DROUGHT-INDUCED EPIGENETIC MODULATION

AND TRANSCRIPTIONAL VARIATION

OF WINTER WHEAT

By

ALEXANDER LIM

Bachelor of Science in Biochemistry

Oklahoma State University

Stillwater, OK

2015

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 2020

DROUGHT-INDUCED EPIGENETIC MODULATION

AND TRANSCRIPTIONAL VARIATION

OF WINTER WHEAT

Thesis Approved:

Dr. Charles Chen

Thesis Adviser

Dr. Patricia Canaan

Dr. Peter Hoyt

Dr. Lan Zhu

# ACKNOWLEDGEMENTS

First, I would like to acknowledge and thank my advisor Dr. Charles Chen for his support, guidance, and invaluable instructions throughout my Masters journey. Secondly, I would like to thank my committee members, Dr. Peter Hoyt, Dr. Patricia Canaan, and Dr. Lan Zhu for their helpful criticism and their expertise. Immense thanks is deserved to my fellow co-worker Bryan Naidenov for his advice, comradeship, and motivation. I would also like to thank my fellow former co-workers in the Translational Genomics Laboratory, Dr. Karyn Willyerd, Dr. Shuzhen Sun, and Chi-Ping Liao. Also, I would like to thank the Biochemistry and Molecular Biology Department and Faculty as well as my fellow cohorts for making my educational journey that much more enjoyable and smoother. Lastly, I would like to thank and acknowledge all my friends and family who believed I could I do this and supported me throughout my ups and downs.

This research and opportunity would not have been possible without the funding from the Oklahoma Center for the Advancement of Science and Technology, the Oklahoma Wheat Commission and the National Science Foundation.

Name: ALEXANDER LIM

Date of Degree: DECEMBER, 2020

Title of Study: DROUGHT-INDUCED EPIGENETIC MODULATION AND TRANSCRIPTIONAL VARIATION OF WINTER WHEAT

Major Field: BIOCHEMISTRY AND MOLECULAR BIOLOGY

Abstract:

The worsening climate uncertainty has exacerbated environmental stress, and the accompanying extreme weather disasters have damaged crop production across the globe. The yield production of wheat, one of the world's most in-demand crops, has endured substantial yield losses due to large-scale drought as a result. This thesis aims to identify the adaptive genetic changes and the epigenomic regulation machinery employed by winter wheat amid water stress similar to the 2014 nation-wide drought. A 50% water reduction was imposed immediately before the transition from vegetative growth to the reproduction stage for two hard red winter wheat genotypes, Duster and DH169, the latter being a derived progeny from a Duster x Billings DH population. As a derived progeny, DH169 demonstrated higher yielding capacity under the 2014 drought compared with Duster. In this study, DH169 was considered drought 'tolerant' and Duster by comparison was drought 'avoidant'. Transcriptomics and reduced representation bisulfite sequencing (RRBS) were used to quantify the genome-wide alteration in gene expression and methylation induced by the imposed water deficit.

This study showed that, under drought, 719 more differentially expressed genes were detected in DH169 compared with the Duster genotype. The majority of the differentially expressed genes were associated with response to oxidative stress and bract morphogenesis. Overall, Duster exhibited more significant methylation changes than DH169 and a greater extent of methylation in drought than control, whereas the methylation of DH169 was found higher under the well-watered condition. Finally, gene body hypermethylation was found associated with down-regulation in DH169; however, the positive association of up-regulation of gene expression and gene-body hypomethylation can only be seen in Duster.

These findings suggest that under a water deficit, the drought-tolerant DH169 genotype undergoes significant transcriptional changes but, less so epigenetically. As a drought 'avoidant' winter wheat, Duster, on the other hand, demonstrated a much more extensive genome-wide epigenetic modification compared with the variation identified at the genetic level. To summarize, my study reveals various genetic adaptation mechanisms employed by two closely related winter wheat genotypes. The whole-genome recombination events during the hybridization process might have disrupted the epigenomic regulatory machinery, requesting DH169 to respond to the imposed water deficit with a more expensive transcriptional variation. It is, therefore, the interrelationships of all the components in the central dogma of molecular biology, rather than any individual process, that modulate a genome's changes in fitness when stressed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I


INTRODUCTION


Wheat (*Triticum aestivum*), as the largest primary commodity, currently provides 19% of the total calories consumed according to the Food and Agriculture Organization of the United Nations (Reeves et al., 2016). The demand of global wheat production has increased from 700 million tonnes from 200 million in 1961 to 2020 (Reeves et al., 2016). With the global population estimated to reach 9.1 billion by 2050, the FAO projects the global food production needs to be increased by 70% to accommodate the population increase, and the need for animal feed.

However, in the last decade, environmental stress and extreme weather events have caused large-scale damage in crop production world-wide (Lesk et al., 2016, Coumou et al., 2012, Barriopedro et al., 2011). For example, from 1980 to 2015, a 21% and 40% global yield reduction in wheat and maize, respectively, was attributed to the global drought (Daryanto et al., 2016); and in Lesk et al., 2016, the continued water deprivation has added another 7% in yield losses in all cereal crops including wheat. Furthermore, changes in weather and environmental stress are often associated with the rise in the global mean temperature (Lindsey and Dahlman, 2018). With each degree Celsius increased in global temperature, it is projected that an additional ~6.0% loss will occur in global wheat production (Zhao et al., 2017).

In the southern Great Plains, winter wheat is the number one agriculture system for states like Oklahoma, occupying 4.5 million acres of farmland and generating 500 million dollars in revenue (USDA, 2019). However, the 2020 Drought Monitor (Svoboda et al., 2002)

revealed that 69.82% of Oklahoma land has been exposed to intense drought in the last 10 years. By June 2014, the United States was strangled by multi-year drought, and wheat production had decreased to 51 million bushels, half of the 2013 production and had not been seen since 1957 (Silva, 2014). Since then, the United States has experienced the most widespread drought in history, with over 45% of the lower 48 states currently under exceptional dry conditions (Climate Prediction Center). Understanding and developing drought genetic resistance for winter wheat is thus vital to secure the United States' rural economy and long-term agricultural sustainability.

Drought stress can induce a series of physiological and biochemical changes all functioning as fine-tuned machinery in order for a plant to adapt to the changes in the growing condition. Global gene expression is a well understood biological mechanism critical for adaptation (López-Maury et al., 2008, Lasky et al., 2014). Transcriptomics, a powerful tool to uncover overall gene expression, utilizes deep-sequencing technologies to provide a quantitative measurement of contracting genetic responses to environmental variation, such as water shortage (Wang et al., 2009). Transcriptomic analysis of plant responses to water stress has successfully revealed up-regulation of transcripts involved in carbohydrate metabolism, oxidative responses and hormone metabolism, while most of the down-regulated transcripts are found to be associated with photosynthesis (Kakumanu et al., 2012; Liu et al., 2015; Bowman et al., 2013; Hübner et al., 2015; Fracasso et al., 2016; Chen et al., 2016). Amongst these studies, signaling molecules (abscisic acid), secondary messengers (reactive oxygen species, ROS), and metabolic processes (proline metabolism), are considered crucial for tolerance to stress in crop plants (Varoquaux et al., 2019, Tang et al., 2013, Johnson et al., 2014, Dudziak et al., 2019). For example, abscisic acid (ABA) is a well-categorized plant hormone found in plants that are experiencing abiotic stress. ABA acts as a signaling molecule capable of inducing expression changes in genes that influence membrane stability and water retention (Skriver and Mundy, 1990). These findings confirm that drought could elicit changes in the expression level. Furthermore, the potential role for

transcription factors and chaperones to modulate physiological responses in combined stressed was studied in Johnson et al., 2014. In wheat, by examining the water shortage induced ROS production, Dudziak et al. 2019 showed that intervarietal substitution lines (ICSCLs) of chromosomes 3B, 5A, 7B, and 7D had significant impacts on global gene expression for all tested genes (MAPK3, MAPK6, CAT, APX, GPX, PC5S PC5R), substantially influencing the genetic function controlling grain production and quality parameters.

Although great progress has been made to decipher variation of crop performance under water stress using both genomic and transcriptomic approaches, little is known about how the environmental cues modify these genomic variants. Recent research has demonstrated that changes in tissue-specific and genotype dependent DNA methylation patterns can be induced in relation to abiotic stress (Liang et al., 2014 Wang et al., 2003). For instance, research of faba beans (Vicia faba) showed that the DNA methylation of the root tissues changed from 23.43% to 19.80% under drought stress for the drought-tolerant genotype and increased from 16.53% to 17.19% in the



Figure 1. Preliminary Grain Yield Trials for Duster and DH169 in 2014 and 2016 - density plot of the Duster and DH169 population and their associated grain yield in the drought year of 2014 and normal conditions of 2016. Marked are the average grain yield.

drought sensitive genotype (Abid et al., 2017). Additionally, studying the changes of DNA methylation on gene expression in rice cultivars with contrasting drought responses, Garg et al.,
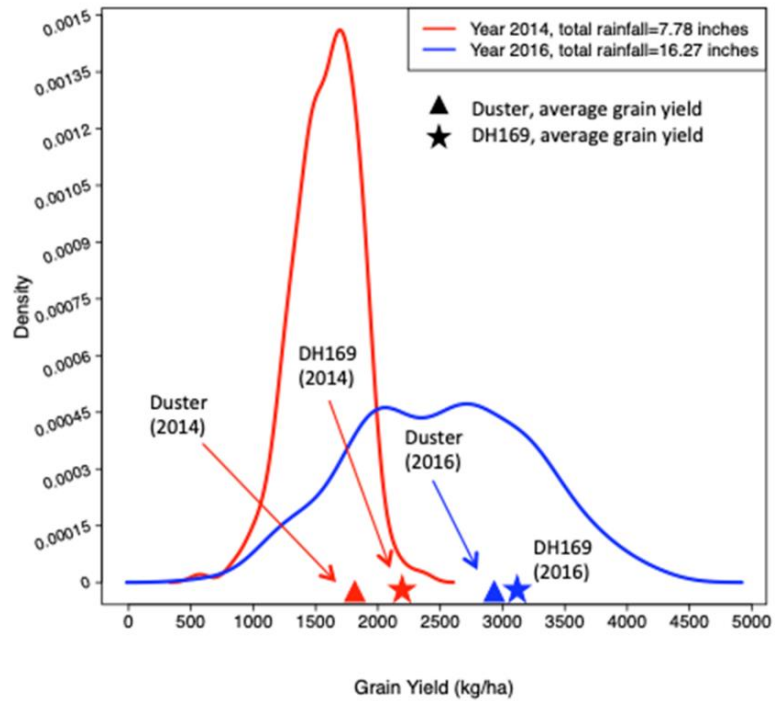
2015 and Zheng et al., 2013 provided evidence of differentially methylated regions (DMRs) in

proximal transposable elements (TE) regulating gene expression through methylation and

demethylation. In Li et al., 2012, the role of DNA methylation in transcriptional termination

regions was examined; however, the results suggested that DNA methylation difference can only

account for limited gene expression between cultivated and wild rice. These studies focused on

the early seedling developmental stages of rice. Two studies, one in wheat (Kaur et al., 2018) and

the other in rye grass (Tang et al., 2014), revealed that overall DNA demethylation was increased

in a drought-tolerant genotype compared with a drought sensitive one under water stress. More

importantly, in Kaur et al., 2018 using methylation sensitive amplified polymorphism (MSAP),

highlighted the variable changes in DNA methylation between different growth stages under

drought stress. And, the detected methylation polymorphism ranged from 74.8% at anthesis to

88.9% at tillering (Kaur et al., 2018). When comparing genotypes, more were consistently found

in the drought-sensitive wheat in all growth stages, however, these changes were widely variable

ranging from a 200% difference to a 4% difference (Kaur et al., 2018,) suggesting that the

relationship between drought stress and DNA methylation can be dependent on growth stage.

However, the role of methylation and demethylation on drought-associated gene expression in the

late growth stages when plants are in transition from vegetative to reproductive development, has

not yet been well characterized in wheat.

By 2016, the development of new winter wheat genotypes by the Oklahoma Agriculture

Experiment Station had finished preliminary yield trials since 2014. The yield trials of a doubled

haploid (DH) population derived by an intercross of Duster and Billings discovered that under

regular annual rainfall, the yield for Duster and DH169 performed slightly above the average of

the DH population. On the contrary, under the intense drought stress of 2014, as seen in Figure 1,

grain yield in Duster stayed at the overall population mean while DH169 performed at the top 7

percent. In order to compare the wheat genotypes, DH169 with its above average performance in

grain yield under drought stress was considered to be drought 'tolerant'. By contrast Duster with its average performance under drought was considered to be more drought 'avoidant' (Osmolovskaya et al., 2018).

To better understand the genetics and its regulatory mechanisms underlying the observed difference in yielding capacity, *I hypothesize that, as a drought avoidant genotype, Duster prevented water stress damage via a global down-regulation in gene expression by methylation, thus the insignificant change in overall grain yield production in 2014. As a drought tolerant genotype, DH169 responded to water stress by genetically up-regulating expression on transcripts associated with ROS scavenging and metabolic processes, for example, the abscisic acid-responsive element binding protein 1 (AREB 1) subfamily that acts as a positive regulator under drought stress.*

The overarching goal of this research seeks to delineate the interplay of methylation and gene expression crucial to genetic adaptation to drought stress. The water reduction experiment conducted in this study imitated the 2014 drought in Oklahoma. Successful demonstration of these approaches will have a far-reaching impact on the global agricultural sustainability that has been challenged by the unprecedented climate variability, beyond the crop improvement for Oklahoma's wheat production.

**The research objectives of this thesis are:**

(1) To identify the differential expression profiles between a drought-avoidant (Duster) and a drought-tolerant (DH169) winter wheat genotypes to better understand their molecular genetic adaptation to drought stress.

(2) To assess and quantify the changes in the epigenetic methylome associated with Duster and DH169; further, to elucidate the contributions of methylation and demethylation under drought stress, at the whole genome and sub-genome scales.

(3) To illustrate the relationship between the changes in the methylome and the drought-stress associated differential gene expression.

CHAPTER II

LITERATURE REVIEW

## 2.1 GENE EXPRESSION IN DROUGHT

Advancements in technology and techniques have propagated new methods capable of investigating crop performance under stress. In particular, transcriptomics has been a powerful technique in measuring transcript abundance between tens to hundreds of thousands of genes, thus emulating an expression profile. For example, the gene expression profile under abiotic stress such as drought has revealed up-regulation of transcripts involved in carbohydrate metabolism, oxidative responses and hormone metabolism, while most of the down-regulated transcripts were found to be associated with photosynthesis (Kakumanu et al., 2012; Liu et al., 2015; Bowman et al., 2013; Hübner et al., 2015; Fracasso et al., 2016; Chen et al., 2016). More importantly identified gene expression can be influenced by signaling molecules (abscisic acid), secondary messengers (reactive oxygen species, ROS), and metabolic processes (proline metabolism), which are considered crucial for tolerance to stress in crop plants (Dudziak et al., 2019, Varoquaux et al., 2019, Tang et al., 2013, Johnson et al., 2014).

For instance, abscisic acid (ABA) is a well-categorized plant hormone found in plants that are experiencing abiotic stress and acts as a signaling molecule that induces expression changes in genes governing membrane stability and water retention (Skriver and Mundy, 1990). Amongst these, abscisic acid-responsive element binding protein (AREB1) is a transcription factor that is able to bind to the promoter region of ABA-inducible genes but,

Fujita et al., 2008, demonstrated that the expression of intact AREB1 gene was insufficient to effect expression of downstream genes and was dependent on the presence of ABA. By creating a constitutively activated form of AREB1 they were able to demonstrate ABA hypersensitivity and enhanced drought tolerance without the presence of exogenous ABA (Fujita et al., 2008). Also, in a study in rice, overexpression of PYL5 a functional ABA receptor that regulates ABA-dependent gene expression was found to induce many stress-responsive genes under drought conditions (Kim et al., 2014). In summary, gene expression associated with ABA signaling is an important mechanism in response to drought stress. However, though expression of ABA has been considered effective in improving drought tolerance, it showed a negative impact on plant height and severely reduced total seed yield in rice (Kim et al., 2014).

As mentioned previously, Dudziak et al. 2019 revealed that intervarietal substitution lines (ICSCLs) of chromosomes 3B, 5A, 7B, and 7D had significant impacts on global gene expression for all tested genes (MAPK3, MAPK6, CAT, APX, GPX, PC5S PC5R), substantially influencing the genetic function controlling grain production and quality parameters. The genes tested were involved in proline biosynthesis (P5CS and P5CR) and protection of the cell against toxic ROS (Ascorbate peroxidase (APX) and guaicol peroxidase (GPX)) (Das et al., 2014). In rice, OsLG3 overexpression was shown to significantly improve the tolerance of rice plants when under drought (Xiong et al., 2018). Using digital gene expression analysis, the study was able to analyze global gene expression between a transgenic overexpressed OsLG3 line versus an RNA interference line (RNAi), that suppressed OsLG3, and found that the overexpressed line demonstrated up-regulation in ROS scavenging related genes (APX1, APX2, APX4, APX6, APX8, CATB, POD1, SODcc1 and FeSOD) (Xiong et al., 2018). Consequently, the same genes were also found to be down-regulated in the RNAi lines. In summary Xiong et al., 2018 concluded that OsLG3 had a role in controlling ROS homeostasis which was confirmed by RT-

qPCR. Overall, research in drought response in plants reveal that ROS reactive gene expression is another factor to consider when dealing with drought stress.

