Effects of Drought on Gene Expression in Maize Reproductive and Leaf Meristem Tissues as Revealed by Deep Sequencing

Akshay Kakumanu

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Plant Pathology, Physiology, and Weed Science

Ruth Grene, Chair

Glenda Gillaspy

Lenwood S. Heath

T. M. Murali

June 28, 2012

Blacksburg, Virginia

Keywords: Drought, Illumina, RNA-Seq, Maize, Ovaries, Leaf Meristem

Effects of Drought on Gene Expression in Maize Reproductive and Leaf Meristem Tissue Revealed by Deep Sequencing

Akshay Kakumanu

(ABSTRACT)

Drought is a major environmental stress factor that poses a serious threat to food security. The effects of drought on early reproductive tissue at 1-2 DAP (days after pollination) is irreversible in nature and leads to embryo abortion, directly affecting the grain yield production. We developed a working RNA-Seq pipeline to study maize (*Zea mays*) drought transcriptome sequenced by Illumina GSIIx technology to compare drought treated and well-watered fertilized ovary (1-2DAP) and basal leaf meristem tissue. The pipeline also identified novel splice junctions - splice variants of previously known gene models and potential novel transcription units. An attempt was also made to exploit the data to understand the drought mediated transcriptional events (e.g. alternative splicing). Gene Ontology (GO) enrichment analysis revealed massive down-regulation of cell division and cell cycle genes in the drought stressed ovary only. Among GO categories related to carbohydrate metabolism, changes in starch and sucrose metabolism-related genes occurred in the ovary, consistent with a decrease in starch levels, and in sucrose transporter function, with no comparable changes occurring in the leaf meristem. ABA-related processes responded positively, but only in the ovaries. GO enrichment analysis also suggested differential responses to drought between the two tissues in categories such as oxidative stress-related and cell cycle events. The data

are discussed in the context of the susceptibility of maize kernel to drought stress leading to embryo abortion, and the relative robustness of actively dividing vegetative tissue taken at the same time from the same plant subjected to the same conditions. A hypothesis is formulated, proposing drought-mediated intersecting effects on the expression of invertase genes, glucose signaling (hexokinase 1-dependent and independent), ABA-dependent and independent signaling, antioxidant responses, PCD, phospholipase C effects, and cell cycle related processes.

This work was supported by the National Science Foundation Plant Genome Research Program (grant no. DBI0922747), iPlant Collaborative (NSF DBI-0735191) and also NSF ABI1062472.

Dedicated to my grandmother, Prabhavathy Kamisetty

Acknowledgments

I thank my advisor, Prof. Ruth Grene, for believing in me, and providing me with constant support and guidance. I also thank my committee members, Prof. Lenwood S. Heath, Prof. T. M. Murali, and Prof. Glenda Gillaspy for agreeing to be on my committee and providing their invaluable inputs. I like to thank Prof. Andy Pereira and Prof. Ruth Grene for providing me with an opportunity to work on a very interesting project right from my first day of masters at Virginia Tech. Thanks also to Judy Fielder and Donna Frod for helping me with all the departmental matters.

My research being interdisciplinary, I was fortunate to have received help from people with different expertise. Dr. Ruth Grene helped me in making sense of the vast amounts of literature on maize drought and introducing me to the field of plant biology. I also owe her a special thanks for all the help in biological interpretation of the data. Dr. Madan Ambavaram, post-doc in Dr. Pereira′s lab, conducted all the wet-lab experiments, which formed the basis of the whole project. I am very thankful to him, for patiently explaining the experimental setup and for his exquisite work with the initial wetlab experiments. Arjun

Krishnan, senior grad student in Pereira′s lab taught me a great deal by example. My long and frequent conversations with him helped me in the bioinformatics part of my project. Curtis Klumas, lab-mate in Ruth′s lab, helped a great deal in the later parts of the project. He was patient and committed in helping me with critically analyzing the GO enrichment results. I would like to specially thank him for Figures 4.2 and 6.1. I would also like to thank Elijah Myers for his support with MapMan and SBGN (Systems Biology Graphical Notation).

Finally, I would like to thank my family for their support during tough times.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Drought stress on maize (*Zea mays*) reproductive tissue at a very early stage (1-2DAP) leads to embryo abortion, which is irreversible in nature. Such drastic effects are not seen in the vegetative tissue and reproductive tissue at a much later stage (8-10DAP). To gain insights into possible causes of differential sensitivity, we have studied the effects of drought on the transciptome obtained from well-watered and drought stressed basal leaf meristem and pollinated ovaries of a fully sequenced maize (B73) genotype.

Ch.2 of this document updates our current understanding of drought stress on maize tissues. This chapter also describes our motivation in selecting particular vegetative and reproductive tissues for transcriptome sequencing and comparison. A brief introduction of the Illumina GSIIx sequencing technology is given in Ch.3. This chapter also gives an outline of the various tools to perform different steps of RNA-Seq analysis and also gives our rationale for

selecting the tools and procedures which constitute our pipeline.

The Gene Ontology (GO) has two components: the ontologies and the annotations that map gene products to these defined ontologies. Ch.4 of this document describes in detail a few selected biological process GO terms that were enriched with either up or down regulated genes in both the tissues. In the latest maize GO annotations, only 8300 genes are annotated to GO-biological process terms (as of May 2011). Hence, in order to provide an additional level of analysis, the MapMan software tool was selected for its capacity to view statistically significant genes in the context of known metabolic pathways, biological processes, and functional categories. The results obtained from MapMan are explained in detail in Ch.5.

Taking together the interpretations obtained from different steps of our analysis, a working hypothesis has been formulated. Interpretation and conclusions of the working hypothesis are described in Ch.6.

Chapter 2

Maize Drought Biology

2.1 Literature Review

According to the Food and Agriculture Organization (FAO), maize (*Zea mays*) is the most widely grown cereal crop in the United States (`http://www.fao.org/corp/statistics/en/`). Since maize has a wide variety of uses ranging from animal stock to fuels, yield loss results in many undesirable consequences. The yield loss in maize is directly affected by embryo abortion, which takes place when plants experience abiotic stress such as drought during their reproductive phase [1, 2, 3]. Similar effects have also been reported for wheat (*Triticum aestivum*), barley (*Hordeum vulgare*, soybean (*Glycine max*), and chickpea (*Cicer arietinum*), suggesting a widespread phenomenon in the plant world [3]. Embryo abortion is observed even when drought stress is relieved before fertilization has occurred [4], suggesting

that the vulnerability is caused by the influence of drought on the maternal tissue, which develops extensively both before and immediately after fertilization [1]. Ovaries in tissue subjected to drought stress or with a previous history of drought stress stop growth within 1-2 days after pollination [5], suggesting a direct correlation between tolerance to drought stress in female floral parts and yield in maize [3]. In cases where the drought stress was applied before fertilization, increased levels of reactive oxygen species (ROS) are expected to exist in the maternal tissues and participate in other related molecular events, since this is a universal response of plant cells to an abiotic stress such as drought, as reviewed previously [6, 7]. An altered, and potentially more oxidized, environment is thus created, within which the embryo and endosperm begin their development.

The main source of sugar substrates to the developing embryo are known to be supplied by a hexose supply pathway involving soluble and cell wall invertases, and the availability of sugar substrates within the developing embryo is directly correlated with embryo abortion [2]. Expression of *lvr*, a soluble invertase, was shown to be the most vulnerable to drought stress at 3DAP, with cell wall associated invertase also showing a decrease at that time [1]. The levels of invertase substrate sucrose are higher in drought stressed than in well-watered maize ovaries from -6 to +7DAP [1]. A set range of sucrose to hexose ratios has been shown to be crucial for seed development [8], and this ratio appears to be altered during drought stress at this early stage. Sugars function both as a nutrient source and as a signaling molecule in plant cells [9, 10]. Aample evidence exists suggesting distinct cellular glucose and sucrose signaling pathways [11, 12]. Along with the important metabolic role carried out by cell wall

invertases, sugars are also likely part of a signaling pathway, since they are known to form complexes in the nucleus with PIP5K9, a component of the phosphoinositol signaling system [13]. Hence, it is very likely that many pathways regulated by sugar signaling are influenced under drought stress in the reproductive tissue at this early stage.

Another manifestation of drought stress in the reproductive tissue is the increase and accumulation of abscisic acid (ABA) levels. Drought mediated increase in ABA levels has been shown to take place at around 9DAP and has the potential to affect many down stream process [14, 15]. An influence of ABA on apoplastic sugar transport (involving cell wall invertase and hexose transporters) and pollen sterility was observed in cold stressed rice [16]. Similar effects of increase in ABA on the pollen sterility have also been shown in wheat [17]. Increase in ABA levels has also been correlated with a decrease in kernel set [3]. However, no comparable effects of ABA on developing maternal tissue either before, or immediately after, pollination in maize have yet been reported.

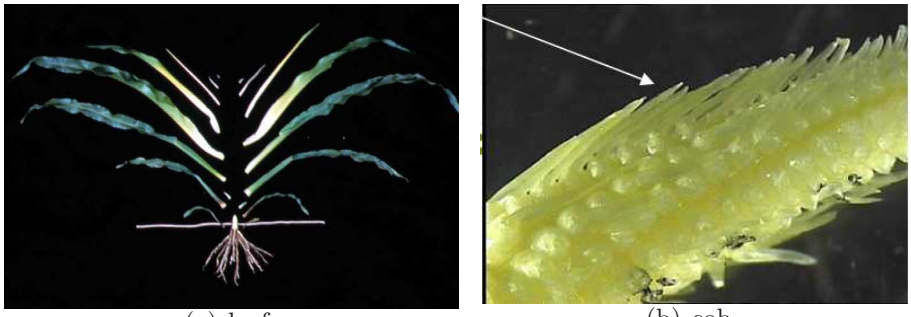
Effects of drought on endosperm cell division, sugar metabolism/signaling, auxin and ethylene-related processes, the behavior of MADS box genes, and polyamine and trehalose metabolism have all been proposed as part of the chain of events surrounding embryo abortion in maize and rice [18, 19, 20, 2, 3]. Most of these studies, such as those cited above, were conducted on fertilized ovaries from 8 days after pollination onward. The only exceptions are the lower levels of starch and the well established early responses of cell wall and soluble invertases to drought stress at 1DAP, at both the transcript and enzyme activity level, immediately after pollination [1, 19].

To date, few genome-wide approaches to this phenomenon have been reported. A small cDNA microarray containing about 2,500 cDNAs from maize was used to monitor gene expression specifically in developing maize endosperm and placenta/pedicel tissues during water deficit and re-watering at 9 days after pollination [18], a time when endosperm and placental tissues could be clearly distinguished. A transcriptomic study of the effect of drought on maize ears and tassels undergoing meiosis revealed the up-regulation of raffinose and trehalose-associated genes, as well as other carbohydrate-related processes [21]. Beyond these observations, little is understood about the chain of molecular events leading to early embryo abortion in drought stressed maize, and virtually nothing is understood about crucial events occurring upon fertilization. Gene expression studies in maize in response to water stress have been investigated in roots [22], seedlings [23], and developing ear and tassel [21]. Genome wide comparisons between the drought stress responses across different developmental stages in maize have not yet been made.

2.2 Tissue Sampling for Transcriptome analysis

Most of the drought experiments on maize as mentioned in the previous section were conducted on fertilized maize ovaries (8DAP), yet irreversible drought mediated embryo abortion is observed immediately after pollination. Effects of drought on the maize transcriptome at this early, and crucial, stage of ovary development have not, as yet, been characterized, and whether or not the various pathways invoked in the studies enumerated above contribute

to the observed cessation of embryo growth in stressed plants as early as 1-2DAP is as yet unknown. To advance our understanding on the effects of drought at this early crucial stage, we have chosen to study drought treated ovaries one day after pollination (Figure 2.1(b)). We wanted to compare the heightened sensitivity of the recently fertilized ovaries to drought stress with an actively dividing non-reproductive tissue in the same plant under the same drought stress condition to gain insights into the possible differential sensitivity. In mono-cotyledon leaves, such as maize, cell division is restricted to a very limited zone near the leaf insertion points [24] (Figure 2.1(a)). Hence, we have chosen basal leaf meristem from the three youngest leaves of maize plant at the 12-leaf stage. Table 2.1 shows comparable gene expression between leaf meristem and ovaries in GO-terms related to active growth, suggesting active development and growth in both the tissues.



(a) leaf (b) cob

Figure 2.1: Images showing snapshots of the tissue used for transcriptome sampling in maize leaf and ovary tissues

Maize plants of the inbred line B73 were grown in 10-L pots with a 1 : 1 : 1 mix of peat : vermiculite : perlite with 6g pulverized limestone, 35g of CaSO4, 42 of powdered FeSO4, and 1g of trace fritted element [25]. The plants were hand irrigated daily to maintain soil

Table 2.1: No of genes with Fragment Per Kilobase of exon per Million fragments mapped (FPKM), an expression measure, greater than 5 in select GO terms related to growth and development. The distribution of FPKM expression values are shown in Figure 3.2

GO Term	Name	Ovary (FPKM $\geq$ 5)	Leaf (FPKM $\geq$ 5)
GO:0007049	Cell Cycle	40/53	21/53
GO:0051301	Cell Division	35/49	21/49
GO:0016568	Chromatin Modification	26/30	23/30
GO:0006260	DNA Replication	50/72	30/72

water content close to field capacity and the nutrients were supplied on a weekly basis with a general purpose fertilizer (15-16-17 Scott-Sierra Horticultural Product Co, Marysville, OH). Plants were grown under well-watered conditions until they reached the reproductive stage (at the onset of silk emergence), when irrigation was withheld for half of the plants. Two to three days after irrigation was withheld, the plants were hand pollinated, and 24 hours after pollination measurements and samples were collected for transcriptome analysis. At this stage, drought stressed plants had undergone three or four days of drought stress, while controls were well watered throughout this period. The degree of drought stress was determined by monitoring soil moisture content, relative water content (RWC), chlorophyll fluorescence

Table 2.2: Analysis of the physiological parameters in maize affected by mild and severe drought at the reproductive stage.

	Soil Moisture Content (%)(cm3/cm3)	Chlorophyll Fluorescence (Fv/Fm,PSII)	CO2 Exchange (mol m-2s-1)	Relative Water Content (%)	SeedNumber
Well-Watered	49.2	0.779, 0.71	25.9	97.15	139.25
Mild Drought	21.5	0.741, 0.53	20.25	83.42	104.75
Severe Drought	10.7	0.721, 0.42	14	66.5	34.75
LSD(0.05)	5.1707	0.017, 0.012	4.1215	4.025	12.23
P value	< 0.0001	<0.0001	< 0.0001	< 0.0001	< 0.0001

(Fv/Fm and PS II) and CO2 gas-exchange in the leaves (Table 2.2). Soil moisture content decreased from 49.2 cm3cm-3 in well-watered controls to 21.5 for moderate drought, and to 10.7 for severe drought. RWC dropped from 97.15% (control) to 83.4 (moderate drought) and to 66.5 (severe drought). These levels of drought reduced photosynthetic rates and PS II function dramatically. Seed set was significantly affected by severe drought at 1DAP (Table 2.2).

