissues, indicative of the early stage of development that root tissue was collected (seven days after sowing) and therefore it's similarity to its progenitor tissue, embryo. The rest of the other tissues clustered near other similar tissues based on the PCA analysis (Figure 4.8). Samples were therefore assigned to the three tissue types; seed, vegetative and reproductive, for further evaluation and comparison of expression patterns.

4.3.3 Genes expressed in triticale tissues

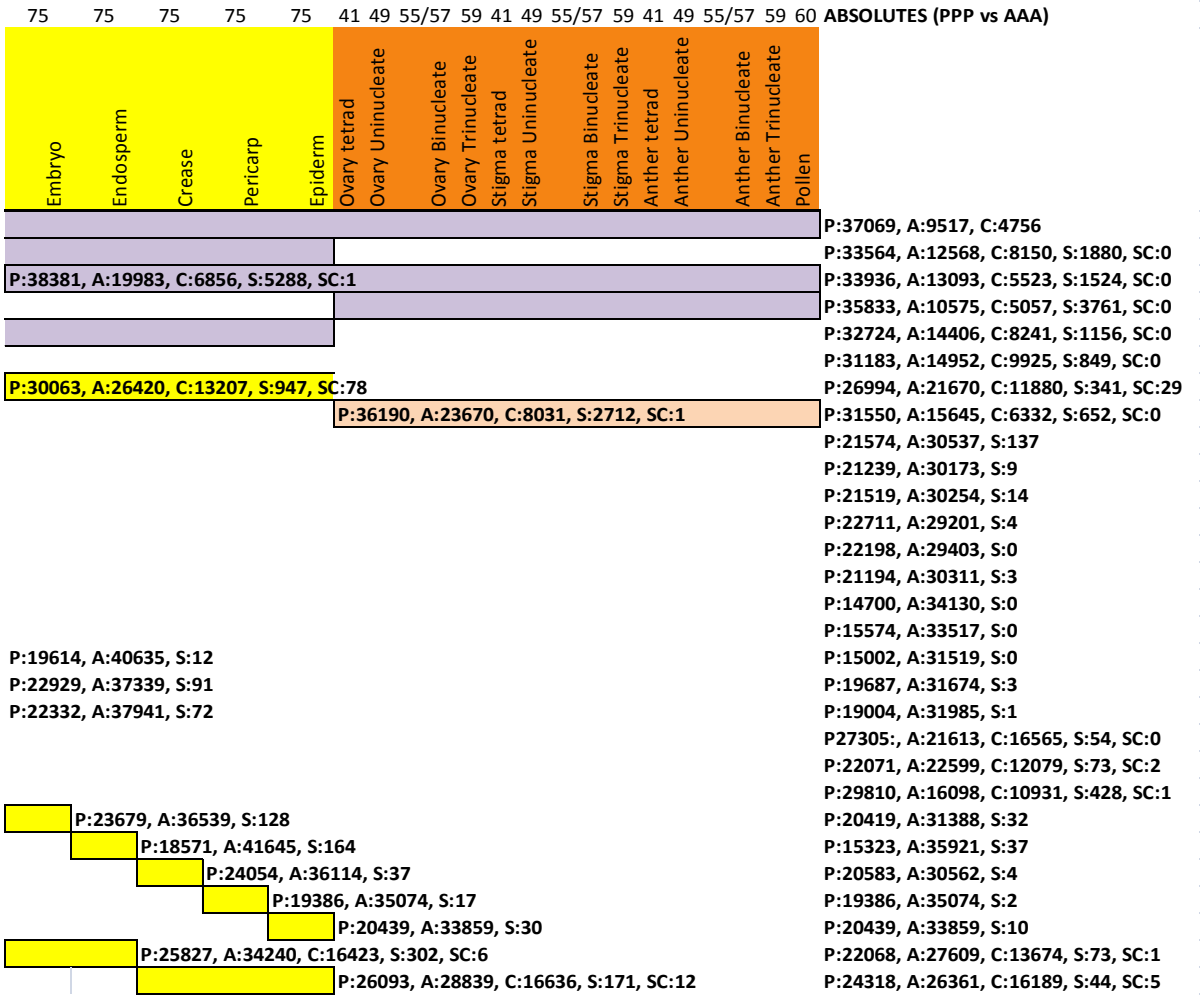
MAS 5.0 consensus detection calls were used to determine the genes expressed in each of the tissues studied. The numbers of probe sets that showed gene expression (P) are listed in Table 4.5. From the entire array, a maximum of 41,918 probe sets were found to show expression in at least one of the tissues studied (using majority P consensus detection calls) and 17,022 probe sets were not represented in the triticale transcriptome from the tissues studied (Table 4.5, A, using majority A consensus detection calls). The remaining 2,175 probe sets had detection calls that did not allow the determination of presence or absence of expression by majority detection call (included marginal [M] detection calls which resulted in neither majority P or majority A consensus detection call). The set of 41,918 probe sets is therefore considered the reference set of probe sets for the determination of functional classifications and enrichment analysis for the triticale tissues in this study.

Table 4.5 Numbers of probe sets in gene lists. Numbers represent presence of tissues (C), specific expression to a particular tissue or tissue group (S) and specifically identified using the stringent protocol of absolute consensus detection calls and all other column of table are probe sets consensus detection calls.



P = expression Present in the tissue or group of tissues
A = expression Absent in the tissue or in all tissues within the group of tissues
C = Constitutively expressed in all tissues in the group of tissues
S = expressed Specifically within the tissue or group of tissues and not expressed (Absent) in all other tissues
SC = Specifically and Constitutively expressed in all tissues in the group of tissues and not expressed in all other tissues

expression (P), absence of expression (A), constitutive expression across a group of
and constitutively expressed across a group of tissues (SC). Numbers listed in far right
numbers represent probe sets identified using the permissive protocol of majority



The probe sets that represent gene expression in the triticale tissues studied were classified according to their GO slim (generic, developed by the GO consortium, www.geneontology.org) functional annotations (refer to Chapter 3 for wheat array annotation information). Figures 4.9-4.11 show the functional classifications of sequences expressed (majority P consensus detection calls) in any of the tissues studied as well as the numbers of sequences found to be expressed in the three main groups of tissues (vegetative, seed and reproductive). Due to the generalization of terms in the GO slim categories used for summarization, the trends in numbers of genes expressed in each category are similar between the tissue types and are proportional to the total numbers of genes expressed in each group of tissues (+/- up to 1.14% per category).

The 17,022 probe sets that were not found to be expressed in any of the tissues studied may represent wheat transcripts not present in triticale (from the D genome for example) and transcripts expressed under environmental conditions not examined in this study (biotic and abiotic stresses for example). Functional analysis of the probe sets not expressed revealed that, indeed, some of the biological processes that were over-represented in the list of non-expressed probe sets when compared to expressed probe sets included stress response terms such as response to organic cyclic substances, programmed cell death and apoptosis, spore germination, mycelium development, instar larval or pupal development and a number of processes that involve symbiotic organisms (See Supplementary Table S4.4 for a complete list of enriched terms).

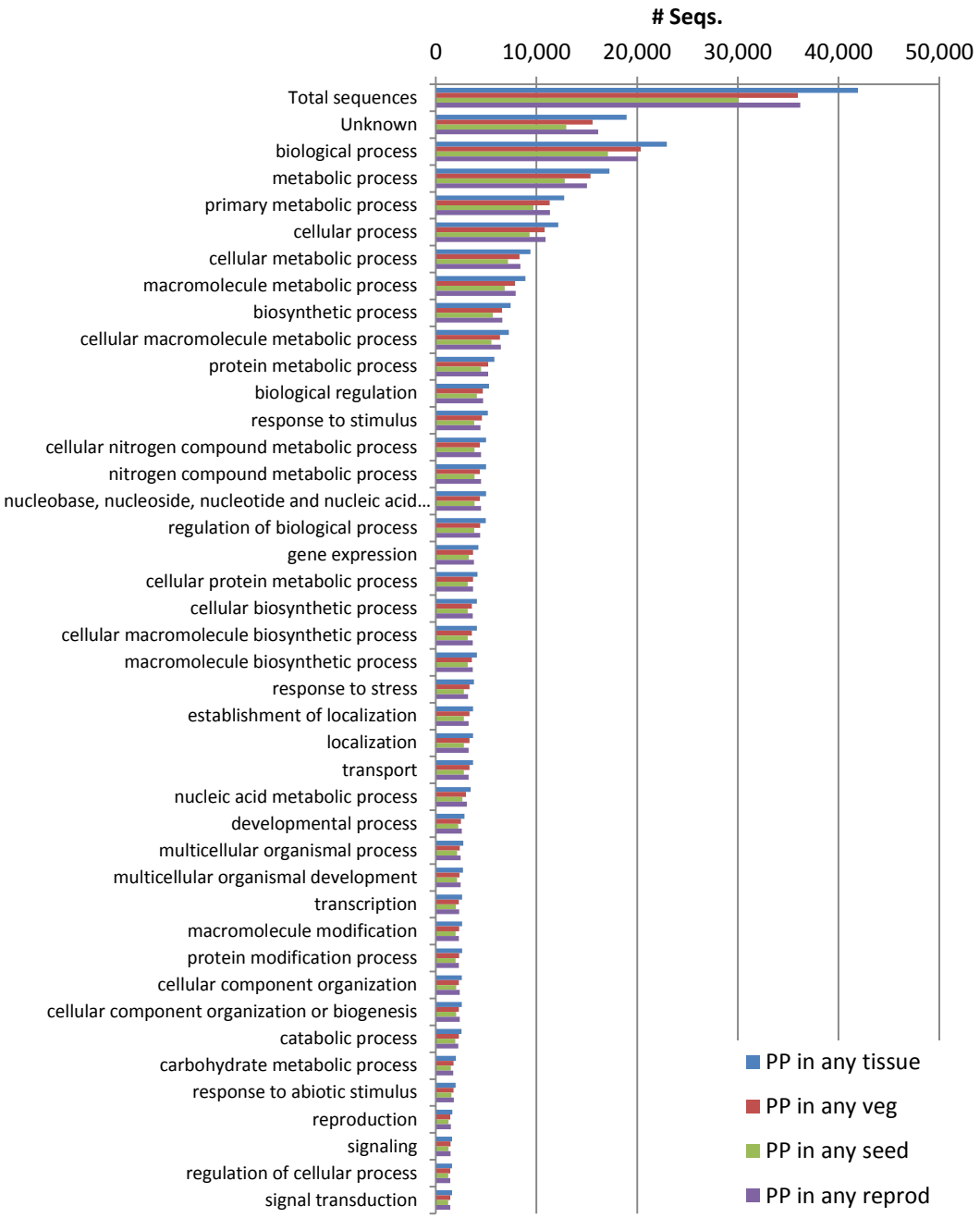


Figure 4.9 Biological processes represented by probe sets found to represent expression in triticale tissues. Only the top 40 most represented biological process GO slim terms are shown. PP indicates probe sets with majority P consensus detection calls (at least 2 of 3 biological replicates have a P detection call) were used for functional classification.

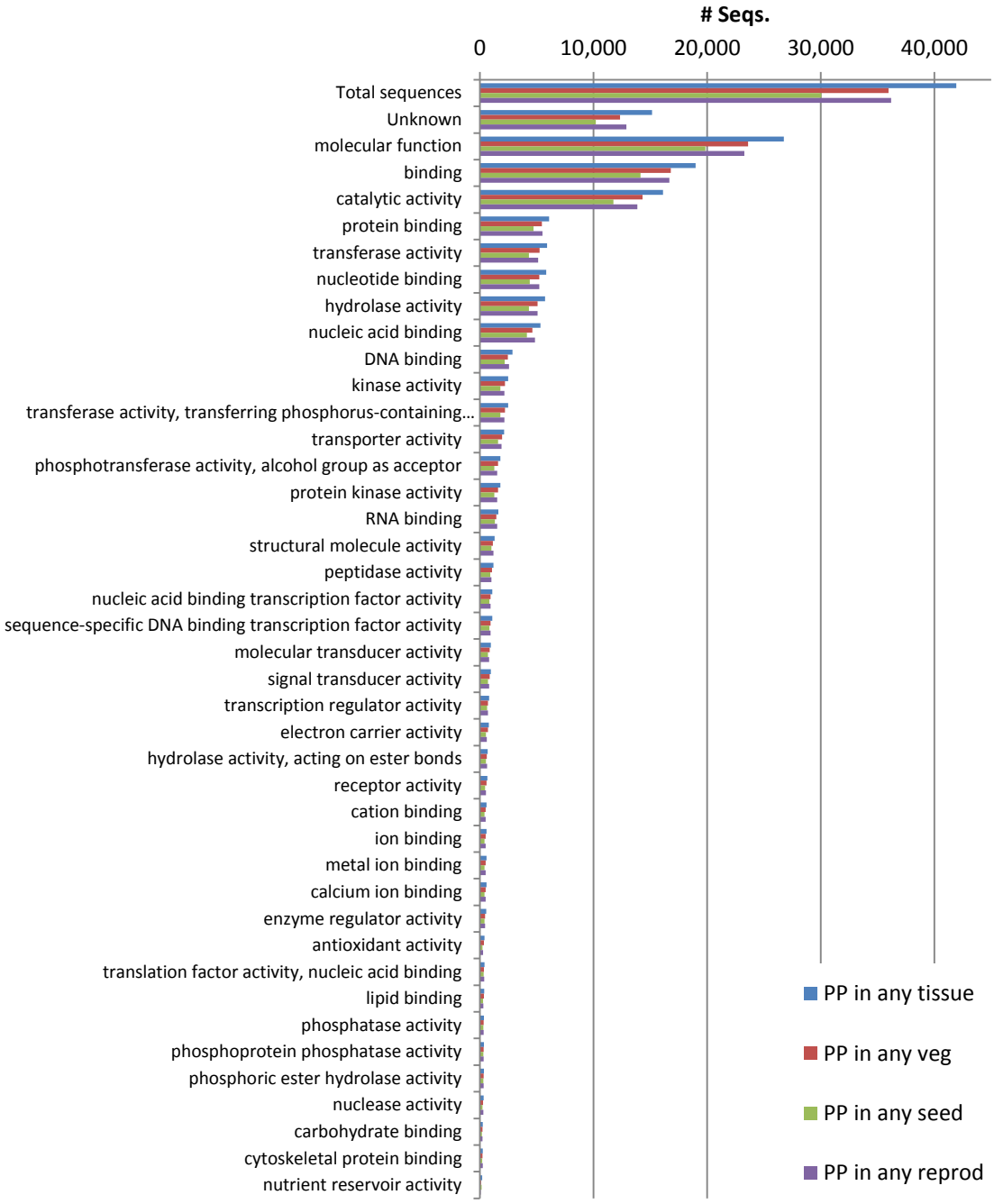


Figure 4.10 Molecular functions represented by probe sets found to represent expression in triticale tissues. Only the top 40 most represented molecular function GO slim terms are shown. PP indicates probe sets with majority P consensus detection calls (at least 2 of 3 biological replicates have a P detection call) were used for functional classification.

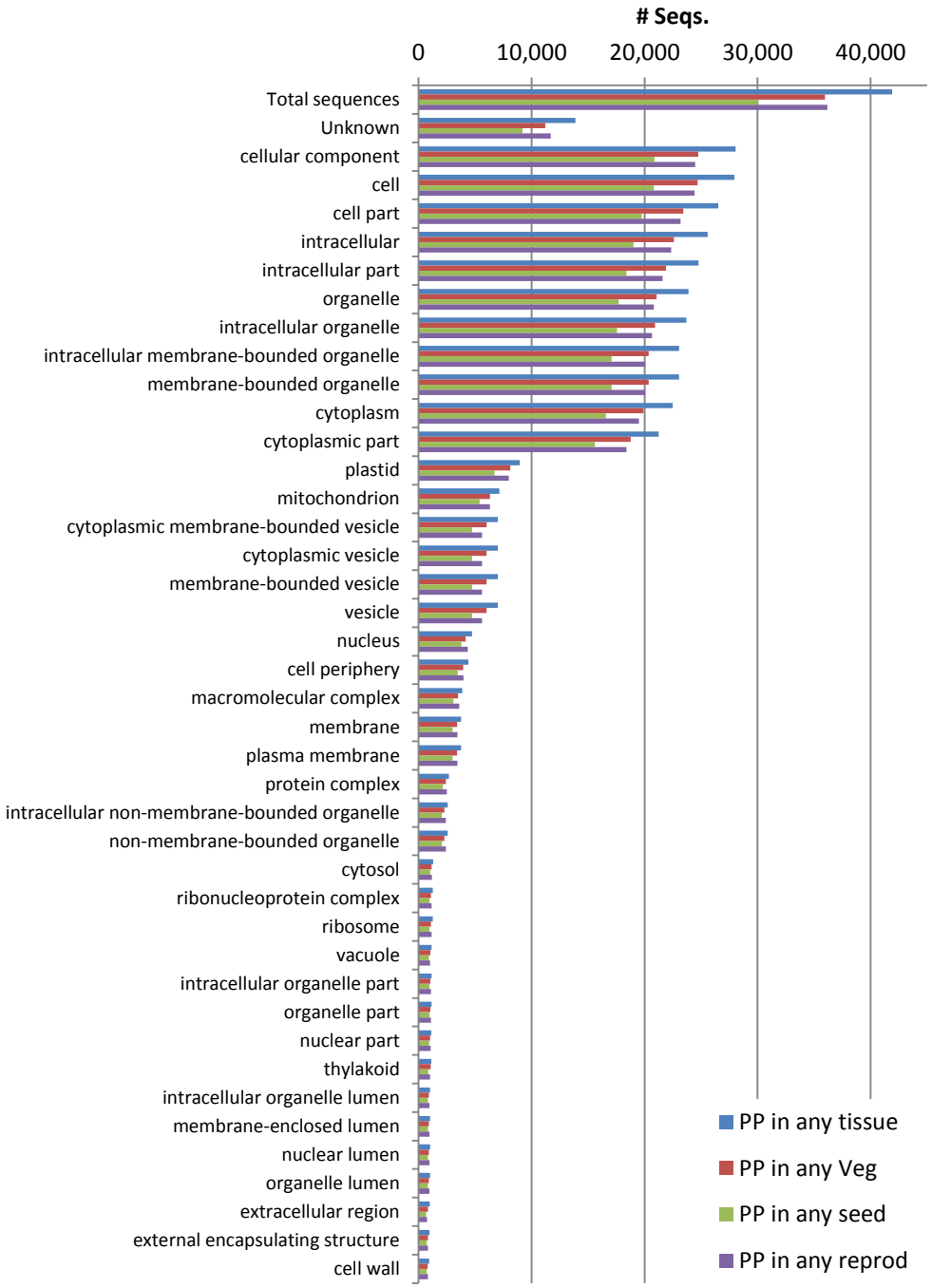


Figure 4.11 Cellular components represented by probe sets found to represent expression in triticale tissues. Only the top 40 most represented cellular component GO slim terms are shown. PP indicates probe sets with majority P consensus detection calls (at least 2 of 3 biological replicates have a P detection call) were used for functional classification.

4.3.4 Triticale genes with constitutive expression

Probe sets that were present in all tissues studied (C) were determined to be those representing transcripts that are constitutively expressed (Table 4.5). Transcripts that were constitutively expressed in all tissues studies were represented by 6,256 probe sets (Table 4.5, permissive protocol). It was expected that many of these probe sets represent housekeeping genes and therefore are essential to plant cellular functions. Functional analysis of the annotations of these probe sets in comparison to all probe sets expressed in any tissue revealed that there was enrichment in several basic molecular functions such as translation initiation and elongation factor activities, GTP binding, electron carrier activity, enzyme binding and DNA polymerase activity in the constitutively expressed gene set. Biological processes that are enriched in this gene list include protein folding, post-translational protein modification, cellular membrane organization, ATP synthesis coupled proton transport, nucleosome assembly and glycolysis to name a few. For a complete list of enriched terms in the constitutively expressed gene list see Supplementary Table S4.5.

Constitutive expression was examined not only for the entire set of tissues examined but also for groups of tissues such as vegetative and seed tissues to identify genes expressed throughout tissue groups (Table 4.5, C). It was found that as more tissue samples were included in a tissue group, the fewer probe sets were found to show constitutive expression across all tissues within the group. The seed tissue group consists of 5 different sample types and 13,207 probe sets represent transcripts that are constitutively expressed across these tissues, while the reproductive tissue group consists

of 13 different sample types and only 8,031 probe sets were constitutively expressed across all reproductive tissues (Table 4.5).

4.3.5 Tissue-specific expression

Table 4.5 lists the numbers of probe sets that were identified as being expressed (P), not expressed (A) and specific (S) in each of the individual vegetative (green) and seed (yellow) tissues examined. This table also displays the numbers of probe sets that are expressed (P), not expressed (A), constitutively expressed (C), specifically expressed (S) and constitutively expressed and specific (SC) to groups of tissues. While reproductive tissues were not examined individually in this study, the results as a group of tissues are included for comparison and because identification of the probe sets not expressed in these tissues (A) was critical to determine tissue-specificity in vegetative and seed tissues. The numbers of probe sets identified in each of these categories is presented for both the permissive (majority consensus detection calls) and stringent (absolute consensus detection calls) strategies in Table 4.5.

Supplementary Table S4.6 contains all the data for all gene lists shown in Table 4.5. This supplementary table lists all probe sets on the wheat array in rows and all gene lists from Table 4.5 in columns. Each probe set either belongs to a gene list and therefore labeled as TRUE, or is absent in a gene list and labeled as FALSE in a particular column of the table. Consensus MAS5.0 detection calls and mean RMA expression values are also provided in Supplementary Table S4.6 for reference. As well, the merged annotations from Chapter 3 are also provided in this table for functional annotation information alongside a few annotation columns from the Affymetrix® annotation release

31 (August 11, 2010). This table can be sorted by any of the columns to provide a list of probe sets that meet any of the expression criteria listed in Table 4.5 or to examine a tissue or group of tissues of interest fitting multiple criteria.

Probe sets representing tissue-specific expression (S) were identified as those having a present consensus detection call (P) in the tissue of interest and having an absent consensus detection call (A) in all other tissues. For the individual vegetative and seed tissue samples studied, the individual tissue sample with the largest number of tissue-specifically expressed probe sets is the root tissue (Table 4.5, S:440) followed by endosperm (S:164), young stem tissue (Stem21, S:131) and embryo tissue (S:128) respectively (permissive protocol, majority consensus detection call).

Table 4.5 also lists the numbers of tissue-specific probe sets identified using the stringent protocol of absolute consensus detection calls (far right column). Supplementary Table S4.6 can be filtered for probe sets meeting either the permissive or stringent protocol criteria and can be used to find the overlap between these protocols.

While the probe sets identified as tissue-specific using the stringent protocol are considered to represent tissue-specific expression with a higher level of confidence than those identified using the permissive protocol, there are very few tissue-specific probe sets identified and it is possible that a high number of false negatives were identified using the stringent protocol. On the other hand, the permissive method may include false positives. For individual tissues all tissue-specific (S) probe sets identified using the stringent method were also identified using the permissive method and this overlap represent the probe sets that should be considered as true positives with the highest degree of confidence. The permissive protocol is presented to provide the highest

number of probable tissue-specific expression with a reasonable level of confidence. It may be advisable to perform supplementary transcriptional quantification such as qRT-PCR to verify the expression patterns of candidate genes chosen for further investigation based on only the permissive protocol.

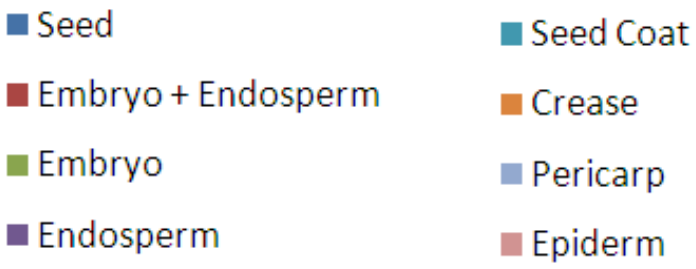
Tissue type-specific probe sets were identified as those having a consensus detection call of present (P) in at least one tissue within the tissue group of interest and having detection consensus calls of absent (A) in all tissues not within the tissue group. Vegetative, seed, reproductive, stem, leaf, aerial vegetative, progeny (embryo and endosperm) and seed coat are all tissue groups that were examined for expression specificity (Table 4.5, S). For the three main groups of tissues (vegetative, seed and reproductive) the reproductive tissues had the highest number of tissue type-specifically expressed probe sets (2,712) followed by vegetative (2,555) and seed (947).

Functional enrichment analysis was performed on tissue-specific and tissue type-specific probe set lists. Functional annotations for these lists were compared to the functional annotations of the reference list of expressed probe sets (41,918 probe sets) using Fisher's exact test in the Blast2Go software (Conesa and Götz, 2008). This analysis identified biological process, molecular function and cellular component GO annotation terms that are enriched in the genes specifically expressed in a particular tissue or tissue type and provides insight into tissue differentiation and development. Enrichment results were filtered by a p-value < 0.05, the most specific terms in each branch of the ontology and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations.

4.3.5.1 Functional enrichment in seed-specific genes

Figures 4.12-4.14 show the results of the functional analysis of seed tissue-specific probe set lists. Each chart shows the enriched terms belonging to one of the three GO branches, biological processes (Figure 4.12) molecular functions (Figure 4.13), and cellular components (Figure 4.14). A number of biological processes were found to be over-represented in the seed tissues as well as some processes that were under-represented (Figure 4.12). The under-represented processes are processes that are generally required for cellular functions in all plant cells including organelle organization, cellular catabolic processes, response to chemical stimulus and cellular component biogenesis to name a few from Figure 4.12. Therefore, these processes are underrepresented by seed-tissue-specific genes when compared to their representation by genes constitutively expressed in all tissues. While many of the biological processes that were found to be enriched in the embryo were expected based on the developmental stage of the embryo tissue sampled, including gibberellic acid mediated signaling pathway, negative regulation of photomorphogenesis and negative regulation of seed germination, it was expected that there be more macromolecule biosynthesis processes, representing seed storage compound biosynthesis, be enriched in the endosperm than were revealed by the results. Glycogen biosynthesis was the only macromolecule biosynthetic process enriched in the endosperm specific genes and only at a level of 2.56% higher than the reference set. While this is surprising, it may be that these processes are represented at a higher frequency by the genes that are not specific to the endosperm but by genes that are up-regulated in the endosperm. The majority of biological processes that were enriched in seed-specific gene lists were represented by seed coat-specific genes. The enrichment

Figure 4.12 Biological processes enriched in seed tissue-specific probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe sets (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of probe sets. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents tissue-specific gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list. One category ends with … due to space limitations and should read as "positive regulation of gene-specific transcription from RNA polymerase II promoter."



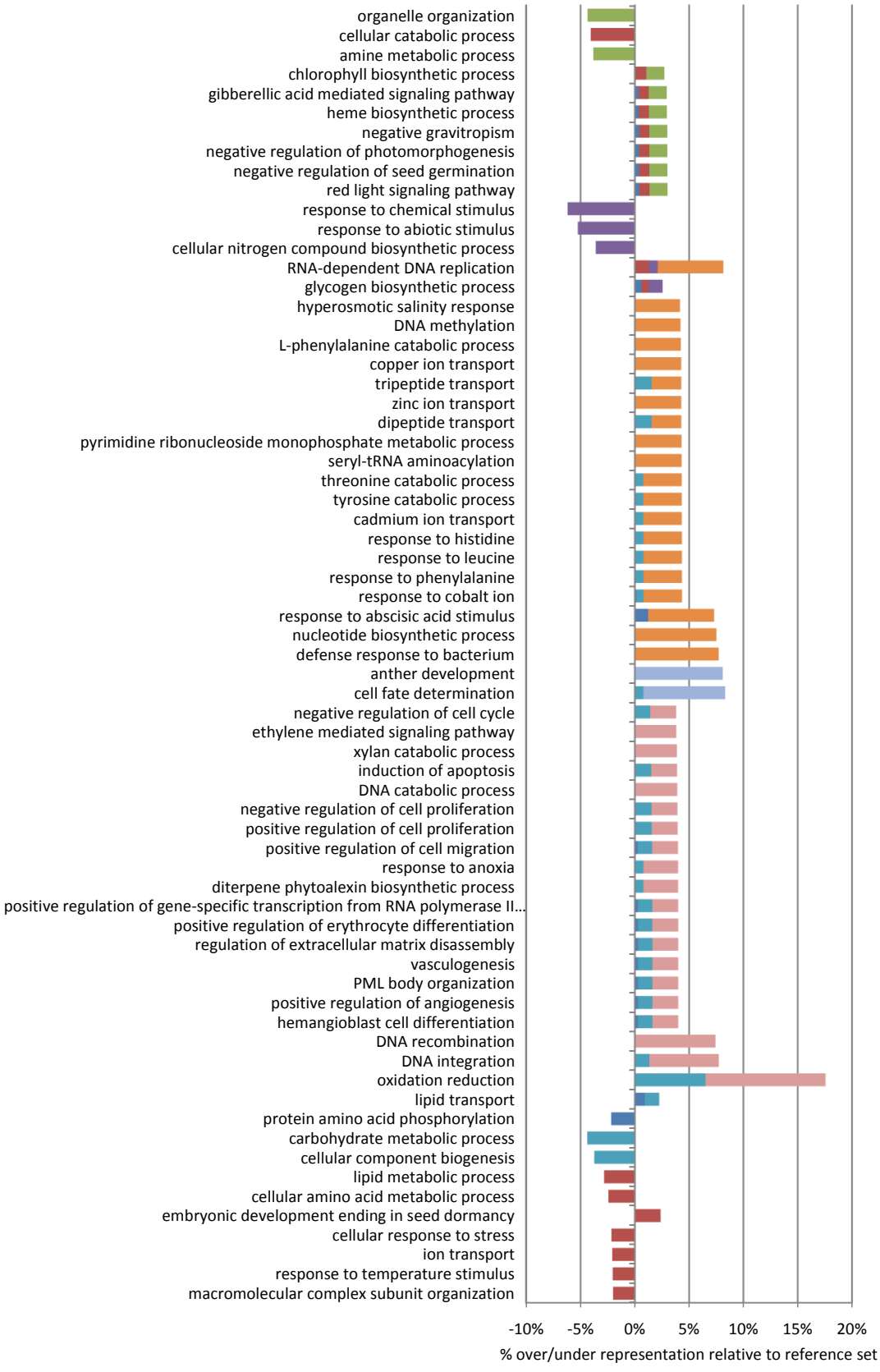
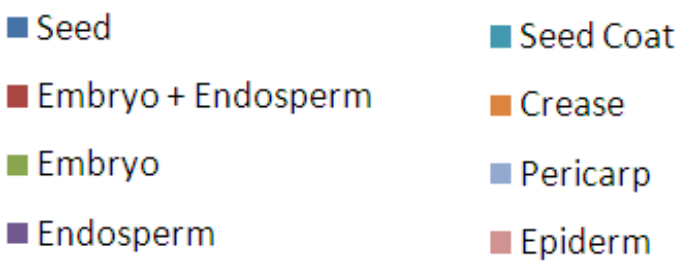

organelle organization
cellular catabolic process
amine metabolic process
chlorophyll biosynthetic process
gibberellic acid mediated signaling pathway
heme biosynthetic process
negative gravitropism
negative regulation of photomorphogenesis
negative regulation of seed germination
red light signaling pathway
response to chemical stimulus
response to abiotic stimulus
cellular nitrogen compound biosynthetic process
RNA-dependent DNA replication
glycogen biosynthetic process
hyperosmotic salinity response
DNA methylation
L-phenylalanine catabolic process
copper ion transport
tripeptide transport
zinc ion transport
dipeptide transport
pyrimidine ribonucleoside monophosphate metabolic process
seryl-tRNA aminoacylation
threonine catabolic process
tyrosine catabolic process
cadmium ion transport
response to histidine
response to leucine
response to phenylalanine
response to cobalt ion
response to abscisic acid stimulus
nucleotide biosynthetic process
defense response to bacterium
anther development
cell fate determination
negative regulation of cell cycle
ethylene mediated signaling pathway
xylan catabolic process
induction of apoptosis
DNA catabolic process
negative regulation of cell proliferation
positive regulation of cell proliferation
positive regulation of cell migration
response to anoxia
diterpene phytoalexin biosynthetic process
positive regulation of gene-specific transcription from RNA polymerase II...
positive regulation of erythrocyte differentiation
regulation of extracellular matrix disassembly
vasculogenesis
PML body organization
positive regulation of angiogenesis
hemangioblast cell differentiation
DNA recombination
DNA integration
oxidation reduction
lipid transport
protein amino acid phosphorylation
carbohydrate metabolic process
cellular component biogenesis
lipid metabolic process
cellular amino acid metabolic process
embryonic development ending in seed dormancy
cellular response to stress
ion transport
response to temperature stimulus
macromolecular complex subunit organization
-10%
-5%
0%
5%
10%
15%
20%
% over/under representation relative to reference set

Figure 4.13 Molecular functions enriched in seed tissue-specific probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe sets (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of probe sets. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-specific gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list.



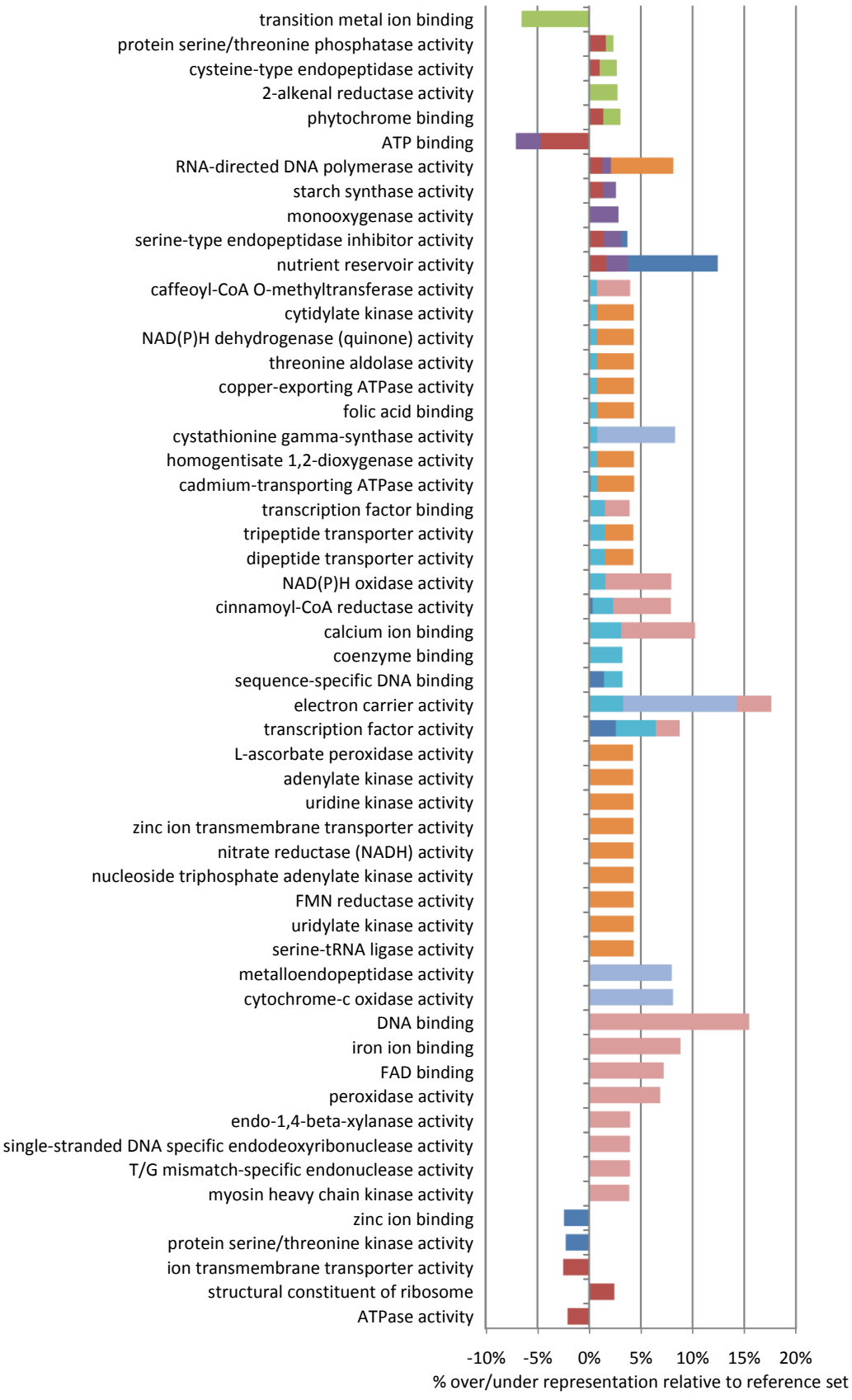
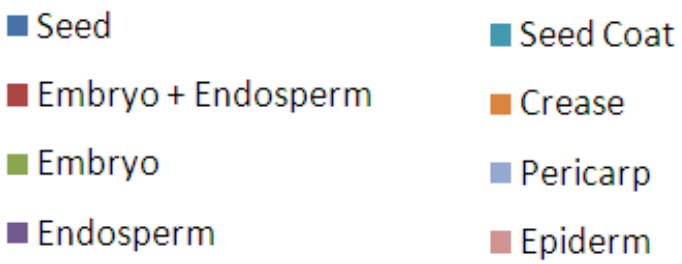
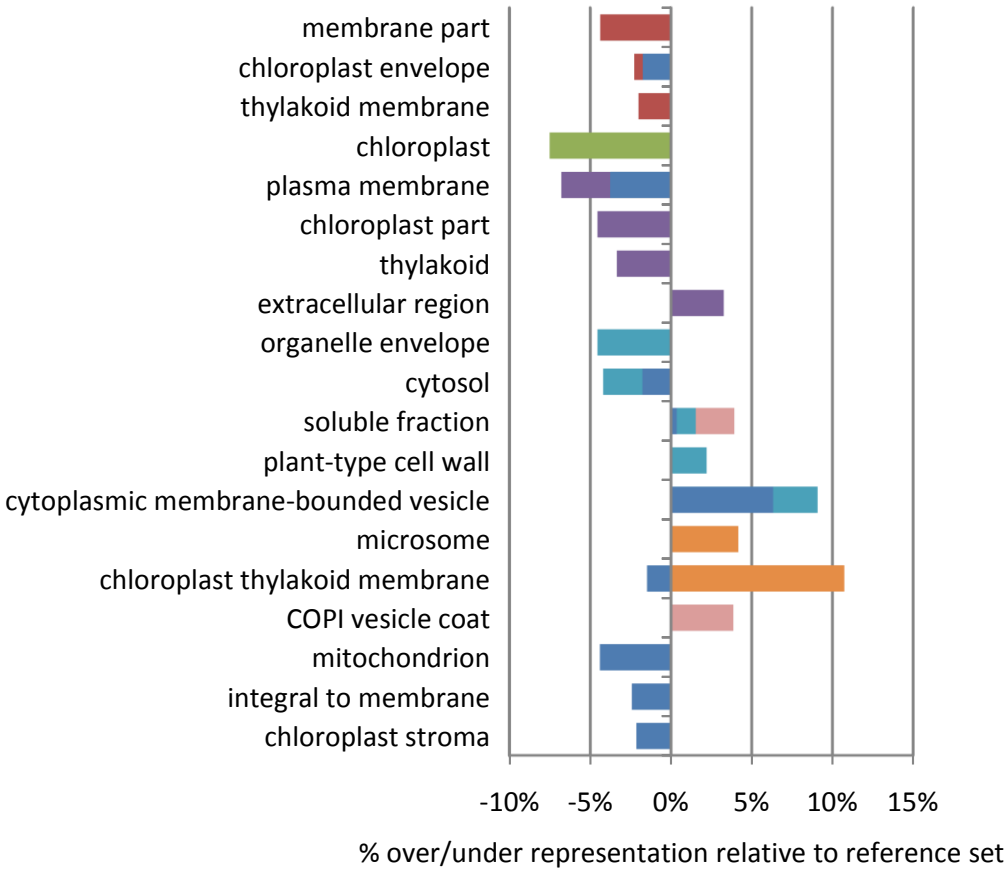

transition metal ion binding
protein serine/threonine phosphatase activity
cysteine-type endopeptidase activity
2-alkenal reductase activity
phytochrome binding
ATP binding
RNA-directed DNA polymerase activity
starch synthase activity
monooxygenase activity
serine-type endopeptidase inhibitor activity
nutrient reservoir activity
caffeoyl-CoA O-methyltransferase activity
cytidylate kinase activity
NAD(P)H dehydrogenase (quinone) activity
threonine aldolase activity
copper-exporting ATPase activity
folic acid binding
cystathionine gamma-synthase activity
homogentisate 1,2-dioxygenase activity
cadmium-transporting ATPase activity
transcription factor binding
tripeptide transporter activity
dipeptide transporter activity
NAD(P)H oxidase activity
cinnamoyl-CoA reductase activity
calcium ion binding
coenzyme binding
sequence-specific DNA binding
electron carrier activity
transcription factor activity
L-ascorbate peroxidase activity
adenylate kinase activity
uridine kinase activity
zinc ion transmembrane transporter activity
nitrate reductase (NADH) activity
nucleoside triphosphate adenylate kinase activity
FMN reductase activity
uridylate kinase activity
serine-tRNA ligase activity
metalloendopeptidase activity
cytochrome-c oxidase activity
DNA binding
iron ion binding
FAD binding
peroxidase activity
endo-1,4-beta-xylanase activity
single-stranded DNA specific endodeoxyribonuclease activity
T/G mismatch-specific endonuclease activity
myosin heavy chain kinase activity
zinc ion binding
protein serine/threonine kinase activity
ion transmembrane transporter activity
structural constituent of ribosome
ATPase activity
-10% -5% 0% 5% 10% 15% 20%
% over/under representation relative to reference set

Figure 4.14 Cellular components enriched in seed tissue-specific probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe sets (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of probe sets. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-specific gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list.




membrane part
chloroplast envelope
thylakoid membrane
chloroplast
plasma membrane
chloroplast part
thylakoid
extracellular region
organelle envelope
cytosol
soluble fraction
plant-type cell wall
cytoplasmic membrane-bounded vesicle
microsome
chloroplast thylakoid membrane
COPI vesicle coat
mitochondrion
integral to membrane
chloroplast stroma
-10%
-5%
0%
5%
10%
15%
% over/under representation relative to reference set

of the biological process oxidation reduction was the highest of any process in any seed tissue at 17.56% over representation in the epiderm. Flavenoid biosynthesis occurs in the seed coat and produces a number of compounds that contribute to a number of physiological functions including seed coat color, dormancy, viability and disease resistance (Debeaujon et al., 2000; Lepiniec et al., 2006). The oxidation of flavonoids such as proanthocyanidin during the course of seed desiccation leads to the formation of brown pigments that confer color to the mature seed (Debeaujon et al., 2003). Oxidation reactions during flavonoid biosynthesis processes in the seed coat may be the cause of such high over-representation of oxidation-reduction in the epiderm.

While only a few molecular functions were found to be enriched in the probe sets that were specifically expressed in the seed (including serine-type endopeptidase inhibitor activity, nutrient reservoir activity, sequence-specific DNA binding and transcription factor activity) these terms and several others were found to be enriched within individual seed tissue-specific probe set lists and in some cases at higher levels than in the entire seed-specific probe set list (Figure 4.13). For example transcription factor activity was over-represented in the seed by 2.58% but is over-represented in the seed coat by 6.45% and in the epiderm by 8.76% higher than in the reference set (Figure 4.13). These results tell us that enrichment of transcription factor activity in the seed is primarily within the epiderm tissue. The epiderm tissue-specific gene list was also highly enriched in DNA binding (15.48% over representation) suggesting a high level of transcriptional regulation in the epiderm tissue. One surprise in the molecular function results of the enrichment analysis is that the nutrient reservoir activity was enriched at a high level in the seed at 12.43% but was not enriched at an even higher level in the storage compartment of the

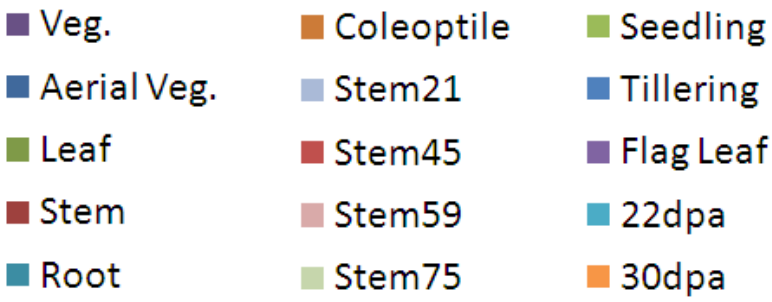
seed, the endosperm but at a lower level of only 3.77%. A few functions were under-represented in the seed tissues including transition metal ion binding in the embryo tissue and ATP binding in the embryo and endosperm tissue. This under-representation is indicative that these functions are not specific to these seed tissues but are found in plant tissues other than seed tissues.

The seed-specific gene lists are also over and under-represented in a number of GO terms associated with cellular components (Figure 4.14). The majority of chloroplast related GO's terms were under-represented in the seed tissue-specific gene lists as expected due to the reduced amount of photosynthesis that occurs in seed tissues. The seed is primarily a storage organ and the products of photosynthesis are transported to the seed from other more photosynthetically active tissues such as leaf tissue (Taiz and Zeiger, 2002). The embryo and endosperm make up the majority of the seed volume and are not chlorophyll-containing green tissues so it is reasonable that genes expressed specifically within these tissues are under-represented in the chloroplast related GO term annotations. A few cellular component terms were over-represented including extracellular region in the endosperm, microsome and chloroplast thylakoid membrane in the crease, soluble fraction and COPI vesicle coat in the epiderm, and plant-type cell wall and cytoplasmic membrane-bounded vesicle in the seed coat (Figure 4.14).

4.3.5.2 Functional enrichment in vegetative-specific genes

Gene lists specific to vegetative tissues and groups of vegetative tissues were also analyzed for functional enrichment (Figures 4.15-4.17). The biological processes that are enriched within vegetative tissue-specific gene lists shed light on some of the

Figure 4.15 Biological processes enriched in vegetative tissue-specific probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe sets (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of probe sets. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-specific gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list. One category ends with … due to space limitations and should read as "negative regulation of abscisic acid mediated signaling pathway."



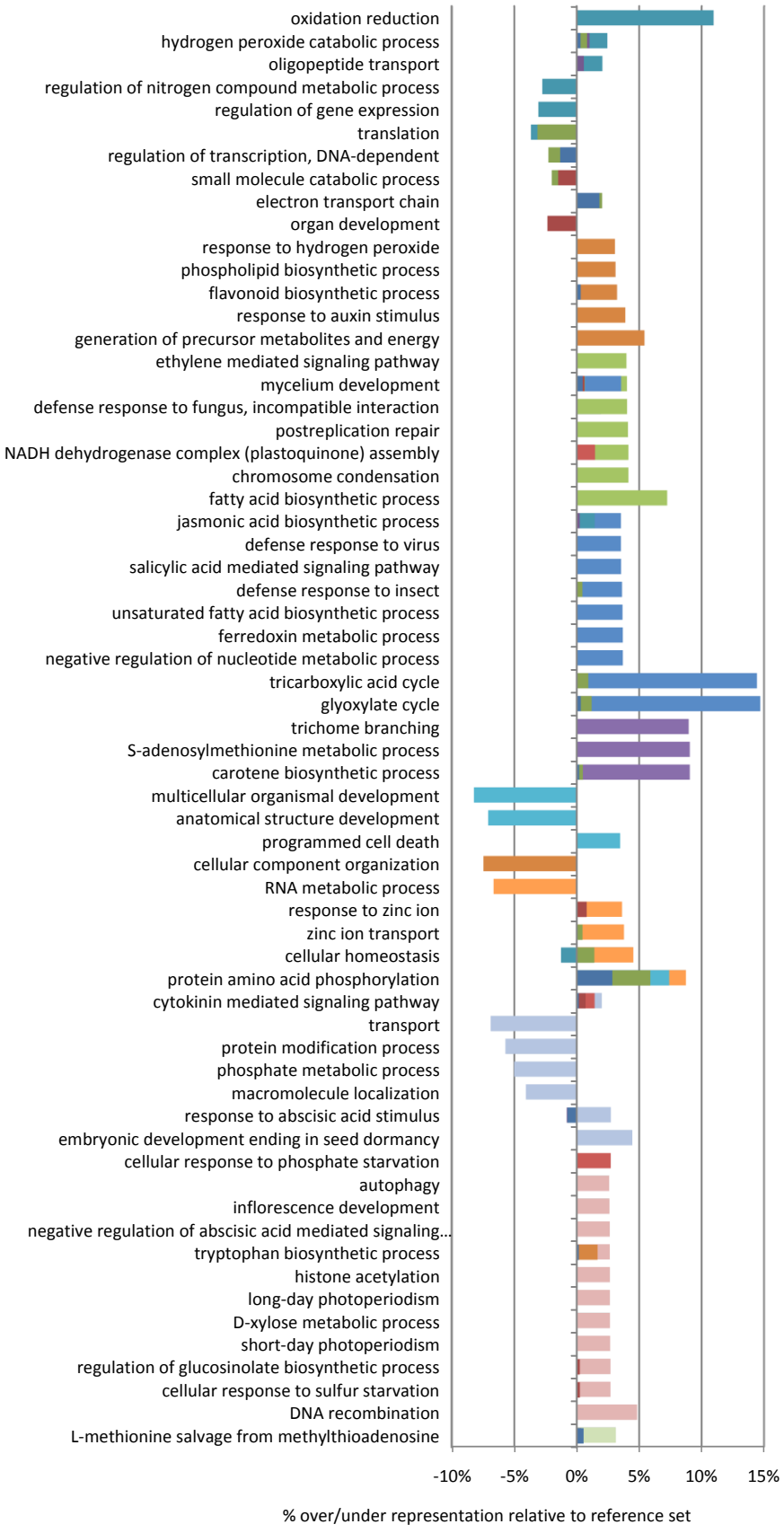
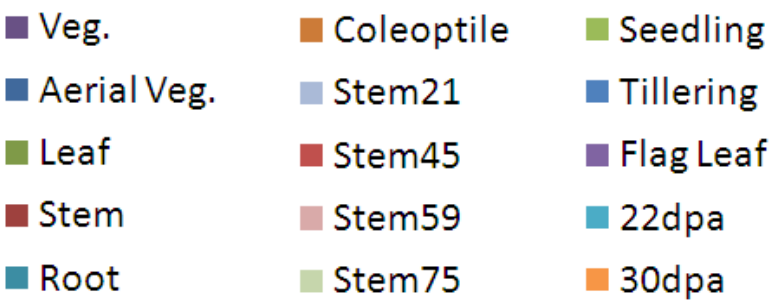

oxidation reduction
hydrogen peroxide catabolic process
oligopeptide transport
regulation of nitrogen compound metabolic process
regulation of gene expression
translation
regulation of transcription, DNA-dependent
small molecule catabolic process
electron transport chain
organ development
response to hydrogen peroxide
phospholipid biosynthetic process
flavonoid biosynthetic process
response to auxin stimulus
generation of precursor metabolites and energy
ethylene mediated signaling pathway
mycelium development
defense response to fungus, incompatible interaction
postreplication repair
NADH dehydrogenase complex (plastoquinone) assembly
chromosome condensation
fatty acid biosynthetic process
jasmonic acid biosynthetic process
defense response to virus
salicylic acid mediated signaling pathway
defense response to insect
unsaturated fatty acid biosynthetic process
ferredoxin metabolic process
negative regulation of nucleotide metabolic process
tricarboxylic acid cycle
glyoxylate cycle
trichome branching
S-adenosylmethionine metabolic process
carotene biosynthetic process
multicellular organismal development
anatomical structure development
programmed cell death
cellular component organization
RNA metabolic process
response to zinc ion
zinc ion transport
cellular homeostasis
protein amino acid phosphorylation
cytokinin mediated signaling pathway
transport
protein modification process
phosphate metabolic process
macromolecule localization
response to abscisic acid stimulus
embryonic development ending in seed dormancy
cellular response to phosphate starvation
autophagy
inflorescence development
negative regulation of abscisic acid mediated signaling...
tryptophan biosynthetic process
histone acetylation
long-day photoperiodism
D-xylose metabolic process
short-day photoperiodism
regulation of glucosinolate biosynthetic process
cellular response to sulfur starvation
DNA recombination
L-methionine salvage from methylthioadenosine
-10%
-5%
0%
5%
10%
15%
% over/under representation relative to reference set

Figure 4.16 Molecular functions enriched in vegetative tissue-specific probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe sets (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of probe sets. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-specific gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list. One category ends with … due to space limitations and should read as "[ribulose-bisphosphate carboxylase]-lysine N-methyl transferase activity."



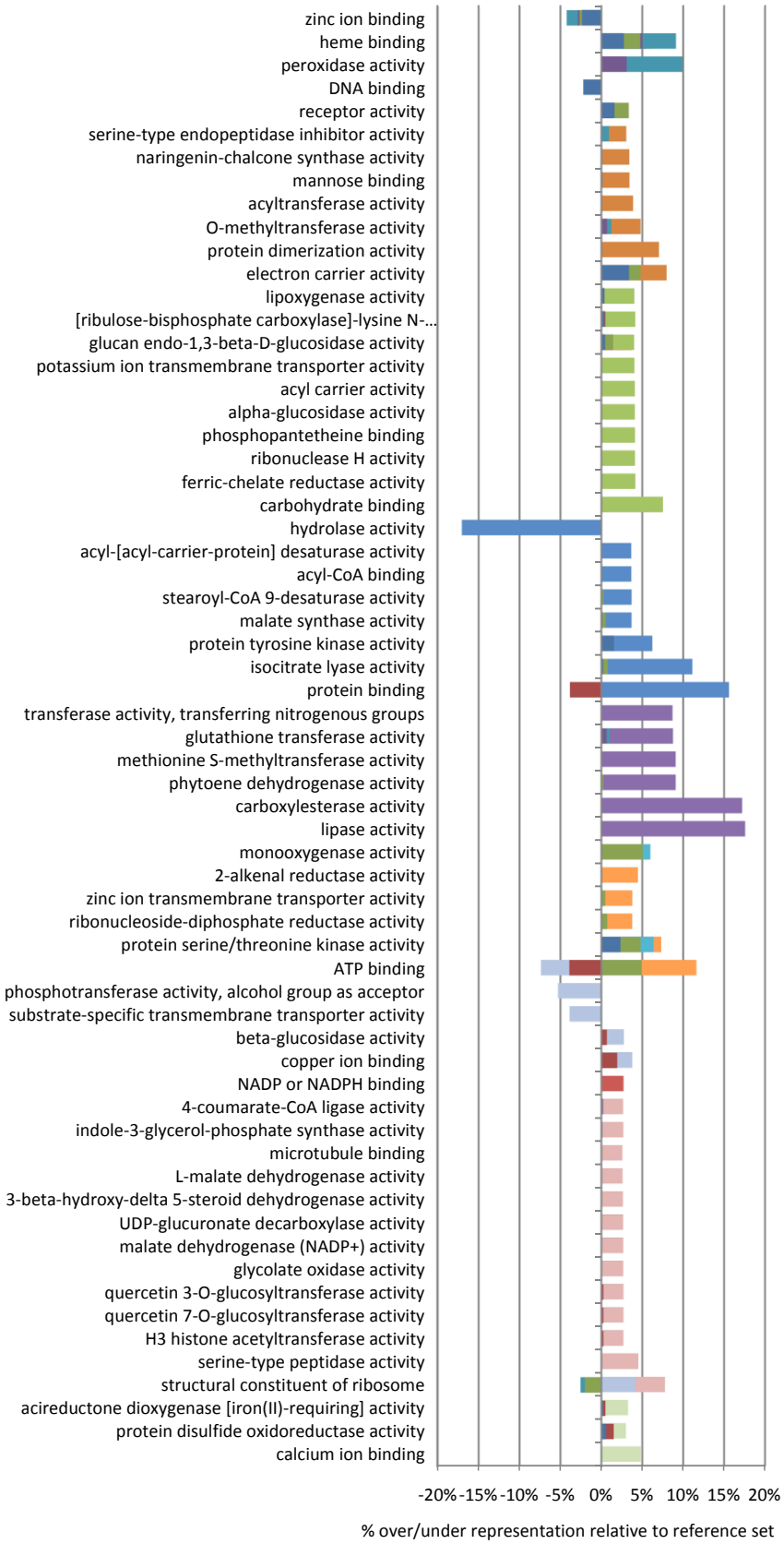
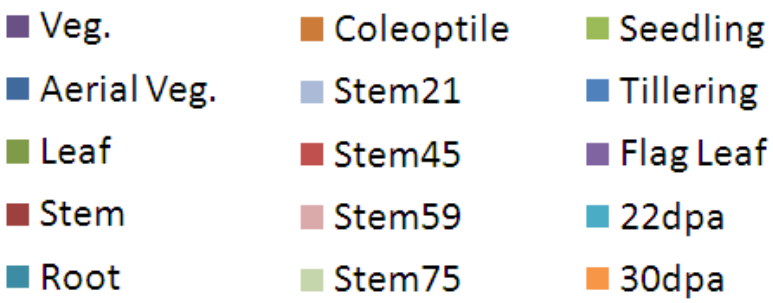
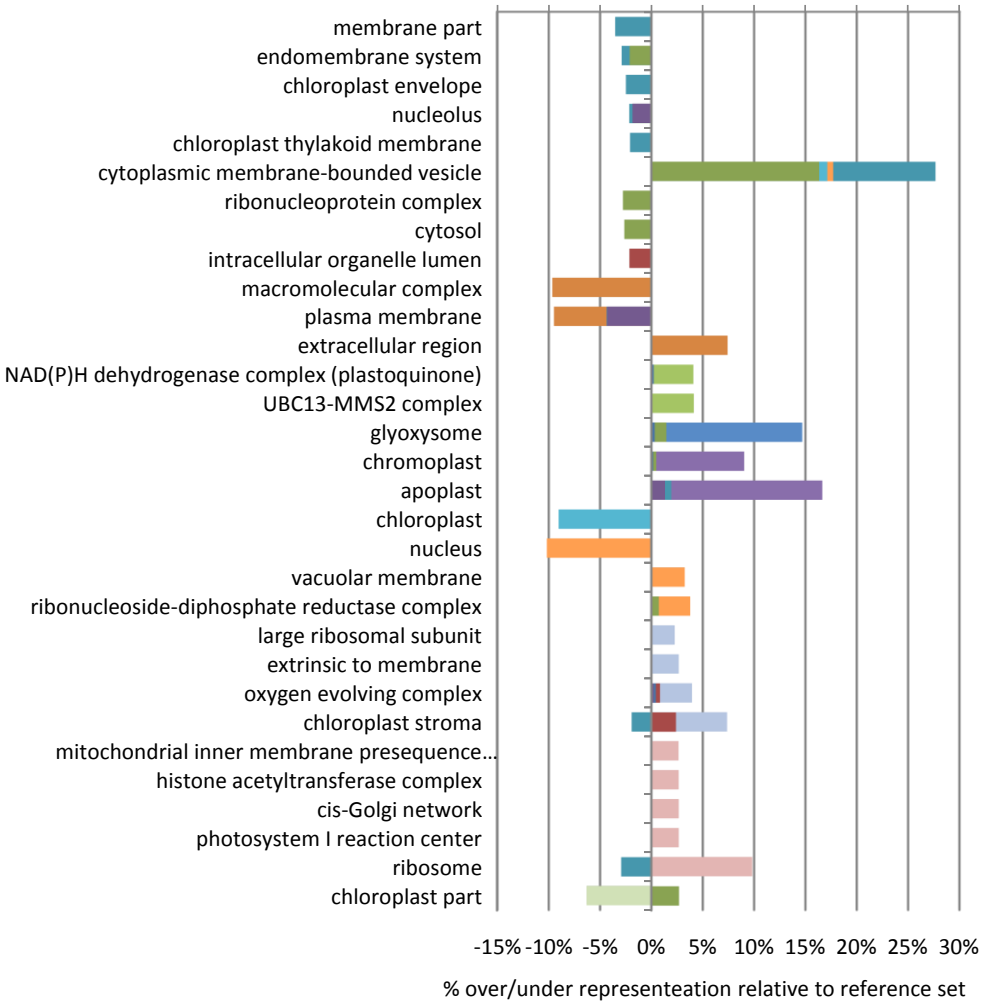

zinc ion binding
heme binding
peroxidase activity
DNA binding
receptor activity
serine-type endopeptidase inhibitor activity
naringenin-chalcone synthase activity
mannose binding
acyltransferase activity
O-methyltransferase activity
protein dimerization activity
electron carrier activity
lipoxygenase activity
[ribulose-bisphosphate carboxylase]-lysine N-...
glucan endo-1,3-beta-D-glucosidase activity
potassium ion transmembrane transporter activity
acyl carrier activity
alpha-glucosidase activity
phosphopantetheine binding
ribonuclease H activity
ferric-chelate reductase activity
carbohydrate binding
hydrolase activity
acyl-[acyl-carrier-protein] desaturase activity
acyl-CoA binding
stearoyl-CoA 9-desaturase activity
malate synthase activity
protein tyrosine kinase activity
isocitrate lyase activity
protein binding
transferase activity, transferring nitrogenous groups
glutathione transferase activity
methionine S-methyltransferase activity
phytoene dehydrogenase activity
carboxylesterase activity
lipase activity
monooxygenase activity
2-alkenal reductase activity
zinc ion transmembrane transporter activity
ribonucleoside-diphosphate reductase activity
protein serine/threonine kinase activity
ATP binding
phosphotransferase activity, alcohol group as acceptor
substrate-specific transmembrane transporter activity
beta-glucosidase activity
copper ion binding
NADP or NADPH binding
4-coumarate-CoA ligase activity
indole-3-glycerol-phosphate synthase activity
microtubule binding
L-malate dehydrogenase activity
3-beta-hydroxy-delta 5-steroid dehydrogenase activity
UDP-glucuronate decarboxylase activity
malate dehydrogenase (NADP+) activity
glycolate oxidase activity
quercetin 3-O-glucosyltransferase activity
quercetin 7-O-glucosyltransferase activity
H3 histone acetyltransferase activity
serine-type peptidase activity
structural constituent of ribosome
acireductone dioxygenase [iron(II)-requiring] activity
protein disulfide oxidoreductase activity
calcium ion binding
-20% -15% -10% -5% 0% 5% 10% 15% 20%
% over/under representation relative to reference set

Figure 4.17 Cellular components enriched in vegetative tissue-specific probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe sets (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of probe sets. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-specific gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list. One category ends with … due to space limitations and should read as "mitochondrial inner membrane presequence translocase complex."




membrane part
endomembrane system
chloroplast envelope
nucleolus
chloroplast thylakoid membrane
cytoplasmic membrane-bounded vesicle
ribonucleoprotein complex
cytosol
intracellular organelle lumen
macromolecular complex
plasma membrane
extracellular region
NAD(P)H dehydrogenase complex (plastoquinone)
UBC13-MMS2 complex
glyoxysome
chromoplast
apoplast
chloroplast
nucleus
vacuolar membrane
ribonucleoside-diphosphate reductase complex
large ribosomal subunit
extrinsic to membrane
oxygen evolving complex
chloroplast stroma
mitochondrial inner membrane presequence...
histone acetyltransferase complex
cis-Golgi network
photosystem I reaction center
ribosome
chloroplast part
-15% -10% -5% 0% 5% 10% 15% 20% 25% 30%
% over/under representeation relative to reference set

characteristics of these tissues at different developmental stages. For example, defense response to fungus is enriched in the seedling tissue while defense response to virus and insect are both enriched in the tillering tissue (Figure 4.15) indicating that defense pathways to different biotic stresses may be activated at different stages of leaf development. Also, organization and structure developmental processes including multicellular organismal development, anatomical structure development and cellular component organization are under-represented in the early and late senescent leaf tissues (22DPA and 30DPA) while programmed cell death is enriched in the early senescent tissue as would be expected. In contrast to the senescent leaf tissues, the generation of precursor metabolites and energy is enriched in the coleoptile-specific gene list (Figure 4.15).

Several molecular function GO terms were found to be enriched in the vegetative tissue-specific gene lists (Figure 4.16). While several molecular functions were enriched in a particular vegetative tissue stage, heme binding and peroxidase activity were found to be enriched in aerial vegetative tissues and root tissues indicating a general enrichment in the vegetative tissue type. Protein binding is highly enriched at the tillering stage (15.59%) and carboxylesterase activity and lipase activity are highly enriched (17.20% and 17.57% respectively) in the flag leaf tissue indicating that these are highly important functions for later leaf development. A few functions were found to be under-represented in the vegetative tissue-specific probe set lists in comparison to the reference set including hydrolase activity which was under-represented by 17.04% in the tillering-specific probe set list. This under-representation of hydrolase activity can therefore be considered a characteristic of leaf tissue at the tillering stage of development. A number

of hydrolases are active during leaf senescence (He and Gan, 2002; Schelbert et al., 2009). Leaf tissue at the tillering stage is still undergoing active growth and development so this activity may be inhibited at this stage resulting in its under-representation.

The enrichment of cellular component terms in the annotations of vegetative tissue-specific gene lists also reveals characteristics of these tissues (Figure 4.17). Glyoxysome is enriched in tillering leaf tissue (almost 15 %, Figure 4.17), indicating that the glyoxylate pathway may be more active at this stage of development. This is also consistent with the enrichment of the biological process glyoxylate cycle (almost 15 % Figure 4.15). The chromoplast and apoplast are enriched in the flag leaf tissue indicative of the later stages of leaf development in which pigments other than chlorophyll, and vascular tissue for transporting photosynthetic products, are important for plant function (Gregersen et al., 2008). The cellular component terms chloroplast and nucleus are underrepresented in the early and late senescent leaf tissue-specific gene lists respectively, indicative of the reduction of photosynthetic activity and cellular degeneration at these stages of leaf development (Lee et al., 2004; Gregersen and Holm, 2007; Lim et al., 2007).

4.3.6 Triticale tissue-enriched transcripts

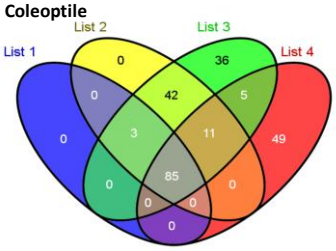
To identify tissue-enriched transcripts, the probe sets were filtered for 2-fold higher expression between the tissue of interest and all other tissues following ANOVA, FDR and the conversion of RMA expression values to zero in the cases of absent expression. Of the 61,115 probe sets on the wheat array, 41,918 were determined to be present (majority P consensus detection call) in at least one tissue in the study (Table

4.5). Following the ANOVA and FDR it was found that 58,891 probe sets have expression that significantly differs across the 29 tissue samples (adjusted p-value ≤ 0.05) and 41,532 of these also have a majority P consensus detection call in at least one tissue. These 41,532 probe sets were therefore considered for further evaluation for tissue-enriched expression.

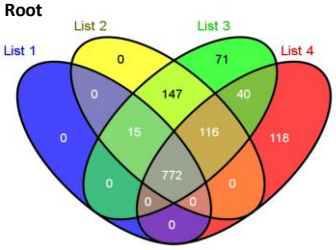
Four different strategies for the identification of tissue-enriched probe sets were evaluated and Figure 4.18 shows the numbers of probe sets identified as tissues enriched using each of these strategies for each tissue as well as the overlap among the four strategies. Probe sets that were identified using only strategies 1 or 2 were found not to have a majority present consensus detection call in the tissue of interest and were therefore discounted as likely false positives. Evaluation of probe sets identified by only strategy 4 also revealed that these are likely false positives due to the majority present detection calls in tissues other than the tissue of interest and the less than 2 fold difference in mean RMA values compared to the tissue of interest. These probe sets were therefore discounted as likely false positives and removed from further analysis. While the probe sets identified by strategy 3 (RMA values converted to zero when consensus detection call is non majority P, all probe sets encompassed by green ovals in Figure 4.18) are considered those representing tissue-enriched gene expression, the probe sets identified by strategies 1, 2 and 4 as are also presented to show the overlap in the strategies. It should also be noted that probe sets that were identified as tissue-specific (section 3.5) were encompassed within the probe sets identified as tissue-enriched by strategy 3 and were not found within the tissue-enriched probe sets identified by only strategies 1, 2 or 4. It was expected that tissue-specific probe sets would also be

Figure 4.18 Venn diagrams showing the overlap of probe set lists identified as displaying tissue-enriched expression using four different analysis methods (List 1 to 4). List 1 represents probe sets identified as tissue-enriched using pairwise comparisons of mean RMA expression values. List 2 used mean RMA expression values for pairwise comparisons following the conversion of these values to 0 when the majority of replicate detection calls were A (>50% absent). List 3 used mean RMA expression values for pairwise comparisons following the conversion of these values to 0 when the majority of replicate detection calls were not P (<50% present). List 4 used mean RMA values for pairwise comparisons following the conversion of these values to 0 when the absolute consensus among replicate detection calls was not P (<100% present). With an increase in overlap among the four lists there is an increase confidence that the probe sets identified truly represent tissue-enriched expressed genes.