Additionally, a study in maize inbred lines utilized transcriptome analysis to identify 555 and 2,558 genes that were detected to specifically respond to drought in a tolerant and sensitive line, respectively (Zhang et al., 2017). Zhang et al., 2017 were able to further identify 13 genes that showed opposite expression patterns between the drought tolerant and sensitive lines under moderate and severe drought stress. Another comparative analysis revealed that in sugarcane leaf during the elongation stage 1,501 genes were found to be differentially expressed under drought stress (Li et al., 2016). In Li et al., 2016, out of these 1,501 differentially expressed genes (DEGs), 821 were found to be up-regulated and 680 were down-regulated (Li et al., 2016). Through functional annotation and pathway analysis the DEGS were found to be associated with biosynthesis of secondary metabolites, ribosomes and carbon metabolism, suggesting that they may play a role in the water-deficit tolerance. (Li et al., 2016). In another study in soybean leaf tissue, Chen et al., 2016 were able to identify 289 transcription factors out of 2,724 DEGs under drought stress that demonstrated motifs known to be involved with the stress tolerance mechanism. In the study, the transcription factors were basic helix-loop helix (bHLH), ethylene response factors (ERFs), myeloblastosis (MYB), no apical meristem (NAC), and WRKY amino acid motif (WRKY) type major families, which are known to be involved in the mechanism of stress tolerance. Furthermore, ABA related transcript factors such as the aforementioned MYB, and NAC were up-regulated in drought conditions and have been extensively studied (Chen et al., 2016, Nakashima et al., 2012, Dubos et al., 2010). In conclusion, current research suggests using differential gene expression is useful in identifying candidate genes and transcription factors associated with the drought stress response that may aid in improving drought tolerance in plants.

## 2.2 DNA METHYLATION IN DROUGHT

In 1925, a study by Johnson and Coghill, reported the first early discoveries of nucleotide modification, specifically 5-methyl-cytosine. Not until the early 1940s and 1950s did they establish a methodology to isolate 5-methylcytosine (Hotchkiss et al., 1948). Discovery of the cytosine methylation mechanism has laid the foundation for studies on epigenetic regulation. By the 1970s and 1980s many studies began focusing on the impact of methylation on gene expression (Mandel et al., 1979, McGhee and Ginder, 1979, Shen et al., 2013, Razin and Cedar, 1991). Regions that are susceptible to methylation are dinucleotide repeats rich in cytosine such as CpG islands. Methylation of CpG rich regions located near the transcription start sites of genes (Liang et al., 2014), within gene bodies (Bewick et al., 2014) and in the intergenic regions (Suzuki and Bird, 2008) can cause gene inactivation and activation. In order to ascertain these sites of methylation that can regulate gene expression, several methodologies were created, these can be categorized into three main categories: 5-methylcytosine (5mC) immunoprecipitation methods, restriction enzyme usage, or sodium bisulfite modification (Marzese and Hoon, 2015). Methods that consist of isolation of 5mC DNA or proteins can be targeted using specific antibodies. Due to direct extraction no data processing is required for conversion methods however, this methodology is highly sensitive to the quality and potential of cross-reactivity of 5mC antibodies (Marzese and Hoon, 2015). Methylation-sensitive restriction enzymes methods boast high-throughput for relative methylation using microarray hybridization (Irizarry et al., 2008). For example, Irizarry et al., 2008 were able to identify novel DMRs that were identified in colon cancer with 90% specificity. Though methylation microarray is limited based on the fragmentation sites, incomplete digestion and lack single CpG resolution (Marzese and Hoon, 2015). Lastly, methods that implement sodium bisulfite conversion allow generation of genome-wide maps of 5mC at single-base resolution. Specifically, whole-genome bisulfite sequencing and reduced representation bisulfite sequencing are able to generate genome-wide information of the

methylation status amongst millions of CpG sites at single base resolution. Instead of the restriction on incomplete chemical reaction, this bisulfite sequencing methodology is highly computationally demanding and dependent on the efficiency of bisulfite conversion.

In a *Populous trichocharpa* study, Liang et al., 2014 revealed that DNA methylation in 100-2,000 base pair upstream of the transcription start site paired with gene body methylation was positively associated with gene expression. Specifically, the study was able to identify gene-body methylated genes that had significantly higher expression levels when compared with gene body-unmethylated genes (Liang et al., 2014). Moreover, methylation within the transcription start site (TSS), transcriptional termination regions (TTR) and downstream regions were found to have negative correlations with gene expression levels (Liang et al., 2014). Further, research in soybean, through whole-genome bisulfite sequencing, has shown that genome-wide DNA methylation levels can undergo dynamic changes within seed maturation (An et al., 2017). These methylation changes demonstrated an average of 66% CpG, 45% CHG and 9% CHH contexts, was methylated in cotyledons, and that CHH methylation levels changed from 6% to 11% between the early stage and late stage of seed development (An et al., 2017). And, transcribed genes were two-fold more likely to be differentially methylated than non-transcribed genes (An et al., 2017). An et al., 2017, identified 40, 66, and 2,136 genes containing differentially methylated regions (DMR) with negative correlation between their expression and methylation in the CpG, CHG, and CHH contexts, respectively. To demonstrate the diverseness of methylation patterns, evidence in gymnosperms such as *Selaginella moellendorffii* showed a lack of DNA methylation in TSS and gene bodies (Bartels et al., 2018, Feng et al., 2010, Zemach et al., 2010).

Utilizing DNA methylation-sensitive amplified polymorphisms Wang et al., 2011, discovered that drought-induced genome-wide DNA methylation accounted for ~12.1% of the methylation differences in the rice genome. They highlighted three major properties of DNA methylation patterns caused by drought-induction: genotypic specificity, reversibility after

recovery, and developmental and tissue specificity (Wang et al., 2011). Similar patterns of

drought-induced methylation with respect to drought stress tolerance can be seen in faba bean

(Abid et al., 2017). The findings in Abid et al., 2017 suggested, a greater number of methylation

in leaves than in roots highlighting the tissue specificity of DNA methylation under drought

stress. The study also reported a decrease in DNA methylation in both the drought-tolerant

genotype and drought sensitive genotype with a reduction of 6.81% and 3.07% respectively.

Tissue specific methylation was also documented, the root tissue of the drought-tolerant genotype

diminished in DNA methylation by 3.63% whereas the drought-sensitive genotype increased by

0.66% under drought stress (Abid et al., 2017). In conclusion, analysis of differentially

methylated regions with respect to their expression found five high homology putative proteins

(LOX, CDPK, ABC, GH and PEPC), suggesting their important to the drought-tolerant genotype

under drought stress. A study on perennial ryegrass demonstrated demethylation under drought

treatment with a reduction in global DNA methylation by 10.28% (Tang et al., 2014). Out of 652

sites amplified for cytosine methylation detection, 61.08% were found to have experienced

methylation/demethylation changes when under drought stress, the remaining sites underwent no

change. These studies implemented methylation sensitive amplification polymorphism as their

primary tool to investigate methylation under drought. Due to the size and complexity of plant

genomes, it remains challenging for genome-wide methylation investigation in response to

drought stress.

CHAPTER III


MATERIALS AND METHODS

## 3.1 WINTER WHEAT

From 2012 to 2015 the hard red winter wheat cultivar Duster was the number one wheat variety in Oklahoma and continues to be in the top two (Reg. No. CV-1065, PI 644016) (Edwards et al., 2012). Duster was created by the Oklahoma Agriculture Experiment Station (OAES) and the USDA-ARS (Edwards et al., 2012). Duster has remained as a top variety due to several desirable qualities such as above average grain yield, disease resistance and insect tolerance. For continued improvements, DH169 was a drought tolerant progeny derived directly from the Duster x Billings doubled-haploid population (Poudel, 2016).

## 3.2 DROUGHT EXPERIMENT

Seeds produced from a 2016 field trial in Stillwater, OK were sown and germinated in flats under stable greenhouse conditions (22°C day, 18°C night; 16-hour light cycle). Three weeks after germination, seedlings were transplanted into 5L tree pots (12.7x12.7x30.5cm) in a soil mixture of sand, vermiculite, and peat (4:2:1) with the addition of a standard slow release fertilizer (Osmocote). Plants were allowed to adapt post-transplant for one week prior to vernalization in a cold room (4°C; 10-hour light cycle). After a five-week vernalization period, plants were returned to greenhouse conditions where they remained until full maturity.

Water was withheld from the treatment group at Feekes 10 (in boot) to mimic the drought period experienced in the 2014 field season. Water reduction treatment was conducted for a period of 6 days while control pots continued with regularly scheduled watering. Daily water loss was tracked gravimetrically. Leaf samples were collected from control and treatment plants after the 6-day treatment. After treatment, pots returned to regularly scheduled watering through maturity. Seed heads were collected for analysis of grain yield production-related phenotypes.

## 3.3 RNA SEQUENCING OF WINTER WHEAT

The two winter wheat genotypes tested in the drought experiment and selected for RNA sequencing was Duster and DH169. For RNA extraction, flag leaves were sampled from Duster and DH169, immediately submerged in RNAlater (Invitrogen), and stored at -20°C after saturation at 4°C overnight.

Total RNA was extracted from flag leaf tissue of four biological replicates per genotype for each experimental condition using Trizol (Invitrogen) in accordance with the manufacturer's instructions. Samples were sent to Oklahoma Medical Research Foundation Sequencing Facility (Oklahoma City, OK). RNA library preparation was performed using TruSeq Stranded Total RNA with Ribo-Zero Plant to remove overwhelming amounts of Rubisco and sequenced on the Illumina HiSeq3000 PE150.

## 3.4 RNA-SEQ ALIGNMENT FOR TRANSCRIPT ASSEMBLY

Preparation of the differential expression analysis began with mapping of the RNA-seq reads for both Duster and DH169 to the Chinese spring wheat (CSW) v2.0 reference genome (Appels et al., 2018). The RNA-seq alignment was performed using hisat2.2.1 (Kim et al., 2015) with the following parameters -p 32 --mp 1,0 --sp 3,0 --pen-noncansplice 20 --dta_cufflinks.

1. The –p switch designates the number of threads to be used for parallel processing.
2. The –mp switch sets the maximum and minimum mismatch penalties.

3. The –sp switch sets the maximum and minimum penalties for soft-clipping per base.
4. The –pen-noncansplice switch sets the penalty for each pair of canonical splice sites.
5. The –dta_cufflinks switch adds a necessary column of information of the alignment that cufflinks requires for transcript assembly.

Alignment using hisat2.2.1 of the respective sample yielded a sequence alignment map (SAM) file that was sorted and converted to a binary alignment map (BAM) file for space efficiency.

## 3.5 TRANSCRIPT ASSEMBLY USING CUFFLINKS

Cufflinks, a reference-guided transcript assembler, assembled Duster and DH169 transcripts (Trapnell et al., 2012). A transcript assembly is required for each sample and biological replicate. The default parameters were kept for transcript assembly except for specifying the number of threads in which case was 32.

Cuffmerge (Trapnell et al., 2012) with default parameters, consolidated a master transcriptome using all the individual transcript assemblies. It automatically filters artifacts and merges isoforms and genes that are shared, thus maximizing the quality of the final transcriptome assembly (Trapnell et al., 2012). Transcript and gene abundance was calculated as fragments per kilobase of transcript per million mapped reads (FPKM). This value was calculated for both the drought and control samples as $FPKM_y$ and $FPKM_x$ respectively. The first filtering step applied by cufflinks prior to differential expression analysis was a test evaluation that determines if a transcripts abundance calculation has one of the following classifications: OK, NOTEST, LOWDATA, HIDATA or FAIL; details are included in the cufflinks documentation (Trapnell et al., 2012). Only the transcripts and abundance calculations with a test status of OK were selected for further downstream analysis. Details of the likelihood function and model used for abundance calculation are found in the cufflinks documentation (Trapnell et al., 2012). Additionally, in order to decrease computational load, cuffquant is able to perform abundance calculation and store the results in a binary file called CXB (Trapnell et al., 2012).

Cuffdiff, takes the resulting .cxb binary files from cuffquant as input. Cuffdiff produces files including FPKM tracking files, count tracking files, read group tracking files and a tab delimited file that shows the results of the differential expressions for isoforms and genes (Trapnell et al., 2012). To determine if a transcript was up-regulated or down-regulated cuffdiff outputs a log 2 transformation of the fold change between drought and control. The fold change was represented by the drought FPKM expression divided by the control FPKM expression value. The transcripts were filtered for a log 2 fold change with a p-value and FDR cut-off of 0.05, those that passed this filtering were considered significant (Trapnell et al., 2012). Visualization and validation of the differentially expressed transcripts were performed using Cummerbund (Goff et al., 2013). A volcano plot was created to compare the drought and control samples given both DH169 and Duster with the -log10(pvalue) over the log2(fold change). As mentioned previously, the cut-off value was 0.05.

Gene Ontology (GO) enrichment of the significant differentially expressed transcripts revealed biological significance by classification into three categories: biological processes, cellular components and molecular function. Prior to GO enrichment, manual extraction of the .fasta sequences for all the transcripts was required. The transcript DNA sequences were extracted from CSW v2.0 using the positional information found in the gene expression cuffdiff output file. The extracted transcripts were then re-aligned to the CSW wheat gene annotation v1.1 using BLAST. The BLAST results were then filtered for three separate metrics: e-value, sequence identity and query coverage. The blast filtering identified the unique gene mapping of the CSW v2.0 sequence with the CSW v1.1 high confidence (HC) gene annotation. Due to BLAST matching several genes additional filtering was required.

This filtering was performed knowing that CSW v2.0 and the associated v1.1 annotation would share the same chromosomal information. Therefore, if an aligned transcript identified more than one v1.1 gene, only the genes that shared the same chromosome information were

selected. In the case that the alignment identified more than one gene in the same chromosome then another filtering step was performed. Genes of the same chromosome were selected based on whichever was closer position wise to the transcript, thus curating the final gene set. The gene names selected from the BLAST results for the significant transcripts were then used as input in AgriGO, a GO-enrichment analysis (Du et al., 2010). To identify the associated GO terms, AgriGO incorporates singular enrichment analysis (SEA). As documented by AgriGO, SEA provides a means to identify enriched GO terms given a gene list. This gene list is compared to a query list of the overall genes from the selected organism. To aggregate the ontology terms REViGO is used to semantically associate like terms and their p-values (Supek et al., 2011).

## 3.6 DNA SAMPLE PREPARATION FOR REDUCED REPRESENTATION BISULFITE SEQUENCING

As reported in Table 1, a total of 6 wheat ground leaf samples were prepared for reduced representation bisulfite sequencing (RRBS) analysis. Each respective genotype sample was represented by their condition, control or drought, and their replicate number. For example, Dus_C1 represented the Duster genotype (Dus) followed by the control condition (C) and replicate number (1).

The individuals selected for RRBS were isolated from the same plant source as the RNA-seq samples. All of the 6 samples had their DNA extracted from the ground leaf using a DNeasy Plant Mini kit (QIAGEN). Each tissue sample was chemically lysed and cleaned of RNA using RNase digestion. All sample precipitates including cell debris, proteins and polysaccharides were filtered and removed from the samples by homogenization through centrifugation with a QIAshredder spin column (QIAGEN). The lysates were loaded onto a DNeasy Plant mini spin column and spun down. The silicia membrane in the spin column bound the DNA while allowing other contaminates to pass through. Wash steps were performed to further purify the DNA bound

to the membrane and aid in the further removal of any remaining contaminants and enzyme inhibitors. After washing several times, the cleaned DNA was eluted with a low-salt buffer. All extracted DNA samples were then saved and a quality check was performed to determine DNA concentration using a Fluostar Optimate plated reader (BMG Labtech) with the Quant-IT Picogreen dsDNA assay kit (ThermoFisher Scientific) on 480/520 nm. Furthermore, to identify any signs of DNA degradation a 1% E-gel EX agarose gel was performed (ThermoFisher Scientific).

Table 1. Wheat Sample ID and Associated Treatments

| Sample ID | Treatment |
|-----------|-----------|
| Dus_C1 | Control |
| Dus_D1 | Drought |
| DH169_C1 | Control |
| DH169_C2 | Control |
| DH169_D1 | Drought |
| DH169_D2 | Drought |

## 3.7 DNA LIBRARY PREPARATION AND SEQUENCING FOR REDUCED REPRESENTATION BISULFITE SEQUENCING

Library preparation of the RRBS samples were performed using the Premium RRBS kit (Diagenode) (https://www.diagenode.com/files/products/kits/Premium_RRBS_kit_manual.pdf). Approximately 100-130 ng of DNA was isolated from each sample. Each sample was then enzymatically digested by the restriction enzyme MspI at 37°C for 12 hours. End preparation of the DNA samples were performed and unique adaptors for each sample was added, adaptor ligation was performed using DNA ligase. DNA size selection was performed on the adaptor-ligated DNA fragments using Agencourt AMPure XP beads (Beckman Coulter). Several wash steps using the DNA was eluted in resuspension buffer. A small sample of the eluted DNA was subjected to qPCR using 2X KAPA HiFi HotStart ReadyMix (Kapa Biosystems) for

quantification and pooling of the 6 samples. The pooling was performed at two set parameters: the Ct value and adaptor ID of each sample. Another DNA cleanup using AMPure XP beads was performed following pooling to reduce sample volume. Bisulfite treatment was performed and the bisulfite-converted DNA was eluted in BS elution buffer twice. Additionally, part of the bisulfite converted library was used in qPCR for determination of the optimal cycle number for enrichment. 2X MethylTaq Plus Master Mix was used for the final PCR amplification as well as a final bead cleanup using AMPure XP beads. Quality of the final DNA libraries was assessed by an Agilent 2100 High Sensitivity DNA chip. The concentration was determined by performing qPCR on the samples using a dilution of PhiX index3 as the standard. DNA sequencing was performed by Nxt-DX using the Illumina HiSeq 4000 for 50bp paired end reads.

### 3.8 DNA METHYLATION ALIGNMENT AND CALLING

Bismark was used to call methylation and non-methylated cytosines (unmethylation) (Krueger et al., 2011). Bismark performs alignment by using Bowtie as its base aligner and proceeds to align the bisulfite treated reads to the reference genome. Bismark performs two types of nucleotide conversions prior to alignment, the first conversion employs C-to-T read conversion and genome conversion (Krueger et al., 2011). The second conversion utilizes G-to-A conversion. The converted reads undergo parallel alignment between the converted bisulfite reads and the reference genome with both the forward and reverse strands. Out of the four different alignments the best alignment is selected and the methylation call determined both of which are reported in a SAM alignment file.