Before proceeding to the library preparation for transcriptome sequencing, the expression levels of invertases were compared with previously observed patterns in drought treated developing ovaries (1DAP). Figure 2.2 shows the fold changes in the expression values of the invertases, which matched with previous observations.

Two biological replicates were used for all RNA-seq experiments from each tissue type. The total RNA from the leaf meristem and ovary tissues was extracted using Trizol reagent

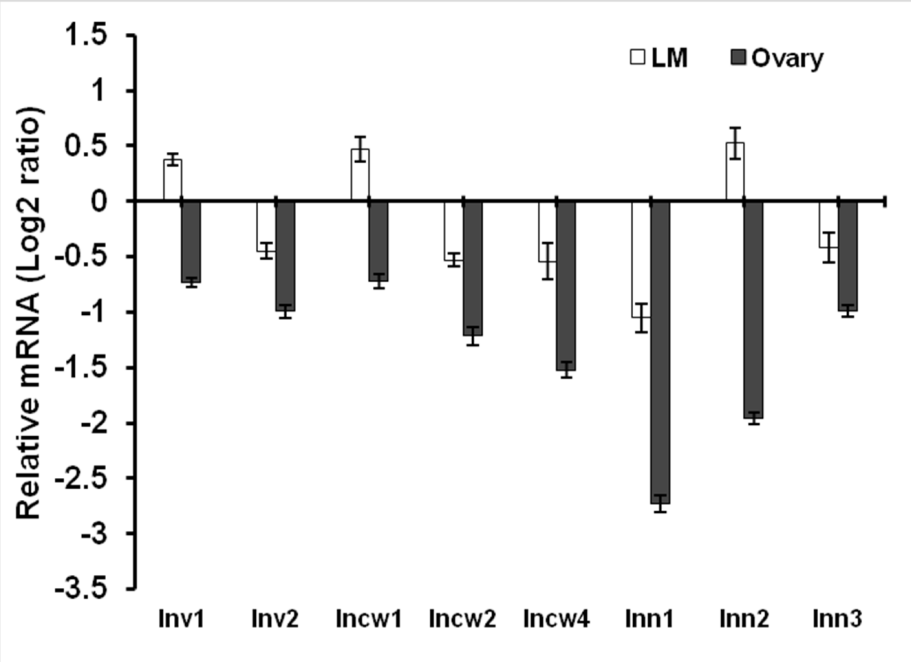


Figure 2.2: Expression analysis of maize vacuolar, cell wall and neutral invertases assessed through qRT-PCR

(Invitrogen) and purified using the RNAeasy Plant Mini kit (Qiagen). On-column DNase digestion was performed according to the manufacturer's protocol (Qiagen). The integrity and quality of the total RNA was checked using NanoDrop 1000 Spectrophotometer and formaldehyde-agarose gel electrophoresis. The Poly(A) RNA was isolated from purified total RNA using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces under elevated temperature, and the cleaved RNA fragments copied into first strand cDNA using reverse transcriptase and random primers. Second strand cDNA synthesis was done using DNA Polymerase I and RNaseH, the cDNA fragments processed for end repair, an addition of a single A base, and ligation of the adapters. These products were then purified and enriched by PCR to create the final cDNA library and

sequenced on the Illumina Genome Analyser IIx system according to the manufacturer's recommendations (Illumina).

Chapter 3

RNA-Seq Pipeline

3.1 Preprocessing Illumina Reads

The RNASeq pipeline shown in Figure 3.5 was implemented in this anlysis. This and the following sections in this chapter explain different steps of the pipeline. As mentioned in the previous chapter the transcriptomes of two tissues, the leaf meristem, and ovaries, with one biological replicate, were sequenced using the Illumina Genome Analyzer IIx. We used a strand specific mRNA-Seq protocol that preserves the information about which strand was originally transcribed. Eight libraries were sequenced in total and the library names are abbreviated as follows :-

MCC-1 :- Maize Ovaries Control Biological Replicate 1

MCC-2 :- Maize Ovaries Control Biological Replicate 2

MCD-1 :- Maize Ovaries Drought Treated Biological Replicate 1

MCD-2 :- Maize Ovaries Drought Treated Biological Replicate 2

MLC-1 :- Maize Leaf Meristem Control Biological Replicate 1

MLC-2 :- Maize Leaf Meristem Control Biological Replicate 2

MLD-1 :- Maize Leaf Meristem Drought Treated Biological Replicate 1

MLD-2 :- Maize Leaf Meristem Drought Treated Biological Replicate 2

The initial length of the reads in each library was 76bp long, contaminated with varied length of adapter sequences. The first preprocessing step performed was to remove the adapter sequences and low quality bases, which resulted in reads with varied lengths ranging from reads whose length was as low as 11bp to reads 76bp long (Figure 3.1). Reads with length less than 25bp were discarded and not considered for further analysis mainly because it is very difficult to accurately map reads with such short lengths. This did not result in much loss of data as there were less than 0.002 % reads with less than 25bp long in each library.

From Figure 3.1 it is clear that in most of the libraries there are fewer than 10% of the reads with length less than 30bp, thus the read lengths are ideal for short read mapping tools like BWA [26], Tophat [27], and Bowtie [28]. The read length distribution is also very similar among different libraries.

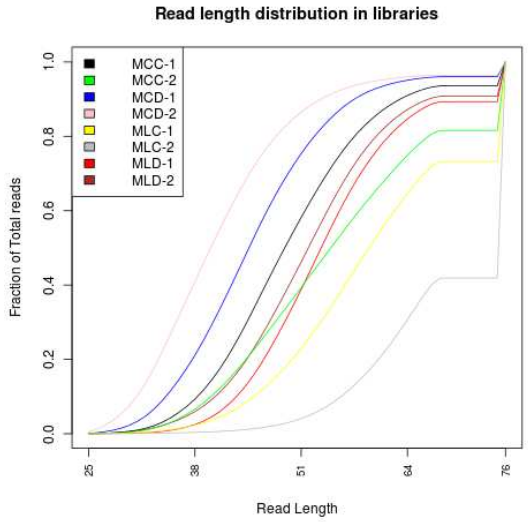


Figure 3.1: Read Length Distribution in Maize Ovary and Leaf Meristem Libraries

3.2 Mapping Illumina Reads

One of the essential and basic tasks of a RNASeq analysis is mapping the reads to a reference genome if there is one. Many tools have been developed in recent years to do this task [29]. Most of these tools can be classified into two main categories: unspliced aligners and spliced aligners. Unspliced aligners implement mapping algorithms that do not accommodate large gaps. Some of the most commonly used unspliced aligners are BWA [26], Stampy [30] and Bowtie [28]. Alternatively, spliced aligners use algorithms that can map reads to the whole genome and across all splice junctions. Unspliced aligners fall under two main categories: 'seed and extend' and 'exon first'. In 'seed and extend' methods, all the reads are first broken into small fragments and then aligned to the genome. Based on contiguity of the alignments of the seeds, the tool identifies potential intron spanning reads. Alternatively,

'exon first' methods adopt a two step process. In the first step, the reads are mapped to the genome using an unspliced aligner. In the second step the unmapped reads are split in smaller segments and aligned independently. The genomic regions coverd by the mapped reads are then searched for potential splice sites. The 'exon first' methods are much faster than the 'seed extend' methods but result in fewer splice junctions. In our RNASeq analysis, we have used Tophat [27], which uses a spliced 'exon first' algorithm. Tophat is built on an unspliced aligner Bowtie.

The maize reference genome, like most of other plant genomes, has many transposable elements. In fact, nearly 85% of the maize genome is composed of transposable elements [31]. Since the reads in a RNASeq experiment come from mRNA, it very unlikely that they will map to the transposable elements. Hence a masked maize genome, with transposable elements masked, was used. The latest release of the maize masked genome downloaded from `http://ftp.maizesequence.org/release-5b/` was used.

Default Tophat parameters, which allow up to 2 mismatches and report up to 40 alignments for reads mapping at multiple positions, were used. Based on the alignments, the reads were categorized into three classes. Uniquely mapped reads are those that map to only one position in the genome, and spliced reads are those that span across a splice junction. Multi-mapping reads are those that map to more than one position in the reference genome. An in-house Perl script (mapping-statistcs-from-tophat-samfile.pl) was used to calculate the mapping statistics shown in the Table 3.1. The details and usage of all the Perl scrip introduced in the rest of this chapter are given at the end of the thesis.

Table 3.1: Read Mapping Statistics

Sample	Sequenced Reads	Unique Mapping Reads	Spliced Reads	Multi Mapping Reads	Total Mapping Reads
MCC-1	38,720,518	29,006,929	3,377,444	3,334,737	32,341,666
MCC-2	32,396,207	21,765,134	829,439	2,202,487	23,967,621
MCD-1	38,617,367	29,275,556	2,854,178	3,885,381	33,160,937
MCD-2	38,362,194	28,779,699	2,093,744	4,245,354	33,025,053
MLC-1	37,577,043	27,108,369	3,396,442	2,585,335	29,693,704
MLC-2	30,596,778	19,382,018	2,870,407	1,696,101	21,078,119
MLD-1	37,264,930	27,552,879	2,793,741	2,970,863	30,523,742
MLD-2	36,699,121	27,216,792	2,800,586	3,113,970	30,330,762

3.3 Transcriptome Reconstruction and Quantification

The next logical step in RNASeq analysis is to reconstruct the transcriptome. Transcriptome reconstruction is a process in which all expressed isoforms of a gene are reconstructed using the reads and/or read alignments. There are two main approaches for transcriptome reconstruction: genome guided and genome independent. Genome independent methods assemble the transcriptome solely based on the reads and not on the read alignments. These methods are used in cases where the genome of the species of interest is not yet sequenced. Alternatively, genome guided methods first map the reads to the genome and then, based on the alignments of the overlapping reads, the transcriptome is reconstructed. In our pipeline, we have adopted a genome guided approach; Cufflinks [32] was used to assemble the transciptome. Cufflinks initially constructs a graph connecting all the bases in the genome that are either contiguously connected or connected by a spliced read. It than constructs the tran-

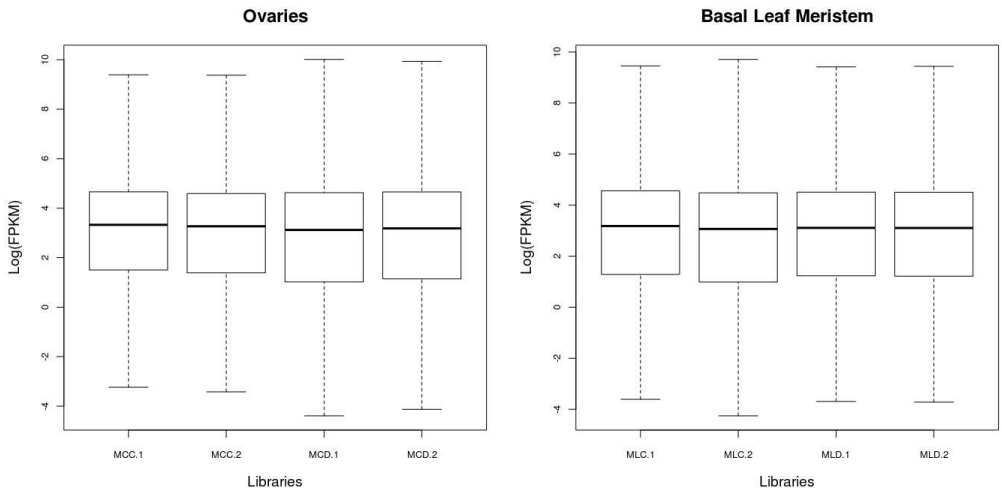


Figure 3.2: Boxplot of Log FPKM expression values to base 2

scriptome by parsing the graph and choosing a minimal set of paths such that all the reads are included in at least one transcript. Cufflinks quantifies the transcriptome abundance by calculating the FPKM (fragments per kilobase of exon per million fragments mapped). The box-plot distributions of the log FPKM values are shown in the Figure 3.2. The 25% quartile, median, and the 75% quartile are almost the same in all the libraries. The distribution of the expression values among the libraries being considered for differential expression must be the same, hence it is essential to check this before proceeding further in the analysis. Since in each tissue there are four libraries, four different assemblies will be generated by Cufflinks. The reconstructed gene models in each of these assemblies are more or less the same, as they come from the same tissue. Hence in order to take the union of these transcripts, we used a tool called Cuffmerge in the Cufflinks software. After constructing the union of the transcripts, Cuffmerge overlays the reference annotations and classifies all the assembled transcripts into different categories. Each category is given a class code. In-house

Table 3.2: Assembly Statistics

Class	% of Total Transcripts in Ovary	% of Total Assembled Reads in Ovary	% of Total Transcripts in Leaf Meristem	% of Total Assembled Reads in Leaf Meristem
Match to Known Transcripts	67.2	73.3	68.3	73.6
Novel Isofroms of Known Genes	17.9	17.3	19.4	16.3
Intergenic	8.3	2.9	5.7	2.53
Artifacts	6.4	7.12	6.5	8.3

Perl scripts (cuffmerge.pl, and cal-adj-length.pl) were written to classify these transcripts into 3 main categories, and to calculate the coverage in each category. Table 3.2 shows the assembly statistics. 'The Match to Known Transcripts' group comprises of all assembled transcripts that matched exactly with a known reference gene model, the reference annotations were downloaded from `http://ftp.maizesequence.org/current/working-set/`. Transcripts that had atleast one new splice junction were grouped into 'Novel Isoforms of Known Genes'. All the transcripts that were assembled in the intergenic regions according to the existing annotations were considered as Intergenic transcripts. According to Table 3.2, majority of assembled transcripts in the leaf and ovary matched to existing gene models. There were far more novel isoforms resulting from alternative splicing than entirely new transcription units. The 'Artifacts' group comprises of transcripts that were assembled by the reads that mapped to the opposite strand, indicating possible miRNA and siRNA contamination.

3.4 Novel Splice Sites and Transcription Units

Novel splice sites are those splice junctions that were supported by at least 2 reads and that have not been characterized earlier. In-house Perl script (novel-splice-junction-stats.pl), was written to extract these novel and existing splice sites from the assembled transcripts. Out of the 0.3 billion reads obtained by RNA-Seq, 21 million reads mapped to splice junctions, revealing 76k novel splice junctions that account for 23% of the total number of splice junctions identified. Of the 76k novel junctions, 36k were supported in both ovary and leaf meristem tissues while the remaining reads were supported in either the ovary or leaf meristem. It is quite clear from Figure 3.3 that there are far fewer novel splice sites than existing ones. Another interesting observation from Figure 3.3 is that the majority of annotated and novel splice sites are supported by 50-80 reads.