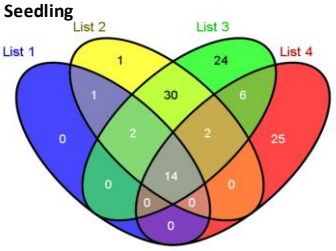
Coleoptile



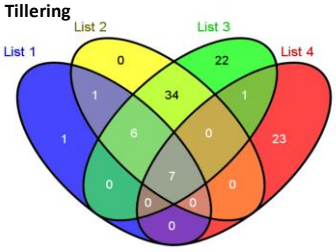
Root



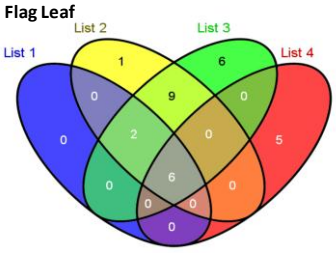
Seedling



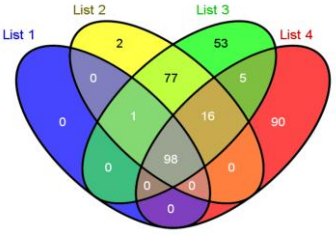
Tillering



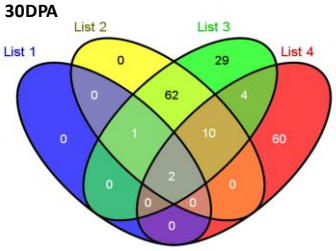
Flag Leaf



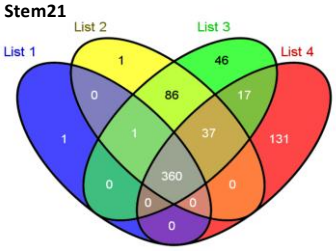
22DPA



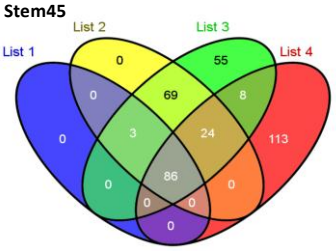
30DPA



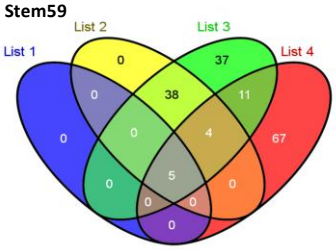
Stem21



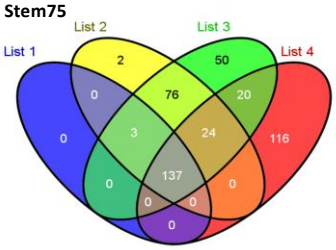
Stem45



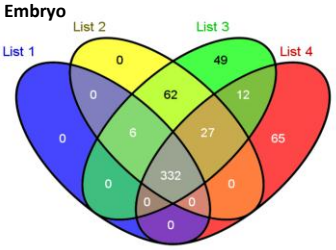
Stem59



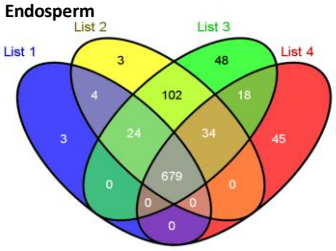
Stem75



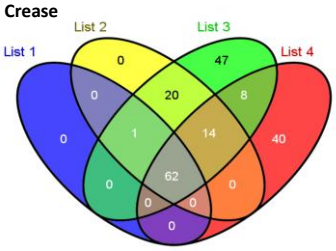
Embryo



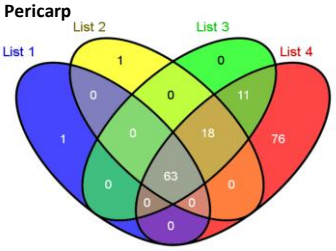
Endosperm



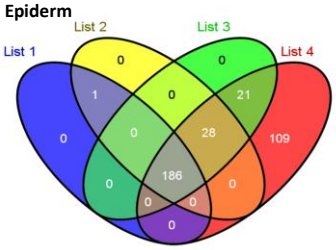
Crease



Pericarp



Epiderm

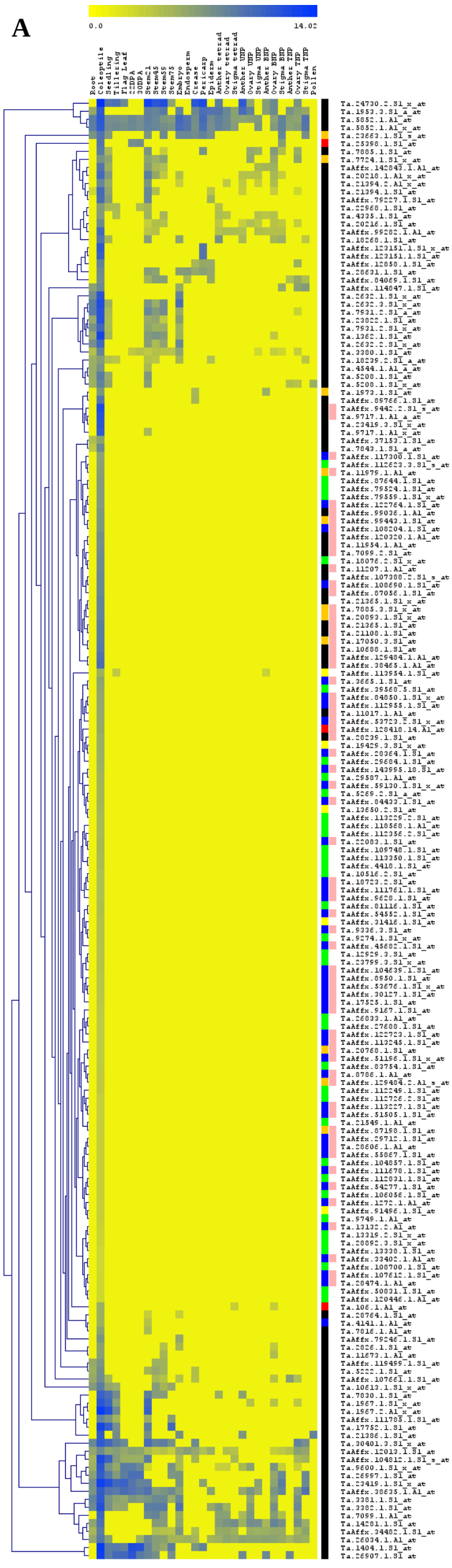


identified as tissue-enriched. This was the case for most tissue-specific probe sets with the exception of a few of tissue-specific probe sets that did not have significantly different expression across the 29 tissue sample types (adjusted p-value ≤ 0.05 filter), and were therefore not included in the tissue-enriched set. The remainder of the tissue-specific probe sets were identified as tissue-enriched using strategy 3. This provides further evidence that strategy 3 identified the true tissue-enriched probe sets.

While probe sets that were only identified by strategies 1, 2 or 4 were removed as false positives, it was also considered that the greater the overlap among the strategies, the higher the confidence in the probe sets identified as representing true tissue-enriched gene expression. Therefore, the probe sets identified by all four strategies (grey, center of Venn diagrams, Figure 4.18) represent the probe sets believed to be representative of tissue-enriched expression with the highest level of confidence.

The lists of tissue-enriched probe sets were used for hierarchical clustering of their expression patterns to identify similar patterns of expression among these lists. Mean RMA values following the conversion of non majority P consensus detection calls to zero were used for clustering by HCL. The trend in clustering results was similar for all tissue-enriched lists and a couple of the clustered expression heat maps are provided as representative examples in Figure 4.19. The trees resulting from the HCL showed clusters at the top and bottom of the tree of probe sets with varied expression across the tissues studied, while there was a cluster in the centre of each tree of tissue-specific probe sets. To the right of the heat maps in Figure 4.19 is a colored column showing the overlap in the four strategies evaluated for the identification of tissue-enriched probe sets to show the relative level of confidence that the probe set is truly tissue-enriched. To the

Figure 4.19 Expression heat maps of coleoptile (A) and crease-enriched (B) probe sets. Probe set lists were clustered using HCL in the MeV package of the TM4 software suite. The tree to the left of the heat map shows the relationships of the probe sets following clustering. The first vertical colored bar to the right of the heat map represents the lists from the Venn diagrams in Figure 4.18 that the probe sets belong to (green: list 3 only; blue: list 2 and 3; red: lists 1, 2 and 3; yellow: lists 3 and 4; orange: lists 2, 3, and 4; black: lists 1, 2, 3 and 4). The second vertical colored bar to the right of the heat map (pink) identifies probe sets that were also identified as tissue-specific (using the majority present detection calls in the tissue of interest and majority absent detection calls in all other tissues). To the right of the vertical colored bars are the IDs of the probe sets represented in the heat map. Mean RMA expression values following the conversion of these values to 0 when the majority of detection calls were not P (< 50% present) were used to construct the heat map. Each row represents the probe set expression values across the tissues studied and each column represents one of the tissue types studied. The horizontal color scale at the top of the heat map represents the expression values with yellow representing no expression and dark blue representing maximum expression. The mean expression values for each probe set in a particular tissue are represented by the intensity of color.


A
0.0
14.02
Root
Coleoptile
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Stem21
Stem45
Stem55
Stem75
Embryo
Endosperm
Crease
Pericarp
Epiderm
Anther tetrad
Ovary tetrad
Stigma tetrad
Anther UMP
Ovary UMP
Stigma UMP
Anther BMP
Ovary BMP
Stigma BMP
Anther TMP
Ovary TMP
Stigma TMP
Pollen
Ta.24730.2.S1_x_at
Ta.1953.2.S1_a_at
Ta.5852.1.A1_at
Ta.5852.1.A1_x_at
Ta.23663.1.S1_s_at
Ta.25398.1.S1_at
Ta.7885.1.S1_at
Ta.7724.1.S1_x_at
TaAffx.142843.1.A1_at
Ta.20218.1.A1_x_at
Ta.21394.2.A1_x_at
Ta.21394.1.S1_at
TaAffx.79227.1.S1_at
Ta.22568.1.S1_at
Ta.4238.1.S1_at
Ta.20216.1.S1_at
TaAffx.99282.1.A1_at
Ta.18268.1.S1_at
TaAffx.123151.1.S1_x_at
TaAffx.123151.1.S1_at
TaAffx.12858.1.S1_at
Ta.28691.1.S1_at
TaAffx.84069.1.S1_at
TaAffx.114847.1.S1_at
Ta.2632.1.S1_x_at
Ta.2632.3.S1_x_at
Ta.7931.2.S1_a_at
Ta.23822.1.S1_at
Ta.7931.2.S1_x_at
Ta.1362.1.S1_at
Ta.2632.2.S1_x_at
Ta.3380.1.S1_at
Ta.18239.2.S1_a_at
Ta.4544.1.A1_a_at
Ta.5208.1.S1_at
Ta.5208.1.S1_x_at
Ta.1973.1.S1_at
TaAffx.89766.1.S1_at
TaAffx.9442.2.S1_s_at
Ta.9717.1.A1_a_at
Ta.23419.3.S1_x_at
Ta.9717.1.A1_x_at
TaAffx.97159.1.S1_at
Ta.7842.1.S1_a_at
TaAffx.117300.1.S1_at
TaAffx.112623.3.S1_s_at
Ta.11579.1.A1_at
TaAffx.87644.1.S1_at
TaAffx.79524.1.S1_at
TaAffx.79559.1.S1_x_at
TaAffx.122764.1.S1_at
TaAffx.99036.1.A1_at
TaAffx.99443.1.S1_at
TaAffx.108204.1.S1_at
TaAffx.120320.1.A1_at
Ta.11554.1.A1_at
Ta.7059.2.S1_at
Ta.18076.2.S1_x_at
Ta.11207.1.A1_at
TaAffx.107388.2.S1_s_at
TaAffx.108690.1.S1_at
TaAffx.87056.1.S1_at
Ta.21365.1.S1_x_at
Ta.7885.3.S1_x_at
Ta.20893.1.S1_x_at
Ta.21365.1.S1_at
Ta.21108.1.S1_at
Ta.17050.3.S1_at
Ta.10688.1.S1_at
TaAffx.129484.1.A1_at
TaAffx.38465.1.A1_at
TaAffx.113954.1.S1_at
Ta.3665.1.S1_at
TaAffx.39568.5.S1_at
TaAffx.84850.1.S1_x_at
TaAffx.112955.1.S1_at
Ta.11017.1.A1_at
TaAffx.53723.2.S1_x_at
TaAffx.128418.14.A1_at
Ta.28239.1.S1_at
Ta.19429.3.S1_x_at
TaAffx.28364.1.S1_at
TaAffx.29694.1.S1_at
TaAffx.143595.18.S1_at
Ta.29587.1.A1_at
TaAffx.59130.1.S1_x_at
Ta.5269.2.S1_a_at
TaAffx.84433.1.S1_at
Ta.13650.2.S1_at
TaAffx.113229.2.S1_at
TaAffx.118568.1.A1_at
TaAffx.112356.2.S1_at
Ta.22083.1.S1_at
TaAffx.109748.1.S1_at
TaAffx.113350.1.S1_at
TaAffx.4418.1.S1_at
Ta.10516.2.S1_at
Ta.18729.2.S1_at
TaAffx.111761.1.S1_at
TaAffx.9628.1.S1_at
TaAffx.81116.1.S1_at
TaAffx.54552.1.S1_at
TaAffx.31416.1.S1_at
Ta.9336.3.S1_at
Ta.9274.1.S1_x_at
TaAffx.45682.1.S1_at
Ta.12929.3.S1_at
Ta.23799.3.S1_x_at
TaAffx.104639.1.S1_at
TaAffx.8950.1.S1_at
TaAffx.53676.1.S1_x_at
TaAffx.30127.1.S1_at
Ta.17525.1.S1_at
TaAffx.9167.1.S1_at
Ta.26833.1.A1_at
TaAffx.27688.1.S1_at
TaAffx.152723.1.S1_at
TaAffx.112245.1.S1_at
Ta.20768.1.S1_at
TaAffx.51196.1.S1_x_at
TaAffx.83754.1.S1_at
Ta.8786.1.A1_at
TaAffx.129484.2.A1_s_at
TaAffx.112249.1.S1_at
TaAffx.112726.2.S1_at
TaAffx.113227.1.S1_at
TaAffx.51505.1.S1_at
Ta.21549.1.A1_at
TaAffx.87158.1.S1_at
TaAffx.29712.1.S1_at
Ta.28606.1.A1_at
TaAffx.55867.1.S1_at
TaAffx.104857.1.S1_at
TaAffx.111678.1.S1_at
TaAffx.112831.1.S1_at
TaAffx.54277.1.S1_at
TaAffx.106056.1.S1_at
TaAffx.1272.1.A1_at
TaAffx.91496.1.S1_at
Ta.9749.1.A1_at
Ta.13132.2.A1_at
Ta.13319.2.S1_x_at
Ta.28892.3.S1_x_at
TaAffx.13238.1.S1_at
TaAffx.33402.1.A1_at
TaAffx.108700.1.S1_at
TaAffx.107612.1.S1_at
Ta.28474.1.A1_at
TaAffx.50831.1.S1_at
TaAffx.120446.1.A1_at
Ta.106.1.A1_at
Ta.28764.1.S1_at
Ta.4141.1.A1_at
Ta.7816.1.A1_at
TaAffx.79246.1.S1_at
Ta.2826.1.S1_at
Ta.11673.1.A1_at
TaAffx.119499.1.S1_at
Ta.5222.1.S1_at
TaAffx.107661.1.S1_at
Ta.10613.1.S1_x_at
Ta.7830.1.S1_at
Ta.1967.1.S1_x_at
Ta.1967.2.A1_x_at
TaAffx.111785.1.S1_at
Ta.17752.1.S1_at
Ta.21386.1.S1_at
Ta.30401.3.S1_x_at
TaAffx.12019.1.S1_at
TaAffx.104812.1.S1_s_at
Ta.9600.1.S1_x_at
Ta.26997.1.S1_at
Ta.23419.1.S1_x_at
TaAffx.38635.1.A1_at
Ta.3381.1.S1_at
Ta.3382.1.S1_at
Ta.7099.1.A1_at
Ta.14281.1.S1_at
TaAffx.34482.1.S1_at
Ta.26034.1.A1_at
Ta.1404.1.S1_at
Ta.26907.1.S1_at
B
0.0
13.83
Root
Coleoptile
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Stem21
Stem45
Stem55
Stem75
Embryo
Endosperm
Crease
Pericarp
Epiderm
Anther tetrad
Ovary tetrad
Stigma tetrad
Anther UMP
Ovary UMP
Stigma UMP
Anther BMP
Ovary BMP
Stigma BMP
Anther TMP
Ovary TMP
Stigma TMP
Pollen
Ta.5650.1.S1_a_at
Ta.5650.2.S1_x_at
Ta.11599.2.S1_a_at
Ta.5650.3.S1_x_at
Ta.11265.1.S1_at
Ta.7567.1.A1_at
Ta.9031.1.S1_at
Ta.25919.1.S1_at
Ta.9031.1.S1_s_at
TaAffx.84130.1.S1_at
Ta.7495.3.S1_at
TaAffx.8139.1.S1_s_at
Ta.8792.2.A1_a_at
Ta.9166.1.S1_at
Ta.20524.1.A1_at
Ta.20524.1.A1_a_at
Ta.7899.1.A1_at
Ta.5335.1.A1_x_at
Ta.5335.1.A1_at
Ta.3234.1.S1_at
Ta.9973.1.S1_at
Ta.9060.1.S1_at
Ta.22430.1.S1_at
TaAffx.76007.1.S1_at
TaAffx.24695.1.S1_at
Ta.10087.1.A1_at
Ta.5664.1.A1_at
TaAffx.25818.1.S1_at
Ta.9958.1.S1_at
TaAffx.52557.1.S1_at
Ta.26902.1.S1_at
Ta.20102.1.S1_x_at
TaAffx.70168.1.A1_at
Ta.14033.2.S1_at
TaAffx.25090.1.S1_at
Ta.20102.2.S1_at
TaAffx.92973.1.A1_at
Ta.30028.1.S1_s_at
TaAffx.52008.1.S1_at
Ta.21049.1.S1_at
Ta.5578.1.S1_x_at
TaAffx.96414.1.S1_s_at
TaAffx.25418.1.S1_at
Ta.10885.1.A1_at
Ta.20064.1.S1_at
TaAffx.25303.1.S1_at
Ta.23673.1.S1_a_at
TaAffx.11110.1.S1_at
Ta.20362.1.S1_s_at
Ta.24105.1.A1_s_at
TaAffx.25533.1.S1_at
TaAffx.85380.1.S1_at
TaAffx.107532.1.S1_at
Ta.13160.2.S1_x_at
TaAffx.4918.1.S1_at
TaAffx.94689.1.S1_at
Ta.18662.1.S1_x_at
TaAffx.87028.1.S1_at
TaAffx.24323.1.S1_at
Ta.1916.1.A1_at
TaAffx.84978.1.S1_at
TaAffx.116069.4.S1_at
TaAffx.58105.1.S1_at
Ta.4399.1.S1_x_at
TaAffx.80040.1.S1_at
Ta.6258.1.S1_at
TaAffx.107116.1.S1_at
Ta.1229.3.S1_x_at
Ta.22111.2.S1_at
TaAffx.16182.1.S1_at
TaAffx.97673.2.A1_at
Ta.4808.3.S1_at
TaAffx.84175.1.S1_at
TaAffx.106718.1.S1_at
TaAffx.4998.1.S1_at
TaAffx.52949.1.S1_at
Ta.6043.3.S1_x_at
TaAffx.91748.1.S1_at
TaAffx.57748.1.S1_at
TaAffx.87595.1.S1_at
TaAffx.79432.1.S1_at
Ta.177.1.S1_at
TaAffx.113462.1.S1_at
Ta.1602.1.S1_at
Ta.20047.1.S1_at
Ta.22393.1.S1_at
TaAffx.8456.1.S1_at
TaAffx.30466.1.S1_at
TaAffx.85116.1.S1_at
TaAffx.112194.1.S1_at
TaAffx.105134.1.S1_at
TaAffx.56217.1.S1_at
TaAffx.108513.1.S1_at
TaAffx.112294.1.S1_at
TaAffx.121152.1.S1_at
TaAffx.111824.1.S1_at
TaAffx.24118.2.S1_at
Ta.24730.3.S1_at
Ta.22650.2.S1_at
Ta.16323.2.S1_x_at
Ta.3413.1.S1_at
TaAffx.113156.1.S1_x_at
TaAffx.83154.1.S1_at
TaAffx.27105.1.S1_at
TaAffx.26793.1.S1_at
TaAffx.90192.1.S1_at
TaAffx.111023.1.S1_at
TaAffx.31309.1.S1_at
Ta.6231.2.A1_at
TaAffx.111341.1.S1_at
TaAffx.80038.1.S1_at
TaAffx.79804.1.S1_at
Ta.20304.2.S1_at
TaAffx.58220.1.S1_at
Ta.10141.1.A1_at
TaAffx.52046.1.S1_at
Ta.19017.1.S1_at
TaAffx.90017.1.S1_at
TaAffx.51584.1.S1_at
TaAffx.6518.1.S1_at
TaAffx.113565.1.S1_at
TaAffx.24663.1.S1_at
TaAffx.80223.1.S1_at
TaAffx.111167.1.S1_at
Ta.6185.1.S1_at
Ta.11450.1.A1_at
Ta.28545.1.S1_at
Ta.19449.1.S1_at
Ta.10565.1.A1_at
TaAffx.80461.1.S1_at
TaAffx.106051.1.S1_at
TaAffx.52872.1.S1_at
TaAffx.4629.1.S1_at
Ta.160.2.A1_at
TaAffx.7064.2.S1_at
TaAffx.37805.2.S1_at
Ta.24043.1.A1_at
TaAffx.51625.1.S1_at
TaAffx.6223.1.S1_at
Ta.5578.1.S1_a_at
Ta.5650.3.S1_at
Ta.10732.1.A1_at
TaAffx.52445.1.S1_at
Ta.10268.1.S1_at
Ta.14733.2.S1_a_at
Ta.14733.2.S1_x_at
TaAffx.110589.3.S1_s_at
TaAffx.79574.1.S1_at
Ta.4450.1.S1_at
TaAffx.81871.1.S1_at
TaAffx.32294.1.S1_at
TaAffx.70160.1.A1_at

right of the multicolored column is another column (pink) that identifies the probe sets that were also identified as tissue-specific (majority P consensus detection call) in the tissue-specific analysis. Clustering of the tissue-enriched probe sets identified using the tissue-enriched analysis revealed an increase in the number of putative tissue-specific probe sets identified as compared to the tissue-specific analysis.

The results from the tissue-enriched expression analysis are also found in Supplementary Table S4.7. This supplementary table has the results for all the criteria used for tissue-enriched expression identification including which probe sets were identified as tissue-enriched using each of the four strategies evaluated for each seed and vegetative tissue in the study. While only the probe sets that were identified using the third strategy (list 3 in Table S4.7) and have an adjusted p-value less than or equal to 0.05 and a majority P consensus detection call in the tissue of interest are considered tissue-enriched, the results from the other strategies are provided for comparison and to show the overlap among the strategies. Probe sets identified as tissue-specific by the tissue-specific expression analysis above are also listed in this table for each of the vegetative and seed tissues for comparison. This table also has the MAS5.0 consensus detection calls, RMA normalized expression values, the results of the ANOVA and FDR, as well as the mean RMA expression values following conversion of absent expression values to zero when the majority consensus detection call is not present (>50% P) for reference. Supplementary Table S4.7 can be filtered according to the methods used to identify probe sets with enriched expression in all triticale vegetative and seed tissues studied to allow independent researchers to determine if genes of their own interest are also tissue-enriched in triticale.

Functional enrichment analysis was performed on probe set lists identified as tissue-enriched in vegetative and seed tissues. Functional annotations for these lists were compared to the functional annotations of the reference list of expressed probe sets (41,918 probe sets) using Fisher's exact test in the Blast2Go software (Conesa and Götz, 2008). This analysis identified biological process, molecular function and cellular component GO annotation terms that are enriched in the genes up-regulated in a particular seed or vegetative tissue and may provide further insight into tissue differentiation and development along with the enrichment analysis performed for tissue-specific probe sets. Enrichment results were filtered by a p-value < 0.05, the most specific terms in each branch of the ontology and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set annotations.

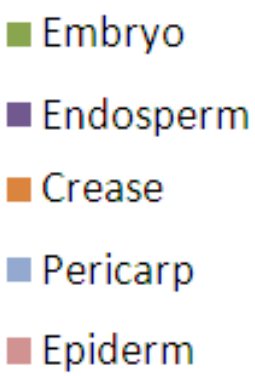
4.3.6.1 Functional enrichment in seed-enriched transcripts

The annotations from the probe sets that were identified as seed tissue-enriched were analyzed for functional enrichment of biological process (Figure 4.20), molecular function (Figure 4.21) and cellular component (Figure 4.22) terms. One observation was that despite the fact that a larger number of probe sets were identified as seed tissue-enriched, compared to seed tissue-specific, there were fewer functional terms found to be enriched in the annotations of seed tissue up-regulated probe sets. This may be due to the fact that tissue-enriched probe sets are also expressed in other tissues and therefore there are fewer terms with a difference in representation between tissue-enriched probe set lists and the reference set. As predicted, some of the functional terms that were over represented in the tissue-specific probe set lists were also over-represented in the tissue-

enriched probe set lists and occasionally at a higher percentage. For example, lipid transport was over-represented in epiderm and seed-specific lists at just over 2% (Figure 4.12) but was over represented in the pericarp-enriched list at 11.37% (Figure 4.20). Other functional terms had little difference in the level of representation between the tissue-specific and tissue-enriched lists such as glycogen biosynthetic process which was over-represented by 2.56% and 2.86% respectively (Figures 4.12 and 4.20). The functional analysis of the seed-enriched lists revealed a significant under representation of a few terms relating to transcriptional activity in the seed coat tissues (gene expression, nucleic acid metabolic process, regulation of cellular process) which was not apparent in the functional analysis of tissue-specific lists. This under-representation suggests that at the stage of seed tissue sampling, the seed coat tissues may no longer be very transcriptionally active. This is an observation that could not have been made solely on the functional analysis of tissue-specific lists and supports the importance of examining the tissue-enriched genes as well as tissue-specific genes in understanding tissue development and differentiation.

Among the molecular functions that were found to be enriched in seed up-regulated probe sets were also some terms that were similar to those found to be enriched in seed tissue-specific lists. Starch synthase activity enrichment was found to be slightly lower in endosperm-enriched transcripts (2.01%, Figure 4.21) compared to endosperm-specific transcripts (2.60%, Figure 4.13). While nutrient reservoir activity was found to be enriched in the endosperm-specific transcripts by only 3.77% but enriched in seed-specific transcripts by 12.43% (Figure 4.13), the functional analysis of endosperm up-regulated probe sets show that this enrichment is largely due to the endosperm as the term

Figure 4.20 Biological processes enriched in seed tissue-enriched probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-enriched gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list.



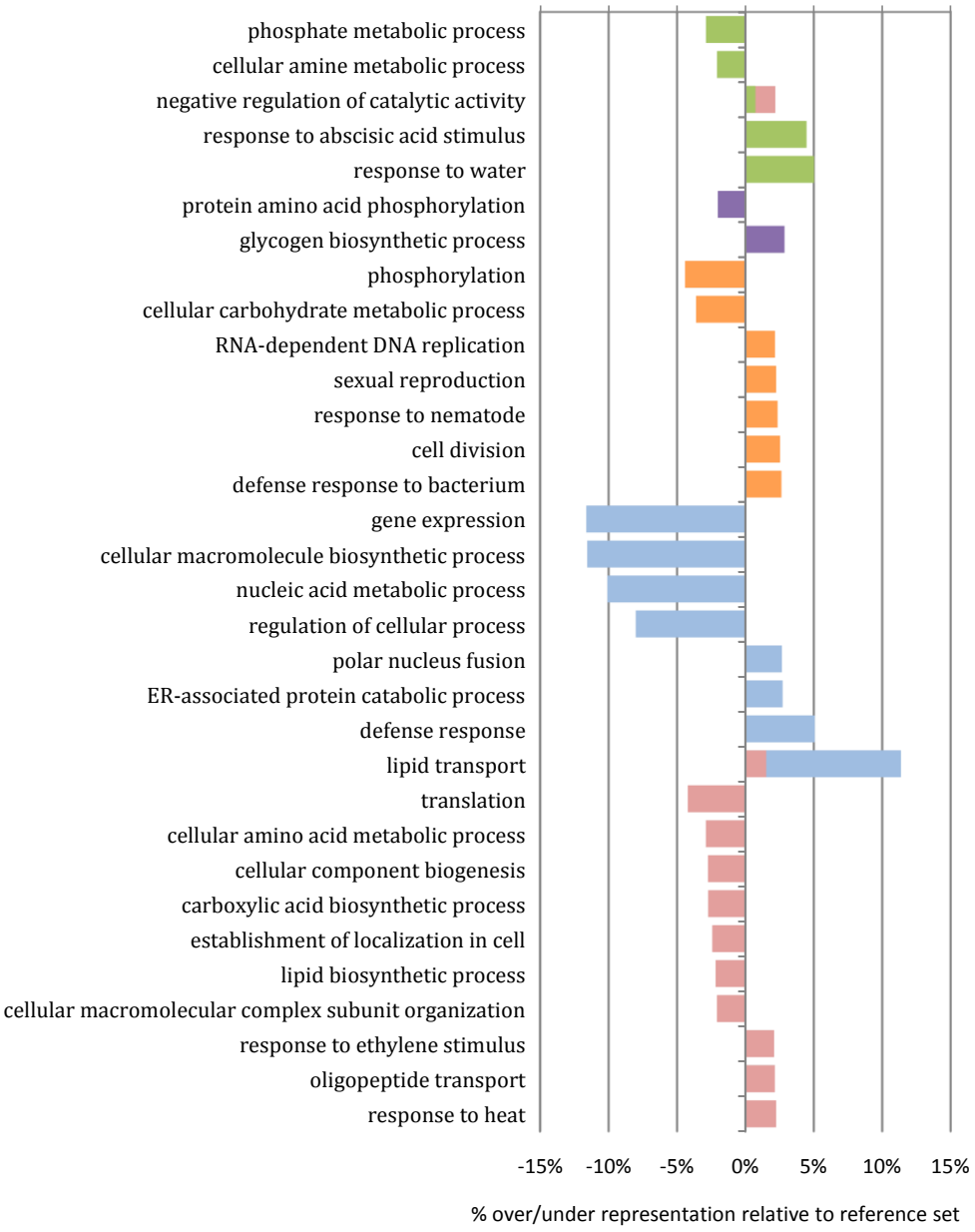

phosphate metabolic process
cellular amine metabolic process
negative regulation of catalytic activity
response to abscisic acid stimulus
response to water
protein amino acid phosphorylation
glycogen biosynthetic process
phosphorylation
cellular carbohydrate metabolic process
RNA-dependent DNA replication
sexual reproduction
response to nematode
cell division
defense response to bacterium
gene expression
cellular macromolecule biosynthetic process
nucleic acid metabolic process
regulation of cellular process
polar nucleus fusion
ER-associated protein catabolic process
defense response
lipid transport
translation
cellular amino acid metabolic process
cellular component biogenesis
carboxylic acid biosynthetic process
establishment of localization in cell
lipid biosynthetic process
cellular macromolecular complex subunit organization
response to ethylene stimulus
oligopeptide transport
response to heat
-15%
-10%
-5%
0%
5%
10%
15%
% over/under representation relative to reference set

Figure 4.21 Molecular functions enriched in seed tissue-enriched probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-enriched gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list.

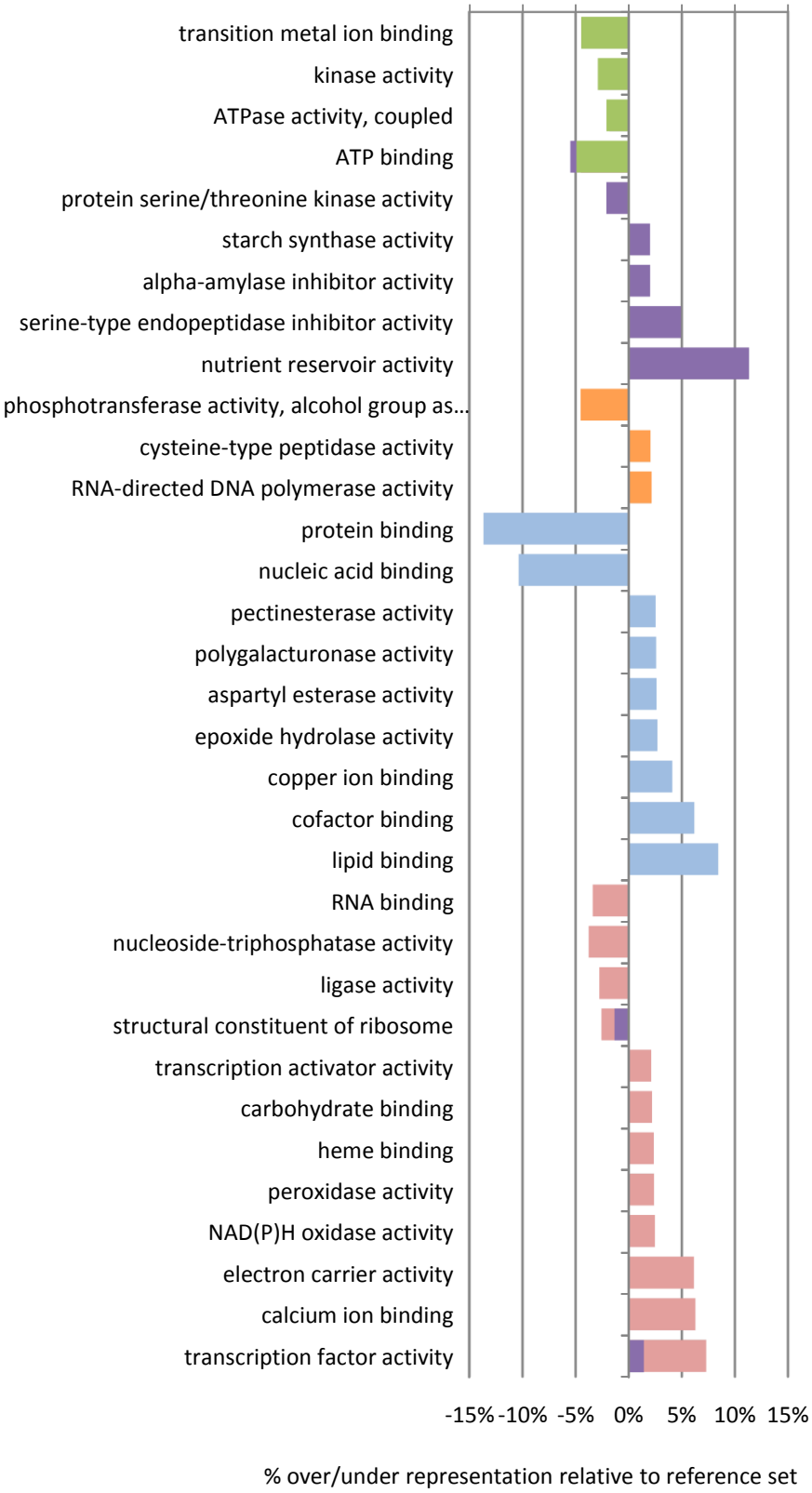
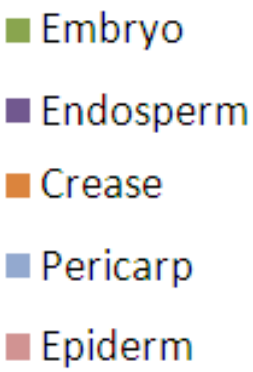
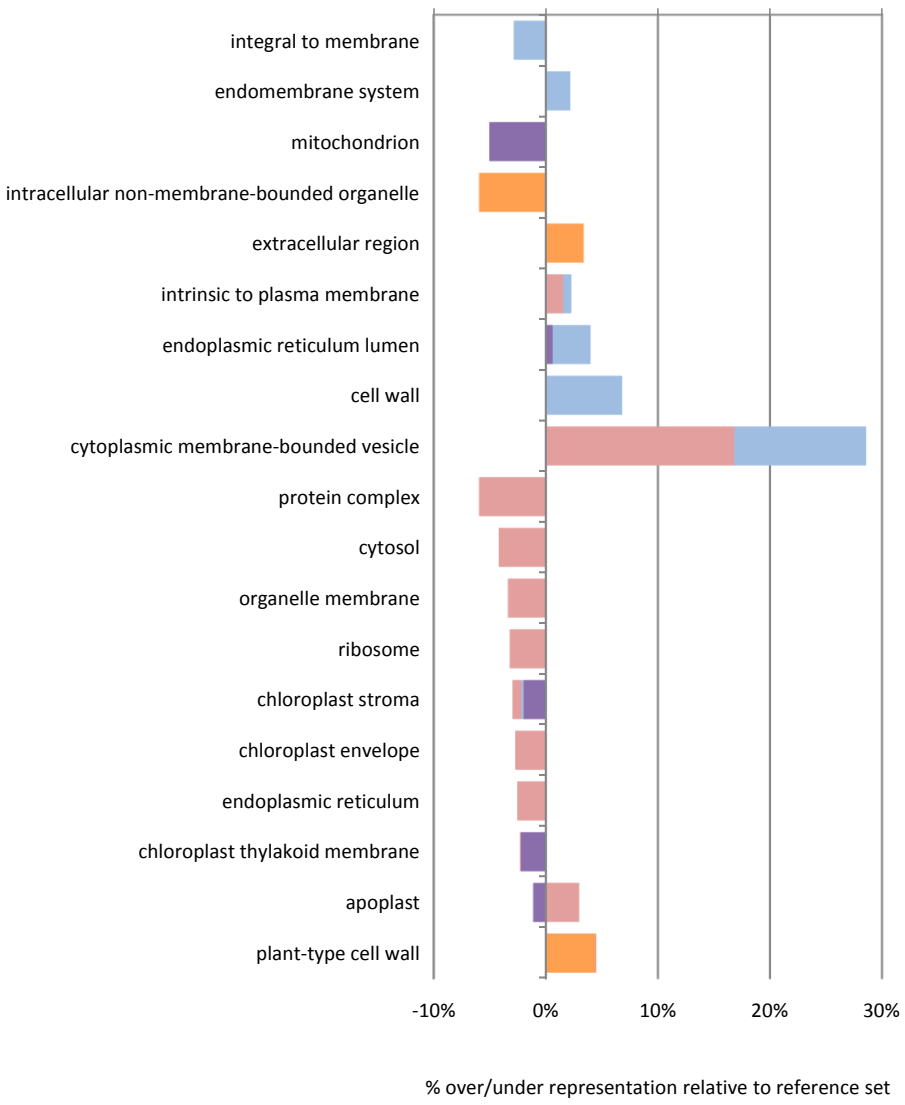

transition metal ion binding
kinase activity
ATPase activity, coupled
ATP binding
protein serine/threonine kinase activity
starch synthase activity
alpha-amylase inhibitor activity
serine-type endopeptidase inhibitor activity
nutrient reservoir activity
phosphotransferase activity, alcohol group as...
cysteine-type peptidase activity
RNA-directed DNA polymerase activity
protein binding
nucleic acid binding
pectinesterase activity
polygalacturonase activity
aspartyl esterase activity
epoxide hydrolase activity
copper ion binding
cofactor binding
lipid binding
RNA binding
nucleoside-triphosphatase activity
ligase activity
structural constituent of ribosome
transcription activator activity
carbohydrate binding
heme binding
peroxidase activity
NAD(P)H oxidase activity
electron carrier activity
calcium ion binding
transcription factor activity
-15% -10% -5% 0% 5% 10% 15%
% over/under representation relative to reference set

Figure 4.22 Cellular components enriched in seed tissue-enriched probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-enriched gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list.




integral to membrane
endomembrane system
mitochondrion
intracellular non-membrane-bounded organelle
extracellular region
intrinsic to plasma membrane
endoplasmic reticulum lumen
cell wall
cytoplasmic membrane-bounded vesicle
protein complex
cytosol
organelle membrane
ribosome
chloroplast stroma
chloroplast envelope
endoplasmic reticulum
chloroplast thylakoid membrane
apoplast
plant-type cell wall
-10%
0%
10%
20%
30%
% over/under representation relative to reference set

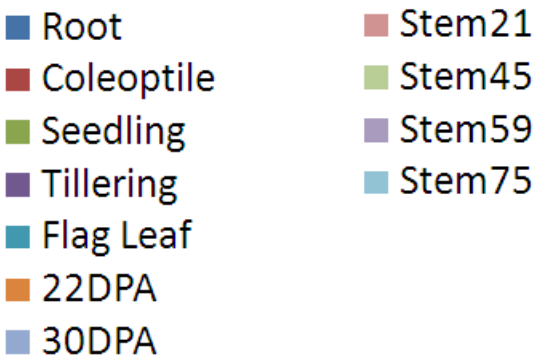
is enriched by 11.32% in endosperm-enriched probe sets (Figure 4.21). Both electron carrier activity and transcription factor activity were also highly enriched in the epiderm-enriched probe set list but at lower levels than in the epiderm-specific probe set list. A lack of transcriptional activity in the seed coat is also supported by the significant under-representation of the molecular function terms protein binding and nucleic acid binding in the pericarp-enriched probe set list (Figure 4.21).

The cellular component term enrichment analysis of seed-enriched probe sets showed an under-representation of chloroplast related terms similar to the tissue-specific probe sets (Figure 4.22). One of the most striking differences between the cellular component functional analyses of seed-enriched probe sets compared to seed-specific probe sets was the difference in the representation of the term chloroplast thylakoid membrane. This term was over-represented in crease-specific probe sets by over 10% but was not over-represented in crease-enriched probe sets. The other highly over-represented cellular component term in the seed-specific probe sets was cytoplasmic membrane-bounded vesicle (9.08% in seed coat, Figure 4.14), and this term was also highly over-represented in both pericarp (28.58%) and epiderm (16.81%) enriched probe sets (Figure 4.22).

4.3.6.2 Functional enrichment in vegetative-enriched transcripts

Functional analysis of vegetative tissue-enriched probe set lists was also performed for biological processes, molecular functions and cellular component GO annotations (Figures 4.23 to 4.25). While fewer biological process terms were enriched than in the vegetative tissue-specific probe set lists, some of the enriched terms were

Figure 4.23 Biological processes enriched in vegetative tissue-enriched probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-enriched gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list.



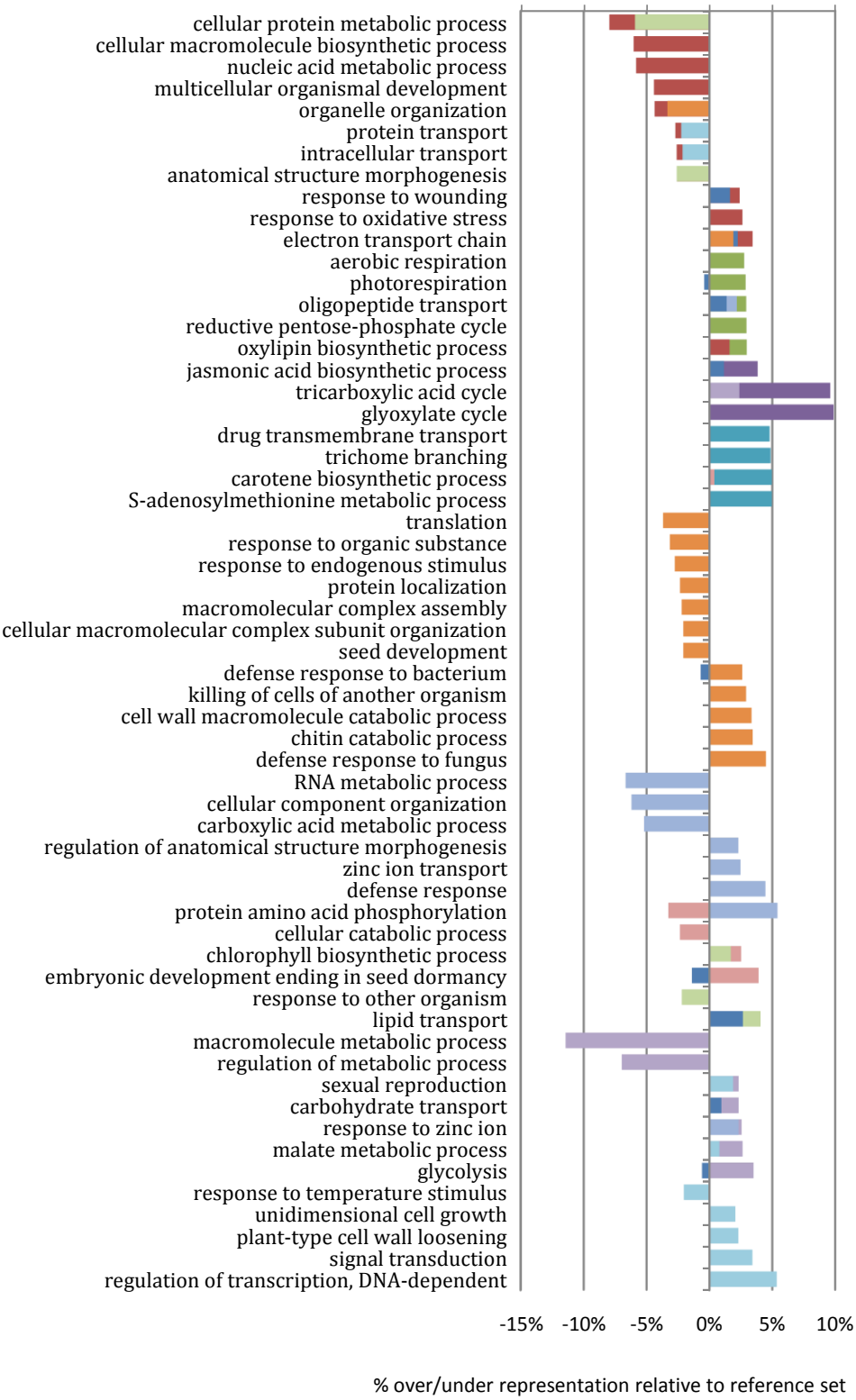
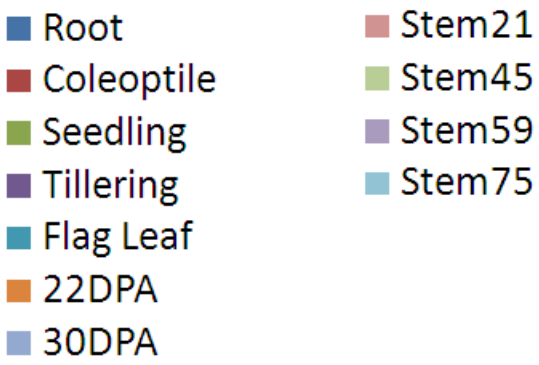

cellular protein metabolic process
cellular macromolecule biosynthetic process
nucleic acid metabolic process
multicellular organismal development
organelle organization
protein transport
intracellular transport
anatomical structure morphogenesis
response to wounding
response to oxidative stress
electron transport chain
aerobic respiration
photorespiration
oligopeptide transport
reductive pentose-phosphate cycle
oxylipin biosynthetic process
jasmonic acid biosynthetic process
tricarboxylic acid cycle
glyoxylate cycle
drug transmembrane transport
trichome branching
carotene biosynthetic process
S-adenosylmethionine metabolic process
translation
response to organic substance
response to endogenous stimulus
protein localization
macromolecular complex assembly
cellular macromolecular complex subunit organization
seed development
defense response to bacterium
killing of cells of another organism
cell wall macromolecule catabolic process
chitin catabolic process
defense response to fungus
RNA metabolic process
cellular component organization
carboxylic acid metabolic process
regulation of anatomical structure morphogenesis
zinc ion transport
defense response
protein amino acid phosphorylation
cellular catabolic process
chlorophyll biosynthetic process
embryonic development ending in seed dormancy
response to other organism
lipid transport
macromolecule metabolic process
regulation of metabolic process
sexual reproduction
carbohydrate transport
response to zinc ion
malate metabolic process
glycolysis
response to temperature stimulus
unidimensional cell growth
plant-type cell wall loosening
signal transduction
regulation of transcription, DNA-dependent
-15%
-10%
-5%
0%
5%
10%
% over/under representation relative to reference set

Figure 4.24 Molecular functions enriched in vegetative tissue-enriched probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-enriched gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list.



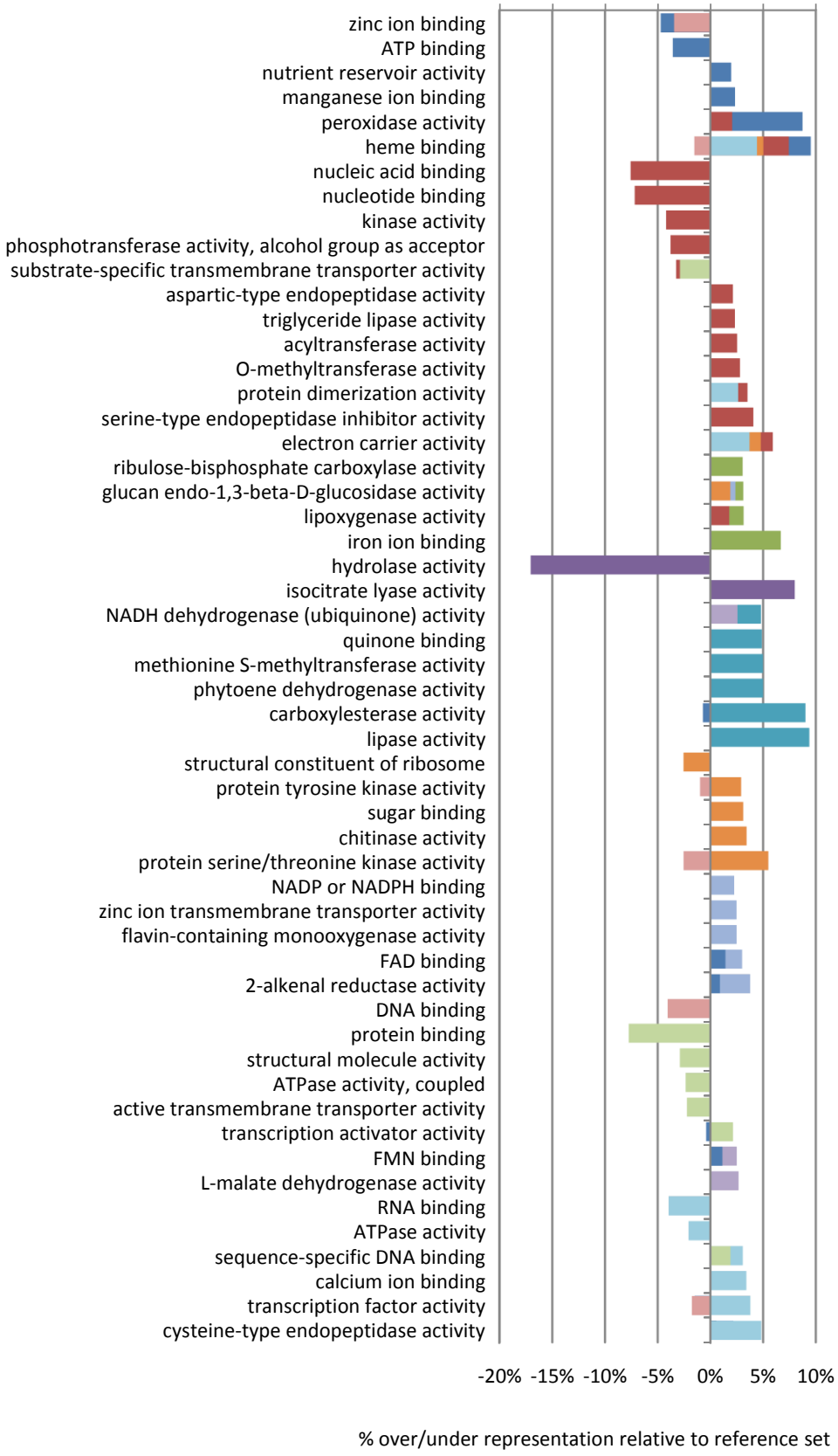
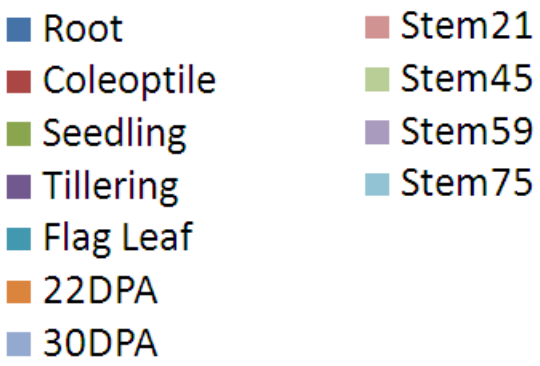
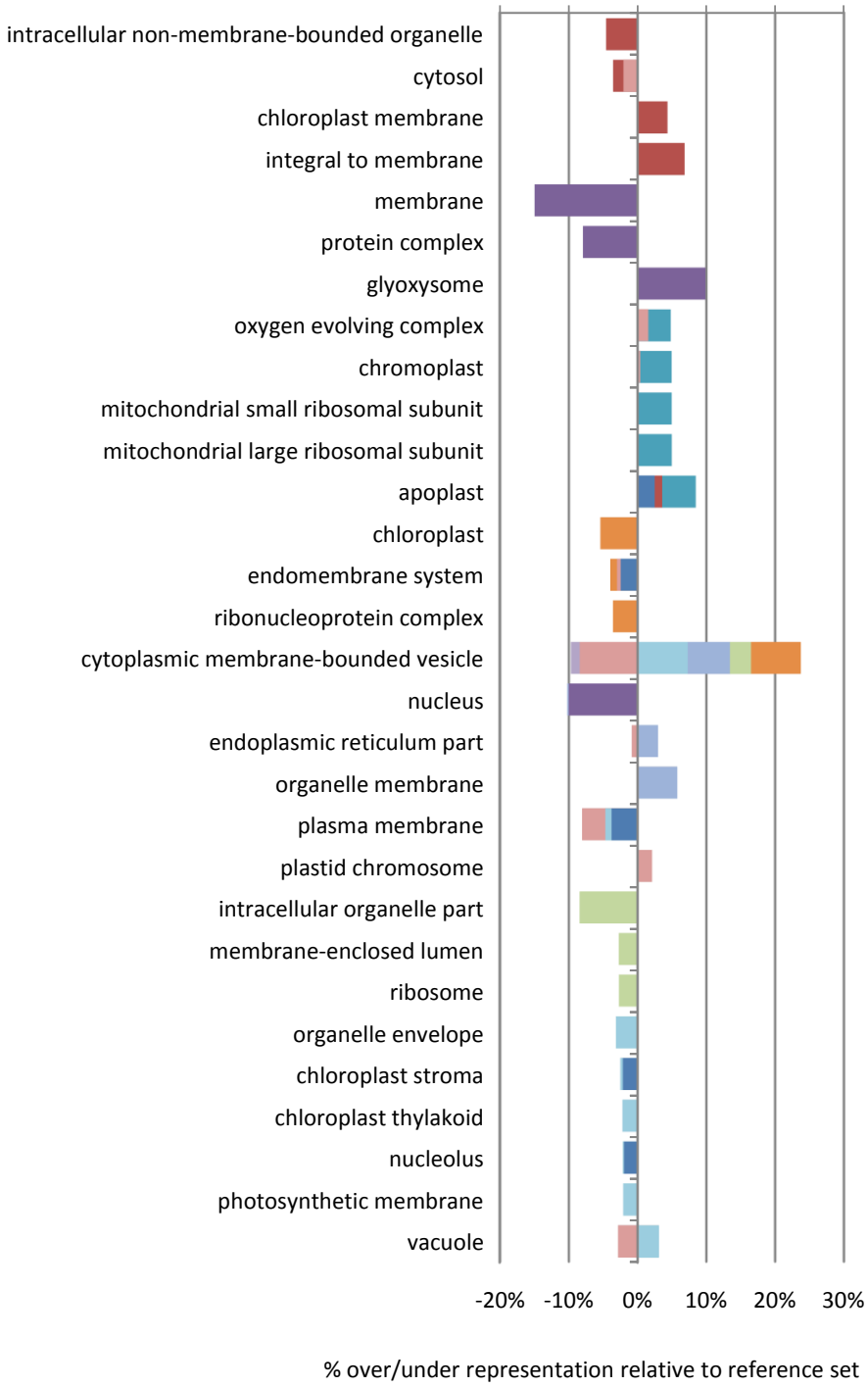

zinc ion binding
ATP binding
nutrient reservoir activity
manganese ion binding
peroxidase activity
heme binding
nucleic acid binding
nucleotide binding
kinase activity
phosphotransferase activity, alcohol group as acceptor
substrate-specific transmembrane transporter activity
aspartic-type endopeptidase activity
triglyceride lipase activity
acyltransferase activity
O-methyltransferase activity
protein dimerization activity
serine-type endopeptidase inhibitor activity
electron carrier activity
ribulose-bisphosphate carboxylase activity
glucan endo-1,3-beta-D-glucosidase activity
lipoxygenase activity
iron ion binding
hydrolase activity
isocitrate lyase activity
NADH dehydrogenase (ubiquinone) activity
quinone binding
methionine S-methyltransferase activity
phytoene dehydrogenase activity
carboxylesterase activity
lipase activity
structural constituent of ribosome
protein tyrosine kinase activity
sugar binding
chitinase activity
protein serine/threonine kinase activity
NADP or NADPH binding
zinc ion transmembrane transporter activity
flavin-containing monooxygenase activity
FAD binding
2-alkenal reductase activity
DNA binding
protein binding
structural molecule activity
ATPase activity, coupled
active transmembrane transporter activity
transcription activator activity
FMN binding
L-malate dehydrogenase activity
RNA binding
ATPase activity
sequence-specific DNA binding
calcium ion binding
transcription factor activity
cysteine-type endopeptidase activity
-20% -15% -10% -5% 0% 5% 10%
% over/under representation relative to reference set

Figure 4.25 Cellular components enriched in vegetative tissue-enriched probe set lists. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the tissue-enriched gene lists that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one tissue-enriched probe set list.




intracellular non-membrane-bounded organelle
cytosol
chloroplast membrane
integral to membrane
membrane
protein complex
glyoxysome
oxygen evolving complex
chromoplast
mitochondrial small ribosomal subunit
mitochondrial large ribosomal subunit
apoplast
chloroplast
endomembrane system
ribonucleoprotein complex
cytoplasmic membrane-bounded vesicle
nucleus
endoplasmic reticulum part
organelle membrane
plasma membrane
plastid chromosome
intracellular organelle part
membrane-enclosed lumen
ribosome
organelle envelope
chloroplast stroma
chloroplast thylakoid
nucleolus
photosynthetic membrane
vacuole
-20%
-10%
0%
10%
20%
30%
% over/under representation relative to reference set

similar. The most highly enriched terms were the tricarboxylic acid cycle and glyoxylate cycle in the tillering-enriched probe set list (Figure 4.23), the same terms that were most highly enriched in the tillering-specific probe set list (Figure 4.15). The most under-represented term was macromolecule metabolic process in the Stem59-enriched probe set list (Figure 4.23), an effect that was not observed in the Stem59-specific probe set list (Figure 4.15).

Functional analysis of vegetative tissue-enriched probe set list showed similar patterns of over and under-representation of molecular function terms (Figure 4.24) compared to vegetative tissue-specific probe set lists (Figure 4.16). Hydrolase activity was under-represented in both tillering-enriched and tillering-specific probe set lists by the same amount (17.05% and 17.04% respectively) suggesting that probe sets that are tillering-enriched but not tillering-specific do not contribute to this effect. Heme binding and peroxidase activity are both highly over-represented in root-enriched probe sets and carboxylesterase activity and lipase activity are both highly over-represented in flag leaf-enriched probe sets (Figure 4.24), though these terms are not as highly over-represented as they were in the tissue-specific probe set lists (Figure 4.16).

Many of the cellular component terms that were found to be over or under-represented in the vegetative tissue-enriched probe set lists (Figure 4.25) were similar to the terms over and under-represented in the vegetative tissue-specific probe set lists (Figure 4.17) however not all observations were the same. For example, there was an over-representation of ribosome in the Stem59-specific probe set list but there was no over-representation of ribosome observed in the Stem59-enriched probe sets. Cytoplasmic membrane-bounded vesicle was observed to be the most highly over-

represented term in both the vegetative tissue-specific and vegetative tissue-enriched probe set lists (Figures 4.15 and 4.23); however, in the tissue-specific lists this term was over-represented in leaf and root tissues, while in the tissue-enriched lists this term was over represented in leaf and stem tissues but also under-represented in some stem tissues as well. The pattern of oscillating representation of this term within the stem tissues (under in Stem21, over in Stem45, under in Stem59 and over in Stem75) is quite puzzling and may require further investigation to confirm this effect or determine the cause of such an effect.

Overall the functional analysis revealed a number of biological processes, molecular functions and cellular components that were over and under-represented in the tissue-enriched probe set lists and most of the terms seemed to make intuitive sense as to the tissues they represent. Together with the functional analysis of the tissue-specific probe set lists the results provide a picture of cellular state within each of the tissues and developmental stages studied.

4.3.7 Developmentally-regulated gene expression

To identify genes that were developmentally-regulated in leaf and stem tissues (leaf and stem DRGs), probe sets were filtered for significant variation among these tissues (adjusted p-value ≤ 0.05) and a minimum of 2 fold difference in mean expression values across the tissues following the conversion of expression values to zero in the case of absent expression (non majority P consensus detection call). Using the filtering criteria, 8,208 probe sets were found to represent leaf DRGs and 17,294 probe sets were found to represent stem DRGs. The complete results from the leaf and stem

developmentally regulated gene expression analysis are also presented in Supplementary Table S4.8. This table lists all probe sets on the wheat array and whether or not the probe set fits the criteria used to identify leaf and stem developmentally regulated genes. Supplementary Table S4.8 also contains the consensus detection calls, RMA normalized expression values, results from the ANOVA and FDR of leaf and stem tissues as well as the mean RMA values following the conversion to zero in the case of absent expression (non majority P consensus detection call) for reference. Supplementary Table S4.8 can be filtered for any of the criteria used to generate the leaf or stem DRGs probe set lists.

Groups of probe sets identified as tissue-enriched have similar expression patterns (namely up-regulated in the tissue of interest) and therefore the aggregative clustering method HCL was used to cluster these probe sets according to their level of similarity. The lists of probe sets identified as representing DRGs have several different patterns of expression and therefore the self organizing tree algorithm (SOTA), which is a divisive clustering method, was chosen to separate these genes into groups sharing similar patterns of expression (Herrero et al., 2001). To divide the DRGs into the appropriate number of clusters to reveal differences in expression patterns, the diversity as a percentage of the initial diversity was evaluated after each division for a maximum of 50 cycles of division (Figures 4.26 and 4.27). Based on the diversity history, it was observed that the diversity reached a near asymptote after approximately 30% of the initial diversity was reached in the resulting clusters. Increased cycles of division, after this plateau was reached, failed to separate the probe sets in a biologically meaningful way. Therefore, the number of cycles of division was chosen that would reduce the diversity to below 30% of the diversity of the initial cluster (leaf or stem DRGs).

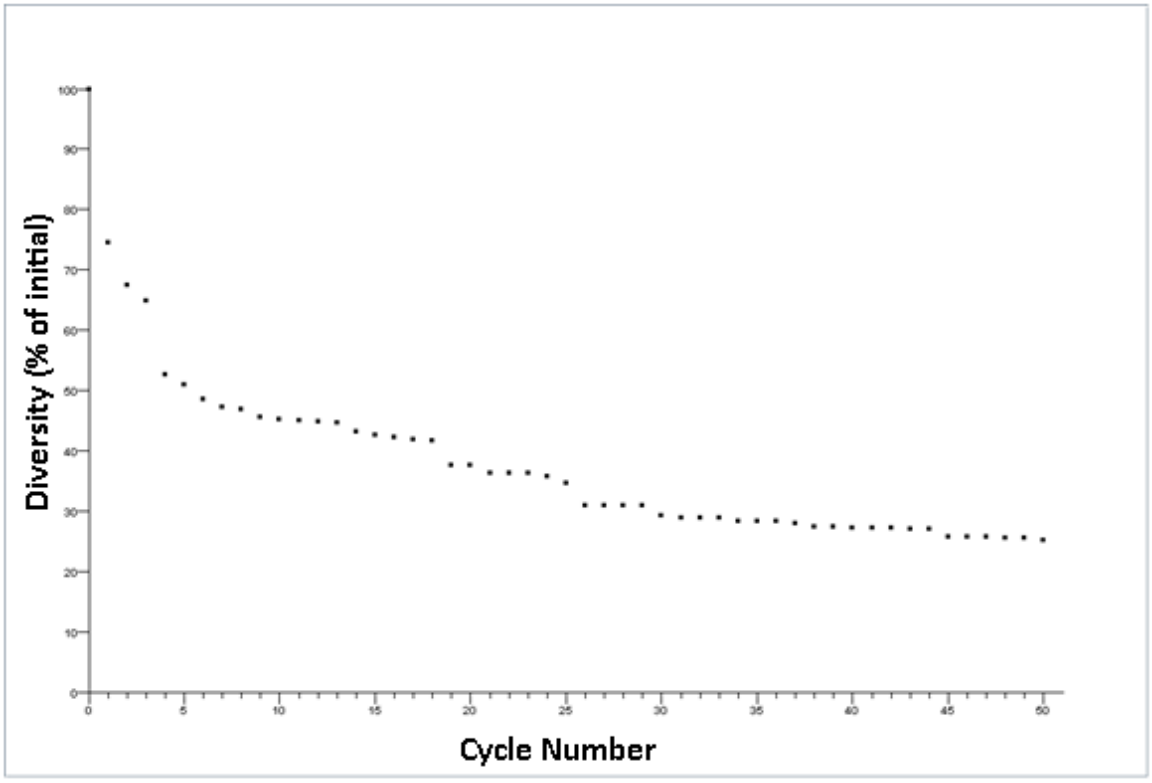


Figure 4.26 Tree diversity history plot for self organizing tree algorithm (SOTA) results of leaf differentially expressed probe sets. The diversity of the resulting clusters are plotted following each cycle of division relative to the diversity of the initial cluster as a percentage. The number of cycles of division was chosen based on the threshold of less than or equal to 30% diversity. The leaf developmentally regulated probe sets underwent 30 cycles of division to reach this threshold which resulted in 31 clusters of leaf developmentally regulated probe sets.

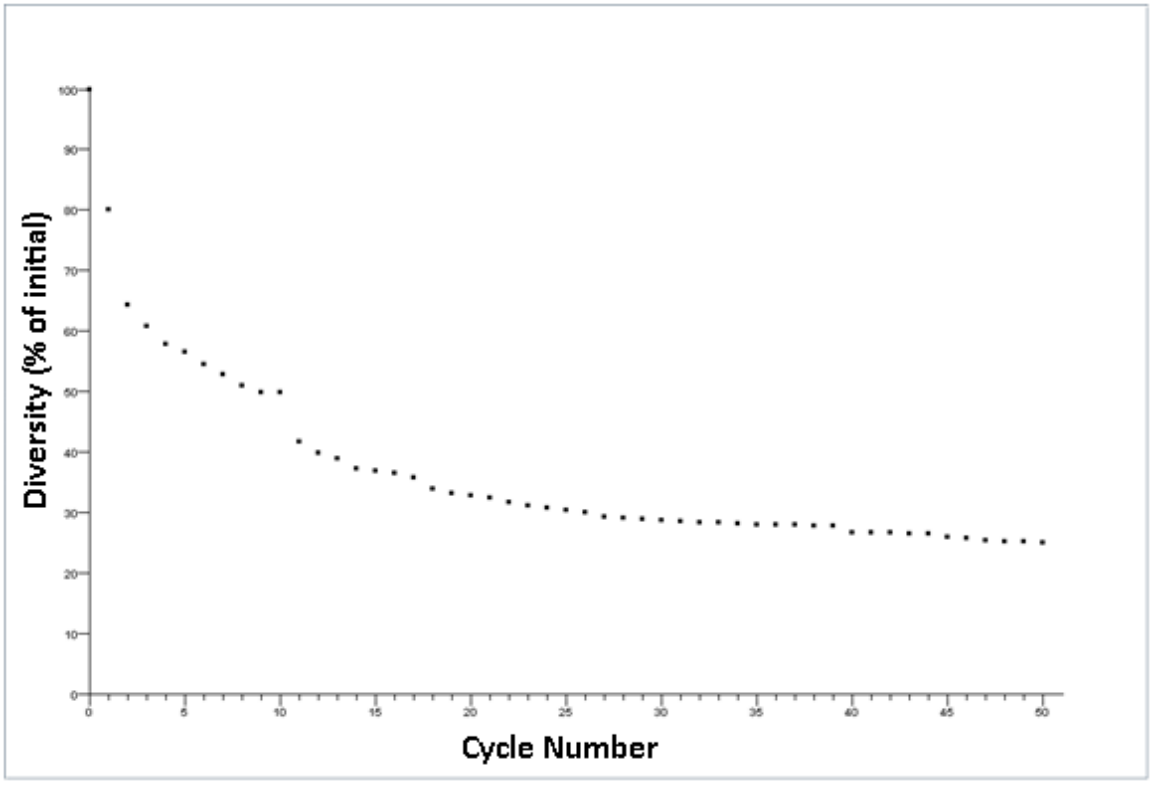


Figure 4.27 Tree diversity history plot for self organizing tree algorithm (SOTA) results of stem differentially expressed probe sets. The diversity of the resulting clusters are plotted following each cycle of division relative to the diversity of the initial cluster as a percentage. The number of cycles of division was chosen based on the threshold of less than or equal to 30% diversity. The stem developmentally regulated probe sets underwent 27 cycles of division to reach this threshold which resulted in 28 clusters of stem developmentally regulated probe sets.

For leaf DRGs this was 30 cycles which resulted in 31 clusters and for stem DRGs this was 27 cycles which resulted in 28 clusters (Figures 4.26 and 4.27). The resulting clusters contain probe sets with expression patterns that are more similar to each other than to the expression patterns of probe sets in the other SOTA clusters. The resulting SOTA clusters are summarized by centroid expression patterns (Figures 4.28 and 4.29). The leaf DRGs were divided 30 times into 31 clusters. These clusters ranged in number of probe sets from 4 to 1,600 (Figure 4.28, clusters 14 and 29). The stem DRGs were divided 27 times into 28 clusters which ranged from 109 to 2,978 probe sets (Figure 4.29, clusters 4, 11 and 25). Figures 4.30 and 4.31 show the expression graphs of the probe sets in each of the leaf and stem SOTA clusters respectively. The centroid expression is shown in magenta in these figures and highlights the different expression patterns present within the DRGs. The various patterns of expression that are observed in the leaf DRGs centroid expression graphs are up-regulation throughout development (cluster 25), down-regulation throughout development (clusters 13, 14, 16), up-regulation at a single developmental stage (clusters 1, 2, 6, 21), down-regulation at a single developmental stage (clusters 15, 18, 19, 20, 27), up-regulation during late stages of leaf development (clusters 7, 22, 24), down-regulation during late stages of leaf development (4, 8, 9, 10, 11, 12, 15, 17, 18), low expression throughout development (clusters 7, 22), mid expression throughout development (clusters 28, 29, 30) and high expression throughout development (cluster 31, Figure 4.30). Similar variation in patterns of expression were observed among the stem DRGs centroid expression graphs (Figure 4.31). The results of the leaf and stem SOTA clustering can also be found in Supplementary Table S4.8.