Duster was a single replicate sample. Each sample had a higher sequencing yield compared with DH169. Dusters sequencing output was on average 1.36 times greater than the control samples in DH169. As for drought, Duster had nearly twice as many reads compared with DH169. Sequencing and replicate differences can cause skew towards Duster. In order to

maintain equivalency in read count and replicates between Duster and DH169, subsampling of the Duster reads was required. With subsampling, two replicates were created in control and drought for Duster, thus matching DH169. Additionally, for each replicate the read count was maintained at 25,000,000 reads. Subsampling was performed randomly by assigning an integer index value for each read and randomly sampling integers for read selection. This was performed for each replicate creating two randomized groups for Duster control and drought and allowing a fairer comparison between Duster and DH169.

## 3.9 DIFFERENTIAL METHYLATION ANALYSIS

The alignment SAM file from methylation calling was further processed in an R package called methylKit for differential methylation analysis (Akalin et al., 2012). MethylKit is able to filter, merge, and select the differentially methylated cytosines as well as identify islands of methylation. Differentially methylated cytosines are methylated or de-methylated cytosine nucleotides with respect to drought versus control. Each alignment file was loaded for processing in methylKit using methread, a function designed to interpret the Bismark alignment file. MethylKit's methread automatically performs read filtering and requires a minimum coverage of 10 reads. This is performed on all replicates and samples as well for each methylation type.

A uniform methylation table was created to report all cytosines that were found across all replicates before calculating differential methylation. The unified methylation table records the position and number of methylations and demethylations per replicate. Differential methylation was calculated using the calculateDiffMeth() function provided by methylKit. CalculateDiffMeth() determines p-values given each methylated site using a statistical significance test. Additionally, p-values were corrected to q-values, using the sliding linear model (SLIM) method to correct for the multiple hypothesis testing problem, details are in the documentation (Akalin et al., 2012). For each coinciding q-value, the percent methylation

difference was calculated as well. To ensure quality, differential methylation was filtered for a percent methylation difference larger than 25% and a q-value less than 0.01, thus reducing the likelihood of Type I errors (Akalin et al., 2012). This filtering step allows for categorization of the methylation into two groups hyper- or hypo- methylated. Hypermethylation indicates that the cytosine is methylated in drought and hypomethylated is demethylated in drought.

Furthermore, to investigate methylation localization, the differentially methylated cytosines (DMCs) were grouped into methylation islands (DMI) or regions (DMR); these were identified by methylKit. A methylation island was determined by windows of 1,000 base pairs in length with a step-size of 1,000 base pairs. As mentioned previously, calculateDiffMeth() by methylKit also determines the p-value given the methylations within a tile. The methylation percent difference was also calculated for each given island.

Differentially methylated islands (DMI) and their proximity to genetic features such as the gene body and the 5' upstream region are tabulated by position in the genome. The 5' upstream region was estimated to be 2,000 base pairs upstream of the transcription start site (TSS). The start and end positions of the methylation islands and the genetic features were used to determine overlap between the two. Given every instance of the overlap they were categorized into two groups, islands associated with the gene body and islands associated with the 5' upstream region. These differentially expressed genes (DEGs) that overlapped with DMI's whether it be the 5' upstream region and/or the gene body, were termed as differentially methylated genes (DMG). Every instance of the group was then plotted by the methylation percent difference over the expression value for the respective gene. A regression model was fitted through the points to determine if there was a predictive relationship between the methylation percent difference and the associated gene expression.

CHAPTER IV

RESULTS

## 4.1 DROUGHT EXPERIMENT

Figure 2 confirms that a drought response occurred in phenotype for the wheat samples where a continuous 2-day time stamp over a six-day period of drought reveals an increasingly gravimetric loss under said conditions. By the sixth day, drought-treated (DT) samples suffered 8.7 % gravimetric loss compared with the well-watered (WW) samples that only loss 1.1 % (Figure 2A). In the case of relative water content, the WW samples saw no change after 6 days, however, the DT samples decreased by 14% (Figure 2B). Additionally, in photosynthesis (Pn), transpiration (Tr) and instantaneous water use efficiency (IWUE) between the WW and DT conditions, revealed that the WW samples underwent greater photosynthesis and transpiration while the DT samples had greater IWUE (Table 2). Significance was calculated using a t-test.

Table 2. Mean Differences in Photosynthesis, Transpiration and IWUE between well-watered (WW) and drought-treated (DT) samples

| Measurement | Mean | Standard Deviation |
|:---:|:---:|:---:|
| *Pn(WW-DT) | 2.15 | 0.51 |
| Tr(WW-DT) | 2.66 | 0.88 |
| IWUE(WW-DT) | -1.57 | 1.13 |

*Pn – Photosynthesis; Tr – Transcription; IWUE – Instantaneous water use efficiency.

Figure 2. Assessment of Mean Phenotypes - Pot water status was assessed by gravimetric loss (A). The mean loss of drought pots (red) was significantly greater ($p < 0.001$) than that of well-watered pots (blue) which maintained relatively stable pot weights throughout the treatment period. Relative water content (RWC) was used as a measurement of plant water status (B). No difference in mean RWC was detected in well-watered plants (blue) before (0d) and after (6d) treatment. Mean RWC of drought pots (red) shows significant difference ($p < 0.01$) after treatment period (6d).
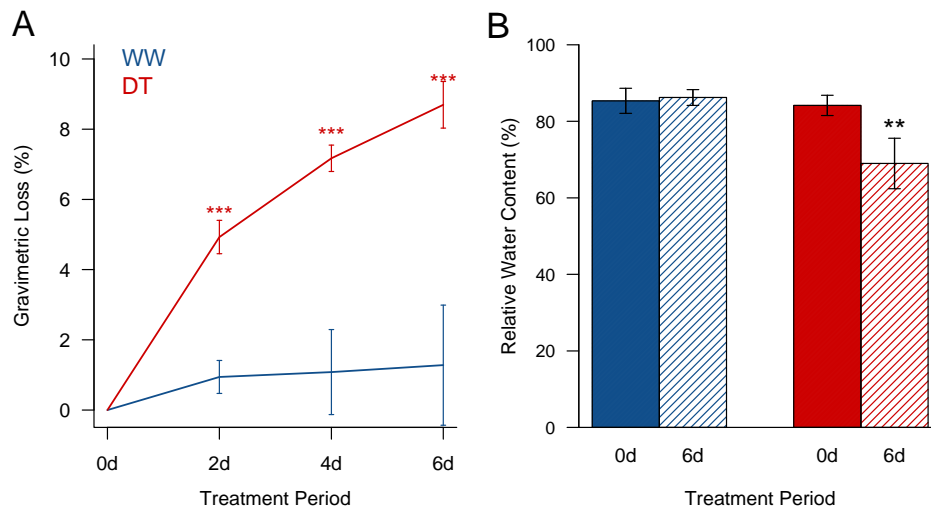


## 4.2 RNA-SEQ ALIGNMENT TO CSW V2.0

The Illumina RNA-seq yielded ~84 million reads across the 4 control replicates in Duster and ~85 million reads across the 4 drought replicates. The average read count and standard deviation for the control replicates were 21,543,176 reads with a standard deviation of 1,526,729.30. While for drought, the average was 21,970,155 reads with a standard deviation of 2,224,428.69. Furthermore, an average of 64.35 % of the control reads were uniquely aligned once, while drought had an average of 66.95 % of uniquely mapped reads. Additionally, in Table 3 of the Duster RNA-seq read alignments using hisat2.2.1, the mapping rate between both the control and drought reads had an average of 96.78% and 96.44%, respectively. Additional break

23

down of the mapping rates, alignments and percent of total reads analyzed between each replicate and condition are identified in Table 3.

DH169 had similar sequencing yield in control at ~92 million sequencing reads. However, in drought, total sequencing yield was increased at ~103 million sequencing reads. The average read count and standard deviation for the control samples of DH169 was 23,264,284 with a standard deviation of 3,057,913.88. For the drought reads, the average across the 4 replicates was 26,013,748 with a standard deviation of 3,249,961.95. An average of 61.94% of the reads uniquely aligned for the control replicates and 67.74% for drought. This demonstrated that a majority of the read pairs have a single unique alignment. Subsequently, the mapping rate between both the control and drought reads were 96.66% and 95.70%, respectively. Overall, DH169 and Duster had similar mapping efficiency.

Given the total average of reads in drought and control for Duster and DH169, DH169 exhibited a higher number of reads compared with Duster with a difference of 3 million reads on average between drought and control (Table 3).

Table 3. Control and Drought Sample Read Alignment Using hisat 2.2.1 for Duster and DH169

| Duster | Control 1 | Control 2 | Control 3 | Control 4 |
|---|---|---|---|---|
| Total Reads Analyzed | 20,589,001 | 22,725,088 | 22,955,645 | 19,902,971 |
| Read pairs aligned 0 times | 3,930,134 (19.09%) | 4,047,945 (17.81%) | 4,133,695 (18.01%) | 3,657,728 (18.38%) |
| Read pairs aligned 1 time | 13,056,256 (63.41%) | 15,483,842 (68.14%) | 14,661,879 (63.87%) | 12,331,199 (61.96%) |
| Read pairs aligned >1 time | 3,602,611 (17.50%) | 3,193,301 (14.05%) | 4,160,071 (18.12%) | 3,914,044 (19.67%) |
| Overall Alignment Rate (%) | 96.17 | 96.37 | 96.60 | 96.62 |
| | Drought 1 | Drought 2 | Drought 3 | Drought 4 |
| Total Reads Analyzed | 22,375,352 | 24,863,918 | 19,663,167 | 20,978,184 |
| Read pairs aligned 0 times | 4,365,901 (19.51%) | 3,832,579 (15.41%) | 3,832,579 (15.41%) | 3,527,294 (16.81%) |
| Read pairs aligned 1 time | 14,263,598 (63.75%) | 16,843,735 (67.74%) | 16,843,735 (67.74%) | 14,382,710 (68.56%) |
| Read pairs aligned >1 time | 3,745,853 (16.74%) | 4,187,604 (16.84%) | 4,187,604 (16.84%) | 3,068,180 (14.63%) |
| Overall Alignment Rate (%) | 96.22 | 97.41 | 96.82 | 96.67 |

| DH169 | Control 1 | Control 2 | Control 3 | Control 4 |
|---|---|---|---|---|
| Total Reads Analyzed | 21,329,088 | 20,414,999 | 27,196,569 | 24,116,480 |
| Read pairs aligned 0 times | 4,412,909 (20.69%)* | 4,290,934 (21.02%) | 6,592,401 (24.24%) | 5,986,004 (24.82%) |
| Read pairs aligned 1 time | 13,592,599 (63.73%) | 12,699,945 (62.21%) | 16,818,914 (61.84%) | 14,462,868 (59.97%) |
| Read pairs aligned >1 time | 3,323,580 (15.58%) | 3,424,120 (16.77%) | 3,785,254 (13.92%) | 3,667,608 (15.21%) |
| Overall Alignment Rate (%) | 97.74 | 95.23 | 97.30 | 96.37 |
| | Drought 1 | Drought 2 | Drought 3 | Drought 4 |
| Total Reads Analyzed | 25,667,561 | 30,006,874 | 26,308,269 | 22,072,287 |
| Read pairs aligned 0 times | 3,428,295 (13.36%) | 6,287,506 (20.95%) | 3,837,649 (14.59%) | 3,876,037 (17.56%) |
| Read pairs aligned 1 time | 17,270,468 (67.29%) | 19,724,251 (65.73%) | 18,208,636 (69.21%) | 15,165,396 (68.71%) |
| Read pairs aligned >1 time | 4,968,798 (19.36%) | 3,995,117 (13.31%) | 4,261,984 (16.20%) | 3,030,854 (13.73%) |
| Overall Alignment Rate (%) | 95.60 | 96.65 | 94.19 | 96.35 |

*Percentage of total reads analyzed

## 4.3 TRANSCRIPT ASSEMBLY

The assembled transcripts for Duster had an N50 of 2,118 with the largest transcript being 15,651 base pairs long. The average GC content across all the transcripts was 55.87%. In Table 4, the transcript assembly of the Duster reads to CSW resulted in a total of 62,240 genes which were composed of 99,218 transcripts and 599,342 exons. For Duster, the exon average for a gene was 9.63 exons with a standard deviation of 16.46. One gene XLOC_004272 was identified to have a large number of exons, 444. Contrastingly, the median exon count for the genes was 4. Additionally, the gene was comprised of two transcripts and found in chromosome 1B in Duster. The 444 exons were located within the base pairs 412,301,221 to 412,321,517. The gene length was found to be only 20,296 base pairs long and was revealed to not be significantly differentially expressed between drought and control. The FPKM value for the genes was ~5.0 in control and drought for 4.65.

Overall, DH169 had similar results to Duster, the assembled transcripts had an N50 of 2,102 with a maximum transcript size of 17,808 base pairs. The GC content for the transcripts in DH169 was higher than Duster at 56.27%. Transcript assembly was greater in DH169 in all three categories at 63,781 genes, 101,280 transcripts and 611,178 exons. The average gene was composed of 9.58 exons with a standard deviation of 16.38. The median exon count for the DH169 genes was 4, the same as Duster. The gene XLOC_004350 was found to have the largest number of exons at 445. This was comparable to the 444 found in Duster. Furthermore, XLOC_004350 was composed of two transcripts and identified in chromosome 1B. It was located between the base pairs of 412,301,152 to 412,321,554. The length of the gene was 20,402 base pairs with an FPKM control value of 4.12 and an FPKM drought value of 4.09, it is also not significantly differentially expressed. Additionally, in both Duster and DH169 all genes reported a minimum of 1 exon.

To conclude, general transcript assembly was similar between Duster and DH169. Only slight differences were found. To reiterate, DH169 had more assembled transcripts compared with Duster and the most expressed gene found in both wheat genotypes was only found to be slightly less expressed in DH169 with a difference of 0.88 in control and 0.56 in drought.

Table 4. Feature counts for Genes, Exons and Transcripts of the Cufflinks Transcript Assembly for Duster and DH169

| Genotype | Genes | Transcripts | Exons |
|----------|-------|-------------|-------|
| Duster | 62,240 | 99,218 | 599,342 |
| DH169 | 63,781 | 101,280 | 611,178 |

**4.4 DIFFERENTIAL EXPRESSION ANALYSIS BETWEEN DUSTER AND DH169**

In Duster, regarding differences in overall expression of the abundance calculation in FPKM, the average FPKM values for control filtered by the test status OK was 13.35 whereas the average FPKM for drought was lower at 11.84. The median calculation for Duster in control was 3.01 and in drought 2.82. The gene that was expressed the most for control in Duster was XLOC_029321 with an FPKM value of 3,350.93. This gene was located in chromosome 4A with a base pair position at 569,721,583 to 569,847,379 with a large gene length of 125,796 bp. For drought, the gene that was most expressed was XLOC_051466, found in chromosome 6D. Its expression value was comparatively higher at 4,444.07 and was located at 492,523,358 to 492,558,731 and possessed a gene length of 35,373 bp.

As for the drought-tolerant DH169, the average FPKM values for control and drought were higher at 36.17 and 31.75, respectively. The median calculation for FPKM expression was 2.75 for control and 2.69 for drought. The gene XLOC_014624 was the most expressed in control

and had an FPKM value of 1,287,230. It was found in chromosome 2B with a base pair location at 50,508,505 to 50,508,655 and a small gene length of 150 base pairs. For drought, the maximum FPKM value was high at 1,086,010 and reports the same gene as control. As both of these maximum expression was over 280 times greater than the second highest expression value, and the gene length reported was quite small at 150 bp they were considered outliers and excluded. Therefore, the second highest values for control and drought were 4,539.89 and 3,665.19 respectively. For control, the gene id was XLOC_000003 found in chromosome 1A with a base pair location of 347,224 to 352,064, the gene length was 4,840 bp. As for drought conditions the gene id was XLOC_052772 found in chromosome 6D with a base pair location of 492,523,379 to 492,559,401 and a gene length of 36,022. One important note which will be addressed in detail later is that both XLOC_000003 and XLOC_052772 found in DH169 were considered significantly differentially expressed genes whereas the two maximum expressed genes found in Duster were not.

Generally, Duster and DH169 report similar levels of overall expression with only a small difference, Duster being higher, of 0.26 and 0.12 for control and drought, respectively. Out of the genes that were most expressed, DH169 and Duster both shared the same gene under drought conditions found in chromosome 6D, however, it is only differentially expressed in DH169. They both identified different genes for the highest expressed transcripts for control conditions.

Dusters A sub-genome median for the control and drought samples were 3.10 and 2.90, reflecting little change between whole genome and A sub-genome expression. For the B sub-genome, the median was comparatively lower at 2.88 for control and 2.71 for drought. In contrast for sub-genome D they increased to 3.04 in control and 2.86 in drought. In comparison, for DH169 the median for the control and drought samples were 2.83 and 2.77 for the A sub-genome, 2.64 and 2.59 for the B sub-genome, and finally, 2.78 and 2.73 for the D sub-genome. Overall the

28

sub-genome expression values for both control and drought were decreased in DH169 compared with Duster.

For Duster, in the A sub-genome, the maximum control expression value was 3,350.93. The gene name associated with this expression is XLOC_029321 and the corresponding chromosome was chromosome 4A. Its position was 569,721,583 to 569,847,379 and had a gene length of 125,796 bp. In drought condition the maximum expression value is 2,897.27 reporting the same transcript as the control expression. In the B sub-genome, the maximum expression values for control and drought was 2,652.46 and 2,231.25, respectively. The gene associated with both expression values was XLOC_030820 located in chromosome 4B with a gene location at 187,901,783 to 187,909,081 and a gene length of 7,298 bp. Lastly, in the D sub-genome the maximum expression values for control and drought were 2,917.84 and 4,444.07. The gene associated with maximum control expression was XLOC_051466 found in chromosome 6D with its location at 492,523,358 to 492,558,731 and a gene length of 35,373. Similarly, XLOC_051466 is the same gene for the maximum drought expression.