Table 3.3: Intergenic Transcripts Blastx Results

Tissue	Total Intergenic Transcripts	Unique Match to rice proteins	Unique Match to sorghum proteins
Ovary	7288	1561	2581
Leaf Meristem	4879	1331	2513

BLASTX [33] was used to align the intergenic transcripts onto rice and sorghum proteins to filter for potential novel transcription units. BLASTX translates a DNA sequence in all six possible reading frames and compares it with protein sequences. All BLASTX hits were processed using the Per script (process-blastx.pl), to identify the best hit for each input

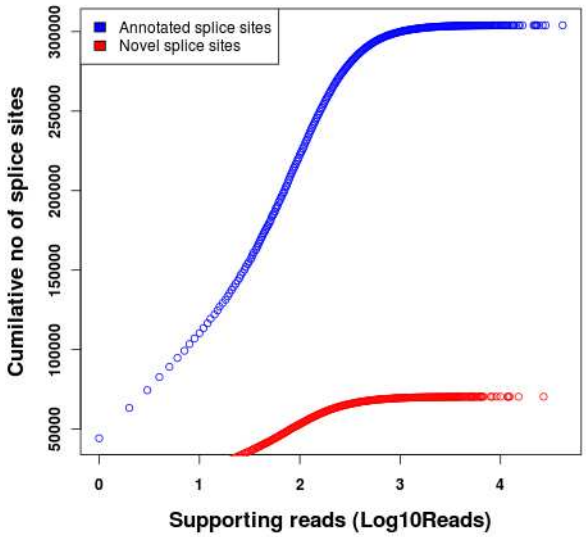


Figure 3.3: Cumulative no of reads supporting novel and annotated splice sites

intergenic transcript (Table 3.3).

3.5 Differential Expression

Cufflinks quantifies the transcript abundance of each transcript in terms of FPKM. The tool also calculates the abundance of each gene by combining the expression of all the transcripts of a gene. Hence, the result of a transcript quantification analysis is a list of genes and their FPKM expression values in each library. To check the extent of variability among the biological replicates, the coefficient of variation (standard deviation/mean) of FPKM values for each gene, between the replicates, was calculated. Nearly 75% of the genes, in ovaries and leaf meristem, had a coefficient of variation of less than 20% (Figure 3.4), indicating

slight variability among the biological replicates. In order to identify differentially expressed genes between two conditions, it is essential to compare the distribution of the expression profiles of genes in the two data sets. The most suited parameter to estimate this is the T-statistic. Intuitively, the T-statistic calculates the difference in the means of the expression distributions also taking into consideration their variance. Since there is only one biological replicate the calculation of the variance is not robust. The Limma R package [34] was used.

Limma make the analysis more robust for a small number of samples by fitting a linear model to the expression data of each gene by borrowing information across genes using an empirical Bayesian method. The package was downloaded from `http://bioconductor.org/packages/release/bioc/html/limma.html`. The lmFit function was used to estimate fold change and standard error by fitting a linear model for each gene. The eBayes function was used to apply empirical Bayes smoothing of the standard errors. Limma outputs a T-Statistic and a p-value for each gene. The QVALUE R package [35], downloaded from `http://genomics.princeton.edu/storeylab/qvalue/`, was used to calculate the q-values from the p-values obtained from Limma. Only those genes that had at least 10 reads mapping in at least one of the libraries were considered for differential expression testing. This lead to the elimination of 800 genes in both the tissues, and as a result, 25,800 genes in the leaf meristem, and 26,300 genes in the reproductive tissue were considered for differential expression testing. Identification of differentially expressed genes in response to drought revealed 4328 genes up-regulated and 6044 genes down-regulated for a q-value cutoff of 0.05 and log_2fold cutoff of 1 in the maize ovary tissue; a much smaller number of genes, 462(up-

regulated) and 220 (Down-regulated), responded in the basal leaf meristem for the same cutoff.

The final step in the pipeline, the GO-Enrichment analysis, is explained in chapter 4 of this thesis.

3.6 Perl Scripts

- **mapping-statistcs-from-tophat-samfile.pl**

  The usage of this script can be obtained by using the "help" option. The input for this script is a sam mapping file. The "accepted-hits.bam" file, which is the output of Tophat, should first be converted into a sam file using Samtools. The output for this script is the list of all reads which span a splice junctions and the mapping statistics that includes: no. of unique alignments, no. of splice alignments, and no. of reads mapping to multiple positions in the reference genome.

- **cuffmerge.pl**

  The input files for this script are: merged.gtf file (an output file from cuffmerge), reference annotations in gft file format, isoforms.fpkm-tracking file (an output file from cufflinks), and a file name to write the output. The file identifies the exon structure of all the novel isoforms assembled by cufflinks. The usage and the format of the output file can be obtained by using the "help" option.

- **cal-adj-length.pl**

In order to calculate the FPKM, cufflinks calculates the adjusted length for each transcript [32]. The Perl script "cal-adj-length.pl", calculates the adjusted length and uses it to calculate the % coverage of each class of transcripts mentioned in section 3.2. The input files for this script are: the output file from cuffmerge.pl, the fpkm expression values (from cufflinks), and a file name to write the output. The usage and the format of the output file can be obtained by using the "help" command.

- **novel-splice-junction-stats.pl**

  This identifies the novel splice junctions in the assembled transcripts. The input files for this script are: the output file from cuffmerge.pl, the reference annotations in gtf format, file names to write the output. The output files contain, the location (base pair position) of all the novel splice sites identified.

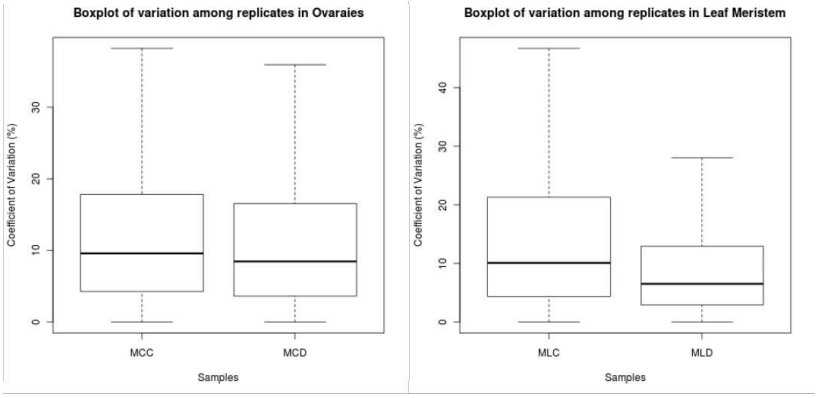


Figure 3.4: Boxplot distribution of coefficient of variation, in percentage, among biological replicates.

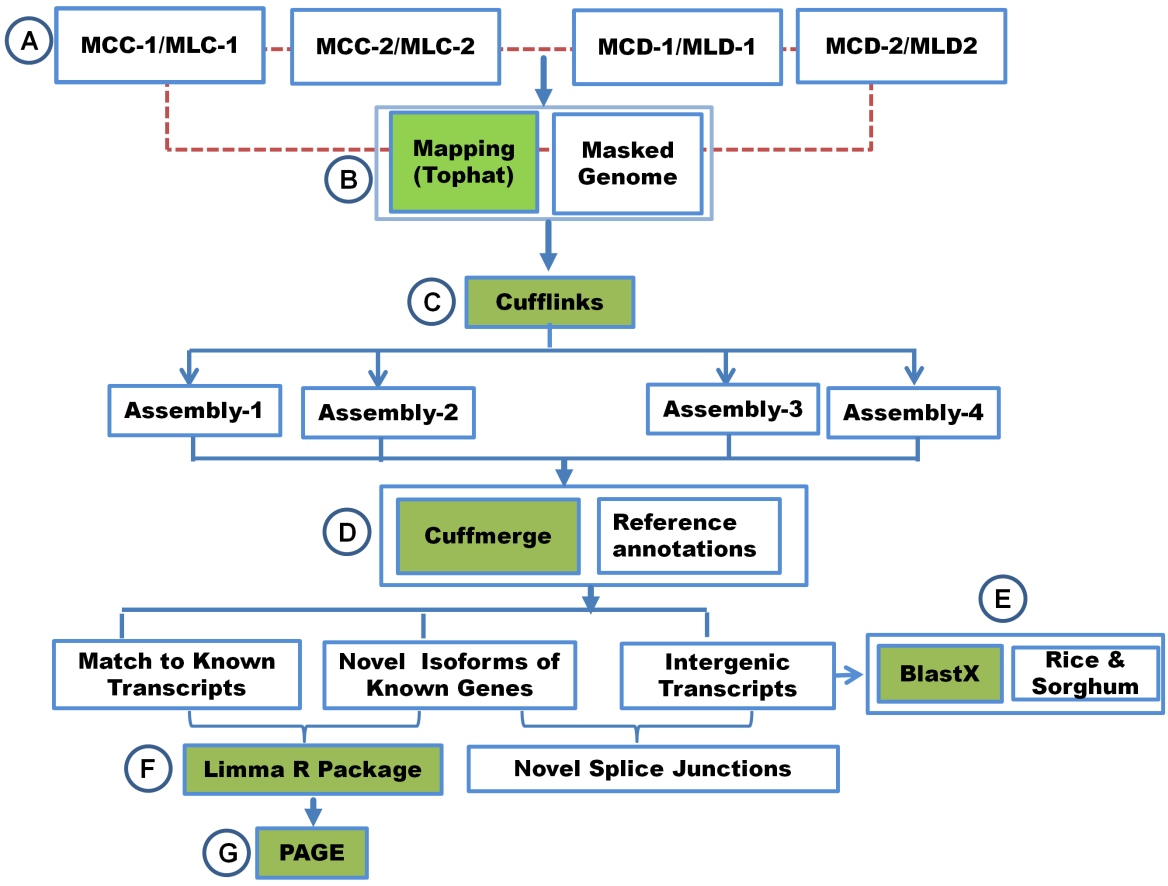


Figure 3.5: RNA-Seq workflow showing different steps implemented from raw reads to GO functional enrichment

Chapter 4

Analysis of Selected Enriched GO Terms

This chapter gives a detailed analysis of cell-cycle-related and antioxidant-gene-related GO terms. To obtain a better understanding of drought-mediated embryo abortion, it is essential to look at the expression of "Programmed Cell Death" genes. This chapter also gives a detailed analysis of the expression of PCD, phospolipase, and PI signaling genes, obtained by searching the RNASeq data for known gene from literature. This method was adopted because of the absence of related GO annotations in maize.

4.1 Introduction

Gene Ontology is a collaborative major bioinformatics initiative to define a controlled vocabulary of terms to represent gene and gene product attributes across different organisms (`http://www.geneontology.org/`). Every gene product in an organism that has defined gene ontologies has three attributes: molecular process, biological process, and cellular component. The molecular process attribute of a gene describes molecular activity, such as enzymatic and catalytic activity that occurs at a molecular level. The biological process attribute of a gene gives the series of events accomplished by one or more ordered assemblies of molecular functions. The cellular component attribute gives the location of the gene product in a larger cellular component. The GO ontologies are defined by a team of experts. Each ontology is a directed acyclic graph (DAG) with explicit parent-child relationships. Each annotation has an evidence code attached to it describing the source of the annotation. All the evidence codes fall under the following four categories; Experimental Evidence Codes, Computational Analysis Evidence Codes, Author Statement Evidence Codes, and Curator Statement Evidence Codes. Maize GO functional annotations were downloaded from the AgBase website (`http://www.agbase.msstate.edu/cgi-bin/information/Maize.pl`). Only the biological process category was considered for the functional enrichment analysis. In the maize annotation used, approximately 8830 genes were annotated to GO-Biological Process terms (as of May 2011). There was a high degree of overlap between some terms, so, in order to reduce the redundancy, the following procedure was adopted to merge terms. The terms were first sorted in descending order based on the number of genes annotated to them.

Then, for each term in the sorted list, the Jaccard index [36] was calculated between the selected term and each of the next four terms in the sorted list. The terms were then merged into one if the Jaccard index between them was greater than 0.95. These merged terms were then considered for functional enrichment.

4.2 Parametric Analysis of Gene Set Enrichment

For GO enrichment analysis, the algorithm in [37], also known as PAGE, was implemented. The input for this algorithm is a list of genes and a parameter that can be either a test-statistic or log-fold change; here the T-statistic was used as the parameter. Suppose a GO-term has m genes annotated to it. According to the central limit theorem, which is the basis for PAGE, the mean of the T-statistic of these m genes is normally distributed with mean μ (same as the parent distribution) and standard deviation δ/m (δ is the standard deviation of the parent distribution) if m is greater than 5. Hence only those terms which are annotated to at least 5 genes or more were considered. The z-score for each term is given by the following formula: $Z = (Sm - \mu) * m^{1/2}/\delta$; where Sm is the mean of the T-statistic value for the given set of genes. P-values were calculated for each GO-Term from the normal distribution and were corrected for multiple hypothesis testing using the QVALUE R package [35]. A q-value cutoff of 0.05 was used to select enriched GO-terms. The GO-enrichments results are represented in the form of a heat map (Figure 4.1). Multiple Experiment Viewer downloaded from `http://www.tm4.org/mev/` was used to draw the heat

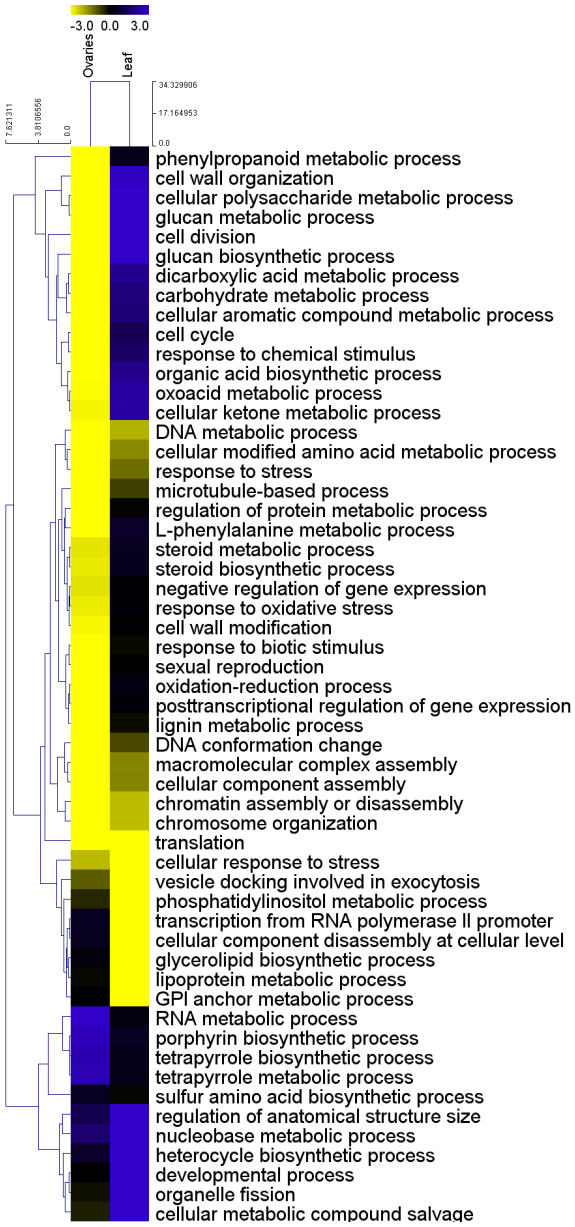


Figure 4.1: Heat map showing the different biological process GO terms enriched in maize reproductive and vegetative tissue under drought stress

map. From the heat map representation (Figure 4.1) it is evident that meristem cells, which give rise to leaf tissue, are less affected by drought than the ovary tissue at 1DAP. Several

distinct patterns of gene expression specific to one of the two drought stressed tissues are shown in Figure 4.1. GO categories corresponding to various aspects of carbon metabolism (e.g. 'cellular polysaccharide metabolic process', 'glucan biosynthetic process', and 'organic acid biosynthetic process') were enriched in down-regulated genes in the case of the drought stressed ovary tissue, whereas the opposite result was observed for leaf meristem tissue. A similar contrast was observed for the GO categories 'cell division' and 'cell cycle'. In the case of the GO categories 'response to stress', 'response to oxidative stress', and 'oxidation-reduction process', there was no significant enrichment for the leaf meristem tissue, but the ovary tissue showed enrichment for down-regulation. The category 'translation' showed down-regulation in both tissues. For the category 'vesicle docking involved in exocytosis' and several lipid-associated categories 'phosphatidylinositol metabolic process', 'lipoprotein metabolic process', and 'glycerolipid biosynthetic process', enrichment for down-regulation was observed in the leaf meristem tissue with no significant enrichment in the ovary tissue. A select few of these categories were chosen for detailed examination.