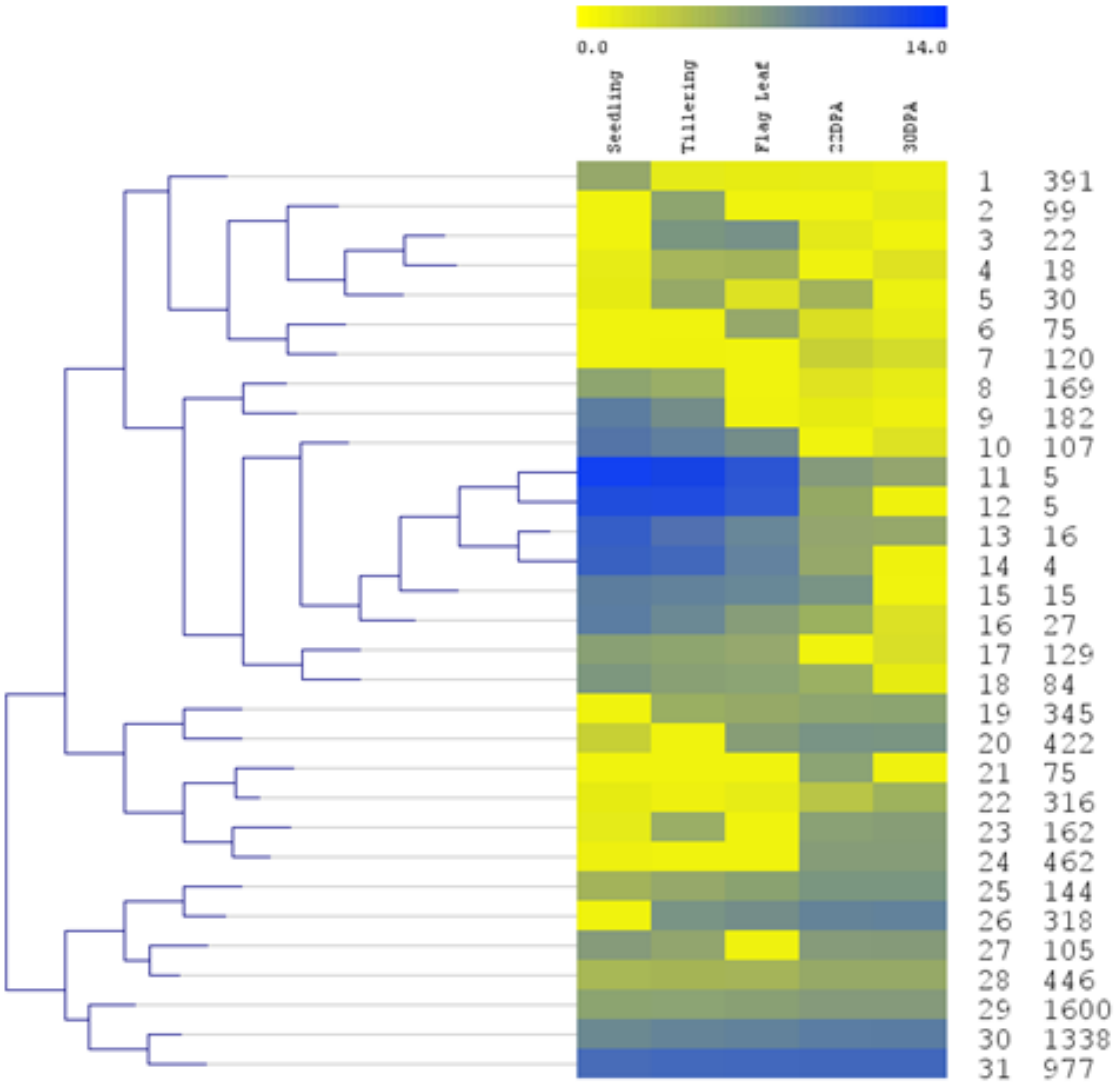


Figure 4.28 Centroid expression heat map and dendogram of leaf self organizing tree algorithm (SOTA) clusters. The dendogram to the left of the heat map shows the order of cluster division. The blue in the dendogram branches represents the order in which clusters were divided. The heat map shows the centroid expression patterns of the clusters of leaf developmentally regulated genes resulting from the SOTA. The numbers to the right of the heat map are: cluster numbers in the first column to the right; number of probe sets in the cluster in the second column to the right. The horizontal color scale at the top of the heat map represents the expression values with yellow representing no expression and dark blue representing maximum expression. The centroid expression value for each cluster in a particular tissue is represented by the intensity of color.

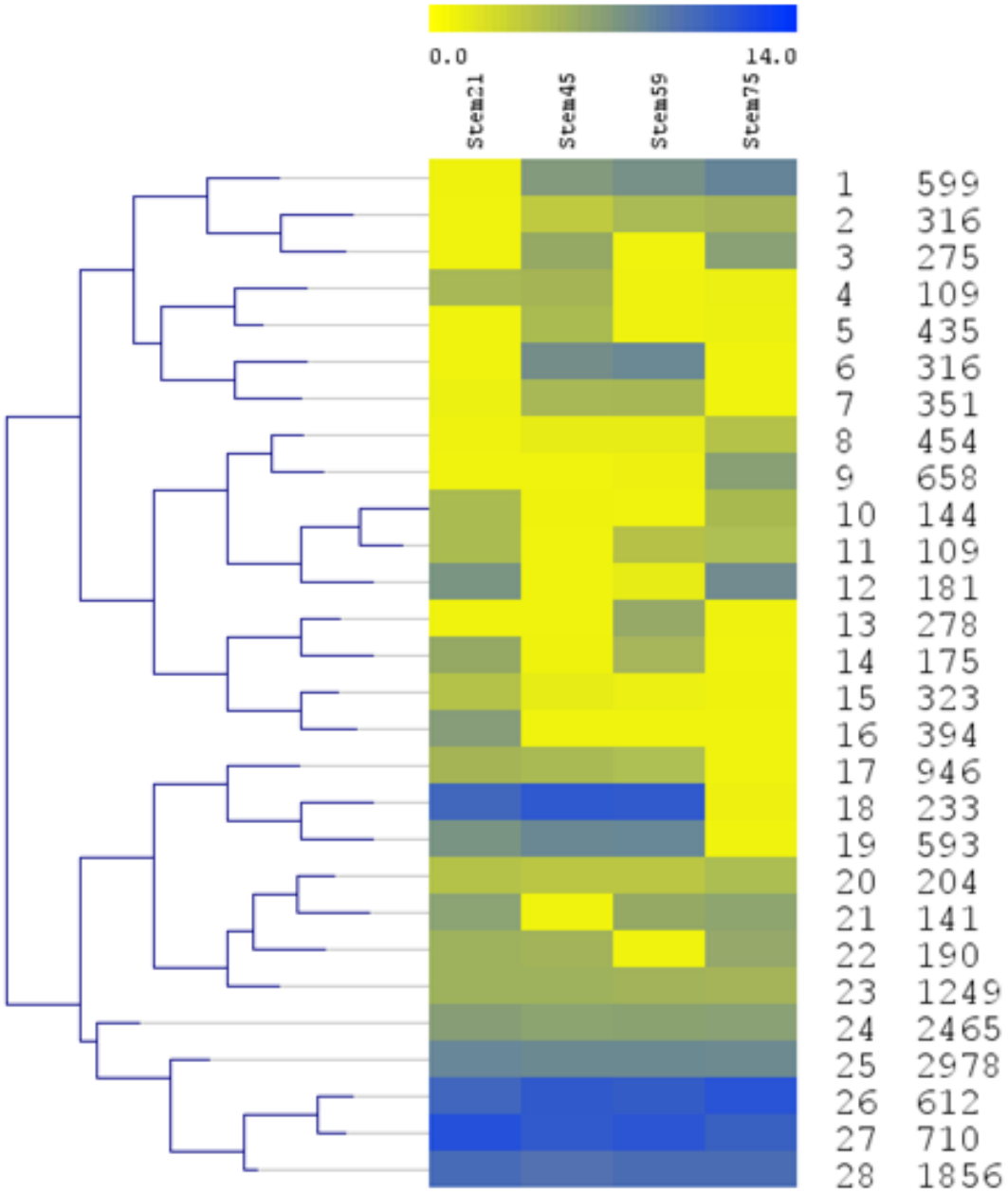


Figure 4.29 Centroid expression heat map and dendogram of stem self organizing tree algorithm (SOTA) clusters. The dendogram to the left of the heat map shows the order of cluster division. The blue in the dendogram branches represents the order in which clusters were divided. The heat map shows the centroid expression patterns of the clusters of stem developmentally regulated genes resulting from the SOTA. The numbers to the right of the heat map are: cluster numbers in the first column to the right; number of probe sets in the cluster in the second column to the right. The horizontal color scale at the top of the heat map represents the expression values with yellow representing no expression and dark blue representing maximum expression. The centroid expression value for each cluster in a particular tissue is represented by the intensity of color.

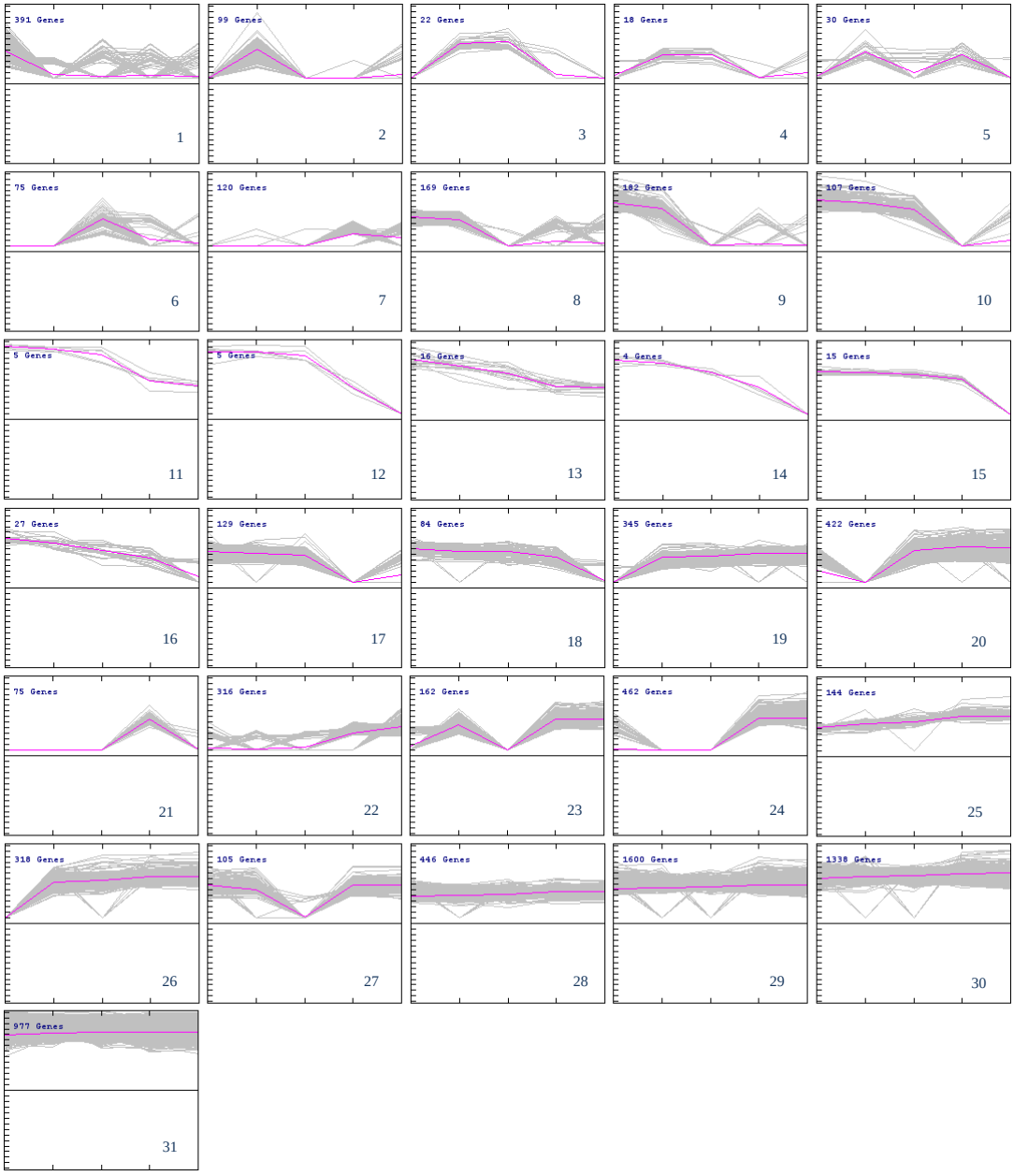


Figure 4.30 Expression graphs for the self organizing tree algorithm (SOTA) clusters of leaf developmentally regulated genes. The 31 clusters (numbered in the lower right of each graph) resulting from the SOTA of all leaf developmentally regulated genes are shown with the number of probe sets in each cluster listed in the upper left corner of each expression graph. The grey lines represent the expression pattern of each probe set in a cluster. The magenta lines represent the centroid expression of all probe sets in a cluster. Vertical tic marks on the horizontal axis represent the leaf tissues at the five developmental stages (left to right; Seedling, Tillering, Flag Leaf, 22DPA and 30DPA). Horizontal tick marks on the vertical axis represent RMA expression values with distance between each tic representing a difference in expression of 1 ($\log_2$ scale).

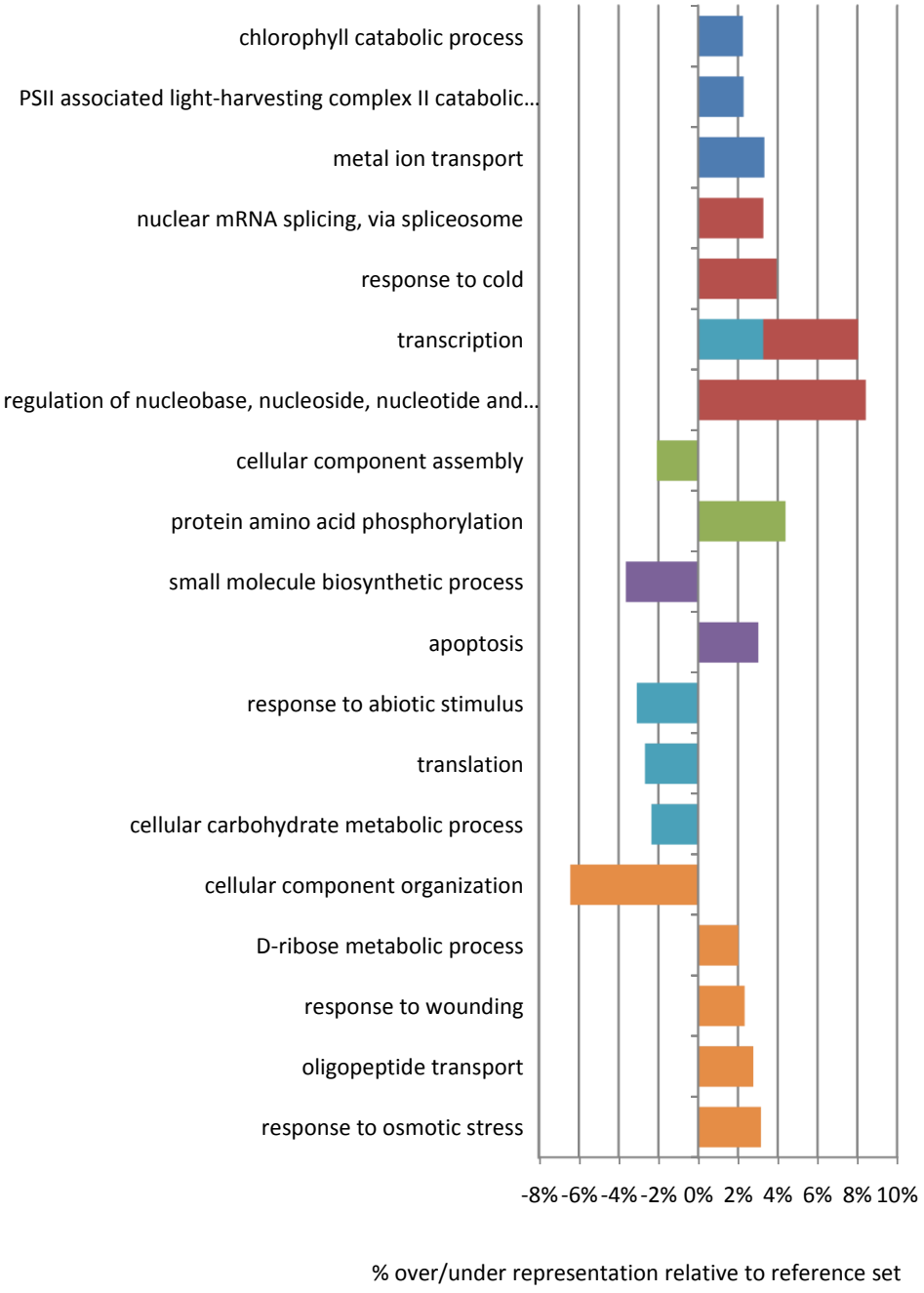
Figure 4.31 Expression graphs for the self organizing tree algorithm (SOTA) clusters of stem developmentally regulated genes. The 28 clusters (numbered in the lower right of each graph) resulting from the SOTA of all stem developmentally regulated genes are shown with the number of probe sets in each cluster listed in the upper left corner of each expression graph. The grey lines represent the expression pattern of each probe set in a cluster. The magenta lines represent the centroid expression of all probe sets in a cluster. Vertical tic marks on the horizontal axis represent the stem tissues at the four developmental stages (left to right; Stem21, Stem45, Stem59, Stem75). Horizontal tick marks on the vertical axis represent RMA expression values with distance between each tic representing a difference in expression of 1 ($\log_2$ scale).

4.3.7.1 Functional analysis of leaf senescence associated genes

The division of leaf DRGs into clusters of similar expression pattern revealed a number of clusters that appear to be leaf senescence associated genes (SAGs) with up-regulation of expression at later stages of leaf development. Clusters 7, 21, 22, 23, 24 and 27 all show patterns of expression that may be indicative of leaf SAGs (Figure 4.30). To demonstrate the utility in dividing DRGs into clusters of similar expression patterns, functional enrichment analysis of the probe sets from these leaf SAG clusters was performed to identify molecular functions and biological processes that may be important in leaf senescence.

Functional analysis revealed an over-representation of biological processes involved in cellular degradation and cell death processes including chlorophyll catabolic process, PSII associated light-harvesting complex II catabolic process, nuclear mRNA splicing via spliceosome and apoptosis (Figure 4.32) as would be expected during senescence. There was also an under-representation of organizational processes including cellular component assembly, small molecule biosynthetic process and cellular component organization. A number of stress responses were over-represented including response to cold, response to wounding and response to osmotic stress which highlights the overlap between senescence and stress response gene expression (Lim et al., 2007; Krupinska and Humbeck, 2008; Kudryakova, 2009). Interestingly transcription was over-represented in clusters 21 and 24 and regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process was over-represented in cluster 21 (Figure 4.32). These clusters show expression up-regulation at the early senescence leaf development stage (22DPA) suggesting that the transcriptional response that initiates leaf

Figure 4.32 Biological processes enriched in leaf senescence associated probe sets. SOTA clusters of leaf developmentally regulated probe sets (DRGs) showing patterns indicative of senescence associated genes (SAGs) were subjected to functional enrichment analysis. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the leaf DRGs SOTA clusters that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one SOTA cluster.


chlorophyll catabolic process
PSII associated light-harvesting complex II catabolic...
metal ion transport
nuclear mRNA splicing, via spliceosome
response to cold
transcription
regulation of nucleobase, nucleoside, nucleotide and...
cellular component assembly
protein amino acid phosphorylation
small molecule biosynthetic process
apoptosis
response to abiotic stimulus
translation
cellular carbohydrate metabolic process
cellular component organization
D-ribose metabolic process
response to wounding
oligopeptide transport
response to osmotic stress
-8% -6% -4% -2% 0% 2% 4% 6% 8% 10%
% over/under representation relative to reference set

senescence occurs during early leaf senescence and that the timing of leaf tissue sampling was able to capture this response.

Functional analysis also revealed a number of molecular functions that were over and under-represented in the leaf senescence DRGs (Figure 4.33). Molecular functions that were over-represented include a number of enzyme activities including ATP-dependent helicase, protein serine/threonine kinase, exonuclease, monooxygenase, ribokinase, fructokinase, acetolactate synthase, inositol 3-alpha-galactosyltransferase, muconate cycloisomerase and serine-type carboxypeptidase activities (Figure 4.33). The over-representation of such a large number of enzyme activities may reflect the active degradative processes that occur during leaf senescence (Gregersen and Holm, 2007; Lim et al., 2007). An over-representation of electron carrier activity and ATP binding reflect the energy requirements provided by the mitochondria in leaf tissue during programmed developmental senescence (Figure 4.33). It has been shown that during senescence when photosynthetic activity decreases due to the degradation of chloroplasts, there is an increase in ATP/ADP ratio but an overall decline in adenylate content as there is an increased dependence on mitochondrial respiration (Keskitalo et al., 2005).

Cellular components under and over-represented in leaf senescence DRGs support the overall picture provided by the biological processes and molecular functions. Mitochondrial inner membrane, cajal body and nucleolus terms are over-represented in early senescence associated clusters highlighting the energy and transcriptional requirements of the senescence initiation (Figure 4.34). Cellular components are for the most part under-represented in senescence associated DRGs as would be expected from the general cellular degradation that occurs during senescence (Gregersen and Holm,

Figure 4.33 Molecular functions enriched in leaf senescence associated probe sets. SOTA clusters of leaf developmentally regulated probe sets (DRGs) showing patterns indicative of senescence associated genes (SAGs) were subjected to functional enrichment analysis. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the leaf DRGs SOTA clusters that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one SOTA cluster.

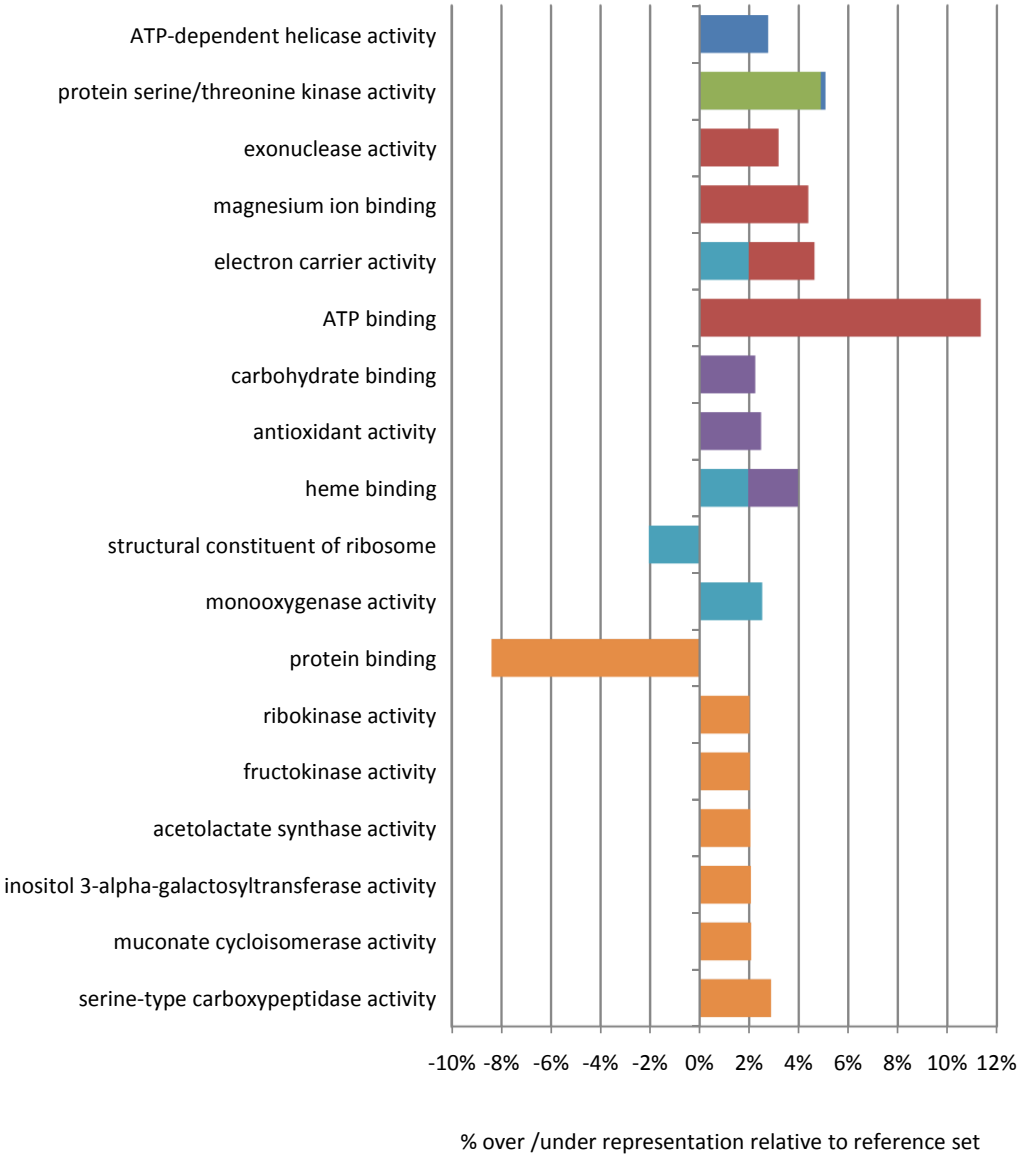
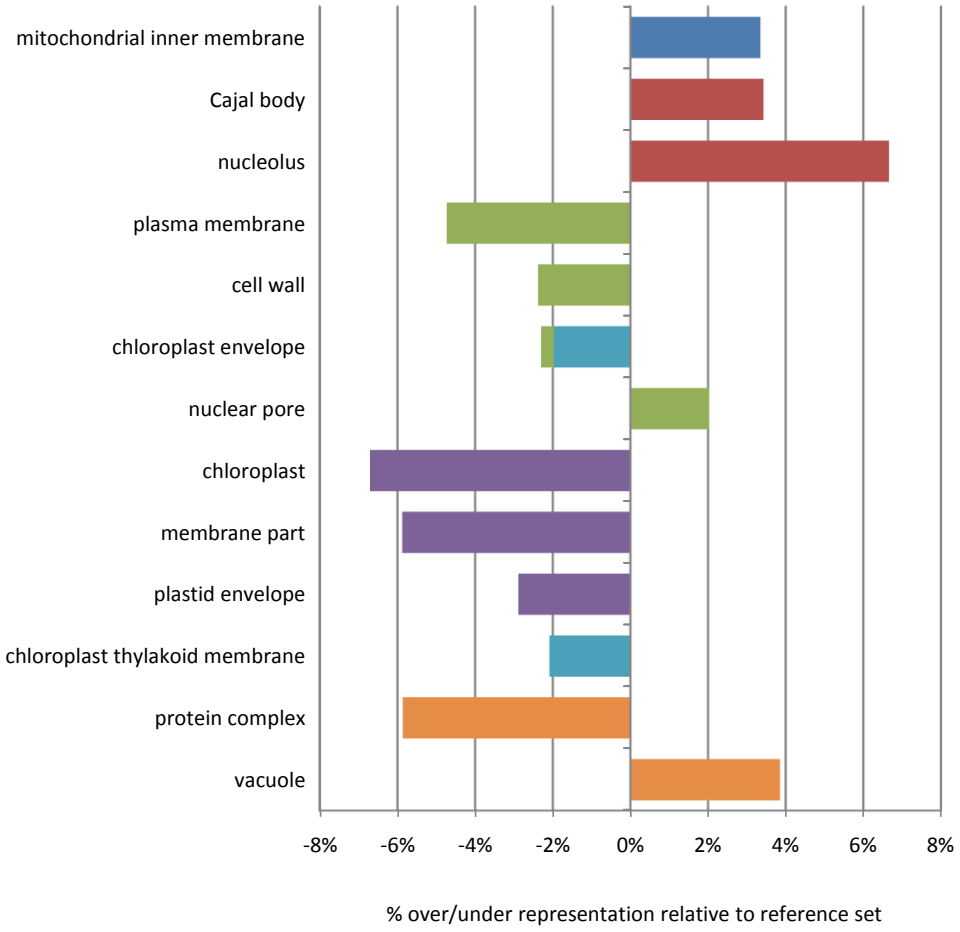

ATP-dependent helicase activity
protein serine/threonine kinase activity
exonuclease activity
magnesium ion binding
electron carrier activity
ATP binding
carbohydrate binding
antioxidant activity
heme binding
structural constituent of ribosome
monooxygenase activity
protein binding
ribokinase activity
fructokinase activity
acetolactate synthase activity
inositol 3-alpha-galactosyltransferase activity
muconate cycloisomerase activity
serine-type carboxypeptidase activity
-10% -8% -6% -4% -2% 0% 2% 4% 6% 8% 10% 12%
% over /under representation relative to reference set

Figure 4.34 Cellular components enriched in leaf senescence associated probe sets. SOTA clusters of leaf developmentally regulated probe sets (DRGs) showing patterns indicative of senescence associated genes (SAGs) were subjected to functional enrichment analysis. Enrichment is represented as the % over or under represented relative to the reference set of probe set annotations (those that are expressed in any of the triticale tissues studied). Enrichment results were filtered by a p-value < 0.05 and a difference of greater than or equal to 2% occurrence of the term between the test set and the reference set of annotations. Only the most specific terms in each branch of the ontology are shown for each gene list to avoid redundancy of parent/child terms. The legend below represents the leaf DRGs SOTA clusters that were used for functional enrichment analysis. The bars in the chart are overlapped to show when a term is enriched in more than one SOTA cluster.


mitochondrial inner membrane
Cajal body
nucleolus
plasma membrane
cell wall
chloroplast envelope
nuclear pore
chloroplast
membrane part
plastid envelope
chloroplast thylakoid membrane
protein complex
vacuole
-8%
-6%
-4%
-2%
0%
2%
4%
6%
8%
% over/under representation relative to reference set

2007; Lim et al., 2007). The vacuole however, is over-represented in cluster 27 which is up-regulated during early (22DPA) and late (30DPA) senescence (Figure 4.30). The vacuole is the site of storage of cellular degradation enzymes and is employed as a lytic organelle in nutrient recycling during senescence (Hopkins et al., 2007). The functional analysis results provide a picture of the processes, functions and cellular components that are involved in leaf senescence and show the effectiveness of SOTA clustering of DRGs in understanding plant development.

4.3.8 qRT-PCR

From the generated lists of tissue-specific tissue-enriched and developmentally-regulated probe sets, several candidate probe sets were chosen for verification of expression profiles using qRT-PCR within the leaf and seed tissues (Table 4.6). Most tissue-specific and tissue-enriched candidates were chosen for their high levels of expression in the tissue of interest, a property that is useful for the isolation of promoters to drive transgene expression. Within the tissue samples surveyed by qRT-PCR the microarray mean RMA expression values showed highly differential expression patterns (Figure 4.35).

Because of the range of tissue types used in the experiment and the concern that expression of a single housekeeping gene may not be constant across all tissues, the geometric mean of five putative housekeeping genes was used as an internal reference. All five housekeeping genes were evaluated for their expression stability in leaf and seed tissues and all had a stability measure ≤ 1.11 with an average value of 0.95 (Table 4.7). Normalization factors for tissue samples calculated using the geometric mean of the

Table 4.6 Probe sets chosen for verification of microarray expression patterns using qRT-PCR.

Probe set ID	GenBank Accession no.	Annotation	Microarray Expression Pattern
TaAffx.99992.1.S1_at	CA593620	isocitrate lyase	Tillering-specific
Ta.27399.3.S1_at	CA718775	glycine-rich protein / late embryogenesis abundant protein	Embryo-enriched
Ta.2798.1.S1_s_at	BJ290016	early-methionine-labelled polypeptide	Embryo-enriched
Ta.5720.1.S1_at	BJ296692	defensin precursor	Embryo-enriched
TaAffx.69777.1.S1_at	J02961.1	w heat germ agglutinin isolectin complexes	Embryo-enriched
Ta.23142.5.S1_x_at	CD868513	low molecular w eight glutenin	Endosperm-enriched
Ta.11127.2.S1_a_at	CA708523	starch brancing enzyme 3	Endosperm-enriched
Ta.13160.2.S1_x_at	CA635509		Crease-enriched
Ta.20064.1.S1_at	CA666265	ubiquitin-conjugating enzyme	Crease-specific
Ta.30755.1.S1_at	CN007962	nonspecific lipid-transfer protein precursor	Pericarp enriched
Ta.590.2.S1_x_at	CA601369	high light protein; kda jasmonate-induced protein	Leaf developmentally regulated. Dow n regulated throughout development

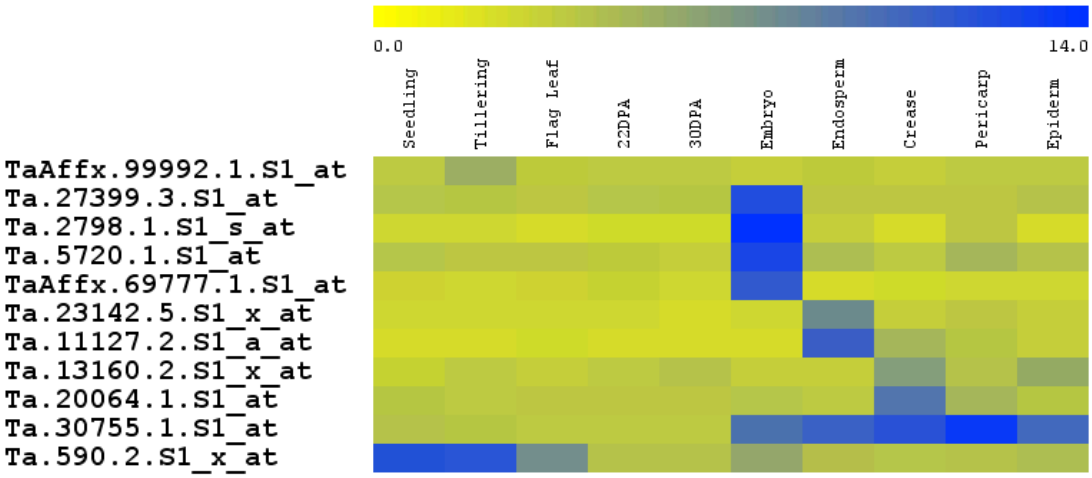


Figure 4.35 Expression heat maps of tissue-specific, enriched and developmentally-regulated probe sets chosen for verification of expression by qRT-PCR. Tissue-specificity was determined by a majority present consensus detection call (>50% P) in the tissue of interest and a majority absent consensus detection call (>50% A) in all other tissues. Tissue-enrichment was determined by probe sets having expression significantly different in the tissue of interest from other tissues (ANOVA and FDR adjusted p-value ≤ 0.05) and a minimum of 2-fold change in expression between the tissue of interest and all other tissues. Developmental down-regulation was determined by significant changes in expression across the leaf developmental stages (ANOVA and FDR adjusted p-value ≤ 0.05). All expression values were transformed to a logarithmic scale (base 2).

Table 4.7 Reference gene quality control for putative housekeeping genes. Five genes were chosen for use as an internal control for qRT-PCR. The coefficient of variation (CV) represents the variation of the normalized relative quantities of a reference gene across all samples. Lower CV values denote higher stability. The stability measure value is the gene expression stability parameter as calculated by geNorm (Vandesompele et al., 2002). The lower the stability measure value, the more stable expression is for the reference gene.

Gene	CV	Stability Measure (GeNorm)
Actin	41.4%	1.04
ALD	58.5%	1.11
Contig 1	35.0%	0.83
Contig 5	40.5%	0.94
Ef1a	27.0%	0.81
Mean	40.5%	0.95

reference genes are shown in Figure 4.36. The normalization factors are distributed around 1.0 and range from 0.26 (± 0.03) to 2.14 (± 0.09). Large deviations from 1.0 (i.e., >5) indicate large differences in starting material quantity or quality or a problem with a reference gene (Hellemans et al., 2007). Therefore, these are acceptable deviations from 1.0 considering the diversity in the samples (diverse tissue types).

The expression of the tillering-specific candidate gene, putative isocitrate lyase transcript (TaAffx.99992.1.S1_at) showed over 10 fold and 100 fold higher expression in tillering tissue compared to other tissues by microarray analysis and qRT-PCR respectively (Figure 4.37). The tillering-specific expression was underestimated by microarray analysis which may be a consequence of the compression of the microarray data due to the normalization.

Embryo-enriched genes showed varying degrees of tissue-specific up-regulation by qRT-PCR. The putative glycine-rich protein transcript (Ta.27399.3.S1_at) had over 100 fold higher expression in embryo than all other tissues in the microarray and qRT-PCR results (Figure 4.38 A). However, expression of this transcript was almost undetectable by qRT-PCR in all other leaf and seed tissues assayed. The early-methonine-labelled (Em) protein transcript (Ta.2798.1.S1_s_at), defensin transcript (Ta.5720.1.S1_at) and wheat germ agglutinin (TaAffx.69777.1.S1_at) on the other hand, all had approximately 1000 fold higher expression in the embryo compared to all other tissues in the qRT-PCR data and had similar levels of fold change as in the microarray data (Figure 4.38 B to D). Expression levels were most similar between qRT-PCR and microarray for the Em protein and the putative defensin while levels of expression for the glycine-rich protein and wheat germ agglutinin showed larger differences between

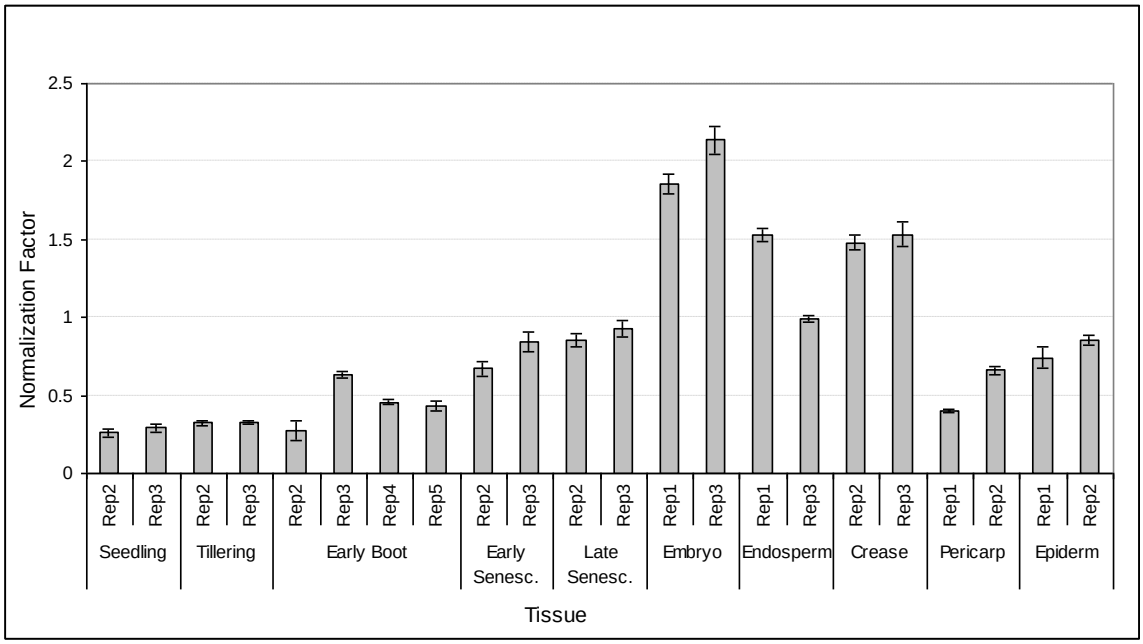


Figure 4.36 Distribution of sample normalization factors. Normalization factors were calculated using the geometric mean of the five reference genes. Two biological replicate samples were used for each tissue type.

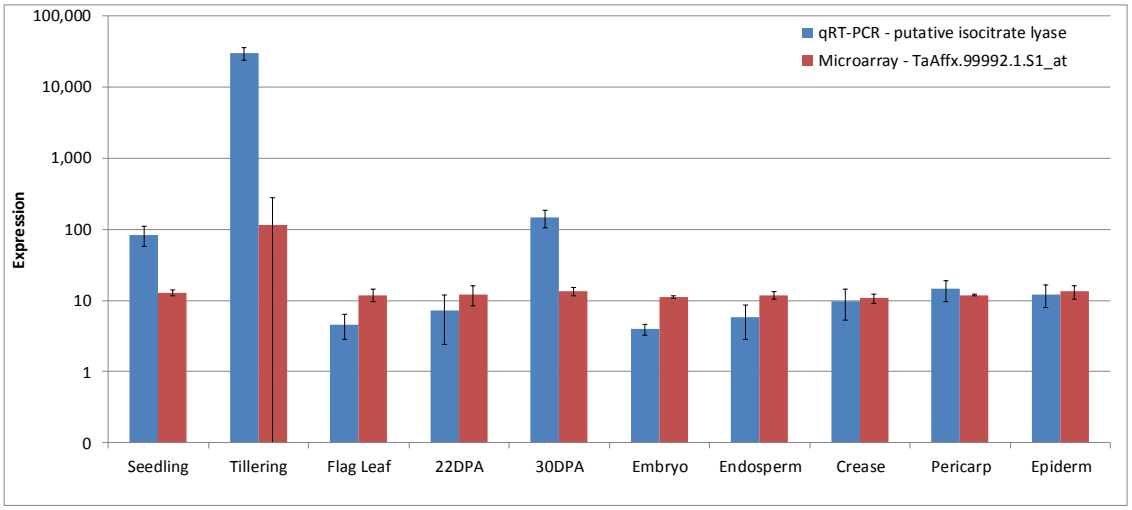
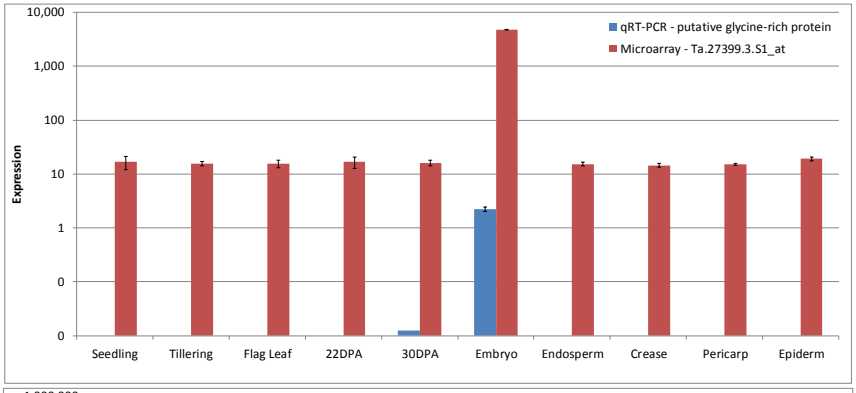


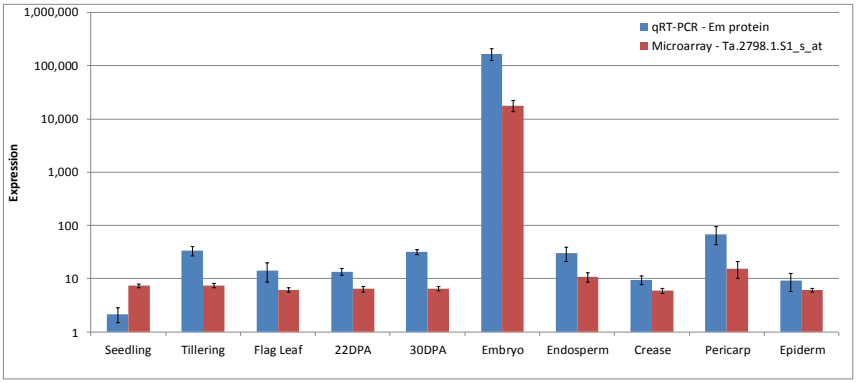
Figure 4.37 Relative expression of a candidate tillering-specific gene. Probe set TaAffx.99992.1.S1_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST CA593620, a putative isocitrate lyase. Values shown are mean expression values (microarray n = 5 for Flag Leaf, 2 for pericarp and epiderm and 3 for all other tissues, qRT-PCR n = 6 for all tissues, 3 technical replicates each for 2 biological replicates of each tissue). Error bars represent 1 standard deviation from the mean and have different lengths above and below the mean due to the log scale of the y axis. The large error bar shown for the Tillering tissue microarray data is due to a lack of detection in one of the microarray reps (still tillering-specific using the majority P criteria). Expression is only shown for tissues that were assayed using qRT-PCR.

Figure 4.38 Relative expression of four candidate embryo-enriched genes. A) Probe set Ta.27399.3.S1_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST CA718775, a putative glycine-rich protein transcript. B) Probe set Ta.2798.1.S1_s_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST BJ290016, an early methionine-labelled polypeptide transcript. C) Probe set Ta.5720.1.S1_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST BJ296692, a defensin precursor transcript. D) Probe set TaAffx.69777.1.S1_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST J02961.1, a wheat germ agglutinin isolectin complex protein transcript. Values shown are mean expression values (microarray n = 5 for Flag Leaf, 2 for pericarp and epiderm and 3 for all other tissues, qRT-PCR n = 6 for all tissues, 3 technical replicates each for 2 biological replicates of each tissue). Error bars represent 1 standard deviation from the mean and have different lengths above and below the mean due to the log scale of the y axis. Expression is only shown for tissues that were assayed using qRT-PCR.

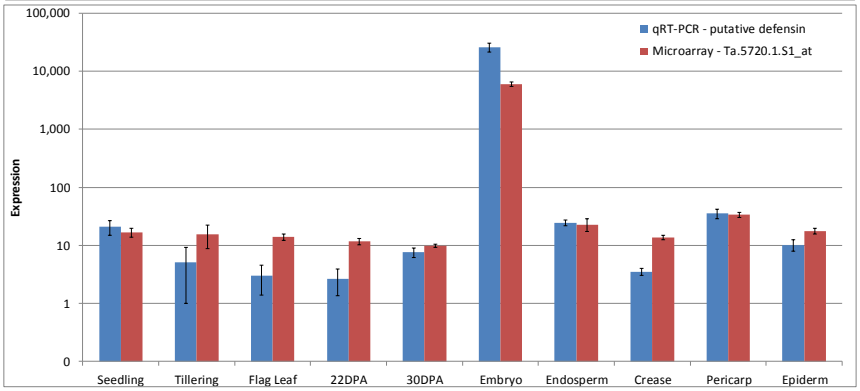
A


qRT-PCR - putative glycine-rich protein
Microarray - Ta.27399.3.S1_at
Expression
10,000
1,000
100
10
1
0
0
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Embryo
Endosperm
Crease
Pericarp
Epiderm

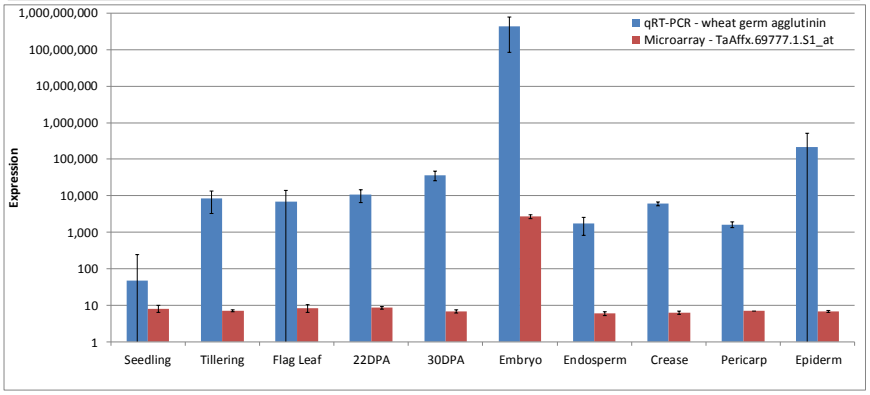
B


qRT-PCR - Em protein
Microarray - Ta.2798.1.S1_s_at
Expression
1,000,000
100,000
10,000
1,000
100
10
1
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Embryo
Endosperm
Crease
Pericarp
Epiderm

C


qRT-PCR - putative defensin
Microarray - Ta.5720.1.S1_at
Expression
100,000
10,000
1,000
100
10
1
0
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Embryo
Endosperm
Crease
Pericarp
Epiderm

D


qRT-PCR - wheat germ agglutinin
Microarray - TaAffx.69777.1.S1_at
Expression
1,000,000,000
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
100
10
1
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Embryo
Endosperm
Crease
Pericarp
Epiderm

microarray and qRT-PCR results. The glycine-rich protein expression was overestimated, while the wheat germ agglutinin expression was underestimated by microarray analysis in comparison to qRT-PCR.

Although the glycine-rich protein embryo qRT-PCR expression values are lower compared to the other embryo-enriched genes (Figure 4.38 A, y axes), the glycine-rich protein may be the most tissue-specific among the leaf and seed tissues, as the expression values in the other tissues are relatively low (average less than 1) and no expression was detected in some tissues by qRT-PCR. Wheat germ agglutinin showed the highest levels of expression in the embryo with expression values over 100 million but also showed higher levels of expression in other tissues compared to the other embryo candidate genes (Figure 4.38 D).

The two endosperm-enriched candidate genes showed at least 10 fold higher expression in endosperm tissue than in all other tissues in both microarray and qRT-PCR results (Figure 4.39 A and B). The low molecular weight (LMW) glutenin (Ta.23142.5.S1_x_at) showed greater differences in levels of expression between endosperm tissue and leaf and embryo tissues than between endosperm tissue and seed coat tissues (Figure 4.39 A). This may be a result of contamination of seed coat tissues by endosperm cells during dissection. The starch branching enzyme IIb transcript (Ta.11127.2.S1_a_at) demonstrated particularly high levels of endosperm-specific up regulation among the tissues assayed in the qRT-PCR results (Figure 4.39 B) with expression levels less than 1 in all other tissues. Expression of this transcript was however detected in the other seed coat tissues, which again suggests that there may be small amounts of tissue contamination due to the dissection process.

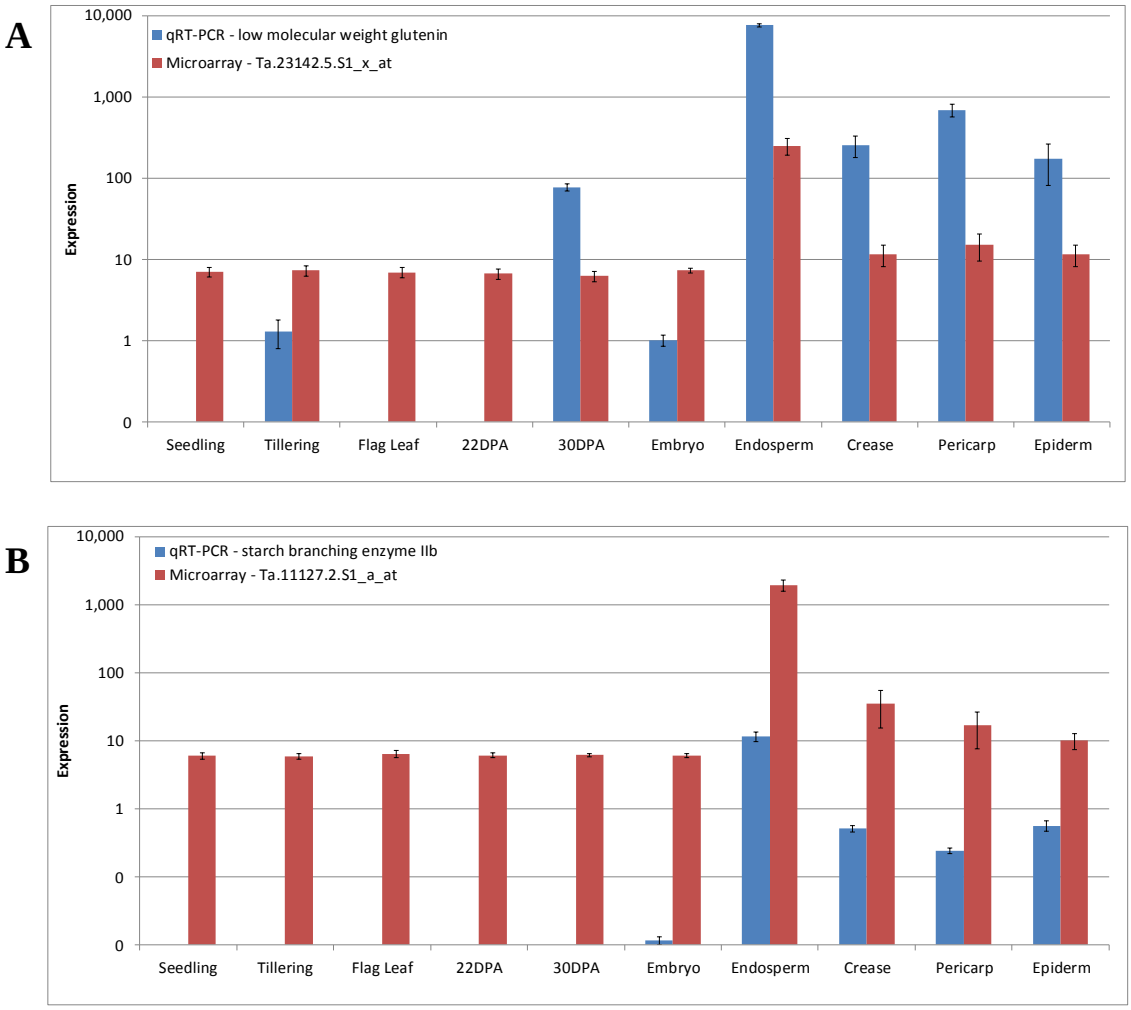


Figure 4.39 Relative expression of two candidate endosperm-enriched genes. A) Probe set Ta.23142.5.S1_x_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST CD868513, a low molecular weight glutenin transcript. B) Probe set Ta.11127.2.S1_a_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST CA708523, a starch branching enzyme 3 transcript. Values shown are mean expression values (microarray n = 5 for Flag Leaf, 2 for pericarp and epiderm and 3 for all other tissues, qRT-PCR n = 6 for all tissues, 3 technical replicates each for 2 biological replicates of each tissue). Error bars represent 1 standard deviation from the mean and have different lengths above and below the mean due to the log scale of the y axis. Expression is only shown for tissues that were assayed using qRT-PCR.

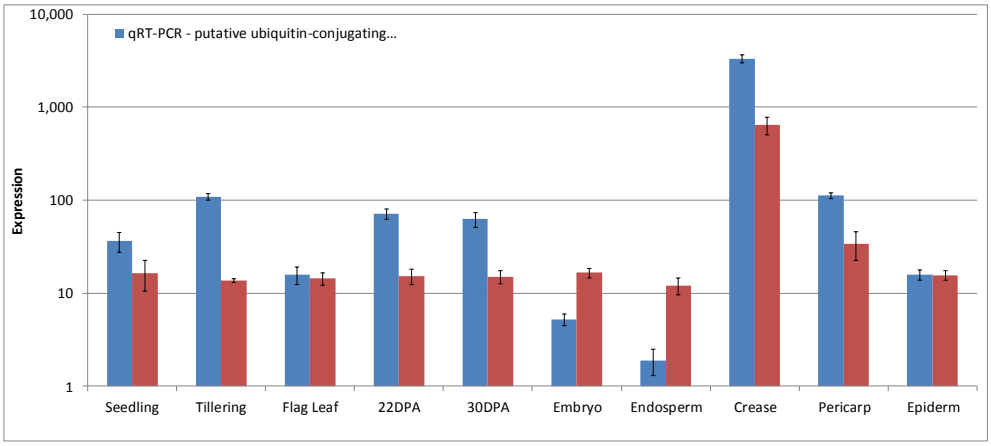
The two chosen crease candidate genes showed varying levels of expression as well as varying levels of specificity (Figure 4.40 A and B). The putative crease-specific ubiquitin-conjugating enzyme transcript (Ta.20064.1.S1_at, Figure 4.40 A) had higher levels of expression than the putative protein kinase transcript (Ta.13160.2.S1_x_at, Figure 4.40 B) in the crease tissue in both microarray analysis and qRT-PCR. Microarray analysis expression values in other tissues were very similar between the two transcripts; however, qRT-PCR expression values for other tissues were higher for the putative ubiquitin-conjugating enzyme than the putative protein kinase (3 fold to over 100 fold higher).

A putative lipid transfer protein transcript was identified by the microarray analysis as a pericarp-enriched candidate. The representative probe set showed consistently low levels of expression in all leaf tissues and expression in all seed tissues almost 100 fold higher than all leaf tissues by microarray analysis (Figure 4.40 C). Verification of expression by qRT-PCR revealed that expression of the lipid transfer protein did indeed have over 10 fold higher expression in all seed tissues but also that the expression was higher than estimated by the microarray analysis in both seed and leaf tissues. The qRT-PCR expression values are over 100 fold higher than the microarray expression values. Pericarp enrichment of the lipid transfer protein transcript was confirmed by qRT-PCR.

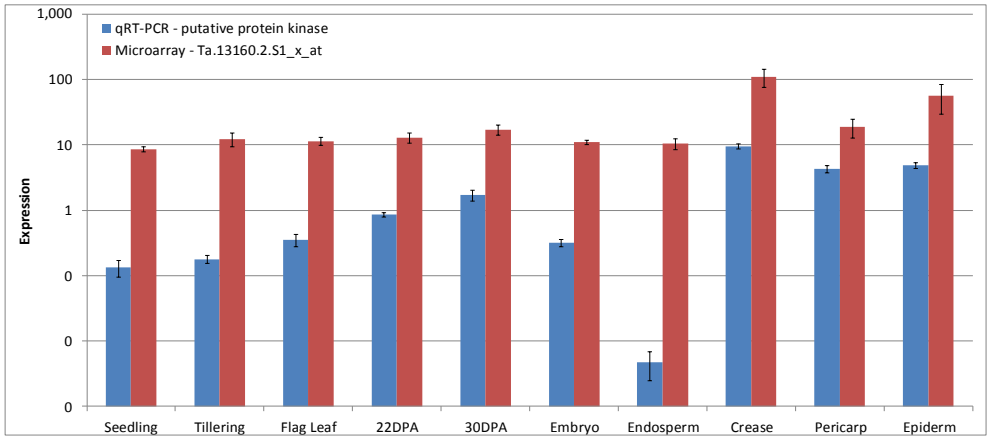
The probe set having an expression pattern consistent with down-regulation throughout development (Ta.590.2.S1_x_at) showed a strong decline in expression from each leaf developmental stage to the next in the microarray data (Figure 4.41). The smallest decline was from early senescence (22DPA) to late senescence (30DPA). The

Figure 4.40 Relative expression of candidate seed coat tissue-specific and enriched genes. A) Crease-specific probe set Ta.20064.1.S1_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST CA666265, a ubiquitin-conjugating enzyme transcript. B) Crease-enriched probe set Ta.13160.2.S1_x_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST CA635509, a putative protein kinase. C) Pericarp-enriched probe set Ta.30755.1.S1_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST CN007962, a nonspecific lipid-transfer protein precursor transcript. Values shown are mean expression values (microarray n = 5 for Flag Leaf, 2 for pericarp and epiderm and 3 for all other tissues, qRT-PCR n = 6 for all tissues, 3 technical replicates each for 2 biological replicates of each tissue). Error bars represent 1 standard deviation from the mean and have different lengths above and below the mean due to the log scale of the y axis. Expression is only shown for tissues that were assayed using qRT-PCR.

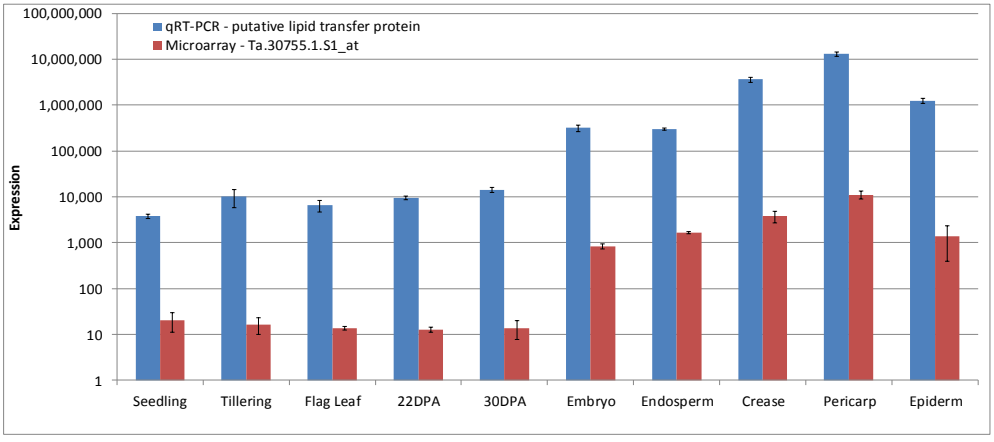
A


qRT-PCR - putative ubiquitin-conjugating...
Expression
10,000
1,000
100
10
1
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Embryo
Endosperm
Crease
Pericarp
Epiderm

B


qRT-PCR - putative protein kinase
Microarray - Ta.13160.2.S1_x_at
Expression
1,000
100
10
1
0
0
0
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Embryo
Endosperm
Crease
Pericarp
Epiderm

C


qRT-PCR - putative lipid transfer protein
Microarray - Ta.30755.1.S1_at
Expression
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
100
10
1
Seedling
Tillering
Flag Leaf
22DPA
30DPA
Embryo
Endosperm
Crease
Pericarp
Epiderm

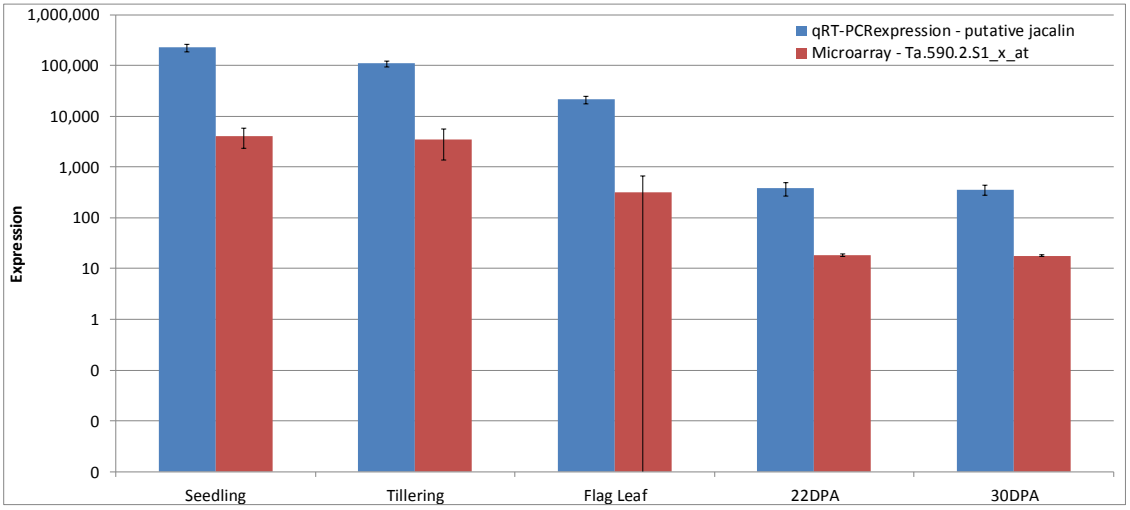


Figure 4.41 Relative expression of a candidate leaf developmentally-regulated gene. Down-regulated throughout development probe set Ta.590.2.S1_x_at expression data from the microarray experiment is shown in comparison to the qRT-PCR expression data of its representative sequence, EST CA601369, a high light protein, kda jasmonate-induced protein transcript. Values shown are mean expression values (microarray n = 5 for Flag Leaf and 3 for all other tissues, qRT-PCR n = 6 for all tissues, 3 technical replicates each for 2 biological replicates of each tissue). Error bars represent 1 standard deviation from the mean and have different lengths above and below the mean due to the log scale of the y axis. The large error bar shown for the Flag Leaf microarray data is due to a large variation among the replicates. Despite the large variation all Flag Leaf replicates had lower expression than all Seedling and Tillering replicates and higher expression than all 22DPA and 30DPA replicates. Expression is only shown for tissues that were assayed using qRT-PCR.

putative jacalin representative transcript of this probe set showed a very similar pattern of expression by qRT-PCR. The expression levels were underestimated by microarray analysis compared to qRT-PCR, however the levels of decline throughout development were very similar.

Genes chosen as being tissue-specific or tissue-enriched based on the microarray results showed high levels of up-regulation in the tissue of interest compared to all other tissues and this tissue-specific up-regulation was verified by qRT-PCR. In most cases the qRT-PCR showed higher levels of expression than the microarray results, especially when expression levels were particularly high, such as for the tissue-specific up regulation. The expression patterns of two transcripts by qRT-PCR were different from the microarray analysis results, the putative embryo-enriched glycine-rich protein transcript had lower expression than expected and the putative embryo-enriched wheat germ agglutinin transcript was not as specifically enriched as expected. However, most expression patterns were confirmed by qRT-PCR suggesting that the microarray results are valid.

4.4 DISCUSSION

4.4.1 A triticale expression atlas

Here I have presented a comprehensive analysis of transcript expression in triticale. The dataset provided serves as an 'expression atlas' for hexaploid triticale, the first of its kind. Combined with the transcriptional analysis of the reproductive tissues by my colleagues at ECORC, this comprehensive dataset may be used by other triticale or cereal researchers to explore transcript accumulation, timing, co-regulation and function. Tissues in this study were also grown under controlled and optimal growth conditions, allowing this expression atlas to serve as a baseline of comparison for transcript expression under biotic and abiotic stress conditions examined in other studies.

The Affymetrix GeneChip® Wheat Genome Array was the largest available resource for expression analysis in triticale at the time of this study. While the wheat GeneChip was not designed specifically for triticale expression analysis, the level of sequence homology between the cereals allowed its use to survey gene expression in triticale. We have provided a complete dataset of the triticale expression results across the entire wheat GeneChip allowing comparison to other studies using the wheat GeneChip and even cross-species comparisons to explore transcriptional conservation and differences, as is now becoming the focus among some plant researchers (Ma et al., 2005b; Krom and Ramakrishna, 2008; Schreiber et al., 2009). The use of the wheat GeneChip to survey triticale, the hybrid between wheat and rye, will allow comparison to wheat transcriptional patterns for the investigation of the effects of hybridization on gene expression among cereal species.

The transcriptional analysis presented here covers a broad range of tissues to more precisely identify tissue-specific expression and up-regulation. The tissues included are also the focus of academic and commercial interest. Leaf development is important to the yield of cereal plants since the leaves, and the flag leaf in particular, provide the majority of photosynthetic products that are transported to and stored in the caryopsis (Gong et al., 2005). Understanding gene expression patterns in senescing leaf tissues are of particular interest as it is during this stage when carbohydrates and proteins in the vegetative tissues are recycled and exported to sink organs such as the seed (Kichey et al., 2005). The timing and regulation of this process can affect the duration of crop productivity and ultimately yield. Stem tissues are receiving increased interest for their potential use in cellulosic bio-fuel production. Stem tissues are mainly a low value by-product of cereal production in which the grain is the primary valuable product. There is potential in converting this by-product to cellulosic bio-fuels, while adding value to the triticale crop. Understanding the changes in transcription that occur throughout stem development can lead to a better understanding of the processes that regulate changes in cell wall composition and structure, thereby allowing researchers to select targets for crop improvement for more efficient conversion of biomass to fuels through a change in cell wall composition or increased yield of biomass (Carroll and Somerville, 2009). Understanding regulation of root tissue development is important to agronomic aspects such as nitrogen use efficiency, water uptake, drought, salinity and mineral response to name a few. This study has provided a baseline for root gene expression that can be used for comparison when examining selected target genes involved in these processes in depth.

The caryopsis is the single major source of carbohydrates in the human diet, making cereals the most agriculturally and economically important crops. As well as a food source, seeds are used for feed, ethanol production and the production of a large number of industrial and commercial bio-products (Goddijn and Pen, 1995). An understanding of the processes involved in the development of the cereal caryopsis is therefore of significant agricultural and economic importance (Drea et al., 2005). This study has provided an atlas of gene expression within the seed in the middle of the grain filling period (soft dough stage) of development and can be used to select candidates for further study throughout caryopsis development.

Most angiosperm flowers consist of both male and female reproductive organs. The mechanism of development and differentiation of these organs and how they differ from each other and the vegetative plant tissues is of great interest to the plant research community. Understanding the gene regulation during sexual reproduction in plants has direct applications in agriculture such as the production of hybrid seed through male sterility (Endo et al., 2002). Comparison of the gene expression profiles from this gene expression study to studies in other plant species may allow the identification of regulatory factors that are common among the angiosperms, differ between dicotyledonous and monocotyledonous species and are unique to triticale and/or cereals. While the focus of this study was on transcript expression in vegetative and seed tissues, our colleagues at ECORC have provided a comprehensive analysis of the transcript expression in the reproductive tissues. We have also included the reproductive tissue expression data in our study for reference which can be queried to identify expression patterns of candidate transcripts for hypothesis driven research regarding these tissues.

Inclusion of the expression data from all of the major triticale tissue types allows this data set to be interrogated for patterns and levels of gene expression. This data set is provided as a reference that can be used to test specific biological hypotheses in triticale and among cereals and their progenitor species.