As for DH169, the maximum expression values for the control and drought in the A sub-genome was 4,539.89 and 2,907.28. The gene associated with the maximum control was XLOC_000003, found in chromosome 1A with a base pair location of 347,224 to 352,064, the gene length was 4,840 bp. For drought, the gene id was XLOC_030052 located in chromosome 4A with a base pair location at 448,403,947 to 448,411,264 and a gene length of 7,317 base pairs. Continuing with the B sub-genome, the maximum control and drought expression values are 3,374.65 and 2,823.96, respectively. Both values are associated with the same gene, the id being XLOC_031588 located in chromosome 4B and its location at 187,906,503 to 187,909,026 with a gene length of 2,523. Finally, for the D sub-genome, the maximum control and drought expression values were 3,530.05 and 3,665.19. The gene associated for control has an id of XLOC_044814 and is located in chromosome 5D with a base pair location at 269,111,562 to

269,113,253 and a gene length of 1,691 bp. As for drought, the gene id was identified as XLOC_052772 located in chromosome 6D with a base pair location at 492,523,379 to 492,559,401 and a gene length of 36,022.

To retain important zero counts in the FPKM calculation, prior to log transformation, a constant value of 0.001 was added to all FPKM values. The original FPKM values are log normal distributed and to assume normalization the FPKM values were then log transformed and a two-tailed t-test was performed to determine if the mean difference between control and drought were significantly different. For Duster, the resulting p-value was 3.41E-13 and for DH169, the p-value calculated was 0.036.

Furthermore, a volcano plot was created for both Duster and DH169 as shown in Figure 3. These were created to visually identify a proportion of statistically significant genes with large log 2 fold changes. The red coloration in Figure 3 delineates the significant genes from the non-significant genes with a FDR cut-off of 0.05.

When looking at a genome-wide perspective for the differentially expressed genes (DEG), overall, DH169 shows a greater number of DEG's at 2,429 versus Duster at 1,710 and when averaged per chromosomes, Duster was 77.73 and DH169 was 110.41. Duster had a median FPKM value of 5.28 for control and 3.39 for drought. In contrast, DH169 was lower at 4.21 for control and higher at 4.23 for drought. Out of the DEGs in Duster, 523 were up-regulated in drought and 1,160 DEGs were down-regulated in drought. As for DH169 1,187 DEG's were up-regulated in drought and 1,180 were down-regulated in drought. Therefore, the difference between Duster and DH169 are largely attributed in up-regulated genes, 664 more in DH169.

To better understand the relationship between DEGs and their genome location the DEGs were re-categorized into their respective sub-genomes. When evaluating the DEGs across sub-genomes both Duster and DH169 demonstrated a higher average of differentially expressed genes

in the D sub-genome with a value of 607 for Duster and 844 in DH169. As evident in Figure 4, the relative proportion of DEGs in the D sub-genome was 35.50% for Duster and 34.75% for DH169, both of which represent the sub-genome with the highest percentage of DEGs. However, to determine if this difference of DEGs between sub-genomes was statistically significant a one-way ANOVA was performed for both Duster and DH169. The p-values, 0.37 for Duster and 0.42, showed that there was no significant difference between the sub-genomes with respect to the number of DEGs.

The sub-genome breakdown of the expression values was also evaluated. For Duster the median value for control and drought expression was 5.23 and 3.26 for the A sub-genome. The B sub-genome had median values of 4.76 and 3.02 for control and drought. Lastly, for the D sub-genome the median expression was 5.34 and 3.66 for control and drought. On the other hand, DH169 had median values of 4.12 and 3.96 for control and drought in the A sub-genome. For the B sub-genome, the median control and drought expression values were 3.97 and 4.20, respectively. Finally, for the D sub-genome the median expression values for control and drought were 4.18 and 4.54. It is apparent across all sub-genome expression levels that DH169 was greater in drought compared with Duster. By contrast, Duster displayed higher levels of expression in control compared with DH169.

For Duster given sub-genome A there are 155 up-regulated DEG's and 359 down-regulated DEG's. In sub-genome B, 176 DEG's were up-regulated and 360 down-regulated DEG's. Finally, in the D sub-genome 187 DEG's were up-regulated and 413 DEG's were down-regulated. In DH169 the A sub-genome has 390 up-regulated DEG's and 392 down-regulated DEG's. In the B sub-genome there are 367 up-regulated genes and 355 down-regulated genes. Lastly, in the D sub-genome there are 425 up-regulated genes and 405 down-regulated genes.

To further understand the importance of the significant DEG's, a breakdown between all the chromosomes was reported in Table 5 for both Duster and DH169. Out of the individual chromosomes of Duster, the greatest number of DEG's were found to be in chromosomes 2D and 5D. Chromosomes 2D and 5D were at 116 and 110, respectively, both of which are greater than the average by nearly 40%. In comparison, DH169 shows the greatest number of DEG's in chromosomes 7D and 1A at 155 and 142. The number of DEG's in 7D and 1A are approximately ~40% and 30% greater than the average number of DEG's across all chromosomes. With respect to the CSW v1.1 annotation, the chromosomes associated with the highest number of DEG's in DH169, 7D and 1A, have less genes at 9,743 compared with Dusters, 2D and 5D, with 11,428.

Figure 3. Volcano Plots of Duster and DH169 Transcripts – two volcano plots generated for the Duster and DH169 wheat genotypes that highlight the transcripts that are considered significant with an FDR cut-off of 0.05.
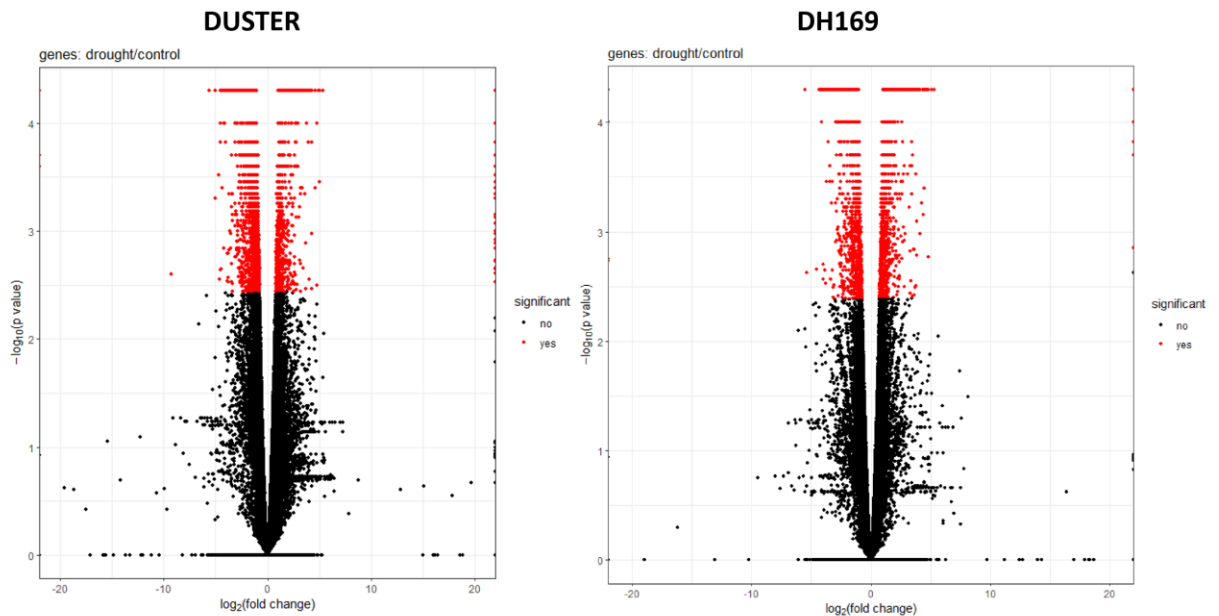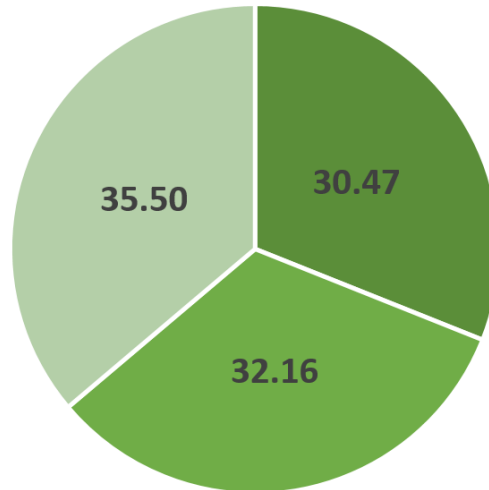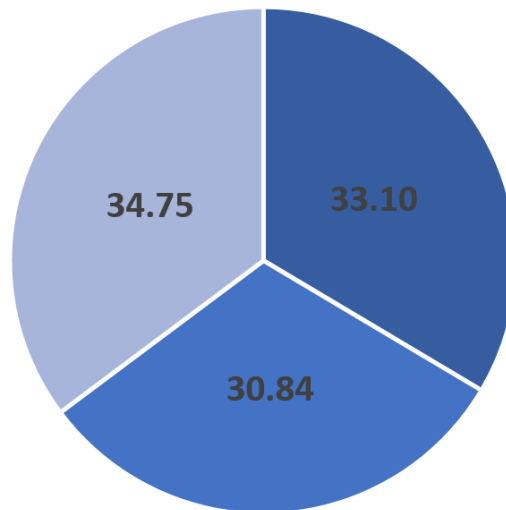
Figure 4. Percentage of DE Transcripts Per Sub-genome between Duster and DH169 – Pie charts describing the DEG percentage per sub-genome of Duster and DH169

## DE Percentage Per Subgenome Duster



- A-Subgenome
- B-Subgenome
- D-Subgenome

## DE Percentage Per Subgenome DH169



- A-Subgenome
- B-Subgenome
- D-Subgenome

Table 5. Number of the Significant Differentially Expressed Genes between Duster and DH169 – Chromosome breakdown of the total number of significantly differentially expressed genes identified for Duster and DH169.

|  | Duster to CSW | DH169 to CSW |
|---|---|---|
| **Chr1A** | 92 | 142 |
| **Chr1B** | 85 | 96 |
| **Chr1D** | 75 | 107 |
| **Chr2A** | 88 | 131 |
| **Chr2B** | 79 | 124 |
| **Chr2D** | 116 | 130 |
| **Chr3A** | 84 | 115 |
| **Chr3B** | 82 | 122 |
| **Chr3D** | 93 | 128 |
| **Chr4A** | 51 | 106 |
| **Chr4B** | 68 | 99 |
| **Chr4D** | 57 | 93 |
| **Chr5A** | 84 | 123 |
| **Chr5B** | 87 | 120 |
| **Chr5D** | 110 | 130 |
| **Chr6A** | 57 | 85 |
| **Chr6B** | 71 | 83 |
| **Chr6D** | 69 | 101 |
| **Chr7A** | 65 | 102 |
| **Chr7B** | 78 | 105 |
| **Chr7D** | 87 | 155 |
| **ChrUn** | 32 | 32 |
| **Total** | 1,710 | 2,429 |

**4.5 ANNOTATION AND GENE ONTOLOGY ENRICHMENT OF DUSTER AND DH169**

Based on the BLAST results and gene filtering, Duster only had 1,041 DEGs that had associated gene names. As for DH169 there was a total 1,636 DEGs that were mapped to an associated gene. Out of the 1,041 Duster DEGs that aligned, only 659 uniquely mapped to a gene name leaving 489 that mapped to multiple genes. As for DH169, out of the 1,636 DEGs that aligned to a gene name 878 mapped uniquely once and 758 mapped multiple times.

AgriGO was able to identify 1,808 GO terms for the Duster genes. When filtered for an FDR value greater than 0.05 only 643 of the 1,808 GO terms were considered significant. Out of the 643 significant GO terms 477 were classified with biological processes, 41 with cellular components, and 125 with biological processes. The most enriched GO term for biological processes was response to oxygen-containing compound (GO:1901700) with a FDR value of 2.4E-24. The best for molecular function was chitinase activity (GO:0004568) with a FDR value of 2.9E-07. Lastly, for cellular component, was intrinsic component of plasma membrane (GO:0031226) with an FDR value of 5.1E-11.

As for DH169 there were a greater number of GO terms identified at 2,073. However, only 514 were considered significant. Of the identified 514 significant GO terms 377 were identified with biological process, 24 with cellular components and 112 with molecular function. This implies that a majority of the DEGs found in DH169 were grouped functionally.

The most significant GO term identified in the biological processes for DH169 was response to oxygen-containing compounds (GO:1901700) with a FDR value of 1.5E-44. As for molecular functions the most overrepresented GO term was identified to be glutathione transferase activity (GO:0004364) with a FDR value of 1.8E-07. Finally, for cellular components the number one term was identified to be apoplast (GO:0048046) with a FDR value of 1.6E-09.

All three top-hits for DH169 are associated with stress response, more specifically oxidative stress.

The most enriched ontology terms were identified by dividing the number of genes found within a GO term over the total number of genes used as input. For Duster the most enriched GO term in molecular function representing 10.69% of the genes was transmembrane transporter activity (Figure 5). For biological processes the most significant was response to stimulus with approximately 60.69% of the genes (Figure 5). Finally, in cellular components 49.78% of the genes were associated with membrane (Figure 5).

Given DH169, the most significant GO term for molecular function with 15.08% of the genes was oxidoreductase activity (Figure 6). For biological processes the best result was response to stimulus with 60.91% of the genes (Figure 6). Lastly, for cellular components was membrane with 44.55% of the genes (Figure 6). The 9 enriched GO ontology terms as revealed in Figure 5 and 6, demonstrated similar function compared to the semantic associated results in Figure 7.

Figure 7's depiction of the semantic aggregation of the functional enrichment with consideration of p-value revealed that for Duster, the largest group for biological processes was associated with response to oxygen-containing compound followed by hexose transmembrane transport and third monocarboxylic acid biosynthesis. For molecular function the primary group was substrate-specific transmembrane transporter activity, then UDP-glucosyltransferase activity and third linoleate 13S-lipoxygenase activity. For the last ontology group cellular components, the top group was nucleosome followed by intrinsic component of plasma membrane and lastly intrinsic component of membrane

As for DH169, the largest group associated with biological processes was response to oxygen-containing compounds followed by secondary metabolism and thirdly, regulation of

biological quality. For molecular function the largest group was glutathione S-conjugate-exporting ATPase activity then glutathione transferase activity and third iron ion binding. Lastly, in cellular components the representative group was intrinsic component of plasma membrane which is followed by photosystem II antenna complex then plant-type vacuole.

Given the results, both Duster and DH169 identified with strong evidence across GO term enrichment, semantic association (Figure 7), and genes per GO term (Figure 5 and 6), that response to stimulus was the greatest identified mechanism. Interestingly, when reviewing the secondary categories, DH169, highlights oxidative stress as its primary target for the DEGs. In contrast, Duster distinguished more terms relating to plant membrane and its internal components.

Figure 5. Most Significant Ontology Terms for Duster - Bar graph representation of the 9 most enriched terms from the differentially expressed genes, in relation to the gene percentage associated with the ontological term.
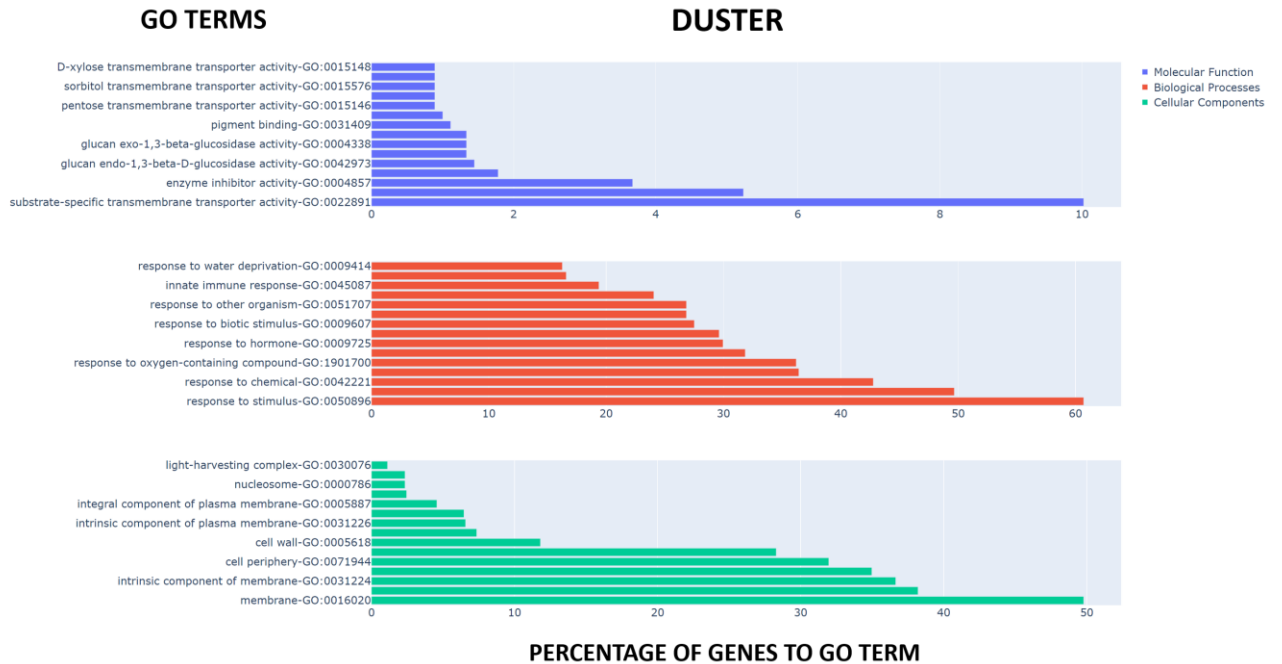
Figure 6. Most Significant Ontology Terms for DH169 - Bar graph representation of the 9 most enriched ontology terms of the differentially expressed genes found in DH169 in relation to the gene percentage associated with the ontological term.