4.3 Cell Cycle and Cell Divisions

The cell cycle of a plant cell has four major phases; G1, S, G2, and M. G1 and G2 are gap phases and are highly variable in length. The S phase is the DNA synthesis phase where the whole genome of the plant cell is replicated. M is the mitosis phase where cell mitosis takes place. S and M phases are constant in length. Once the cell enters the M phase through the

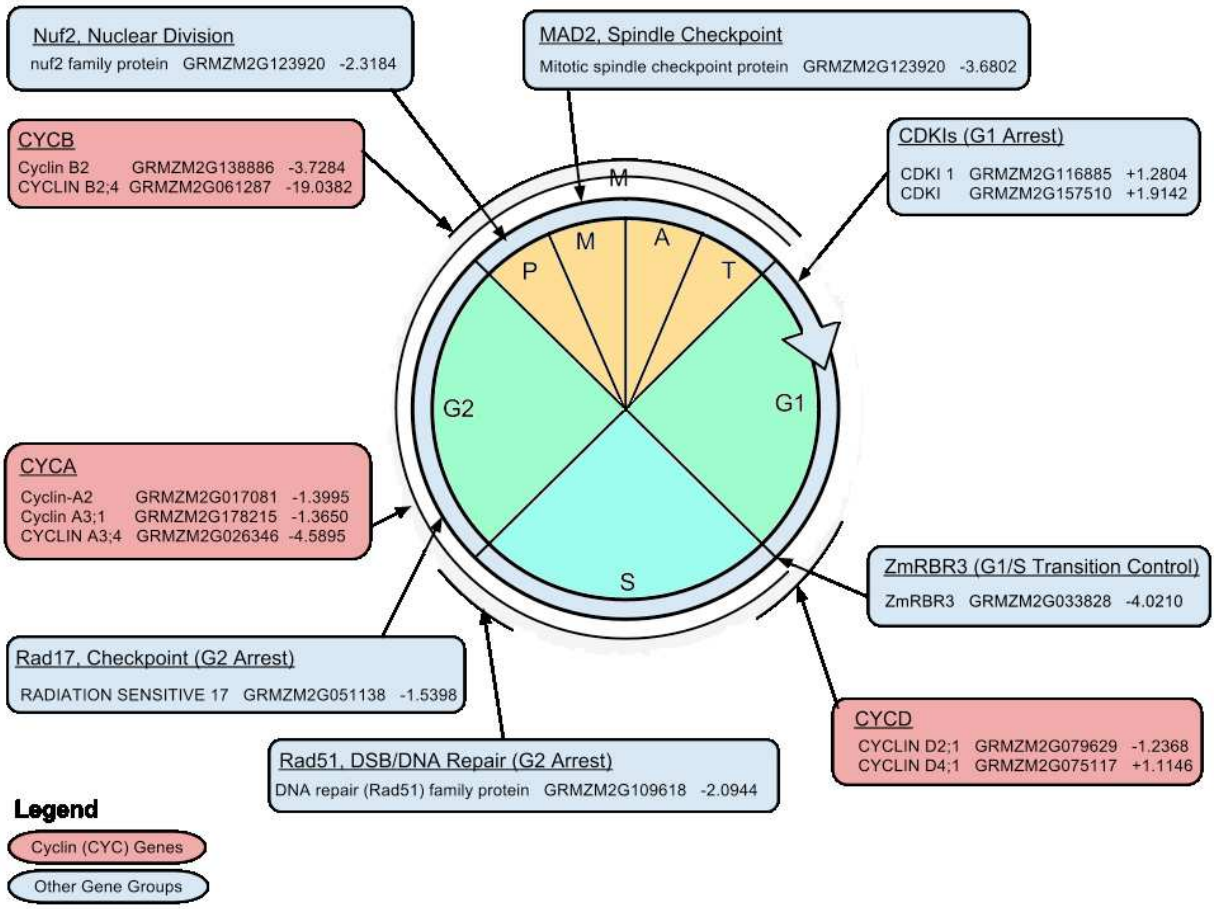


Figure 4.2: RNA-Seq expression data in drought stressed ovary tissue with associated cell cycle phases

S phase it is committed to complete the cell cycle. If the cell is not capable of completing the cycle, due to insufficient cellular resources, the result can be cell death. Because of the severity of the outcome, cell division is highly regulated and coordinated [38]. In order to understand the effects of drought on the highly regulated cell cycle, the expression of the genes annotated to the "cell-cycle" (GO:007049) GO-term were analyzed. The "cell-cycle" GO-term annotated to 49 genes was enriched with down-regulated genes with a z-score of -3.6 in the ovaries, whereas it was enriched with few up-regulated genes with a z-score of 1.2 in the

leaf meristem. Genes with a log-fold value less than 1 and with a q-value greater than 0.05 were excluded from further analysis, the expression of the remaining genes is shown in Table 4.1. The most important cell cycle factors are the cyclin dependent kinase (CDK) which are regulated by the cyclins (CYCs). Plant cyclins are typically categorized into four classes, A-D. A-type cyclins are present in S phase to M phase and type-B are maximized at the G2-M transition and M phase. D-type cyclins regulate G1-S transition (Figure 4.2) [39]. In the current data set, eleven cyclins were identified, with nine down-regulated and one, CYCLIN D4 (GRMZM2G075117), up-regulated in the ovary tissue. Two genes (GRMZM2G140633 and GRMZM2G476685) with the same cyclin delta-2 annotation were up-regulated in the leaf meristem tissue (Figure 4.2 and Table 4.1). Similar down-regulation of cyclins was observed in wheat (*Triticum aestivum*) by [40], leading to accumulation of the cells in G1 and G2 phases under drought stress.

Another important family of genes involved in cell cycle regulation in plants are the RBR (Plant Retinoblastoma-related) genes. There are 3 genes belonging to this family in maize. ZmRBR1 inhibits the cell cycle through effects on chromatin remodeling and ZmRBR3 exerts a positive influence through controlling DNA replication and minichromosome maintenance [41]. The transition from mitotic cell cycle to the endoreduplication cell cycle (G/S phases producing triploid DNA content) in endosperm is associated with low ZmRBR3 and increased ZmRBR1 gene expression [42]. ZmRBR2, whose role is unknown, is constitutively expressed at fifty percent of ZmRBR1 levels [43]. Our experimental results show down-regulation of ZmRBR3 and non-significant up-regulation of ZmRBR1 and ZmRBR2. This drought-

stressed expression pattern does not match healthy cell cycle or endoreduplication cycle expectations.

Cyclin dependent kinase inhibitors (CDKIs) promote G1 arrest in the cell cycle. In maize endosperm tissue, mutants overexpressing the CDKI, Zeama;KRP;1, counter retinoblastoma-related protein function display endoreduplication without division [44]. Modulation of the CDKs plays an important role in the transition from mitotic cell cycle to the endoreduplication cell cycle [42]. Two CDKIs (cyclin dependent kinase inhibitor and cyclin dependent kinase inhibitor 1) were up-regulated in ovary tissue (Table 4.1). A third CDKI, cyclin dependent kinase inhibitor 2, showed up-regulation below the cutoff value. Up-regulation of the CDKIs is consistent with disrupted cell cycle progression in the drought stressed ovary tissue.

In yeast and humans, RAD17 participates in a checkpoint protein complex involved in DNA damage G2 cell cycle arrest. In Arabidopsis, RAD17 homologue (AtRAD17) mutants show increased sensitivity to DNA damaging chemicals and delays in double strand break (DSB) repair. The frequency of inter-chromosomal homologous recombination increases along with general deregulation of repair [45]. An Arabidopsis rad9-rad17 double mutant showed no induction of ribonucleotide reductase subunit (AtRNR2b) and could not sustain nucleotide production for repair [46]. Our data show RAD17 expression was down-regulated in ovary tissue and unresponsive in leaf meristem.

In yeast, RAD51 plays a key role in DSB/DNA repair and homologous recombination. In *Zea mays*, RAD51 has a similar role. Double mutants of the RAD51 homologs show male

sterility, reduced seed set, and a high susceptibility to DSBs [47]. MAD2 was originally identified in budding yeast, as a highly conserved protein involved in the spindle checkpoint system ensuring that the metaphase is complete before the progression to anaphase. A similar role has been demonstrated for the Zea mays MAD2 homolog with the localization of the MAD2 homolog to the prometaphase kinetochore [48]. RAD51 was down-regulated in maize drought stressed ovary tissue and undetected in leaf meristem. The down-regulation of both RAD17 and RAD51 suggests disruption of cell cycle control.

Nuf2, in yeast, is associated with the spindle pole body (SPB) and a candidate for SPB separation. Work with mutants demonstrated that Nuf2 is required for nuclear division. Using antibodies to yeast Nuf2, a similar protein is found in the SPB of mammals [49]. There is no direct evidence for the role of Nuf2 homolog in plants, except sequence homology. Nuf2 exhibited down-regulation in drought-stressed ovary tissue, contrary to the expected up-regulation in a rapidly dividing tissue. Expression of Nuf2 in the leaf meristem was not detected.

Taken together (Figure 4.2), it is clear that, in ovaries, the expression of genes regulating the cell cycle under drought stress do not match that of a healthy cell cycle. On the other hand no such responses were seen in the leaf meristem.

4.4 Responses of Antioxidant Genes

In order to study the responses of the antioxidant genes, the GO term "response to oxidative stress" was analyzed in detail. Many genes annotated to this category were down-regulated by drought in the ovary tissue but the opposite response was seen in the leaf meristem. Out of the 42 differentially expressed genes annotated to this category in the ovary tissue, all but 6 were down-regulated; homologs of catalase 2 and APX4 were among the up-regulated ones. The majority of the down-regulated genes belonged to the peroxidase superfamily (Table 4.2). Homologs of several members of the ascorbate peroxidase group, such as APX1, were down-regulated. Genes for thioredoxin reductase, a glutathione-S-transferase, a cytosolic Cu/Zn, and a mitochondrial Mn superoxide dismutase were also down-regulated (Table 4.2). The down-regulation of these genes in the ovary tissue suggests that the antioxidant defense process was failing in the ovary tissue under the level of drought stress imposed. Specifically, the down-regulation of the anti-oxidant gens results in a failure to scavenge the accumulated ROS, thus resulting in ROS buildup. An analogous down-regulation of anitxidant genes was observed in a more drought sensitive accession of Andean potato, when compared to a more drought-tolerant accession [50, 51]. On the other hand, the drought stress did not have this effect in the antioxidant genes in the leaf meristem, where the antioxidant defense mechanism has occurred as has been observed many times.

Overall, the down regulation of several key peroxidase genes might have resulted in the suppression of well-established resistance pathways that can essentially mount a resistance

to oxidative stress, including the protection of chromosomal DNA that arise in response to drought [52, 7].

4.5 Programmed Cell Death Gene Responses

Programmed cell death (PCD) genes encode proteins for the pathway that removes the damaged tissue is an organized and a safe way. When the drought stress is severe, to the point that ROS scavenging is ineffective, PCD plays a crucial role by organized destruction of the cells that is essential for organ and tissue formation. The expression profiles of several known PCD genes, in the current dataset, were identified by searching the RNASeq data for known PCD gene terms (lsd, lol, lis, metacaspas, dad, mlo, and bax/bak). Twenty six genes were identified in the data set, of which 11 were differentially expressed in the ovaries and 3 in the leaf meristem (Table 4.3).

The differentially expressed genes, in both tissues, belonged to 4 main classes; lsd (lesions simulating disease resistance) family, metacapase, mlo (modulator of defense and death) family, and the bax inhibitors.

LSD1 is a negative regulator of cell death initiated by localized superoxide production and may reduce cell death spread (Dietrich et al 1994). Drought stressed ovary tissue showed a marked down regulation of the lsd1 homolog.

Metacaspases are cysteine dependent arginine/lysine specific proteases found in plants and play a similar role to animal PCDs [53]. In the drought stressed ovary tissue, a mixed

response of the metacaspase was seen, with up-regulation of 2 matacaspases (annotated as *Lol3* and *Metacaspase 1)*), and down regulation of 2 metacaspases (annotated as *Metacaspase 5* and *Metacaspase type II*). While, in the leaf meristem, only one gene (*Metacaspase 5*) was up-regulated.

The Mlo family of genes are plant specific genes known to be associated with stress in several species. Several loss of gene mutant analyses in barley demonstrated spontaneous PCD, indicating the negative regulation of the Mlo gene family on PCD [54]. In the current RNASeq dataset 5 genes belonging to the Mlo gene family were down-regulated. Only one gene, *Mlo4*, was up-regulated in the leaf meristem.

Two BAX inhibitor genes were differentially expressed in either the leaf or the ovary tissue. BAX inhibitor I, is a trans-membrane protein in the endoplasmic reticulum, and is ubiquitous in the eukaryotes as a cell death suppressor [55]. This gene was up-regulated in the leaf meristem but was not differentially expressed in the ovary tissue. The other Bax inhibitor gene was down-regulated only in the ovary tissue, possibly contributing to PCD.