4.4.2 Specificity of the wheat GeneChip

One of the potential issues with this transcript expression analysis is the use of two different triticale cultivars. AC Certa and AC Alta are both spring triticale varieties developed by Agriculture and Agri-Food Canada Research Station at Swift Current (McLeod et al., 1996a; McLeod et al., 1996b). While there are some genetic differences between these varieties of triticale, it is expected that their differences are less than the difference between triticale and wheat when surveying global expression in triticale using the wheat GeneChip. Therefore, all samples types were normalized together to expand the range of tissues surveyed and to better define the transcriptional differences between tissues. Pearson correlation values reveal greater differences between coleoptile and root tissues of the same cultivar (AC Alta), than between coleoptile and seedling tissues of the two different cultivars (AC Alta and AC Certa respectively, Supplementary Table S4.3). This suggests that the tissue effect is greater that the cultivar effect and that inclusion of all tissues was the best approach to increasing the spatial resolution for the purposes of identifying tissue-specific gene expression patterns with the data available. Ideally, the same tissues would be sampled for both cultivars to determine the extent of the cultivar effect and potentially identify single feature polymorphisms (SFPs) or expression-level polymorphisms (ELPs) on the wheat GeneChip that could be used to distinguish between

cultivars (Coram et al., 2008c). The probe sets on the wheat GeneChip were designed to conserved regions of consensus sequences and it has been shown that the wheat GeneChip is unable to distinguish between contributions from individual homoeologues (Schreiber et al., 2009). Therefore, we expect that it would be difficult to separate the differences that are caused by the difference in cultivar. High throughput sequencing studies of both triticale varieties may allow the identification of sequence differences between these cultivars and features on the wheat array that will allow the separation of the contribution of each to the expression profiles.

The fact that the wheat GeneChip cannot distinguish the contributions from individual homeologues (Schreiber et al., 2009) may also explain the detectable expression differences between the microarray and the potentially more specific qRT-PCR. It has been shown that there is generally a weak reproducibility in cross-platform transcriptomics studies and this is further weakened in wheat due to the complexity of its allohexaploid genome (Poole et al., 2007). We expect this effect to be very similar in the comparison of microarray using the wheat GeneChip and qRT-PCR analysis of triticale transcripts due to its similarly complex hexaploid genome. qRT-PCR primers are designed to be specific to a particular transcript and therefore will differentiate between homeologues better than the GeneChip. This could account for some of the expression level differences observed between the microarray and qRT-PCR results when the qRT-PCR expression levels are lower in comparison to the microarray due to the fact that the microarray is possibly detecting expression from multiple homeologues. On the other hand, when the specificity is not improved by qRT-PCR, the increased sensitivity of

qRT-PCR (Valasek and Repa, 2005) in comparison to microarray could explain the cases where the qRT-PCR expression values are higher than the microarray.

4.4.3 Genes expressed in triticale

The number of detected genes in the tissues studied was represented by over 68% of the probe sets on the array. This is likely a small under-representation of the total genes expressed throughout the developmental life cycle of triticale grown under optimal growth conditions due to the lack of some cell types included in the study (i.e., crown tissue, late development root tissue, early and late development caryopsis tissues) and the fact that there may be some triticale transcripts from the rye genome and even the wheat genomes that are not represented on the array. Complete genome sequences for wheat, rye and triticale are not available to date due to their large size and complexity, and therefore the full complement of their transcriptomes has not yet been confirmed. The numbers of genes detected in each tissue ranged from 30.39% (endosperm) to 42.63% (stem45) of genes on the array with the exception of pollen in which only 20.08% of genes were detected. The lower detection rates in pollen, and even endosperm, are similar to a study in Arabidopsis in which expression has been tested across a wide range of tissue types (Schmid et al., 2005). Pollen has been shown to stand apart from other tissues in the number of detected transcripts and the overall expression profile (Schmid et al., 2005), a result which is supported by the PCA analysis performed in this study (Figure 4.8).

The number of probe sets showing significant binding to transcripts (majority present detection call) in any tissue in this study (41,918 probe sets) is higher than in

studies where a smaller range of tissues was surveyed (wheat caryopsis at 10 stages of development, Wan et al., 2008) as would be expected. Our ability to detect triticale transcripts using the wheat GeneChip is therefore comparable to other studies of cereal expression. In our study, 30,063 probe sets on the wheat GeneChip showed significant binding to transcripts in the triticale seed tissues sampled at only one stage of development, whereas approximately 33,000 probe sets showed significant binding to transcripts in a study that surveyed expression in wheat caryopsis at ten stages of development (Wan et al., 2008). We expect that if an increased number of seed developmental stages were sampled the number of detected triticale transcripts would increase to similar numbers or even exceed those previously detected in wheat.

While many genes were expressed in all or many tissues sampled, less than 1% were determined to be specific to any single tissue sample. More tissue-specific genes were found for root, embryo and endosperm tissues than for other tissue types which may reflect the more specialized nature of these tissues, a situation also identified in other plant transcriptome analyses (Schmid et al., 2005). It is predicted that the number of tissue-specific transcripts presented is larger than the number of cell-type specific transcripts that would be identified with an increased spatial resolution over which gene expression is monitored. Isolation of individual cell types using methods such as laser capture micro-dissection (LCM) or cell sorting with subsequent expression analysis by microarray, high-throughput sequencing or qRT-PCR can be used to improve the spatial resolution in cases where the current resolution is inadequate. For example, increased spatial resolution would be needed to identify transcripts that differentiate between leaf cell types (trichomes, stomata, mesophyll) or spatial differences within the endosperm

(starchy endosperm vs aleurone cells). However, the spatial resolution achieved in this study is substantial considering the breadth of tissues and developmental stages covered, particularly in the vegetative tissues. In wheat, transcriptomic studies have covered meiosis and microsporogenesis (Crismani et al., 2006), grain development (Wan et al., 2008), cold tolerance (Laudencia-Chingcuanco et al., 2011), salt tolerance (Mott and Wang, 2007), aluminum tolerance (Houde and Diallo, 2008), hessian fly resistance (Liu et al., 2007), rust fungi resistance (Hulbert et al., 2007; Coram et al., 2008a; Coram et al., 2008b) and non-host interactions to the barley pathogen *Ustilago hordei* (Gaudet et al., 2010) using the wheat GeneChip. While each of these studies has identified transcripts of interest in their respective systems, they have not provided sufficient analysis of global gene expression under normal growth conditions as a baseline for reference in other studies. Analysis of grain development (Gregersen et al., 2005), leaf senescence (Gregersen and Holm, 2007), dehydration response (Mohammadi et al., 2007), salt response (Kawaura et al., 2008) and temperature response (Gaudet et al., 2011) has also been performed using custom cDNA arrays. While the use of cDNA can be effective at identifying differentially regulated transcripts, the variation in array design can make cross-study comparisons difficult. Our study has encompassed many of the tissues sampled in the above mentioned studies, under optimal growth conditions and examined the expression using the wheat GeneChip, a platform that will allow comparison to other studies of cereal gene expression.

Recently an analysis of gene expression throughout barley development using the Affymetrix Barley1 GeneChip® was conducted (Druka et al., 2006) and repeated in wheat using the same tissue types and the wheat GeneChip for comparative analysis among the

Triticeae (Schreiber et al., 2009). While the wheat and barley analyses of tissue transcriptomics were the most comprehensive survey of global gene expression in the *Triticeae* species to date, this study includes more vegetative developmental tissue stages (stem and leaf stages), an increased number of reproductive tissues and developmental stages and separation of seed tissues. While there are several differences in tissues used in comparison to this study, there is also some similarity in the tissues chosen (root, coleoptiles, caryopsis and reproductive tissues) which may allow a combination and comparison of results in the future to expand the analysis of expression among the *Triticeae*.

The in-depth analysis to identify genes up-regulated in a tissue-specific manner (tissue-enriched gene expression) also revealed an increased number of putative tissue-specific transcripts. The method of analysis chosen to identify these transcripts (conversion of RMA values to zero in the case of absence of expression) is unique and described for the first time in this study. Some aspects of our analysis method have been used in similar studies but those studies have been more limited in terms of the strength or the depth of analysis. For example, in a global analysis of gene expression in Arabidopsis seed development, consensus detection calls were used to determine the presence and absence of expression in a given tissue (Le et al., 2010). This study was limited due to the fact that only two biological replicates were used for each sample. The use of three biological replicates for most sample types in this study allowed a more in depth analysis by allowing the identification of putative false positives (Figures 4.4 and 4.5) and gave more confidence in the consensus detection calls and increased statistical strength in determining expression variation. The consensus detection calls, which take

into account the mis-match hybridization, were used to determine transcript detection in each tissue type. I chose to combine the detection calls with the RMA normalized expression data to determine tissue-enriched transcript expression. Using this method, identification of those transcripts that were enriched in a particular tissue but still expressed at low levels was possible (Figure 4.5). Other studies have chosen arbitrary cutoffs to determine presence and absence of expression, such as two times the background expression (Druka et al., 2006) and expression values greater than or less than chosen values (i.e., gcRMA >6 = present and <4 = absent, Schmid et al., 2005). By using the consensus detection calls to determine presence of expression and converting expression values to zero in the case of absence of expression, we were able to identify transcripts that were expressed at low levels but still tissue-specific and tissue-enriched. Most of these transcripts with low level expression would not have been identified using arbitrary cutoff values.

Functional analysis of tissue-specific and tissue-enriched transcripts provided information on the molecular functions and biological processes involved in tissue development and differentiation. The information from the functional analysis can now be used to identify biological pathways important to tissue development and differentiation. This information can be used to investigate specific biological questions about the regulation of these pathways. Transcripts with similar levels and patterns of gene expression (identified by HCL of tissue-enriched transcripts for example) are often considered to be co-regulated (Druka et al., 2006). By identifying co-regulation between transcripts without functional information and those with functional annotations, novel candidate transcripts can be identified as being involved in the same regulatory process or

having a similar function. While further characterization studies will be needed to confirm predictions made from co-regulation analysis, strong indications of which candidates to interrogate can be made through a quick electronic examination of the data presented here prior to performing arduous experimentation to identify candidates.

Identification of developmentally-regulated transcript expression patterns proved informative in discovery of processes important to organ development such as leaf senescence. Clustering of the developmentally regulated transcripts revealed several patterns of transcript accumulation throughout leaf and stem development (Figures 4.30 and 4.31). While clustering the probe sets representing developmentally regulated transcripts revealed candidate genes that may be important to aspects of leaf and stem development, we have shown that the functional analysis of these clusters can reveal a large amount of information regarding a particular stage of development (Figures 4.32 to 4.34). Functional analysis improves upon the identification of candidate transcripts by identifying biological processes and molecular functions that may play important roles in developmental programs and may lead to the elucidation of regulatory networks. This type of analysis is made possible by using the improved annotation set created in Chapter 3. Previously, one of the problems in interpreting the biological meaning of expression data from wheat GeneChip experiments was a lack of functional annotations. By using the improved annotation set for the wheat GeneChip analysis we have been able to perform functional enrichment analysis of several gene lists to identify processes and functions important in individual tissues or stages of development. Previously, due to a lack of available detailed annotation data, many cereal transcriptomics studies have classified transcripts according to very general functional categories such as respiration,

photosynthesis, catabolism, stress etc. (Gregersen and Holm, 2007; Wan et al., 2008). Our improved annotation of the wheat GeneChip (Chapter 3) allows classification of transcripts with an increased depth of informative description as well as the comparison of lists of transcripts to identify statistically over and under-represented functional terms. Functional analysis of senescence associated transcripts was performed to highlight the potential of this type of analysis.

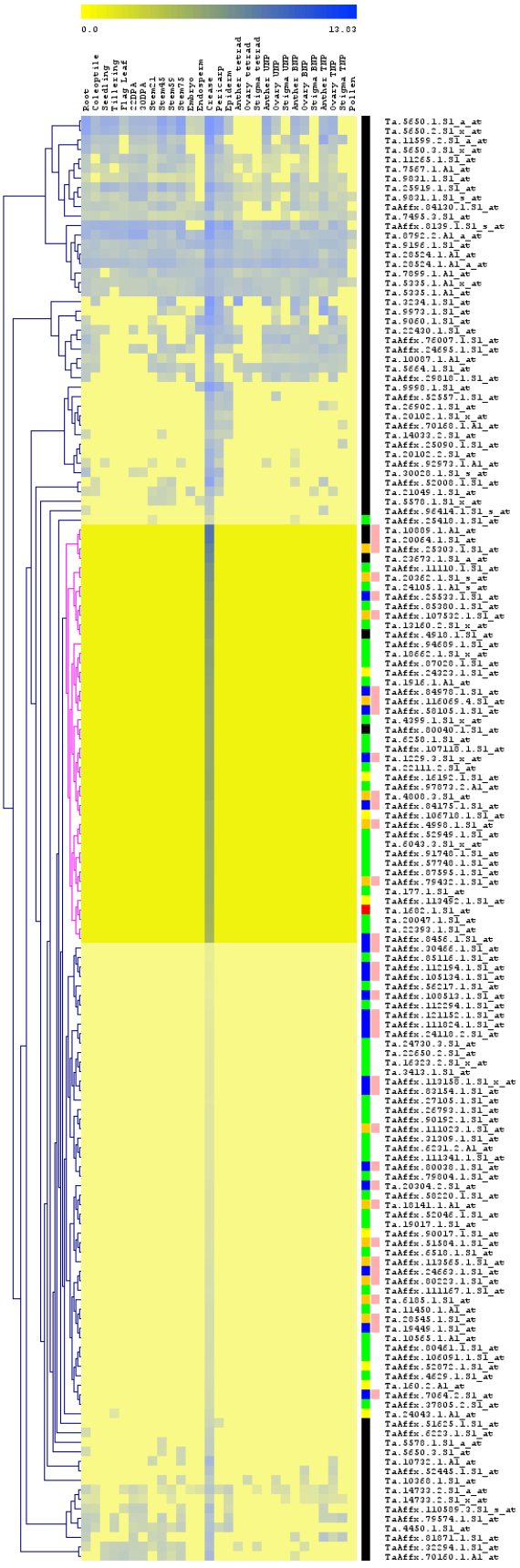
The senescence process is not a simple deterioration of plant tissues but a controlled process involving the activation of metabolic pathways that regulate the controlled degradation of cellular organelles and their components and mobilize these nutrients to other plant tissues. The timing, control, and efficiency of remobilization are important to the plant life cycle and of great interest to agriculture as this process affects grain yield (Gan and Amasino, 1997; Gong et al., 2005; Peñarrubia and Moreno, 2009; Guiboileau et al., 2010). Among the leaf developmentally regulated probe sets, at least six clusters (containing 1,240 of the 8,208 leaf DRG probe sets) had senescence associated expression patterns (Figure 4.28). Functional analysis revealed that while these probe sets were distributed in all major functional groups, they were enriched in certain functional terms. Enriched terms included stress response terms, transcription, cell death and degradation processes and enzyme activities (Figures 4.30-4.31). Enrichment of transcriptional activity has been shown in studies of senescence in wheat leaves (Gregersen and Holm, 2007) and Arabidopsis (Buchanan-Wollaston et al., 2003; Buchanan-Wollaston et al., 2005) suggesting that there may be some conservation among the regulatory pathways controlling senescence. Focusing on the transcripts involved in this transcriptional activity may lead to the understanding of the regulatory processes

controlling developmental senescence. The functional enrichment analysis of the SAGs has shown that this type of analysis provides meaning to the lists of transcripts with particular expression patterns in the context of a broader biological system.

4.4.4 Candidates for crop improvement

Through our transcript expression profiling we have identified a number of promising candidates for promoter characterization for the purposes of directed transgene expression studies. Ideally, promoters driving high levels of expression in a tissue-specific manner are needed for production of desired proteins in plants. The hierarchical clustering of the tissue-enriched probe set expression patterns (which include the tissue-specific probe sets) rapidly identifies transcripts with the desired expression patterns (Figure 4.19). From these clusters transcripts with tissue-specific expression and those that are highly expressed in the desired tissue can be visually separated from those that are expressed at more moderate or low levels. If the expression information is combined with the functional information from the annotations, the number of ideal candidates can be further reduced to those transcripts that are involved in a particular biological pathway. For example, in the crease-enriched transcripts (Figure 4.19), a cluster of 44 probe sets show crease-specific expression with relatively high levels of expression (Figure 4.42). If the biological processes of these probe sets are examined, there are 5 that are involved in transport (supplementary Table S4.9). These probe sets are representative of the following transcripts: a cytochrome b5; a transmembrane expressed citrate transporter family protein; a qltg3-1; a heavy metal transporter 3 zn cd p -type

Figure 4.42 Cluster of highly expressed crease-specific probe sets. Expression heat map of crease-enriched probe sets following HCL clustering in MeV. The cluster of highly expressed crease-specific probe sets is highlighted with a pink dendogram to the left while the remainder of the heat map has been faded and has a blue dendogram to the left to distinguish the cluster of interest. The dendogram to the left of the heat map shows the relationships of the probe sets following HCL clustering. The first vertical colored bar to the right of the heat map represents the lists from the Venn diagrams in Figure 16 that the probe sets belong to (green: list 3 only; blue: list 2 and 3; red: lists 1, 2 and 3; yellow: lists 3 and 4; orange: lists 2, 3, and 4; black: lists 1, 2, 3 and 4). The second vertical colored bar to the right of the heat map (pink) identifies probe sets that were also identified as tissue-specific (using the majority present detection calls in the tissue of interest and majority absent detection calls in all other tissues). To the right of the vertical colored bars are the IDs of the probe sets represented in the heat map. Mean RMA expression values following the conversion of these values to 0 when the majority of detection calls were not P (< 50% present) were used to construct the heat map. Each row represents the probe set expression values across the tissues studied and each column represents one of the tissue types studied. The horizontal color scale at the top of the heat map represents the expression values with yellow representing no expression and dark blue representing maximum expression. The mean expression values for each probe set in a particular tissue are represented by the intensity of color (prior to fading of probe sets not in the cluster of interest).


0.0
13.83
Root
Coleoptile
Seedling
Flag leaf
Flag leaf
Flag leaf
22DPA
30DPA
Stem21
Stem45
Stem55
Stem75
Rachis
Endosperm
Crease
Pericarp
Nucellus
Anther tetrad
Ovary tetrad
Stigma tetrad
Anther UNP
Ovary UNP
Stigma UNP
Anther BMP
Ovary BMP
Stigma BMP
Anther TMP
Ovary TMP
Stigma TMP
Pollen

ATPase and a peptide transporter protein (TaAffx.84978.1.S1_at, TaAffx.107118.1.S1_at, TaAffx.8456.1.S1_at, TaAffx.4998.1.S1_at, TaAffx.107532.1.S1_at, respectively). This demonstrates the usefulness of our data set and how with the information provided, researchers can quickly reduce the number of candidates for further investigation to those with the desired expression pattern and involved in the desired process or function.

Promoters of genes that are expressed in leaf tissues throughout the development of the plant may be useful for improving a number of agronomic traits. Alternatively, promoters of genes that are expressed in seed tissues and not in leaf tissues may be useful for the expression of proteins, lipids or starches to be extracted from the harvested grains. Utilization of a leaf-specific promoter in conjunction with a seed-specific promoter for expression of two products has the potential for extraction of one product from grain while the remainder of the biomass is used for the extraction of the other product, thus maximizing the potential economic return on investment. A number of probe sets were identified as being leaf-specific (489) and seed-specific (947). For use of the entire above ground biomass, the extensive aerial vegetative specific probe set list (1,723) can be considered for candidate selection. This provides a large set of candidate genes to choose from for isolating vegetative or seed promoters. This is a higher number of potential candidates than is presented in the expression analysis of other cereal crops where only single tissue-specific or specifically up-regulated transcripts are considered and the transcripts specific to the above-ground vegetative plant tissues are not considered as a whole (Druka et al., 2006; Schreiber et al., 2009).

The genes associated with the identified developmentally-regulated probe sets are interesting for crop improvement purposes. For example, promoters of these genes may be useful for targeting plant defense against diseases that increase or decrease in risk and pathogenicity throughout plant development or for increasing or decreasing the expression of a particular product either early or late in development when it is not normally highly expressed. For instance, genes that increase to their maximum levels of expression at the early boot developmental stage then decrease throughout further development may be desired for altering or improving the storage of photosynthates in leaf tissues.

A representative sequence of a probe set displaying decreasing expression throughout development, a putative jacalin, was chosen for further characterization via qRT-PCR (Figure 4.41). Comparison to sequences in UniGene with protein references revealed only weak similarity (49.7%) to a rice protein (Os05g0143600) containing a jacalin-like lectin domain. Proteins containing this domain are lectins. Plant lectins are carbohydrate-binding proteins and jacalin-related lectins are those that are structurally and evolutionary related to the jack-fruit lectin and have an affinity for the disaccharide β-galactose-1,3-N-acetylgalactosamine (Galβ(1,3)GalNAc, Van Damme et al., 1998). The similarity is in the region of the jacalin domain and suggests that the putative jacalin has carbohydrate-binding function. This gene showed declining expression from each stage of development to the next by microarray and qRT-PCR expression values. Expression of this transcript was detected in triticale by the microarray analysis (majority P detection calls) in the coleoptile, seedling, tillering, flag leaf, stem21, stem45, stem59, anther UNP, stigma BNP and stigma TNP samples. This sequence was also found in

stem, leaf and flower and mixed tissues of wheat using UniGene's EST ProfileViewer (http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi). While the expression was found in similar tissues in our study and the UniGene EST ProfileViewer, our study has improved the spatial resolution of the expression of this transcript as well as determining the level of expression in each tissue and throughout development by examining a larger number of tissue types. This result emphasizes the importance of the type of global expression analysis carried out in this study to provide a baseline expression profile.

In addition to choosing candidates for promoter characterization for transgene expression purposes, researchers are particularly interested in dissecting the regulatory networks that control plant developmental processes. Transcripts that are unique to particular tissues or developmental stages are predicted to play roles in the differentiation and development of these tissues (Le et al., 2007; Le et al., 2010). Of particular interest, will be the transcription factors that are unique to each of the tissues, since they are the key regulators of many developmental processes. By combining the expression profiles from this study with the gene ontology information from the previous chapter (Chapter 3), probe sets representing transcription factor transcripts can easily be identified. For example, if the seed-specific transcripts are sorted for those with the ontology transcription factor activity, the 947 probe sets can be reduced to a set of 40 probe sets representing transcripts with transcription factor activity (Table 4.8). From this analysis we observe that seed-specific transcription factors are diverse and include NAC, NAM, WRKY, Zinc finger and AP2 domain containing transcription factors. If the list of searched ontologies is expanded to transcription factor activity; transcription regulator

Table 4.8 Seed-specific probe sets with transcription factor activity gene ontology and their sequence descriptions.

Probe set_ID	Seq. Description
Ta.10054.1.S1_at	ABA response element binding factor
Ta.10834.1.A1_at	ABA response element binding factor
Ta.10852.1.S1_at	AP2 domain containing expressed
Ta.10948.1.A1_at	DNA binding
Ta.1272.1.S1_at	---NA---
Ta.13257.2.S1_at	DNA binding
Ta.13257.2.S1_x_at	DNA binding; protein
Ta.17523.1.S1_at	Box-like protein
Ta.24093.1.S1_at	HAP3 transcriptional-activator
Ta.25258.1.S1_at	NAC domain protein NAC6
Ta.27782.1.S1_at	Protein spatula
Ta.8284.2.A1_at	---NA---
Ta.8284.2.A1_x_at	---NA---
Ta.893.1.S1_at	Storage protein activator
TaAffx.106600.1.S1_at	Transcription Factor 1
TaAffx.106851.1.S1_at	DNA binding; homeobox-like resistance
TaAffx.113415.1.S1_at	cbf-like transcription factor
TaAffx.116069.1.S1_at	DOF Zinc finger protein 3
TaAffx.117198.1.S1_at	ER33 protein
TaAffx.117376.1.S1_at	NAC domain-containing protein 21; NAC domain ipr003441
TaAffx.117676.1.S1_at	NAC domain protein NAC6; NAC domain-containing protein 21
TaAffx.122144.1.S1_at	KNOX family class 2 homeodomain protein
TaAffx.123327.1.S1_x_at	Ethylene-Responsive factor-like protein 1
TaAffx.128544.1.S1_at	Zinc finger
TaAffx.128544.2.S1_x_at	Zinc finger
TaAffx.128733.1.S1_x_at	AtMYB4-like protein; MYB transcription factor
TaAffx.16192.1.S1_at	AP2 erf domain-containing transcription factor
TaAffx.32069.2.S1_at	AP2-domain dna-binding
TaAffx.37139.1.S1_at	ANAC079 ANAC080 AtNAC4
TaAffx.5112.1.S1_at	WRKY23 - superfamily of TFs having WRKY and Zinc finger domains
TaAffx.52152.1.S1_at	NAM (no apical meristem)-like protein
TaAffx.52554.1.S1_at	GL2-type homeodomain protein
TaAffx.52812.1.S1_at	Protein spatula
TaAffx.52812.1.S1_s_at	Protein spatula
TaAffx.56048.1.S1_at	Hypothetical protein OsJ_02965; Homeobox transcription
TaAffx.6787.1.S1_at	WRKY7 - superfamily of tfs having WRKY and Zinc finger domains
TaAffx.71942.1.A1_at	Retrotransposon unclassified
TaAffx.78436.1.S1_at	AP2 erf domain-containing transcription factor
TaAffx.79432.1.S1_at	bZIP transcription factor
TaAffx.93080.1.S1_at	GATA transcription factor 25

activity; regulation of transcription; transcription activator activity; negative regulation of transcription; regulation of transcription, DNA-dependent; sequence specific DNA binding; and DNA binding, the list is reduced from 947 to 82 probe sets (Supplementary Table S4.10). This method quickly reduces the list of transcripts on which to focus when investigating the regulation of seed-specific developmental processes. Relatively few transcription factors have been characterized in cereals particularly in terms of their target genes and biological roles (Wan et al., 2008). Combining our improved annotation data (Chapter 3) with our expression analysis data allows the identification of putative novel transcription factors and the localization of their expression within tissues and throughout development which may allow characterization of an increased number of transcription factors important to plant development and tissue differentiation. Our dataset allows researchers to quickly identify the candidate transcripts for further characterization to uncover regulatory networks in triticale and related cereals.

4.4.5 Future Directions

While this study is the most comprehensive transcriptional analysis in triticale to date, future experiments would be valuable additions to the current data set. Additional tissues that would add valuable information to the data set include triticale caryopsis at different stages of development to identify developmentally important transcripts in fertilization, seed development and seed tissue differentiation and root tissue at different stages of development to identify developmentally important transcripts. In addition to completing the expression atlas by sampling an increased number of tissue types, expression analysis of the same tissues under various environmental conditions and their

comparison to the baseline expression in this study would identify transcripts important in environmental response. Investigation of these transcripts should lead to the identification of transcription factors and the elucidation of biological pathways and regulatory networks that control these processes.

High throughput sequencing is quickly replacing microarray analysis as a global transcriptional analysis method due to decreasing costs and the large amount of resulting data. If gene expression in additional tissue types, developmental stages and environmental conditions are surveyed it is likely that high-throughput sequencing will be the method of choice. Creating a method to relate expression results from sequencing to GeneChip expression results will allow their combination to create a comprehensive data set without the need to re-sample tissues that have already been analyzed by microarray.

The tissues included in this study are complex and contain a number of different cell types that are individually specialized to perform a range of functions. For example, leaf tissue includes specialized cell types, such as stomata guard cells, trichomes, epidermis, mesophyll and vascular cells. The transcript expression profiles we have identified are an average across all component cell types within a sample type. To identify the precise cellular specificity and abundance of expression among the cell types contained within a tissue sample, it is recommended that further studies, such as *in situ* hybridizations or qRT-PCR of laser capture micro-dissected cell types, are carried out for individual genes to improve the spatial resolution of expression patterns and to address specific biological questions.

Following the precise determination of transcript expression profiles, an important next step will be the determination of protein abundance and the identification interactions in each tissue to correlate the transcript expression data with protein function and plant development.

4.4.6 Conclusion

This study of the triticale transcriptome has provided a gene expression atlas that includes all major tissue types and several developmental stages in hexaploid triticale. We have also performed a comprehensive analysis of this data to identify transcripts that are expressed in a tissue-specific, tissue-enriched, and developmentally regulated manner. As well as identifying transcripts with specific expression patterns, functional analysis of these transcripts was performed to determine which biological processes and molecular functions within each tissue and at different stages of development. This analysis allowed the identification of tissue-specific transcription factors and the biological processes and molecular functions involved in developmental processes such as leaf senescence. While much of the results of our analysis are presented within, we have also provided our data set as a reference for queries of transcript expression, accumulation, timing, function and co-regulation by researchers within the triticale, cereal, and plant communities. We expect that our broad coverage of transcript expression throughout triticale development will further the understanding of the regulation of triticale development and tissue differentiation as well as the evolution and conservation of these processes when compared to transcriptional profiling studies in other plant species.

4.5 LITERATURE CITED

Altenbach, S.B., and Kothari, K.M. (2004). Transcript profiles of genes expressed in endosperm tissue are altered by high temperature during wheat grain development. Journal of Cereal Science **40,** 115-126.

Andersson, A., Keskitalo, J., Sjodin, A., Bhalerao, R., Sterky, F., Wissel, K., Tandre, K., Aspeborg, H., Moyle, R., Ohmiya, Y., Bhalerao, R., Brunner, A., Gustafsson, P., Karlsson, J., Lundeberg, J., Nilsson, O., Sandberg, G., Strauss, S., Sundberg, B., Uhlen, M., Jansson, S., and Nilsson, P. (2004). A transcriptional timetable of autumn senescence. Genome Biology **5,** R24.

Becerra, C., Puigdomenech, P., and Vicient, C. (2006). Computational and experimental analysis identifies Arabidopsis genes specifically expressed during early seed development. BMC Genomics **7,** 38.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y., and Udvardi, M.K. (2008). A gene expression atlas of the model legume *Medicago truncatula*. The Plant Journal **55,** 504-513.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B **57,** 289-300.

Blazejczyk, M., Miron, M., and Nadon, R. (2007). FlexArray: A statistical data analysis software for gene expression microarrays. Genome Quebec, Montreal, Canada, URL http://genomequebec.mcgill.ca/FlexArray

Boodley, J.W., and Sheldrake Jr., R. (1977). Cornell peat-lite mixes for commercial plant growing. New York State College Agricultural Life Science Information Bulletin **43,** 8.

Buchanan-Wollaston, V., Earl, S., Harrison, E., Mathas, E., Navabpour, S., Page, T., and Pink, D. (2003). The molecular analysis of leaf senescence - a genomics approach. Plant Biotechnology Journal **1,** 3-22.

Buchanan-Wollaston, V., Page, T., Harrison, E., Breeze, E., Lim, P.O., Nam, H.G., Lin, J.-F., Wu, S.-H., Swidzinski, J., Ishizaki, K., and Leaver, C.J. (2005). Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. The Plant Journal **42,** 567-585.

Canadian Biotechnology Advisory Committee (2002). Improving the regulation of genetically modified foods and other novel foods in Canada. Report to the Government of Canada Biotechnology Ministerial Coordinating Committee. (Ottawa ON: Candadian Biotechnology Advisory Committee).

Carroll, A., and Somerville, C. (2009). Cellulosic Biofuels. Annual Review of Plant Biology **60,** 165-182.

Conesa, A., and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics **2008,** 1-12.

Coram, T.E., Settles, M.L., and Chen, X. (2008a). Transcriptome analysis of high-temperature adult-plant resistance conditioned by *Yr39* during the wheat–*Puccinia striiformis* f. sp. *tritici* interaction. Molecular Plant Pathology **9,** 479-493.

Coram, T.E., Wang, M., and Chen, X. (2008b). Transcriptome analysis of the wheat–*Puccinia striiformis* f. sp. *tritici* interaction. Molecular Plant Pathology **9,** 157-169.

Coram, T.E., Brown-Guedira, G., and Chen, X. (2008c). Using transcriptomics to understand the wheat genome. CAB reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources **3**.

Crismani, W., Baumann, U., Sutton, T., Shirley, N., Webster, T., Spangenberg, G., Langridge, P., and Able, J. (2006). Microarray expression analysis of meiosis and microsporogenesis in hexaploid bread wheat. BMC Genomics **7,** 267.

Debeaujon, I., Leon-Kloosterziel, K.M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. Plant Physiology **122,** 403-414.

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., and Caboche, M. (2003). Proanthocyanidin-accumulating cell in Arabidopsis testa: regulation of differentiation and role in seed development. The Plant Cell **15,** 2514-2531.

Draghici, S., Khatri, P., Eklund, A.C., and Szallasi, Z. (2006). Reliability and reproducibility issues in DNA microarray measurements. Trends in Genetics **22,** 101-109.

Drea, S., Leader, D.J., Arnold, B.C., Shaw, P., Dolan, L., and Doonan, J.H. (2005). Systematic spatial analysis of gene expression during wheat caryopsis development. The Plant Cell **17,** 2172-2185.

Druka, A., Muehlbauer, G., Druka, I., Caldo, R., Baumann, U., Rostoks, N., Schreiber, A., Wise, R., Close, T., Kleinhofs, A., Graner, A., Schulman, A., Langridge, P., Sato, K., Hayes, P., McNicol, J., Marshall, D., and Waugh, R. (2006). An atlas of gene expression from seed to seed through barley development. Functional & Integrative Genomics **6,** 202-211.

Endo, M., Matsubara, H., Kokubun, T., Masuko, H., Takahata, Y., Tsuchiya, T., Fukuda, H., Demura, T., and Watanabe, M. (2002). The advantages of cDNA microarray as an effective tool for identification of reproductive organ-specific genes in a model legume, *Lotus japonicus*. FEBS Letters **514,** 229-237.

Fisher, R.A. (1922). On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. Journal of the Royal Statistical Society **85,** 87-94.

Furutani, I., Sukegawa, S., and Kyozuka, J. (2006). Genome-wide analysis of spatial and temporal gene expression in rice panicle development. The Plant Journal **46,** 503-511.

Galbraith, D.W., and Birnbaum, K. (2006). Global studies of cell type-specific gene expression in plants. Annual Review of Plant Biology **57,** 451-475.

Gan, S., and Amasino, R.M. (1997). Making sense of senescence (molecular genetic regulation and manipulation of leaf senescence). Plant Physiology **113,** 313-319.

Gaudet, D.A., Wang, Y., Penniket, C., Lu, Z.X., Bakkeren, G., and Laroche, A. (2010). Morphological and molecular analyses of host and nonhost interactions involving barley and wheat and the covered smut pathogen *Ustilago hordei*. Molecular Plant-Microbe Interactions **23,** 1619-1634.

Gaudet, D.A., Wang, Y., Frick, M., Puchalski, B., Penniket, C., Ouellet, T., Robert, L., Singh, J., and Laroche, A. (2011). Low temperature induced defence gene expression in winter wheat in relation to resistance to snow moulds and other wheat diseases. Plant Science **180,** 99-110.

The Gene Ontology Consortium, Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics **25,** 25-29.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C., and Ohlrogge, J. (2000). Microarray analysis of developing Arabidopsis seeds. Plant Physiology **124,** 1570-1581.

Giunta, F., Motzo, R., and Deidda, M. (1999). Grain yield analysis of a triticale (× *Triticosecale* Wittmack) collection grown in a Mediterranean environment. Field Crops Research **63,** 199-210.

Goddijn, O.J.M., and Pen, J. (1995). Plants as bioreactors. Trends in Biotechnology **13,** 379-387.

Gong, Y.H., Zhang, J., Gao, J.F., Lu, J.Y., and Wang, J.R. (2005). Slow export of photoassimilate from stay-green leaves during late grain-filling stage in hybrid winter wheat (*Triticum aestivum* L.). Journal of Agronomy and Crop Science **191,** 292-299.

Gregersen, P., Brinch-Pedersen, H., and Holm, P. (2005). A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. Transgenic Research **14,** 887-905.

Gregersen, P.L., and Holm, P.B. (2007). Transcriptome analysis of senescence in the flag leaf of wheat (*Triticum aestivum* L.). Plant Biotechnology Journal **5,** 192-206.

Gregersen, P.L., Holm, P.B., and Krupinska, K. (2008). Leaf senescence and nutrient remobilisation in barley and wheat. Plant Biology **10,** 37-49.

Guiboileau, A., Sormani, R., Meyer, C., and Masclaux-Daubresse, C. (2010). Senescence and death of plant organs: Nutrient recycling and developmental regulation. Comptes Rendus Biologies **333,** 382-391.

Guo, Y., Cai, Z., and Gan, S. (2004). Transcriptome of Arabidopsis leaf senescence. Plant, Cell & Environment **27,** 521-549.

He, Y., and Gan, S. (2002). A gene encoding an acyl hydrolase is involved in leaf senescence in Arabidopsis. The Plant Cell Online **14,** 805-815.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biology **8,** R19.

Herrero, J., Valencia, A., and Dopazo, J.n. (2001). A hierarchical unsupervised growing neural network for clustering gene expression patterns. Bioinformatics **17,** 126-136.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biology **5,** 1-13.

Hopkins, M., McNamara, L., Taylor, C., Wang, T.-W., and Thompson, J. (2007). Membrane dynamics and regulation of subcellular changes during senescence. In Senescence processes in plants, S. Gan, ed (Ames, Iowa: Blackwell, Pub.), pp. 39-68.

Horlein, A., and Valentine, J. (1995). Triticale (× *Triticosecale*). In Cereals and psuedocereals, W. JT, ed (London: Chapman and Hall), pp. 187-221.

Houde, M., and Diallo, A. (2008). Identification of genes and pathways associated with aluminum stress and tolerance using transcriptome profiling of wheat near-isogenic lines. BMC Genomics **9,** 400.

Hulbert, S.H., Bai, J., Fellers, J.P., Pacheco, M.G., and Bowden, R.L. (2007). Gene expression patterns in near isogenic lines for wheat rust resistance gene *Lr34/Yr18*. Phytopathology **97,** 1083-1093.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Research **31,** e15.

Kawashima, T., Wang, X., Henry, K.F., Bi, Y., Weterings, K., and Goldberg, R.B. (2009). Identification of *cis*-regulatory sequences that activate transcription in the suspensor of plant embryos. Proceedings of the National Academy of Sciences **106,** 3627-3632.

Kawaura, K., Mochida, K., and Ogihara, Y. (2008). Genome-wide analysis for identification of salt-responsive genes in common wheat. Functional & Integrative Genomics **8,** 277-286.

Keskitalo, J., Bergquist, G., Gardeström, P., and Jansson, S. (2005). A cellular timetable of autumn senescence. Plant Physiology **139,** 1635-1648.

Kichey, T., Le Gouis, J., Sangwan, B., Hirel, B., and Dubois, F. (2005). Changes in the cellular and subcellular localization of glutamine synthetase and glutamate dehydrogenase during flag leaf senescence in Wheat (*Triticum aestivum* L.). Plant Cell Physiology **46,** 964-974.

Kondou, H., Ooka, H., Yamada, H., Satoh, K., Kikuchi, S., Takahara, Y., and Yamamoto, K. (2006). Microarray analysis of gene expression at initial stages of rice seed development. Breeding Science **56,** 235-242.

Kreps, J.A., Wu, Y., Chang, H.-S., Zhu, T., Wang, X., and Harper, J.F. (2002). Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. Plant Physiology **130,** 2129-2141.

Krom, N., and Ramakrishna, W. (2008). Comparative analysis of divergent and convergent gene pairs and their expression patterns in Rice, Arabidopsis, and *Populus*. Plant Physiology **147,** 1763-1773.

Krupinska, K., and Humbeck, K. (2008). Senescence processes and their regulation. Plant Biology **10,** 1-3.

Kudryakova, N. (2009). Leaf senescence and gene expression. CAB reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources **4**.

Large, E.C. (1954). Growth stages in cereals illustration of the Feekes scale. Plant Pathology **3,** 128-129.

Laudencia-Chingcuanco, D., Ganeshan, S., You, F., Fowler, B., Chibbar, R., and Anderson, O. (2011). Genome-wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (*Triticum aestivum* L.). BMC Genomics **12,** 299.

Le, B.H., Wagmaister, J.A., Kawashima, T., Bui, A.Q., Harada, J.J., and Goldberg, R.B. (2007). Using genomics to study legume seed development. Plant Physiology **144,** 562-574.

Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G.N., Fischer, R.L., Okamuro, J.K., Harada, J.J., and Goldberg, R.B. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences of the United States of America **107,** 8086-8070.

Lee, R.-H., Lin, M.-C., and Chen, S.-C. (2004). A novel alkaline α-galactosidase gene is involved in rice leaf senescence. Plant Molecular Biology **55,** 281-295.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., and Caboche, M. (2006). Enetics and biochemistry of seed flavenoids. Annual Review of Plant Biology **57,** 405-430.

Lim, P.O., Nam, H.G., and Gerald, P.S. (2005). The molecular and genetic control of leaf senescence and longevity in Arabidopsis. Current Topics in Developmental Biology **67,** 49-83.

Lim, P.O., Kim, H.J., and Gil Nam, H. (2007). Leaf senescence. Annual Review of Plant Biology **58,** 115-136.

Lin, J.-F., and Wu, S.-H. (2004). Molecular events in senescing Arabidopsis leaves. Plant Journal **39,** 612-628.

Liu, X., Bai, J., Huang, L., Zhu, L., Liu, X., Weng, N., Reese, J., Harris, M., Stuart, J., and Chen, M.-S. (2007). Gene expression of different wheat genotypes during attack by virulent and avirulent hessian fly (*Mayetiola destructor*) larvae. Journal of Chemical Ecology **33,** 2171-2194.

Lu, X.C., Gong, H.Q., Huang, M.L., Bai, S.L., He, Y.B., Mao, X.Z., Geng, Z., Li, S.G., Wei, L.P., Yuwen, J.S., Xu, Z.H., and Bai, S.N. (2006). Molecular analysis of early rice stamen development using organ-specific gene expression profiling. Plant Molecular Biology **61,** 845-861.

Ma, L., Sun, N., Liu, X., Jiao, Y., Zhao, H., and Deng, X.W. (2005a). Organ-specific expression of Arabidopsis genome during development. Plant Physiology **138,** 80-91.

Ma, L.G., Chen, C., Liu, X.G., Jiao, Y.L., Su, N., Li, L., Wang, X.F., Cao, M.L., Sun, N., Zhang, X.Q., Bao, J.Y., Li, J., Pedersen, S., Bolund, L., Zhao, H.Y., Yuan, L.P., Wong, G.K.S., Wang, J., Deng, X.W., and Wang, J. (2005b). A microarray analysis of the rice transcriptome and its comparison to Arabidopsis. Genome Research **15,** 1274-1283.

McLeod, J.G., DePauw, R.M., Clarke, J.M., and Pfeiffer, W.H. (1996a). AC Certa spring triticale. Canadian Journal of Plant Science **76,** 333-335.

McLeod, J.G., DePauw, R.M., Clarke, J.M., and Townley-Smith, T.F. (1996b). AC Alta spring triticale. Canadian Journal of Plant Science **76,** 139-141.

Mohammadi, M., Kav, N.N., and Deyholos, M.K. (2007). Transcriptional profiling of hexaploid wheat (*Triticum aestivum* L.) roots identifies novel, dehydration-responsive genes. Plant, Cell & Environment **30,** 630-645.

Mott, I.W., and Wang, R.R.C. (2007). Comparative transcriptome analysis of salt-tolerant wheat germplasm lines using wheat genome arrays. Plant Science **173,** 327-339.

Peñarrubia, L., and Moreno, J. (2009). Senescence in plants and crops. In Handbook of Plant and Crop Physiology (CRC Press).

Poole, R., Barker, G., Wilson, I., Coghill, J., and Edwards, K. (2007). Measuring global gene expression in polyploidy; a cautionary note from allohexaploid wheat. Functional & Integrative Genomics **7,** 207-219.

Quackenbush, J. (2005). Using DNA microarrays to assay gene expression. In Bioinformatics: A practical guide to the analysis of genes and proteins, Third edition, A.D. Baxevanis and B.F.F. Ouellette, eds (Hoboken, New Jersey: John Wiley & Sons, Inc).

Rabbani, M.A., Maruyama, K., Abe, H., Khan, M.A., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M., Shinozaki, K., and 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 Physiology **133,** 1755-1767.

Repellin, A., Båga, M., Jauhar, P., and Chibbar, R. (2001). Genetic enrichment of cereal crops via alien gene transfer: New challenges. Plant Cell, Tissue and Organ Culture **64,** 159-183.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In Methods in Molecular Biology: Bioinformatics Methods and Protocols, S.A. Krawetz and S. Misener, eds (Totowa, NJ: Humana Press Inc.), pp. 365-386.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, A., Vinsavich, A., Trush, V., and Quackenbush, J. (2003). TM4: A free, open-source system for microarray data management and analysis. Biotechniques **34,** 374-378.

Salmon, D.F. (2004). Production of triticale on the Canadian Prairies. In Triticale improvement and production. FAO plant production and protection paper 179, M. Mergoum and H. Gomez-Macpherson, eds (Rome: Food and Agriculture Organization of the United Nations), pp. 99-102.

Schelbert, S., Aubry, S., Burla, B., Agne, B., Kessler, F., Krupinska, K., and Hörtensteiner, S. (2009). Pheophytin pheophorbide hydrolase (Pheophytinase) is involved in chlorophyll breakdown during leaf senescence in *Arabidopsis*. The Plant Cell **21,** 767-785.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. Nature Genetics **37,** 501-506.

Schreiber, A., Sutton, T., Caldo, R., Kalashyan, E., Lovell, B., Mayo, G., Muehlbauer, G., Druka, A., Waugh, R., Wise, R., Langridge, P., and Baumann, U. (2009). Comparative transcriptomics in the *Triticeae*. BMC Genomics **10,** 285.

Singh, G., Kumar, S., and Singh, P. (2003). A quick method to isolate RNA from wheat and other carbohydrate-rich seeds. Plant Molecular Biology Reporter **21,** 93a-93f.

Taiz, L., and Zeiger, E. (2002). Plant Physiology. (Sunderland, Massachusetts: Sinauer Associates, Inc.).

Takaiwa, F. (2004). Health-promoting transgenic rice suppressing life-related disease and type-I allergy. In Rice is Life: scientific perspectives for the 21st century. Proceedings of the World Rice Research Conference (Tsukuba, Japan).

Valasek, M.A., and Repa, J.J. (2005). The power of real-time PCR. Advances in Physiology Education **29,** 151-159.

Van Damme, E.J.M., Peumans, W.J., Barre, A.J., and Rouge, P.J. (1998). Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Critical Reviews in Plant Sciences **17,** 575-692.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology **3,** research0034.1-0034.11.

Wan, Y., Poole, R., Huttly, A., Toscano-Underwood, C., Feeney, K., Welham, S., Gooding, M., Mills, C., Edwards, K., Shewry, P., and Mitchell, R. (2008). Transcriptome analysis of grain development in hexaploid wheat. BMC Genomics **9,** 121.

Wang, L., Xie, W., Chen, Y., Tang, W., Yang, J., Ye, R., Liu, L., Lin, Y., Xu, C., Xiao, J., and Zhang, Q. (2010). A dynamic gene expression atlas covering the entire life cycle of rice. The Plant Journal **61,** 752-766.

Wellmer, F., Riechmann, J.L., Alves-Ferreira, M., and Meyerowitz, E.M. (2004). Genome-wide analysis of spatial gene expression in Arabidopsis flowers. The Plant Cell **16,** 1314-1326.

Wilson, I.D., Barker, G.L.A., Beswick, R.W., Shepherd, S.K., Lu, C., Coghill, J.A., Edwards, D., Owen, P., Lyons, R., Parker, J.S., Lenton, J.R., Holdsworth, M.J., Shewry, P.R., and Edwards, K.J. (2004). A transcriptomics resource for wheat functional genomics. Plant Biotechnology Journal **2,** 495-506.

Yamakawa, H., Hirose, T., Kuroda, M., and Yamaguchi, T. (2007). Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. Plant and Cell Physiology **48,** S62-S62.

Zadoks, J.C., Chang, T.T., and Konzak, C.F. (1974). A decimal code for the growth stages of cereals. Weed Research **14,** 415-421.

Zhou, J., Wang, X., Jiao, Y., Qin, Y., Liu, X., He, K., Chen, C., Ma, L., Wang, J., Xiong, L., Zhang, Q., Fan, L., and Deng, X. (2007). Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. Plant Molecular Biology **63,** 591-608.

CHAPTER 5 Localization of triticale transcripts using *in situ* hybridization

5.1 INTRODUCTION

The cereal caryopsis is the product of a double fertilization event that results in a single seeded fruit that consists of the embryo and the carbohydrate rich endosperm which are surrounded by tissues layers that become the seed coat. The cereal caryopsis is the primary source of carbohydrate in the human diet, necessitating an understanding of its development to design crop improvement strategies. Embryogenesis also establishes the essential features of the mature plant; therefore a profound understanding of developmental processes during caryopsis development is critical to the understanding of the differentiation, patterning and development of the mature plant. An in depth analysis of the molecular processes regulating triticale caryopsis development at the cellular level and comparison to other cereals and angiosperms will provide insight into the mechanisms responsible for cellular differentiation and patterning in the triticale and cereal caryopsis. This knowledge may also provide insight into the evolution of the caryopsis in cereals and more globally in angiosperms.

Despite the economic importance of cereal caryopses, there is still relatively little known about their developmental regulation. Efforts have been made to characterize grain development in cereals including wheat (Drea et al., 2005; Wegel et al., 2005), barley (Doan et al., 1996), rice (Brown et al., 1996), maize (Becraft and Asuncion-Crabb, 2000; Gomez et al., 2002; Monjardino et al., 2007) and brachypodium (Opanowicz et al., 2011) and many have focused on the endosperm due its unique triploid nature and uncertain evolution. These studies have shown that while there are similarities in the developmental patterns there are also distinctive differences in tissue types,

developmental timing, organization and gene expression patterns among the cereals. Functional analysis of gene expression in the triticale caryopsis would help to understand the developmental processes involved in triticale grain development and its similarities and differences to other cereals.

In cereals, the female gametophyte is surrounded by the maternal tissues of the ovule. The gametophyte consists of the nucellus and the embryo sac and these are surrounded by the two integuments which are in turn surrounded by the carpel wall. These maternal layers develop to become the coat. Double fertilization results in the fusion of one sperm cell with the egg cell to form the zygote which develops into the embryo and the other sperm cell fertilizes the central cell by fusing with the two polar nuclei to form the triploid endosperm (Esau, 1977).

Following fertilization the embryo, endosperm and surrounding tissues undergo many changes to form the mature caryopsis. The majority of the nucellus becomes absorbed. The outer integument disintegrates while the inner integument and the nucellar epidermis are compressed and form the pigmented seed coat layer. The cuticle and the thin walled parenchyma cells of the carpel undergo cell wall thickenings, cell death and reabsorption during caryopsis development until they form the outer seed coat (Esau, 1977, WHEAT: The Big Picture, http://www.wheatbp.net/cgi-bin/grain3.pl). The zygote undergoes a single cell division to produce an apical and a basal cell marking the beginning of cell differentiation that determines the patterning of the next sporophytic generation. Cell divisions of the apical cell form the globular embryo which continue to divide and differentiate into the shoot and root poles and the scutellum which serves to transfer nutrients from the endosperm to the embryo. After fertilization, the endosperm

undergoes a number of rounds of mitosis without cytokinesis to form the ceonocytic endosperm before undergoing cellularization (Olsen, 2001). Cellularization of the endosperm begins on the outside of the endosperm (nearest to the nucellus) and progresses inward (Olsen, 2001, 2004). As endosperm development progresses, regions differentiate to form the aleurone layer, the outermost layer of endosperm; transfer cells, the modified aleurone cells near the scutellum and nucellar projection that develop transfer cell morphology; and the starchy endosperm, the inner endosperm cells that serve as starch and protein storage cells (Olsen, 2004).

Functional analysis of the mechanisms regulating the development of the triticale caryopsis is difficult due to its hexaploid nature, low transformation efficiency, lack of a completely sequenced genome and the inability to generate mutant screens. Microarrays have been used to study gene expression of cereal caryopsis in barley (Druka et al., 2006), wheat (Gregersen et al., 2005; Wan et al., 2008; Schreiber et al., 2009), rice (Yamakawa et al., 2007; Jiao et al., 2009; Wang et al., 2010), maize (Sekhon et al., 2011), many as part of larger gene expression analyses throughout plant development. Gene expression analysis using microarrays often has limited spatial resolution as tissue samples collected for these studies often include entire plant parts such as the whole caryopsis or the entire plant. While sampling methods such as tissue dissection or laser capture micro-dissection can improve the spatial resolution they are tedious and time consuming and therefore less common in global gene expression studies. In many cases, these global expression analyses are used to identify interesting genes which are then subjected to further functional characterization.

In situ hybridization of RNA transcripts can be used to improve the spatial resolution and define the cellular specificity of gene expression patterns. This technique has been recently been used in a large scale spatial analysis of gene expression in the wheat caryopsis (Drea et al., 2005) demonstrating its utility in the functional characterization of gene expression in cereal caryopsis. Here we have developed a method for the spatial analysis of gene expression throughout the development of the triticale caryopsis. We have developed this approach to suit laboratories without access to automated high-throughput equipment and accompanying techniques. Procedures for probe preparation were modified to allow specific hybridization to regions of transcripts that may enable distinction among gene families and similar sequences that may have different spatial expression patterns. Distinct spatial expression patterns were observed that indicate specialized function in cellular differentiation and caryopsis development. We have shown that this method can improve the spatial resolution of previous gene expression analysis methods and can provide a cellular context for further functional analysis.

5.2 MATERIALS AND METHODS

5.2.1 Plant material preparation

Triticale (x *Triticosecale* Wittm. *cv*. AC Certa) was planted into 1 gallon pots of Cornell mix, a steam-pasteurized (121˚C for 8 h) mixture of loam, sand, and peat (3:1:1, by volume, Boodley and Sheldrake, 1977). 3 seeds were planted per pot 2 cm deep then seedlings were thinned to 2 per pot after 1 week growth. Starting seed material was produced by Agriculture and Agri-Food Canada in Swift Current, Saskatchewan as part of their annual seed increase. All plants were grown in growth cabinets (Conviron, Winnipeg, Canada) in order to maintain a disease and stress free environment. Growth conditions were 16 h photoperiods at 18˚C and 8 h dark periods at 15˚C. Light intensities of 350 µmole m^{-2} s^{-1} of photosynthetic photon flux density were provided by cool white fluorescent tubes and incandescent bulbs. Individual spikes (heads) were tagged daily at anthesis and whole triticale caryopses were harvested at 3, 6, 9 and 12 days post anthesis (DPA). Spikelet differentiation begins in spikelets that at maturity are positioned in the central region of the spike then proceeds acropetally within a spikelet and acropetally and basipitally along the rachis simultaneously and fertilization patterns within a spike follow differentiation patterns (McMaster et al., 1992). Therefore, only caryopsis from the mid-section of the spike and the outer two florets of each spikelet were chosen to maximize developmental consistency. Brush ends of carypsis (stigmas) were trimmed to prevent entrapment of air bubbles which prevents sinking in the fixative solution and can cause uneven fixation and embedding. Embryos with attached scutella from 12 DPA caryopses were also dissected from the remainder of the caryopses under a stereomicroscope.

Collected caryopses and embryos were placed in 10 mL vials containing 5 mL of fixative solution (4% v/v formaldehyde, sodium-phosphate buffer pH 7.5 [8 mM NaH_2PO_4, 42 mM Na_2HPO_4], 50 mM NaCl) and transferred to a closed vessel for fixation at 35˚C overnight (between 16 and 18 h) under vacuum (>50 cm Hg). Following fixation, the dehydration process was as follows: 70% ethanol for 1 h at 35˚C, 80% ethanol for 1.5 h at 35˚C, 90% ethanol for 2 h at 35˚C, 100% ethanol for 1 h at 35˚C, 100% ethanol for 1.5 h at 35˚C, 100% ethanol for 2 h at 35˚C, 100% ethanol overnight at 4˚C (no vacuum), 100% ethanol for 2 h at 35˚C, 100% xylene for 0.5 h at 35˚C, 100% xylene for 1 h at 35˚C, 100% xylene for 1.5 h at 35˚C. All dehydration steps were performed under vacuum (>50cm Hg) unless otherwise stated. Molten paraffin wax was slowly added to the xylene containing the caryopsis for an infiltration process as follows: ~70:30 xylene:paraffin for 1 h with increasing temperature to 60˚C, ~50:50 xylene:paraffin for 1.5 h at 60˚C, 100% paraffin for 1 h at 60˚C. Several changes of paraffin were made as follows: 100% paraffin for 1 h at 60˚C, 100% paraffin overnight at 60˚C, 100% paraffin for 2 h at 60˚C, 100% paraffin for 2 h at 60˚C, 100% paraffin for 2 h at 60˚C. All paraffin infiltration steps were performed under vacuum (>50cm Hg). Individual caryopses were then transferred to fresh molten paraffin in an embedding block at 60˚C. Samples were aspirated under vacuum for 15 min at 60˚C to remove any air bubbles. Embedding blocks were then placed at room temperature to solidify. Sample blocks containing individual caryopses were stored in sealed containers at 4˚C until sectioning. All vials, tools and embedding blocks were cleaned with Absolve™ (Perkin Elmer Life and Analytical Sciences, Cat. No. 6NE9711) then rinsed with RNase-free water prior to use to eliminate RNA degradation by exogenous RNases.

5.2.2 Section preparation

Ten-micrometer sections from triticale samples at the required stages were cut on a Leica Microtome (RM2235; Wetzlar, Germany) and organized on Fisherbrand® Superfrost®/Plus pre-cleaned microscope slides (Fisher Scientific, Canada). Slides were divided into wells using a paraffin crayon to allow placement of sections from caryopses at multiple stages of development onto the same slide with separation. This allowed multiple hybridizations on a single slide. Sections were placed onto the slide by floating on RNase-free water in the wells created by the paraffin crayon then removal of the water by wicking with a kimwipe. After drying down at 42˚C overnight, slides were checked by visual inspection using a dissecting microscope. To remove wax and prepare the slides for hybridization, the following pretreatment steps were performed; three washes in HistoChoice® clearing agent (Sigma-Aldrich, St. Louis Missouri) for 15 min each; then through a 100, 100, 95, 85, 70, 50, and 30% ethanol series (30 s each); two rinses in H_2O (30 s each) and a 5 min wash in 2X SSC (1X SSC is 0.15 M NaCl and 0.015 M sodium citrate). Slides were then centrifuged at 200 x g for 1 min in a swinging bucket rotor (centrifuge) to remove SSC and immediately hybridized with DIG labeled probes.

5.2.3 Sequence analysis and primer design for DNA templates and RNA probes

Two sequences previously shown to have distinctive expression patterns by *in situ* hybridization in wheat seeds (Drea et al., 2005) were chosen to design probes for use as positive controls in our study. These probes were designed based on the histone H4 transcript (transcript ID 701992946, Drea et al., 2005) which had a high level of expression with a distinctive spotty pattern and a putative pectinesterase (transcript ID

702008330, Drea et al., 2005) which was present throughout the nucellus at 3 DPA and maintained in the abaxial nucellar projection from 6 to 9 DPA. The remaining probe sequences chosen for probe design to investigate cellular specificity of expression patterns within the seed, were based on expression patterns observed in the microarray analysis of triticale gene expression (Chapter 4). All transcripts used to generate templates for probe synthesis and their expected cellular expression patterns are listed in Table 5.1.

Gene-specific forward and reverse primers used for amplification of primary DNA templates from the selected transcripts are listed in Table 5.2. All primers were designed using Primer3 (Rozen and Skaletsky, 2000) with lengths near 21 nucleotides, GC content near 50%, melting temperatures near 60˚C. Amplicons between 140 and 160 bp were targeted to allow probe specificity and eliminate the need for probe hydrolysis. Primers used for expression level quantification by qRT-PCR in Chapter 4 were also used for amplification of templates for *in situ* probe synthesis in cases where the same transcripts were studied (Ta.13160.2.S1_x_at, Ta.20064.1.S1_at and Ta.69777.1.S1_at).

5.2.4 Generating DNA templates and DIG labeled RNA probes

Triticale cDNA generated for qRT-PCR in Chapter 4 was used to generate DNA templates required for generation of RNA probes for *in situ* hybridization. A pool of all seed cDNA samples was made and 10 µL of a 1:50 dilution used in primary sequence amplification reactions. Primary amplification of the sequence of interest was performed by standard hot start PCR using (50 µL reactions: 1X PCR buffer, 200 µM each dNTP,

Table 5.1 Transcripts chosen for amplification of DNA templates for *in situ* probe synthesis.

Probe Name	Source and transcript used for primer design*	Expected Target Cells
Histone H4	Drea et al. 2005 (transcript ID 701992946)	High expression with distinctive spotty pattern
Pectinesterase	Drea et al. 2005 (transcript ID 702008330)	present throughout nucellus at 3 DPA and mintained in the abaxial nucellar projection from 6 to 9 DPA
Ta.20064.1.S1_at	Microarray (Chapter 4), CA666265	Crease cells
Ta.13160.2.S1_x_at	Microarray (Chapter 4), CA635509	Crease cells
Ta.69777.1.S1_at	Microarray (Chapter 4), J02961.1	Embryo cells
Ta.1751.2.S1_x_at	Microarray (Chapter 4), CA611575	Embryo cells

* Accession numbers for transcripts from Chapter 4 are NCBI EST accession numbers

Table 5.2 Gene specific primers used to generate primary DNA templates used for *in situ* probe synthesis.

Probe Name	Gene Specific Forward Primer 5' to 3'	Gene Specific Reverse Primer 5' to 3'	PCR Product size (bp)
Histone H4	ACGTCGTCTACGCGCTCAA	CAGAGGAAATAGCGAACTAAAC	151
Pectinesterase	AACTGATGGCAGGGTTAAGTG	GTTTTCTACCGAAGCCACTGC	158
Ta.20064.1.S1_at	GCTCGTCTCCTTGCCATCTA	CCATTAAGCATCCGAACATGA	146
Ta.13160.2.S1_x_at	CTGACGTCCAGAACCCATC	CTCCTTGGTCTCCTCAACCTC	158
Ta.69777.1.S1_at	GCAGGGTTTGCACTAACAACTA	CATTCTGCGAGAAGAGTGGAG	148
Ta.1751.2.S1_x_at	GGCGCCTCCTAATAAAGGGCTCCA	TGCTTCGGAAGGAGTAACCAAGCAC	162

0.3 µM forward primer, 0.3 µM reverse primer, 1X QIAGEN Q-solution, 1.25 units QIAGEN HotStarTaq ®). PCR cycling was as follows: 95˚C for 15 min, then 40 cycles of 94˚C for 1 min, 56˚C for 1 min and 72˚C for 1 min; and a final extension of 72˚C for 10 min. PCR products were purified using the High Pure PCR Product Purification Kit (Roche Applied Science, Mannheim, Germany) according to the kit manual and products were eluted in 50 µL H_2O. For *in situ* probe labeling by *in vitro* transcription, a T7 RNA polymerase site was added to the amplified PCR products (20 ng) using the BLOCK-iT™ T7-TOPO® Linker kit (Invitrogen™ Life Technologies, Carlsbad, California) according to the kit manual. Linked PCR products (1 µL) were immediately used for a secondary amplification using the BLOCK-iT™ T7 primer (5'-GATGACTCGTAATACGACT CACTA-3') and either the gene-specific reverse or gene-specific forward primer (listed in Table 5.2) to generate sense and antisense templates respectively. Aside from the change in primers, the secondary amplification reaction mixture and cycling conditions were the same as for the primary amplification. Secondary PCR products were analyzed for quantity and quality using agarose gel electrophoresis then purified using the High Pure PCR Product purification kit (Roche Applied Science, Germany) according to the kit manual. Products were eluted in 50 µL of H_2O and quantified using UV spectrophotometry.

Generation of DIG labeled sense and antisense ssRNA probes from the sense and antisense DNA templates was performed by *in vitro* transcription with T7 RNA polymerase using the DIG RNA Labelling Kit (SP7/T7) from Roche (Roche Applied Science, Mannheim, Germany) according to the kit manual. DIG labeled probes were purified by adding 4 µL 4 M LiCl and 50 µL 100% ethanol and incubating at -80˚C for

1 h then centrifuging at maximum speed (approximately 16,000 x g) for 10 min in a bench top microfuge to pellet the probe and enable removal of the ethanol mixture before washing probes with 70% ethanol. Probes were centrifuged for another 10 min, ethanol removed, then allowed to air dry and resuspended in 50 µL of RNase-free H_2O and stored at -80˚C. DIG labeled RNA probes were quantified using a spot test according to the DIG RNA Labelling Kit (SP7/T7) kit manual. Colorimetric detection was performed using anti-DIG AP, Fab fragments (Roche Applied Science, Mannheim, Germany), NBT/BCIP (Roche Applied Science, Mannheim, Germany) and the Roche Wash and Block Buffer Set (Roche Applied Science, Mannheim, Germany). After cross-linking with UV light the spotted membrane was washed 5 minutes in 1X washing buffer, incubated for 60 min in blocking buffer then 60 min in anti-DIG AP solution (75 mU/mL in 1X maleic acid buffer), rinsed with washing buffer then washed with washing buffer two times for 15 minutes each then incubated in coloring reaction solution (18.75 mg/mL NBT, 9.4 mg/mL BCIP) in the dark until detection of probe spots was sufficient (up to 24 h with periodic checking). All detection steps occurred under constant agitation.

5.2.5 Northern blots with DIG labeled probes

Northern blot analysis was used to test the RNA binding ability and capacity for detection of the control DIG-labeled RNA probes (histone H4 and pectinesterase) designed above as the method for synthesizing these probes is different than that described by Drea et al. (2005). Whole triticale caryopses at 3, 6, 9 and 12 DPA were collected for RNA isolation as above but were immediately placed in liquid N_2 and stored at -80˚C until RNA extraction. RNA was extracted using the modified protocol of Singh

et al. (2003) with RNeasy® RNA cleanup and on-column DNA digestion used for endosperm and embryo samples in Chapter 4. Quantity and quality of RNA samples were assessed by UV spectrophotometry and agarose gel electrophoresis respectively. RNA (either 3 or 15 µg) was separated on a formaldehyde agarose gel, stained with ethidium bromide then blotted onto a positively charged membrane (Whatman Nytran™ SPC nylon membrane, GE Healthcare Life Sciences). The blotted membrane was UV cross-linked and stained with methylene blue prior to hybridization. Membranes were pre-hybridized in 15 mL of hybridization buffer [50% formamide, 0.25% (w/v) SDS, 300 mM NaCl, 10 mM Tris-HCl (pH 7.5), 1 mM (w/v) EDTA (pH 8.0), 1X Denhardt's solution, 125 µg/mL denatured herring sperm DNA, 125 µg/mL yeast tRNA, 10% w/v dextran sulfate] for 30 minutes prior to hybridization overnight in 3.5 mL of hybridization buffer with 600 ng of DIG-labeled probe. Following hybridization the membranes were washed two times in 2X SSC, 0.1% SDS at room temperature for 5 min and two times in 0.1X SSC, 0.1% SDS at 68˚C for 15 min. The washed membrane was developed with the immunodetection system using anti-DIG alkaline phosphatase and NBT-BCIP (Roche).

5.2.6 Hybridization of DIG labeled probes to tissue sections and washing of slides

Adhesive hybridization chambers (HybriWell™ HBW2240, Grace bio-Labs, Bend, Oregon, U.S.A.) were applied to the slides over the tissue sections and then filled with 200 µL of preheated (68˚C) hybridization solution [50% formamide, 0.25% (w/v) SDS, 300 mM NaCl, 10 mM Tris-HCl (pH 7.5), 1 mM (w/v) EDTA (pH 8.0), 1X Denhardt's solution, 125 µg/mL denatured herring sperm DNA, 125 µg/mL yeast tRNA, 10% w/v dextran sulfate] for pre-hybridization. Slides were incubated at 68˚C for 1 hour

in a wetting box with 50 mL of wetting solution (50% formamide and 50% 2X SSC prepared from Sigma 20X SSC for molecular biology, Sigma-Aldrich, St. Louis Missouri). Meanwhile probe solutions were prepared by diluting DIG labelled cRNA probes to 20 ng/µL using RNase-free H_2O then to 0.5 ng/µL in 200 µL of preheated (68˚C) hybridization solution, boiled for 5 min at 100˚C and then incubated on ice for 2 min immediately prior to applying to the slides. After the one hour incubation with hybridization solution, the pre-hybridization chambers were removed from the slides and the pre-hybridization solution was carefully blotted away using laboratory wipes. The edges of the slides were carefully wiped without disturbing the tissue sections to free the edges from hybridization solution. Fresh hybridization chambers were placed onto the slides and filled with the pre-boiled hybridization solution containing the DIG-labeled probe. The hybridization chambers were sealed and the slides incubated overnight at 68˚C in a wetting box containing 50 mL of wetting solution.

Following hybridization, chambers were removed from slides which were immediately placed in 4X SSC at room temperature. Slides were washed using the DIG Wash and Block Buffer set (Roche Applied Science, Mannheim, Germany) and the following series, two 15 min washes in 2X SSC, 0.1% SDS at room temperature, two 15 min washes in 0.1X SSC, 0.1% SDS at 68°C, 5 minutes in Roche washing buffer (diluted to 1X with RNase-free H_2O) at room temperature, 60 min in Roche blocking buffer (diluted to 1X using Roche 1X maleic acid buffer), 60 min in anti-DIG-AP solution (75 mU Anti-DIG-AP, Fab fragments/mL in 1X Roche blocking buffer, Roche Applied Science, Mannheim, Germany) at room temperature, rinsed with Roche washing buffer, two 15 min washes with Roche washing buffer at room temperature and 5 min in Roche

detection buffer (diluted to 1X with H_2O) at room temperature. All washing steps were with constant agitation. Slides were then transferred to shallow trays where they covered with coloring reaction solution (180 µL of NBT/BCIP in 10 mL of 1X detection buffer, Roche Applied Science, Mannheim, Germany). Slides were incubated in the dark with constant agitation for up to 24 h with periodic checking until coloring reaction had proceeded to allow transcript detection. Slides were then washed two times for 10 min in H_2O to stop the reaction and allowed to dry. Slides were prepared for imaging by permanently mounting in Aqua PolyMount (Polysciences Inc. Warrington, PA) under glass coverslips.

5.2.7 Imaging

Representative sections for each probe and their respective negative control probe were imaged using a stereomicroscope (Zeiss, SteREO Discovery.V20, Carl Zeiss MicroImaging GmbH, Göttingen, Germany) mounted charge coupled device color camera (AxioCam, Carl Zeiss) controlled by dedicated software. Images were captured under brightfield conditions.