Figure 7. Gene Ontology Semantic Relationships Between DH169 and Duster - gene ontology clustering of biological process, molecular function and cellular component terms through semantic comparison. Sizes of the squares are showed as the level of the statistical significance, larger indicates a more significant p-value. Colored blocks are associated terms that are semantically similar.

**4.6 METHYLATION ALIGNMENT FOR DUSTER AND DH169**

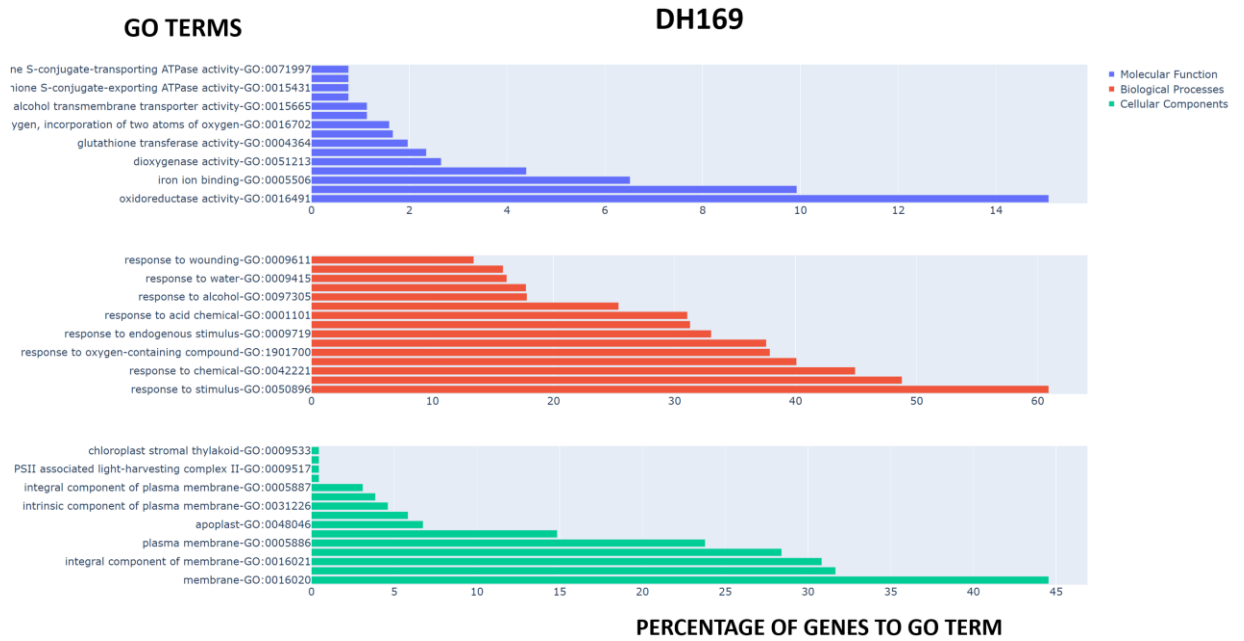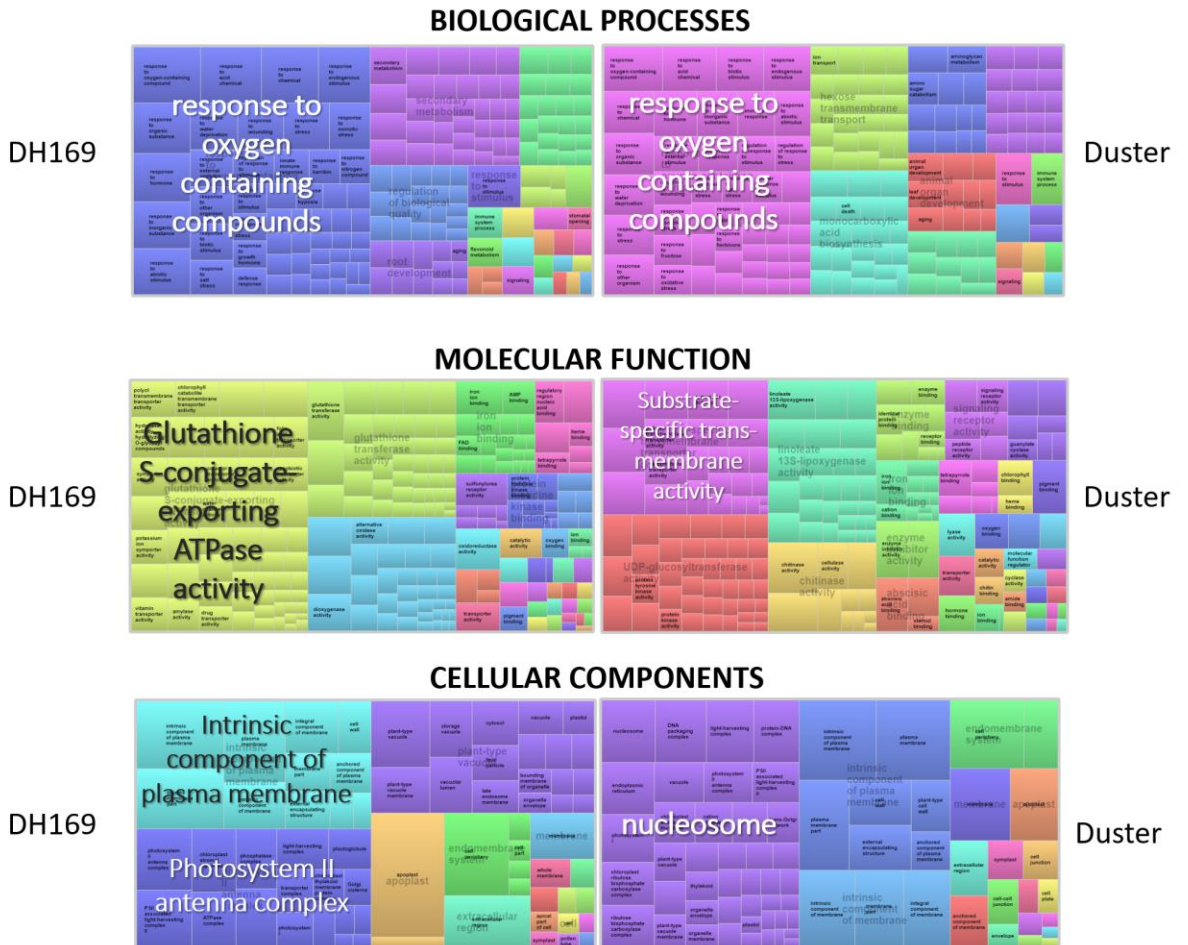As mentioned in methods the Duster reads have been partitioned into two replicates while preserving similar distribution of reads as compared to DH169 (Table 6). The original read count for Duster was 56,224,030 for drought and 33,439,660 for control. In Table 6 below, for Duster, the greatest number of unique matches was identified in the control 1 sample with 12,834,864 reads. As for DH169 the greatest number of unique alignments was in control 2. The average number of unique mappings for Duster in control was 12,833,271 reads with a standard deviation of 2,252.84 and in drought the average was 12,567,928 with a standard deviation of 2,287.49. For DH169 the average number of reads for control was 25,82,194 and drought was 24,985,174. Given the pair-end reads that had a unique alignment the average for control was 12,720,594 with a standard deviation of 2,628,307.4. For drought the average number of unique hits is 12,949,730 with a standard deviation of 1,662,945.4. Comprehensively, subsampling resulted in similar read alignment between Duster and DH169 thereby maintaining a fair comparison for further downstream analysis.

Table 6. Duster and DH169 Methylation Alignment Report - methylation read alignment performed by Bismark with respect to replicates and winter wheat genotypes. Categories of alignment performance are listed, and the numbers of sequencing reads that did not map to the CSW reference genome are also shown.

| Duster | Control 1 | Control 2 | Drought 1 | Drought 2 |
|---|---|---|---|---|
| Total sequence pairs analyzed | 25,000,000 | 25,000,000 | 25,000,000 | 25,000,000 |
| Alignments with a unique best hit | 12,834,864 | 12,831,678 | 12,566,310 | 12,569,545 |
| No alignments under any condition | 4,756,822 | 4,758,376 | 4,340,374 | 4,341,093 |
| Pairs that did not map uniquely | 7,408,314 | 7,409,946 | 8,093,316 | 8,089,362 |
| Mapping efficiency | 51.30% | 51.30% | 50.30% | 50.30% |

| DH169 | Control 1 | Control 2 | Drought 1 | Drought 2 |
|---|---|---|---|---|
| Total sequence pairs analyzed | 22,731,337 | 27,633,050 | 27,452,119 | 22,518,228 |
| Alignments with a unique best hit | 10,862,100 | 14,579,088 | 14,125,610 | 11,773,850 |
| No alignments under any condition | 7,062,327 | 6,844,015 | 6,470,664 | 5,751,684 |
| Pairs that did not map uniquely | 4,806,910 | 6,209,947 | 6,855,845 | 4,992,694 |
| Mapping efficiency | 47.80% | 52.80% | 51.50% | 52.30% |

Furthermore, in Table 6, the overall methylation alignment performed by Bismark revealed an average mapping efficiency of 50.8% in Duster where over half of the paired-end alignments identified a unique alignment across all the samples. DH169 demonstrated comparable mapping efficiency with an average of 51.1%.

## 4.7 METHYLATION CALLING

Methylation calls identified by Bismark for Duster and DH169 are as highlighted in Table 7.  For Duster the average number of cytosines analyzed in control was 308,236,165.5 and for drought the average was 295,739,628, approximately 13 million more cytosines analyzed in control versus drought. Calculated from Table 7 the average total number of methylations for control was 72,223,331 and for drought it was 82,924,199. More importantly, given by the average, more methylation was found in drought as evident by drought having nearly 10.7 million more methylated cytosines than control. Overall, the relative proportion of the cytosines that were methylated in control was 23.43% compared with drought which had more at 28.04%.

As for DH169, the average number of cytosines found in control was 311,435,823, whereas in drought more was found at 320,197,298. Out of the cytosines, 43,128,280 were identified to be methylated in control and for drought less so at 41,131,606, a difference of approximately 2 million. In terms of relative proportion, 13.85% of the cytosines in control were found to be methylated conversely to drought where only 12.85% were methylated. This contrasts with Duster where more methylation was actually found in drought versus control.

Taking in consideration the different methylation types CpG, CHH, and CHG (Table 7), CpG methylation was the most abundant in both control and drought and for both Duster and DH169. This was followed by CHG being the second highest and CHH being the lowest reported methylation. This can be further emphasized in Figure 8 detailing the methylation proportion for Duster and DH169.

41

To determine if the different methylation type plays a role in drought response, Figure 8 also details the methylation proportion with respect to control and drought. As evident in Figure 8, CpG methylation accounts for 70% of the methylation types in control and 71% drought. DH169 follows suit in control and drought with 69% and 68%, respectively. Comprehensively, only small methylation changes occurred between control and drought, typically only a 1% difference as mentioned above. However, although minute, when comparing Duster and DH169, Duster has a greater CpG and CHG contribution, whereas DH169 has a greater proportion of CHH methylation.

To conclude, Duster undergoes a greater degree of methylation in drought compared with control. DH169 reveals the opposite with less methylation occurring in drought compared with control. Genome-wide, Duster has a greater degree of methylation across the genome than DH169. Finally, CpG, CHH, and CHG methylation undergo little changes between control and drought and Duster and DH169.

Table 7. Raw Methylation Count of Duster and DH169 Reads to CSW - the methylated and unmethylated cytosine determination by Bismark. The table shows the total methylation and unmethylation in the context of CpG, CHG and CHH for Duster and DH169 and the replicates in control (well-watered) and drought conditions.

| Duster | Control 1 | Control 2 | Drought 1 | Drought 2 |
|---|---|---|---|---|
| **Total C's analyzed** | 308,272,488 | 308,199,843 | 295,686,364 | 295,792,892 |
| | | | | |
| **Methylated C's** | | | | |
| **CpG** | 50,357,295 | 50,345,161 | 58,511,653 | 58,521,816 |
| **CHG** | 18,323,657 | 18,320,135 | 20,889,335 | 20,894,206 |
| **CHH** | 3,550,978 | 3,549,435 | 3,517,088 | 3,514,299 |
| | | | | |
| **Unmethylated C's** | | | | |
| **CpG** | 43,262,434 | 43,255,656 | 31,131,207 | 31,150,693 |
| **CHG** | 43,887,298 | 43,874,109 | 38,135,985 | 38,150,595 |
| **CHH** | 148,890,826 | 148,855,347 | 143,501,096 | 143,561,283 |
| | | | | |
| **DH169** | | | | |
| **Total C's analyzed** | 258,353,016 | 364,518,629 | 345,128,039 | 295,266,557 |
| | | | | |
| **Methylated C's** | | | | |
| **CpG** | 28,165,259 | 31,015,052 | 34,335,401 | 21,898,779 |
| **CHG** | 10,002,862 | 10,809,177 | 11,967,309 | 7,532,724 |
| **CHH** | 2,798,924 | 3,466,366 | 3,605,788 | 2,923,211 |
| | | | | |
| **Unmethylated C's** | | | | |
| **CpG** | 50,035,909 | 83,256,903 | 71,649,530 | 70,764,265 |
| **CHG** | 41,657,364 | 63,873,116 | 57,989,479 | 52,887,204 |
| **CHH** | 125,692,698 | 172,098,015 | 165,580,532 | 139,260,374 |

Figure 8. Relative Proportion of Methylation and Unmethylation Levels in Duster and DH169 -
Pie chart representation of the relative methylation identified between control and drought
conditions for Duster and DH169.
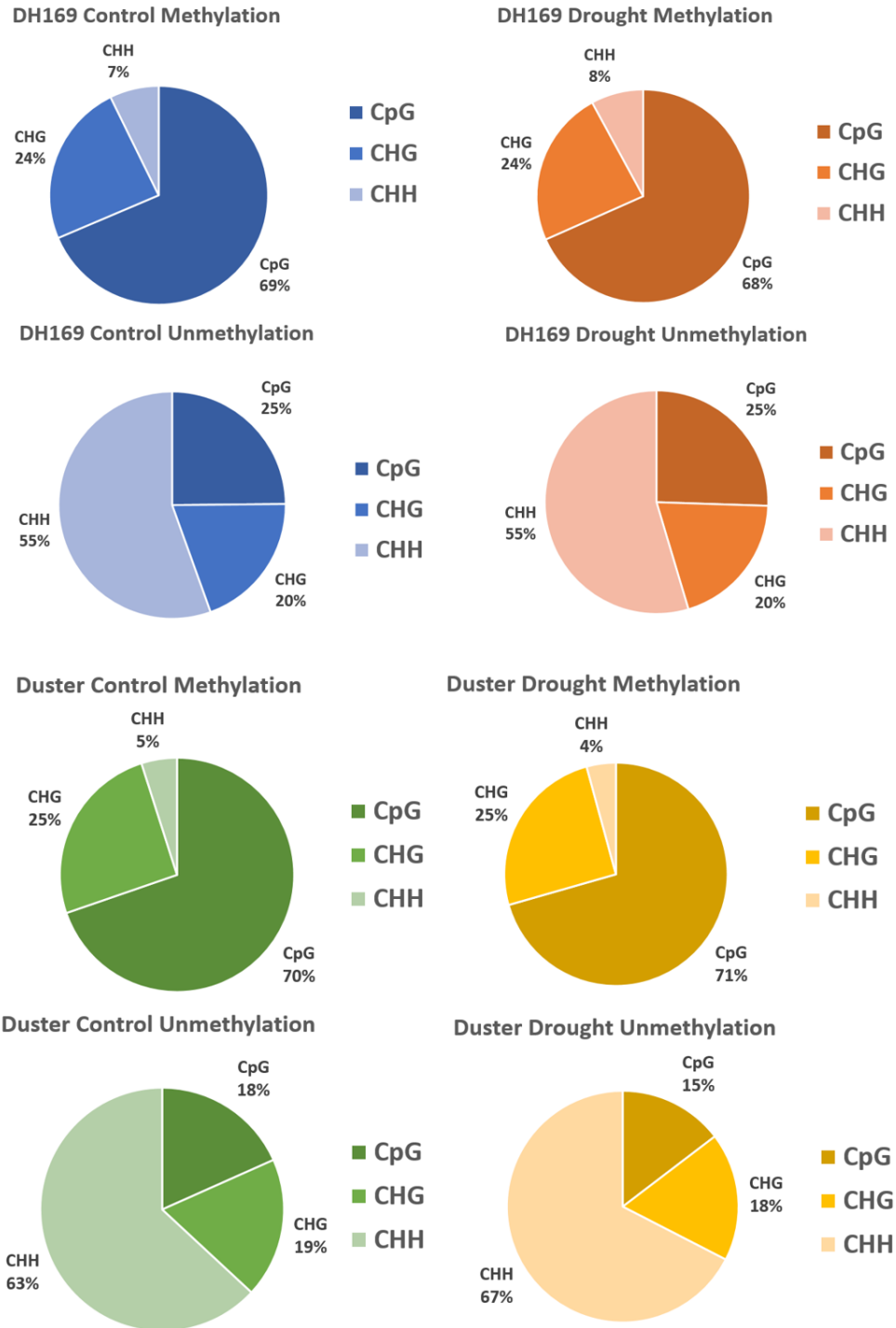
## 4.8 DIFFERENTIAL METHYLATION BETWEEN DROUGHT AND CONTROL FOR DUSTER AND DH169

Identifying differentially methylated cytosines (DMC) across the genome in both DH169 and Duster lay bare that DH169 undergoes less methylation switching than Duster, as portrayed in Figure 9A. The total number of methylation changes that occurred in drought condition was 78,225 for Duster and 31,287 for DH169. In Figure 9A for DH169, CpG is the dominant methylation type in both hyper- and hypo- methylation followed by CHH and lastly CHG. Similarly, in Duster, CpG is also the dominant DMC but, CHH and CHG seemed to maintain an equal ratio.