Taken together, the expression profiles of the PCD genes indicates increased PCD occurring in the ovary tissue, while no such responses were evident in the leaf meristem.

4.6 Phospholipase and PI Signaling Associated Genes

Phospholipase C catalyzes the conversion of membrane bound phosphotidylinositol species into phosphorylated inositol isoforms and diacylglycerol (DAG), each of which participate

in PI signaling. In the current data set, seventeen genes annotated to PLC were identified (Table 4.4), all the genes were categorized into 4 broad classes: PLAs, PLCs, PLD,s and miscellaneous. PLC transcript levels have shown to vary under drought conditions [56]. In our dataset, of the four PLCs, three showed significant down-regulation in ovary tissue and insignificant changes in the leaf meristem, while one gene, *Phosphatidylinositol-specific phospholipase C4*, showed an insignificant change in expression in the ovary and a massive down-regulation in the leaf meristem. This down regulation is likely to have a significant effect on the availability of $Ins(1,4,5)P_3$ for signal transduction in both the tissues. Hence, it is unlikely that $Ins(1,4,5)_3$ associated signaling plays a role in the differential drought sensitivity of the ovary tissue, compared to the leaf meristem. 5PTase and PIP5K, other genes involved in $Ins(1,4,5)P_3$ signaling, showed mixed responses, yielding inconclusive results.

Five genes annotated as PLDs were differentially expressed in the ovary tissue (Table 4.4). Of the differentially expressed PLD genes, 3 of them were down-regulated in the ovary tissue, while the remaining two were up-regulated. In contrast, in the leaf meristem, only one gene was up-regulated. PLDs play a very important role in plant growth, development and environmental stress responses [57, 58, 59].Enzymatic activity of the PLD gene family results in the production of PA, another signaling molecule that responds to abiotic stress [60]. It is also known that, in the case of drought stress, increased PLD alpha activity results in activation of ABA signaling [59]. Arabidopsis plants down-regulated with respect to PLD alpha expression were found to be more drought sensitive than their corresponding wild types [50]. However, due to the mixed expression of the PLD gene family genes in the

ovary tissue, the data is difficult to interpret. In contrast, phospholipase D delta, which is known to be involved in ROS responses in Arabisopsis, were found to be down-regulated in the ovary tissue but not in the leaf meristem. This is in agreement with the finding that cell cycle genes and major ROS scavenging genes were down-regulated in the ovary tissue and not in the leaf meristem. Overall, the data (Table 4.4) suggests that the signaling pathway in which PLD delta participates was specifically down-regulated in the ovary tissue under drought stress. The repression of this pathway might have resulted in the failure of the antioxidant processes, and of signaling mediated through regulatory genes such as APX1, leading to arrest of the cell cycle and eventual cell death.

Table 4.1: Maize drought stress gene expression for cell cycle related genes. Annotation sources: (Zm) Zea mays (MaizeSequence.org), (Os) Oryza sativa (Rice Genome Annotation Project), (At) Arabidopsis thaliana (TAIR). Expression values are log-fold values and a dash (-) indicates failure to meet significance cutoff or undetected

Gene Product Category	Gene	Annotation	Ovary(Log2 Ratio)	Leaf Meristem (Log2 Ratio)
Cyclins	GRMZM2G073671	Cyclin3^{Zm}	-18.59	-
	GRMZM2G140633	Cyclin.delta2^{Zm}	-1.5	1.01
	GRMZM2G017081	Cyclin-A2^{Zm}	-1.4	-
	GRMZM2G034647	Cyc1—cylin1^{Zm}	-19.3	-
	GRMZM2G138886	CylinB2^{Zm}	-3.73	-
	GRMZM2G476685	Cylin.delta2^{Zm}	-	1.38
	GRMZM2G026346	Cylin.A3.1^{At}	-1.37	-
	GRMZM2G079629	Cylin.D2.1^{At}	-1.24	-
	GRMZM2G061287	Cylin.B2.4^{At}	-19.04	-
	GRMZM2G026346	Cylin.A3.4^{At}	-4.59	-
	GRMZM2G075117	Cylin.D4.1^{At}	1.11	-
Cyclin Dependent Kinase Inhibitor	GRMZM2G116885	cyclin dependent kinase.inhibitor1^{Zm}	1.28	1.11
	GRMZM2G157510	cyclin dependent kinase.inhibitor^{Zm}	1.91	-
Miscellaneous cell cycle products	GRMZM2G033828	Retinoblastoma-related protein 3 (ZmRBR3)^{Zm}	-4.02	-
	GRMZM2G109618	DNA repair (Rad51) family protein^{At}	-2.09	-
	GRMZM2G051138	Radiation Sensitive 17^{At}	-1.54	-
	GRMZM2G047143	Mitotic spindle checkpoint protein MAD2^{Zm}	-3.68	16.34
	GRMZM2G123920	nuf2 family protein^{Zm}	-2.32	

Table 4.2: Genes annotated to Response to Oxidative Stress GO-term. Annotation Sources: (Zm) Zea mays (MaizeSequence.org), (At) Arabidopsis thaliana (TAIR). The expression values are log-fold and (-) indicates failure to meet the significance cutoff or undetected

Gene Product Category	Gene	Annotation	Ovary(Log2 Ratio)	Leaf Meristem (Log2 Ratio)
Ascorbate Peroxidases	GRMZM2G460406	Ascorbate Peroxidase 3^{At}	1.13	-
	GRMZM2G156227	Ascorbate Peroxidase 4^{At}	19.52	-
	GRMZM2G004211	APx3-Peroxisomal Ascorbate Peroxidase^{Zm}	-1.62	-
	GRMZM2G014397	Stromal Ascorbate Peroxidase^{At}	-1.62	-
	GRMZM2G137839	Ascorbate Peroxidase 1^{Zm}	-1.89	-
Miscellaneous	GRMZM2G059991	Superoxide Dismutase^{Zm}	-1.62	-
	GRMZM2G079348	Catalase Isozyme 3^{Zm}	-1.68	-
	GRMZM2G088212	Catalase Isozyme 1^{Zm}	1.44	-
	GRMZM2G116273	gst1— Glutathione S-transferase 1^{Zm}	-3.11	-
	GRMZM2G422750	Chalcone Synthase C2^{Zm}	-24.5	
	GRMZM2G135893	GP Protein^{Zm}	-1.49	

Table 4.3: Maize gene expression for Programmed Cell Death (PCD) related genes in drought stressed ovary and leaf meristem tissues. Annotation Sources: (Zm) Zea mays (MaizeSequence.org), (Os) Oryza sativa (Rice Genome Annotation Project), (At) Arabidopsis thaliana (TAIR). Ovary expression and leaf meristem expression are log-fold values and a dash (-) indicates failure to meet significance cutoff or undetected

Gene Id	Annotation	Ovary(Log2 Ratio)	Leaf Meristem (Log2 Ratio)
GRMZM2G055135	LSD1 zinc finger domain containing protein^{Os}	-3.86	-
GRMZM2G120069	LOL3^{Zm}	3.55	-
GRMZM2G066041	Metacaspase 5^{At}	-1.04	1.4
GRMZM2G083016	Metacaspase type II^{Zm}	-4	-
GRMZM2G120079	Metacaspase 1^{At}	4.75	-
GRMZM2G040441	Barley mlo defense gene homolog2^{At}	-2.26	-
GRMZM2G051974	MLO-like protein 14^{At}	-2.82	-
GRMZM2G089259	Barley mlo defense gene homolog4^{At}	-2.1	1.13
GRMZM5G881803	Barley mlo defense gene homolog8^{At}	-2.12	-
GRMZM2G031331	Barley mlo defense gene homolog3^{At}	-1.14	-
GRMZM2G029087	BAX inhibitor 1^{At}	-	1.7
GRMZM2G095898	Transmembrane BAX inhibitor motif-containing protein^{Os}	-2.0	-

Table 4.4: Genes annotated to phospholipase in the dataset. Annotation Sources: (Zm) Zea mays (MaizeSequence.org), (At) Arabidopsis thaliana (TAIR). The expression values are log-fold and (-) indicates failure to meet the significance cutoff or undetected.

Gene Id	Annotation	Ovary(Log2 Ratio)	Leaf Meristem (Log2 Ratio)
GRMZM2G035421	Patatin-like phospholipase family protein^{At}	1.76	-
GRMZM2G087612	Patatin-like phospholipase family protein^{At}	1.44	-
GRMZM2G137435	Phosphoinositide-specific phospholipase C family protein^{At}	-1.41	-
GRMZM2G072578	Phospholipase A2 family protein^{At}	-4.13	-
GRMZM2G166971	Phospholipase A2 family protein^{At}	2.2	-
GRMZM2G139041	non-specific phospholipase C1^{At}	-1	-
GRMZM2G081719	non-specific phospholipase C6^{At}	-3.63	-
GRMZM2G114354	phosphatidylinositol-specific phospholipase C4^{At}	-	-17.2
GRMZM2G154523	phospholipase A 2A^{At}	-2.22	3.1
GRMZM2G349749	phospholipase A 2A^{At}	1.74	1.28
GRMZM2G019029	phospholipase D alpha 1^{At}	2.5	1
GRMZM2G054559	phospholipase D alpha 1^{At}	-1.1	-
GRMZM2G179792	phospholipase D alpha 1^{At}	2.03	-
GRMZM2G442551	phospholipase D alpha 2^{At}	1.32	-
GRMZM2G108912	phospholipase D delta^{At}	-1.3	-
GRMZM2G451672	phospholipase; galactolipase^{At}	-1.16	-
GRMZM2G451672	triacylglycerol lipase^{Zm}	-2.24	-

Chapter 5

Analysis of Selected Enriched MapMan Bins

This chapter gives a detailed analysis of MapMan [61] bins involved in starch and sucrose metabolism and catabolism. Another important class of enzymes involved in maintaining hexose/sucrose ration are the invertases; this chapter gives a detailed analysis of the invertases that were differentially expressed. MapMan bins related to raffinose, trehalose and ABA metabolism are also discussed in this chapter.

5.1 Starch and Sucrose Metabolism

In order to understand the effects of drought on the starch and sucrose metabolism in the maize tissues, the genes overlaid on the sucrose and starch synthesis bins in the primary

metabolism mapping file in MapMan were studied. Several genes encoding enzymes involved in the synthesis of starch, namely, ADP-glucose pyrophosphorylase, starch synthase, starch-branching enzyme, and starch debranching enzyme were significantly down-regulated in the drought stressed ovary (Figure 5.1). These results validate the finding of [5], which observed a lower level of starch in water stressed ovaries. It has long been known that AGPase activity represents the rate limiting step in starch biosynthesis in all the tissues examined to date, with maize endosperm being no exception [42]. Our results showed the down regulation of 12 genes encoding five major enzymes for starch synthesis in the ovary tissue. The down-regulated classes included four isoforms of AGPase, one soluble starch synthase, three granular-bound starch synthase (GBSS), two isoforms of starch branching enzyme (SBE), and two isoforms of starch debranching enzymes (DBE) (Figure 5.1). In contrast, none of the genes showed significant changes in the drought stressed leaf meristem.

The several classes of enzymes involved in starch biosynthesis have very specific functions. In cereals, the GBBS class of genes are encoded in the *Waxy* locus and have an important effect on amylose synthesis and is exclusively bound to a starch granule [62]. Maize endosperm contains at least five starch synthase isoforms that are categorized according to conserved sequence relationships. In our study, we observed the down-regulation of multiple starch synthase isoforms (Figure 5.1). Starch synthase isoforms have unique roles in starch synthesis, and this has been experimentally validated through mutant studies in different plants. For example, SSIIIa mutants of rice and maize are deficient in long chains of amylopectin, suggesting their specific role in chain elongation.

From the data, it is clear that genes encoding proteins/enzymes at different steps of the starch biosynthesis were down-regulated, along with the rate limiting step carried by the AGPase class of enzymes, suggesting a mechanism that coordinates this effect. Other effects of repressing AGPase expression were reported in AGPase knockdown pea (*Pisum sativum*) seeds [63]. Seeds of the knockdown plants showed an expected metabolic shift in composition from starch towards the accumulation of sugars, but also the repression of cell wall and vacuolar invertase gene expression, an increase in sucrose synthase expression, increases in ROS, and changes in the interactions of ABA and sugar signaling. These changes resonate with our data on coordinated regulation, although the pea seeds studied were quite mature compared to the maize ovaries, and no imposed abiotic stress was involved.

Several isoforms of alpha and beta amylases were up-regulated in the stressed ovaries (Figure 5.1; bin 5). The breakdown of starch in plant storage organs is catalyzed by amylase, this being the only enzyme known to attack raw starch granules. The cooperative action of a range of other degrading enzymes, including β-amylase, debranching enzyme, and starch phosphorylase, result in the dissolution of the stored starch leading to the production of a number of oligosaccharides, and finally to maltose, glucose or glucose-1-phosphate as shown in Figure 5.1. The gene encoding maltose excess protein 1-like (GRMZM2G156356) was evidently down-regulated at 1DAP (Figure 5.1). Genes encoding sucrose synthase1 (GRMZM2G089713) and sucrose synthase2 (GRMZM2G318780) were up-regulated by 1.7 and 1.5 fold, respectively, at 1DAP in the stressed ovaries (Figure 5.1). The increase in the transcript levels of the two sucrose synthases suggests the possibility of sucrose hydrolysis

and the entrance of hexose as the primary sugar in endosperm metabolism. However, since low glucose signaling (Table 5.2), which is a response to carbon deprivation, appears to have been activated, not every tissue or cell type in the ovary at 1DAP are likely to have received sufficient hexose to meet its metabolic needs.

Overall, the data suggests the down-regulation of the starch and sucrose biosynthesis and up-regulation of the starch and sucrose degradation pathway in the drought stressed ovaries, with no comparable change in the leaf meristem.

5.2 Invertase in its Metabolic and Likely Signaling Role

Invertase plays a key role in providing hexose to the embryo and endosperm to support the cell cycle and cell division [64]. Furthermore, our current RNASeq data revealed several invertases that were differentially expressed in the ovary tissue and the leaf meristem. As the activities of the invertases are critical and since many invertases were differentially expressed in both tissues, the expression of members of the invertases family were validated using qRT-PCR in both tissues (Figure 2.2). As shown in Figure 2.2, the expression of genes for two soluble invertase (*Ivr1* and *Ivr2*), three cell wall invertase (*Incw1, Incw2* and *Incw4*) and three neutral invertase (*Inn1, Inn2* and *Inn3*) were significantly down-regulated in the drought stressed ovary tissue, while, in the leaf meristem, one invertase (*Inv1*) was slightly up-regulated. In addition to the differentially expressed genes detected by qRT-PCR, three more genes encoding proteins similar to neutral invertase (GRMZM2G007277,

GRMZM2G115451, and GRMZM2G18737) were significantly down-regulated in the ovary tissue. It has been widely reported that invertase activity in the developing seeds of maize is mainly due to two cell wall invertase genes [65]. In a mutant deficient in *INCW2* reduced cell number and cell size were also observed [66]. An equivalent result was observed in our data where the expression of *Incw1* and *Incw2* is lowered (Figure 2.2), the cell cycle appears to be inhibited, and ultimately seed number was reduced (Table 2.2).