5.3 RESULTS

5.3.1 Probe synthesis and Quantification

Amplification of DNA templates from seed cDNA pools using the primers in Table 5.2 produced single products that were used for annealing of the T7 polymerase priming site and secondary amplification using the T7 primer. Generation of DIG-labeled sense and antisense single stranded RNA (ssRNA) probes from the sense and antisense DNA templates by *in vitro* transcription with T7 RNA polymerase was successful and detection of the probes with anti-DIG-AP and NBT/BCIP was possible as determined by the spot test (Figure 5.1). The *in vitro* transcription produced probes with concentrations between 100 and 400 ng/ µL (Table 5.3) as determined by spot intensity comparison to the control labeled RNA spotted alongside each probe in Figure 5.1.

5.3.2 Northern Blots

Histone H4 and the pectinesterase probes were generated for use as positive controls for *in situ* hybridization of RNA transcripts. Northern blot analysis of these probes revealed that the probe synthesis methods described above were successful at producing DIG-labeled antisense RNA probes that would hybridize to triticale seed RNA transcripts. It also demonstrated that the DIG-labeled sense RNA probes would serve as negative controls for each transcript hybridization. The spot blot results showed that DIG-labeling of both sense and antisense probes was successful but following the hybridization, washing and detection steps for the northern blot, only the antisense probe

Figure 5.1 Spot blots of DIG-labeled probes for *in situ* hybridizations. 1 µl of dilutions of the DIG labeled probes were spotted onto positively charged Nytran membrane. The serial dilution factors of the antisense (A) and sense (S) probes were (from left to right); 1:20, 1:200, 1:2X10^4, 1:4 X10^4, 1:2 X10^5, 1:4 X10^5, 1:2X10^6, 1:4 X10^6, 1:2 X10^7. 1 µl volumes of serial dilutions of control labeled RNA (C) were also spotted alongside each probe for estimation of quantity of DIG-labeled probes by comparison of spot intensities following detection with anti-DIG-AP fragments and coloring reaction using NBT-BCIP. Quantities of controlled labeled RNA are listed below each blot. Spots used to estimate the quantity of DIG-labelled antisense and sense probes are marked by a black dash.

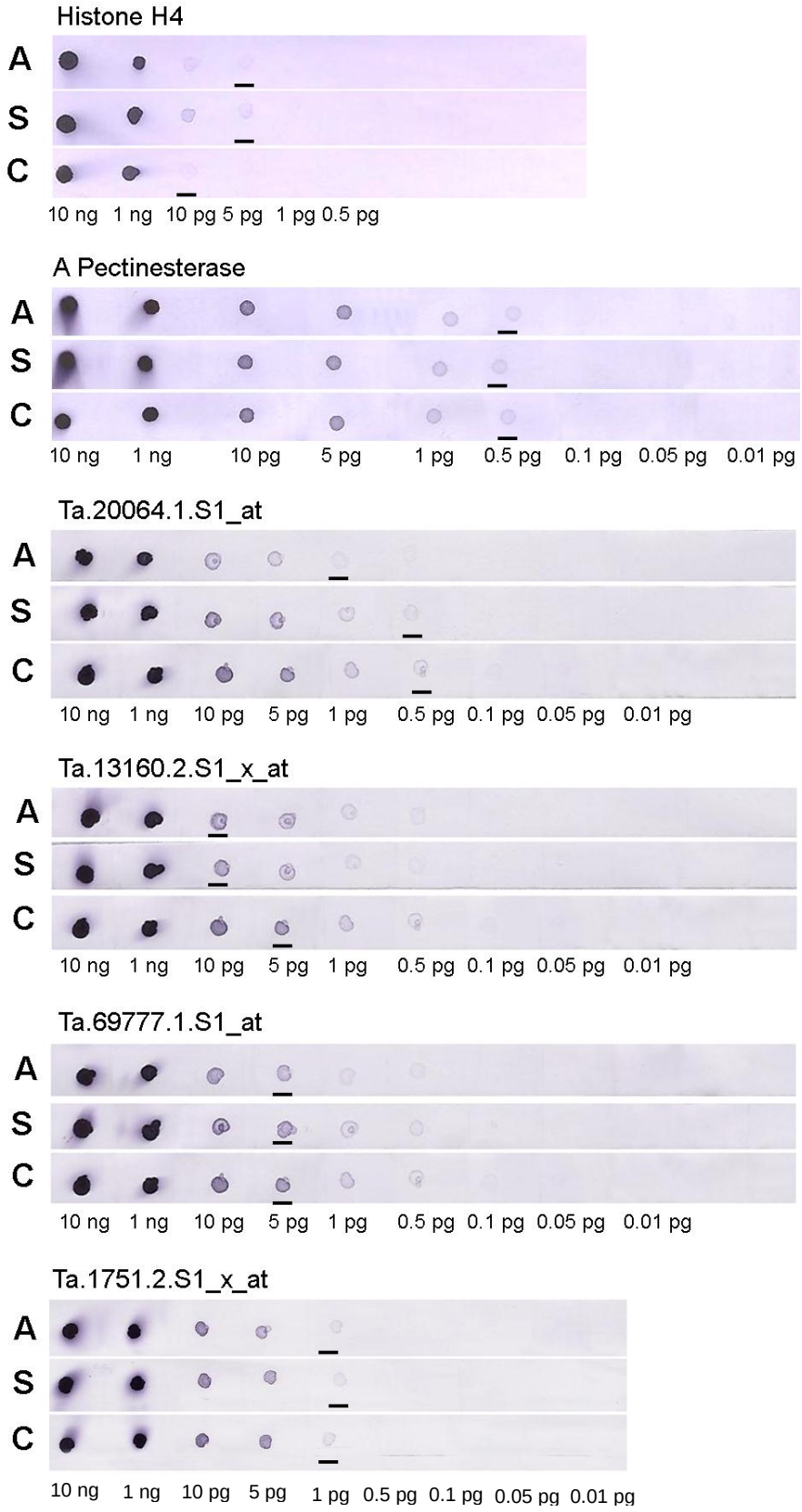

Histone H4
A
S
C
10 ng 1 ng 10 pg 5 pg 1 pg 0.5 pg
A Pectinesterase
A
S
C
10 ng 1 ng 10 pg 5 pg 1 pg 0.5 pg 0.1 pg 0.05 pg 0.01 pg
Ta.20064.1.S1_at
A
S
C
10 ng 1 ng 10 pg 5 pg 1 pg 0.5 pg 0.1 pg 0.05 pg 0.01 pg
Ta.13160.2.S1_x_at
A
S
C
10 ng 1 ng 10 pg 5 pg 1 pg 0.5 pg 0.1 pg 0.05 pg 0.01 pg
Ta.69777.1.S1_at
A
S
C
10 ng 1 ng 10 pg 5 pg 1 pg 0.5 pg 0.1 pg 0.05 pg 0.01 pg
Ta.1751.2.S1_x_at
A
S
C
10 ng 1 ng 10 pg 5 pg 1 pg 0.5 pg 0.1 pg 0.05 pg 0.01 pg

Table 5.3 Concentrations of DIG-labeled RNA probes. Probe concentrations were estimated by comparison to control labeled RNA on a spot blot following detection with anti-DIG AP and NBT/BCIP.

Probe Name	Antisense Probe	Sense Probe
Histone H4	~400 ng/µL	~400 ng/µL
Pectinesterase	~200 ng/µL	~200 ng/µL
Ta.20064.1.S1_at	~100 ng/µL	~200 ng/µL
Ta.13160.2.S1_x_at	~100 ng/µL	~100 ng/µL
Ta.69777.1.S1_at	~200 ng/µL	~200 ng/µL
Ta.1751.2.S1_x_at	~200 ng/µL	~200 ng/µL

resulted in a color reaction. The detection of transcripts by the antisense probes and not the sense probes indicates that hybridization between the seed RNA transcripts and the antisense RNA probes was successful and that no naturally occurring antisense histone H4 or pectinesterase transcripts could be detected in the seed RNA. The lack of detection of hybridization by the sense RNA probes also means that the hybridization and washing conditions were stringent enough to prevent non-specific hybridization to RNA transcripts.

The histone H4 antisense probe hybridized to all four seed developmental stages examined and most strongly to RNA from the early seed developmental stages (3 and 6 DPA, Figure 5.2). Expression of the histone H4 transcript was reduced slightly in the 9DPA stage and more so in the 12 DPA stage as compared to the 6 DPA stage of development. Because the amount of RNA collected at the 3 DPA stage of development was lower than that collected from the other stages of development, a smaller amount of total RNA at this stage was loaded onto the gel used for the northern blot (3 µg instead of the 15 µg used for all other stages). Even with the smaller amount of total RNA used for this stage of development, the expression of histone H4 at 3 DPA was relatively high compared to that at 6 DPA and later stages. The histone H4 sense probe did not hybridize to triticale seed RNA at any of the four stages of development analyzed (Figure 5.2). The histone H4 transcript was expected to be detectable at high levels throughout seed development according to previous findings by Drea et al. (2005). While we detected high expression of this transcript throughout triticale seed development similar to that seen in wheat seed we also determined that the expression levels do decline throughout triticale seed development.

Figure 5.2 Northern blot of histone H4 DIG-labeled RNA probes to triticale seed RNA. Total triticale seed RNA was separated on a formaldehyde agarose gel (3 µg of 3DPA RNA and 15 µg of 6, 9 and 12 DPA RNA, top) alongside 6 µg of RNA ladder (0.24 to 9.5 Kb, Invitrogen Life Technologies) then transferred to positively charged nylon membrane (Whatman Nytran™ SPC nylon membrane, GE Healthcare Life Sciences, middle). Membranes were hybridized with DIG-labelled antisense and sense histone H4 ssRNA probes (bottom left and right, respectively). Detection of hybridization was by anti-DIG-AP and coloring reaction using NBT/BCIP. RNA ladder bands were marked on the blots with a graphite pencil. Band sizes are (top to bottom): 9.49, 7.46, 4.40, 2.37, 1.35 Kb.

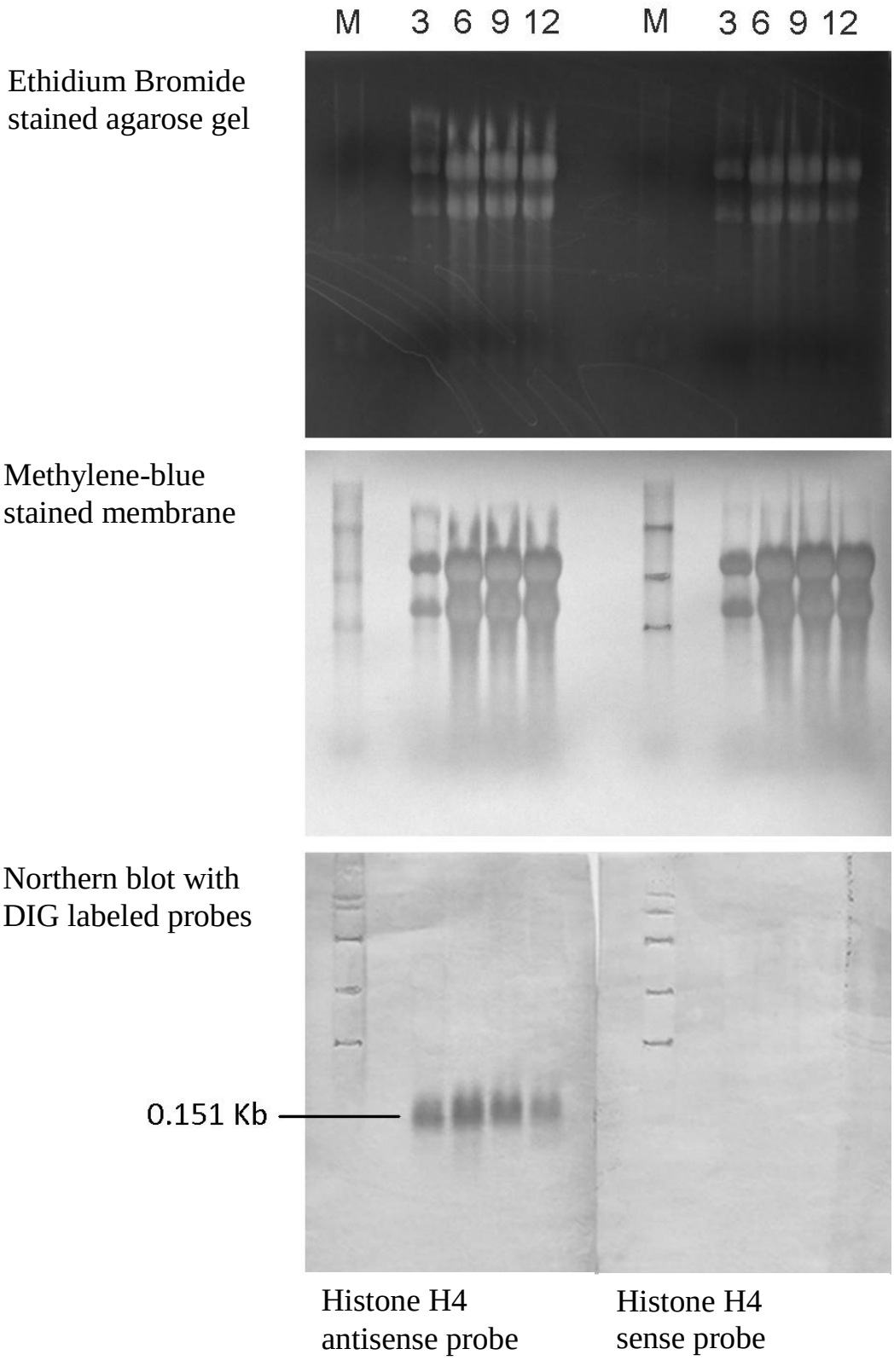
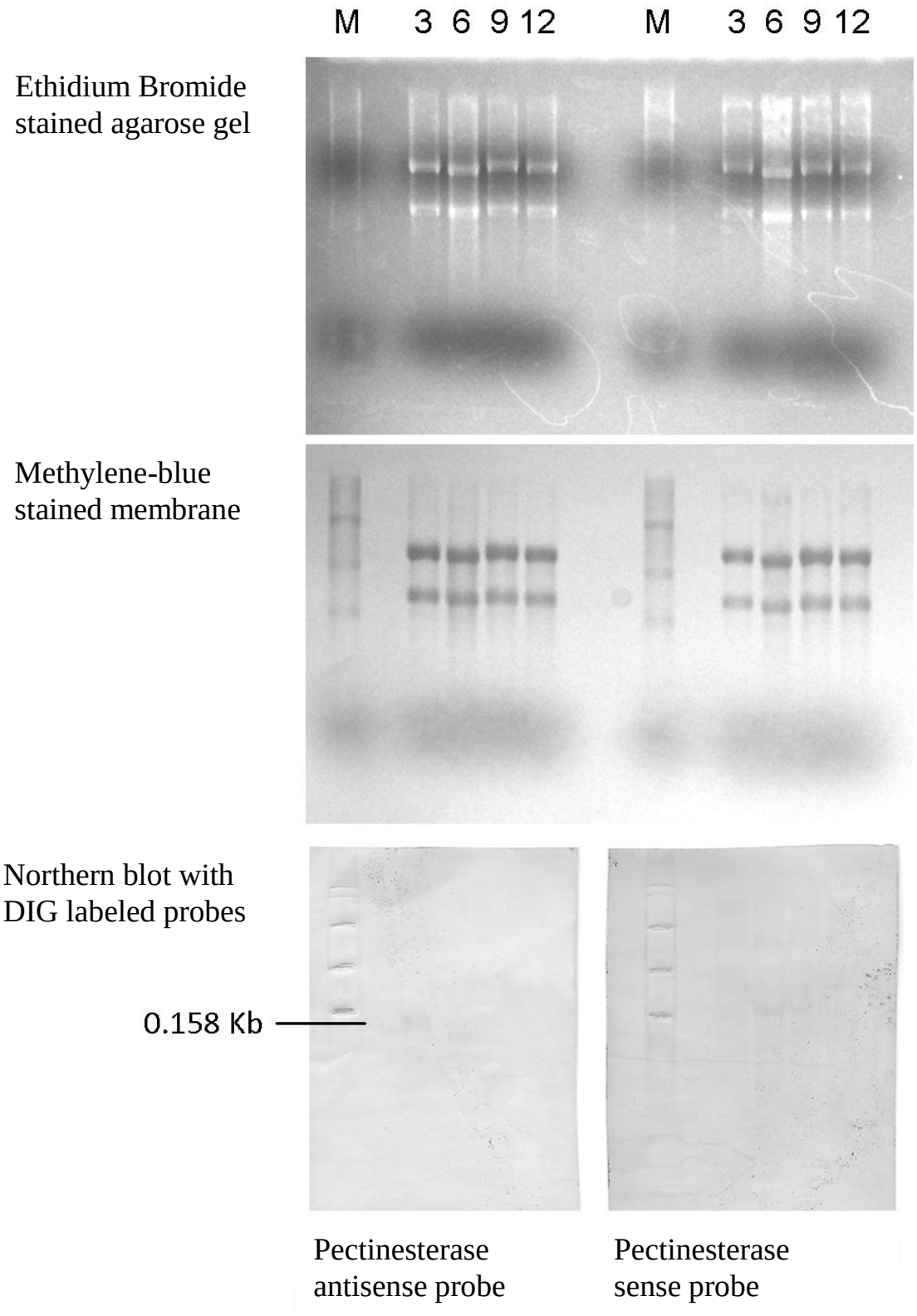

M 3 6 9 12
M 3 6 9 12
Ethidium Bromide stained agarose gel
Methylene-blue stained membrane
Northern blot with DIG labeled probes
0.151 Kb
Histone H4 antisense probe
Histone H4 sense probe

Figure 5.3 Northern blot of pectinesterase DIG-labeled probes to triticale seed RNA. Total triticale seed RNA was separated on a formaldehyde agarose gel (15 µg of RNA from 3 6, 9 and 12 DPA RNA, top) alongside 6 µg of RNA ladder (M, 0.24 to 9.5 Kb, Invitrogen Life Technologies) then transferred to positively charged nylon membrane (Whatman Nytran™ SPC nylon membrane, GE Healthcare Life Sciences, middle). Membranes were hybridized with DIG-labelled antisense and sense histone pectinesterase ssRNA probes (bottom left and right, respectively). Detection of hybridization was by anti-DIG-AP and coloring reaction using NBT/BCIP. RNA ladder bands were marked on the blots with a graphite pencil. Band sizes are (top to bottom): 9.49, 7.46, 4.40, 2.37, 1.35 Kb.

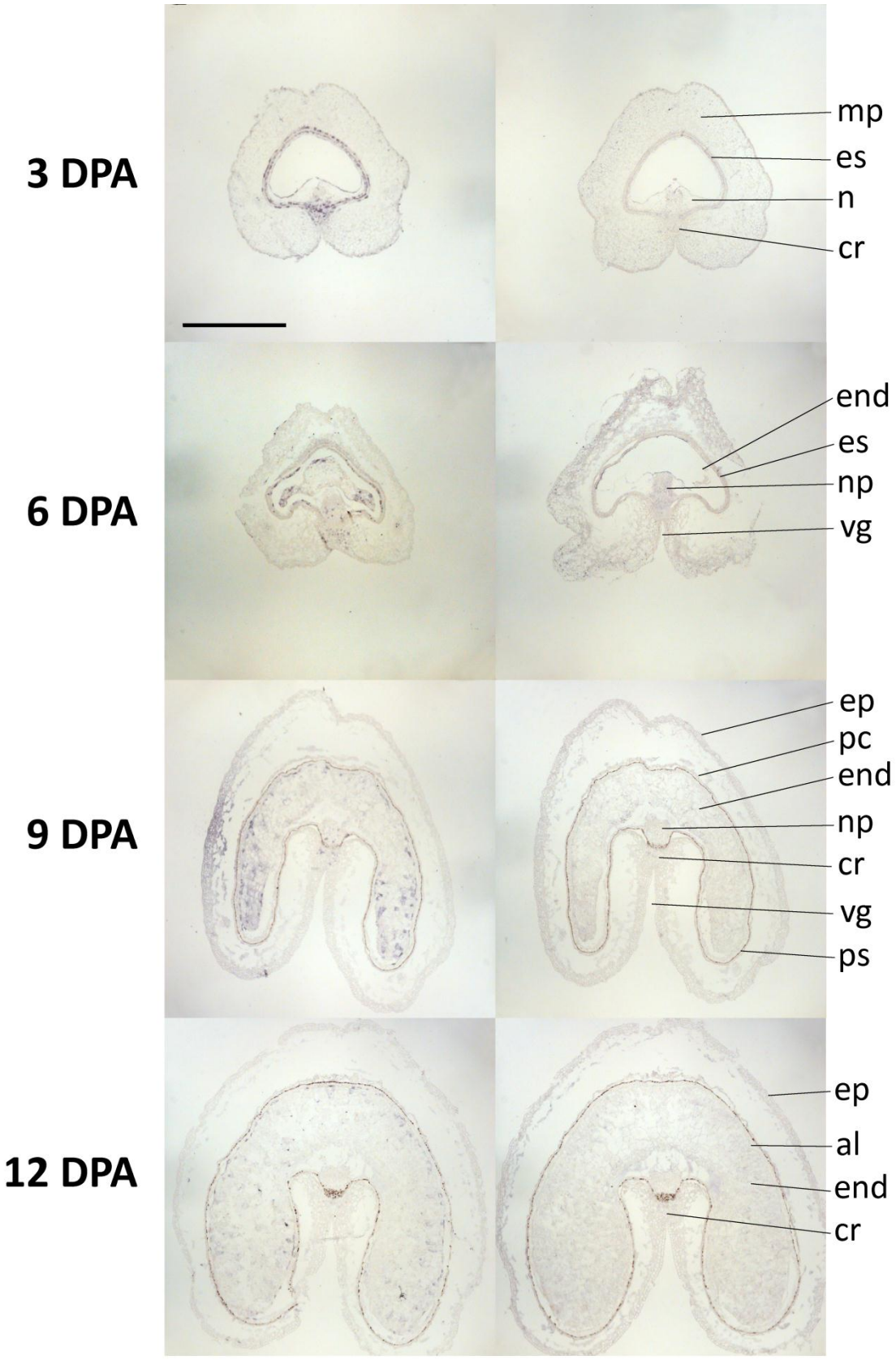

M 3 6 9 12
M 3 6 9 12
Ethidium Bromide stained agarose gel
Methylene-blue stained membrane
Northern blot with DIG labeled probes
0.158 Kb
Pectinesterase antisense probe
Pectinesterase sense probe

The pectinesterase antisense probe hybridized to 3 DPA seed RNA (Figure 5.3) but at much weaker levels than the histone H4 antisense probe. Only faint detection of the antisense probe hybridization to the 3 DPA RNA sample was possible. Detection of the pectinesterase antisense probe was not visible in the later stages of seed development (6, 9 and 12 DPA). There was no detection of the pectinesterase sense probe at any stage of triticale seed development (Figure 5.3). The expression pattern of the pectinesterase was expected to be detected at 3 DPA and maintained throughout 6 and 9 DPA seed developmental stages according to previous findings by Drea et al. (2005). Expression of this transcript was not detectable in triticale seed RNA at 6 DPA and later stages suggesting that the timing of expression pattern may vary between wheat and triticale.

5.3.3 Histone H4 is expressed in areas of active cell proliferation

The *in situ* hybridization of the histone H4 DIG-labeled RNA probes to triticale seed tissue sections at 3, 6, 9 and 12 DPA is shown in Figure 5.4. As in the northern blot, the expression was most apparent at early caryopsis developmental stages (3 DPA) and becomes reduced at the late stages of caryopsis development (12 DPA). The expression at 3 DPA is most apparent in the embryo sac wall tissue layers, the nucellar epidermis and the nucellus. Expression was also detected in the vascular tissues of the crease and the nucellar projection. A lower level of expression was detectable in the epidermal cell layers and the maternal pericarp. Transcript hybridization by the sense probe was not detectable at this stage or later stages of development. At 6 DPA, histone H4 expression is maintained in the nucellus, the crease and the nucellar projection, though to a lesser extent than at 3 DPA. Expression in remaining embryo sac wall layers appeared to be

Figure 5.4 *In situ* hybridization of histone H4 transcript. Cell type specific labeling of histone H4 at 3, 6, 9 and 12 days post anthesis (DPA) as shown by *in situ* hybridization of transverse sections of triticale caryopsis probed with digoxygenin-labeled antisense RNA (left side) and sense (right side) histone H4 RNA and viewed under Bright-Field Optics. Tissue types are annotated as follows: aleurone (al), crease (cr), embryo sac (es), endosperm (end), epiderm (ep), maternal pericarp (mp), nucellar projection (np), nucellus (n), pericarp (pc), pigment strand (ps) and ventral groove (vg). All images are in the same scale. Scale bar = 1 mm.


3 DPA
6 DPA
9 DPA
12 DPA
mp
es
n
cr
end
es
np
vg
ep
pc
end
np
cr
vg
ps
ep
al
end
cr
Antisense
Sense

reduced. At 6 DPA expression is also detectable in the developing endosperm, particularly in the areas furthest from the nucellar projection on the dorsal side and in the cheeks. At 9 DPA the majority of histone H4 expression is found within the endosperm and the crease. Expression in the endosperm cells is highest in the cheeks of the grain on either side of the ventral groove and the cells on the dorsal side of the endosperm. At 9 DPA, these areas of the endosperm are still undergoing active cell division and expansion during the grain filling stage development. Some expression is still visible in the crease and nucellar projection areas but is reduced in comparison to earlier stages of development. At 12 DPA, expression of the histone H4 transcript is no longer visible in the crease area. Since histone H4 is a marker for the S phase of cell division (Fobert et al., 1994) this suggests that cell division in this area is near completion at this stage of grain development. Expression is still visible in the endosperm and is more concentrated in the aleurone and starchy endosperm cells near the aleurone than in the central endosperm, although there are still a few cells showing expression throughout the endosperm (Figure 5.4, 12 DPA). The histone H4 sense probe showed no visible detection above background levels. Starting at 9 DPA the pigment strand is visible in the inner pericarp in both the sense and antisense probed sections as a dark cell layer outside of the endosperm and nucellus. This is not considered detection of transcript expression as this pigment strand is also visible in sections that underwent the same hybridization, washing and detection steps but without any DIG-labeled probes (either sense or antisense) added to the hybridization buffer (data not shown). At 12 DPA the pigment strand appears fully developed through the crease area with an increased number of cell layers in this area compared to the 9 DPA sections (Figure 5.4).

Blast analysis of the histone H4 sequence against the wheat GeneChip target sequences revealed 10 probe sets with matching target sequences (all ten probe sets beginning with Ta.10329 and all with an expect (*e*)-value less than or equal to 1*e*-144, Table 5.4). All ten of these probe sets are annotated as histone H4 (annotations are from Chapter 3 and all were previously shown by microarray to be expressed in the majority of the seed tissues; majority P consensus detection calls, with the exception in some cases of the epiderm tissue, Chapter 4). The majority of these probe sets also had fairly high levels of expression throughout most of the seed tissues with the exception of one probe set which had quite low expression in all seed tissues (mean RMA expression values less than 3, logarithmic scale, base 2, Table 5.4). The expression patterns observed by *in situ* hybridization therefore, confirmed the expression pattern of histone H4 observed by microarray analysis. However, compared to the microarray analysis, the expression pattern was examined over a broader range of seed developmental stages using *in situ* hybridization. This showed that there is a gradual decline in the levels of expression, and therefore cell proliferation in the triticale caryopsis throughout seed development.

5.3.4 The pectinesterase expression is only detectable at early caryopsis development

Expression of the putative pectinesterase transcript was visible in the 3 DPA stage of triticale caryopsis development by *in situ* hybridization of transverse caryopsis sections probed with DIG-labeled antisense RNA probes (Figure 5.5). Overall differences in detection between the antisense probe and sense probe were less drastic than for the histone H4 expression, similar to the trend observed by northern blot. At 3

Table 5.4 Wheat GeneChip probe sets with target sequences matching histone H4 and pectinesterase sequences. Sequences for histone H4 and the putative pectinesterase transcript were retrieved from the Supplementary Table 1 from Drea et al. (2005) and Blast homology search using the NetAffyx™ analysis center on the Affymetrix® website was used to identify probe sets on the wheat GeneChip with homologous target sequences. The annotations for the matching probe sets are from Chapter 3.

			MAS5.0 Consensus detection calls					Mean RMA expression values					
	Probe set_ID	***e*-value**	Embryo	Endosperm	Crease	Pericarp	Epiderm	Embryo	Endosperm	Crease	Pericarp	Epiderm	Annotations
Histone H4 matching probe sets	Ta.10329.42.S1_at	1.00E-158	PPP	PPP	PPA	PP	PA	9.11	6.89	6.92	7.77	6.36	histone h4
	Ta.10329.34.S1_x_at	1.00E-156	PPP	PPP	PPP	PP	PP	11.22	9.08	7.90	9.54	7.13	histone h4
	Ta.10329.22.S1_x_at	1.00E-155	PPP	PPP	PPP	PP	PA	6.20	6.06	5.70	6.27	5.87	histone h4
	Ta.10329.20.S1_x_at	1.00E-151	PPP	PPP	PPP	PP	PP	11.18	9.66	8.84	9.85	8.07	histone h4
	Ta.10329.44.S1_x_at	1.00E-150	PPP	PPP	PPP	PP	PA	9.06	7.76	7.30	8.31	6.46	histone h4
	Ta.10329.13.S1_a_at	1.00E-147	PPP	PPP	PPP	PP	PP	13.65	12.08	11.71	13.09	9.24	histone h4
	Ta.10329.29.S1_x_at	1.00E-146	PPP	PPP	PPP	PP	PP	12.59	10.38	10.34	11.62	8.77	histone h4
	Ta.10329.22.S1_at	1.00E-145	PPA	AAP	MPP	MP	PP	2.68	2.60	2.80	2.94	2.86	histone h4
	Ta.10329.19.S1_x_at	1.00E-144	PPP	PPP	PPP	PP	PP	12.68	11.54	11.41	11.94	10.72	histone h4
	Ta.10329.25.S1_x_at	1.00E-144	PPP	PPP	PPP	PP	PP	11.70	8.87	8.47	9.58	7.13	histone h4
Pectinesterase matching probe sets	TaAffx.5634.1.S1_at	1.00E-123	AAA	PPP	PPP	PP	PP	5.20	9.84	9.46	9.43	6.74	pectinesterase family protein
	Ta.8439.1.A1_x_at	3.00E-78	AAA	PPP	PPP	PP	PA	3.42	6.67	5.81	8.02	4.18	pectinesterase family protein
	TaAffx.98928.1.A1_at	0.26	AAA	AAA	PAP	PP	MP	6.15	5.39	6.81	7.80	6.94	yggt family protein

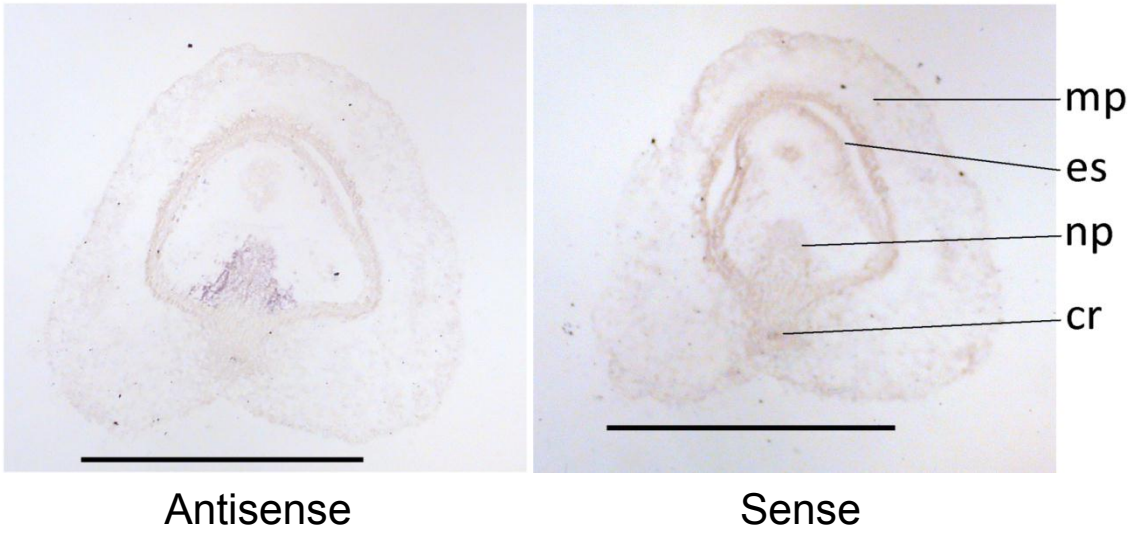


Figure 5.5 *In situ* hybridization of a putative pectinesterase transcript. Cell type specific labeling of a putative pactinesterase transcript at 3 days post anthesis (DPA) as shown by *in situ* hybridization of transverse sections of triticale caryopsis probed with digoxygenin-labeled antisense (left) and sense (right) pectinesterase RNA and viewed under Bright-Field Optics. Tissue types are annotated as follows: crease (cr), embryo sac (es), maternal pericarp (mp) and nucellar projection (np). Both images are in the same scale. Scale bar = 1 mm.

DPA, pectinesterase expression was clearly detected by the antisense probe in the cells of the nucellar projection and in a few cells of the nucellus inside the embryo sac. At the 6, 9 and 12 DPA expression was not detectable beyond background levels seen with hybridization of the sense probe (data not shown). This pattern of expression throughout development is consistent with that observed by northern blot (Figure 5.3). It was expected that the pectinesterase transcript would be expressed throughout the nucellus at 3 DPA based on previous result in wheat that showed the pectinesterase transcript to be expressed throughout the nucellus at 3 DPA followed by a decline in expression levels and restriction to the nucellar projection by 6 DPA (Drea et al., 2005). While the expression pattern in triticale is similar to those observed in wheat, a slight difference in timing may be responsible for the restriction of most of the expression to the nucellar projection by 3 DPA in triticale. Hybridization to caryopses sections sampled before fertilization, between fertilization and 3 DPA and between 3 DPA and 6 DPA may reveal very similar spatial patterns of expression of pectinesterase between wheat and triticale with a slight shift in timing.

Blast analysis of the putative pectinesterase sequence (transcript ID 702008330 from Drea et al, 2005) against the wheat GeneChip target sequences revealed 3 probe sets with sequence homology (Table 5.4). Two of these probe sets had blast hits with low *e*-values and pectinesterase family protein annotations. The two probe sets with pectinesterase annotations were previously found to be expressed in the endosperm, crease and pericarp tissues in triticale seeds by microarray analysis (Chapter 4) which would make their expression pattern consistent with that found by Drea et al (2005) in wheat caryopsis. However, the two probe sets that had target sequences with a high

percentage of similarity to the putative pectinesterase transcript from which the RNA probes were designed, were expressed in the endosperm, crease and pericarp tissues at the time of caryopsis tissue sampling for microarray analysis which was near 12 DPA (Chapter 4). Based on these microarray results we expected to detect pectinesterase transcript within the caryopsis sections sampled at 6, 9 and even 12 DPA.

5.3.5 The crease-specific putative protein kinase transcript is expressed in the crease vascular tissue

In situ hybridization of RNA probes designed to hybridize to the putative protein kinase transcript represented by probe set Ta.13160.2.S1_x_at on the wheat GeneChip, revealed that this transcript is expressed in the crease vascular tissues at 9 and 12 DPA (Figure 5.6). According to the microarray expression data this transcript was only expressed in the crease tissue within the seed at the time of sampling (Table 5.5). The *in situ* hybridization results support the microarray analysis results and show expression in the vascular tissues in the crease region of the triticale caryopsis at 9 and 12 DPA. There were strong differences in the intensity of the color reaction following the detection steps between the antisense and sense probes in the vascular crease region. At earlier stages of development (3 and 6 DPA) it appears that there may be some differences in transcript detection between the antisense and sense DIG-labeled RNA probes but the overall background levels make it difficult to determine any cell-specific expression differences. If the differences in overall detection levels are attributed to the detection of transcript expression and not differences in background level, then the putative protein kinase transcript expression is not restricted to cellular or tissue-specific expression patterns

Figure 5.6 *In situ* hybridization of a putative protein kinase transcript. Cell type specific labeling at 3, 6, 9 and 12 days post anthesis (DPA) as shown by *in situ* hybridization of transverse sections of triticale caryopsis probed with digoxygenin (DIG)-labeled antisense RNA (left side) and sense RNA (right side) and viewed under Bright-Field Optics. DIG labeled RNA probes were designed from the putative protein kinase transcript (EST accession number CA635509) represented by probe set Ta.13160.2.S1_x_at on the Affymetrix GeneChip® Wheat Genome Array. Tissue types are annotated as follows: crease (cr), embryo sac (es), endosperm (end), epiderm (ep) modified aleurone (ma), nucellar projection (np), pericarp (pc), pigment strand (ps), vascular strand (vs) and ventral groove (vg). All images are in the same scale. Scale bar = 1 mm.

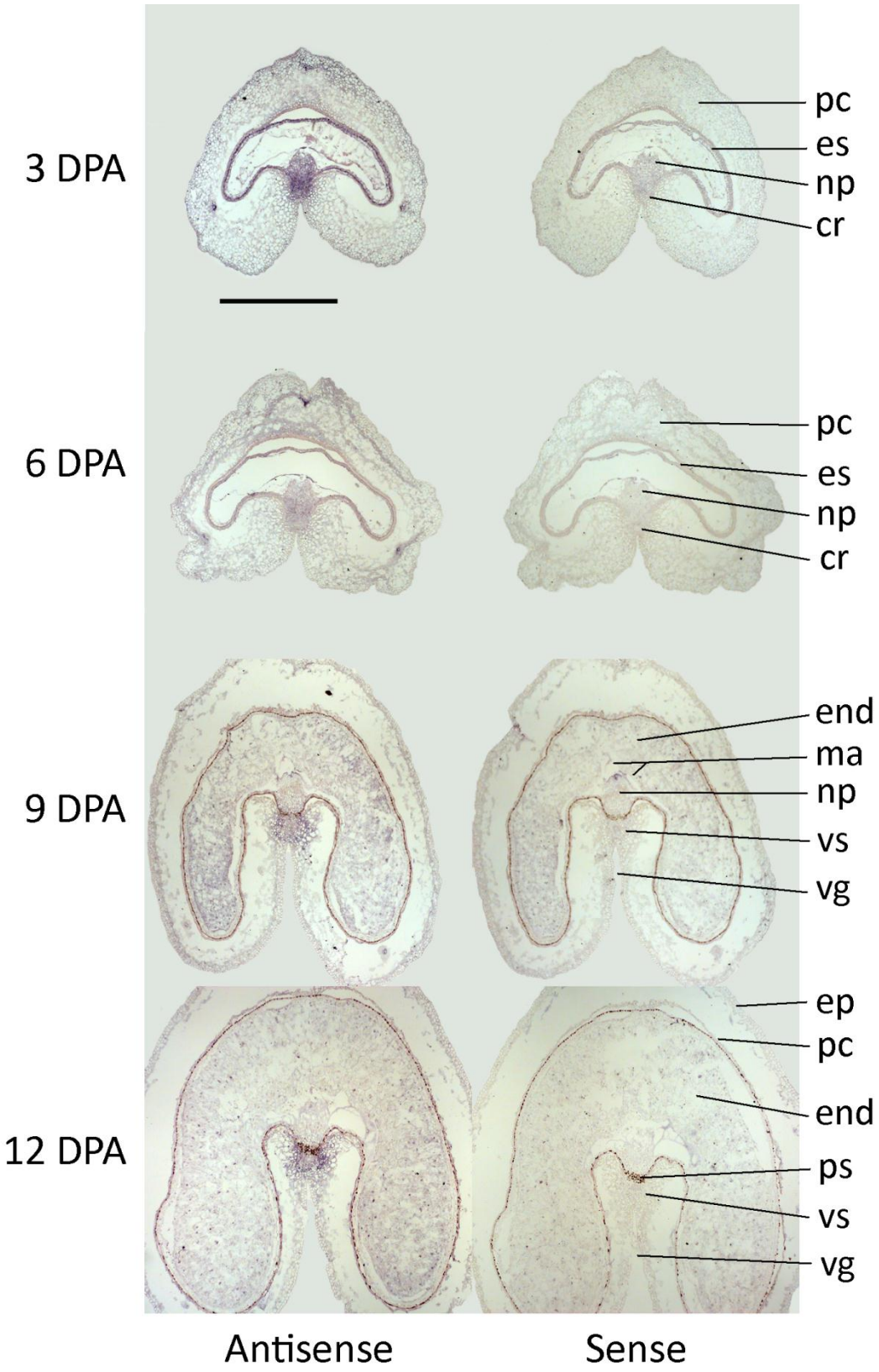

3 DPA
6 DPA
9 DPA
12 DPA
pc
es
np
cr
pc
es
np
cr
end
ma
np
vs
vg
ep
pc
end
ps
vs
vg
Antisense
Sense

Table 5.5 Expression patterns observed by microarray (Chapter 4) for probe sets chosen to design RNA probes for *in situ* hybridizations to triticale seed sections.

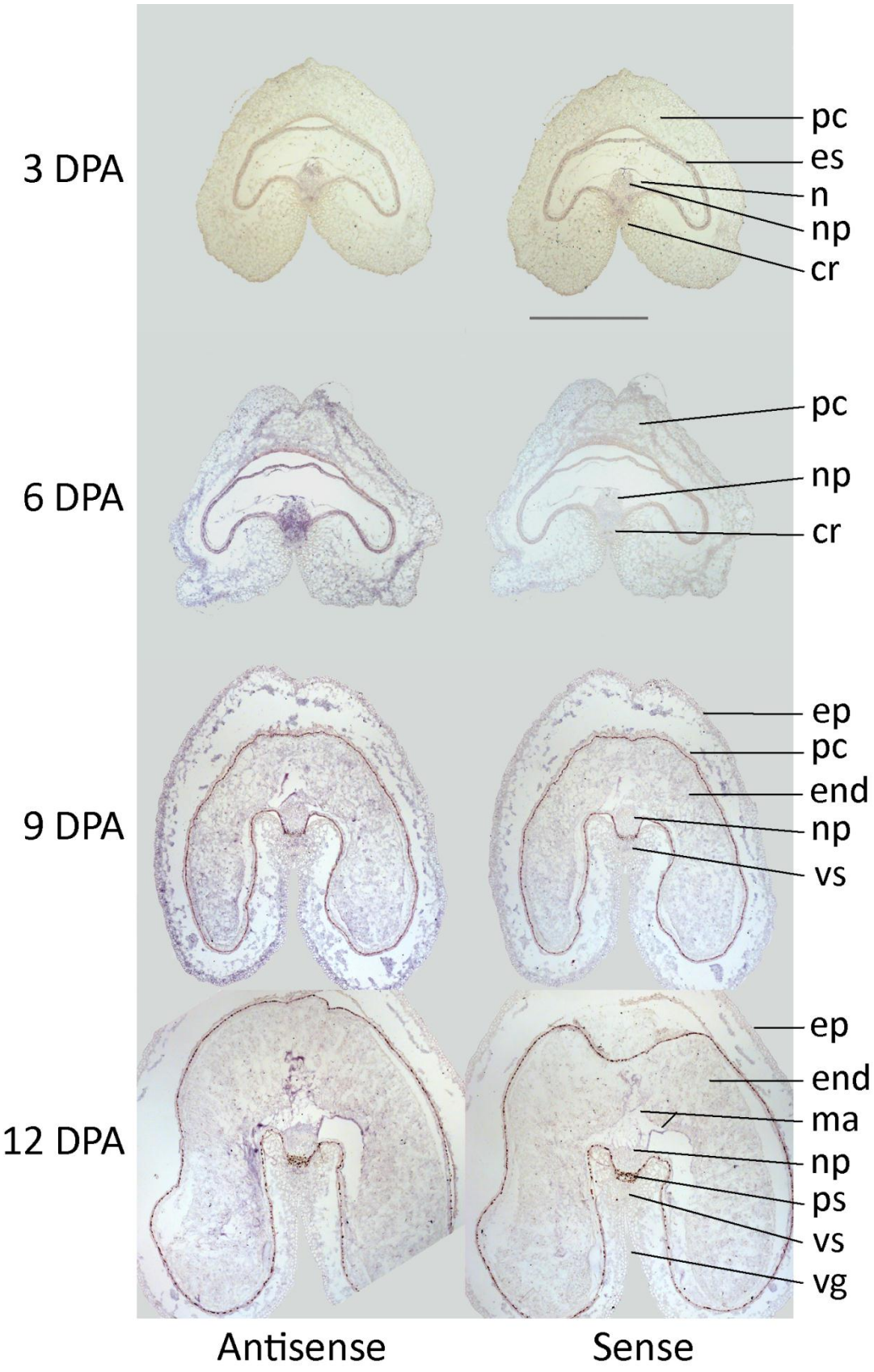
Probe set_ID	Expression pattern within triticale seed.	detection calls					Mean RMA expression values					Annotations
		Embryo	Endosperm	Crease	Pericarp	Epiderm	Embryo	Endosperm	Crease	Pericarp	Epiderm	
Ta.13160.2.S1_x_at	Crease-specific	AAA	AAA	PMP	AA	MA	3.46	3.37	6.73	4.19	5.73	putative pritein kinase
Ta.20064.1.S1_at	Crease-specific	AAA	AAA	PPP	AA	AA	4.04	3.58	9.31	5.05	3.96	ubiquitin-conjugating enzyme
TaAffx.69777.1.S1_at	Embryo-specific	PPP	AAA	AAA	AA	AA	11.38	2.56	2.64	2.80	2.75	wheat germ agglutinin isolectin 3
Ta.1751.2.S1_x_at	Embryo-specific	PPP	AAA	AAA	AA	AA	10.85	2.97	2.93	3.05	2.89	glutamine synthetase

within the caryopsis during early stages of seed development but is non-specific within the seed and that as development progresses, the expression of this transcript becomes restricted to the crease region. Additional repetitions of the *in situ* hybridizations and further assaying of tissues by microarray or qRT-PCR analysis at earlier time points may help to confirm this observation.

5.3.6 The crease-specific ubiquitin-conjugating enzyme transcript is expressed in the endosperm near the ventral groove at late stages of caryopsis development

In situ hybridization of RNA probes designed to hybridize to the putative ubiquitin-conjugating enzyme-transcript represented by probe set Ta.20064.1.S1_at on the wheat GeneChip, revealed that this transcript is expressed in endosperm cells near the crease region at 12 DPA (Figure 5.7). There was little difference in the overall intensity levels in the caryopsis sections hybridized with the antisense probe compared to those hybridized with the sense probe in the 3 DPA tissue sections. At 6 DPA it appeared as though there might be hybridization of the antisense probe to the crease tissue and endosperm tissue near the nucellar projection, but overall intensity of the sections hybridized with the antisense probe was greater than the sections hybridized with the sense probe, making it difficult to conclude if the observed hybridization was specific or just background. At 9 DPA the situation was similar to 6 DPA with an overall higher intensity of the antisense probed sections compared to sense probed sections but with some specific hybridization in the endosperm near the nucellar projection. At the 12 DPA stage of caryopsis development, there is a visible difference in detection intensity in

Figure 5.7 *In situ* hybridization of a ubiquitin-conjugating enzyme transcript. Cell type-specific labeling at 3, 6, 9 and 12 days post anthesis (DPA) as shown by *in situ* hybridization of transverse sections of triticale caryopsis probed with digoxygenin (DIG)-labeled antisense RNA (left side) and sense RNA (right side) and viewed under Bright-Field Optics. DIG labeled RNA probes were designed from the ubiquitin-conjugating enzyme transcript (EST accession number CA666265) represented by probe set Ta.20064.1.S1_at on the Affymetrix GeneChip® Wheat Genome Array. Tissue types are annotated as follows: crease (cr), embryo sac (es), endosperm (end), epiderm (ep), modified aleurone (ma), nucellar projection (np), nucellus (n), pericarp (pc), pigment strand (ps), vascular strand (vs), and ventral groove (vg). All images are in the same scale. Scale bar = 1 mm.


3 DPA
6 DPA
9 DPA
12 DPA
pc
es
n
np
cr
pc
np
cr
ep
pc
end
np
vs
ep
end
ma
np
ps
vs
vg
Antisense
Sense

the endosperm tissue near the nucellar projection and on either side of the caryopsis ventral groove.

5.3.7 The wheat germ agglutinin isolectin 3 transcript is expressed specifically within the root cap of the embryo

In situ hybridization of the wheat germ agglutinin (WGA) transcript was carried out on sections of embryos sampled at 12 DPA (Figure 5.8). Embryos were sectioned longitudinally in the dorsal plane (Figure 5.8, top) and in the ventral plane (Figure 5.8, bottom). Clear differences were observed in the sections hybridized with the antisense RNA probes compared to the sections hybridized with the sense RNA probes following the detection steps. Detection of expression is distinctly visible in the root cap and epidermis of the embryonic root and the rudimentary primary roots. Expression was not detected in any of the embryonic shoot or scutellum tissue regions. The annotation for the WGA transcript probed is wheat germ agglutinin isolectin 3 which corresponds to the WGA from the B genome.

5.3.8 The glutamine synthetase transcript is expressed in the procambial cells in the embryo and scutellum

In situ hybridization of the glutamine synthetase transcripts was carried out on sections of embryos sampled at 12 DPA (Figure 5.9). Embryos were sectioned longitudinally along the dorsal plane and the ventral plane (Figure 5.9, top and bottom respectively). Detection of expression of the glutamine synthetase transcript by the DIG-labeled antisense RNA probe (Figure 5.9, left side) clearly distinguished that this

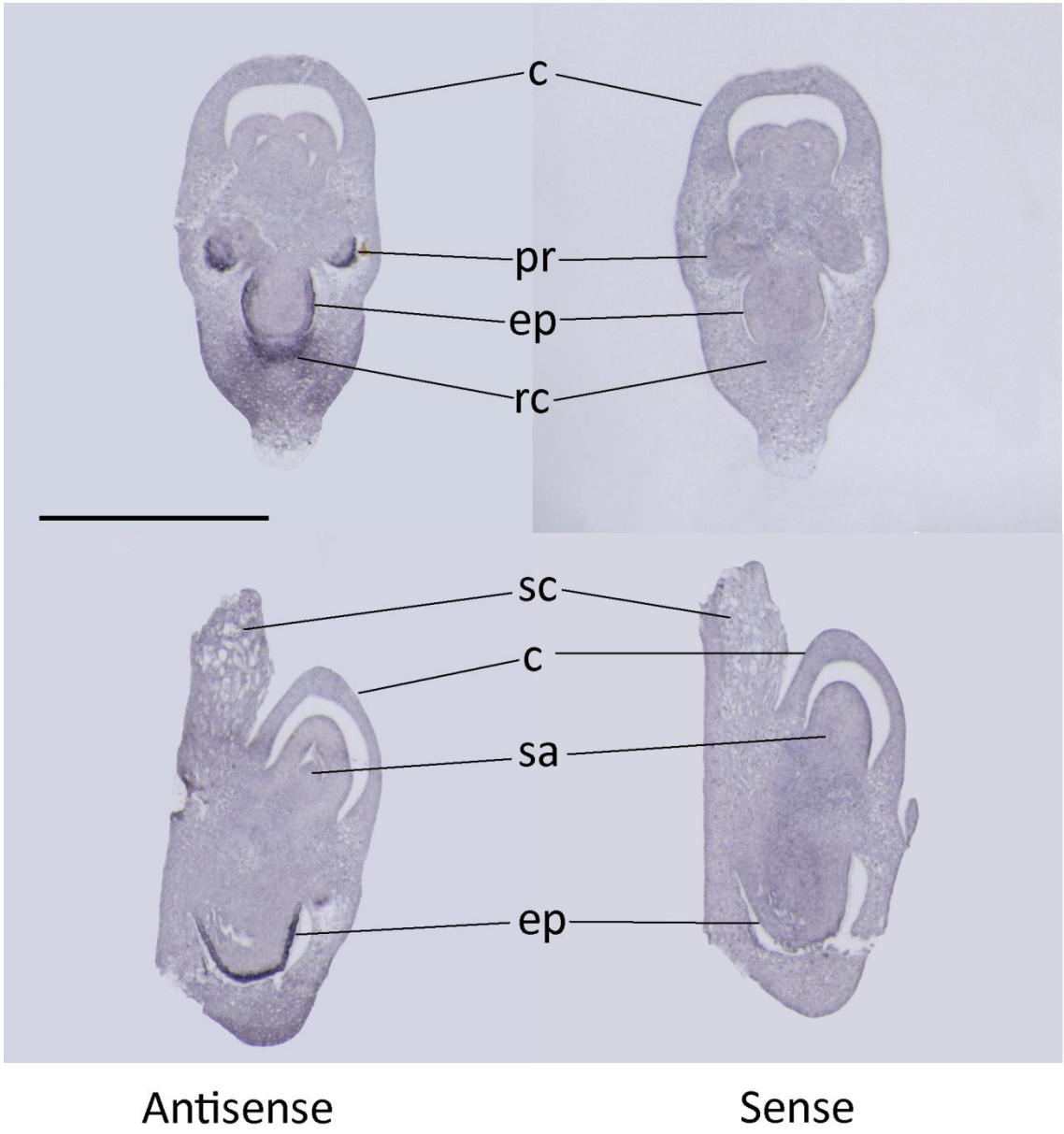


Figure 5.8 *In situ* hybridization of a wheat germ agglutinin transcript. Cell type specific labeling of embryos at 12 days post anthesis (DPA) as shown by *in situ* hybridization of dorsal (top) and ventral (bottom) sections of triticale embryos probed with digoxygenin-labeled antisense RNA (left side) and sense RNA (right side) and viewed under Bright-Field Optics. DIG labeled RNA probes were designed from the wheat germ agglutinin transcript (accession number J02961.1) represented by probe set TaAffx.69777.1.S1_at on the Affymetrix GeneChip® Wheat Genome Array. Tissue types are annotated as follows: coleoptiles (c), root epidermis (ep), primordial root (pr), root cap (rc), shoot apex (sa) and scutellum (sc). All images in the figure are in the same scale. Scale bar = 1 mm.

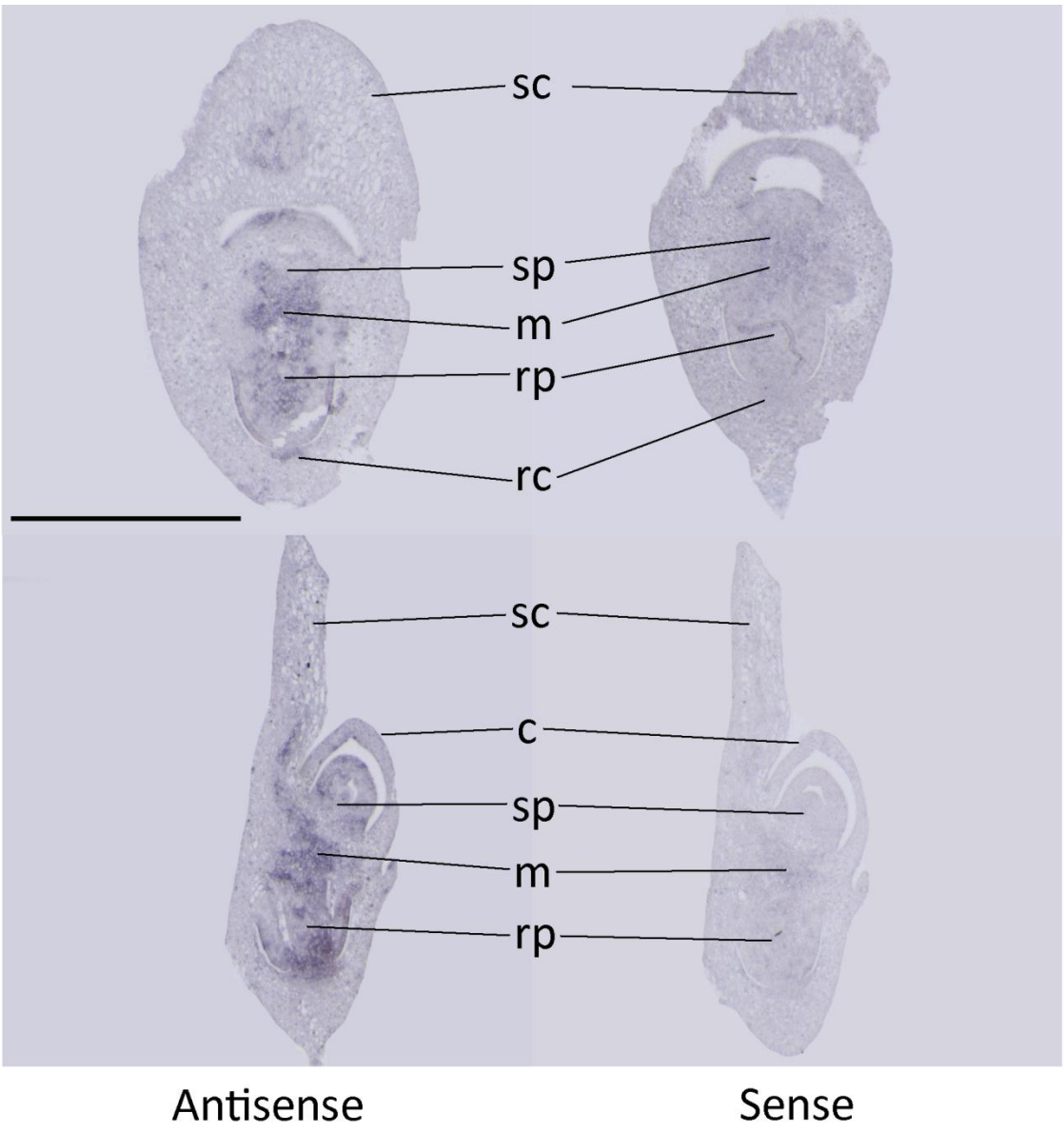


Figure 5.9 *In situ* hybridization of a glutamine synthetase transcript. Cell type-specific labeling of embryos at 12 days post anthesis (DPA) as shown by *in situ* hybridization of dorsal (top) and ventral (bottom) sections of triticale embryos probed with digoxygenin-labeled antisense RNA (left side) and sense RNA (right side) and viewed under Bright-Field Optics. DIG labeled RNA probes were designed from the glutamine synthetase transcript (EST accession number CA611575) represented by probe set Ta.1751.2.S1_x_at on the Affymetrix GeneChip® Wheat Genome Array. Tissue types are annotated as follows: coleoptile (c), mesocotyl (m), root cap (rc), root pole (rp), shoot pole (sp) and scutellum (sc). All images in the figure are in the same scale. Scale bar = 1 mm.

transcript is expressed within the vascular tissues of the scutellum and the embryonic shoot and root. Expression was also visible in the epidermis of the root and primordial leaves. Expression of the glutamine synthetase was shown to be specific to the embryo and root tissues by microarray analysis (Chapter 4). The root tissue used for the microarray analysis was sampled during early root development at 7 days after germination. Expression was not detected by microarray in shoot tissues at early or later stages of development.

5.4 DISCUSSION

I have developed a method for using *in situ* hybridization of RNA transcripts to survey gene expression in triticale at the cellular level throughout seed development. This method of DIG labeled ssRNA probe synthesis is a simple and effective method to generate short and specific probes that can target transcripts of interest. By designing primers to amplify the cDNA templates from regions containing sequence differences, this method can easily be adapted to generate probes that will hybridize to differential regions of transcripts originating from gene families in order to distinguish their expression patterns. We have also shown that these probes can be used to localize transcript expression to cellular regions within the broader context of a tissue. For example expression within the embryo can be localized to a particular cell layer or to the root versus shoot embryonic tissues.

Plants are known for having complex genomes with genome duplications being a common phenomenon in their evolution. The triticale used in this study is an allohexaploid containing two wheat genomes (the A and B genomes) and the rye genome (R genome) in their diploid state. The wheat and rye genomes are both complex genomes containing many large gene families. Triticale, is therefore especially complex, containing sequences from both wheat and rye which may be homologues, orthologues and/or paralogues of the same gene within its genome. Genes that may have diverged between wheat and rye may be reunited in triticale. Because of this complexity, distinguishing between members of a gene family can be problematic. Probe generation for *in situ* hybridization often entails the amplification of a cDNA clone, labeling the complimentary sequence with digoxygenin during reverse transcription followed by

hydrolysis of the labeled product to generate probes of an appropriate length (often between 100 and 200 bases, Darby and Hewitson, 2006). Rather than using an entire cDNA sequence for the generation of *in situ* probes, I propose that probes be designed to be the desired length without hydrolysis and to target the sequence of interest in a region that may differentiate between similar genes as described in the methods above. I have shown this method to be effective at targeting specific members of gene families. Hybridization to the wheat germ agglutinin isolectin B (WGA-B) transcript was specific enough to only hybridize to the root specific WGA-B (Figure 5.4). The protein kinase and glutamine synthetase transcripts analyzed are also members of gene families with diverse expression patterns but we were able to detect the tissue-specific expression of our target transcripts without cross hybridization to other protein kinases and glutamine synthetases in other tissues. Furthermore, this may be a useful technique to determine if there are any subfunctionalization effects of the wheat and rye genome combinations in triticale by performing simultaneous hybridizations in wheat, rye and triticale caryopses of candidate transcripts and observing any differences in cellular expression patterns.

Understanding the regulatory pathways involved in tissue differentiation and development requires the characterization of the genes involved in these pathways. Sequence analysis and annotation through sequence homology can provide partial information about the function of a gene. Many molecular functions can be accurately determined through sequence homology to genes from other species in the absence of experimental information about a gene of interest. While these molecular functions are useful information, they often do not explain the physiological role of the gene product in tissue development and differentiation. Pairing the information gained from sequence

homology searches with expression location information can improve the understanding of how these functions are relevant to the development of a tissue. While microarray and high-throughput sequencing techniques often sample various tissues from a plant, these tissue samples are composites of several cell types, each with their own unique characteristics. Localizing expression of a gene to specific cells within a tissue, is much more informative for determining its potential role in plant tissue development. We have demonstrated the usefulness in improving the spatial resolution of expression patterns observed by microarray analysis through *in situ* hybridization to RNA transcripts. Through this analysis we have proposed putative functions of the transcripts studied in more detail than was provided through previous annotation and expression analysis efforts.

While *in situ* hybridization provides a more detailed view of gene expression patterns, broader global gene expression analysis methods such as microarray analysis, qRT-PCR and/or high-throughput sequencing are useful for identification of candidate transcripts for further investigation. These global methods can also direct further efforts to specific plant organs or tissues that should be analyzed for cellular specific transcript expression by *in situ* hybridization since the tissue preparation methods can be time consuming. Together with data sets in wheat (Drea et al., 2005) and other cereals, and combined with other developmental and genomic resources, our *in situ* results should provide a descriptive and informative framework to understand gene expression timing and to predict gene function. Further characterization of the regulatory sequences of these genes may improve the information about their function by indicating the mode of regulation of expression of their transcripts.

While the cellular localization of transcripts can help to indicate the role of transcripts with little functional information, transcripts with well characterized functions can be used to compare developmental programs in triticale to other cereals. The histone H4 transcript was detected throughout the seed developmental stages examined by *in situ* hybridization. While the histone H4 transcript was used as a positive control for transcript detection by *in situ* hybridization in triticale caryopsis sections, it was also helpful in characterizing the cell cycle activity of the various seed tissues throughout their development. Histone H4 is one of the 5 main histone proteins involved in eukaryotic chromatin structure. The histone H4 transcript can be used as an indicator of cell cycle activity since only cells in the S phase of the cell cycle express histone H4 (Fobert et al., 1994). It was apparent from the results which cell layers within the seed were undergoing active cell proliferation. At early stages of seed development most cell layers showed some histone H4 transcript expression, while at later stages of seed development, tissues including the outer pericarp and epidermis showed very little histone H4 expression indicating that these tissues were not as actively proliferating and had entered a quiescent or even senescent stage of development. The histone H4 transcript expression was used to characterize the pattern of cell proliferation throughout wheat caryopsis development (Drea et al., 2005). It was expected that similar patterns of histone H4 expression would be observed in triticale caryopsis development. In wheat the expression of histone H4 was apparent in most tissues at 3 DPA and declined in pericarp and integuments by 6 DPA similar to what was observed here for triticale. By 9 DPA, the histone H4 expression in wheat was almost totally restricted to the endosperm and abaxial edge of the nucellar projection (Drea et al., 2005). While our results show a

similar pattern of restricted expression we observed that there was still expression in the vascular tissues of the crease at this stage. Expression in the nucellar projection progresses from the abaxial to the adaxial in our results similar to the results of Drea et al. (2005); however, in our results the timing of this progression is slightly delayed when compared to that of wheat, indicative of a prolonged growth phase of the grain due to cell division. The mature grain size of triticale is generally larger than that of commonly grown wheat varieties (Mergoum and Gomez-Macpherson, 2004). The increased grain size and the timing of histone H4 expression suggests that prolonged cell division contributes to the growth of the grain which results in the increased size. Reduction in transcript expression in the endosperm facing the nucellar projection, the cells that become the modified aleurone, was observed in both wheat and triticale, highlighting the importance of early differentiation of these cells to the cereal grain filling process (Thompson et al., 2001).

Identification of the cellular localization of candidate transcript expression can increase or decrease interest in further characterization for crop improvement purposes. The crease is the site where phloem is unloaded from the vascular tissue to the post-phloem transport pathway towards the developing endosperm (Patrick and Offler, 1995). The transcripts analyzed in this area are interesting for the effect that they may have on the regulation of phloem transport, sink strength and ultimately grain yield. The protein kinase transcript localized to the crease vascular tissues is interesting because of its possible role in signaling pathways that control phloem unloading. Protein kinases are involved in a wide range of signaling pathways. Grain filling is a regulated process that requires phloem transport from the flag leaves to the developing caryopsis. It is thought

that there are a number of signals that regulate this process including hormones such as auxin and abscisic acid (Darussalam et al., 1998; Yang et al., 2004) and even sucrose has been shown to act as a signal for phloem unloading into sink tissues (Barker et al., 2000). Based on the location of expression and the putative function of the protein kinase transcript, it is possible that this protein has a role in the signalling of phloem unloading from the crease vascular tissues to the developing endosperm during grain filling.

High levels of enzymes used in the breakdown of sucrose in the sink tissue during the grain filling period, increases the sink strength and also the sink capacity (Yang et al., 2004). In small cereal grains such as wheat and barley, the layer of endosperm cells facing the nucellar projection develop transfer cell morphology to enable solute assimilation from the crease into the endosperm (Thompson et al., 2001). Based on the localization of the ubiquitin-conjugating enzyme transcript in the endosperm transfer cells near the nucellar projection, it may be one of the enzymes involved in the breakdown of sucrose or have some function in the uptake of photosynthates released from the crease and nucellar projection for assimilation into the starchy endosperm. Regulation of this transcript may have an effect on triticale grain yield. The ubiquitin conjugating function of the transcript suggests that the gene product is involved in the degradation of assimilates into compounds for long term storage in the endosperm. The endosperm transfer cells are a modified aleurone layer that may have been removed with the crease tissue during the manual dissections performed for sample preparation for microarray analysis of transcript expression (Chapter 4). This would account for the observed expression in the nucellar projection of the endosperm rather than in the crease as expected from the microarray analysis results (Table 5.5).

Pectinesterase is a ubiquitous cell-wall-associated enzyme that facilitates plant cell wall modifications and subsequent breakdown. Plant pectinesterase is known to be present in several isoforms encoded by a family of genes (Micheli, 2001). It is possible that the microarray cannot distinguish between these isoform transcripts, but that the *in situ* probes that we have designed are specific to pectinesterase that is only expressed in the nucellar projection during early seed development. Pectinesterase functions by either altering the localized pH of the cell wall resulting in alterations in cell wall integrity and contributing to cell wall loosening or by giving rise to block of free carboxyl groups which can create a pectate gel and contribute to cell wall stiffening (Micheli, 2001). The nucellar projection has specialized transfer cells that facilitate the transfer of solutes across the maternal/filial interface to the developing endosperm. These transfer cells develop a unique morphology with a number of cell wall ingrowths that help to facilitate this transfer (Wang et al., 1994; Thompson et al., 2001). Based on the timing and location of expression of the pectinesterase transcript studied, it seems likely that this enzyme is involved in cell wall modification as part of nucellar transfer cell wall differentiation. Regulation of the nucellar-specific pectinase may also affect grain yield as this tissue is part of the pathway from the crease to the endosperm and this enzyme is required for transfer cell development in this tissue to allow adequate phloem transport from the nucellus near the crease to the endosperm cavity (Micheli, 2001; Thompson et al., 2001).

Plant defense is another important focus of many crop improvement programs. Wheat germ agglutinins are plant lectins which are carbohydrate binding proteins. Wheat germ agglutinin is a chitin binding lectin that binds N-acetyl-D-glucosamine specifically

(Van Damme et al., 1998) but can also bind phytohormones including cytokinins, abscisic acid, and gibberellic acid which may serve as a means of regulating free versus stored phytohormones in other tissues (De Hoff et al., 2009). The molecular structure of plant lectins has been well characterized, but their physiological role within the plant is still not well understood. Some plant lectins have displayed anti-pathogenic and anti-predation functions against microbes, insects and herbivores (De Hoff et al., 2009). The ability of WGA to bind to chitin and its synthesis as a propeptide that is secreted (Mishkind et al., 1983) suggests that WGA-B may serve as a mediator of plant defense in the root tip. The wheat germ agglutinin transcript was found to be specific to the developing root tissues within the embryo at 12 DPA by *in situ* hybridization. Expression in the root and coleorhiza were expected based on previous studies of WGA-B expression patterns in embryos and young seedlings observed by *in vivo* labeling and *in situ* hybridization in *Triticum aestivum* (Raikhel et al., 1988). Our observations showed very similar pattern of expression in the embryo. Expression of this transcript was also detected by microarray analysis in triticale root tissue at 7 days after germination (Chapter 4) which also correlates with the root expression in 3 day old seedlings of *Triticum aestivum* (Raikhel et al., 1988). The expression patterns observed appear to confirm the annotation that the transcript is WGA-B. The localization of this transcript to the root tissue and the chitin binding properties of the WGA protein make it a possible candidate for plant defense strategies against soil borne fungal pathogens (De Hoff et al., 2009). Further evaluation of the expression of this transcript at earlier (3, 6 and 9DPA) and later stages of embryo development and throughout root development following

germination, will be helpful in determining its suitability as a candidate for crop improvement.