Sub-genome breakdown of the DMC's was represented in Figure 9B and 9C. In DH169, Figure 9B, the sub-genome differences of the total hypermethylation across all three types revealed that the D sub-genome has less methylation than the A and B sub-genome. This is true with a total count of 4,005 in the D sub-genome which contrasts to A which has 5,414 and B with 5,977. This is also true for hypomethylation where the D sub-genome has the least amount of total methylation with 4,140 compared with the A sub-genome has 5,359 and B sub-genome with 6,003. Additionally, as revealed by Figure 9A, the sub-genome representation follows the same pattern as the genome-wide perspective in terms of CpG, CHH, and CHG. Once again, the dominant DMC for DH169 is CpG for both hyper- and hypo- methylation. After CpG it is followed by CHH then CHG.

For Duster, Figure 9C, given hypermethylation the D sub-genome suffers the lowest methylation changes with 10,089 compared with the A sub-genome with 13,349 and B sub-genome with 15,642. When evaluating hypomethylation for Duster, the D sub-genome had the least number of hypomethylation at 9,672 sites compared with the A sub-genome with 12,989 and B sub-genome with 15,199. Additionally, the B sub-genome had the greatest total number of

differentially methylated cytosines in Duster and DH169 for both hyper- and hypo- methylation. Moreover, the dominant DMC amongst sub-genomes was CpG. This is similarly followed with the genome-wide view where the A sub-genome exhibits similar ratios of CHH and CHG. However, differences appear in the B and D sub-genomes. In the B sub-genome, there is more CHG methylation than CHH with CHG totaling 4,148 and CHH with 3,513. In contrast, in the D sub-genome this relationship is flipped with more CHH methylation than CHG at 2,776 and CHH 2,404. This relationship is also revealed in the hypomethylation.

Most importantly to recapitulate, in Figure 9B and 9C, the D sub-genome demonstrates depreciation in hypermethylation and hypomethylation changes in both Duster and DH169 highlighting a possible sub-genome epigenetic bias.  Just as well, the B sub-genome highlights the greatest number of hypermethylation and hypomethylation. Additionally, given the DMCs Duster continues to be more methylated and undergoes more methylation switching than DH169.

Figure 9. Single Differentially Methylated Cytosine Proportion Between DH169 and Duster - (a) Overall count of the number of DMCs between Duster and DH169 and hypermethylation and hypomethylation. (b) DH169 sub-genome count of the hypermethylation and hypomethylation between each methylation type. (c) Duster sub-genome count of the hypermethylation and hypomethylation between different methylation types (CpG, CHH, CHG).
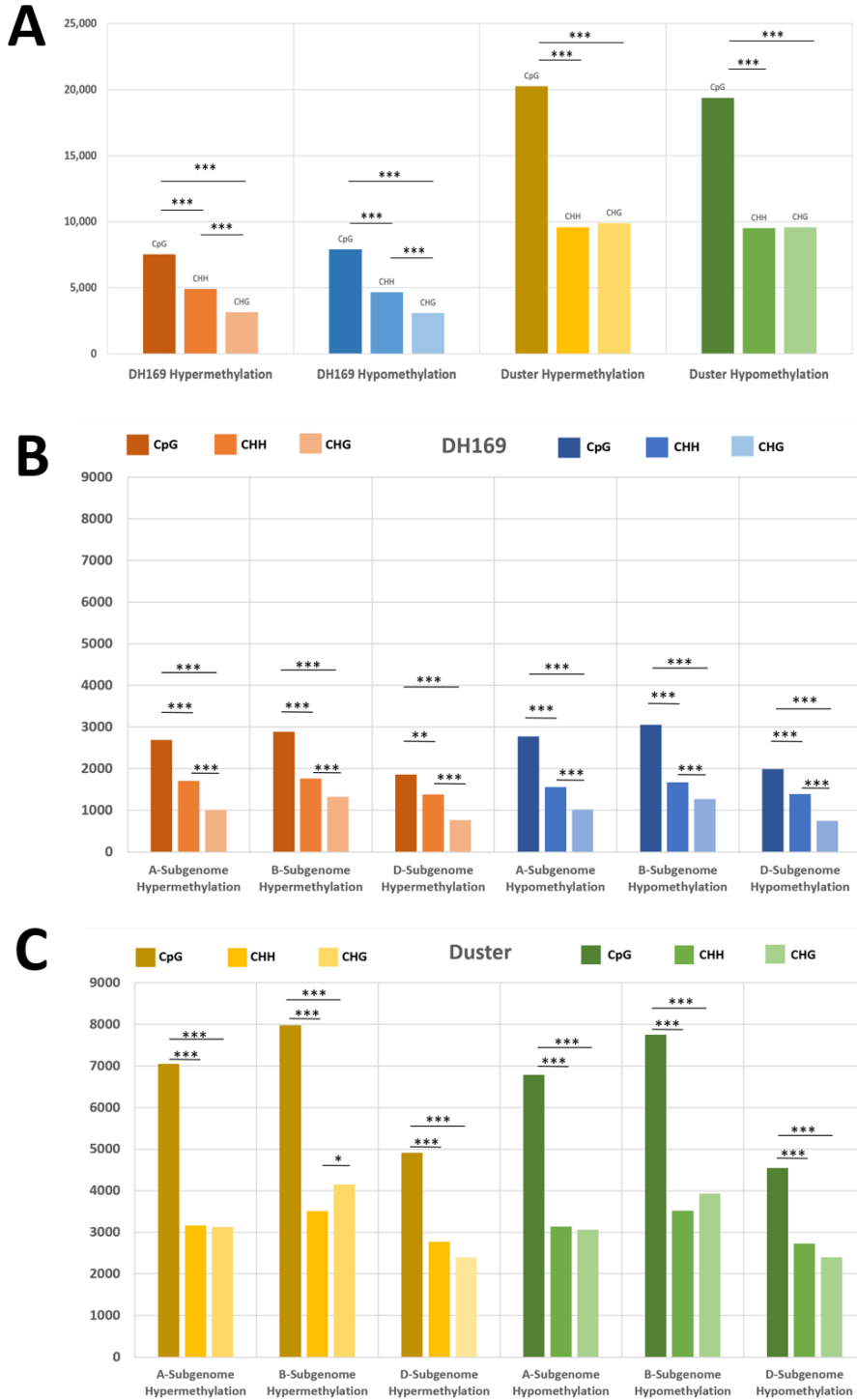
## 4.9 DIFFERENTIAL METHYLATION REGIONS BETWEEN DROUGHT AND CONTROL FOR DUSTER AND DH169

Figure 10 highlights the aggregation of methylation into islands demonstrating any differences between methylation context, drought and control, and DH169 and Duster. To note, the first contrast of the DMIs versus the DMCs was that the overall DMIs in Figure 10A revealed that CpG is no longer the dominant methylation type. In fact, CHH has replaced it being the new dominant methylation, with CHG once again being the least. This is seen in both Duster and DH169, their respective sub-genomes and between hyper- and hypo- methylation. Additionally, compared with the D sub-genome depreciation in overall methylation for DMCs, DMIs actually see that the D sub-genome undergoes the greatest methylation switches in drought to control.

To determine if the differences between the number of hypermethylation and hypomethylation are significant the Shapiro-Wilks test is first applied to determine if the methylations are normally distributed across the chromosomes. Out of all the tests only CpG hyper and hypomethylation and CHH hyper and hypomethylation imply a normal distribution with p-values greater than 0.05. Using the t-test to validate if the difference between differential methylation was significant it was determined that only CpG hyper- and hypo- methylation had significant change with a p-value of 0.049. To determine if the CpG, CHH or CHG methylation were significantly different a t-test was performed across all the methylation types as reported in Figure 7.

To better understand if there were any sub-genome differences Figure 10B focused on the DH169 sub-genome breakdown. One particular note was the depletion of the CHH island hypomethylation count. To determine if this change was significant a single factor ANOVA was performed between the sub-genomes and CHH hypomethylation resulting in a p-value of 0.02.

Interestingly, for Duster in Figure 10C, once again in the B sub-genome, CpG

hypermethylation depicts an overall increase in the total number of hypermethylation compared

with the other sub-genomes. A single-factor ANOVA was performed between the sub-genomes

and CpG hypermethylation which evaluated a p-value of 0.027. To rehash, Duster experiences a

greater number of genome-wide methylation switching in drought compared with DH169.

However, when accounting for methylation islands both Duster and DH169 were relatively

similar with the exception of CpG where DH169 presented a greater number of methylation

island switching compared with Duster.

Figure 10. Differential Methylation of 1,000 Basepair Islands Between DH169 and Duster – (a) Overall count of the methylation islands found in Duster and DH169 and their respective hypermethylation and hypomethylation. (b) DH169 sub-genome count of the hypermethylation and hypomethylation islands between each methylation type. (c) Duster sub-genome count of the hypermethylation and hypomethylation islands between different methylation types (CpG, CHH, CHG).
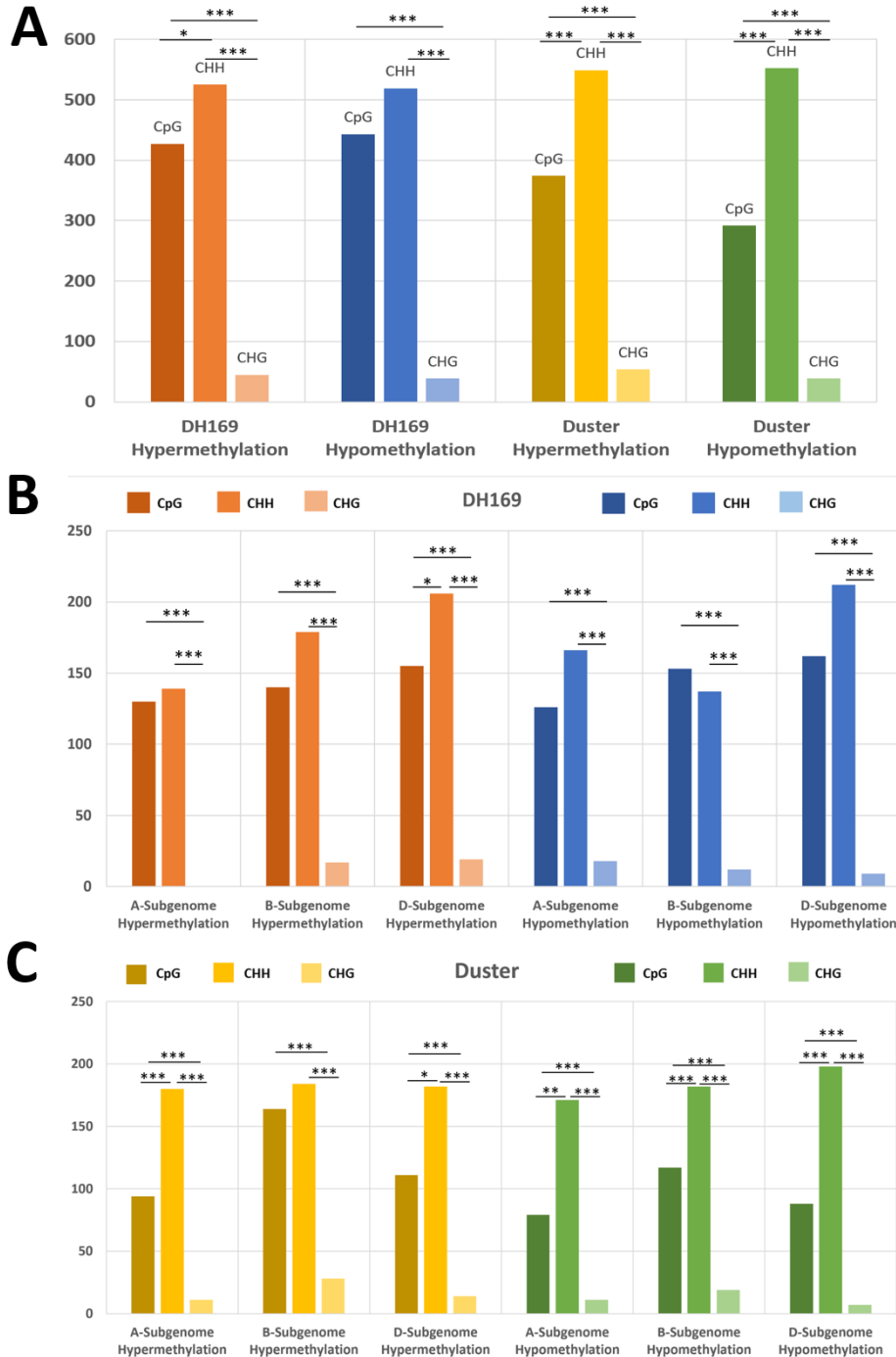
**4.10 DIFFERENTIALLY METHYLATED GENES (DMG) AND REGIONS (DMR)**

To identify the relationship between DNA methylation and gene differential expression the methylation percent difference was mapped to the expression value. Given Duster, the number of DMGs with respect to the 5' upstream region that were identified was 208. Out of the 208 DMGs only 16 were considered differentially expressed. Of the 16, 4 were associated with CpG methylation and 12 were associated with CHG. Of the CpG DMGs 2 were up-regulated and 2 were down-regulated. For the CHG, DMGs, 8 were down regulated and 4 were up-regulated.

As for DH169 there was a total of 296 DMGs that overlapped with a methylation island in its 5' region. Of these 296, only 26 were determined to be differentially expressed, 12 were associated with CpG, 13 were associated with CHH and 1 was associated with CHG methylation. For the CpG methylated DMGs 4 were down-regulated and 8 were up-regulated. The 1 CHG methylated DMG was down-regulated. Lastly, given the CHH DMGs, 6 were up-regulated and 7 were down-regulated.

For the gene body overlap with methylation islands, in Duster there was a total of 138 DMGs. Only 5 out of the 138 DMGs were found to be differentially expressed. Of these 5, 2 were associated with CHG and 3 were tied with CpG methylation. The DMGs that had CpG hypermethylation, 2 were found to be up-regulated, 1 down-regulated, no gene names were identified for them. For the 2 that coincided with CHH methylation both were down-regulated, however, one of the DMGs, XLOC_048400, identified in chromosome 6B with a fold-change expression of -2.32 had a gene name identified, TraesCS6B02G290500. This gene, is considered to play a role in transmembrane transport (GO:0055085).

As for DH169, there was a total of 192 DMGs that were overlapped in the gene body. Slightly more DEGs were found in DH169 compared with Duster at 14. Within these 14, 2 were associated with CpG methylation, 1 with CHH methylation 11 with CHG methylation. There was

1 up-regulated DMG and 1 down-regulated DMG for CpG. For the singular CHH methylated DMG it was up-regulated. Out of the 11 CHG methylated DMGs 6 were down-regulated and 5 were up-regulated. Out of the 14 DMGs, 7 were found to have an annotated gene name. There were 4 genes that identified oxidation-reduction process as the primary GO term (GO:0055114). The other three were beta-amylase (GO:0000272), DNA integration (GO:0015074) and proteolysis (GO:0006508).

To evaluate if differential gene expression was a function of local methylation changes, a regression model was fitted using the log2 fold change of the DMGs with the difference of the associated DMR. The results were shown in Figure 11. As CpG methylation was the dominant methylation, the analyses were performed only with the CpG context. The first significant regression relationship in Figure 11 was demonstrated by DH169 with a correlation coefficient of 0.37 indicating possible minor correlation between a CpG DMR hypermethylation and DMG expression. In Duster, a positive relationship was found between the CpG DMR hypomethylation and DMG up-regulation shown in Figure 11, the correlation coefficient was 0.56.

Figure 11. Regression Analysis of CpG Differentially Methylated Regions (DMR) and Differential Expression of DMGs - Scatterplot of the changes in methylation in the gene-body and the differential expression of the associated DMGs in drought condition for Duster (green dots) and DH169 (blue dots). The lines indicate the significant associations between the DMI under drought and the differential expression in the winter wheat genes.
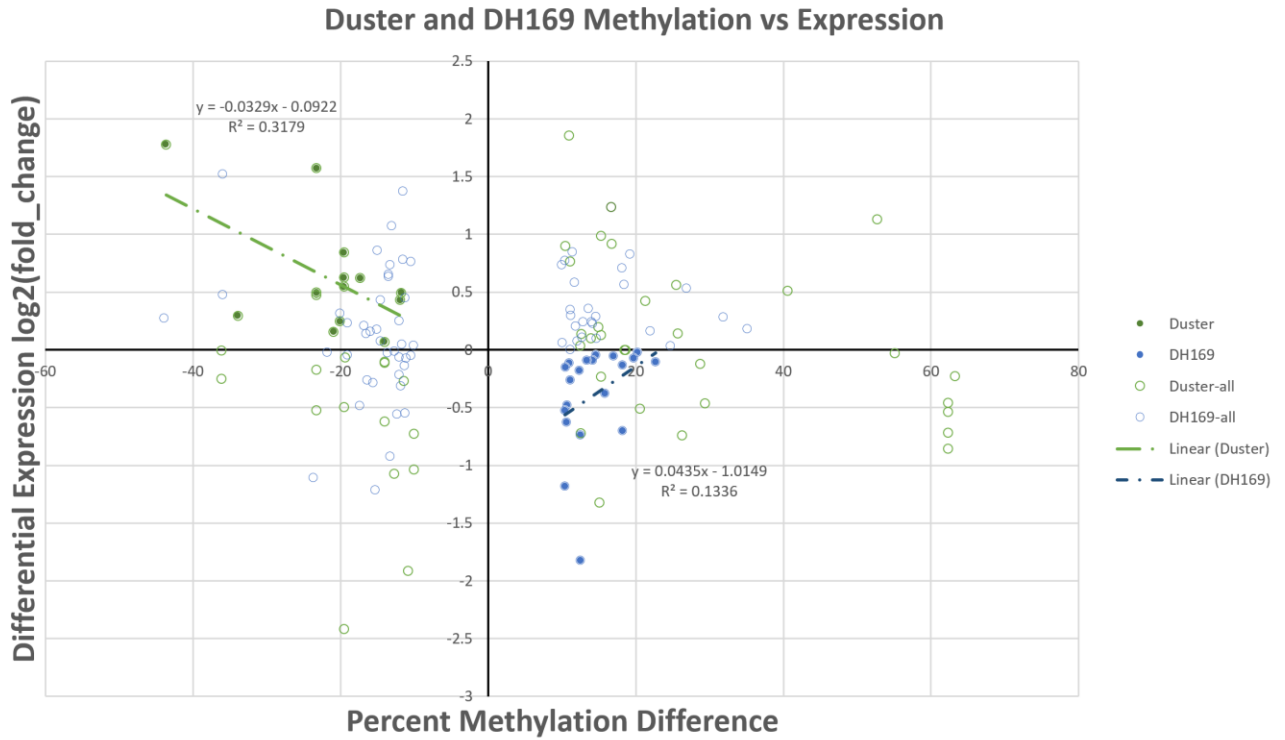
CHAPTER V


DISCUSSION


Understanding abiotic stresses on plant productivity is more crucial than ever in this ever-changing climate. Boyer (1982) highlighted the early limitations in only studying plant productivity with respect to reproductive success and warned that future availability of water would decrease as demand for production increased. He concluded that the genetic potential for plant adaptation was high and that further understanding of the underlying biological mechanisms to environmental stress would be crucial for future agricultural production (Boyer, 1982). In fact, over the years from 1939 to 1978 abiotic stress had accounted for 71% of crop losses (Boyer, 1982). Abiotic stress can be categorized into many groups such as salt stress, cold stress, heat stress, and drought stress. To unveil the genetic mechanisms of abiotic stress in plants, a number of studies highlighted novel mechanisms such as calcium signaling (Knight et al., 1999), glycinebetaine accumulation, (Giri, 2011), drought stress and antioxidant activity (Sairam et al., 2005). Over the last decade, transcriptomic analysis has been in the forefront to identify genes (Hubner et al., 2005), genetic pathways/networks (Haak et al., 2017), regulatory proteins (Ma et al., 2017) and their associated targets (Schroeder et al., 2001, Zhang et al., 2014, Van Ha et al., 2014) for plants genetic adaptation to abiotic stress. To further illustrate, Padmalatha et al., 2012 used genome-wide transcriptomic analysis in cotton to ascertain the regulation of genes in fibre elongation under drought, Lenka et al., 2011 compared the transcriptomic profiles of different drought responsive genotypes in rice and found that a drought sensitive phenotype experienced down-regulation of genes that confer drought tolerance and a drought tolerant phenotype had

up-regulation of pathways that were associated with carbon fixation and glycolysis/gluconeogenesis. Liu et al., 2012 distinguished stress-related genes and transcription factors that were responsive to heat stress and recovery, and Le et al., 2012 studied upregulated genes such as DREB, NAC, AREB and ZAT/STZ transcription factors in soybean leaf tissues under drought stress.