Both cell wall and soluble invertases were down-regulated under drought stress in the ovary, whereas one soluble invertase was up-regulated in the leaf meristem. This suggests a distinct difference in the response of sugar metabolism and most likely sugar signaling between the two tissues under drought stress. The down-regulation of invertases in the ovary echoes the results of [1], which also reported an accompanying buildup of sucrose in that tissue under drought stress. The consequence of this is most likely a decrease in glucose levels, one of the products of invertase activity, disrupting the hexose/sucrose ratios in the ovary, as discussed by [1]. Furthermore, the buildup of sucrose in drought stressed ovaries demonstrated by [1], makes it less likely that the supply of photoassimilate to the ovary is a crucial player in the early embryo abortion phenomenon, as opposed to later stages of seed development when demand for carbon skeletons is much greater.

Cell wall invertases may also play a signaling role since, in Arabidopsis, a cell wall invertase was demonstrated to bind to phosphatidyl monophosphate 5-kinase, a component of a signaling pathway associated with the regulation of growth and the mediation of cellular drought signaling. However, phospholipase C genes, which catalyze the crucial IP3 generat-

ing step in PI signaling, were down-regulated in both tissues, suggesting that differences in PI signaling may not form part of the mechanism(s) that underlies the relative sensitivity of early maize embryos to drought stress. The result of the repression of the soluble invertase will be less substrate available to associate with mitochondrial hexokinase to provide ADP to maintain coupled electron transport. Hexokinase, as a signaling molecule, senses a decrease in available glucose and activates the SnRK1 pathway; however, in the drought-stressed ovaries, hexokinase most likely cannot function metabolically because glucose is not available. Under these conditions, increased ROS arises in the mitochondrion and, most likely, other subcellular locations (due to a lack of substrate) to the point where the ROS scavenging genes are down-regulated, a symptom of stress susceptibility. Comparable events were not detected in the leaf meristem; in contrast, in the leaf meristem the increase in invertase activity can be interpreted as a response to a decrease in available carbon skeletons.

5.3 Hexokinase 1 in its Metabolic and Signaling Role

Hexokinase 1 exerts the same catalytic function as other members of the gene family [67]. A continuous and efficient supply of glucose to hexokinases, including the enzyme that is associated with the mitochondrion, is necessary to maintain its activity at a rather constant level, which is needed to control the flux through the mitochondrial electron transport chain and avoid ROS buildup [68]. The maize homolog of HXK1, the sensor of low glucose levels, was up-regulated in the ovary, as were HXK1 targets such as a homolog of ABI5, while a

homolog of CAB2 was down-regulated, as predicted by the HXK1 sensing pathway (Figure 5.2). This is consistent with signaling through the low glucose signaling pathway established in Arabidopsis [69, 67, 70, 71]. The response of the CAB homolog in a non-photosynthetic tissue, such as the maize ovary, suggests that this gene can function as a signaling component in an environment in which the encoded protein is not fulfilling its better known role as part of the photosynthetic apparatus. The low glucose sensing pathway functions under stress conditions, and, under carbon deprivation with low glucose, it is reported that hexokinase1 signaling is uncoupled from stress responses and from its interaction with ABA-related events [67]. This may have occurred in the drought stressed maize ovaries. However, since the experimental system employed by those workers is so far removed from the maize system, it is difficult to determine whether or not such an uncoupling took place under drought in the drought-stressed ovaries.

5.4 Responses in Raffinose and Trehalose Metabolism

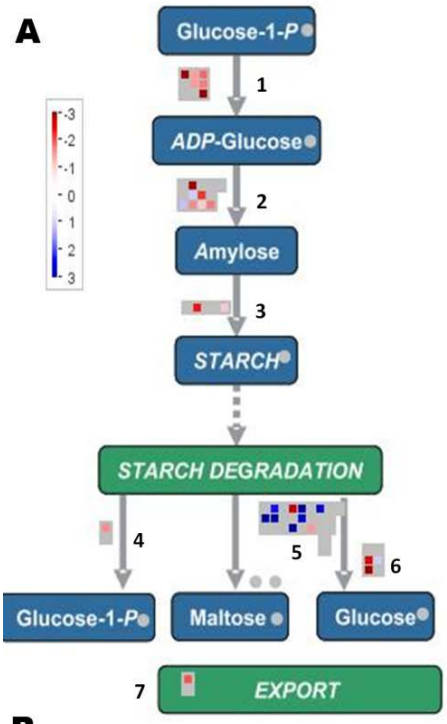
The MapMan bin corresponding to raffinose synthesis genes was studied to understand the stress response on raffinose metabolism in both tissues. As shown in Figure 5.3, two genes (GRMZM5G872256 and GRMZM2G165919) encoding galactinol synthase were up-regulated in the ovary tissue, whereas no galactinol synthase genes were differentially expressed under drought stress in the leaf meristem. Four genes annotated to seed imbibition, were up-regulated in the ovary, while two genes belonging to the same class were down-regulated.

Only one gene annotated to seed imbibition was differentially up-regulated in the leaf meristem (Figure 5.3). The up-regulation of these genes in the ovary tissue suggests an increase in raffinose and galactinol synthesis. Increased levels of galactinol and raffinose are known to be involved in ROS (Reactive Oxygen Species) scavenging [72, 73, 69]. However, the ROS scavenging action provided by raffinose may not have compensated for the loss of ROS scavenging action by down-regulated anitoxidant genes (Table 4.2), since embryo abortion still took place. Galactinol and raffinose synthesis genes have a W-box cis element in the promoter region, which is bound by SnRk1. SnRK1, along with bZip and TP6, forms a regulatory complex that is known to be functional under carbon starvation conditions [69, 74]. The data shows the up-regulation of the SnRk1 gene (Figure 5.2) in the ovary tissue, which also suggests a possible mechanism for up-regulation of the raffinose synthesis genes by SnRk1 through W-box cis element. Two trehalose phosphate synthase (TPS) genes and several putative TPS genes were up-regulated in the ovary tissue, suggesting increased TPS levels in the ovary tissue. The data suggests the activation of the SnRk1, bZip, and TPS regulatory network in the ovary tissue. One of the influences of this regulatory network is the down-regulation of the carbon metabolism-associated gene involved in biosynthesis and up-regulation of the genes involved in catabolism. The pattern observed for starch and sucrose metabolism-associated genes in drought stressed ovaries suggests such an influence of the SnRK1 complex.

5.5 ABA Related Effects

ABA is a major plant hormone that plays a key role in adaptive responses to drought stress [75]. ABA regulates a large number of genes in Arabidopsis [76]. The first committed step in ABA synthesis is catalyzed by 9-cis-epoxycarotenoid dioxygenase (NCED gene family) [77, 78]. The extent of stress-mediated synthesis of ABA is also influenced by the ZEP gene family [79]. A drought-mediated increase in ABA levels has been reported at the time of anthesis in maize exposed to stress for longer periods before pollination [18, 3]. In agreement with these findings, the data reported here show up-regulation of ABA biosynthesis genes in the ovary tissue (Figure 5.4). Genes with weak similarity to the AAO class in Arabidopsis and known to catalyze the last step in ABA synthesis were down-regulated in ovary tissue. However, AAO3, which is the gene specifically responsible for ABA synthesis in seeds [80], was unaffected by exposure to drought in either tissue. In fact, previous studies have shown that single loss of function mutants of AAO1 and AAO4 [80] (down-regulated in the present study) did not change endogenous ABA levels when compared to the wild type. A large number of transcription factors (TFs) are known to be involved in ABA mediated responses. Major positive regulators of ABA mediated response include the TFs belonging to the bZIP class [76]. Members of the bZIP gene family are known to regulate important cellular processes in all eukaryotes [76]. An important transcription factor in this class is ABI5, a positive regulator of ABA signaling, which is expressed in seeds and mainly serves as a developmental checkpoint [81]. The up-regulation of the genes involved in ABA synthesis and the ABI5 transcription factor in the drought stressed ovary tissue suggest the activation

of the ABA mediated response pathway, which induces the expression of many genes that play an important role in drought responses [82]. Members of the ABF gene family are known to be induced in vegetative tissue under drought stress [76], this is in accordance with the up-regulation of ABF4 gene in the leaf meristem observed in this thesis. The late embryogenesis abundant proteins (LEA) up-regulated in the ovary tissue belongs to another of the many classes of ABA-induced genes.

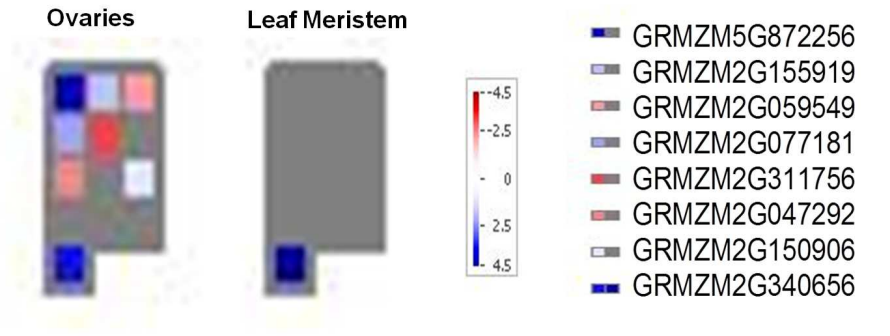


Bin ID	Maize ID	Annotation	Ovaries (Log2Ratio)
1	GRMZM2G144002	Glucose-1-phosphate adenylyltransferase large subunit 3 precursor	-3.5204
	GRMZM2G429899	Glucose-1-phosphate adenylyltransferase large subunit 1	-1.5003
	GRMZM2G027955	Glucose-1-phosphate adenylyltransferase large subunit 2	-1.5395
	GRMZM2G163437	Glucose-1-phosphate adenylyltransferase small subunit Fragment	-1.8696
	GRMZM2G068506	Similar to ADP glucose pyrophosphorylase 1	-1.7372
	GRMZM2G391936	Plastid ADP-glucose pyrophosphorylase large subunit	-4.8558
2	GRMZM2G008263	Granule bound starch synthase IIa precursor	-3.9202
	GRMZM2G130043	Starch synthase V	1.071
	GRMZM2G348551	Similar to starch synthase 2	-2.178
	GRMZM2G024993	Granule-bound starch synthase 1	-1.106
	GRMZM2G141399	Starch synthase DULL1	-1.746
3	GRMZM2G032628	1,4-alpha-glucan-branching enzyme 2	-2.334
	GRMZM2G073054	Starch branching enzyme IIa Fragment	-1.128
4	GRMZM2G074158	Phosphorylase	-1.638
5	GRMZM2G081502	Similar to alpha-amylase	3.758
	GRMZM2G025833	Beta-amylase	4.706
	GRMZM2G103055	Alpha-amylase	4.93
	GRMZM2G074781	Similar to alpha-amylase	3.733
	GRMZM2G138468	Alpha-amylase	2.59
	GRMZM2G058310	Beta-amylase	-1.543
	GRMZM2G082034	Beta-amylase	2.617
6	GRMZM2G136803	Similar to disproportionating enzyme 2	-2.565
	GRMZM2G013811	Similar to disproportionating enzyme 2	-3.12
7	GRMZM2G156356	Maltose excess protein 1-like	-1.98
8	GRMZM2G008507	Similar to Sucrose-phosphate synthase family protein	-2.2
9	GRMZM2G083248	Similar to sucrose transporter 2	2.358
	GRMZM2G081589	Similar to sucrose transporter 2	-3.04
	GRMZM2G034302	sucrose transporter1	3.183
10	GRMZM2G089713	Sucrose synthase 1	1.763
	GRMZM2G045171	sucrose synthase, putative, expressed	-0.6
	GRMZM2G152908	Sucrose synthase 2	-3.075
	GRMZM2G318780	sucrose synthase2	1.5119
11	GRMZM2G115451	Plant neutral invertase family protein	-2.989
	GRMZM2G477236	Similar to light harvesting like protein Lil3	-1.475
	GRMZM2G007277	Similar to Plant neutral invertase family protein	-2.64
	GRMZM2G118737	Similar to Plant neutral invertase family protein	-1.035
	GRMZM2G119689	Similar to cell wall invertase 2	-3.72
	GRMZM2G174249	Siimilar to cell wall invertase 4	-2.08
	GRMZM2G089836	Beta-fructofuranosidase 1	-1.171
	GRMZM2G394450	Beta-fructofuranosidase 1 Precursor	-3.394
12	GRMZM2G051677	Fructokinase-2	-2.986
	GRMZM2G072091	Similar to pfkB-like carbohydrate kinase family protein	-3.11
	GRMZM2G466780	Protein kinase	1.93
13	GRMZM2G104081	Similar to hexokinase 1	2.42
	GRMZM2G432801	Hexokinase-1	-1.648

Figure 5.1: Effects of drought stress on the expression of genes associated with starch and sucrose metabolism. Drought stress responses associated with starch (A) and sucrose metabolism (B) in ovaries with a table showing the gene names, putative function and expression values