Higher nitrogen use efficiency is becoming an increasingly important goal for crop improvement due to the need to increase yields in a future that will allow less land and resources to grow our food sources. The regulation of enzymes involved in nitrogen fixation will be important to exploring increases in nitrogen use efficiency. Glutamine synthetase is a nitrogen metabolism enzyme that catalyzes the condensation glutamate and ammonia to form glutamine. Glutamine synthetase genes were recently cloned and studied in wheat (Bernard et al., 2008). This study identified ten glutamine synthetase (GS) sequences in wheat belonging to four sub families: GS2, GS1, GSr and GSe. Interestingly, GSr sequence expression was shown to be confined to vascular cells in leaf and stem tissue by *in situ* hybridization (Bernard et al., 2008). While my results do not show expression of the glutamine transcript in aerial tissues, the expression I observed in triticale embryos by *in situ* hybridization also appeared to be confined to the vascular tissues. *In situ* hybridization analysis of embryos and root tissues at later stages of development would be informative to determining if the confinement to vascular tissues persists. In wheat, GSr expression was detected in root tissue (Zakoks' 12), but also in flag leaf, peduncle and glume (Zadoks' 60, 75 and 75 respectively) by qRT-PCR (Bernard et al., 2008), while we did not detect any expression in aerial vegetative tissues by microarray analysis (Chapter 4). It is possible that this transcript is an embryo and root specific glutamine synthetase. Glutamine synthetase transcripts have been shown to be differentially expressed in wheat and a detailed study of glutamine synthetase transcripts in triticale may highlight if there are differences in the patterns of expression

compared to wheat, perhaps due to a contribution from the rye genome in triticale. This enzyme may be of interest due to its ability to modulate nitrogen metabolism and its potential in nitrogen use efficiency improvement particularly if its expression is root specific. While the transcript was localized to the vascular tissue in the embryo, it was also found to be expressed in the root at early development (Chapter 4), and may be a target for improvement of uptake of soil applied fertilizer during early plant growth making further characterization of this enzyme and its regulation in triticale of interest.

This study has demonstrated the usefulness of *in situ* hybridization of RNA transcript to increase the spatial and temporal resolution of expression patterns observed by other global expression analyses techniques. We have also demonstrated how an improved temporal resolution can reveal developmental regulation of transcript expression. By identifying the cellular localization and timing of transcript expression we have demonstrated how this knowledge can provide information on the putative functions of gene products in tissue development and differentiation. This information can direct hypotheses for further studies into the physiological role of genes in plant development.

5.5 LITERATURE CITED

Barker, L., Kuhn, C., Weise, A., Schulz, A., Gebhardt, C., Hirner, B., Hellmann, H., Schulze, W., Ward, J.M., and Frommer, W.B. (2000). SUT2, a putative sucrose sensor in sieve elements. The Plant Cell **12,** 1153-1164.

Becraft, P.W., and Asuncion-Crabb, Y. (2000). Positional cues specify and maintain aleurone cell fate in maize endosperm development. Development **127,** 4039-4048.

Bernard, S., Møller, A., Dionisio, G., Kichey, T., Jahn, T., Dubois, F., Baudo, M., Lopes, M., Tercé-Laforgue, T., Foyer, C., Parry, M., Forde, B., Araus, J., Hirel, B., Schjoerring, J., and Habash, D. (2008). Gene expression, cellular localisation and function of glutamine synthetase isozymes in wheat (*Triticum aestivum* L.). Plant Molecular Biology **67,** 89-105.

Boodley, J.W., and Sheldrake Jr., R. (1977). Cornell peat-lite mixes for commercial plant growing. New York State College Agricultural Life Science Information Bulletin **43,** 8.

Brown, R., Lemmon, B., and Olsen, O.-A. (1996). Development of the endosperm in rice (*Oryza sativa* L.): Cellularization. Journal of Plant Research **109,** 301-313.

Darby, I.A., and Hewitson, T.D. (2006). *In situ* hybridization protocols. Third edition. (Humana Press).

Darussalam, Cole, M.A., and Patrick, J.W. (1998). Auxin control of photoassimilate transport to and within developing grains of wheat. Functional Plant Biology **25,** 69-78.

De Hoff, P., Brill, L., and Hirsch, A. (2009). Plant lectins: the ties that bind in root symbiosis and plant defense. Molecular Genetics and Genomics **282,** 1-15.

Doan, D.N.P., Linnestad, C., and Olsen, O.-A. (1996). Isolation of molecular markers from the barley endosperm coenocyte and the surrounding nucellus cell layers. Plant Molecular Biology **31,** 877-886.

Drea, S., Leader, D.J., Arnold, B.C., Shaw, P., Dolan, L., and Doonan, J.H. (2005). Systematic spatial analysis of gene expression during wheat caryopsis development. The Plant Cell **17,** 2172-2185.

Druka, A., Muehlbauer, G., Druka, I., Caldo, R., Baumann, U., Rostoks, N., Schreiber, A., Wise, R., Close, T., Kleinhofs, A., Graner, A., Schulman, A., Langridge, P., Sato, K., Hayes, P., McNicol, J., Marshall, D., and Waugh, R. (2006). An atlas of gene expression from seed to seed through barley development. Functional & Integrative Genomics **6,** 202-211.

Esau, K. (1977). Anatomy of Seed Plants. (New York: Wiley).

Fobert, P.R., Coen, E.S., Murphy, G.J.P., and Doonan, J.H. (1994). Patterns of cell division revealed by transcriptional regulation of genes during the cell cycle in plants. The EMBO Journal **13,** 616-624.

Gomez, E., Royo, J., Guo, Y., Thompson, R., and Hueros, G. (2002). Establishment of cereal endosperm expression domains: Identification and properties of a maize transfer cell-specific transcription factor, *ZmMRP-1*. The Plant Cell **14,** 599-610.

Gregersen, P., Brinch-Pedersen, H., and Holm, P. (2005). A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. Transgenic Research **14,** 887-905.

Jiao, Y., Lori T., S., Gandotra, N., Sun, N., Liu, T., Clay, N.K., Ceserani, T., Chen, M., Ma, L., Holford, M., Zhang, H.-y., Zhao, H., Deng, X.-W., and Nelson, T. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nature Genetics **41,** 258-263.

McMaster, G.S., Morgan, J.A., and Wilhelm, W.W. (1992). Simulating winter wheat spike development and growth. Agricultural and Forest Meteorology **60,** 193-220.

Mergoum, M., and Gomez-Macpherson, H. (2004). Triticale improvement and production. FAO plant production and protection paper, **179**.

Micheli, F. (2001). Pectin methylesterases: cell wall enzymes with important roles in plant physiology. Trends in Plant Science **6,** 414-419.

Mishkind, M.L., Palevitz, B.A., and Raikhel, N.V. (1983). Localization of wheat germ agglutinin-like lectins in various species of the *Gramineae*. Science **220,** 1290-1292.

Monjardino, P., Machado, J., Gil, F.S., Fernandes, R., and Salema, R. (2007). Structural and ultrastructural characterization of maize coenocyte and endosperm cellularization. Canadian Journal of Botany **85,** 216-223.

Olsen, O.-A. (2001). Endosperm development: Cellularization and cell fate specification. Annuual Review of Plant Physiology and Plant Molecular Biology **52,** 233–267.

Olsen, O.-A. (2004). Nuclear endosperm development in cereals and *Arabidopsis thaliana*. The Plant Cell Online **16,** S214-S227.

Opanowicz, M., Hands, P., Betts, D., Parker, M.L., Toole, G.A., Mills, E.N.C., Doonan, J.H., and Drea, S. (2011). Endosperm development in *Brachypodium distachyon*. Journal of Experimental Botany **62,** 735-748.

Patrick, J.W., and Offler, C.E. (1995). Post-sieve element transport of sucrose in developing seeds. Australian Journal of Plant Physiology **22,** 681-702.

Raikhel, N.V., Bednarek, S.Y., and Wilkins, T.A. (1988). Cell-type-specific expression of a wheat-germ agglutinin gene in embryos and young seedlings of *Triticum aestivum*. Planta **176,** 406-414.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In Methods in Molecular Biology: Bioinformatics Methods and Protocols, S.A. Krawetz and S. Misener, eds (Totowa, NJ: Humana Press Inc.), pp. 365-386.

Schreiber, A., Sutton, T., Caldo, R., Kalashyan, E., Lovell, B., Mayo, G., Muehlbauer, G., Druka, A., Waugh, R., Wise, R., Langridge, P., and Baumann, U. (2009). Comparative transcriptomics in the *Triticeae*. BMC Genomics **10,** 285.

Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Buell, C.R., de Leon, N., and Kaeppler, S.M. (2011). Genome-wide atlas of transcription during maize development. The Plant Journal **66,** 553-563.

Singh, G., Kumar, S., and Singh, P. (2003). A quick method to isolate RNA from wheat and other carbohydrate-rich seeds. Plant Molecular Biology Reporter **21,** 93a-93f.

Thompson, R.D., Hueros, G., Becker, H.-A., and Maitz, M. (2001). Development and functions of seed transfer cells. Plant Science **160,** 775-783.

Van Damme, E.J.M., Peumans, W.J., Barre, A.J., and Rouge, P.J. (1998). Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Critical Reviews in Plant Sciences **17,** 575-692.

Wan, Y., Poole, R., Huttly, A., Toscano-Underwood, C., Feeney, K., Welham, S., Gooding, M., Mills, C., Edwards, K., Shewry, P., and Mitchell, R. (2008). Transcriptome analysis of grain development in hexaploid wheat. BMC Genomics **9,** 121.

Wang, H.L., Offler, C.E., and Patrick, J.W. (1994). Nucellar projection transfer cells in the developing wheat grain. Protoplasma **182,** 39-52.

Wang, L., Xie, W., Chen, Y., Tang, W., Yang, J., Ye, R., Liu, L., Lin, Y., Xu, C., Xiao, J., and Zhang, Q. (2010). A dynamic gene expression atlas covering the entire life cycle of rice. The Plant Journal **61,** 752-766.

Wegel, E., Pilling, E., Calder, G., Drea, S., Doonan, J., Dolan, L., and Shaw, P. (2005). Three-dimensional modelling of wheat endosperm development. New Phytologist **168,** 253-262.

Yamakawa, H., Hirose, T., Kuroda, M., and Yamaguchi, T. (2007). Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. Plant and Cell Physiology **48,** S62-S62.

Yang, J., Zhang, J., Wang, Z., Xu, G., and Zhu, Q. (2004). Activities of key enzymes in sucrose-to-starch conversion in wheat grains subjected to water deficit during grain filling. Plant Physiology **135,** 1621-1629.

CHAPTER 6 Identification and characterization of three triticale tissue-specific gene promoters

6.1 INTRODUCTION

Cereals are the main agricultural products and provide the main source of carbohydrates in the human diet in Canada. In addition to human food supply, cereals can be used to produce a range of industrial and commercial products. Genetic transformation provides a means of diversifying the range of these products produced from cereals by altering the relative amounts or types of proteins or starches in the cereal grain or plant. Application of this technique can be used to make new or improved processed products. In order to achieve these objectives, promoters directing transgene expression, preferably at high levels, will be required (Goddijn and Pen, 1995).

A promoter is a regulatory region of a gene located upstream of the coding sequence that contains specific sequences recognized by transcription factor proteins to signal transcription initiation (Griffiths et al., 1996). Careful selection of the promoters used to target transgene expression can ensure the necessary level of spatial and temporal control of gene expression. Genes expressed in a tissue-specific manner make excellent candidates for isolating such promoters. Promoters of these genes will allow the appropriate level of gene expression to be established in both a spatial and temporal manner within the plant. A number of potential tissue-specific genes have already been identified in cereals through various methods including transcriptional analysis of tissue extracted RNA (Ma et al., 2005; Druka et al., 2006; Mark et al., 2007; Schreiber et al.,

2009 to name a few), Southern and northern blot analysis (Raikhel et al., 1988), *in situ* hybridization of RNA transcripts (Raikhel et al., 1988) and protein immunolocalization (Mishkind et al., 1983; Raikhel and Pratt, 1987; Rogers et al., 1997). The isolation and characterization of promoters from such genes, and the regulatory motifs within them, should provide suitable promoters for transgene expression in related cereals such as triticale.

Wheat germ agglutinin (WGA) is a cereal lectin belonging to the chitin-binding class of lectins (Van Damme et al., 1998). Cereal lectins are sugar binding proteins that are localized in a tissue-specific manner in embryos and root tips of adult plants (Lerner and Raikhel, 1989). WGA-like lectins are found in wheat (*Triticum aestivum*), rye (*Secale cereale*), barley (*Hordeum vulgare*) and rice (*Oryza sativa,* Mishkind et al., 1983). In hexaploid wheat, there are three unique WGA isolectins (WGA-A, WGA-B and WGA-D) encoded by the respective diploid genomes (Smith and Raikhel, 1989). Wheat lectin is localized in peripheral portions of wheat embryos, coleorhiza and outer layers of the radicle, coleoptile, scutellum and the periphery of the root cap (Mishkind et al., 1983). Hexaploid triticale contains the A and B genomes of wheat and the R genome from rye and therefore likely contains WGA-A and WGA-B isolectins and any rye WGA-like lectins. Protein and nucleotide sequence are available for the wheat WGA; however, no genomic sequence data is available for the rye chitin binding lectin (Van Damme et al., 1998). Only protein and cDNA sequences are available for WGA-like lectins and little is known about the promoters that direct the tissue-specific expression of these proteins.

Aleurain is a barley thiol protease that is closely related to mammalian cathepsin H (Whittier et al., 1987). Aleurain is secreted by the aleurone layer during germination and is primarily an aminopeptidase with poor general endoprotease activity (Holwerda and Rogers, 1992). Aleurain is compartmentalized into separate vacuoles from storage proteins, presumably to protect the storage proteins from active proteases in the acidified aleurain-containing vacuoles (Rogers et al., 1997). Aleurain is a single copy gene in barley although some polyploid species have multiple genes for aleurain homologues (Rogers et al., 1997).

The secalins are seed storage proteins encoded by a family of genes located at the Sec-1 locus on the short arm of rye chromosome 1R (Clarke et al., 1996). The ω-secalin genes are expressed at high levels in the endosperm of rye and wheat-rye translocation lines (Clarke and Appels, 1998). The ω-secalin gene family encodes proteins that are related to the gliadins of wheat and the C-hordeins of barley (Clarke and Appels, 1999). The genes are well characterized in rye and each gene is contained in a 9.2 kb repeated unit (Clarke et al., 1996). The units consist of a 1.1 kb gene and an 8.1 kb spacer sequence separating the genes and 15 of these units are arranged in a head to tail manner. The sequence of these gene units has been studied and compared including a 300 bp portion of the upstream non-coding sequences (Clarke and Appels, 1999). However, there is little information available regarding the majority of the upstream or spacer sequences and possible regulatory elements contained therein.

Genes previously characterized as showing tissue-specific expression in cereal species were chosen from public databases for isolation of promoter sequences in hexaploid triticale (AC Certa) by walking upstream of the transcription start sites. This

technique is aimed at isolating the promoter of the triticale homologue of the chosen gene. The upstream sequences will be investigated as potential candidates for use in transgene experiments in hexaploid triticale. Isolation and characterization of the promoter regions may help shed light on the regulation of the tissue-dependent expression patterns of these genes or their homologues in triticale and other plant species.

Genes showing tissue-specific expression in three seed tissues (embryo, aleurone and endosperm) were chosen as candidates for promoter sequence isolation and characterization. The publicly available coding sequence for WGA isolectin A (GenBank accession number M25536), genomic DNA sequence for barley thiol protease aleurain (GenBank accession number X05167) and genomic DNA rye ω-secalin (GenBank accession number AF000227), expressed in the embryo, aleurone, and endosperm respectively, were selected as candidates for the characterization of their regulatory sequences in the hexaploid triticale (AC Certa) genome. This chapter describes the genome walking approach used to sequence the upstream region of these genes and discusses the characteristics of the identified sequences and their potential for use as promoters for transgene expression. Preliminary results on the ability of upstream promoter fragments from these sequences to direct transient transgene expression in triticale seed and leaf tissues are also presented.

6.2 MATERIALS AND METHODS

6.2.1 Genomic DNA isolation

Genomic DNA (440 µg) was extracted from 5 g of triticale (X *Triticosecale* Wittm. *cv* AC Certa) seedling leaf tissue according to the method described by Sharp et al. (1988). Young leaf tissue was ground in liquid nitrogen. Purified genomic DNA was quantified using UV spectrophotometry at 260 nm.

6.2.2 Genome walking library preparation

Using the extracted genomic DNA, BD GenomeWalker™ Libraries were prepared using the BD GenomeWalker™ Universal Kit (BD Biosciences, Cat. No. 638904). For digestion, 2.5 µg aliquots of genomic DNA were restricted for 18 hours using 80 U of restriction enzyme. Blunt end digests were carried out with each of *Dra*I, *Stu*I, *Eco*RV and *Pvu*II supplied with the kit. Spermidine was added to a final concentration of 0.4 mM to reactions with *Stu*I and *Pvu*II to ensure complete digestion (Bloch and Grossman, 2003). The digested DNA was purified and the BD GenomeWalker™ Adaptor was ligated according to the manufacturer's instruction. The prepared libraries were used for all subsequent genome walking nested PCRs.

6.2.3 DNA walking PCR primers

Primers used for PCR are listed in Table 6.1. AGG358r and AGG110r were designed from the WGA isolectin A sequence (accession no. M25536) in the public database. AGG-706r and AGG-905r were designed from the 5' end of the sequence

Table 6.1 Genome walking primer sequences. Listed primers were used for genome walking PCR and sequencing of cloned genome walking PCR products.

Primer Name	Primer Sequence 5' to 3'
Adaptor-specific primers	
AP1	GTAATACGA CTCACTATAGGGC
AP2	ACTATAGGGCACGCGTGGT
Gene-specific primers	
AGG358r	CAGCAGAGGTTGTTCGGGCATAGCTT
AGG110r	GTACTGGCTGCAGCAGAGGTTGTTGG
AGG-706r	TAGCGTTTCTATTGGAGATGCTCTTAG
AGG-905r	GCGAGATGTTAGATGTAGTTTTGAGAG
CyPr344r	CTCGAGGCTCTCGGAGAAGAT
CyPr294r	CTCTCGTAGCTCTTGCCGTAC
OmSec171r	CTGTGGTTGATAGGGGGAATATTGT
OmSec67r	ATTGCAACTCTTGTTCGCTAGGGTT
Sequencing primers	
TOPO TA M13 Forward	GTAAAACGACGGCCAG
TOPO TA M13 Reverse	CAGGAAACAGCTATGAC
Universal M13 Forward	CGCCAGGGTTTTCCCAGTCACGAC
Universal M13 Reverse	ACGGGATAACAATTTCACACAGGA
3pev566FOR	CGGGAAACGACAATCTGCTA
3pev964REV	GAACTTCAGGGTCAGCTTGC

identified in the first round of genome walking. CyPr344r and CyPr294r were designed from the barley gene for thiol protease aleurain sequence (accession no. X05167). OmSec171r and OmSec67r were designed from the rye ω-secalin gene sequence (accession no. AF000227).

6.2.4 DNA walking

In order to identify sequences upstream of the chosen genes, genome walking was employed. All DNA walking PCR reactions were performed using the BD Advantage™ 2 PCR kit (BD Biosciences, Cat. No. 639206). For walking upstream of WGA-A sequence (GenBank accession no. M25536) PCR using primers AP1 and AGG358r was performed according to the protocols described in the BD GenomeWalker™ Universal Kit manual on the constructed triticale libraries. Secondary nested PCR was performed using primers AP2 and AGG110r. Initially, secondary nested PCR failed to produce a large isolated fragment, so agarose plugs of the largest primary PCR products were taken from products electrophoresed on a 1% agarose gel. The 1 mm by 3 mm plugs were frozen in water at -20°C overnight then thawed and 1 µl was used for secondary PCR. Secondary PCR was performed according to the BD GenomeWalker™ Universal Kit manual with 5 extra cycles (94°C for 25 s and 67°C for 3 min) to increase the amount of product amplified. To increase specificity, another round of secondary PCR was performed using agarose plugs of the largest secondary PCR products and touchdown cycling conditions as described in the BD GenomeWalker™ Universal Kit manual. A fragment approximately 1.3 kb was excised, purified, cloned into pGEM®-T vector (Promega, Cat. No. A3600) and sequenced. A second round of genome walking was

performed using gene-specific primers designed from the 5' end of the sequenced clone insert to get more sequence upstream of wheat germ agglutinin isolectin A. For the second round of genome walking, primary PCR was performed using AP1 and AGG-706r according to the kit manual and secondary PCR was performed using AP2 and AGG-905r according to the kit manual but using the touchdown cycling conditions described in the manual. A fragment approximately 2 kb was gel purified, cloned and sequenced.

For walking upstream of the barley thiol protease aleurain sequence (GenBank accession no. X05167), and rye ω-secalin gene sequence (GenBank accession no. AF000227) primary PCR was performed according to the protocols described in the BD GenomeWalker™ Universal Kit manual on the constructed triticale libraries. Primers AP1 and CyPr344r were used for aleurain and AP1 and OmSec171r for ω-secalin. Secondary nested PCR was performed with the touchdown cycling conditions described in the kit manual using primers AP2 and CyPr294r for aleurain and AP2 OmSec67r for ω-secalin. Fragments approximately 1.7 kb were gel purified from products of each secondary PCR, purified and cloned into pGEM®-T vector (Promega, Cat. No. A3600) for sequencing. Transformants were screened and positive clones were sequenced.

6.2.5 Purification and cloning of PCR products

Selected secondary PCR products were excised after agarose gel electrophoresis using a clean scalpel blade then purified using the QIAquick® gel extraction kit according to the kit manual (QIAGEN, Cat. No. 28706). A-addition reactions were performed on all purified PCR products prior to cloning using the QIAGEN® A-addition kit (QIAGEN,

Cat. No. 231994). Agglutinin A-addition products were cloned into pCR®4-TOPO® vectors supplied with the TOPO TA Cloning® Kit for Sequencing (Invitrogen, Cat. No. 45-0030) and 2 µl of cloning reaction transformed into 50 µl of One Shot® TOP10 Competent cells (Invitrogen, Cat. No. K4575-01) according to the kit manual. Aleurain and ω-secalin A-addition products were cloned into the pGEM®-T vector supplied with the pGEM®-T vector System I (Promega, Cat. No. A3600) according to kit manual instructions with the exception that a 10X ligation buffer (Promega, Cat. No. C1263) was used in the ω-secalin ligation reaction in place of the 2X supplied buffer in order to minimize ligation reaction volumes. Maximum volumes of A-addition products were added to pGEM®-T cloning reactions which resulted in an insert:vector molar ratio of 2:1. 2 µl of pGEM®-T cloning reactions were transformed into 50 µl of Max Efficiency® DH5™ competent cells (Invitrogen, Cat. No. 18258-012).

6.2.6 Sequencing of cloned PCR products

Transformation products of all cloning reactions were plated on Luria Broth (LB) agar with ampicillin (100 µg/ml). Agglutinin transformants were screened via colony PCR using M13 primers supplied with the TOPO TA Cloning® Kit for Sequencing (Table 6.1). Aleurain and ω-secalin transformants were screened via colony PCR using universal M13 primers (Table 6.1). LB broth (5 ml) with ampicillin (100 µg/ml) in a sterile 50 ml disposable tube, was inoculated with transformants positive for an insert and incubated overnight at 200 rpm and 37°C. Plasmid DNA was purified from the overnight cultures using the QIAprep® Spin miniprep kit (QIAGEN, Cat. No. 27104) according to the kit manual and eluted in 50 µl of water.

For sequencing, 0.1 µg per 1000 bp (total size of plasmid and insert) and 3.2 pmol of primer in 12 µl of water were sent to the University of Calgary, University Core DNA (UCDNA) Services Laboratory. Purified agglutinin clone inserts from the first and second round of genome walking were sequenced from both ends using the M13 forward and reverse primers supplied with the TOPO TA Cloning® Kit for Sequencing. Purified aleurain and ω-secalin clone inserts were sequenced from both ends using the universal M13 forward and reverse primers.

Sequencing reactions at the UCDNA laboratory were done using one of the following ABI PRISM Cycle Sequencing Kits: BigDye Terminators Version 3.1 or dGTP BigDye Terminators Version 3.1. Cycling conditions were 96°C for 10 s to denature then 30 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. After cycling, samples were purified using Sephadex G-50 spin columns. Sequencing samples were then dried down and resuspended in 10 µl HiDi Formamide (Applied Biosystems, Cat. No. 4311320), heated to 95°C for 2 min then cooled to 4°C for 2 min. Samples were then placed in a 3730XL DNA Analyzer (Applied Biosystems) for electrophoresis. The UCDNA sequencing lab uses POP-7 Polymer (Applied Biosystems, Cat. No. 4335615) in a 50 cm capillary array and BigDye sequencing buffer with EDTA (Applied Biosystems, Cat. No. 4335613).

6.2.7 Sequence analysis and Identification of *cis*-regulatory elements

Upstream sequences identified using DNA walking were submitted to PlantCARE (Lescot et al., 2002) under 'Search for CARE' to identify *cis*-acting regulatory elements. Upstream sequences were run through the Basic Local Alignment Search Tool (BLAST,

Altschul et al, 1990, http://www.ncbi.nlm.nih.gov/BLAST/) to find regions of similarity to sequences in the public database and in the patent database.

6.2.8 Cloning of full length WGA promoter sequence

The full length WGA and promoter sequence was amplified from triticale genomic DNA using Phusion® High-Fidelity DNA polymerase (New England Biolabs, Cat. No. M0530S). Reaction conditions were 1X Phusion® GC Buffer, 0.5 µM primers (Forward primer 5'-GAATTTCCCAAGAAGACATGAAAT-3', Reverse primer 5'-CCCTGGTGCTCATCATCTTC-3'), 200 µM dNTPs, 3% DMSO and 1 unit of Phusion DNA Polymerase, and 25 ng genomic DNA in a total reaction volume of 50 µl. Cycling conditions were: denaturing at 98°C for 30 s; 35 cycles of 98°C for 10 s, 63°C for 30 s and 72°C for 45 s; a final extension at 72°C for 10 min. Secondary amplification using primary PCR products (10 µL) as template was performed to further amplify the amount of promoter sequence. Reaction conditions and cycling conditions were the same as for the primary PCR. PCR products were purified using the QIAquick PCR purification kit (QIAGEN, Cat. No. 28104) according to the kit manual. To improve cloning efficiency, A-addition reactions were performed on all purified PCR products prior to cloning using the QIAGEN® A-addition kit (QIAGEN, Cat. No. 231994) to ensure they had a single A overhang. PCR products were then ligated into 50 ng of the pGEM®-T vector (Promega, Cat. No. A3600) using a 3:1 insert:vector molar ratio and T4 DNA ligase (3U, Promega, Cat. No. M1801) in 10 µL reactions overnight at 4°C. 2 µl of pGEM®-T cloning reactions were transformed into 50 µl of Max Efficiency® DH5™ competent cells (Invitrogen, Cat. No. 18258-012) and plated onto LB agar plates with ampicillin (100

µg/mL) and incubated at 37°C overnight. Colonies were screened for inserts using standard PCR and universal M13 primers. Positive clones were cultured for plasmid preps using the QIAprep Spin Miniprep Kit (QIAGEN, Cat. No. 27104) according to the kit instructions. Plasmids containing the appropriately sized inserts were sequenced using the universal M13 primers to ensure the promoter sequences matched the expected previously analysed promoters.

6.2.9 Generating expression constructs with promoter deletion fragments

The Gateway® Technology (Invitrogen, Cat. No. 12535-019) was used to generate vectors containing the promoter fragments of interest and transfer these promoter fragments using site-specific recombination into expression vectors containing the green fluorescent protein (GFP) gene for transient expression studies. Primers were designed to amplify promoter deletion fragments of varying lengths from the pGEM-T clones containing the full length promoter sequences. Locations of primer design were chosen based on the location of *cis*-elements that were predicted to play a role in the regulation of seed-specific gene expression. Primers were designed with site-specific attachment sites (*att*B) to allow integration into the donor vector pDONR™221 (Invitrogen, Cat. No. 12536-017) to generate entry clones (Table 6.2). Amplification of promoter deletion fragments with recombination sites was performed by PCR using Phusion High Fidelity DNA polymerase (New England BioLabs Inc., Cat. No. M0530S) and the following reaction conditions: 1X Phusion® GC Buffer, 0.5 µM primers, 200 µM dNTPs and 1 unit of Phusion® DNA Polymerase, and 10 ng plasmid DNA in a total

Table 6.2 Promoter deletion fragment amplification primers for Gateway® cloning. Primers used to amplify promoter deletion fragments with attB sites for Gateway® cloning into the pDONR™221 donor vector (Invitrogen, Cat. No. 12535-019).

Primer Name	Primer Sequence 5' to 3'
Triticain	
Trit-Rev-1-attB	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCGCCGGCGGATGGGTG
Trit-For-218-attb	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAGACAAAGAATAATCCTCAC
Trit-For-576-attb	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTGGTCTTATGGGCTTTGAT
Trit-For-1059-attb	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTGAAAACAAACACCCCCTTA
Trit-For-1228-attb	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGGCTATTGAATGTATGGAG
Trit-For-1348-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTGTCTCTTCCTGTTTGTTTTCAG
Wheat Germ Aggutinin	
WGA-Rev-1-attB	GGGGACCACTTTGTACAAGAAAGCTGGGATCTTCATGCTTTTTGCTTCCTCT
WGA-For-278-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTCAGTGCAGTGCCACGTCTAT
WGA-For-598-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTGCGCGCTATAACTAATGCAG
WGA-For-1087-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGATGGTTCGATTGACTGA
WGA-For-1570-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGACGTTTATTTTGGGATGG
WGA-For-2214-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTATAATTTTGCATGGCAGCAC
WGA-For-2485-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGAGAGCCATTTGAATTTCG
WGA-For-2734-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAATTTCCCAAGAAGACATGAAAT
ω-Secalin	
OmSec-Rev-3-attB	GGGGACCACTTTGTACAAGAAAGCTGGGTTGGATTTGTGGTTACTAATGCTTG
OmSec-For-208-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTGTGCCATCAAACACAACA
OmSec-For-287-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTATCGAGCATGCCTTACAACC
OmSec-For-360-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAATTCCAAAAGAAGGCAAA
OmSec-For-475-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAACAATCTTCTCATTTATTGTGTGA
OmSec-For-1092-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCAGTTTTCCCTCTTCTTG
OmSec-For-1275-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTCCTCACATGGGTCTTCCT
OmSec-For-1523-attB	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTTCAATTGCTAATCAGTGATAAAAC

reaction volume of 20 µl. Cycling conditions were: denaturing at 98°C for 30 s; 35 cycles of denaturing at 98°C for 10 s, annealing at 64°C for 30 s and 72°C for 45 s; a final extension extension at 72°C for 10 min. PCR products with the attB sites were purified using the QIAquick PCR purification kit (QIAGEN, Cat. No. 28104) according to kit instructions then used in Gateway® BP recombination reactions with the pDONR™221 vector according to the Gateway® manual (Version E, Invitrogen). Recombination reaction products were transformed into 50 µl of Max Efficiency® DH5™ competent cells (Invitrogen, Cat. No. 18258-012) and plated onto LB agar plates with kanamycin (50 µg/mL) and incubated at 37°C overnight. Colony PCR using M13 primers was performed to identify the presence of the appropriate size promoter deletion fragments and positive clones were cultured and plasmid DNA was purified using the QIAprep Miniprep Plasmid Purification kit (QIAGEN, Cat. No. 27104) according to the kit manual. Purified entry clones were sequenced to verify the intended promoter fragment was present then used in LR recombination reactions according to the Gateway® manual. The destination vector used was the pLRCWHEAT3-pevAgdv destination vector supplied by the Laroche Lab containing the gateway cassette, the green fluorescent protein gene and the nos terminator (Figure 6.1). LR reaction products were transformed into 50 µl of Max Efficiency® DH5™ competent cells (Invitrogen, Cat. No. 18258-012) and plated onto LB agar plates with ampicillin (100 µg/mL) and incubated at 37°C overnight. Colony PCR using destination vector-specific primers 3pev566FOR (5'-CGGGAAACGACAATCTGCTA-3') and 3pev964REV (5'-GAACTTCAGGGTCAGCTTGC-3') was performed to verify the presence of the appropriate size promoter deletion fragments (promoter deletion fragment length + 399

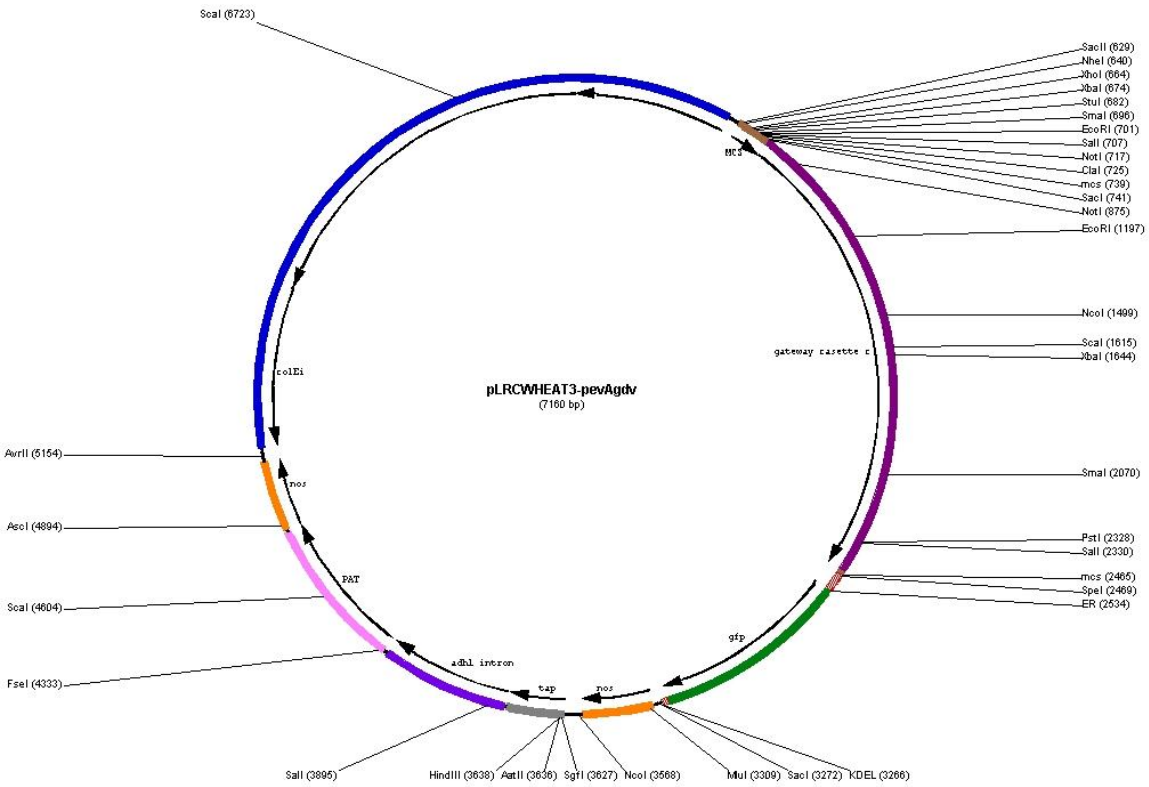


Figure 6.1 Map of Gateway® destination vector pLRCWHEAT3-pevAgdv. This destination vector was designed using the Gateway® Vector Conversion System (Invitrogen, Cat. No. 11828-029) to convert a green fluorescent protein (GFP) expression vector into a Gateway® destination vector. Gateway® cassette c (burgundy) contains *att*R recombination sites flanking a *ccd*B gene and a chloramphenicol-resistance gene blunt-end cloned into the multiple cloning site of pLRCWHEAT3. The vector can be propagated in One Shot® *ccdB* Survival™ 2 T1^R competent cells (Invitrogen, supplied with Gateway® Vector Conversion kit). The pLRCWHEAT3 vector parent contains the following: gfp (green), green fluorescent protein gene; nos (orange), nopaline synthase (nos) terminator; tap (grey), truncated aleurone promoter (a constitutive promoter); adh1 (purple), maize ADH1 intron; PAT (pink), phosphinothricin acetyl transferase (PAT) enzyme gene (glufosinate herbicide resistance gene); colEi (dark blue), pBluescript II KS+ plasmid backbone; MCS (brown), multiple cloning site. Restriction endonuclease sites and their position are indicated on the exterior of the map.

bp destination vector sequence). Positive clones were cultured in 50 mL of LB broth containing ampicilin (100 µg/mL) and plasmid DNA was purified using the QIAGEN CompactPrep Midi Kit (QIAGEN, Cat. No. 12843) according to the kit manual. Purified plasmid DNAs were sequenced using the 3pev566FOR and 3pev964REV primers (Table 6.1) prior to use in bombardments for transient expression studies to ensure the correct promoter deletion fragments were present.

6.2.10 Plasmid controls used in transient expression

Plasmid pBC17 (CaMV35S:ADH1intron:C1:nos:CaMV35S:DH1intron:Bperu: nos) contains cDNAs of C1 (1.1 kb) and B-peru (2.2 kb) genes for anthocyanin production, each under the control of the CaMV35S (0.43 kb) promoter with the maize ADH1 intron (0.6 kb) for constitutive expression in monocots (Chawla et al., 1999). This plasmid was used as a positive control plasmid for co-bombardments with the expression vectors containing the putative seed-specific promoter deletion fragments. Plasmid pLRCWheat1 (actin:gfp:nos), kindly provided by Dr. André Laroche, carries GFP under the control of actin promoter and was used as a positive control for the expression of GFP in each tissue.

6.2.11 Tissue preparation for transient expression

Immature seeds (12 days post anthesis) were carefully removed from the heads of triticale plants and placed into 50 mL tubes without damaging any tissues. In a laminar flow hood seeds were submerged in 70% ethanol and gently agitated 30 s then in 10% bleach (0.475% sodium hydroxide) for 3 min then rinsed 3 times (one min each) with

sterile H_2O to surface sterilize. Seeds were then spread on a sterile absorbent surface in the laminar flow hood to remove excess H_2O. Surface sterilized seeds were dissected and placed onto Murashige and Skoog nutrient mix (MS, Murashige and Skoog, 1962) agar plates for bombardment. Seeds were dissected longitudinally along the ventral groove into halves, laterally across the ventral groove in 3 to 4 mm sections or endosperm halves removed from remaining seed tissue. For leaf tissues, individual leaves were marked and bombarded without removal from the plant.

6.2.12 Cartridge preparation and biolistic procedures

Cartridges were prepared according to the Helios® Gene Gun System instruction manual (Bio-Rad Laboratories, 1996, Cat. No. 1652431 and 1652432) by using 12.5 mg of plasmid DNA from GFP expression vectors containing promoter fragments or 12.5 mg of promoter:GFP expression vectors along with 12.5 mg plasmid DNA of pBC17 construct. Plasmid DNA was added to 25 mg gold particles (1.0 µm) and 0.005 mg/L polyvinylpyrrolidone (mol. wt. 360 000). Seed sections were bombarded using the Helios® Gene Gun set at a delivery pressure of 140 psi under sterile conditions. Bombardment of leaf tissues was done on intact plants using a delivery pressure of 200 psi. Tissues were examined 48 h after bombardment using a stereo microscope under white light for the detection of anthocyanin pigmentation and under blue light (488 nm) with an emission filter (509 nm) to detect GFP. In addition to bombardments with the promoter fragment:GFP expression vectors and co-bombardements with promoter fragment:GFP expression vectors and pBC17, bombardments were also done with

cartridges containing gold particles without any DNA and cartridges without any gold particles or DNA as negative controls.

6.2.13 Imaging

Image capture was carried out using a stereo microscope (Zeiss, SteREO Discovery.V20, Carl Zeiss MicroImaging GmbH, Göttingen, Germany)-mounted charge coupled device color camera (AxioCam, Carl Zeiss) controlled by dedicated software. The control plasmid pLRCWheat1 was used as a control for exposure times during imaging to avoid confusing autofluorescence with positive transient expression in leaf tissues.

6.3 RESULTS

6.3.1 Transient expression of control expression vectors

Bombardments using the control plasmids pBC17 and pLRCWheat1 demonstrated that anthocyanin and GFP expression were clearly visible in seed tissues when under the control of promoters capable of directing transient expression (Figures 6.2 A and B and 6.3). Bombardment controls containing no DNA and controls containing no gold or DNA resulted in only background levels of autofluorescence (data not shown). In all seed bombardments, including controls not using any DNA coated gold particles, higher levels of autofluorescence were observed in embryo and seed coat tissues in comparison to the endosperm tissues (visible in Figure 6.2 A and B and 6.3); however, GFP expression was easily distinguished from background autofluorescence in tissues bombarded with the pLRCWheat1 control plasmid. The pLRCWheat1 control plasmid was therefore used as a positive control for transient expression of GFP in seed tissues. Anthocyanin expression was clearly visible in embryo, endosperm and seed coat tissues (Figure 6.3) and was therefore used as a positive control for tissue penetration by the DNA coated gold particles and bombardment success in seed tissues. GFP expression could not be quantified in seed tissues due to the variability in number of cells showing expression among different cartridges, the variability in fluorescence intensity between cells at different depths relative to the surface bombarded, the variability in number of cells showing expression between different tissues and the variability in background autofluorescence among the different tissues. In cells where both GFP and anthocyanin

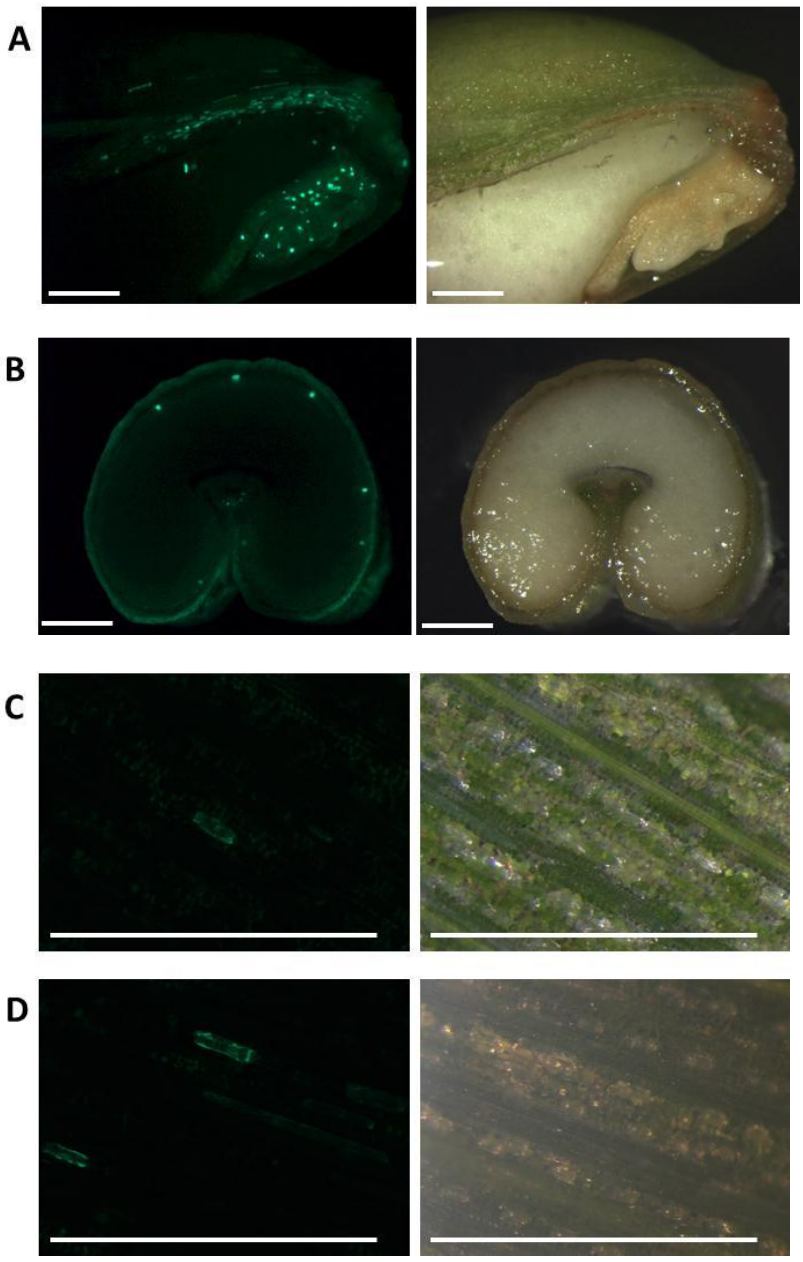


Figure 6.2 Control plasmid pLRCWheat1 expression in triticale tissues. Transient expression of green fluorescent protein (GFP, bright green) directed by plasmid pLCRWheat1 (actin:gfp:nos) was observed under blue light (left) and the same tissues were imaged under white light conditions (right). Seed tissues (A and B) and leaf tissues (C and D) were observed 48 hours after bombardment with the Helios® Gene Gun (BioRad) for transient expression detection. Seeds were dissected either longitudinally along the ventral groove (A) or laterally across the ventral groove (B) and placed on Murshige and Skoog agar plates and then bombarded. Leaf tissue was bombarded without removal from the plant. Scale bars (white) = 1mm. White light images are variable in clarity and color since tissues imaged were focused on the GFP expressing cells under blue light conditions.

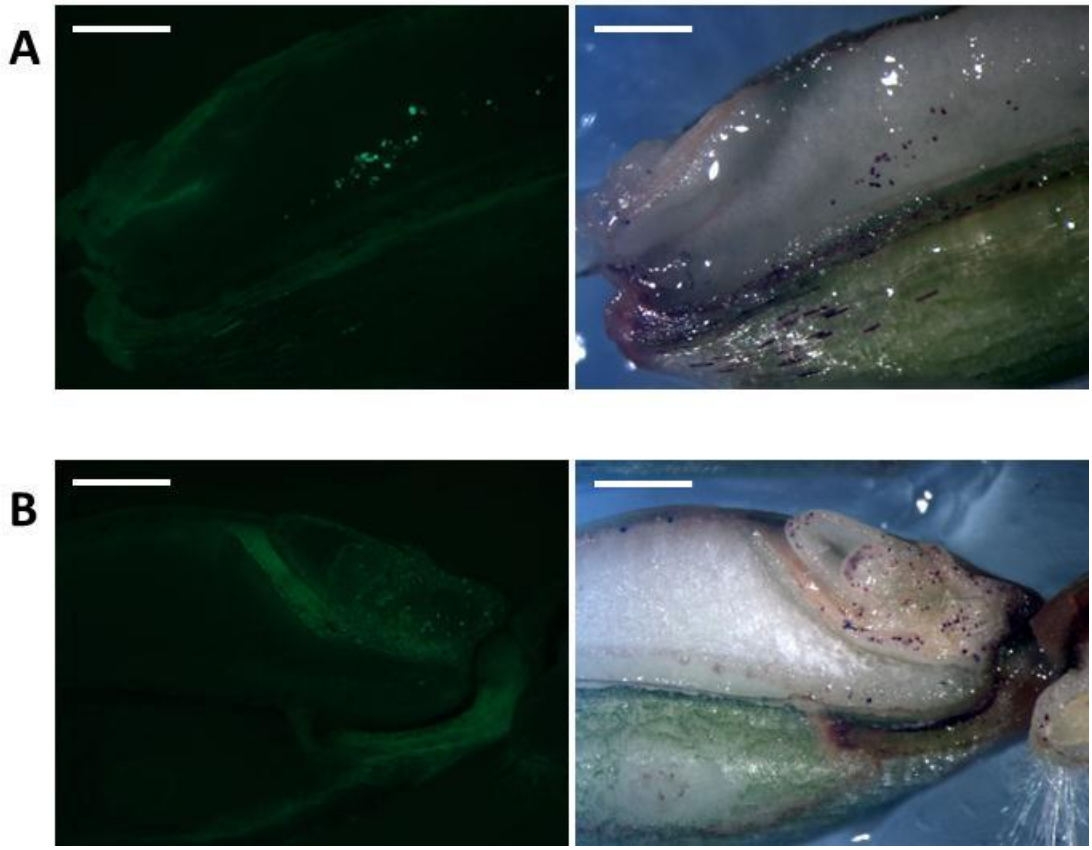


Figure 6.3 Co-bombardment of control plasmids pBC17 and pLRCWheat1 in triticale seed tissues. Transient expression of green fluorescent protein (GFP, bright green) directed by plasmid pLCRWheat1 (actin:gfp:nos) was observed under blue light (left). Anthocyanin (dark purple pigmentation) directed by plasmid pBC17 (CaMV35S:ADH1intron:C1:nos: CaMV35S:ADH1intron:Bperu:nos) was observed under white light conditions (right). Transient expression was observed in endosperm and seed coat tissues (A) and embryo tissues (B) 48 hours after bombardment with the Helios® Gene Gun (BioRad). Seeds were dissected longitudinally along the ventral groove and placed on Murshige and Skoog agar plates and then bombarded. Scale bars (white) = 1mm.

expression were present, the anthocyanin also occludes some of the fluorescence emitted by GFP. For these reasons tissues were screened for presence or absence of GFP expression rather than levels of expression.

In leaf tissues, anthocyanin expression was more difficult to observe due to the increased chlorophyll pigmentation and was therefore not used for assessing bombardment success in leaf tissues. However, the pLRCWheat1 expression vector was also bombarded into leaf tissue and used as a positive control for GFP expression. Transient GFP expression in the leaf tissue using the pLRCWheat1 vector was clearly visible under blue light and clearly distinguishable from autofluorescence (Figure 6.2 C and D).

6.3.2 Triticale wheat germ agglutinin homologue

6.3.2.1 Triticale wheat germ agglutinin homologue promoter sequence

The first round of genome walking generated 1,158 bp of genomic DNA sequence upstream of the AGG110r primer and the second round generated 1,911 bp of genomic DNA sequence. The 3' end of the sequence from the second round of genome walking overlapped with the 5' end of the sequence from the first round for a total of 2,843 bp of genomic DNA sequence upstream of the AGG110r primer site (Figure 6.4). When the sequence was queried using nucleotide BLAST (BLASTn), 109 bp at the 3' end (+1 to +109, Figure 6.4) aligned well with wheat WGA-A sequence (GenBank accession no. M25536) with 87% identity (Figure 6.4, differences shown in grey highlighting). The same 3' region of identified sequence also aligned to WGA-D (87% identity) and to

```
                                                  O2 site
-2734                              GAATTTCCCAAGAAGACATGAAATTCATAATGTT

-2700 TTTCAAATAGCAATTCAGTTTTTATTCGTCGAATATACATCCATGCTCAAGTTGTCCTTC

       O2 site
                Skn-1 motif
              GCN4 motif
-2640 CGATCATGTGGTGAGTCATGACACATGTTGAGTAGCAAAGGATCAGGCTTCAAATAAGAT
                      GCN4 motif
                        Skn-1 motif

-2580 CTGCTACACATGTACTGATGATTGTCAAGAGATGCAACAATCTGAGGATGGTAGTATTTA

-2520 GAGAGGTGGAACACTTGATTAAGAGATAGTAAATTGGAGAGCCATTTGAATTTCGATTCT

-2460 TGAAGAATAGAACAAAGCATGACCACACATACACAAGATTTAAGTAGCCAAACTAGGCAA
                      Skn-1 motif

-2400 AATGCTTTGGTGCCATAGGATGTAGGACCACAAAAATTTAGTTAGTTTGACAATGACAAA
                                                          Skn-1 motif

-2340 GCGGGAAGGCGCAACACTGGTGCAAAGTACATTAAGAAAAGTAATTCTCATCAAAAATAT

-2280 TGATGGAAGAATGAATACTTTTGGACAAGGAGGCAGGATATAATGCACAGATAATCACCA

-2220 AAGCCTATAATTTTGCATGGCAGCACAAACGGTGATAAAAGAAGTTACTATATGAAATAA

-2160 ATTTGCCATTAAATATAAGTACTTTGGACTAAAGCCTAAAGATACACTTAAATTGGAAGT

-2100 TTATTTTATCAAAATTTACCATGTTTCTTTGAGATGCATCTTTGGCTACAATGTTATATG

-2040 TATATTGATTGGCACATAAAATCCAAGCTAGAGATAATATATCCAATAATAGTTAACAAT

-1980 ACATGGCTGAAAATGATAGGAAAGGAAAGCATACTCACAATTCTCTGATTGGGCCAAGAC
```

Figure 6.4 Triticale homologue of wheat germ agglutinin (T-WGA) sequence. Sequence of WGA isolectin A (WGA-A, GenBank accession no M25536, nucleotides 109 to 968) with upstream sequence identified in triticale (AC Certa) by genome walking is shown with putative features including; translation start codons (at nucleotides 1 and 4 boxed), TATA box (red nucleotides) CAAT boxes (green nucleotides). Positive numbers on the left refer to the nucleotide position beginning with the putative translation start codon. Negative numbers (also on the left) refer to nucleotide position of genomic sequence upstream of the coding sequence beginning with the putative start codon. Nucleotides highlighted in yellow are genome walking gene-specific primer complementary regions. Some cis-acting elements identified using PlantCARE (Lescot et al., 2002) are named and are shown by ____ above sequence if element is in + strand (shown) and by ____ below sequence if element is in – strand (complementary). Nucleotides highlighted in grey denote differences between sequence identified by genome walking and WGA-A sequence with the nucleotides from WGA-A above the T-WGA sequence. Lower case sequence represents the translated regions. The translation stop codon (at 631) is underlined. Light blue nucleotides are sequence that is encompassed by an existing patent (US patent # 7214786, Kovalic et al., 2007).

-1920 CTTTATCTAACATGTGTGCTTATTTGTACAACTCTGCACCTAGTCAACTAGACAGTAGAG
AACA

-1860 TTTGTTAGCTCTTTTTACCCTCTATATTTAACATTTAACTAAAAATGCAGGGAAAATATG
motif

-1800 CAGTATGCCAAGACCTACTGTTTGTTTCGCCTCAAATGTATTTTTTAGTTACCAATATGA
Skn-1 motif

-1740 CAACACATAATACATCTAGACACACAAATGAGCGGGTAATCCAAGCAAATCCATGTTAAA

CCGTCC box
-1680 AGCAATCCCGGCAAAACAAAGTCTTGGGCATAGTAATAATATGCTCCCTCCGTCCCAAAA

-1620 TAAGTGTCGCTGATTTAGTACAAACTTCTACTAACTTTGCACTAAATCAACGACATTTAT

-1560 TTTGGGATGGAAGGAGTAGTAAGCATTGAAAATATTTGTCTGTAGAGTAAACCAAAGAGA

-1500 TGAATATTTTGAACCCATTAATTTGAGAGTACTTGCTATGATACTCACACTACAGGAGGT

-1440 CTTAACTCGAGCTTATATTGGTGAACCAACATGTGGTTGAATGATTAGAAGTACAGTGGT

-1380 ATTCCCAGCCCACCAGAATTTAAGTATTTTCTGGATTTATTTCAAGCCTTTCGACAATGT

-1320 GCTCTTAGTAGGAGGAGACGTTCTCGCCAACTACGAAGGTGTCTGTAATGACTTCGTCAA
Skn-1 motif
O2 site
-1260 TTTCAAGATGATGTGCTCGCTCAGTCTCTTGAAGGTGATCATAGATGTAGGGAGTATCTG

-1200 TGTGTATATGGGCATTTGCGTCTGTATTATATTAAAAAAAACTCGAGCTTATATTCAATT

-1140 GTACTCTAGCCAATTACTCTCCCGAGAAAAAAAAATTGTAGCCAACTACTACTTCGATGG

-1080 TTCGATCGACTGAACCAAGATGTGGATCAGGCCTGCCTTACCATCGCAATATTCCTAACA

CAT box
-1020 ACCGACCATTTTTCCTAACAAAGGATATCCGGTGCGAGATGCCACTGCCATTTTTTCCAT

RY element
-960 TAACTGGTCACGTGCATGCATTGCTCTCCTCTCCATGCATGTCACGCCCACACGACTCTC

-900 AAAACTACATCTAACATCTCGCATCATAGTACTTGCTCTCATGAGAAGAAAAAGAAGATA
Py-rich stretch

-840 AGATGATCAATATCATGGACGCAGAGCACAACGTTCGCATCGATCAGATGCTATGCACGC

-780 GACGTGTCCGTTGGGGTGCTTCTGTGGATGGGGACGACGAGGTGTAATCACATGTGTATG

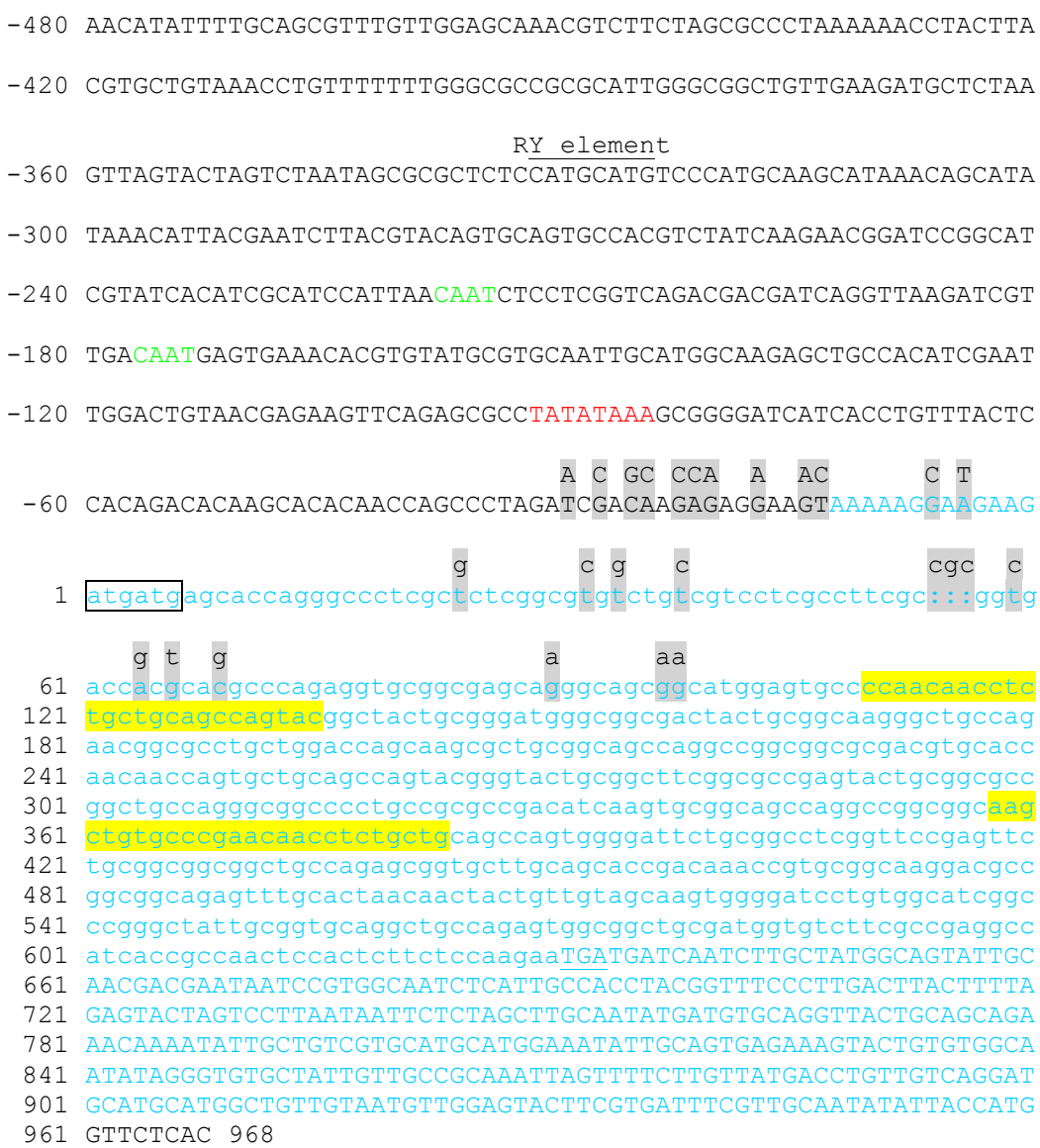
-720 CGTACGTATTAGTTCTAAGAGCATCTCCAATAGAAACGCTATATGAGCGTCGCGCGGTAT

OCT
-660 AAAAACCTATTTTAGCGCGCGCGGATCCGGTTTGAAGTCTCCAACGGCCGCACAAAAAAC

-600 GCGCGCGCTATAACTAATGCAGCGCGCGCAGCATATTTAGTGCGTCCGGTCGCGCGCGCT

-540 GCAAACTCTGGACTGCTGCAGATGAGACCGCTTGATTGGGACCGTTGGACTCGCGCGCAC

Figure 6.4 Continued



barley root-specific lectin mRNA (90% identity). Interestingly the differences between the identified sequence and WGA-A at positions 90, 97 and 98 in the identified sequence correspond to nucleotides in WGA-B (Smith and Raikhel, 1989). The sequence for WGA-B published by Smith and Raikhel (1989) is not complete and does not cover the remaining sequence differences between the identified genomic DNA sequence and WGA-A but the similarities between the identified sequence and the WGA-B sequence indicate that the identified sequence may be a triticale homologue of WGA-B. A total of 2,734 bp of the identified sequence was upstream of a putative ATG start codon.

Two different possible ATG codons were identified by consensus to GenBank sequence M25536 (boxed nucleotides, Figure 6.4). Although several TATA boxes were identified in upstream sequence using PlantCARE, only one TATA box was within 300 bp of the proposed START codon, at position -14 bp, but on the negative strand which may be downstream of, or too close to, the transcription start site for transcription of the triticale WGA homologue (Table 6.3). There is however, a TATA box-like region between -92 and -84 bp that was not identified using PlantCARE (red nucleotides, Figure 6.4). In a blast of the identified triticale sequence against a triticale RNA library generated from RNA-seq data the transcriptional start site appeared to be at nucleotide -8 (blast data not shown, Xu et al., 2011). This would make the TATA box-like region at nucleotides -92 to -84 the most likely location for the TATA box in the core promoter. Several CAAT boxes were also identified within 300 bp of the start codon (Table 6.3), however only two were on the positive strand (-177 bp, -218 bp, green nucleotides, Figure 6.4).

Several *cis*-acting regulatory elements were identified using PlantCARE (Table

Table 6.3 Triticale wheat germ agglutinin homologue promoter sequence elements. *Cis*-elements and motifs identified using PlantCARE (Lescot et al., 2002) in the promoter sequence upstream of wheat germ agglutinin (GenBank accession no. 25536) ATG start codon identified in triticale (AC Certa) through genome walking.

Site Name	Organism	Position	Strand	Matrix score.[a]	sequence	function
A box	*Petroselinum crispum*	-1631	+	6	CCGTCC	cis-acting regulatory element
AACA motif *	*Oryza sativa*	-1865	-	11	TAACAAACTCCA	involved in endosperm-specific negative expression
AE box	*Arabidopsis thaliana*	-2080	-	8	AGAAACAT	part of a module for light response
3-AF1 binding site	*Solanum tuberosum*	-26	+	10	TAAGAGAGGAA	light responsive element
ARE	*Zea mays*	-1512	-	6	TGGTTT	cis-acting regulatory element essential for the anaerobic induction
ABRE	*Arabidopsis thaliana*	-164	+	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Arabidopsis thaliana*	-263	-	7	ACGTGGC	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Arabidopsis thaliana*	-422	+	6	TACGTG	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Hordeum vulgare*	-782	+	9	GCAACGTGTC	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Arabidopsis thaliana*	-952	+	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
as-2 box	*Nicotiana tabacum*	-844	+	9	GATAatGATG	involved in shoot-specific expression and light responsiveness
ATCT motif	*Zea mays*	-198	-	9	AATCTGATCG	part of a conserved DNA module involved in light responsiveness
ATCT motif	*Zea mays*	-800	-	9	AATCTGATCG	part of a conserved DNA module involved in light responsiveness
Box I	*Pisum sativum*	-2700	+	7	TTTCAAA	light responsive element
Box 4	*Petroselinum crispum*	-1484	+	6	ATTAAT	part of a conserved DNA module involved in light responsiveness
CAAT box	*Arabidopsis thaliana*	-122	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-123	-	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-218	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-149	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-177	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-218	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-387	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-506	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-693	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-694	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-833	+	4	CAAT	common cis-acting element in promoter and enhancer regions

* element/motif shown in Figure 6.4

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.3 Continued

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
CAAT box	*Hordeum vulgare*	-941	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1034	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1106	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-1107	-	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-1130	+	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-1131	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1143	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-1144	-	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-1145	+	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1187	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-1263	+	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1326	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-1424	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1480	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1527	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1536	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1678	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1716	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1748	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-1749	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1768	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1899	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-1943	+	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1984	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1997	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-1998	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-2033	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-2037	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-2052	+	4	CAAT	common cis-acting element in promoter and enhancer regions

* element/motif shown in Figure 6.4

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.3 Continued.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
CAAT box	*Arabidopsis thaliana*	-2109	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-2110	-	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-2160	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-2282	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-2350	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-2477	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-2488	-	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-2489	-	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-2543	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-2560	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-2590	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-2690	+	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-2697	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAT box*	*Arabidopsis thaliana*	-980	+	6	GCCACT	cis-acting regulatory element related to meristem expression
CCAAT box	*Hordeum vulgare*	-499	-	6	CAACGG	MYBHv1 binding site
CCAAT box	*Hordeum vulgare*	-619	+	6	CAACGG	MYBHv1 binding site
CCAAT box	*Hordeum vulgare*	-773	-	6	CAACGG	MYBHv1 binding site
CCGTCC box*	*Arabidopsis thaliana*	-1631	+	6	CCGTCC	cis-acting regulatory element related to meristem specific activation
CGTCA motif	*Hordeum vulgare*	-1266	+	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
G box	*Zea mays*	-1413	+	9	GACATGTGGT	cis-acting regulatory element involved in light responsiveness
G Box	*Pisum sativum*	-165	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
G box	*Arabidopsis thaliana*	-165	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
G box	*Brassica napus*	-166	+	8	ACACGTGT	cis-acting regulatory element involved in light responsiveness
G box	*Zea mays*	-267	+	6	CACGTC	cis-acting regulatory element involved in light responsiveness
G Box	*Antirrhinum majus*	-422	-	6	CACGTA	cis-acting regulatory element involved in light responsiveness
G box	*Daucus carota*	-422	+	6	TACGTG	cis-acting regulatory element involved in light responsiveness
G box	*Zea mays*	-780	-	6	CACGTC	cis-acting regulatory element involved in light responsiveness
G box	*Zea mays*	-910	+	6	CACGAC	cis-acting regulatory element involved in light responsiveness
G Box	*Pisum sativum*	-952	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
G box	*Arabidopsis thaliana*	-952	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness

* element/motif shown in Figure 6.4

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.3 Continued.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
GAG motif	*Hordeum vulgare*	-699	-	7	GGAGATG	part of a light responsive element
GAG motif	*Spinacia oleracea*	-1505	+	7	AGAGATG	part of a light responsive element
GAG motif	*Spinacia oleracea*	-2553	+	7	AGAGATG	part of a light responsive element
GATA motif	*Arabidopsis thaliana*	-846	+	10	AAGATAAGATT	part of a light responsive element
GATA motif	*Solanum tuberosum*	-2603	+	9	AAGGATAAGG	part of a light responsive element
GATA motif	*Arabidopsis thaliana*	-1966	+	7	GATAGGA	part of a light responsive element
GC motif	*Oryza sativa*	1100	+	9	CGCCGCGCA	?
GCC box	*Arabidopsis thaliana*	-383	-	7	AGCCGCC	
GCN4 motif *	*Oryza sativa*	-2629	+	7	TGAGTCA	cis-regulatory element involved in endosperm expression
GCN4 motif *	*Oryza sativa*	-2622	-	7	TGTGTCA	cis-regulatory element involved in endosperm expression
GT1 motif	*Arabidopsis thaliana*	-187	+	6	GGTTAA	light responsive element
I box	*Triticum aestivum*	-846	+	9	aAGATAAGA	part of a light responsive element
MBS	*Zea mays*	-209	+	6	CGGTCA	MYB Binding Site
MBS	*Arabidopsis thaliana*	-960	+	6	TAACTG	MYB binding site involved in drought-inducibility
MBSII	*Petunia hybrida*	-2368	+	11	AAAAGTTAGTTA	MYB binding site involved in flavonoid biosynthetic genes regulation
MSA-like	*Catharanthus roseus*	-500	-	9	TCCAACGGT	cis-acting element involved in cell cycle regulation
MSA-like	*Catharanthus roseus*	-776	-	8.5	(T/C)C(T/C)AACGG (T/C)(T/C)A	cis-acting element involved in cell cycle regulation
OCT*	*Arabidopsis thaliana*	-641	+	8	CGCGGATC	cis-acting regulatory element related to meristem specific activation
O2 site*	*Zea mays*	-1254	+	9	GATGATGTGG	cis-acting regulatory element involved in zein metabolism regulation
O2 site*	*Zea mays*	-2639	+	9	GATGATGTGG	cis-acting regulatory element involved in zein metabolism regulation
O2 site*	*Zea mays*	-2723	+	9	GATGACATGA	cis-acting regulatory element involved in zein metabolism regulation
RY element *	*Helianthus annuus*	-927	+	8	CATGCATG	cis-acting regulatory element involved in seed-specific regulation
RY element *	*Helianthus annuus*	-332	+	8	CATGCATG	cis-acting regulatory element involved in seed-specific regulation
Skn-1 motif*	*Oryza sativa*	-1273	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif *	*Oryza sativa*	-1744	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif *	*Oryza sativa*	-2348	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif *	*Oryza sativa*	-2442	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif *	*Oryza sativa*	-2623	-	5	GTCAT	cis-acting regulatory element required for endosperm expression

* element/motif shown in Figure 6.4

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.3 Continued.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
Skn-1 motif *	*Oryza sativa*	-2626	+	5	GTCAT	cis-acting regulatory element required for endosperm expression
Sp1	*Oryza sativa*	-384	+	6	GGGCGG	light responsive element
TATA box	*Lycopersicon esculentum*	-14	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-302	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Brassica oleracea*	-303	+	6	ATATAA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-347	+	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-435	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-592	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-651	+	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-661	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-663	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-681	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-714	-	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1151	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1152	-	5	TATAA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1168	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-1171	-	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1173	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1174	-	5	TATAA	core promoter element around -30 of transcription start
TATA box	*Brassica napus*	-1175	+	6	ATTATA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-1176	-	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1196	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Helianthus annuus*	-1198	-	6	TATACA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1427	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1428	-	5	TATAA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-1644	+	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-1647	+	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1684	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-1733	+	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1758	+	5	TTTTA	core promoter element around -30 of transcription start

* element/motif shown in Figure 6.4

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.3 Continued.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
TATA box	*Lycopersicon esculentum*	-1821	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1848	+	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1848	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1921	-	9	taTATAAAgg	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2003	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Brassica napus*	-2004	+	6	ATATAT	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-2006	+	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-2024	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2040	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Helianthus annuus*	-2042	-	6	TATACA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2046	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2047	-	5	TATAA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-2097	+	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2147	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Brassica oleracea*	-2148	+	6	ATATAA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2172	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-2185	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2215	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Daucus carota*	-2217	+	9	ccTATAAATT	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2242	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Brassica oleracea*	-2243	+	7	ATATAAT	core promoter element around -30 of transcription start
TATA box	*Nicotiana tabacum*	-2526	-	9	tcTATAAAta	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-2667	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-2681	+	5	TTTTA	core promoter element around -30 of transcription start
TC-rich repeats	*Nicotiana tabacum*	-971	+	9	ATTTTCTCCA	cis-acting element involved in defense and stress responsiveness
TCA element	*Brassica oleracea*	-858	+	9	GAGAAGAATA	cis-acting element involved in salicylic acid responsiveness
TCA element	*Brassica oleracea*	-1504	+	9	GAGAAGAATA	cis-acting element involved in salicylic acid responsiveness
TCCACCT motif	*Petroselinum hortense*	-2517	-	7	TCCACCT	
TCT motif	*Arabidopsis thaliana*	-287	+	6	TCTTAC	part of a light responsive element

* element/motif shown in Figure 6.4

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.3 Continued.