## 5.1 Drought Response Mechanisms and Gene Expression

Reactive oxygen species (ROS) are the result of partial reduction of atmospheric $O_2$. One of the predominant consequences of drought stress is an increase in ROS production (Kaur et al., 2017). An accumulation of ROS can cause negative effects like cellular damage and death, however, besides its destructive component it has been found that they can act as secondary messengers in a plethora of cellular processes (Sharma et al., 2012). Several studies have demonstrated that drought can induce oxidative stress in plants and trigger several signaling pathways to promote physiological and biochemical changes as a defensive mechanism (Sgherri et al., 1995, Moran et al., 1994, Boo et al., 1999, Qi et al., 2018, Cruz de Carvalho, 2008).

To mitigate the overabundance of ROS caused by drought, ROS-scavengers such as superoxide dismutase (SOD), catalase (CAT), glutathione reductase (GR), and proline are commonly employed by plants to combat cellular damage (Das et al., 2014, Karuppanapandian et al., 2011, Ahmad et al., 2010). Amongst these, proline accumulation has been associated with the physiological response to drought by its ability to influence the signaling pathways, act as an osmolyte, and perform as a chemical chaperone (Liang et al., 2013). Specifically, proline in plants are known to prevent protein aggregation and stabilization under extreme temperatures (Samuel et al., 2000, Rajendrakumar et al., 1994), protect nitrate reductase in osmotic stress (Sharma et al., 2005), stabilize ribonucleases (Mishra et al., 2006), and act as an antioxidant under environmental stress including drought (Szabados and Savoure, 2010).

In this study, 503 genes in DH169 and 469 in Duster, were strongly related to proline synthesis and oxidative stress. Shown in Figure 7, the primary ontology group associated with drought induced DEGs were dominated by response to oxygen containing compounds. The genes associated to this GO term for example, XLOC_003251 was found to be down-regulated in DH169. Its query coverage was 92% with 100% identity to the gene TraesCS1B02G223300. This bread wheat gene was annotated as Proline dehydrogenase through a homology comparison (PRODH) which has a predominant role in proline metabolism. Proline dehydrogenase is an enzyme used to catalyze the conversion of proline back to glutamate, preventing proline from accumulating and responding to drought stress (Liang et al., 2013). Furthermore, several of the other up-regulated genes are also associated to proline metabolism. As suggested in these studies, proline needs to accumulate in the endogenous membrane in plants to facilitate the stress induced response (Szabados and Savoure, 2010). The DEG's and GO terms identified in this study both exhibited the importance of proline metabolism and ROS-scavenging; more importantly, DH169 in particular showed 252 more up- and down-regulated transcripts with respect to response to oxygen-containing compounds compared with Duster, possibly contributing to its overall drought response and consequently, the yield performance in the drought year.

As a secondary messenger of ROS, the presence of abscisic acid (ABA) content plays a strong role in response to stress, which subsequently leads to stomatal closure, and restricts plant growth, as a result (Mittler and Blumwald, 2015). Stomatal closure is a defense mechanism produced by the plant to prevent water loss when in the state of desiccation (Brodribb and Holbrook, 2003). When selecting the genes identified in the GO terms for water deprivation, over 87% of the DEG's of DH169 are found to be associated with the plant membrane. In fact, a majority of the ABA signaling occurs at the plasma membrane and proteins such as SLAC1 and SnRK2 mediate ABA-induced stomatal closure under drought stress (Geiger et al., 2009, Zhu, 2016). Therefore, a majority of the DEGs in DH169 may play a role in maintaining ABA

signaling in the plant membrane and contributing to more efficient stomatal closure to prevent water loss in wheat when experiencing a water deficit.

In addition, several significantly up-regulated unique homeo-domain genes were found only in DH169. These were: TraesCS5D02G323100, TraesCS5A02G316800, TraesCS5B02G317400, and TraesCS2A02G389400. These genes currently do not have a definitive function. However, recent research has found that the overexpression of homeodomain-leucine zipper family genes, such as TaHDZ members (Li et al., 2020) and GLABROUS 11 in Arabidopsis (Ruan et al., 2012) demonstrated enhanced drought tolerance, lower water loss rates, higher survival rates and higher proline content. ZmHDZ4 is also a homedomain-leucine zipper that conferred tolerance to drought stress in transgenic rice (Wu et al., 2016).

Of the up-regulated homeo-domain genes identified in DH169, analysis on GO ontology enriched the association with plant morphogenesis and tissue formation, for example, bract morphogenesis (GO:0010433), bract formation (GO:0010434), and bract development (GO:0010432). The water reduction experiment of this study was performed at Feekes 10 of wheat growth. Feekes 10 or the boot stage occurs at heading prior to flowering and grain development or pre-anthesis, which is the growth stage where the formation of spikes occurs. Under drought, flag leaf exhibits less photosynthesis, chlorophyll content, relative water content, antioxidants and antioxidative enzyme activities than spikes (Lou et al., 2018). Formation of spikes is thus important since bracts or the glumes and lemmas in the same study showed that there was greater ROS-scavenging capabilities compared with leaf tissue. To summarize, ROS-scavenging machineries in spike and bract development could be the key advantage that DH169 employed to mitigate the effect of adverse growing conditions like drought.

Many studies have revealed that heat stress and drought stress on wheat had varying degrees of change in the transcriptomic and proteomic expression (Altenbach and Kothari, 2004,

Wan et al., 2008, Szucs et al., 2010, Chauhan et al., 2011, Altenbach, 2012). Such variation in genetic responses can also be seen in other plants such as maize (Moser et al., 2006), chickpea (Gunes et al., 2007), and in Szucs et al., 2010, where only few differences regarding gene expression was found between wheat genotypes under water deprivation. In this study, the overall expression was similar between Duster and DH169 under drought. The average median FPKM expression in drought was 2.82 in Duster, and 2.69 for DH169. Furthermore, the overall GO enrichment categories were also indifferent, with the majority gene overrepresentation categorized under response to stimulus (GO:0050896) at 60.69% and 60.91% for Duster and DH169, respectively. However, 719 more differentially expressed genes were identified in DH169 compared with that of Duster. DH169 is a derived genotype with, in theory, 50% of the Duster genome. It is speculated that the extensive genetic response of DH169 can be attributed to the dysregulation of hypomethylation as a result of hybridization of Duster and Billings genotypes.

Finally, many studies in drought have demonstrated that water deprivation and heat should be considered together. For instance, Johnson et al., 2014 were able to highlight unique transcript changes that were only seen using heat and drought gene probes when under combined stress treatment (Johnson et al., 2014). Given that in Oklahoma, a state located in a marginal region between humid eastern and semi-arid western climate, drought events are typically in association with heat. Therefore, I reckon it is important to understand not just the impact of water deprivation but, also the role heat can play in relation, as well as the regulatory elements (Wang et al., 2003) that can be deployed as a plants defense mechanism to drought stress.

## 5.2 Methylation in Drought

The ability of plants to cope with environmental changes relies on flexible mechanisms. Although the gene-central mechanisms, such as the above-mentioned changes in transcription

activities, have been well received, growing evidence highlights that it can be largely influenced by epigenetic changes like noncoding RNAs (Lee, 2012), histone modification (Rice and Allis, 2001), and DNA methylation (Jaenisch and Bird, 2003).

For example, research in Arabidopsis has shown a genome-wide interdependence between methylation and transcription (Zilberman et al., 2007); and genomic regions that exhibit asymmetric transcription are influenced by DNA methylation (Ramirez-Gonazalez et al., 2018). Changes in DNA methylation, such as hypermethylation and hypomethylation, are associated with repression and activation of transcripts, respectively (Neves et al., 2017). More interestingly, recurrent drought exposure in citrus plants have demonstrated changes in frequency of methylation and may be associated to gain of memory to stress (Neves et al., 2017).

Similar to other studies in plants (Feng et al., 2010), CpG methylation was also present as a dominant type of methylation in both DH169 and Duster winter wheat (Table 7). Shown in Figure 8, the relative proportion of methylation in different sequence contexts (CpG versus CHG versus CHH) was found conserved between these two genotypes, and across conditions. However, in Table 7, the raw methylated cytosine counts of Duster genotype were persistently greater than that of DH169. When evaluating differential methylation due to the water reduction experiment conducted, no significant changes between hypermethylation and hypomethylation counts were detected within each of the two wheat genotypes (Figure 9). However, the number of DMC's were greater for both hypermethylation and hypomethylation in the drought avoidant Duster versus the drought tolerant DH169 (Figure 9). A study by Kaur et al., 2018 discerned that a drought sensitive genotype underwent more methylation events compared with a drought tolerant genotype. Here I also speculate that when encountering water deprivation, the Duster genotype could undergo an extensive rapid methylation that consequently silences global transcription activities, resulting in the invariant yield production of Duster wheat in the 2014

drought, when the annual total rainfall was less than half of a non-drought year (USDA, Oklahoma Annual Wheat Review).

Mechanistically DNA methylation inhibits transcription indirectly through proteins such as methyl-CpG binding protein (MeCP-1) (Boyes and Bird, 1991). Even though DNA methylation is commonly found in genes, the types and dynamics of DNA methylation are often reflective of expression status of the gene (Bewick et al., 2017). However, the changes of DNA methylation levels in genic regions have shown varying degrees of association with gene expression. For example, gene-body methylation has been identified to be positively associated with expression in rice (Li et al., 2012); and similarly in *Populous trichocharpa*, Liang et al., 2014 revealed that methylation in 100-2,000 base pair upstream of transcription start sites (TSS) and within the gene body were positively associated with gene expression. In this study, the results seem to be inconsistent as seen in Figure 11, aggregation of hyper-methylated CpG islands in the gene body was only slightly positively correlated with the down regulation in DH169; however, the results in DH169 did not support the proposed positive association between gene-body methylation and up-regulation. In contrast, the Duster genotype demonstrated the positive association of gene body DNA hypomethylation with the up-regulation of gene expression in drought, with a correlation coefficient of 0.32 (Figure 11). Partially attributed to the reduced representation of BSseq, detailed functional characterization of changes in methylome might still be challenging for large complex plant genomes (Eichten et al., 2015).

As for crop improvement, the current paradigm to pinpoint genetic targets for improved stress tolerance with genome-wide association studies (GWAS) and transcriptomic analysis could be largely inaccurate if the phenotypic variation due to genotype-by-environment interaction (G x E) is unaccounted for. The current results on stress-modulated DNA methylation offer a novel perspective on changes of crop plant productivity in stress, while no significant amount of changes in nucleotide and expression are required. This research effort can be expanded in

examining the role methylation plays on cis-elements, as proposed in a recent study that gene body methylation could be a byproduct of DNA methylation silencing on repeats (Bewick et al., 2017). Furthermore, when considering transgenerational stress memory, a study in Arabidopsis showed that the DNA methylome remained relatively stable under drought stress (Ganguly et al., 2017). To conclude, a further look into the outside role of DNA methylation and its role on cis-elements, the intergenic region, and its preservation across generations may elucidate novel mechanisms in the drought response.

# CHAPTER VI


## CONCLUSION

In this study, using comparative transcriptomics and differential methylation analysis, I uncovered distinct genetic responses of drought-avoidant Duster and drought-tolerant DH169 winter wheat genotypes to a water stress experiment that simulated the 2014 state-wide drought. The overall FPKM expression between Duster and DH169 were similar. However, the differential expression profile revealed that DH169 identified 719 more significant genes than Duster. Upon further investigation, DH169 found 664 more up-regulated genes than Duster, and that the overall expression for the DEGs was greater in drought in DH169 at 4.23 compared with Dusters. Based on GO term enrichment, functional analysis of the identified DEGs revealed that both Duster and DH169 employed similar enriched molecular genetic mechanisms such as response to oxidative stress (GO:0006979), transmembrane transportation activity (GO:0022857), and response to water deprivation (GO:0009414). However, the GO terms unique to DH169 were associated with bracts such as bract morphogenesis (GO:0010433), bract formation (GO:0010434), and bract development (GO:0010432). Other DH169 unique GO terms were response to peptides (GO:1901652), and ATPase-coupled glutathione S-conjugate transmembrane transporter activity (GO:0015431). Overall, the results showed that DH169 is capable of regulating more genes associated with the stress response compared with Duster, possibly associating with its greater yielding capacity under the 2014 drought.

Considering epigenomic modulation, the drought-tolerant DH169 underwent fewer overall differential methylation changes compared with the drought-avoidant Duster genotype. On the contrary, the drought-avoidant Duster genome displayed a more extensive genome-wide methylation, whereas in DH169, a greater degree of demethylation can only be seen under the drought condition. Additionally, in contrast to other research that reports a global association of methylation and gene expression, results of this research suggested a context- and genotype-dependent functional association of up- and down-regulation with hypo- and hyper-methylation.

In conclusion, this research reckons that differing regulatory machinery employed in the winter wheat genotypes display diverse phenotypic drought responses despite their close relatedness. Finally, these findings emphasize that a comprehensive understanding of genetic adaptation would aid the overall resilience needed for crop productivity, and consequently, benefit the rural economy and long-term agricultural sustainability in this rapidly deteriorating climate.

# REFERENCES

1. Abid, G., Mingeot, D., Muhovski, Y., Mergeai, G., Aouida, M., Abdelkarim, S., ... & Jebara, M. (2017). Analysis of DNA methylation patterns associated with drought stress response in faba bean (Vicia faba L.) using methylation-sensitive amplification polymorphism (MSAP). Environmental and Experimental Botany, 142, 34-44.
2. Ahmad, P., Jaleel, C. A., Salem, M. A., Nabi, G., & Sharma, S. (2010). Roles of enzymatic and nonenzymatic antioxidants in plants during abiotic stress. Critical reviews in biotechnology, 30(3), 161-175.
3. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology, 13(10), 1-9.
4. Altenbach, Susan B. "New insights into the effects of high temperature, drought and post-anthesis fertilizer on wheat grain development." Journal of Cereal Science 56.1 (2012): 39-50.
5. Altenbach, Susan B., and Kerry M. Kothari. "Transcript profiles of genes expressed in endosperm tissue are altered by high temperature during wheat grain development." Journal of Cereal Science 40.2 (2004): 115-126.
6. An, Y. Q. C., Goettel, W., Han, Q., Bartels, A., Liu, Z., & Xiao, W. (2017). Dynamic changes of genome-wide DNA methylation during soybean seed development. Scientific reports, 7(1), 1-14.
7. Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., ... & Ronen, G. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science, 361(6403).
8. Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., & García-Herrera, R. (2011). The hot summer of 2010: redrawing the temperature record map of Europe. Science, 332(6026), 220-224.
9. Bartels, A., Han, Q., Nair, P., Stacey, L., Gaynier, H., Mosley, M., ... & Xiao, W. (2018). Dynamic DNA methylation in plant growth and development. International journal of molecular sciences, 19(7), 2144.
10. Bewick, Adam J., and Robert J. Schmitz. "Gene body DNA methylation in plants." Current Opinion in Plant Biology 36 (2017): 103-110.
11. Blum, A. (2018). Plant breeding for stress environments. CRC press.
12. Boo, Y. C., & Jung, J. (1999). Water deficit—induced oxidative stress and antioxidative defenses in rice plants. Journal of Plant Physiology, 155(2), 255-261.
13. Bowman, M. J., Park, W., Bauer, P. J., Udall, J. A., Page, J. T., Raney, J., ... & Campbell, B. T. (2013). RNA-Seq transcriptome profiling of upland cotton (Gossypium hirsutum L.) root tissue under water-deficit stress. PLoS One, 8(12), e82634.
14. Boyer, J. S. (1982). Plant productivity and environment. Science, 218(4571), 443-448.
15. Boyes, Joan, and Adrian Bird. "DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein." Cell 64.6 (1991): 1123-1134.
16. Brodribb, T. J., & Holbrook, N. M. (2003). Stomatal closure during leaf dehydration, correlation with other leaf physiological traits. Plant Physiology, 132(4), 2166-2173.
17. Chauhan, H., Khurana, N., Tyagi, A. K., Khurana, J. P., & Khurana, P. (2011). Identification and characterization of high temperature stress responsive genes in bread wheat (Triticum aestivum L.) and their regulation at various stages of development. Plant molecular biology, 75(1-2), 35-51.