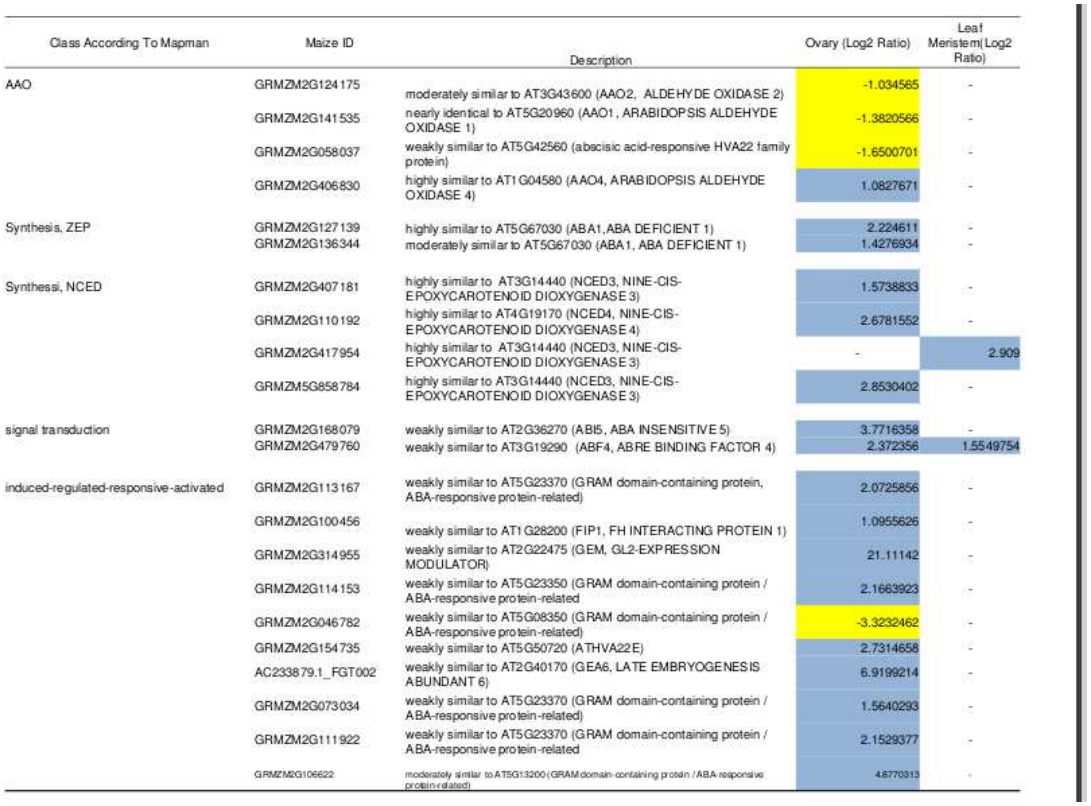
Gene	Glu (+)	Glu (-)	ABA (+)	ABA (-)	HXK Dependent	Annotation	Ovary (Log2 Ratio)	Leaf Meristem (Log2 Ratio)
GRMZM2G119941	◆					Invertase cell wall4 (incw4), mRNA[(At)]	-	1.35
GRMZM2G174249	◆					gpm24 | hypothetical protein LOC100279267 (in At cell wall invertase 4)[(At)]	-2.08	-
GRMZM2G119689	◆					Miniature seed1 (in At cell wall invertase 2)[(At)]	-3.72	–
GRMZM2G422750	◆					c2 | Chalcone synthase C2 (EC 2.3.1.74) (Naringenin-chalcone synthase C2)[(Zm)]	-24.49	-
GRMZM2G104081		◆				Hexokinase 1	2.42	-
GRMZM2G477236	◆				✓	Similar to chlorophyll A-B binding family protein (AT4G17600)	-1.47	-
GRMZM2G128876		◆			✓	Xyloglucan endotransglucosylase/hydrolase 15[(Zm)]	-3.56	1.51
GRMZM2G048085		◆			✓	Senescence-associated protein DIN1	2.21	-
GRMZM2G077278		◆			✓	SNF1 kinase homolog 10[(At)]	1.149	-
GRMZM2G168079		◆	◆		✓	ABI5|ABA response element binding factor (LOC100285149), mRNA[(Zm)]	3.77	–

Figure 5.2: Effects of drought stress on the expression of genes associated with energy level signaling in ovary and leaf meristem tissue. The checkmark indicates known positive regulation or known hexokinase signaling



Maize ID	Annotation	Ovaries (Log2 Ratio)	Leaf Meristem (Log2 Ratio)
GRMZM5G872256	Weakly similar to AT2G47180 (galactinol synthase 1)	4.098	-0.496
GRMZM2G165919	Highly similar to AT2G47180 (galactinol synthase 1)	1.842	-0.065
GRMZM2G059549	Weakly similar to AT3G57520 (seed imbibition 2)		
GRMZM2G077181	Nearly identical to AT3G57520 (seed imbibition 2)	2.21	0.01
GRMZM2G311756	Weakly similar AT3G57520 (seed imbibition 2)	-3.154	0.223
GRMZM2G047292	Moderately similar to AT3G57520 (seed imbibition 2)	-2.54	0.096
GRMZM2G150906	Highly similar to AT5G40390 (seed imbibition 1-like)	1.164	-0.138
GRMZM2G340656	Highly similar to AT3G57520 (seed imbibition 2)	3.53	22.037

Figure 5.3: Differential expression of genes involved in raffinose synthesis in ovary and leaf meristem tissues with a table showing gene names, putative function and expression values.



Class According To Mapman	Maize ID	Description	Ovary (Log2 Ratio)	Leaf Meristem(Log2 Ratio)
AAO	GRMZM2G124175	moderately similar to AT3G43600 (AAO2, ALDEHYDE OXIDASE 2)	-1.034565	-
	GRMZM2G141535	nearly identical to AT5G20960 (AAO1, ARABIDOPSIS ALDEHYDE OXIDASE 1)	-1.3820566	-
	GRMZM2G058037	weakly similar to AT5G42560 (abscisic acid-responsive HVA22 family protein)	-1.6500701	-
	GRMZM2G406830	highly similar to AT1G04580 (AAO4, ARABIDOPSIS ALDEHYDE OXIDASE 4)	1.0827671	-
Synthesis, ZEP	GRMZM2G127139	highly similar to AT5G67030 (ABA1,ABA DEFICIENT 1)	2.224611	-
	GRMZM2G136344	moderately similar to AT5G67030 (ABA1, ABA DEFICIENT 1)	1.4276934	-
Synthessi, NCED	GRMZM2G407181	highly similar to AT3G14440 (NCED3, NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 3)	1.5738833	-
	GRMZM2G110192	highly similar to AT4G19170 (NCED4, NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 4)	2.6781552	-
	GRMZM2G417954	highly similar to AT3G14440 (NCED3, NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 3)	-	2.909
	GRMZM5G858784	highly similar to AT3G14440 (NCED3, NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 3)	2.8530402	-
signal transduction	GRMZM2G168079	weakly similar to AT2G36270 (ABI5, ABA INSENSITIVE 5)	3.7716358	-
	GRMZM2G479760	weakly similar to AT3G19290 (ABF4, ABRE BINDING FACTOR 4)	2.372356	1.5549754
induced-regulated-responsive-activated	GRMZM2G113167	weakly similar to AT5G23370 (GRAM domain-containing protein, ABA-responsive protein-related)	2.0725856	-
	GRMZM2G100456	weakly similar to AT1G28200 (FIP1, FH INTERACTING PROTEIN 1)	1.0955626	-
	GRMZM2G314955	weakly similar to AT2G22475 (GEM, GL2-EXPRESSION MODULATOR)	21.11142	-
	GRMZM2G114153	weakly similar to AT5G23350 (GRAM domain-containing protein / ABA-responsive protein-related	2.1663923	-
	GRMZM2G046782	weakly similar to AT5G08350 (GRAM domain-containing protein / ABA-responsive protein-related)	-3.3232462	-
	GRMZM2G154735	weakly similar to AT5G50720 (ATHVA22E)	2.7314658	-
	AC233879.1_FGT002	weakly similar to AT2G40170 (GEA6, LATE EMBRYOGENESIS ABUNDANT 6)	6.9199214	-
	GRMZM2G073034	weakly similar to AT5G23370 (GRAM domain-containing protein / ABA-responsive protein-related)	1.5640293	-
	GRMZM2G111922	weakly similar to AT5G23370 (GRAM domain-containing protein / ABA-responsive protein-related	2.1529377	-
	GRMZM2G106622	moderately similar to AT5G13200 (GRAM domain-containing protein / ABA-responsive protein-related)	4.8770313	-

Figure 5.4: Maize genes involved in ABA synthesis and ABA mediated response as classified by MapMan. Annotation Source: MapMan. The expression values are log-fold and (-) indicates failure to meet the significance cutoff or undetected

Chapter 6

Conclusions

The RNASeq pipeline developed in this research, as shown in Figure 3.5, has various steps starting from Illumina raw reads processing until GO enrichment results. Nearly 85% of the reads mapped, uniquely or at multiple locations, to the maize masked reference genome in both tissues. This high percentage of mapped reads suggests both accuracy of the mapping tool used (Tophat), and the completeness of the B73 maize genome sequence. Out of the total mapped reads, approximately 73% of the reads mapped to previously annotated transcripts, which account for approximately 68% of the total assembled transcripts. The remaining transcripts comprise 16-17% novel spliced isoforms and 2.5-3% intergenic transcripts. The low number of novel intergenic transcripts identified suggests that the maize gene models are more or less complete. Out of the known transcripts, 13709 and 3942 transcripts in the ovary and leaf meristem, respectively, are drought responsive. Among the novel isoforms identified, 4147 in ovaries and 2404 in leaf meristem are drought responsive. The alternatively

spliced isoforms representing 22% of total assembled reads suggest a relatively high level of post-transcriptional regulation.

The GO annotation of maize gene models are incomplete; to date only 8830 gene models have biological process annotation. Hence, in order to provide additional levels of analysis, MapMan and an extensive literature review on drought related responses were used. The type of analysis described here, gives a good example in dealing with species that have very limited GO annotations.

Taken together, Figure 5.4 suggests that the ABA levels have increased a to greater extent in the drought-stressed ovaries when compared to the leaf meristem. This increase in ABA levels may have caused in the observed down-regulation of genes encoding both cell wall and soluble invertases (Figure 2.2). The repression of the soluble invertase may have compounded the effect of lowered cell wall invertase with respect to its metabolic function, since less substrate was available to associate with mitochondrial hexokinase to provide ADP to maintain coupled electron transport. As Figure 5.2 suggests, hexokinase as a signaling molecule might have sensed a decrease in available glucose and activated the SnRKI pathway. However, the same hexokinase is metabolically inactive because of low glucose levels. Under these substrate deficient conditions in the drought stressed ovaries, it is very likely that the ROS levels have increased to the extent where genes encoding ROS scavenging proteins were down-regulated in the mitochondrion and cytosol (Figure 4.2). The ABA action along with the repression of invertases, in drought stressed ovaries, may have resulted in a process very similar to ABA mediated acceleration of senescence, shown in leaves by others [10].This may be related to

the action of SnRK1, since DIN1, a SEN gene, and SnRK1 target genes, was up-regulated in the stressed ovaries. SnRK1 also may have inhibited AGPase expression, and a co-regulation mechanism to decreases in expression of other genes related to starch biosynthesis.

To conclude, the signature of gene expression observed in the drought stressed maize ovaries at 1DAP is in agreement with the activation of PCD, the halting of the cell cycle, impaired drought, and heightened carbon starvation signaling, mediated by ABA, cell wall and soluble invertases, and by increases in ROS to toxic levels due to mitochondrial malfunctioning and perhaps also in other subcellular locations. The data suggest that these events did not take place in the leaf meristem, where antioxidant defense mechanisms operated successfully, and the downstream changes associated with cell death and PCD did not occur. Interestingly, the gene expression data suggest that PI signaling may be impaired in both tissues, since PLC mRNAs levels were down-regulated by drought in both data sets.

Figure 6.1 shows the overview of drought stress effects on maize ovary tissue at 1DAP. The core figure is drawn according to Systems Biology Graphical Notation (SBGN) Activity Flow specifications (Level 1 Version 1) using Virginia Tech′s Beacon software. Shaded colors group glyphs with similar processes or events. The relationships, or arcs, express the nature of the influence (positive, negative, or unknown) between the glyphs and summarize this thesis findings along with support from the literature.

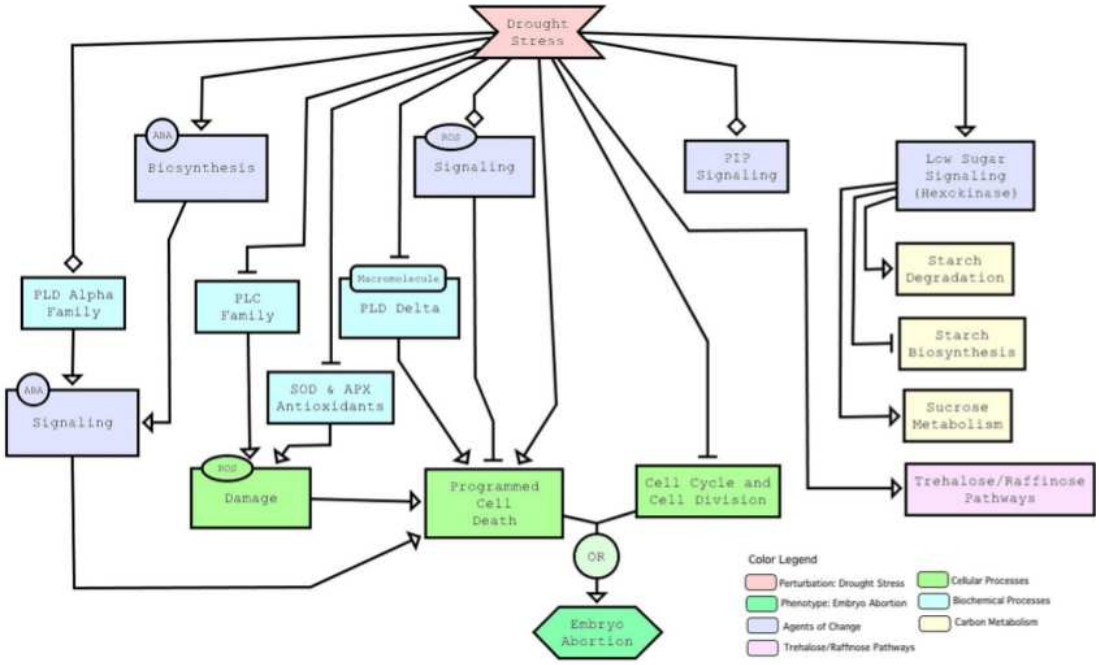


Figure 6.1: Overview of the working hypothesis

Bibliography

[1] Andersen MN, F Asch, Y Wu, C R Jensen, H Naested, V O Mogensen, and K E Koch. Soluble invertase expression is an early target of drought stress during the critical. abortion-sensitive phase of young ovary development in maize. *Plant Physiol*, 130:591–604, 2002.

[2] Feng HY, Wang ZM, Kong FN, Zhang MJ, and Zhou SL. Roles of carbohydrate supply and ethylene, polyamines in maize kernel set. *J Integr Plant Biol*, 53:388–398, 2011.

[3] Setter TL, Yan J, Warnburton M, Ribaut JM, Xu Y, Sawkins M, Buckler ES, Zhang Z, and Gore MA. Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. *J Exp Bot*, 62:701–16, 2011.

[4] Westgate ME and Boyer JS. Transpiration and growth-induced water potentials in maize. *Plant Physiol*, 74:882–889, 1984.

[5] Zinselmeier C, Jeong BR, and Boyer JS. Starch and the control of kernel number in maize at low water potential. *Plant Physiol*, 121:25–36, 1999.

[6] Cruz de Carvalho MH. Drought stress and reactive oxygen species: Production, scavenging and signaling. *Plant Signal Behav*, 3:156–165, 2008.

[7] Suzuki N, Koussevizky S, Mittler R, and Miller G. Ros ans redox signalling in the response of plant to abiotic stress. *Plant Cell Environ*, 35:259–270, 2012.

[8] Wobus U and Weber H. Sugars as signal molecules in plant seed development. *Biol Chem*, 380:937–944, 1999.