Site Name	Organism	Position	Strand	Matrix score.[a]	sequence	function
TGACG motif	*Hordeum vulgare*	-1266	-	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGA element	*Brassica oleracea*	-1572	+	6	AACGAC	auxin-responsive element
5UTR Py-rich stretch *	*Lycopersicon esculentum*	-859	-	9	TTTCTTCTCT	cis-acting element conferring high transcription levels
Unnamed 1	*Zea mays*	-268	-	5	CGTGG	
Unnamed 3	*Zea mays*	-268	-	5	CGTGG	
Unnamed 4	*Petroselinum hortense*	-63	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-214	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-335	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-458	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-622	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-696	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-930	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-935	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1122	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1210	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1307	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1310	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1446	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1548	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1637	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-2252	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-2485	-	4	CTCC	
WUN motif	*Brassica oleracea*	-297	+	9	TCATTACGAA	wound-responsive element

* element/motif shown in Figure 6.4

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

6.3). These include some elements known to be involved in seed-specific regulation such as the AACA motif (Takaiwa et al., 1996), the GCN4 motif (Muller and Knudsen, 1993), the O2 site (Carlini et al., 1999), the RY element (Ezcurra et al., 1999) and the Skn-1 motif (Takaiwa et al., 1991) of which 1, 2, 3, 2, and 6 motifs were identified, respectively (Figure 6.4, Table 6.3). Some of the other elements identified are involved in light response, stress response, hormone response, MYB transcription factor binding, and cell-cycle regulation. Three elements (CAT box, -980 bp, CCGTCC box, -1631 bp and OCT, -641 bp) related to meristem-specific activation were identified. A Py-rich stretch (-844 bp), an element conferring high levels of transcription (Daraselia et al., 1996), was also identified, although on the negative strand and in a position much further upstream than that previously shown to affect transcription levels. Numerous elements were identified throughout the sequence upstream of the translation start site and seed-specific regulatory elements were identified more than 2.6 kb upstream of the putative ATG start codon.

The majority of the promoter sequence identified had no similarity to other sequences in the public databases including patents. Only 13 bp upstream of the putative start codon and the sequence downstream of the ATG start codon to position 960 aligned to sequences in the patent database (light blue nucleotides, Figure 6.4, US patent number 7214786, Kovalic et al., 2007). Therefore we have identified a novel promoter sequence for a triticale homologue of WGA (T-WGA) without patent infringements that may be useful for further studies in triticale and possibly other cereals and even other plant families.

6.3.2.2 Transient expression using triticale wheat germ agglutinin homologue promoter deletion fragments

Seven T-WGA promoter deletion fragments of decreasing lengths were cloned into the GFP expression vector pLRCWHEAT3-pevAgdv and bombarded into triticale seeds dissected longitudinally along the ventral groove and leaf tissues attached to the plant using the Helios® Gene Gun. Tissues were observed 48 hours after bombardment for transient expression of GFP (Table 6.4). Two of the T-WGA promoter fragments (-272 and -2,734 bp fragments) were co-bombarded with the anthocyanin expression vector pBC17 (Figure 6.5). Anthocyanin was clearly visible in the seed tissues following co-bombardment of pBC17 with WGA-272:GFP and WGA-2734:GFP plasmids indicating successful incorporation of the expression vectors into the plant cells using the biolistic approach (Figure 6.5). Anthocyanin expression was observed in all seed tissues in the co-bombardments (Figure 6.5) while GFP expression was variable among the seed tissues dependent upon the T-WGA promoter deletion fragment used (Table 6.4).

Transient expression of GFP was observed under the regulation of all T-WGA promoter fragments in the embryo (Table 6.4). This suggests that the embryo expression is controlled by regulatory elements fairly proximal to the transcriptional start site. The smallest T-WGA promoter fragment (-278 to +1) was able to direct transgene expression in all three seed tissues evaluated. While there is the putative TATA box and putative CAAT box within this region of the promoter, this fragment does not appear to be a minimal promoter capable of ubiquitous transgene expression as there was no expression observed in leaf tissue using this promoter deletion fragment (Table 6.4). Using the -598 promoter deletion fragment, transient GFP expression was observed in the embryo but

Table 6.4 Triticale wheat germ agglutinin homologue (T-WGA) promoter deletion fragment transient expression results. Expression vectors containing T-WGA promoter deletion fragments and the green fluorescent protein (GFP) gene were bombarded into plant tissues using the Helios® Gene Gun (Bio-Rad Laboratories). Tissues were then observed for GFP expression under blue light using a stereo microscope. (+) presence of GFP expression, (-) absence of GFP expression.

WGA Promoter Fragment	Embryo	Endosperm	Seed Coat	Leaf
-278 to +1	+	+	+	-
-598 to +1	+	-	-	-
-1087 to +1	+	+	+	-
-1570 to +1	+	+	-	-
-2214 to +1	+	+	-	-
-2485 to +1	+	-	-	-
-2734 to +1	+	+	+	-

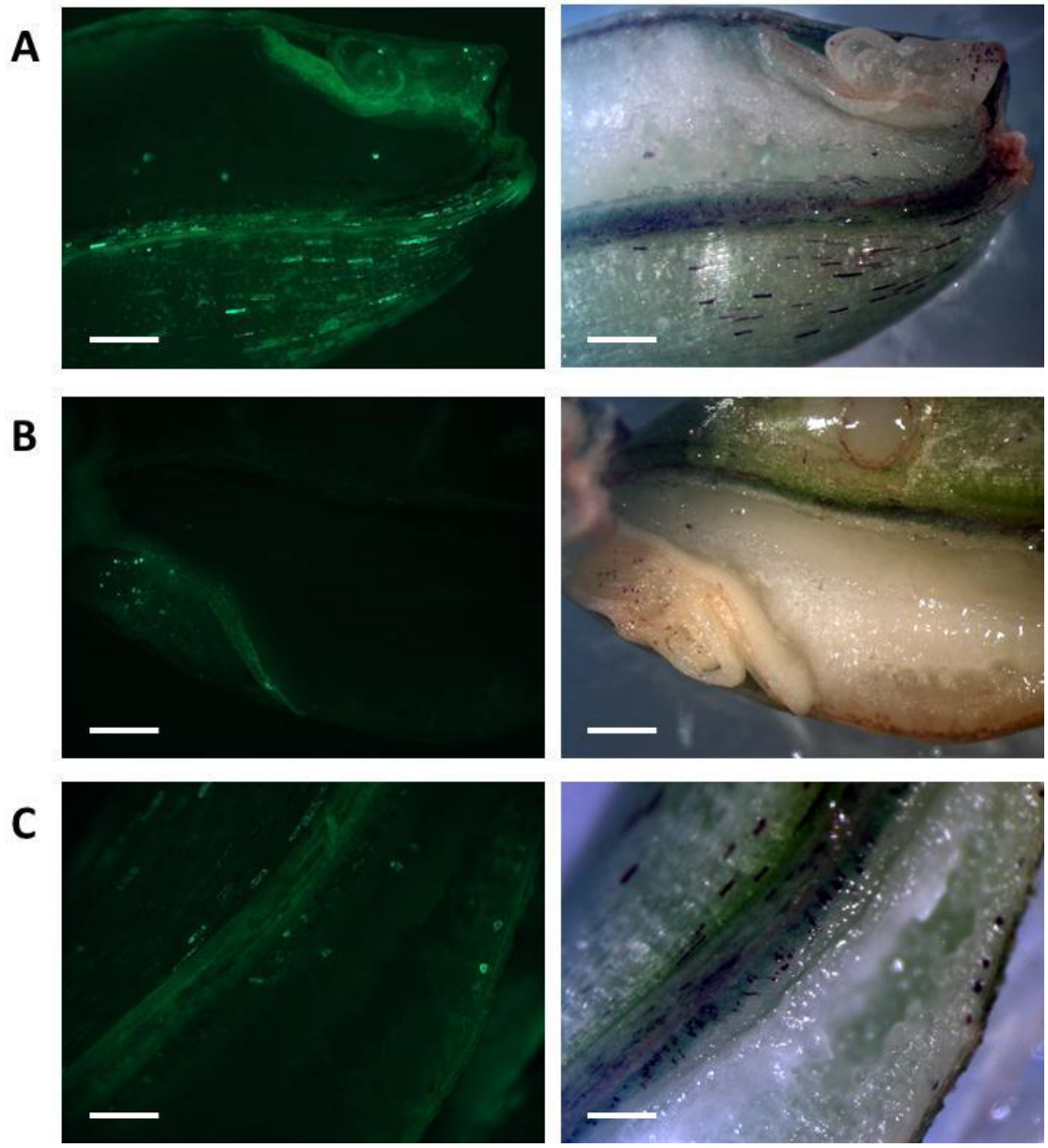


Figure 6.5 Co-bombardment of triticale wheat germ agglutinin (T-WGA) homologue promoter deletion fragment:green fluorescent protein (GFP) expression vectors with a control anthocyanin expression vector. Transient expression of GFP (bright green) in triticale seed tissues directed by T-WGA promoter deletion fragment -272 (A) and complete fragment -2734 (B and C) in expression vector pLRCWHEAT3-pevAgdv was observed under blue light (left). Anthocyanin expression directed by the control vector pBC17 was observed under white light (right A-C) in seed tissues 48 hours after bombardment with the Helios® Gene Gun (BioRad). Seeds were dissected longitudinally along the ventral groove and placed on Murshige and Skoog agar plates and then bombarded. Scale bars (white) = 1mm.

not in any other seed or leaf tissues. This suggests that there is a negative regulator of endosperm and seed coat expression in the promoter region between -598 and -298. Elements in this region include an RY element which is involved in seed-specific regulation, a MYB transcription factor binding site (CCAAT box), a few elements with unknown function (two Unnamed 4 elements and a GCC box), a wounding responsive element (WUN-motif), two light responsive elements (Sp1 and G box), an abscisic acid responsive element (ABRE) and a cell cycle regulation element (MSA-like, Table 6.3). Using the next larger T-WGA promoter fragment (-1,087 to +1) transient GFP expression was observed in all seed tissues suggesting that there is an element between -1087 and -598 that can restore expression to the remaining seed tissues (Table 6.4). There are several putative *cis*-elements in this promoter region including a few which may be of particular interest with respect to the seed-specific expression patterns. These include the OCT element which is related to meristem-specific activation, the Py-rich stretch, an element known to confer high expression levels, an RY element involved in seed-specific regulation and a CAT box, an element related to meristem expression. The next two larger promoter fragments (-1,570 to +1 and -2,214 to +1) were only capable of directing transgene expression in the embryo and endosperm tissues. Expression in seed coat and leaf tissues was not observed using these promoter fragments (Table 6.4). Elements between -1,087 and -2,214 include two Skn-1 motifs which are elements that have been shown to be required for endosperm expression in rice (Washida et al., 1999). Other elements in this region that may be important to the restriction of expression to the endosperm and embryo include the O2 site (maize), CCGTCC box (Arabidopsis) and the AACA motif (rice), elements all related to meristem or seed-specific regulation (Table

6.3). Transient GFP expression was observed in the embryo and not the endosperm, seed coat or leaf tissues using the -2,485 promoter fragment. The largest T-WGA promoter fragment was able to restore expression to the endosperm and seed coat tissues and GFP expression was observed in all triticale seed tissues (Table 6.4). Between -2,214 and -2,485 there are two Skn-1 motifs which may be responsible for the negative regulation in the endosperm tissue. Upstream of -2,485 are two GCN4 motifs (rice), two Skn-1 motifs (rice) and two O2 sites (maize) which may play a role in the restoration of expression in the endosperm and a Box1 and a GATA motif which are both light responsive elements that may have restored expression to the seed coat tissues (Table 6.3). No GFP expression was observed in leaf tissues following bombardment with any of the T-WGA promoter deletion fragment expression vectors (Table 6.4).

6.3.3 Triticale aleurain homologue

6.3.3.1 Triticale aleurain homologue promoter sequence

Genome walking using primers designed from the barley thiol protease aleurain sequence (GenBank accession no. X05167) resulted in 1641 bp of genomic DNA sequence identified upstream of the CyPr294r primer (Figure 6.6). When the sequence was queried using nucleotide BLAST, 460 bp at the 3' end of the sequenced DNA fragment (bp -244 to 117, Figure 6.6) aligned to the barley gene for thiol protease aleurain sequence at its 5'end with 87% identity. However, 211 bp at the 3' end of the sequenced triticale DNA fragment (bp 1 to 211, Figure 6.6) aligned to the wheat mRNA for triticain γ, complete cds (GenBank accession no. AB267409) at its 5' end with 100% identity. The wheat triticain γ differs from the triticain α and β in this region, in

-1348 GTCTCTTCCTGTTTGTTTTCAGCGACTA

-1320 AAAAATCCCTAGTCCTTTCTTAGAGGTTATTAAATGACCATATTGCCCCTAGTATATAGA
Skn-1 motif

-1260 AAAATAACAATCAAACAACACCATGAGGTGGCGGGCTATTGAATGTATGGAGGGGCATTG

-1200 TTGAAAAAGTTCCAAAAAAGACTCTCCTTAAGAGTCTTCTTTATTTAGTCTTAAATGACT
Skn-1 motif

-1140 AGTTTAGTCCCTAAAAAGTCTCTTCCATTTGGTAAAAAAGTCTCTAAAAAGACTTTTTCT

-1080 AGTCCCTACACAAAAAAGTCCCTGAAAACAAACACCCCCTTATATTACTATTACAAAATA

GC motif
-1020 TAATAAAGTACTCATCAGAATTAGATTCCCCCGACAAAACACAAGAATCCATCACAGTGC

-960 ACTCCTCCAAACGAAGCCATTCACTTTGCTGATGTTTCATGCTCTTGCGATTTTCCATGT

-900 TAATTCGGTGCTCAAGGAATGTTATGAAGATTAGGAGTACCTAACATGCACATGTGTTAC

-840 CCACAAGAAAAAATATGCACATGTGTTAGAATCAAGTAATGATCTTTCCTTCCGCCCGCT

-780 GGCATGAGGGGCCGCTAGGGCGAAGAACTCATCCACTTCAGGTTCCGTTGGTGAGCTTGG
CCAAT box

TGACG motif
-720 GATTTGCCTTTCCTTTGTCTGTCGCTCTGACGCTCGCGGGGAGAGAAAATCCGGTGCGTC
CGTCA motif

-660 GGCTCCAGCTAGTAGTTCAGGTTAGAATTTTTGTAGTCTTCGTAGGAGTGACGTTCGGGT

Figure 6.6 Triticale homologue of wheat triticain γ (T-triticain γ) sequence. Sequence of wheat triticain γ (GenBank accession no. AB267409, nucleotides 215 to 1453) with upstream sequence (nucleotides -1348 to -214) identified in triticale (AC Certa) by genome walking is shown with putative features including; transcription start site (nucleotide -60, boxed), translation start codon (nucleotide 1, boxed), TATA box (red nucleotides), CAAT boxes (green nucleotides). Positive numbers on the left refer to the nucleotide position beginning with the start codon. Negative numbers (also on the left) refer to nucleotide position of genomic sequence upstream of the coding sequence beginning with the start codon. Nucleotides highlighted in yellow are genome walking gene-specific primer complementary regions. Some cis-acting elements identified using PlantCARE (Lescot et al., 2002) are named and are shown by ____ above sequence if element is in + strand (shown) and by ____ below sequence if element is in – strand (complementary). Lower case sequence represents the translated regions. The stop codon (nucleotide 1102) is underlined. Violet nucleotides represent an intron identified by genome walking and are not included in sequence numbering. Light blue nucleotides (also including TATA box) are sequence that is encompassed by existing patents in the patent database (nucleotide -92 to 64; US patent # 7132589, Dunn-Coleman et al., 2006 and positions -74 to 1532 excluding the intron; US patent # 7214786, Kovalic et al., 2007).

```
-600 GAATGGCGCCGCTTCATCTTCGAGCTGGTCTTATGGGCTTTGATATTTCTCGACTTTGTC

-540 TGTCGAGACGAAGCCCGGGACGTAGATCCAACAGTTGCGAGATCTTGTGTCAACAGCTTC
                                     MBS

                              TGACG motif
-480 AGATCGATCCAAGAGTTCAACGATGACGGATGCAGCTCCAGCTCCAGGCTGCCGGTTCCT
                        Skn-1 motif
                           CGTCA motif

          ABRE
-420 AGAGACACGTGAAGAATAGTTCTTGGCTACCATCGACAAAGTTAGACCGGCTCCAATAGA

-360 AGAGCAGTGACAATGATCATTCAGTGGTGCTGGCCTCGATGTAATTTTTATTACCAGTGA
           Box III

-300 ACTTTTGTAGGAGTAATATATTAGTCGTATTGAAATAAACATGTGCACGATGTTATCGCG

-240 CTGGATTTTTGGCCTTGAAGATTGAGACAAAGAATAATCCTCACTAGAGCAGTAGGAGCA

-180 GTAGATTATACTTACTGCAGATCACAAATATCTACAGCATTTGGAGAGTCTCAAACAAGT

                                                 GC motif
-120 GGTCAAGCGTGCCGCGGCGCGCTCCCCGATATAAAACCCCCGGCACCGGGACGCGAGCTC
     Box W1         GC motif

 -60 CCTACTGGCCTGCCCCGCCTGCCCAACCCAGTAAACTACTGCCACCCATCCGCCGGCGAA

   1 atggcccacgcccgcatcctcctcctggcgctcgccgccctggccacggccaccgtcgcc
  61 gtcgcctcctcctcctcctcctcctccttcgccgtctccaacccgatccggcccgtcacc
 121 gagcgcgccgcttcgacgctcgagtccaccgtcctcgccgcgctcggccgcacccgccac
 181 gccctccgcttcgcgcgcttcgccgtcaggtCCGTGCCCCAACTAACCCCCCATCGGTCG
 212 ACTCCCCGTCTGCTCCGGCACGCCGCCGTCCCCGTTGACACCCGCTTTCGcaggtacggc
 222 aagagctacgagagacgcggcggaggtgcggaggcggttccggatcttctccgagagcct
 282 cgaggaggtgcgctccaccaaccggaagggcctatcctaccgcctcggcatcaaccgctt
 342 ctcggacatgagctgggaggagttccaggcgacccggctcggcgcggcgcagacctgctc
 402 ggccacgctcgccggcaaccacctgatgcgcgacgccgccgcgctcccggagaccaaaga
 462 ctggagggaggatgggatcgttagccccgtaaaagaccagtcccattgcggctcctgctg
 522 gacgttcagcactactggtgcacttgaggcagcatatactcaggccactggaaagaacat
 582 ctctctttctgagcaacagctggttgactgtgccggcggattcaataatttcggatgcag
 642 cggaggccttccatcccaggcatttgagtacatcaaatacaatggagggatcgataccga
 702 ggagtcctacccttacaagggtgtcaatggcgtctgccattacaaggctgaaaatgccgt
 762 agttcaggttttggactctgttaacatcacactgaatgctgaggacgaactgaaaaacgc
 822 cgtcgggttggttcgcccagtgagtgttgcctttgaggtgatcaacggtttcaggcagta
 882 caagagcggagtttacagtagcgaccattgcggtactacacctgatgacgtgaaccacgc
 942 cgtcttggcggtcgggtacggcgtggagaacggcgtcccgtattggctcatcaagaactc
1002 gtggggcgccgactggggtgacaatggttacttcaagatggaaatgggcaagaacatgtg
1062 tgctgttgctacctgtgcgtcctaccccatcgtagcggcatgaAGCGCCAAATCGACGAG
1122 TTCAGTCTCGCTTCGCTAGACATGAATAATGTGTTTTCAGCCCACAAGACATGAAGATTG
1182 GGGGATACAGAAACTGAAACTCTGCTGTGAATATAAAACCACAAGGACTACAGTTGCAGT
1242 GAACCACGTGCGTGTACTCTTCTTCTGTATGGTTGTATGGCTATATATGTAGGTGATACA
1302 GCAATTTATTGGGGTTTCGCATGTATGTCTACGACATTTTATTTCCGGGACAAATAAAAT
1362 GGCCCATTCATAGTTCACCGAGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
1422 AAAAAAAAAAAAAAAAAAGCGGCCGCGAATTC
```

particular there is a stretch of seven serine residues (codon TCC; nucleotides 67 to 87) that is unique to triticain γ (Kiyosaki et al., 2009). This region, including the stretch of seven serine residues, is identical to the identified triticale sequence. It is therefore concluded that the sequence identified is the triticale homologue to the wheat triticain γ (T-triticain-γ). The wheat mRNA sequence for triticain γ, complete cds was submitted to the database in July 2006 after the barley aleurain was selected for investigation by genome walking. Alignment of the triticale genomic DNA sequence identified by genome walking with the wheat mRNA for triticain γ using the align two sequences function in BLAST, revealed a 79 bp section of the identified sequence just upstream of the secondary PCR primer, CyPr294r, that does not align with the triticain γ mRNA sequence and appears to be an intron (violet colored upper case nucleotides between bp 211 and 212, Figure 6.6). Both gene-specific primers and sequence between the primers and the intron also align with the triticain γ mRNA sequence with 100% identity.

The position of the ATG start codon from the triticain γ sequence was used to determine positions of upstream *cis*-regulatory elements (Figure 6.6, Table 6.5). Several TATA boxes were identified throughout the upstream sequence, with the three closest to the start codon at -89, -91 and -92 (Table 6.5, red nucleotides, Figure 6.6), a reasonable distance for an active TATA box which also match the TATA signal in sequence and position for barley aleurain (Whittier et al., 1987). There is no promoter sequence data available for wheat triticain γ, only mRNA sequence so the TSS for the barley aleurain sequence which corresponds with position -60 bp in the identified triticale sequence was determined to be the putative TSS. The sequence between the putative TATA box and TSS (-92 and -60, Figure 6.6) aligns with the barley aleurain sequence with 100%

Table 6.5 Triticale homologue of wheat triticain γ (T- triticain γ) upstream sequence elements. *Cis*-elements and motifs identified using PlantCARE (Lescot et al., 2002) in the promoter sequence upstream of wheat triticain γ (GenBank accession no. AB267409) identified in triticale (AC Certa) through genome walking.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
ABRE*	*Arabidopsis thaliana*	-415	+	6	CACGTG	cis-acting element involved in the abscisic acid responsiveness
ATCT motif	*Pisum sativum*	-1003	-	9	AATCTAATCC	part of a conserved DNA module involved in light responsiveness
Box III *	*Pisum sativum*	-355	-	11	atCATTTTCACt	protein binding site
Box W1 *	*Petroselinum crispum*	-120	-	6	TTGACC	fungal elicitor responsive element
CAAT box*	*Brassica rapa*	-142	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box*	*Brassica rapa*	-156	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-220	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-272	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-350	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-367	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-368	+	5	CCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-719	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1114	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1204	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1223	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1253	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-1279	-	6	gGCAAT	common cis-acting element in promoter and enhancer regions
CCAAT box*	*Hordeum vulgare*	-736	-	6	CAACGG	MYBHv1 binding site
CGTCA motif*	*Hordeum vulgare*	-457	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
CGTCA motif*	*Hordeum vulgare*	-693	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
G Box	*Pisum sativum*	-415	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
G box	*Arabidopsis thaliana*	-415	+	6	CACGTG	cis-acting regulatory element involved in light responsiveness
GC motif *	*Zea mays*	-84	+	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GC motif *	*Zea mays*	-993	+	6	CCCCCG	enhancer-like element involved in anoxic specific inducibility
GC motif *	*Oryza sativa*	-106	-	9	AGCGCGCCG	
MBS *	*Arabidopsis thaliana*	-509	-	6	CAACTG	MYB binding site involved in drought-inducibility

* element/motif shown in Figure 6.6

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.5 Continued.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
Skn-1 motif *	*Oryza sativa*	-458	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif *	*Oryza sativa*	-1146	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif *	*Oryza sativa*	-1287	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Sp1	*Zea mays*	-19	+	5.5	CC(G/A)CCC	light responsive element
Sp1	*Oryza sativa*	-789	-	6	GGGCGG	light responsive element
TATA box*	*Lycopersicon esculentum*	-89	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box*	*Arabidopsis thaliana*	-91	+	4	TATA	core promoter element around -30 of transcription start
TATA box*	*Brassica oleracea*	-92	+	6	ATATAA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-174	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-175	-	5	TATAA	core promoter element around -30 of transcription start
TATA box	*Brassica napus*	-176	+	6	ATTATA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-282	-	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-284	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Brassica napus*	-285	+	6	ATATAT	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-287	+	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Oryza sativa*	-298	-	7	TACAAAA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-315	+	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Oryza sativa*	-632	-	7	TACAAAA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-1020	+	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1022	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Brassica oleracea*	-1023	+	7	ATATAAT	core promoter element around -30 of transcription start
TATA box	*Oryza sativa*	-1029	+	7	TACAAAA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-1038	-	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1040	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1041	-	5	TATAA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1096	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1108	-	5	TTTTA	core promoter element around -30 of transcription start

* element/motif shown in Figure 6.6

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.5 Continued.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
TATA box	*Lycopersicon esculentum*	-1129	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1266	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1268	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-1293	-	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1322	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1354	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1387	+	4	TATA	core promoter element around -30 of transcription start
TCA element	*Brassica oleracea*	-412	+	9	GAGAAGAATA	cis-acting element involved in salicylic acid responsiveness
TGACG motif*	*Hordeum vulgare*	-457	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG motif*	*Hordeum vulgare*	-693	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
Unnamed 1	*Zea mays*	-376	+	5	CGTGG	
Unnamed 2	*Zea mays*	-83	+	6	CCCCGG	
Unnamed 3	*Zea mays*	-1376	+	5	CGTGG	
Unnamed 4	*Petroselinum hortense*	-63	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-99	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-138	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-186	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-291	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-439	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-445	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-616	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-658	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-681	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-867	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-956	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-959	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1177	+	4	CTCC	
W box	*Arabidopsis thaliana*	-120	-	6	TTGACC	wounding and pathogen response

* element/motif shown in Figure 6.6

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

identity. However, downstream of the barley TSS to the ATG start codon (bp -60 to 1) the identity is reduced to 75%. Several CAAT boxes were also identified, with the nearest to the start codon at -142 and -156 and with reasonable proximity to the TATA region for an active promoter region (green nucleotides, Figure 6.6).

Several other *cis*-acting regulatory elements were identified using PlantCARE (Table 6.5). These elements include three Skn-1 motifs (rice), elements known to be involved in seed-specific regulation (Takaiwa et al., 1991), several elements involved in light response, stress and hormone response, and MYB transcription factor binding sites. GC motifs involved in anoxic-specific response were also identified (Table 6.5, Figure 6.6). The identified elements were dispersed throughout promoter sequence.

The majority of the sequence identified upstream of the ATG start codon had no similarity to other sequences in the database including patents. Only the first 91 nucleotides upstream of the start codon and the downstream wheat triticain γ sequence (not including the proposed intron) aligned to sequences in patent databases (light blue nucleotides, Figure 6.6). The region from the TATA box (-92, Figure 6.6) to nucleotide 64 bp is covered by US patent number 7132589 (Dunn-Coleman et al., 2006). This patent covers the presumed vacuolar targeting signal sequence of wheat triticain γ. The sequence from -74 to 1532 (excluding the intron, Figure 6.6) is covered by US patent number 7214786 (Kovalic et al., 2007). The remaining sequence upstream of the TATA box had no similarity to any sequence in the patent database.

6.3.3.2 Transient expression using triticale triticain γ homologue promoter deletion fragments

Five T-triticain-γ promoter deletion fragments of varying lengths were cloned into the GFP expression vector pLRCWHEAT3-pevAgdv and co-bombarded with anthocyanin protein expression vector pBC17 into triticale endosperm halves and leaf tissues attached to the plant. Endosperm halves were chosen for preliminary transient expression studies since it was expected that the T-triticain-γ homologue promoter would be capable of directing expression in the outermost layer of the endosperm, the aleurone layer, similar to its homologue, barley aleurain (Holwerda and Rogers, 1992). Plant tissues were observed under blue light 48 hours following bombardments and rated for presence or absence of GFP expression (Table 6.6, Figures 6.7, 6.8). The endosperm halves did not contain embryo or seed coat tissues and therefore expression in these tissues was not evaluated. Anthocyanin expression was clearly visible in the endosperm tissue indicating successful incorporation of the expression vectors into the endosperm cells using the particle bombardment approach (Figure 6.7). In all cases where GFP and/or anthocyanin were observed in the endosperm, the expression appeared to be in the periphery of the endosperm, either in or near the aleurone cell layer (Figure 6.7). While this may be an indication of expression specificity to the aleurone cell layer, GFP and anthocyanin expression also showed the same pattern of expression in the aleurone cells when using control vectors (Figures 6.2 and 6.3), suggesting that at the developmental stage used for transient expression studies, transcription is likely reduced in the starchy endosperm and still highly active in the aleurone cells. Triticale endosperm cells have

Table 6.6 Triticale triticain γ homologue (T-triticain-γ) promoter deletion fragment transient expression results. Expression vectors containing T-triticain-γ promoter deletion fragments and the green fluorescent protein (GFP) gene were bombarded into plant tissues using the Helios® Gene Gun (Bio-Rad Laboratories). Tissues were then screened for GFP expression under blue light using a stereo microscope. (+) presence of GFP expression, (-) absence of GFP expression. N/A indicates data is not available as tissues were not bombarded.

Triticain Promoter Fragment	Embryo	Endosperm	Seed Coat	Leaf
-218 to +1	N/A	+	N/A	+
-576 to +1	N/A	+	N/A	+
-1059 to +1	N/A	+	N/A	+
-1228 to +1	N/A	+	N/A	+
-1348 to +1	N/A	+	N/A	+

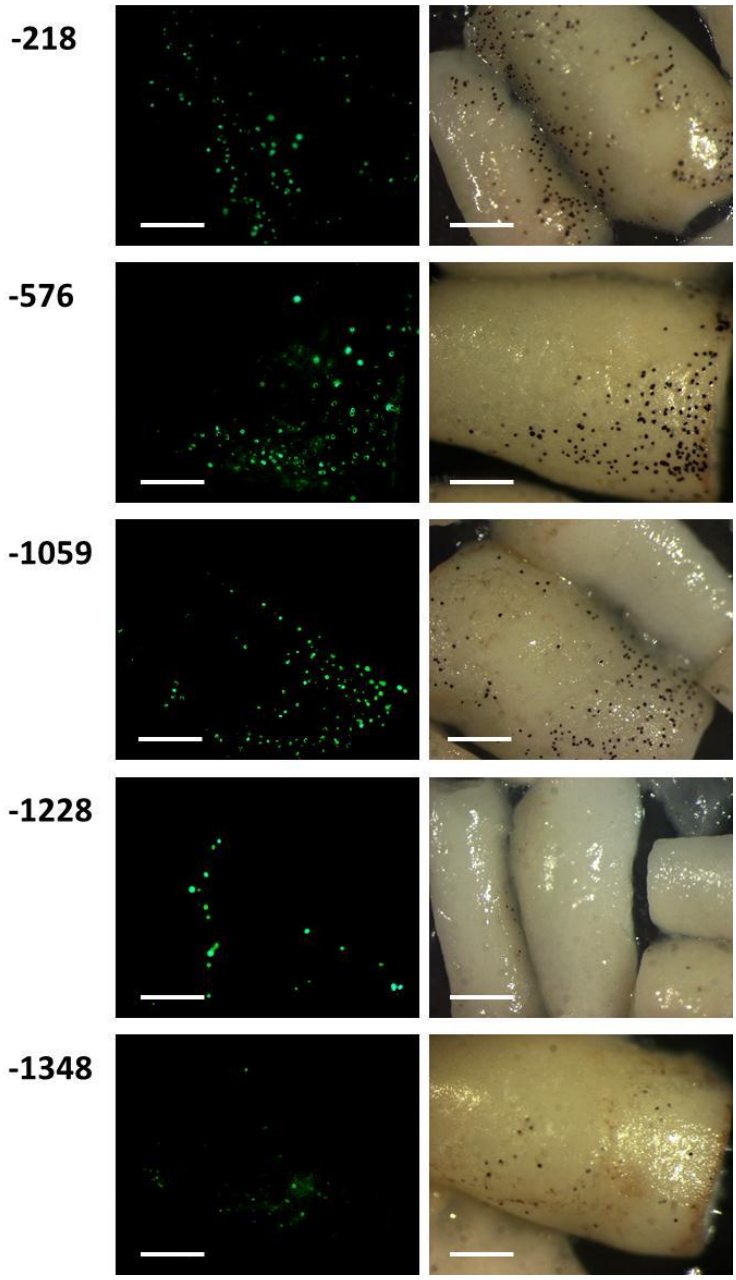


Figure 6.7 Transient expression using triticale triticain γ homologue (T-triticain-γ) promoter deletion fragments in triticale seed tissues. Transient expression of green fluorescent protein (GFP, bright green) directed by T-triticain-γ promoter deletion fragments in the GFP expression vector pLRCWHEAT3-pevAgdv observed under blue light (left) and anthocyanin expression (purple) directed by the control vector pBC17 in the same tissue section observed under white light conditions (right) in triticale endosperm halves. Promoter fragments used are indicated by the length of promoter upstream of the T-triticain-γ start codon (numbers to the left). Endosperms were dissected from triticale seeds 12 days post anthesis and placed on Murshige and Skoog media then bombarded using the Helios® Gene Gun (BioRad). Observations were made 48 hours later. Scale bars (white) = 1 mm.

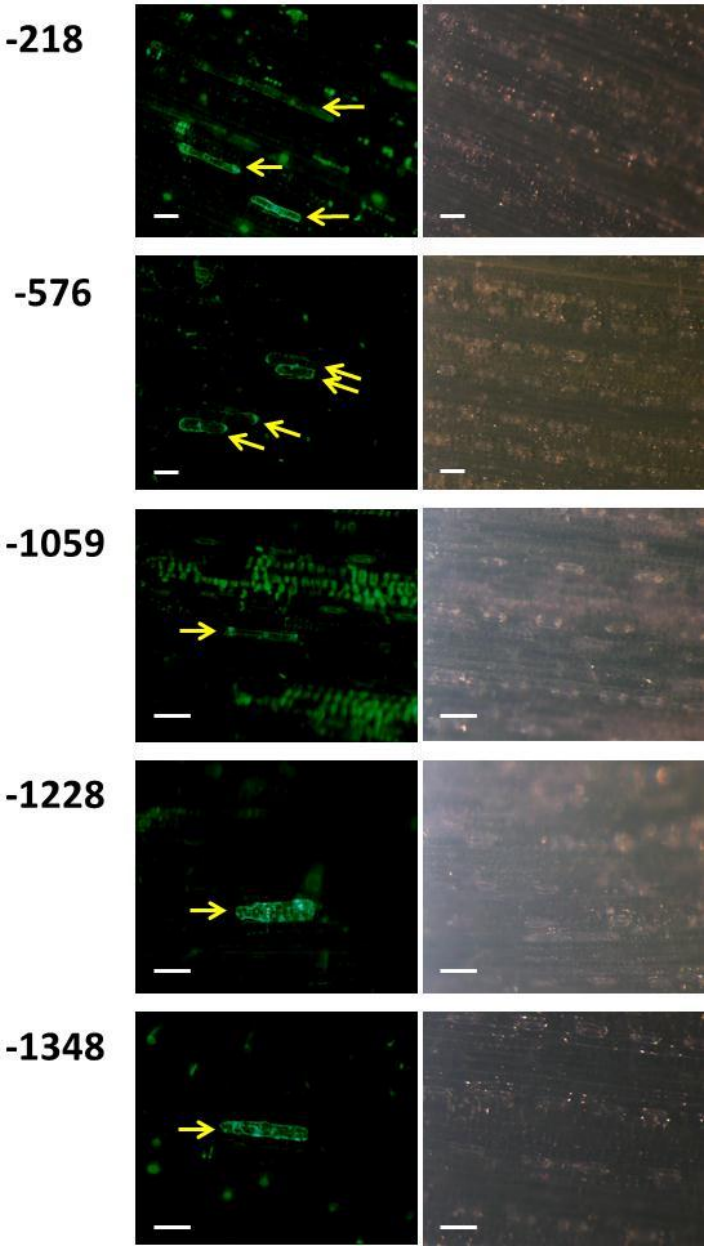


Figure 6.8 Transient expression using triticale triticain γ homologue (T-triticain-γ) promoter deletion fragments in triticale leaf tissues. Transient expression of green fluorescent protein (GFP, bright green) directed by T-triticain-γ promoter deletion fragments in the GFP expression vector pLRCWHEAT3-pevAgdv observed under blue light (left) and the same tissue section observed under white light conditions (right) in leaf tissues. Promoter fragments used are indicated by the length of promoter upstream of the T-triticain-γ start codon (numbers to the left). Leafs were bombarded on live plants using the Helios® Gene Gun (BioRad) then observed 48 hours later for expression. Yellow arrows indicate cells expressing GFP, remaining green in left images is background autofluorescence. Scale bars (white) = 0.1 mm. White light images are variable in clarity and color since tissues imaged were focused on the GFP expressing cells under blue light conditions.

been shown to undergo programmed cell death as early as six days post anthesis (Li et al., 2010), which would indicate our findings are a result of loss of expression capability within the starchy endosperm cells rather than expression specificity. Transient expression of GFP was observed under the regulation of all T-triticain-γ promoter deletion fragments in the endosperm tissue; however, all promoter deletion fragments tested were also capable of directing expression in the leaf tissue (Table 6.6, Figure 6.8). This suggests that the factors controlling the tissue-specificity of triticain expression may be further upstream than the 1348 nucleotides of T-triticain-γ promoter sequence isolated in this study. While an increased number of tissues, including seed coat and embryo, should be tested for transient expression using these T-triticain-γ promoter deletion fragments, these preliminary results suggest that the T-triticain-γ promoter identified in this study may be capable of constitutive transgene expression.

Skn-1 motifs, elements that have been shown to be required for endosperm expression in rice (Washida et al., 1999), were identified in the T-triticain-γ promoter at positions -458, -1,146 and -1,287 bp upstream of the translation start site and were expected to play a role in endosperm specificity of transient expression using T-triticain-γ promoters. While promoter deletion fragments containing these elements did result in transient transgene expression in the endosperm, the smallest T-triticain-γ promoter fragment which does not have a Skn-1 motif was also capable of transgene expression in the endosperm, indicating that this element is not required in the triticain promoter for endosperm expression in triticale.

6.3.4 Triticale ω-secalin homologue

6.3.4.1 Triticale ω-secalin homologue promoter sequence

Genome walking identified 1,602 bp of genomic DNA sequence upstream of the OmSec67r primer (Figure 6.9). This sequence overlapped with the 5' end of the rye ω-secalin mRNA sequence. The 1,602 nucleotides identified by genome walking aligned with the rye ω-secalin gene, genomic sequence (GenBank accession no. AF000227), with 97% identity. The triticale ω-secalin homologue (T-ω-SEC) promoter sequence is shown in Figure 6.9 with the differences with the rye ω-secalin highlighted in grey. There were 45 bp differences (2.9%) in the upstream sequence, the majority of which are single base pair substitutions as well as three single bp insertions, one double bp insertion and two single bp deletions.

The ATG start codon was identified from the ω-secalin gene sequence in the database (GenBank accession no. AF000227) and is shown at nucleotide 1 in Figure 6.9 (boxed). According to this record the mRNA also begins at bp 2751 which would correspond to a transcription start site at nucleotide -1 in the triticale sequence (Figure 6.9). Several TATA boxes and CAAT boxes were identified in the sequence upstream of the start codon using PlantCARE (Table 6.7). Two TATA box rich regions were identified with close proximity to the ATG start codon, one at -87 and another at -107 (red nucleotides, Figure 6.9) which is a conserved TATA region between the published ω-secalin gene and other related storage protein genes (Clarke and Appels, 1999). Only one CAAT box was identified within 100 bp upstream of these TATA box rich regions (green nucleotides, Figure 6.9).

```
                                      A
-1536                          GGTTTGTACCA:CTTTCAATTGCTAATCAGTGATAA

                                     A   T           A
-1500 AACACTATCAGCGCGATGTTTTTCCCTCTGTGGCAAACTCGCCCTAACCCAACGTCTCCT

          C A      A Skn-1 motif                              T
-1440 TGTATCCATGTCCAGTCATCCCTCTAAAATAGATAGTCACAAAGATTAAAACAAATAAAC

                               A                     A
-1380 ACGCACGTCCTGTGTGGCAGCACGGAAAAACGTGATGCCACACCAGTCACGTTCGGCCTG

                         CG
-1320 CTTTGAAGCACGAATCCTGCTTCGCCCAGGCTTCATACGTGTAACCTCCTCACATGGGTC

              G  Skn-1 motif T             G    A
-1260 TTCCTCTAAATACGGTCATATACGGCCTGTCAGGCAGGACGACAAATGGCAATACATCTC

-1200 ATGCACATATCTCCATGCTAGTTTGATGATTCATAGCTACATGCAAACCGGCCCCTACAA

                                                          T
-1140 CATGTTTTCGAGACACACACACCCTTCACTTTTGGACAAATAACACTTCGCAGTTTTCCC

                                     A
-1080 TCTTCTTGCCGGTTGAATGTTCTTAGACCATATGAAAATGGTGAAAATGAACAACATACC

-1020 CAGTTGTTTCTAGATTCTGCTCGCAAAAAAGAATAGTTAAGTTGCAAGCATGTGGGAGAC

       TT                                              C
 -960 AACGTGATTTAGGTTAACTAAAGAGGGTCGGAGGGATCCCTCAGGGATCCAAATTTGCAT

       T                                               C
 -900 GCGTATGGATACTGAGAGGCTTGGGGGGATTCTTGAATTTTTCTTATGAGCCGTGAGAGA
```

Figure 6.9 Triticale ω-secalin homologue (T-ω-SEC) gene sequence. Sequence of rye ω-secalin gene homologous sequence identified in triticale (AC Certa) by genome walking is shown with features including; translation start codon (boxed), putative TATA boxes (red nucleotides), putative CAAT box (green nucleotides) and genome walking gene-specific primer complementary regions (nucleotides highlighted in yellow). Some *cis*-acting elements identified using PlantCARE (Lescot et al., 2002) are named and are shown by ── above sequence if element is in + strand (shown) and by ── below sequence if element is in – strand (complementary). Positive numbers on the left refer to the nucleotide position beginning with the start codon. Negative numbers (also on the left) refer to nucleotide position of genomic sequence upstream of the coding sequence beginning with the start codon. Sequence downstream of the genome walking primers is from rye ω-secalin gene (GenBank accession no. AF000227). Grey highlighted nucleotides denote differences between T-ω-SEC sequence and the rye ω-secalin gene sequence (GenBank accession no. AF000227) with the nucleotides from rye ω-secalin above the T-ω-SEC sequence. Lower case sequence is translated regions. Translation stop codon is underlined. Light blue nucleotides from -54 to 1255 and from 4374 to 4681 are encompassed by US patent # 7214786 (Kovalic et al., 2007).

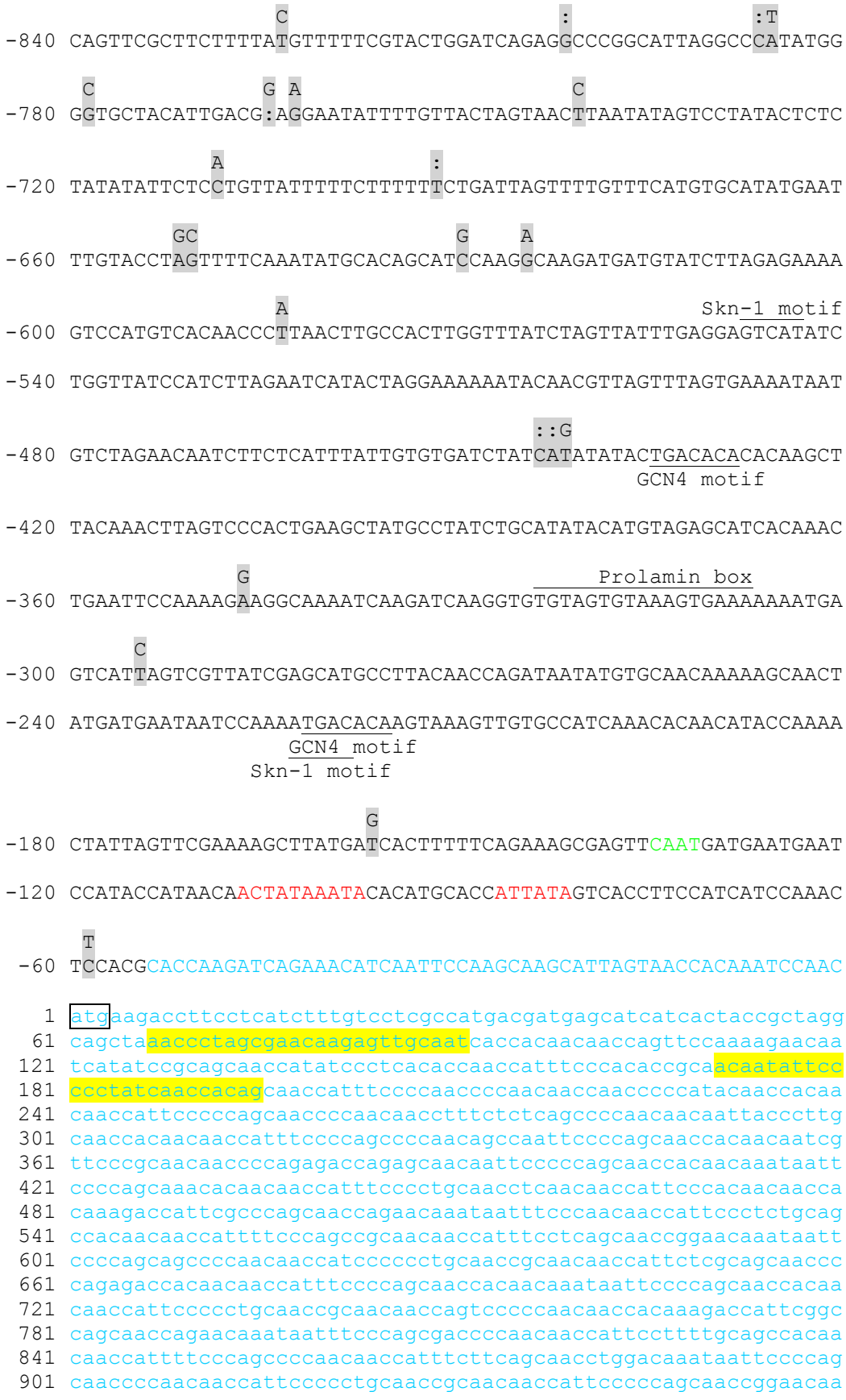

-840 CAGTTCGCTTCTTTTATGTTTTTCGTACTGGATCAGAGGCCCGGCATTAGGCCCATATGG
-780 GGTGCTACATTGACG:AGGAATATTTTGTTACTAGTAACTTAATATAGTCCTATACTCTC
-720 TATATATTCTCCTGTTATTTTTCTTTTTTCTGATTAGTTTTGTTTCATGTGCATATGAAT
-660 TTGTACCTAGTTTTCAAATATGCACAGCATCCAAGGCAAGATGATGTATCTTAGAGAAAA
-600 GTCCATGTCACAACCCTTAACTTGCCACTTGGTTTATCTAGTTATTTGAGGAGTCATATC
Skn-1 motif
-540 TGGTTATCCATCTTAGAATCATACTAGGAAAAAATACAACGTTAGTTTAGTGAAAATAAT
-480 GTCTAGAACAATCTTCTCATTTATTGTGTGATCTATCATATATACTGACACACACAAGCT
GCN4 motif
-420 TACAAACTTAGTCCCACTGAAGCTATGCCTATCTGCATATACATGTAGAGCATCACAAAC
-360 TGAATTCCAAAAGAAGGCAAAATCAAGATCAAGGTGTGTAGTGTAAAGTGAAAAAAATGA
Prolamin box
-300 GTCATTAGTCGTTATCGAGCATGCCTTACAACCAGATAATATGTGCAACAAAAAGCAACT
-240 ATGATGAATAATCCAAAATGACACAAGTAAAGTTGTGCCATCAAACACAACATACCAAAA
GCN4 motif
Skn-1 motif
-180 CTATTAGTTCGAAAAGCTTATGATCACTTTTTCAGAAAGCGAGTTCAATGATGAATGAAT
-120 CCATACCATAACAACTATAAATACACATGCACCATTATAGTCACCTTCCATCATCCAAAC
-60 TCCACGCACCAAGATCAGAAACATCAATTCCAAGCAAGCATTAGTAACCACAAATCCAAC
1 atgaagaccttcctcatctttgtcctcgccatgacgatgagcatcatcactaccgctagg
61 cagctaaaccctagcgaacaagagttgcaatcaccacaacaaccagttccaaaagaacaa
121 tcatatccgcagcaaccatatccctcacaccaaccatttcccacaccgcaacaatattcc
181 ccctatcaaccacagcaaccatttccccaaccccaacaaccaaccccccatacaaccacaa
241 caaccattcccccagcaaccccaacaaccttttctctcagccccaacaacaattacccttg
301 caaccacaacaaccatttccccagccccaacagccaattccccagcaaccacaacaatcg
361 ttcccgcaacaaccccagagaccagagcaacaattcccccagcaaccacaacaaataatt
421 ccccagcaaacacaacaaccatttcccctgcaacctcaacaaccattcccacaacaacca
481 caaagaccattcgcccagcaaccagaacaaataatttcccaacaaccattccctctgcag
541 ccacaacaaccatttcccagccgcaacaaccatttcctcagcaaccggaacaaataatt
601 ccccagcagccccaacaaccatccccccctgcaaccgcaacaaccattctcgcagcaaccc
661 cagagaccacaacaaccatttccccagcaaccacaacaaataattccccagcaaccacaa
721 caaccattcccccctgcaaccgcaacaaccagtcccccaacaaccacaaagaccattcggc
781 cagcaaccagaacaaataatttcccagcgaccccaacaaccattccttttgcagccacaa
841 caaccattttcccagccccaacaaccatttcttcagcaacctggacaaataattccccag
901 caaccccaacaaccattcccccctgcaaccgcaacaaccattcccccagcaaccggaacaa

Figure 6.9 Continued

```
 961 ataatttccgagcaaccccaacaaccattccttctgcaaccacaacaaccgtccccccaa
1021 caaccacaactaccatttccccagccccagcaaccatttgtagtagtggtatagGCATCG
1081 GGGGCCAATCAAACAAAGAGATATAGTACTAGACCGGTGGATCATCGTTGTTTAGTCAAT
1141 GGAGTGTTCAATGTAACGATGATAAATAAAGTGATGTGTACCATCATGTGTAACCCCGAG
1201 CTATACTAGTTCAAACATGAGAATAAAAGAAAGAAAGTTCTTGTCACAAGGACATTGCTG
1261 GTAATTATTAAATTCATGCCATATTCGCATTTTCATCCCAAAAAATAATTTGAGTCATAT
1321 GCATTATCTACCTATTTATGAAGTGAAGTCATATGAGCCTGGCATAATTAATGGTTGTGA
1381 AACTTGGTTTAATTGAAGTACTATCAATTAAGGAGTAATGAGTAATTACAAATTATCTCT
1441 TCGGTGGCTAGCCGTGTGAATAGTGCGACCTTATTATCTGAAACGGTTCCAGATTGTACA
1501 TAGACACATTTTTTACAATTGTTGAGGTGCATATGAAGCAACCTATACTTAAAGATATGG
1561 TGACATATCACTGGATTCATTGAAAATCTCAGTACTAACTTTATATGCAAGGGGGAAGAT
1621 CATGATGACTTTGTGATGCATTTTATCATTATTCAAATGGTCTCTTTAATGGACCGGTTT
1681 CTCAGATTAGAAAATATAGACTTGGGCCATACGCCATACTGAATTCTTGCAAATGGACAA
1741 GATGAAGGCATGATTGAATTTATTCCTTCTATTTCTCCTGCATGTGTAACTTGATATTGC
1801 GCCTGGTACAGTTGCCCAATGGTCAGAAACTTTTATTAAAAGTTTTGTTGGTTACACTAA
1861 GATTACATATATTAAGGGTGTCGGAGAAGAGGAAGAATAAGGTGGCGGAAAGGAAAGGGA
1921 AATGGCAAGCAATGGAGGGTGGGCTGGCTCAGGCTCGTGGCTGGCTCGGGTCGGGGCACA
1981 CTAATGAGCACGGGCATCCTTATGTGGATAAGTTGTTGGTCCTGCACGCATGTGAGTAAG
2041 TGGTGGGTCCCACGTTGATGTGCATAACTTGTGGGTCCCACGTGTCATAACTCAAATCGG
2101 TTACCCCTTGTGTTTTGCATTTCATGGTTAAACAGGGCCATCTCGCATTGGAGCTATAGG
2161 TGGCTTCAAAGATGTCGCACGCGAGCTATTTTCGTCAACTACGTCGCACTTGATCCAATT
2221 CACCTCAAAGGTGTCGCACAGGAGCTATTGACCCTGCCAATGGCCAGTTGGCAAGTTCTC
2281 CAATATTGGAGGAACTTATCATTTTGCACTGCTCGGACCTCAAGTCCATTGGTGGTGCAT
2341 CAGCTGTTGATAAAATGAAGAAAGTGATTATAAACAAGTGCCACGAGCTAAGGAAATACA
2401 TCAGCCACACCATCGTTGGTAGCGTAGCTAGGATTGGCCGACGCCCCGGGCCAAAATTTA
2461 AGGGCCAAACAAATTCTCCCAAGTTCACCCCGGTTTTGAGGCATACCAAAGTCCAGCTAA
2521 TTTGCAACTGATAGAATCTGAACATAATATTTAATATTGCCTCGTCTGATGGATGTAGCG
2581 AATCCCCAAAGAGCATAAAGAGGATAAATAATACAAAAGAGGGTACCAAATCCCAGGGTT
2641 GCTTGATTGAAGCTTCCCACGATTACATCTTCAATAGTGACAGATGATATACCTTCAGGT
2701 ATAAAATTTGATATGCAAGTCAGTGCATAGAGATTTCCTCCAAGATCAATCCTTGCAACC
2761 AATGGAGCTCCCCTTTCATATTGGAAGCCACATCATAACCTTTCCCTCTCAAATTAGCCA
2821 TAGACAAGCCCTGAGTAGCAAACATACCGTCCAAAGCTACCTTGAGTGAAACTGCTCGCG
2881 TGTCAGGAACATACTTGATTGCAATAAATATCTCAATCACCTCTCCTTGTTTGTTCAGAT
2941 GCCTACATAAGACGCAACTGCATGATTACTATAAAAACGGTGTTAGTACCCTGGTTCTCT
3001 TGTTCATATTGTCTTCACGTAAATTGGATTTTTTTAGGAAATTAGTTTCTAATGCAACAA
3061 CTCAACAACTACTCAAAGAAGTATTTTAAAGCATCATTTTGAATGTAACTAATATATTGA
3121 AAGGTAATTTGAAGCATTAATCCAGAGAGAAAAAACATCTAACCGTTGCTACTTGCTACT
3181 GCCTGCTTTCTCAAAGAAAACTAGAATTTAGTAGTGGTCACTGGGCAGAGGCCAATACCT
3241 GCCGGCTGGCTGTACGGTGCGGAGCCGGAGGCGTGGCGAGGTCGCCAGATGCCCGGACGA
3301 CCGATTGCGTGCGGTGATGATGAACCAGAGTCCGGGCAAGGTCGTGGGGCGAGCGGTGTC
3361 ACTGAACGAGAGTTCCGGCGAGGTCGCCGGACGATCAGATCAGATGGACGGGGACGGGGT
3421 CGCCGGATTTGACGAGAGCGGTGGCTGACGGGTGGCCTTTTCTCTGCTTCTGCTCTCCTG
3481 CTTGAACGACAGCTGCTGCGCTCGATGCGTAGGGTTCGATCGGGCCGAGAGGGGCAGGTG
3541 CTGGACTGGGTTGTGCCTACGCCCCGGGCCACTTCTGAAATTTTCTAAAATATAGGCTWA
3601 AAAAATTTGGCAACGCCCGCCATGGCCCTGGTAGCCCGGGGCCTACTACGCCCGGTATAT
3661 GCAATGACGGGCAATTTAACTACTGTATGCACATCATTTGTTCTTCCGAACTATCCTAGA
3721 AAGCTGCATGCATGTAATCTCCCTATGCCGCTCCTGAAAAATAAAATAGATTTAGAACTT
3781 TCAGCTTATTAACCTTTTTCATTTGTCTACCCTGGTTGGTACTTCTGATGATGAAATCTA
3841 AATTTGAAACTTGGAGAAGGGCGTCTCCATTTAGAACACACAATAGGGGAGCCCCTAATC
3901 TGGTTGTCTATAGGAGATGGAATCAGATCAGATGTATGCCGTGCAGGGCTGTTGGAGTAG
3961 TGCAAGGAAAAAAATAAGGGCACTTTGATGCTCCCAGGAGCTTGCTAGTATCCTCTATCA
4021 ACTATAGCAGCAGTGCAGTTGCTTGCTATGCATCAAGGCAAAGTAGAATATGAGCCTTGA
4081 GTAATGTTTGTTTGTGTATCCTGCATTTCTATTGCCCCACTGGAGCATGTTCTGGTGGAG
4141 AAATTATATCTCCTGTTCCTATCCAACTTATTTACCTTATCTTCATGAACCTTACCAATT
```

Figure 6.9 Continued

```
4201 CGTTATCGTAAAATAATGTGCGACTCAGAACATCACAATTCCTGCTTGTTATTTACTTGT
4261 AGCAATATGTATCGAAGTATATGATTAAGACACAACATTTGTTTAGAATAAATCCGAGGC
4321 ATACCGTCGATCATCCGAGGACCAAGCAATCACACGAGGCACACTACACCGAGATTTGTT
4381 AACGAGGTTCACAGATATGGCTACATCCCCGAGGCCTGACTATGGGCGCTCCTTCCCATG
4441 ACACCGCTACAATACCGCACCCTGGCCGCCGTGTGCCAGSACACGCCGCCGGCTACACCG
4501 CGTACCTGTGCTATTATGTTGGCATAGGTTACATCGTGTGTCTACCCTCGCTATATAAGA
4561 GAGGCCTAGGATACAAGTGTCCTACTAGGACACGGCTCCATATCATATCTAAACACAATA
4621 CTACTCAGAGTCCAACGGTAACCTACCTTGTACACAATATTCGACACAACTCTAACAAAC
4681 TCTGCCATGGGAATATTCTCCACCACCTTGAATTCGTCAATGCATCAAACTTTCATGTAC
4741 ATTGGACTTGAGCTTATCCCGTGAGTACCGCTGCTGCACCTGCAACTTGTAGTCCCTTCT
4801 TTTCTTGACCACAATCAACACTCGAGCAAAATTAAATTCCACATTACTCTAGTTAGCGCT
4861 CCCAACTTCGAGAGTATTCGCCCAACGCCATCACACATTGATCATTGACCTGCGTGAAAG
4921 TGAACAACTCACATATTGGGTGTCACACATAAGAGTTACCTGAACTCAACATCACAGCTC
4981 CTTTCTCGACTGTCCGTCTGAAACTTGAAGAAATTTCACCATTGCTTGTAGTCATCCCAA
5041 GTCAAATTCGCAGTTGTCTCACCACATGTATGACGACCAGAGCCCTGGCCCGTTTCCATG
5101 TCCCATGCTTACCGCACGCCTCCCCGCTATTACCGCGTCGAGCCTCCGCTGTCTTGGTCG
5161 AGTCTCAAGGGTCGCGAACCCAAACCACTCAACCCCCACTGTAGAGTACCACCGATCATA
5221 GACGACTGATGACGAGTTTCACGCTTCCATCAGACCACTGGGCTCCAATCTGAACTACGT
5281 GTCACTCGCTTTTCTTACCCGCTGAATAGGCTTCGACTCTCTGTGCACTTCGCCGTTGGC
5341 CCTCAATCCAGCTCAACCATCAACATGACTCCATGGTGGTTGTATGTGCCTCCCCGCGCC
5401 TCATCAACTCCGAGCTCTCATGTGCGCACCGTCTTGAATCAACCCCGCGCCATAGTCTTG
5461 TCGAAGCCACACAAGCCTCGGGCTTGTGACACGTGTTTCCACGTCCAGAGGTCAGCCACC
5521 ATCGGCATCACGCTCCTATGGCGTCGTCGCCGATGTCACCGCCATCTTCTGTACCAACCG
5581 ACTTGCGTCGATTCGTCAGACTGCCTAGTCCGACCCAGCCGAACTCGTTCGACTGTACGA
5641 CGCTTCAAGGCTCCCAAATCATCTTCCATGAATTTCCATCCATCACATGGATAATTCCAT
5701 CTTATCTGAACATGAGTTCCCGCAACGCCTTCAAGCATCCTTGTTGTGCCAACAATCTAA
5761 CACCCATAGTCTGTCATACACCAAGGTTTTCAGTTCCATTGAAGCTTCTCCACCTCAATC
5821 CTGATGACCGATGGCGTCATGATTCTTCGTGCAGCTGACCTTGTGAAAAATTAACGGAGC
5881 CCCACCGCATTGCTCCAAGTACACAAGCAGAATTTCCGCTAACTAATAGCAACTCCAGCC
5941 TAAACTTTGGCCTTCACGTGATCTTCCTTGGCTCAACTATCAATGCTAAGCAAGTTGCTT
6001 TTCTCTGCCAAAGCCTCCTGCTAGATGCGCAAGTACCAATGAAGAAACTTTTTGCCGAGG
6061 TGATCTGCCTCTGTCCGTGTCGTAGGAGGACTCGGTCCTCTTTTTGGGSCAAAGCAAATC
6121 TCAGATGTGTGTATGTACGCCTCGCACGCACACGGGSTYGGCAACAGCMCCACCGCGCCC
6181 GTCCCATGGGCCTGGCTCGCTAGCAACCAGCGCATGCATCGCGCCTGCTGCTCGCACGCA
6241 TTACGCCACCGCCGAGTCGATGTTCCATATGCGGGCTGTATTGACGTACTCGCCTTCCAG
6301 CCGCCGCTTCAGACGCTCGCCCGATTGATCCACAATCGCCGCACGTCTCGCCTGCGGCGG
6361 CCAAGCAACACGCCTCTCTCTAATTGGCTTTTCCGAGAACGTCACAACCGTCAAGCACAT
6421 CTGCTTCTTAGATTGCTCCTTGCAGAACTCATCAACGACGTCTCGCCGCTCTTCTTCCGC
6481 TCCAACTGCAATCAATC
```

Table 6.7 Triticale ω-secalin homologue (T-ω-SEC) gene promoter elements. *Cis*-elements and motifs identified using PlantCARE (Lescot et al., 2002) in the promoter sequence upstream of the rye ω-secalin start codon (GenBank accesion no. AF000227) identified in triticale (AC Certa) through genome walking.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
ABRE	*Arabidopsis thaliana*	-1250	+	7	TACGGTC	cis-acting element involved in the abscisic acid responsiveness
ABRE	*Arabidopsis thaliana*	-1285	+	6	TACGTG	cis-acting element involved in the abscisic acid responsiveness
ACE	*Petroselinum crispum*	-1287	-	9	GACACGTATG	cis-acting element involved in light responsiveness
AE box	*Arabidopsis thaliana*	-44	+	8	AGAAACAT	part of a module for light response
AE box	*Arabidopsis thaliana*	-1017	-	8	AGAAACAA	part of a module for light response
ARE	*Zea mays*	-571	+	6	TGGTTT	cis-acting regulatory element essential for the anaerobic induction
Box I	*Pisum sativum*	-649	+	7	TTTCAAA	light responsive element
Box III	*Pisum sativum*	-492	-	11	atCATTTTCACt	protein binding site
CAAT box	*Brassica rapa*	-10	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Glycine max*	-36	+	5	CAATT	common cis-acting element in promoter and enhancer regions
CAAT box*	*Hordeum vulgare*	-135	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-458	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-472	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-557	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-646	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-772	-	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-908	-	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-911	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1104	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Hordeum vulgare*	-1211	+	4	CAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Arabidopsis thaliana*	-1213	+	6	gGCAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1218	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAAT box	*Brassica rapa*	-1389	+	5	CAAAT	common cis-acting element in promoter and enhancer regions
CAT box	*Arabidopsis thaliana*	-577	+	6	GCCACT	cis-acting regulatory element related to meristem expression
CGTCA motif	*Hordeum vulgare*	-770	-	5	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
G Box	*Pisum sativum*	-960	-	6	CACGTT	cis-acting regulatory element involved in light responsiveness
G Box	*Antirrhinum majus*	-1285	-	6	CACGTA	cis-acting regulatory element involved in light responsiveness
G Box	*Pisum sativum*	-1333	+	6	CACGTT	cis-acting regulatory element involved in light responsiveness
G Box	*Pisum sativum*	-1352	-	6	CACGTT	cis-acting regulatory element involved in light responsiveness

* element/motif shown in Figure 6.8

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.7 Continued.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
G box	*Zea mays*	-1377	+	6	CACGTC	cis-acting regulatory element involved in light responsiveness
GAG motif	*Arabidopsis thaliana*	-726	-	7	AGAGAGT	part of a light responsive element
GCN4 motif*	*Oryza sativa*	-222	-	7	TGTGTCA	cis-regulatory element involved in endosperm expression
GCN4 motif*	*Oryza sativa*	-435	-	7	TGTGTCA	cis-regulatory element involved in endosperm expression
GT1 motif	*Arabidopsis thaliana*	-949	+	6	GGTTAA	light responsive element
MBS	*Arabidopsis thaliana*	-1020	-	6	CAACTG	MYB binding site involved in drought-inducibility
MBS	*Zea mays*	-1248	+	6	CGGTCA	MYB Binding Site
MRE	*Petroselinum crispum*	-952	-	7	AACCTAA	MYB binding site involved in light responsiveness
OBP-1 site	*Arabidopsis thaliana*	-1116	+	10	TACACTTTTGG	cis-acting regulatory element
Prolamin box*	*Hordeum vulgare*	-324	+	16	tgtagTGTAAAGtaaaa	
Skn-1 motif*	*Oryza sativa*	-223	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif*	*Oryza sativa*	-548	+	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif*	*Oryza sativa*	-1246	+	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif*	*Oryza sativa*	-1426	+	5	GTCAT	cis-acting regulatory element required for endosperm expression
Sp1	*Zea mays*	-879	-	5	CC(G/A)CCC	light responsive element
TATA box*	*Arabidopsis thaliana*	-85	+	4	TATA	core promoter element around -30 of transcription start
TATA box*	*Arabidopsis thaliana*	-86	-	5	TATAA	core promoter element around -30 of transcription start
TATA box*	*Brassica napus*	-87	+	6	ATTATA	core promoter element around -30 of transcription start
TATA box*	*Ac*	-105	+	7	TATAAAT	core promoter element around -30 of transcription start
TATA box*	*Nicotiana tabacum*	-107	+	9	tcTATAAAta	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-179	-	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-264	+	5	TAATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-383	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-440	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-442	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Brassica napus*	-443	+	6	ATATAT	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-718	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Brassica napus*	-719	+	6	ATATAT	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-729	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-737	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Glycine max*	-740	+	5	TAATA	core promoter element around -30 of transcription start

* element/motif shown in Figure 6.8

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Table 6.7 Continued.