[9] Couee I, Sulmon C, Gouesbet G, and El Amrani A. Involvement of soluble sugars in reactive oxygen species balance and responses to oxidative stress in plants. *J Exp Bot*, 57:449–459, 2006.

[10] Ruan YL, Jin Y, Yang YJ, Li GJ, and Boyer JS. Sugar input, metabolism, and signaling mediated by invertase: roles in development, yield potential, and response to drought and heat. *Mol Plant*, 3:942–955, 2010.

[11] Bolouri-Maghaddam MR, Le Roy K, Xiang L, Rolland F, and Van den Ende W. Sugar signaling and antioxidant network connections in plant cells. *FEBS J*, 277:2022–2037, 2010.

[12] Matiolli CC, Tomaz JP, Duarte GT, Prado FM, Del Bem LE, Silveria AB, Gauer L, Correa LG, Drumond RD, Viana AJ, Di Mascio P, Meyer C, and Vincentz M. The arabidopsis bzip gene atbzip63 is a sensitive integrator of transient abscisic acid and glucose signals. *Plant Physiol*, 157:692–705, 2011.

[13] Lou Y, Gou JY, and Xue HW. Pip5k9, an arabidopsis phosphatidylinositol monophosphate kinase, interacts with a cytosolic invertase to negatively regulate sugar-mediated root growth. *Plant Cell*, 19:163–181, 2007.

[14] Ober ES, Setter TL, Madison JT, Thompson JF, and Shapiro PS. Influence of water deficit on maize endosperm development : enzyme activities and rna transcripts of starch and zein synthesis, abscisic acid, and cell division. *Plant Physiol*, 97:154–164, 1991.

[15] Setter TL and Parra R. Relationship of carbohydrate and abscisic acid levels to kernel set in maize under postpollination water deficit. *Crop Sci*, 50:980–988, 2010.

[16] Oliver SN, Dennis ES, and Dolferus R. Aba regulates apoplastic sugar transport and is a potential signal for cold-induced pollen sterility in rice. *Plant and Cell Physiology*, 48:1319–1330, 2007.

[17] Ji X, Dong B, Shiran B, Talbot MJ, Edlington JE, Hughes T, White RG, Gubler F, and Dolferus R. Control of abscisic acid catabolism and abscisic acid homeostasis is important for reproductive stage stress tolerance in cereals. *Plant Physiol*, 156:647–662, 2011.

[18] Yu LX and Setter TL. Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. *Plant Physiol*, 131:568–582, 2003.

[19] Boyer JS and McLaughkin JE. Functional reversion to identify controlling genes in multigenic responses: analysis of floral abortion. *J Exp Bot*, 58:267–277, 2007.

[20] Jain M and Khurana JP. Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice. *FEBS J*, 276:3148–3162, 2009.

[21] Zhuang Y, Ren G, Yue G, Li Z, Qu X, Hou G, Zhu Y, and Zhang J. Effects of water-deficit stress on the transcriptomes of developing immature ear and tassel in maize. *Plant Cell Rep*, 26:2137–2147, 2007.

[22] Poroyko V, Spollen WG, Hejlek LG, Hernandez AG, ME LeNoble, Nguyen G, Davis HT, Springer GK, Sharp RE, and Bohnert HJ. Comparing regional transcript profiles from maize primary roots under well-watered and low water potential conditions. *J Exp Bot*, 58:279–289, 2007.

[23] Zheng J, Zhao J, Tao Y, Wang J, Liu Y, Fu J, Gao P, Zhang J, Bai Y, and Wang G. Isolation and analysis of water stress induced genes in maize seedlings by subtractive pcr and cdna macroarray. *Plant Mol Bio*, 55:807–823, 2004.

[24] Tardieu F and Granier C. Quantitative analysis of cell division in leaves: methods, developmental patterns and effects of environmental conditions. *Plant Mol Bio*, 43:555–567, 2000.

[25] Setter TL and Flannigan BA. Water deficit inhibits cell division and expression of transcripts involved in cell proliferation and endoreduplication in maize endosperm. *J Exp Bot*, 52:1401–1408, 2001.

[26] Li H and Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25:1754–60, 2009.

[27] Trapnell C, Pachter L, and Salzeburg SL. Tophat: discovering splice junctions with rna-seq. *Bioinformatics*, 25:1105–1111, 2009.

[28] Langmead B, Trapnell C, Pop M, and Salzberg SL. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol*, 10:R25, 2009.

[29] Graber M, Grabherr MG, Guttman M, and Trapnell C. Computational methods for transcriptome annotation and quantification using rna-seq. *Nature*, 8:469–477, 2011.

[30] Lunter G and Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of illumina sequence reads. *Genome Res*, 21:936–39, 2010.

[31] Schnable PS. The b73 maize genome: Complexity, diversity, and dynamics. *Science*, 326:1112–1115, 2009.

[32] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Baren MLV, Salzburg SL, Wols BJ, and Pachter L. Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, 28:511–515, 2010.

[33] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res*, 25:3389–402, 1997.

[34] Symth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, 3:Article3, 2004.

[35] Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society*, 64:479–498, 2002.

[36] Jaccard P. Distribution de la flore alpine dans le bassin des drouces et dans quelques regions voisines. *Bulletin de la Socit Vaudoise des Sciences Naturelles*, 37:241–272, 1901.

[37] Kim SY and Volsky DJ. Page: parametric analysis of gene set enrichment. *BMC Bioinformatics*, 6:144, 2005.

[38] DeJager SM, Maughan S, Dewitte W, Scofield S, and Murrary JA. The developmental context of cell-cycle control in plants. *Semin Cell Dev Biol*, 16:385–396, 2005.

[39] DeVeylder L, Beeckman T, and Inze D. The ins and outs of the plant cell cycle. *Nat Rev Mol Cell Biol*, 8:655–665, 2007.

[40] Kitsios G and Doonam JH. Cyclin dependent protein kinases and stress responses in plants. *Plant Signal Behav*, 6:204–209, 2011.

[41] Sabelli PA and Larkins BA. The contribution of cell cycle regulation to endosperm development. *Sex Plant Reprod*, 22:207–219, 2009.

[42] Sabelli PA and Larkins BA. The development of endosperm in grasses. *Plant Physiol*, 149:14–26, 2009.

[43] Najera-Martinez M, Ramirez-Parra E, Vazquez-Ramos J, and Plasencia C, Gutierrez ans J. Cloning and molecular characterisation of the maize retinoblastoma gene (zm-rbr2). *Plant Science*, 175:685–693, 2008.

[44] Coelho CM, Dante RA, Sabelli PA, Sun Y, Dilkes BP, Gordon-Kamm WJ, and Larkins BA. Cyclin-dependent kinase inhibitors in maize endosperm and their potential role in endoreduplication. *Plant Physiol*, 138:2323–2336, 2005.

[45] Heitzeberg F, Chen IP, Hartung F, Orel N, Angelis KJ, and Puchta H. The rad17 homologue of arabidopsis is involved in the regulation of dna damage repair and homologous recombination. *Plant Journal*, 38:954–968, 2004.

[46] Roa H, Lang J, Culligan KM, Kellar M, Holec S, Cognat V, Montane MH, Houlne G, and Chaboute ME. Ribonucleotide reductase regulation in response to genotoxic stress in arabidopsis. *Plant Physiol*, 151:461–471, 2009.

[47] Li J, Harper LC, Golubovskaya I, Wang CR, Weber D, Meeley RB, McElver J, and Bowen B. Functional analysis of maize rad51 in meiosis and double-strand break repair. *Genetics*, 176:927–935, 2007.

[48] Yu HG, Muszynski MG, and DaveKelly R. The maize homologue of the cell cycle checkpoint protein mad2 reveals kinetochore substructure and contrasting mitotic and meiotic localization patterns. *J Cell Biol*, 145:425–435, 1999.

[49] Osborne MA, Schlenstedt G, Jinks T, and Silver PA. Nuf2, a spindle pole body-associated protein required for nuclear division in yeast. *J Cell Biol*, 125:853–866, 1994.

[50] Mane SP, Sioson C, Vasquez-Robinet ans AA, Heath LS, and Grene R. Early pldalpha-mediated events in response to progressive drought stress in arabidopsis: a transcriptome analysis. *J Exp Bot*, 58:241–252, 2007.

[51] Mane SP, Robinet CV, Ulanov A, Schafleitner R, Tincopa L, Gaudin A, Nomberto G, Alvardo C, Solis C, Bolivar LA, Blas R, Ortega O, Solis J, Panta A, Rivera C, Samolski I, Carbajulca DH, Bonierbale M, Pati A, Heath LS, Bohnert HJ, and Grene R. Molecular and physiological adaptation to prolonged drought stress in the leaves of two andean potato genotypes. *Functional Plant Biology*, 35:669–688, 2008.

[52] Vanderauwera S, Suzuki N, Miller G, Van de cotte B, Ravanat JL, Hegie A, Triantaphylides C, Shulaev V, Van Montagu MC, Van Breusegem F, and Mittler R. Extranuclear protection of chromosomal dna from oxidative stress. *Proc Natl Acad Sci U S A*, 108:1711–1716, 2011.

[53] Tsiatsiani L, Van Breusegem F, Gallois P, Zavialov A, Lam E, and Bozhkov PV. Metacaspases. *Cell Death Differ*, 18:1279–1288, 2011.

[54] Piffanelli P, Zhou F, Casais C, Orme J, Jarosch B, Schaffrath U, Collins NC, Panstruga R, and Schulze-Lefert P. The barley mlo modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. *Plant Physiol*, 129:1076–1085, 2002.

[55] Ishikawa T, Watanabe N, Nagano M, Kawai-Yamada M, and Lam E. Bax inhibitor-1: a highly conserved endoplasmic reticulum-resident cell death suppressor. *Cell Death Differ*, 18:1271–1278, 2011.

[56] Gillaspy GE. The cellular language of myo-inositol signaling. *New Phytol*, 192:823–839, 2011.

[57] Bargmann BO and Munnik T. The role of phospholipase d in plant stress responses. *Curr Opin Plant Biol*, 9:515–522, 2006.

[58] Xue HW, Chen X, and Mei Y. Function and regulation of phospholipid signalling in plants. *Biochem J*, 421:145–156, 2009.

[59] Hong Y, Zjang W, and Wang X. Phospholipase d and phosphatidic acid signalling in plant response to drought and salinity. *Plant Cell Environ*, 33:627–635, 2010.

[60] M. Li, Hong Y, and Wang X. Phospholipase d- and phosphatidic acid-mediated signaling in plants. *Biochim Biophys Acta*, 1791:927–935, 2009.

[61] Thimm O, Blasing O, Gibson Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, and Stitt M. Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J*, 37:914–39, 2004.

[62] Yan HB, Jiang HW, Pan XX, Li MR, Chen YP, and Wu GJ. The gene encoding starch synthase iic exists in maize and wheat. *Plant Science*, 176:51–57, 2009.

[63] Weigelt K, Kuster H, Rutten T, Fait A, Fernie AR, Miersch O, Wasternack C, Emery RJ, Desel C, Hosein F, Muller M, Saalbach I, and Weber H. 2009. *Plant Physiol*, 149:395–411, 2009.

[64] Bate NJ, Niu X, Wang Y, Reimann KS, and Helentjaris TG. An invertase inhibitor from maize localizes to the embryo surrounding region during early kernel development. *Plant Physiol*, 134:246–254, 2004.

[65] Chourey PS, Jain M, Li QB, and Carlson SJ. Genetic control of cell wall invertases in developing endosperm of maize. *Planta*, 223:159–167, 2006.

[66] Vilhar B, Kladnik A, Blejec A, Chourey PS, and Dermastia M. Cytometrical evidence that the loss of seed weight in the miniature1 seed mutant of maize is associated with reduced mitotic activity in the developing endosperm. *Plant Physiol*, 129:23–30, 2002.

[67] Cho YH, Yoo SD, and Sheen J. Regulatory functions of nuclear hexokinase1 complex in glucose signaling. *Plant Physiol*, 127:579–589, 2006.

[68] Camacho-Pereira J, Meyer LE, Machado LB, Oliveira MF, and Galina A. Reactive oxygen species production by potato tuber mitochondria is modulated by mitochondrially bound hexokinase activity. *Plant Physiol*, 149:1099–1110, 2009.

[69] Valluru R and Van den Ende W. Myo-inositol and beyond–emerging networks under stress. *Plant Science*, 181:387–400, 2011.

[70] Baena-Gonzalez E, Rolland F, Thevelein JM, and Sheen J. A central integrator of transcription networks in plant stress and energy signalling. *Nature*, 448:938–942, 2007.

[71] Xue GP, McIntyre CL, Glassop D, and Shorter R. Use of expression analysis to dissect alterations in carbohydrate metabolism in wheat leaves during drought stress. *Plant Mol Biol*, 67:197–214, 2008.

[72] Foyer CH and Shigeoka S. Understanding oxidative stress and antioxidant functions to enhance photosynthesis. *Plant Physiol*, 155:93–100, 2011.

[73] Knaupp M, Mishra KB, Nedbal L, and Heyer AG. Evidence for a role of raffinose in stabilizing photosystem ii during freeze-thaw cycles. *2011*, 234:477–486, Planta.

[74] Hanson J and Smeekens S. Sugar perception and signaling–an update. *Curr Opin Plant Biol*, 12:562–567, 2009.

[75] Cutler SR, Rodriguez PL, Finkelstein RR, and Abrams SR. Abscisic acid: emergence of a core signaling network. *Annu Rev Plant Biol*, 61:651–679, 2010.

[76] Fujita Y, Fujita M, Shinozaki K, and Yamaguchi-Shinozaki K. Aba-mediated transcriptional regulation in response to osmotic stress in plants. *J PLant Res*, 124:509–525, 2011.

[77] Schwartz SH, Tan BC, Gage DA, Zeevaart JA, and McCarty DR. Specific oxidative cleavage of carotenoids by vp14 of maize. *Science*, 276:1872–1874, 1997.

[78] Dolferus R, Ji X, and Richards RA. Abiotic stress and control of grain number in cereals. *Plant Science*, 181:331–341, 2011.

[79] Frey A, Audran C, Martin E, Sotta B, and Martin-Poll A. Engineering seed dormancy by the modification of zeaxanthin epoxidase gene expression. *Plant Mol Biol*, 39:1267–1274, 1999.

[80] Seo M, Aoki H, Koiwai H, Kamiya Y, Nambara E, and Koshiba T. Comparative studies on the arabidopsis aldehyde oxidase (aao) gene family revealed a major role of aao3 in aba biosynthesis in seeds. *Plant and Cell Physiology*, 45:1694–1703, 2004.

[81] Finkelstein RR and Lynch TJ. The arabidopsis abscisic acid response gene abi5 encodes a basic leucine zipper transcription factor. *Plant Cell*, 12:599–609, 2000.

[82] Finkelstein RR, Gampala SS, and Rock CD. Abscisic acid signaling in seeds and seedlings. *Plant Cell*, 14:S15–45, 2002.