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
TATA box	*Lycopersicon esculentum*	-830	+	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Arabidopsis thaliana*	-1242	+	4	TATA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1394	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1416	-	5	TTTTA	core promoter element around -30 of transcription start
TATA box	*Lycopersicon esculentum*	-1503	-	5	TTTTA	core promoter element around -30 of transcription start
TATCCAT/C motif	*Oryza sativa*	-536	+	7	TATCCAT	
TATCCAT/C motif	*Oryza sativa*	-897	-	7	TATCCAT	
TATCCAT/C motif	*Oryza sativa*	-1438	+	7	TATCCAT	
TC-rich repeats	*Nicotiana tabacum*	-312	-	9	ATTTTCTTCA	cis-acting element involved in defense and stress responsiveness
TC-rich repeats	*Nicotiana tabacum*	-1042	-	9	ATTTTCTCCA	cis-acting element involved in defense and stress responsiveness
TGA element	*Brassica oleracea*	-293	-	6	AACGAC	auxin-responsive element
TGACG motif	*Hordeum vulgare*	-770	+	5	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
Unnamed 1	*Zea mays*	-59	-	5	CGTGG	
Unnamed 3	*Zea mays*	-59	-	5	CGTGG	
Unnamed 4	*Petroselinum hortense*	-551	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-712	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-931	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-966	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1190	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1275	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1445	+	4	CTCC	

* element/motif shown in Figure 6.8

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

Several *cis*-acting regulatory elements were identified using PlantCARE (Table 6.7). These elements include some involved in seed-specific regulation such as the GCN4-motif (rice, Muller and Knudsen, 1993), the prolamin-box (barley, Vicente-Carbajosa et al., 1997) and the skn-1 motif (rice, Takaiwa et al., 1991) of which 2, 1 and 4 were identified of each respectively. Other elements include those involved in light response, hormone and stress response, meristem expression, MYB transcription factor binding, protein binding and anaerobic induction (Table 6.7).

Although the sequence identified in triticale aligned to a sequence already present in GenBank nucleotide database with 97% identity, alignments to sequences in the patent database revealed that only 54 bp of the sequence upstream of the ATG start codon and the downstream coding sequence aligned to sequences in the patent database (light blue nucleotides, Figure 6.9). The majority of these sequences (nucleotides -54 to 1255 and 4374 to 4681) match sequences in US patent number 7214786 (Kovalic et al., 2007, 96% and 88% identity respectively). Two sequences in the patent database; WO patent no 03077643 (Bauer, 2003) and WO patent no. 03078629 (Kock and Bauer, 2003), also aligned to ω-secalin from nucleotides 1-1010 with 77% identity. The T-ω-SEC promoter sequence upstream of these alignments including putative TATA boxes and CAAT boxes and *cis*-regulatory elements did not align to sequences in the patent database. The absence of existing patents for the majority of the T-ω-SEC promoter means that this promoter could be used for further plant studies or bioproduction purposes without infringing on existing patents.

6.3.4.2 Transient expression with triticale ω-secalin homologue promoter deletion fragments

Four T-ω-SEC promoter deletion fragments of decreasing lengths were cloned into the GFP expression vector pLRCWHEAT3-pevAgdv and co-bombarded with the anthocyanin protein expression vector pBC17 into triticale seed longitudinal sections and leaf tissues attached to the plant. Promoter fragments were designed based on the locations of putative regulatory elements expected to play a role in the regulation of ω-secalin seed-specific expression. All four promoter fragments tested were capable of directing transient expression in the seed with varying patterns of expression among seed tissues (Table 6.8, Figure 6.10). Expression of GFP under the control of any of the four T-ω-SEC promoter deletion fragments was not observed in leaf tissue.

The transient expression patterns observed in the seed tissues using the four promoter deletion fragments suggests complex regulatory interactions control the tissue-specific expression patterns of ω-secalin. The smallest promoter fragment (-208 to +1) was capable of transgene expression in all seed tissues while expression under the control of the next largest promoter fragment (-287 to +1) was restricted to the endosperm and seed coat. This suggests that the GCN4 motif found in the region between -208 and -287 bp upstream of the transcription start has a role in the negative regulation of expression in the embryo. The -475 to +1 promoter fragment produced transgene expression in both the embryo and endosperm but not in the seed coat (Table 6.8). This pattern of expression suggests that the prolamin box (at position -324 Figure 6.9), an element previously shown to be important in barley seed storage protein gene expression (Vicente-Carbajosa et al., 1997), is important in fertilized seed tissue (embryo and

Table 6.8 Transient expression results using triticale ω-secalin homologue (T-ω-SEC) promoter deletions. Expression vectors containing T-ω-SEC promoter deletion fragments linked to the green fluorescent protein (GFP) gene were bombarded into plant tissues using the Helios® Gene Gun (Bio-Rad Laboratories). Tissues were then screened for GFP expression under blue light using a stereo microscope 48 hours after bombardment. (+) presence of GFP expression, (-) absence of GFP expression.

ω-Secalin Promoter Fragment	**Embryo**	**Endosperm**	**Seed Coat**	**Leaf**
-208 to +1	+	+	+	-
-287 to +1	-	+	+	-
-475 to +1	+	+	-	-
-1536 to +1	+	+	+	-

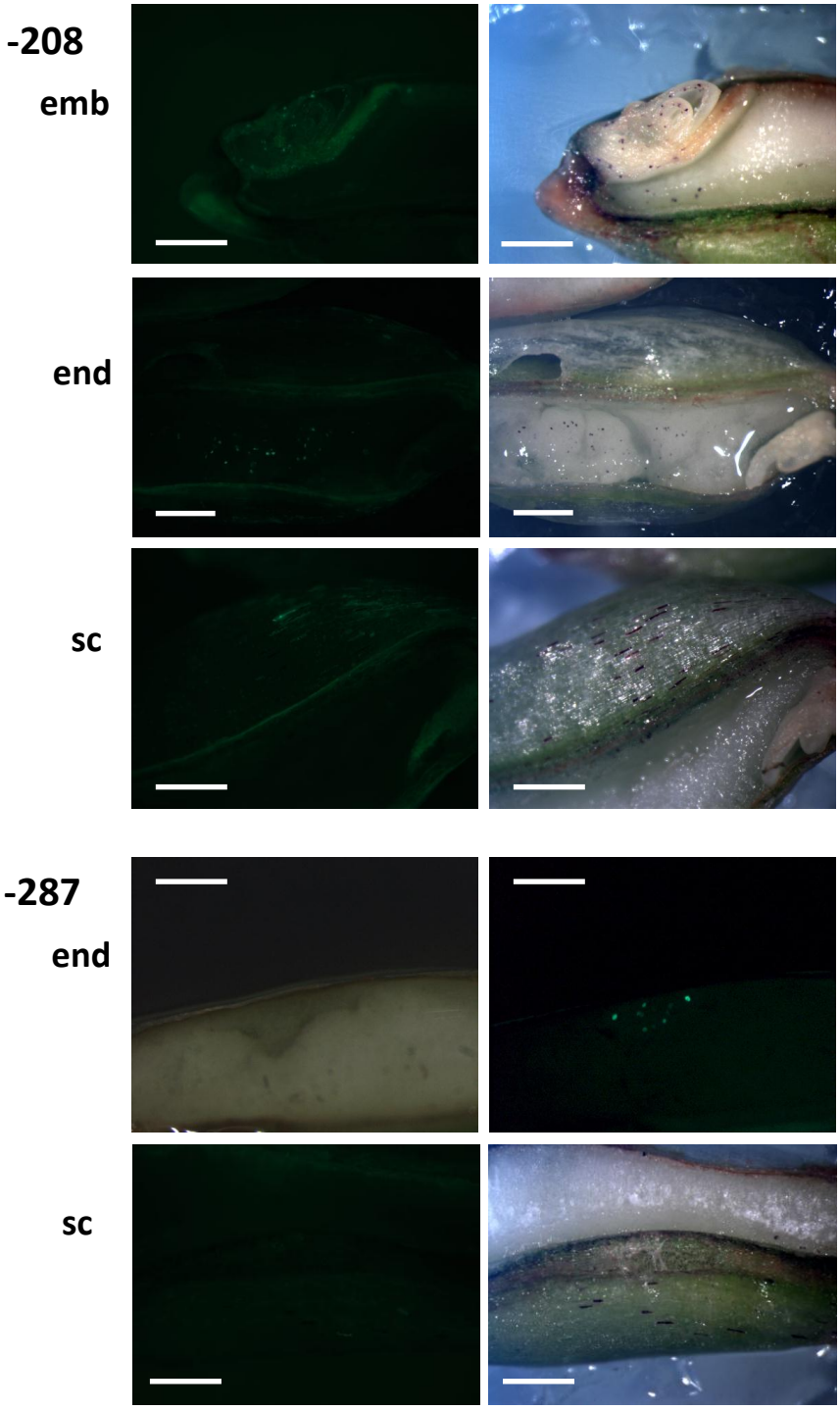
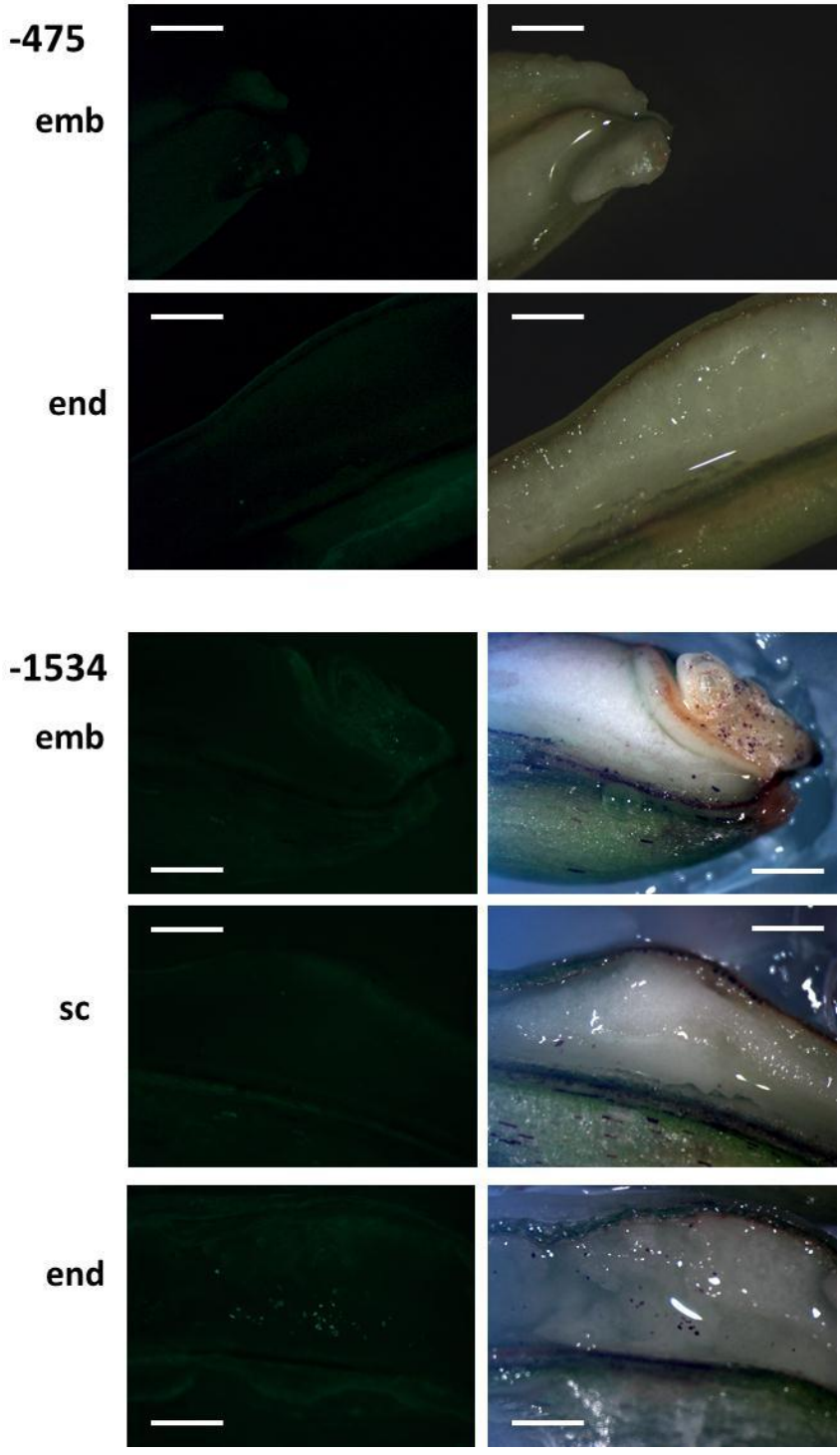


Figure 6.10 Transient expression using triticale ω-secalin homologue (T-ω-SEC) promoter deletion fragments in triticale seed tissues. Transient expression of green fluorescent protein (GFP, bright green) directed by T-ω-SEC promoter deletion fragments in the GFP expression vector pLRCWHEAT3-pevAgdv observed under blue light (left) and anthocyanin expression (purple) directed by the control vector pBC17 in the same tissue section observed under white light conditions (right) in triticale seed



longitudinally dissected halves. Promoter fragments used are indicated by the length of promoter upstream of the T-ω-SEC start codon (numbers to the left). Triticale seeds were dissected longitudinally along the ventral groove 12 days post anthesis and placed on Murshige and Skoog media then bombarded using the Helios® Gene Gun (BioRad). Observations were made 48 hours later. The tissues evaluated are indicated by: emb, embryo; end, endosperm; sc, seed coat. Scale bars (white) = 1 mm.

endosperm) expression rather than in maternally derived tissues (seed coat tissues) and may have a role in negative regulation of expression in the seed coat but that further upstream elements may be required to restrict expression to the endosperm. The full length fragment of the triticale isolated ω-secalin promoter (-1536 to +1 bp) contains three additional Skn-1 motifs dispersed throughout the promoter region (Figure 6.9). These elements have been shown to be required for endosperm expression in rice (Washida et al., 1999); however, expression using this promoter fragment was observed in all seed tissues and was not restricted to the endosperm (Table 6.8). Either additional elements further upstream may be required to complete the complement required for an endosperm-specific expression pattern or the regulatory region involved in ω-secalin expression may be limited to the region between the translation start site and a region between -475 and -1,536 bp. While there were several putative elements identified in this region, using additional fragments with lengths between 475 and 1,536 bp for use in transient expression studies might reveal specific elements in this region responsible for the endosperm-specific expression pattern of ω-secalin.

6.4 DISCUSSION

6.4.1 Triticale agglutinin

Genome walking using primers designed from WGA-A in triticale genomic DNA libraries identified a substantial amount of T-WGA promoter sequence (2.7 kb) upstream of the WGA-A cDNA sequence transcriptional start site (Figure 6.4). Previous to this study, genomic sequences and promoter information were unavailable for WGA isolectins.

The 3' region (109 bp) of the identified sequence is complementary to the 5' end of the open reading frame of the published sequence for WGA-A, however the 14 bp differences observed in the sequence suggests that it may correspond to another WGA isolectin. Three of the identified differences (at nucleotide positions 90, 97 and 98 in Figure 6.4) from WGA-A are consistent with the WGA-B sequence (Raikhel and Wilkins, 1987). The remainder of the differences between the identified sequence and WGA-A are upstream of the sequence complementary to available published sequence for WGA-B (Smith and Raikhel, 1989), leaving the possibility that the promoter sequence identified might be the triticale homologue of WGA-B originating from the wheat B genome in hexaploid triticale. The hexaploid triticale used for this study does not contain the wheat D genome ruling out the possibility that the sequence identified is the promoter sequence to WGA-D despite the same level of identity (89%) of the identified sequence to WGA-A and WGA-D. WGA-like lectins are also found in rye embryos (Mishkind et al., 1983) which are virtually indistinguishable biochemically and immunologically from WGA lectins (Peumans et al., 1982). This suggests that the

sequences could be similar. In a blast of our T-WGA sequence against a reference library assembled rye 454 and Illumina® cDNA reads (Xu et al., 2012, unpublished), one homologous contig was identified. The rye sequence transcription start site aligns to nucleotide position -42 in the T-WGA sequence and there is no promoter sequence available for the rye sequence. In the region between the rye TSS and the genome walking primer AGG 110r, the percent identity to T-WGA is 91%. In the region between the suspected TSS of WGA-A and primer AGG 110r the percent identity between T-WGA and WGA-A is 81%. There is a possibility therefore, that the sequence identified is the promoter sequence from a rye WGA-like lectin in the rye genome of triticale. Using the same genome walking primers and techniques in a genomic library constructed from rye rather than triticale might shed light on whether the sequence identified is from a WGA isolectin or a rye WGA-like lectin and the degree of similarity between the triticale, wheat and rye lectin promoter sequences.

Three possible ATG codons have been previously identified in WGA-A (Smith and Raikhel, 1989). The WGA-A cDNA sequence published in GenBank (accession no. M25536.1) starts with the first of these (-6 bp, Figure 6.4) start codons. The sequence identified differed from WGA-A in the region of the first start codon resulting in only two of the three possible start codons remaining (at nucleotides 1 and 4, Figure 6.4). If the sequence identified is the homologue of WGA-B the resulting protein would differ from WGA-A not only in the relative position of the possible start codons but with an additional seven amino acid differences in the signal peptide region (Figure 6.11). The resulting signal peptide region would be three amino acids shorter than the WGA-A signal peptide region due to the later start codon and a triple nucleotide deletion (Figure

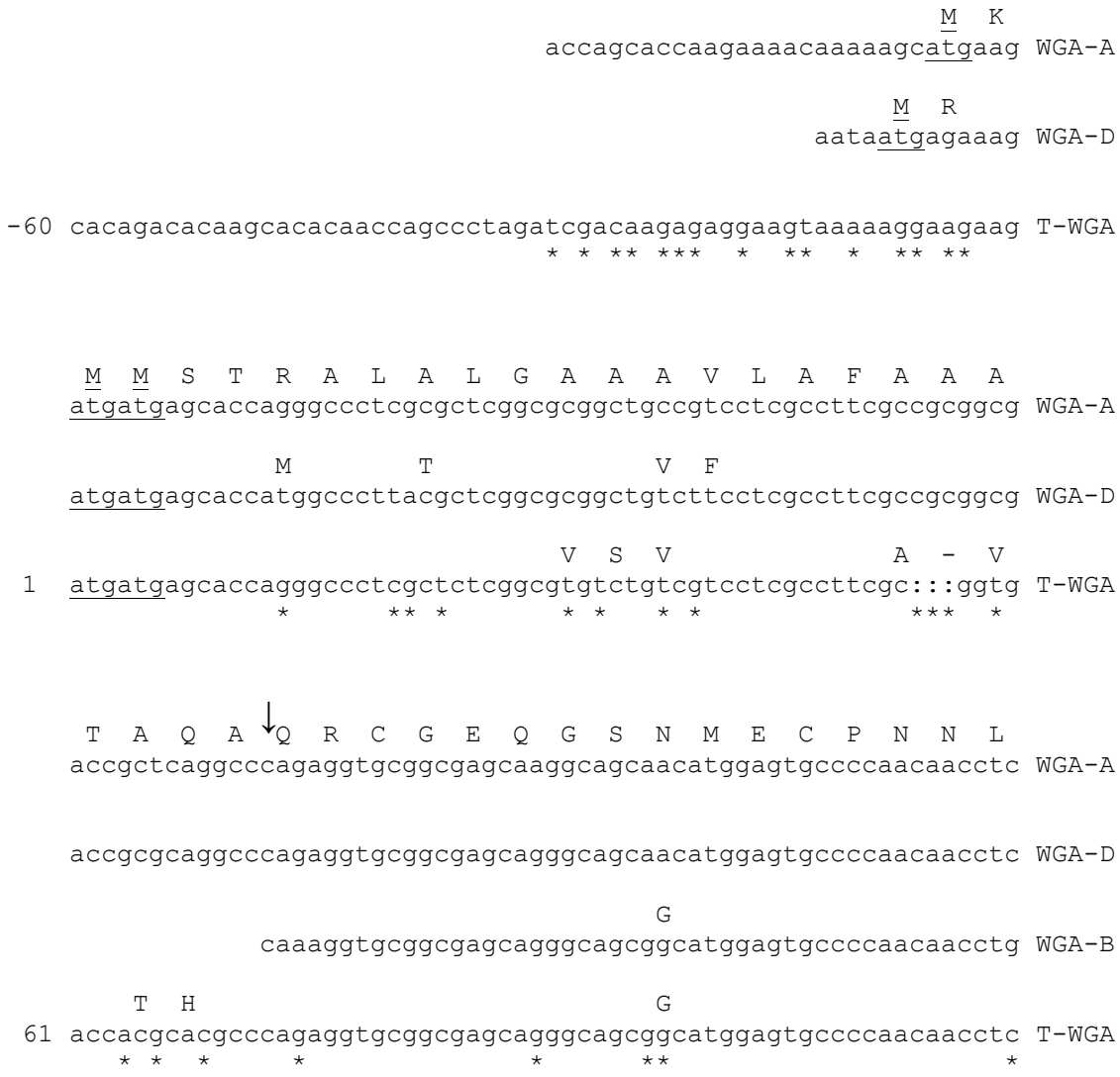


Figure 6.11 Nucleotide sequence and amino acid aligments of the putative triticale wheat germ agglutinin homologue (T-WGA) signal sequence to: WGA isolectins WGA-A (GenBank accession no. M25536.1, Smith and Raikhel, 1989) WGA-D (GenBank accession no. M25537.1, Smith and Raikhel, 1989) and WGA-B (GenBank accession no. J0296.1, Raikhel and Wilkins, 1987) signal peptide sequences. Numbering of nucleotides begins at the putative start codon for T-WGA. The arrow (at nucleotide 73) indicates the beginning of the open reading frames which encode the mature WGA monomers. Asterisks indicate differences in nucleotides between the four isolectins. Amino acid translation is shown above the WGA-A nucleotide sequence with differences shown above the other isolectins. Gaps introduced for optimal alignment are indicated by colons in the nucleotide sequence and dashes in the amino acid sequence. All possible start codons previously identified by Smith and Raikhel (1989) for WGAs are underlined.

6.11, Smith and Raikhel, 1989). It has been suggested that the genes for the individual WGA isolectins are expressed in a differential or tissue-specific manner (Smith and Raikhel, 1989). Differences in the signal peptides rather than differences in gene expression may account for differences in WGA protein expression patterns. The high degree of similarity among the WGA isolectins (90%), in the nucleotide sequences and inbiochemical and immunological behavior, presents difficulties in distinguishing the patterns of expression specific to each of the wheat isolectins (Peumans et al., 1982; Smith and Raikhel, 1989) and among WGA-like lectins in other species of the *gramineae* (Mishkind et al., 1983). In the study of WGA-like lectins, triticale would make an interesting case having WGA-A and WGA-B from wheat and a WGA-like isolectin from rye. Determining differences in gene expression patterns might best be accomplished by using probes designed to hybridize to the differing signal peptide regions or to the 5' and/or 3' untranslated regions.

Published sequences for WGA isolectins are from cDNA clones and therefore do not contain upstream promoter sequences. The promoter sequence identified through genome walking upstream of the transcription start site contain some potential regulatory elements that may be involved in the observed tissue-specific expression patters of WGA (Mishkind et al., 1983). The CAT box (-980 bp, Figure 6.4), CCGTCC box (-1631 bp, Figure 6.4) and the OCT element (-641 bp, Figure 6.4) are *cis*-acting regulatory elements related to meristem-specific activation (Lescot et al., 2002). The WGA-like proteins have been immuno-localized in the periphery of the radicle in wheat, rye, barley and rice embryos and the root cap and coleorhiza of rice embryos (Mishkind et al., 1983), regions all very near to the root meristematic region of the embryo. Removal of these elements

by deletion of the promoter region upstream of position -278 did not eliminate expression in the embryo but did cause a change in the specificity of expression in transient expression assays (Table 6.4). In the absence of these elements, expression was observed in all seed tissues suggesting that these elements play a role in restricting the expression to the embryo (Table 6.4).

Barley contains a lectin that is also a WGA-like lectin but differs in its expression pattern from wheat, rye and rice in that it is root-specific and not expressed in the coleoptiles (Lerner and Raikhel, 1989). Identification of promoter sequence from the barley lectin and performing similar expression studies using this promoter in triticale might help reveal which elements are important to the root-specific expression observed in barley and other *gramineae* species. If the barley lectin promoter sequence also contains these elements, deletion studies in barley may be more informative in the absence of native expression in the coleoptiles, especially since the removal of these elements reduced the specificity of the expression pattern.

Reporter gene fusions may also be used to determine if the gene expression is localized similar to the protein or if the gene is expressed in the meristem and the protein translocated to cells adjacent to the meristem. The presence of signal peptides in WGA-like cDNA clones not present in the mature proteins makes this an interesting point for examination (Lerner and Raikhel, 1989; Smith et al., 2003).

Sequence analysis of the promoter sequence using PlantCARE identified a pyrimidine (PY)-rich stretch (-859 bp, Figure 6.4), a *cis*-acting element conferring high transcription levels. This type of element is an attractive candidate for promoter studies when the goal is to find promoters for transgene expression. The PY-rich stretch was

identified in the promoter for the tomato 3-hydroxy-3-methylglutaryl coenzyme A gene 2 (HMG2) in the 5' untranslated leader sequence downstream of the TATA box (Daraselia et al., 1996). Promoter deletion studies showed that for the promoter fragment that did not include this PY-rich stretch, expression was only observed in the embryo but the next larger fragment that did contain the PY-rich stretch (-1087 to +1) transgene expression was observed in all seed tissues. It appears that in the absence of some elements that confer specificity of expression to the embryo, the PY-rich stretch is able to promote expression throughout the seed. This promoter deletion would be an ideal promoter for transgene expression throughout the triticale seed. However, stable transformation will be needed to determine if the transgene expression will be stable throughout the seed over multiple generations and not expressed in other plant tissues and to quantify the level of transgene expression.

The agglutinin promoter identified contained three Opaque2 (O2) sites (-1,254 bp, -2,639 bp and -2,723 bp) two of which are in close proximity to the GCN4 motifs. It has been shown that the maize (*Zea mays* L.) orthologue of the wheat basic leucine zipper (bZIP) protein EmBP-1 (mEmBP-1), a protein that recognizes the O2 box in zein gene promoters, can activate O2-dependent gene expression in a yeast expression system but demonstrates an inhibitory effect on the zein promoter in transient assays using cultured endosperm cells (Carlini et al., 1999). This could suggest that complex interactions of binding transcription factors contribute to expression in the embryo and inhibition in the endosperm resulting in the specific patterns of expression of WGA proteins. Also, O2-dependent zein activation in transient assays using maize suspension cultures relies on the presence of the neighboring prolamin box (Vicente-Carbajosa et al., 1997) an element

absent in the identified WGA promoter. The transgene expression studies using the WGA promoter fragments supports the theory that there are complex interactions of binding transcription factors that contribute to the embryo-specific expression of WGA. In the absence of the two O2 sites near the GCN4 motifs, expression was observed only in the embryo but using the longer promoter fragment containing these elements, expression was observed in both the embryo and endosperm.

The proximity to GCN4 motifs could also be significant as the O2 sites and GCN4 motif have been shown to be functionally related in wheat (Albani et al., 1997). A recombinant O2 protein has also been shown to bind *in vitro* to a GCN4-like motif of a pea lectin gene necessary for seed-specific expression in transgenic tobacco (de Pater et al., 1993). The seed-specific expression of the wheat lectin WGA and the presence of both O2 sites and GCN4 motifs overlapping each other may therefore be related. The GCN4 motif is known for endosperm expression (Muller and Knudsen, 1993; Albani et al., 1997), but WGA is expressed in embryo cells and not endosperm. When the GCN4 motifs were removed in the transient expression assays, expression was no longer observed in the endosperm supporting the role of the GCN4 motifs in endosperm expression and suggesting negative regulation of these elements during native WGA expression.

The presence of an AACA motif (-1,865 bp) a distal *cis*-regulatory element shown to be involved in the suppression of expression in tissues other than the endosperm (Washida et al., 1999) may either play a different role in the WGA promoter or may also act as a negative regulator, though in tissues other than the embryo. There is also a possibility that because the AACA motif is on the negative strand in the WGA promoter,

it has no role in regulating the expression of WGA. Transient expression studies were unable to clarify the role of this element in WGA expression. Site directed mutagenesis of this element in a few of the promoter fragments may be more informative in determining its role in WGA gene regulation.

Several Skn-1 motifs were also identified in the WGA promoter sequence (Figure 6.4). The Skn-1 motif (A/GTCAT) is a half site of the palindromic sequence recognized by bZIP protein GCN4 and is usually flanked by an AT-rich region (Blackwell et al., 1994). Introducing base substitutions into the Skn-1-like motif of the rice storage protein glutelin gene (*GluB-1*) promoter fused to the GUS reporter gene caused a 35% drop in GUS activity (Washida et al., 1999) suggesting that these elements may play a role in high levels of expression of seed storage proteins. The same study also suggested that a combinatorial effect of at least three of these elements (AACA, GCN4, ACGT and Skn-1) was necessary for high levels of seed-specific expression of *GluB-1* gene. Similar relationships among these elements may be responsible for WGA expression since AACA, GCN4 and Skn-1 motifs are all present in the identified promoter sequence (Figure 6.4). When all of these elements were present in the transient expression assays (in the -2734 to +1 promoter fragment), transgene expression was observed in both the embryo and endosperm but without the GCN4 motifs expression was only observed in the embryo (Table 6.4). Using the full length promoter element sequenced here and site directed mutagenesis in each of these elements individually may help to clarify the precise interactions that occur among these elements to regulate WGA expression in triticale.

Two RY elements, one distal (-927 bp) and one proximal (-332 bp) were also identified in the WGA promoter sequence (Figure 6.4). The RY element has been shown to be necessary for seed-specific regulation in legume species and sunflower (Prieto-Dapena et al., 1999). Seeds in dicotyledonous plants such as legumes and sunflower are composed mainly of the embryo with little endosperm in comparison to monocots such as wheat (Esau, 1977). Thus the presence of this element could be significant to the embryo-specific expression of the WGA. Transient expression studies revealed that the promoter fragment containing only the proximal RY element was able to confer embryo-specific expression suggesting that this element may indeed have a role in restricting WGA expression to the embryo in triticale. The promoter fragment containing both RY elements was able to direct transgene expression in all seed tissues in transient expression studies, but this fragment also contained other elements including the PY-rich stretch.

Several G boxes (8) were identified in the WGA promoter sequence using PlantCARE (Table 4.2). Although, PlantCARE lists these elements as being involved in light responsiveness, they have been shown to play a role in the regulation of plant genes by a variety of environmental signals (e.g. light quality, daylength) and physiological cues (e.g. ABA, MeJA, auxin and sugar response, Williams et al., 1992). These elements were present in several of the promoter fragments used for transient expression. How exactly these elements function in the regulation of WGA is unclear and will require further analysis.

Several of the elements (including AE box, Box I, GAG motif, GATA motif, GCN4 motifs, MBSII, O2 sites, Skn-1 motifs, TCCACCT motif and Unnamed-4 elements) were identified very distal to the gene (more than 2 kb). Transgene expression

studies using promoter fragments indicates that at least some of these distal elements are very likely to play a role in the embryo-specific regulation of WGA expression as changes in expression patterns were observed when these elements were removed (Table 6.4). While short promoter fragments were able to confer embryo-specific expression, mid-length fragments (between 598 and 2,214 bp in length) did not show the same specificity. The use of a longer promoter fragment (2,485 bp) allowed the embryo-specific expression to be restored. Despite the -2,485 WGA promoter having the ability to confer embryo-specific expression, the full length promoter fragment was able to direct expression in both the embryo and endosperm suggesting that there are complex interactions that include some or many of these distal elements in the regulation of WGA gene expression in triticale. Deleting a region of the promoter sequence between these distal elements and the proximal TATA box may also help to understand their function and determine if their distal position is required for their role in WGA gene regulation or if they would have the same or different effect in a more proximal position. In contrast to these distal elements, a 3-AF1 binding site and a GA motif were identified proximal to the gene (-26 bp and -9 bp respectively, Table 4.2) between the TATA box and the ATG start codon. It is unlikely that these elements, previously shown to be involved in light response, play a role in the regulation of WGA expression downstream of the TATA box. However, there are examples of elements downstream of the TATA box that have been shown to play roles as enhancers (Daraselia et al., 1996) and even light responsive mRNA stabilizers (Dickey et al., 1998). In the pea *rbcS-3A* promoter a binding site for 3-AF1 is present between -31 and -51 bp, and although this is directly upstream of the TATA box it is still proximal to the transcription start site (Lam et al., 1990). In the light

responsive pea *rbcS-3A* promoter the 3-AF1 site is near (within 100 bp) to other binding sites (Box II, Box III) that together confer light regulation (Lam et al., 1990). Similarly in the rice *cab1R* promoter, the 3-AF1 site is near to Box III and G box and octamer repeat (OCT-R) elements and deletion in the 3-AF1 site did not eliminate light responsiveness (Luan and Bogorad, 1992). There is a possibility that the 3-AF1 protein binds to the WGA promoter in response to some factor other than light. It has been shown that there may be different forms of the 3-AF1 protein in different plant organs that may bind to different 3-AF1 binding sites supporting an alternative role to light regulation of this site in the WGA promoter (Lam et al., 1990). Transient expression results make it difficult to determine if these very proximal elements have a role in WGA regulation since the smallest promoter fragment used (-278 to +1) enabled expression in all seed tissues and larger promoter fragments showed variation in seed expression patterns indicating that it is the more distal elements that are important to the regulation of WGA expression. Based on its presence in other light-responsive promoters, this 3-AF1 element may not have a role in the tissue-specificity of the WGA but rather in the timing of WGA expression in response to light or other factors.

The presence of several *cis*-acting regulatory elements that have been shown to be important for the seed-specific expression of several other seed storage proteins indicates the identified promoter sequence is likely to play a regulatory role in the expression patterns of WGA. Transient expression of GFP using WGA promoter fragments of increasing length revealed that the regulation of the tissue-specific expression of WGA is complex and likely involves several regulatory elements including some that are over 2 kb upstream of the transcription start site. Deletion or base substitution in each of these

elements will help to further understand their specific role in WGA regulation. Additionally removing sequence between these elements and the transcription start site may also help to define their role in WGA expression regulation in conjunction with other transcription factor binding sites identified in the promoter region.

The WGA promoter sequence identified is mainly free (aside from -1 to -12 bp) of infringing patents. The gene sequence on the other hand, aligns to sequences in the patent database. The region between -12 bp and 960 bp aligns to sequence 7405 in US patent no. 7214786 (Kovalic et al., 2007) with 95% identity (blue nucleotides, Figure 6.4). This region does not include the TATA box or any of the identified regulatory elements presumed to be involved in the tissue-specific expression patterns of WGA. The patent does not make any claims on the promoter sequence identified and therefore the sequence can be used for directed transgene expression.

The transient expression studies allowed us to determine which regulatory elements likely have a role in WGA regulation but have also identified promoters that are suitable for achieving various transgene expression patterns. The small WGA promoter fragment (-278 to +1) is a suitable promoter for seed-specific expression since expression was seen in the embryo, endosperm and seed coat but was not observed in leaf tissues. The -598 to +1 promoter fragment is a suitable promoter for embryo-specific transgene expression since GFP expression was not observed in any other tissues. Stable transformation of promoter:transgene DNA will be necessary to evaluate the specificity of these promoters in an increased number of tissues throughout the plant and to quantify the level of transgene expression by these promoters. Following stable transformation, the evaluation of transcription directed by these promoters should be completed using

quantitative PCR. Gene product expression directed by these promoters should be evaluated using protein isolation techniques such as GUS staining. These evaluations should also provide information to the strength of the promoters in the various tissues evaluated.

6.4.2 Triticale triticain

Genome walking using primers designed to the barley thiol protease aleurain coding sequence (GenBank accession no. X05167) in triticale genomic DNA has identified triticale promoter sequence for the triticale homologue of wheat triticain γ (GenBank accession no. AB267409). A total of 1,348 nucleotides of genomic sequence upstream of the putative translation start site were identified for the triticale homologue of wheat triticain γ (T-triticain-γ). The mRNA sequence for wheat triticain γ was added to the public database on July 30, 2006, after the primers were designed from the barley thiol protease aleurain and the sequence was amplified. No genomic DNA sequence is available in the public database from wheat for the triticain gene. The barley aleurain sequence is interrupted by 7 introns and the genome walking primers were designed within the second exon and consequently identified an intron in the wheat triticain γ sequence as well (violet nucleotides, Figure 6.6). The 5' portion of the T-triticain-γ sequence including intron one has a very high G+C content (around 74%) similar to the 5' portion of the barley aleurain gene (Whittier et al., 1987).

Aleurain homologues are present in many plants, both monocots and dicots, and have conserved regions in their vacuolar targeting domains located between amino acid residues 24 to 45 (corresponds to nucleotides 88 to 123 in Figure 6.6, Rogers et al.,

1997). Sequencing and comparison of upstream promoter sequences from a wide variety of aleurain homologues may also reveal conserved regulatory sequences as well as give insight into the regions responsible for the regulation of these genes in the aleurone layer of seeds. Similarity of the identified T-triticain-γ sequence upstream of the start codon to that of barley aleurain gene (85% identity in the region -1 to –223 bp), including the TATA signal, suggests that there may be conservation among regulatory sequences of aleurain homologues.

In the identified triticale promoter sequence, there were motifs that are known to be involved with hormone response (ABRE, CGTCA and TGACG motifs) that may be active in this promoter for aleurone-specific expression. Barley aleurain has been shown to be expressed in the aleurone cells under the control of plant hormones gibberellic acid (GA) and abscisic acid (ABA, Whittier et al., 1987), so the presence of hormone response elements in the identified triticain gene promoter sequence was not surprising. TGACG motifs, have been identified as methyl jasmonate (MeJA)-responsive regions in the promoter of a barley grain lipoxygenase 1 gene and have been shown to be associated with their complementary sequence 15 bp downstream (Rouster et al., 1997). While the complement to the TGACG motifs (-457 and -693 bp) is not present in the T-triticain-γ promoter in the same spatial context (15 bp), there are two copies of this element in the T-triticain-γ promoter suggesting that depending on the chromatin organization, there may be a similarity in MeJA responsive transcription factor binding to these elements. Reporter gene expression was observed in the aleurone cells in transient expression assays using promoter fragments containing these elements but was also observed using a short promoter fragment that did not contain these elements (Table 6.6). The transient

expression analysis was therefore unable to confirm the role of these elements in T-triticain-γ expression. Additional deletion or mutation studies in conjunction with MeJA treatment may be required to assess their relevance to T-triticain-γ promoter activity.

Three Skn-1 motifs are present in the T-triticain-γ promoter sequence; however, the motifs previously described to be required for cooperative interaction with this motif for endosperm expression (AACA, GCN4, ACGT) are not present (Washida et al., 1999). It is possible that some of these motifs are present further upstream than the sequence identified by genome walking and another round of genome walking would be worthwhile to identify more potential regulatory elements. Alternatively these Skn-1 motifs may not have the same involvement in expression in the aleurone cells, a specialized cell layer of the endosperm, compared to the starchy endosperm. Alternatively, these elements may play a role that does not require the same cooperative interaction as in other seed storage protein gene promoters. While transient expression analysis results showed aleurone cell expression directed by all T-triticain-γ promoter fragments used from 218 bp to 1348 bp, expression was also observed in the leaf tissue using all promoter fragments (Table 6.6). This expression pattern supports a theory that there are additional regulatory elements that control the aleurone-specific expression pattern of triticain by negatively regulating expression in leaf tissues and that these elements were not identified in this study. Based on these results an additional round of genome walking to identify sequence and regulatory elements further upstream should be pursued.

Three GC motifs, elements involved in anoxic response, are present in the T-triticain-γ promoter sequence, one on either side of the TATA box (-84 bp, -106 bp) and

one further upstream (-993 bp). In previous studies of this element, the GC box binding protein was found to be present in both aerobically and anaerobically treated cells at the same level suggesting that it may act more as a general enhancer-like element (Manjunath and Sachs, 1997). Anaerobic response elements have been identified in genes expressed in root tissues during anoxic conditions; however, seeds are not exposed to the same anoxic stress and these elements may therefore play alternate roles in seed tissues. For example, G box elements are found in a wide variety of plant genes associated with diverse expression properties (Manjunath and Sachs, 1997) suggesting that these elements may play different roles in the expression of different genes.

Several motifs were identified as the binding sites for MYB or WRKY transcription factor proteins (Box III, Box W1, CCAAT, MBS, Figure 6.6). Binding of these proteins is often involved in the regulation of gene expression and exploration of their regulatory function in T-triticain-γ may provide more insight into the functional importance of aleurain and its homologues. Interestingly, all of these elements were identified on the negative (complementary) strand and so it may be that binding of their transcription factors would be in the wrong orientation for functional regulation of T-triticain-γ. Point mutations or deletions within these elements in a full length promoter fragment would help confirm their role in T-triticain-γ expression and the importance of strand orientation.

In investigating the possibility of using the identified regulatory sequence as a novel promoter it is important to be sure there the sequence is not covered by existing patents. BLAST searches of the patent database found that the T-triticain-γ sequence, from the TATA box downstream, is covered by existing patents (nucleotides -92 to 64;

US patent # 7132589, Dunn-Coleman et al., 2006 and positions -74 to 1532 excluding the intron; US patent # 7214786, Kovalic et al., 2007) but that the sequence upstream of the TATA box is free of infringing patents. This makes it possible to use the regulatory sequences upstream of the TATA box for tissue-specific transgene expression. The apparent constitutive expression conferred by the T-triticain-γ promoter sequence identified may be useful as an alternative for a constitutive promoter that originates from a cereal crop as many commonly used constitutively promoters have non cereal origins (i.e., CaMV35). Prior to its use for this purpose, transgene expression would have to be surveyed in a wider range of tissues and developmental stages and the level of expression following stable transformation would also have to be determined to ensure adequate levels of expression are achieved.

6.4.3 Triticale ω-Secalin

Genome walking using primers designed to rye ω-secalin in triticale genomic DNA identified over 1.5 kb of triticale genomic sequence upstream of the transcription start site. This T-ω-SEC sequence is homologous to the endosperm storage protein gene rye ω-secalin sequence present in the public database (GenBank accession no. AF000227). This rye sequence also has an additional 1.2 kb of upstream spacer sequence. Genome walking using primers designed from a rye seed storage protein presumably identified the same sequence from the rye genome in triticale. Although wheat has seed storage proteins related to rye ω-secalin (gliadins, Clarke and Appels, 1999), the differences are significant such that their amplification using primers designed from rye ω-secalin is unlikely. Using these primers to amplify sequences from genomic

DNA in other cereal species whose storage proteins are less well studied may make it possible to identify more novel promoter sequences of ω-secalin homologues.

There were some differences between the sequence identified and the ω-secalin sequence in the database (Figure 6.9). The ω-secalins are encoded by a family of 15 genes with slight variations in their spacer sequences (Clarke and Appels, 1999). The identified sequence doesn't precisely match any of the nine sequences previously reported suggesting it may a homologue of one of the remaining six. The -300 element is also highly conserved within the identified sequence (-319 bp to -297 bp, Figure 6.9). The TGAGTCA sequence in the -300 element is referred to as the GCN4 element (Clarke and Appels, 1999); however, it was not identified by PlantCARE. This suggests that although PlantCARE is a very useful database of *cis*-acting regulatory elements, it is not a comprehensive source of all plant elements and analysis of promoter activity is still important to verify the presence and importance of regulatory elements. Within the -300 element, another motif was recognized by PlantCARE, the prolamin box (Vicente-Carbajosa et al., 1997; Guo and Moose, 2003), which suggests it is a highly important region for seed-specific expression despite the lack of recognition of the GCN4 motif. GCN4 motifs as well as Skn-1 motifs were identified on either side of the prolamin box/-300 element (-548, -435, -223 and -222 bp, Figure 6.9). The presence of Skn-1 motifs near the prolamin box and GCN4 motifs suggests that interaction among these regulatory elements may be required for the high levels of endosperm-specific expression of ω-secalin (Washida et al., 1999). Indeed, using transient expression, promoter fragments not containing all three of these elements were capable of expression throughout the seed (Promoter deletion fragments -208 and -287, Table 6.8) but the expression pattern using

longer promoter fragments was more variable supporting their possible role in regulation of the tissue-specific expression pattern. Skn-1 motifs were also identified much further upstream (-1,246 and -1,426 bp). In a transient assay assessing different segments of the rye ω-secalin promoter fused to GUS, an enhancement of *gus* gene expression was observed with the -2.7 kb ω-secalin promoter compared to -0.91 kb promoter (Clarke and Appels, 1998). In our hands, transient expression using a -1,536 T-ω-SEC promoter deletion fragment resulted in a loss of specificity to a broader pattern of expression throughout the seed. My results point to a region between -287 and -475 that is responsible for limiting ω-secalin expression to the endosperm. The -287 fragment eliminated expression in the embryo while the -475 fragment restored expression in the embryo but eliminated expression in the seed coat. A deletion fragment between these two fragments may be capable of endosperm-specific expression, perhaps by eliminating the GCN4 motif (at -435 bp, Figure 6.9) upstream of the Prolamin box (at -324 bp, Figure 6.9). There is a GCN4 motif and Skn-1 motif downstream of the Prolamin box and it is possible the additional GCN4-motif is responsible for the enhancement of expression in tissues other than the endosperm. Further studies using promoter deletion fragments with lengths between -287 and -475 may help to determine if this is the case or if there are other elements responsible for the tissue-specific regulation. Transient expression studies with promoter lengths greater than this, including the -1,536 promoter deletion fragment, should be evaluated for their strength of expression capabilities throughout the seed tissues using quantitative methods such as qRT-PCR and GUS staining. Site directed mutations in the Skn-1 motifs upstream of -0.91 kb may also reveal if they are responsible for the enhancement of expression. Analysis of the homologous published

Table 6.9 *cis*-elements and motifs identified using PlantCARE (Lescot et al., 2002) in the promoter sequence of rye ω-secalin between positions -1,536 bp and -2,756 bp from the ATG start codon (GenBank accession no. AF000227).

Site Name	Organism	Position	Strand	Matrix score.	sequence	function
AC II	*Phaseolus vulgaris*	-2469	+	9	(C/T)T(T/C)(C/T)(A/C)(A/C)C(A/C)A(A/C)C(C/A)(C/A)C	
ARE	*Zea mays*	-1807	-	6	TGGTTT	cis-acting regulatory element essential for the anaerobic induction
Box I	*Pisum sativum*	-2305	-	7	TTTCAAA	light responsive element
Box W1	*Petroselinum crispum*	-2231	+	6	TTGACC	fungal elicitor responsive element
CAT box	*Arabidopsis thaliana*	-2347	-	6	GCCACT	cis-acting regulatory element related to meristem expression
GAG motif	*Hordeum vulgare*	-2732	-	7	GGAGATG	part of a light responsive element
GT1 motif	*Solanum tuberosum*	-2715	+	8	AATCCACA	light responsive element
HSE	*Brassica oleracea*	-2381	-	9	AAAAAATTTC	cis-acting element involved in heat stress responsiveness
HSE	*Brassica oleracea*	-2053	+	9	AAAAAATTTC	cis-acting element involved in heat stress responsiveness
LAMP element	*Spinacia oleracea*	-1810	+	9	CCAAAACCA	part of a light responsive element
LTR	*Hordeum vulgare*	-2419	-	6	CCGAAA	cis-acting element involved in low-temperature responsiveness
MBS	*Arabidopsis thaliana*	-2607	+	6	CAACTG	MYB binding site involved in drought-inducibility
MBS	*Arabidopsis thaliana*	-1755	+	6	TAACTG	MYB binding site involved in drought-inducibility
MBS	*Arabidopsis thaliana*	-2452	+	6	TAACTG	MYB binding site involved in drought-inducibility
RY element	*Helianthus annuus*	-2206	+	8	CATGCATG	cis-acting regulatory element involved in seed-specific regulation
Skn-1 motif	*Oryza sativa*	-2210	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif	*Oryza sativa*	-1606	-	5	GTCAT	cis-acting regulatory element required for endosperm expression
Skn-1 motif	*Oryza sativa*	-1914	+	5	GTCAT	cis-acting regulatory element required for endosperm expression
TC-rich repeats	*Nicotiana tabacum*	-1947	-	9	ATTTTCTTCA	cis-acting element involved in defense and stress responsiveness
TCA element	*Brassica oleracea*	-2004	-	9	CAGAAAAGGA	cis-acting element involved in salicylic acid responsiveness
Unnamed 4	*Petroselinum hortense*	-2748	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1716	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-2111	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1561	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-2384	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1583	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-1864	-	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-2729	+	4	CTCC	
Unnamed 4	*Petroselinum hortense*	-2177	+	4	CTCC	
W box	*Arabidopsis thaliana*	-2231	+	6	TTGACC	

[a] Matrix score refers to the number of nucleotides in the motif that match the core matrix in the database.

sequence (GenBank accession no. AF000227) upstream of the T-ω-SEC sequence identified by genome walking using PlantCARE also identified 3 additional Skn-1 motifs (-1,606 bp, -1,914 bp, -2,210 bp), as well as an RY element (-2,206 bp) which may be involved in the enhancement of expression (Table 6.9). Increasing the promoter sequence from -2.7 kb to -4.5 kb reduced the level of expression in rye (Clarke and Appels, 1998). The rye ω-secalin gene elements are arranged head to tail at the Sec-1 locus of rye (Clarke et al., 1996) and so the upstream untranscribed sequence of one gene is essentially the downstream untranscribed sequence of the gene preceding it. Although there are no obvious negative regulatory elements upstream of the -2.7 kb promoter sequence (Clarke and Appels, 1998), spatial context may play a role in the ω-secalin gene expression. It has been suggested that the 900 bp region around -2.3 kb has characteristics of a scaffold attachment region (SAR) which might create a 7-8 kb loop containing the ω-secalin genes (Clarke and Appels, 1998). This would change the spatial context of the sequence upstream of this SAR to the gene in comparison to the linear format of the -4.5 kb promoter used in transient assays.

The exact mechanisms for regulation of T-ω-SEC expression were not determined in this study; however, the expression patterns observed with the promoter deletion fragments used may be of potential use for transgene expression for agronomic or bioproduction puposes. In particular the -287 promoter fragment enabled expression in the endosperm and seed coat but not the leaf or embryo tissues. This is a useful expression pattern as expression was restricted to the harvestable seed but still not in the embryo, the tissue responsible for the next generation and therefore unlikely to be found

in the early or late developmental root and shoot tissues following germination of seeds not collected by the harvest process.

The ω-secalin sequence aligns in part to sequences in the patent database. The regions between -54 bp and 819 bp and between 4,375 bp and 4,681 bp align to sequences in US patent no. 7214786 (Kovalic et al., 2007) with up to 96 % identity. These regions do not include the TATA box or any of the *cis*-acting regulatory elements presumed to be involved in seed-specific regulation and the patent has no claims on the promoter sequences identified. The region from the start codon to 1,010 bp also aligns to sequences in WO patent no. 03077643 (Bauer, 2003) and WO patent no. 03078629 (Kock and Bauer, 2003) each with 77% identity which is below the 90% homology for which the patents have claims. These patents also have no claims on the promoter sequence. The lack of infringing patents on the promoter sequence of ω-secalin will allow its use for further transgene expression studies or bioproduction purposes in triticale or other plants.

6.4.4 Conclusions

The three sequences identified show potential as candidates for promoters to be used for transgene expression different tissues (i.e., embryo, aleurone and endosperm) of the seed. Several regulatory elements were identified within these sequences and their role was partially evaluated using transient expression studies with promoter deletion fragments. While variability in expression patterns was observed with the deletion of many of these elements in two of the promoter sequences (T-WGA and T-ω-SEC), the precise role of some of these elements will need to be further evaluated through further

promoter deletions/mutations, and quantitative transcript and protein expression analysis using reporter gene expression studies. In addition, further transient expression studies in an increased number of tissues should be performed using the T-triticain-γ promoter deletion fragments to attempt to identify a region of the promoter capable of seed-specific expression. Many of the identified elements in the three promoter sequences are seed-specific elements and are ideal primary targets for further investigation of promoter activity. Walking further upstream of the identified sequences may also locate additional elements that may be involved in their regulation.

These promoter sequences identified are free of infringing patents which add to their potential for use beyond initial promoter studies but for transgene expression for bio-production. The T-WGA promoter shows the most promise as the sequence identified is novel and contains several seed-specific *cis*-acting regulatory elements. Through transient expression studies using promoter deletion fragments we have identified promoter fragments capable of seed-specific expression, as well as several other patterns of tissue-specific expression within the seed. This information provides a new set of tools for use in transgene expression studies that will allow researchers to choose promoter fragments based on their desired expression patterns. The capability to demonstrate different patterns of tissue-specific expression with the promoter deletion fragments also has implications for expression of valuable products in harvestable fractions within the seed for plant bio-production processes.

6.5 LITERATURE CITED

Albani, D., Hammond-Kosack, M.C.U., Smith, C., Conlan, S., Colot, V., Holdsworth, M., and Bevan, M.W. (1997). The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. The Plant Cell **9,** 171-184.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology **215,** 403-410.

Bauer, J. (2003). Methods for increasing the oil content in plants, E.P. Office, ed. European: BASF Plant Science GMBH. Patent no. 03077643.

Blackwell, T.K., Bowerman, B., Priess, J.R., and Weintraub, H. (1994). Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. Science **266,** 621-628.

Bloch, K.D., and Grossman, B. (2003). Digestion of DNA with restriction endonucleases. In: Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. John Wiley & Son, Inc. USA.

Carlini, L.E., Ketudat, M., Parsons, R.L., Prabhakar, S., Schmidt, R.J., and Guiltinan, M.J. (1999). The maize EmBP-1 orthologue differentially regulates Opaque2-dependent gene expression in yeast and cultured maize endosperm cells. Plant Molecular Biology **41,** 339-349.

Chawla, H.S., Cass, L.A., and Simmonds, J.A. (1999). Expression of anthocyanin pigmentation in wheat tissues transformed with anthocyanin regulatory genes. . Current Science **76,** 1365-1370.

Clarke, B., Mukai, Y., and Appels, R. (1996). The *Sec*-1 locus on the short arm of chromosome 1R of rye (*Secale cereale*). Chromosoma **105,** 269-275.

Clarke, B.C., and Appels, R. (1998). A transient assay for evaluating promoters in wheat endosperm tissue. Genome **41,** 865-872.

Clarke, B.C., and Appels, R. (1999). Sequence variation at the *Sec*-1 locus of rye, *Secale cereale* (*Poaceae*). Plant Systematics and Evolution **214,** 1-14.

Daraselia, N.D., Tarchevskaya, S., and Narita, J.O. (1996). The promoter for tomato 3-hydroxy-3-methylglutaryl coenzyme A reductase gene 2 has unusual regulatory elements that direct high-level expression. Plant Physiology **112,** 727-733.

de Pater, S., Pham, K., Chua, N.H., Memelink, J., and Kijne, J. (1993). A 22-bp fragment of the pea lectin promoter containing essential TGAC-like motifs confers seed-specific gene expression. The Plant Cell **5,** 877-886.

Dickey, L.F., Petracek, M.E., Nguyen, T.T., Hansen, E.R., and Thompson, W.F. (1998). Light regulation of Fed-1 mRNA requires an element in the 5' untranslated region and correlates with differential polyribosome association. The Plant Cell **10,** 475-484.

Druka, A., Muehlbauer, G., Druka, I., Caldo, R., Baumann, U., Rostoks, N., Schreiber, A., Wise, R., Close, T., Kleinhofs, A., Graner, A., Schulman, A., Langridge, P., Sato, K., Hayes, P., McNicol, J., Marshall, D., and Waugh, R. (2006). An atlas of gene expression from seed to seed through barley development. Functional & Integrative Genomics **6,** 202-211.

Dunn-Coleman, N., Langdon, T., and Morris, P. (2006). Manipulation of the phenolic acid content and digestibility of plant cell walls by targeted expression of genes encoding cell wall degrading enzymes, U.S.P. Office, ed (United States: Genencor International, Inc.). US Patent no. 7132589.

Esau, K. (1977). Anatomy of Seed Plants. (New York: Wiley).

Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K., and Rask, L. (1999). Interaction between composite elements in the *napA* promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. Plant Molecular Biology **40,** 699-709.

Goddijn, O.J.M., and Pen, J. (1995). Plants as bioreactors. Trends in Biotechnology **13,** 379-387.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C., and Gelbart, W.M. (1996). An introduction to genetic analysis. (New York: W.H.Freeman and Company).

Guo, H., and Moose, S.P. (2003). Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. The Plant Cell **15,** 1143-1158.

Holwerda, B.C., and Rogers, J.C. (1992). Purification and characterization of aleurain: a plant thiol protease functionally homologous to mammalian cathepsin H. Plant Physiology **99,** 848-855.

Kiyosaki, T., Asakura, T., Matsumoto, I., Tamura, T., Terauchi, K., Funaki, J., Kuroda, M., Misaka, T., and Abe, K. (2009). Wheat cysteine proteases triticain α, β and γ exhibit mutually distinct responses to gibberellin in germinating seeds. Journal of Plant Physiology **166,** 101-106.

Kock, M., and Bauer, J. (2003). Constructs and methods for the regulation of gene expression, E.P. Office, ed (BASF Plant Science GMBH). Patent no. 03078629.

Kovalic, D.K., Andersen, S.E., Byrum, J.R., Conner, T.W., Masucci, J.D., and Zhou, Y. (2007). Nucleic acid molecules and other molecules associated with plants and uses thereof for plant improvement, U.S.Patent no. 7214786.

Lam, E., Kano-Murakami, Y., Gilmartin, P., Niner, B., and Chua, N.H. (1990). A metal-dependent DNA-binding protein interacts with a constitutive element of a light-responsive promoter. The Plant Cell **2,** 857-866.

Lerner, D.R., and Raikhel, N.V. (1989). Cloning and characterization of root-specific barley lectin. Plant Physiology **91,** 124-129.

Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P., and Rombauts, S. (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Research **30,** 325-327.

Li, C.-Y., Li, W.-H., Li, C., Gaudet, D.A., Laroche, A., Cao, L.-P., and Lu, Z.-X. (2010). Starch synthesis and programmed cell death during endosperm development in triticale (×*Triticosecale* Wittmack). Journal of Integrative Plant Biology **52,** 602-615.

Luan, S., and Bogorad, L. (1992). A rice cab gene promoter contains separate *cis*-acting elements that regulate expression in dicot and monocot plants. The Plant Cell **4,** 971-981.

Ma, L.G., Chen, C., Liu, X.G., Jiao, Y.L., Su, N., Li, L., Wang, X.F., Cao, M.L., Sun, N., Zhang, X.Q., Bao, J.Y., Li, J., Pedersen, S., Bolund, L., Zhao, H.Y., Yuan, L.P., Wong, G.K.S., Wang, J., Deng, X.W., and Wang, J. (2005). A microarray analysis of the rice transcriptome and its comparison to Arabidopsis. Genome Research **15,** 1274-1283.

Manjunath, S., and Sachs, M.M. (1997). Molecular characterization and promoter analysis of the maize cytosolic glyceraldehyde 3-phosphate dehydrogenase gene family and its expression during anoxia. Plant Molecular Biology **33,** 97-112.

Mark, C.J., Daryl, J.S., and Travis, W.B. (2007). Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci. Plant Biotechnology Journal **5,** 442-453.

Mishkind, M.L., Palevitz, B.A., Raikhel, N.V., and Keegstra, K. (1983). Localization of wheat germ agglutinin-like lectins in various species of the *Gramineae*. Science **220,** 1290-1292.

Muller, M., and Knudsen, S. (1993). The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. The Plant Journal **4,** 343-355.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with Tobacco tissue cultures. Physiologia Plantarum **15,** 473-497.

Peumans, W.J., Stinissen, H.M., and Carlier, A.R. (1982). Isolation and partial characterization of wheat-germ-agglutinin-like lectins from rye (*Secale cereale*) and barley (*Hordeum vulgare*) embryos. Biochemistry Journal **203,** 239-243.

Prieto-Dapena, P., Almoguera, C., Rojas, A., and Jordano, J. (1999). Seed-specific expression patterns and regulation by ABI3 of an unusual late embryogenesis-abundant gene in sunflower. Plant Molecular Biology **39,** 615-627.

Raikhel, N.V., and Wilkins, T.A. (1987). Isolation and characterization of a cDNA clone encoding wheat germ agglutinin. Proceedings of the National Academy of Sciences of the Unites States of America **84,** 6745-6749.

Raikhel, N.V., and Pratt, L.H. (1987). Wheat germ agglutinin accumulation in coleoptiles of different genotypes of wheat. Localization by monoclonal antibodies. Plant Cell Reports **6,** 146-149.

Raikhel, N.V., Bednarek, S.Y., and Wilkins, T.A. (1988). Cell-type-specific expression of a wheat-germ agglutinin gene in embryos and young seedlings of *Triticum aestivum*. Planta **176,** 406-414.

Rogers, S.W., Burks, M., and Rogers, J.C. (1997). Monoclonal antibodies to barley aleurain and homologs from other plants. The Plant Journal **11,** 1359-1368.

Rouster, J., Leah, R., Mundy, J., and Cameron-Mills, V. (1997). Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. The Plant Journal **11,** 513-523.

Schreiber, A., Sutton, T., Caldo, R., Kalashyan, E., Lovell, B., Mayo, G., Muehlbauer, G., Druka, A., Waugh, R., Wise, R., Langridge, P., and Baumann, U. (2009). Comparative transcriptomics in the *Triticeae*. BMC Genomics **10,** 285.

Sharp, P.J., Kreis, M., Shewry, P.R., and Gale, M.D. (1988). Location of *β*-amylase sequences in wheat and its relatives. Theoretical and Applied Genetics **75,** 286-290.

Smith, J.D., Kidwell, K.K., Evans, M.A., Cook, R.J., and Smiley, R.W. (2003). Evaluation of spring cereal grains and wild *Triticum* germplasm for resistance to *Rhizoctonia solani* AG-8. Crop Science **43,** 701-709.

Smith, J.J., and Raikhel, N.V. (1989). Nucleotide sequences of cDNA clones encoding wheat germ agglutinin isolectins A and D. Plant Molecular Biology **13,** 601-603.

Takaiwa, F., Oono, K., Wing, D., and Kato, A. (1991). Sequence of three members and expression of a new major subfamily of glutelin genes from rice. Plant Molecular Biology **17,** 875-885.

Takaiwa, F., Yamanouchi, U., Yoshihara, T., Washida, H., Tanabe, F., Kato, A., and Yamada, K. (1996). Characterization of common *cis*-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. Plant Molecular Biology **30,** 1207-1221.

Van Damme, E.J.M., Peumans, W.J., Barre, A.J., and Rouge, P.J. (1998). Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Crit. Rev. Plant Science. **17,** 575-692.

Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L., and Schmidt, R.J. (1997). A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. Proceedings of the National Academy of Sciences of the Unites States of America **94,** 7685-7690.

Washida, H., Wu, C.-Y., Suzuki, A., Yamanouchi, U., Akihama, T., Harada, K., and Takaiwa, F. (1999). Identification of *cis*-regulatory elements required for endosperm expression of the rice storage protein glutelin gene *GluB-1*. Plant Molecular Biology **40,** 1-12.

Whittier, R.F., Dean, D.A., and Rogers, J.C. (1987). Nucleotide sequence analysis of alpha-arnylase and thiol protease genes that are hormonally regulated in barley aleurone cells. Nucleic Acids Research **15,** 2515-2535.

Williams, M.E., Foster, R., and Chua, N.H. (1992). Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. The Plant Cell **4,** 485-496.

Xu, Y., Badea, C., Tran, F., Frick, M., Schneiderman, D., Robert, L., Harris, L., Thomas, D., Tinker, N., Gaudet, D., and Laroche, A. (2011). Next-Gen sequencing of the transcriptome of triticale. Plant Genetic Resources: Characterization and Utilization **9,** 181-184.

Xu, Y., Penniket, C., Frick, M., Robert, L.S., Harris, L.J., Tran, F., Schneiderman, D., Gulick, P.J., and Laroche, A. (2012). Construction of a rye (*Secale cereale* L.) reference transcriptome. Unpublished.

CHAPTER 7 SUMMARY OF CONCLUSIONS

The presented research has provided a substantial pool of knowledge regarding triticale development, gene function, gene expression and regulation not previously available to triticale researchers. This research has also provided some resources to enable further research in triticale and related cereal crops and has provided tools for transgene expression which will enable development of triticale as a bioproduction platform crop.

Measuring the chlorophyll content in triticale leaves under controlled growth conditions provided a temporal profile of leaf senescence, the first for triticale to my knowledge. Mean convergence points from the linear-linear segmented regressions were used to determine that the relative chlorophyll content begins a rapid decline at 27.4 DPA in flag leaves and 21.5 DPA in third leaves. This temporal profile enabled the selection of sampling times before and after this mean convergence point for tissue sample collection in gene expression analysis experiments. The gene expression analysis experiments confirmed that the 22 and 30 DPA sampling times chosen based on the temporal profile were able to capture early and late changes in expression of transcripts involved in leaf senescence.

Homology searches against various public databases using multiple sequences representing the probe sets on the Affymetrix GeneChip® Wheat Genome array was a successful technique to identify homologous sequences with annotation information. Using assembled cDNA sequences from high throughput sequencing proved to be the most successful in identification of homologous sequences with GO annotation information. By transferring the annotation information that met my stringency criteria

to the probe sets on the wheat GeneChip, I was able to provide GO annotation information for a large number of probe sets that previously did not have any GO annotations. Combining the results from the different homology search techniques allowed me to improve the number of probe sets on the array that had Gene Ontology annotation information from less than 3 % to over 76 %. This vast improvement in annotation of the wheat GeneChip is the highest improvement using an automated method to date to my knowledge. This gene ontology annotations information allowed functional classification of genes identified using the array. By classifying gene lists identified using the wheat GeneChip using this functional information, biological meaning can be given to the results. I was able to classify the probe sets on the wheat GeneChip according to biological process, molecular function and cellular component. Together with expression results, researchers can use this information to make meaningful biological conclusions about their experiments using the wheat GeneChip.

Gene expression profiling of triticale using the Affymetrix GeneChip® Wheat Genome Array resulted in a gene expression atlas for triticale. While my experiment focused on vegetative and seed tissues, by combining my results with the triticale reproductive tissue expression results from our colleagues, the atlas includes all of the major tissue types and several developmental stages. This expression atlas is the first and most comprehensive of its kind for triticale to date. I have identified transcripts that are expressed in triticale, that are not expressed in triticale, that are specifically expressed in each tissue, that have enriched expression in a specific tissue and transcripts expressed in leaf and stem tissues that are developmentally regulated. Together with the functional information provided for the wheat GeneChip by the annotation improvement, I was able

to classify each of these gene lists according to their gene ontologies. Using this functional information I was able to determine which biological processes, molecular functions and cellular components are over-represented in each of the gene lists. Each gene list showed enrichment of particular functions which allowed us to make biologically relevant conclusions from the expression data. By applying the gene ontology annotation information, I was able to identify transcripts with a particular function or involved in a particular biological process that had a particular expression pattern. By identifying transcripts with a transcription factor activity related functional annotation within the seed-specific list of transcripts, I was able to demonstrate the application and relevance of this information. I was able to show that genes with a particular function that have a particular expression pattern could be identified using my results. The expression results, combined with the functional annotation information, allow the selection of candidate genes for further examination. Several seed tissue-specific transcripts were selected for verification of their expression patterns using qRT-PCR. The qRT-PCR results verified the microarray results and provided confidence that our expression analysis is a valuable resource to the plant research community for selection of targets for further investigation.

A novel method for *in situ* hybridization of RNA transcripts in triticale seeds was developed and used for further evaluation of the expression patterns of candidate seed-specific transcripts identified by the gene expression analysis. This method used short DIG labeled ssRNA transcript specific probes produced by *in vitro* transcription of a transcript-specific PCR product with a linked T7 RNA polymerase site. Transcript expression patterns of candidate seed-specific transcripts were evaluated at the cellular

level and through several triticale seed developmental stages using these probes. Using this method I was able to localize the expression of transcripts to particular cell types within a given tissue. The expression of an embryo-specific wheat germ agglutinin transcript was localized to the root epidermis and root cap cells within the embryo. An embryo-specific glutamine synthetase was localised to the procambial cells within the embryo and scutellum. The expression of a crease-specific protein kinase was restricted to the vascular cells in the crease at later stages of seed development. A putative crease-specific ubiquitin conjugating enzyme was found to be expressed in regions of the endosperm near the time of tissue sampling for the microarray analysis. The expression patterns observed allowed conclusions to be drawn about the transcripts potential roles in triticale seed development.

Promoter sequences for three putative tissue-specific triticale gene homologues were sequenced from genomic DNA. Novel promoter sequences were identified for a triticale homologue of wheat germ agglutinin (T-WGA), a triticale homologue of rye ω-secalin (T-ω-SEC) and a triticale homologue of wheat triticain-γ (T-triticain-γ). The novel promoter sequences were partially characterized to identify *cis*-acting regulatory elements that may be responsible for their expression patterns in triticale. Many *cis*-acting sequence elements in the promoters were identified electronically using PlantCARE and the locations of putative seed-specific elements were used to design promoter deletion fragments used in transient expression studies. The transient expression studies revealed that the promoters had variability in their ability to confer expression in the different tissues examined. Only the promoter deletion fragments of T-triticain-γ were able to direct transient expression of a transgene in leaf tissue, while all

promoter deletion fragments of T-WGA and T-ω-SEC were seed specific. Two T-WGA promoter deletion fragments (-598 and -2485) were capable of embryo-specific transgene expression while all other T-WGA promoter deletion fragments directed expression in the embryo and at least one other seed tissue. While none of the T-ω-SEC promoter deletion fragments tested were capable of endosperm-specific expression, one fragment was capable of endosperm and seed coat-specific expression (-287), one was capable of embryo and endosperm-specific expression (-475) and two were capable of seed tissue-specific expression (-208 and -1536). The promoters identified are novel triticale sequences not encompassed by existing patents and are tools that I have identified as capable of directing transgene expression in triticale in a number of desired expression patterns.

The gene ontology data set and the gene expression atlas provided by this research are valuable resources to the triticale research community. In addition to the triticale community, the cereal research community and even the greater plant research community will benefit from these resources. I have developed several techniques for triticale research including: GO annotation improvement using multiple homology searches, determining a temporal profile of senescence using relative changes in chlorophyll content, *in situ* hybridization using transcript-specific probes and transient expression in triticale seed and leaf tissues. These methods are techniques that could be applied to future studies in triticale but could also be applied to research in other cereals and plants. In addition to these techniques, I have identified a number of tools, in the form of promoter deletion fragments, for directing transgene expression in triticale. These tools may even prove to be effective at directing transgene expression in other

cereals or even non-cereal plants. Together the resources, techniques and tools provided by this research provide an enormous improvement to the knowledge base regarding triticale development, gene function, gene expression and regulation. To succeed in making triticale an effective bioproduction platform crop, this knowledge is of utmost importance. This knowledge will be sure to have applications in future research in triticale, cereals and other plants.