18. Chen, W., Yao, Q., Patil, G. B., Agarwal, G., Deshmukh, R. K., Lin, L., ... & Xu, D. (2016). Identification and comparative analysis of differential gene expression in soybean leaf tissue under drought and flooding stress revealed by RNA-Seq. Frontiers in Plant Science, 7, 1044.

19. Coumou, Dim, and Stefan Rahmstorf. "A decade of weather extremes." Nature climate change 2.7 (2012): 491-496.

20. Cruz de Carvalho, M. H. (2008). Drought stress and reactive oxygen species: production, scavenging and signaling. Plant signaling & behavior, 3(3), 156-165.

21. Daryanto, S., Wang, L., & Jacinthe, P. A. (2016). Global synthesis of drought effects on maize and wheat production. PloS one, 11(5), e0156362.

22. Das, K., & Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. Frontiers in environmental science, 2, 53.

23. Du, Z., Zhou, X., Ling, Y., Zhang, Z., & Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research, 38(suppl_2), W64-W70.

24. Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends in plant science, 15(10), 573-581.

25. Dudziak, K., Zapalska, M., Börner, A., Szczerba, H., Kowalczyk, K., & Nowak, M. (2019). Analysis of wheat gene expression related to the oxidative stress response and signal transduction under short-term osmotic stress. Scientific reports, 9(1), 1-14.

26. Edwards, J. T., Hunger, R. M., Smith, E. L., Horn, G. W., Chen, M. S., Yan, L., ... & Osburn, R. D. (2012). 'Duster'wheat: A durable, dual-purpose cultivar adapted to the southern Great Plains of the USA. Journal of Plant Registrations, 6(1), 37-48.

27. Eichten, Steven R., and Nathan M. Springer. "Minimal evidence for consistent changes in maize DNA methylation patterns following environmental stress." Frontiers in plant science 6 (2015): 308.

28. Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., ... & Ukomadu, C. (2010). Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences, 107(19), 8689-8694.

29. Fracasso, A., Trindade, L. M. & Amaducci, S. Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE. BMC Plant Biol. 16, 115 (2016)

30. French, R. J., and J. E. Schultz. "Water use efficiency of wheat in a Mediterranean-type environment. I. The relation between yield, water use and climate." Australian Journal of Agricultural Research 35.6 (1984): 743-764.

31. Fujita, Y., Fujita, M., Satoh, R., Maruyama, K., Parvez, M. M., Seki, M., ... & Yamaguchi-Shinozaki, K. (2005). AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. The Plant Cell, 17(12), 3470-3488.

32. Ganguly, D. R., Crisp, P. A., Eichten, S. R., & Pogson, B. J. (2017). The Arabidopsis DNA methylome is stable under transgenerational drought stress. Plant Physiology, 175(4), 1893-1912.

33. Garg, R., Chevala, V. N., Shankar, R., & Jain, M. (2015). Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response. Scientific reports, 5, 14922.
34. Geiger, D., Scherzer, S., Mumm, P., Stange, A., Marten, I., Bauer, H., ... & Romeis, T. (2009). Activity of guard cell anion channel SLAC1 is controlled by drought-stress signaling kinase-phosphatase pair. Proceedings of the National Academy of Sciences, 106(50), 21425-21430.
35. Giri, J. (2011). Glycinebetaine and abiotic stress tolerance in plants. Plant signaling & behavior, 6(11), 1746-1751.
36. Goff, L., Trapnell, C., & Kelley, D. (2013). cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version, 2(0).
37. Gunes, A., Pilbeam, D. J., Inal, A., Bagci, E. G., & Coban, S. (2007). Influence of silicon on antioxidant mechanisms and lipid peroxidation in chickpea (Cicer arietinum L.) cultivars under drought stress. Journal of Plant Interactions, 2(2), 105-113.
38. Haak, D. C., Fukao, T., Grene, R., Hua, Z., Ivanov, R., Perrella, G., & Li, S. (2017). Multilevel regulation of abiotic stress responses in plants. Frontiers in plant science, 8, 1564.
39. Hotchkiss, R. D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. Journal of Biological Chemistry, 175(1), 315-332.
40. Hubner, N., Wallace, C. A., Zimdahl, H., Petretto, E., Schulz, H., Maciver, F., ... & Musilova, A. (2005). Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. Nature genetics, 37(3), 243-253.
41. Hübner, S., Korol, A. B. & Schmid, K. J. RNA-Seq analysis identifies genes associated with differential reproductive success under drought-stress in accessions of wild barley Hordeum spontaneum. BMC Plant Biol. 15, 134 (2015).
42. Irizarry, R. A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S. A., Jeddeloh, J. A., ... & Feinberg, A. P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). Genome research, 18(5), 780-790.
43. Jaenisch, Rudolf, and Adrian Bird. "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nature genetics 33.3 (2003): 245-254.
44. Johnson, T. B., & Coghill, R. D. (1925). Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus1. Journal of the American Chemical Society, 47(11), 2838-2844.
45. Johnson, S. M., Lim, F. L., Finkler, A., Fromm, H., Slabas, A. R., & Knight, M. R. (2014). Transcriptomic analysis of Sorghum bicolor responding to combined heat and drought stress. BMC genomics, 15(1), 456.
46. Kakumanu, A., Ambavaram, M. M., Klumas, C., Krishnan, A., Batlang, U., Myers, E., ... & Pereira, A. (2012). Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant physiology, 160(2), 846-867.
47. Karuppanapandian, T., Moon, J. C., Kim, C., Manoharan, K., & Kim, W. (2011). Reactive oxygen species in plants: their generation, signal transduction, and scavenging mechanisms. Australian Journal of Crop Science, 5(6), 709.

48. Kaur, G., & Asthir, B. (2017). Molecular responses to drought stress in plants. Biologia Plantarum, 61(2), 201-209.
49. Kaur, A., A. Grewal, and P. Sharma. "Comparative analysis of DNA methylation changes in two contrasting wheat genotypes under water deficit." Biologia plantarum 62.3 (2018): 471-478.
50. Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature methods, 12(4), 357-360.
51. Kim, H., Lee, K., Hwang, H., Bhatnagar, N., Kim, D. Y., Yoon, I. S., ... & Kim, B. G. (2014). Overexpression of PYL5 in rice enhances drought tolerance, inhibits growth, and modulates gene expression. Journal of experimental botany, 65(2), 453-464.
52. Knight, H. (1999). Calcium signaling during abiotic stress in plants. In International review of cytology (Vol. 195, pp. 269-324). Academic Press.
53. Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. bioinformatics, 27(11), 1571-1572.
54. Lasky, J. R., Des Marais, D. L., Lowry, D. B., Povolotskaya, I., McKay, J. K., Richards, J. H., ... & Juenger, T. E. (2014). Natural variation in abiotic stress responsive gene expression and local adaptation to climate in Arabidopsis thaliana. Molecular biology and evolution, 31(9), 2283-2296.
55. Lenka, S. K., Katiyar, A., Chinnusamy, V., & Bansal, K. C. (2011). Comparative analysis of drought-responsive transcriptome in Indica rice genotypes with contrasting drought tolerance. Plant biotechnology journal, 9(3), 315-327.
56. Le, D. T., Nishiyama, R., Watanabe, Y., Tanaka, M., Seki, M., Yamaguchi-Shinozaki, K., ... & Tran, L. S. P. (2012). Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. PloS one, 7(11), e49522.
57. Lee, Jeannie T. "Epigenetic regulation by long noncoding RNAs." Science 338.6113 (2012): 1435-1439.
58. Lesk, Corey, Pedram Rowhani, and Navin Ramankutty. "Influence of extreme weather disasters on global crop production." Nature 529.7584 (2016): 84-87.
59. Li, C., Nong, Q., Solanki, M. K., Liang, Q., Xie, J., Liu, X., ... & Li, Y. (2016). Differential expression profiles and pathways of genes in sugarcane leaf at elongation stage in response to drought stress. Scientific reports, 6, 25698.
60. Li, S., Chen, N., Li, F., Mei, F., Wang, Z., Cheng, X., ... & Mao, H. (2020). Characterization of wheat homeodomain-leucine zipper family genes and functional analysis of TaHDZ5-6A in drought tolerance in transgenic Arabidopsis. BMC Plant Biology, 20(1), 50.
61. Li, X., Zhu, J., Hu, F., Ge, S., Ye, M., Xiang, H., ... & Li, Q. (2012). Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC genomics, 13(1), 300.
62. Liang, D., Zhang, Z., Wu, H., Huang, C., Shuai, P., Ye, C. Y., ... & Yin, W. (2014, December). Single-base-resolution methylomes of Populus trichocarpa reveal the association between DNA methylation and drought stress. In BMC genetics (Vol. 15, No. S1, p. S9). BioMed Central.

63. Liang, X., Zhang, L., Natarajan, S. K., & Becker, D. F. (2013). Proline mechanisms of stress survival. Antioxidants & redox signaling, 19(9), 998-1011.

64. Lindsey, Rebecca, and L. Dahlman. "Climate change: Global temperature." National Oceanic and Atmospheric Administration (2018).

65. Liu, G. T., Wang, J. F., Cramer, G., Dai, Z. W., Duan, W., Xu, H. G., ... & Li, S. H. (2012). Transcriptomic analysis of grape (Vitis viniferaL.) leaves during and after recovery from heat stress. BMC plant Biology, 12(1), 174.

66. Liu, Z., Xin, M., Qin, J., Peng, H., Ni, Z., Yao, Y., & Sun, Q. (2015). Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (Triticum aestivum L.). BMC plant biology, 15(1), 1-20.

67. López-Maury, L., Marguerat, S., & Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nature Reviews Genetics, 9(8), 583-593.

68. Lou, L., Li, X., Chen, J., Li, Y., Tang, Y., & Lv, J. (2018). Photosynthetic and ascorbate-glutathione metabolism in the flag leaves as compared to spikes under drought stress of winter wheat (Triticum aestivum L.). PLoS One, 13(3), e0194625.

69. Ma, J., Li, R., Wang, H., Li, D., Wang, X., Zhang, Y., ... & Li, Y. (2017). Transcriptomics analyses reveal wheat responses to drought stress during reproductive stages under field conditions. Frontiers in Plant Science, 8, 592.

70. Mandel, J. L., & Chambon, P. (1979). DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. Nucleic acids research, 7(8), 2081-2103.

71. McGhee, J. D., & Ginder, G. D. (1979). Specific DNA methylation sites in the vicinity of the chicken β-globin genes. Nature, 280(5721), 419-420.

72. Mishra, S., & Dubey, R. S. (2006). Inhibition of ribonuclease and protease activities in arsenic exposed rice seedlings: role of proline as enzyme protectant. Journal of Plant Physiology, 163(9), 927-936.

73. Mittler, R., & Blumwald, E. (2015). The roles of ROS and ABA in systemic acquired acclimation. The Plant Cell, 27(1), 64-70.

74. Moran, J. F., Becana, M., Iturbe-Ormaetxe, I., Frechilla, S., Klucas, R. V., & Aparicio-Tejo, P. (1994). Drought induces oxidative stress in pea plants. Planta, 194(3), 346-352.

75. Moser, S. B., Feil, B., Jampatong, S., & Stamp, P. (2006). Effects of pre-anthesis drought, nitrogen fertilizer rate, and variety on grain yield, yield components, and harvest index of tropical maize. Agricultural Water Management, 81(1-2), 41-58.

76. Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2012). NAC transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1819(2), 97-103.

77. Neves, D. M., da Hora Almeida, L. A., Santana-Vieira, D. D. S., Freschi, L., Ferreira, C. F., dos Santos Soares Filho, W., ... & da Silva Gesteira, A. (2017). Recurrent water deficit causes epigenetic and hormonal changes in citrus plants. Scientific reports, 7(1), 1-11.

78. Padmalatha, K. V., Dhandapani, G., Kanakachari, M., Kumar, S., Dass, A., Patil, D. P., ... & Leelavathi, S. (2012). Genome-wide transcriptomic analysis of cotton under drought stress reveal significant down-regulation of genes and pathways involved in fibre

elongation and up-regulation of defense responsive genes. Plant molecular biology, 78(3), 223-246.

79. Poudel, P. (2016). Screening of Winter Wheat Double Haploid Population 'Buster'under Heat and Drought Stress (Doctoral dissertation).

80. Osmolovskaya, N., Shumilina, J., Kim, A., Didio, A., Grishina, T., Bilova, T., ... & Frolov, A. (2018). Methodology of drought stress research: Experimental setup and physiological characterization. International journal of molecular sciences, 19(12), 4089.

81. Qi, J., Song, C. P., Wang, B., Zhou, J., Kangasjärvi, J., Zhu, J. K., & Gong, Z. (2018). Reactive oxygen species signaling and stomatal movement in plant responses to drought stress and pathogen attack. Journal of integrative plant biology, 60(9), 805-826.

82. Rajendrakumar, C. S., Reddy, B. V., & Reddy, A. R. (1994). Proline-protein interactions: protection of structural and functional integrity of M4 lactate dehydrogenase. Biochemical and biophysical research communications, 201(2), 957-963.

83. Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., ... & Khedikar, Y. (2018). The transcriptional landscape of polyploid wheat. Science, 361(6403).

84. Razin, A., & Cedar, H. (1991). DNA methylation and gene expression. Microbiology and Molecular Biology Reviews, 55(3), 451-458.

85. Reeves, T. G., Thomas, G., & Ramsay, G. (2016). Save and grow in practice: maize, rice, wheat--a guide to sustainable cereal production. UN Food and Agriculture Organization, Rome.

86. Rice, Judd C., and C. David Allis. "Histone methylation versus histone acetylation: new insights into epigenetic regulation." Current opinion in cell biology 13.3 (2001): 263-273.

87. Ruan, L., Chen, L., Chen, Y., He, J., Zhang, W., Gao, Z., & Zhang, Y. (2012). Expression of Arabidopsis HOMEODOMAIN GLABROUS 11 enhances tolerance to drought stress in transgenic sweet potato plants. Journal of Plant Biology, 55(2), 151-158.

88. Sairam, R. K., Srivastava, G. C., Agarwal, S., & Meena, R. C. (2005). Differences in antioxidant activity in response to salinity stress in tolerant and susceptible wheat genotypes. Biologia Plantarum, 49(1), 85.

89. Schroeder, J. I., Kwak, J. M., & Allen, G. J. (2001). Guard cell abscisic acid signalling and engineering drought hardiness in plants. Nature, 410(6826), 327-330.

90. Sgherri, C. L. M., & Navari-Izzo, F. (1995). Sunflower seedlings subjected to increasing water deficit stress: oxidative stress and defence mechanisms. Physiologia Plantarum, 93(1), 25-30.

91. Sharma, P., Jha, A. B., Dubey, R. S., & Pessarakli, M. (2012). Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. Journal of botany, 2012.

92. Sharma, P., & Dubey, R. S. (2005). Modulation of nitrate reductase activity in rice seedlings under aluminium toxicity and water stress: role of osmolytes as enzyme protectant. Journal of plant physiology, 162(8), 854-864.

93. Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H. L., ... & Zhang, Y. (2013). Genome-wide analysis reveals TET-and TDG-dependent 5-methylcytosine oxidation dynamics. Cell, 153(3), 692-706.

94. Skriver, K., & Mundy, J. (1990). Gene expression in response to abscisic acid and osmotic stress. The Plant Cell, 2(6), 503.

95. Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one, 6(7), e21800.

96. Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., ... & Miskus, D. (2002). The drought monitor. Bulletin of the American Meteorological Society, 83(8), 1181-1190.

97. Szabados, L., & Savoure, A. (2010). Proline: a multifunctional amino acid. Trends in plant science, 15(2), 89-97.

98. Szűcs, A., Jäger, K., Jurca, M. E., Fábián, A., Bottka, S., Zvara, Á., ... & Fehér, A. (2010). Histological and microarray analysis of the direct effect of water shortage alone or combined with heat on early grain development in wheat (Triticum aestivum). Physiologia plantarum, 140(2), 174-188.

99. Tang, S., Liang, H., Yan, D., Zhao, Y., Han, X., Carlson, J. E., ... & Yin, W. (2013). Populus euphratica: the transcriptomic response to drought stress. Plant molecular biology, 83(6), 539-557.

100. Tang, X. M., Tao, X., Wang, Y., Ma, D. W., Li, D., Yang, H., & Ma, X. R. (2014). Analysis of DNA methylation of perennial ryegrass under drought using the methylation-sensitive amplification polymorphism (MSAP) technique. Molecular Genetics and Genomics, 289(6), 1075-1084.

101. Trapnell, Cole, et al. "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks." Nature protocols 7.3 (2012): 562-578.

102. USDA. 2019 STATE AGRICULTURE OVERVIEW; USDA, 2020.

103. Van Ha, C., Leyva-González, M. A., Osakabe, Y., Tran, U. T., Nishiyama, R., Watanabe, Y., ... & Yamaguchi-Shinozaki, K. (2014). Positive regulatory role of strigolactone in plant responses to drought and salt stress. Proceedings of the National Academy of Sciences, 111(2), 851-856.

104. Varoquaux, N., Cole, B., Gao, C., Pierroz, G., Baker, C. R., Patel, D., ... & Yoshinaga, Y. (2019). Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. Proceedings of the National Academy of Sciences, 116(52), 27124-27132.

105. Wan, Y., Poole, R. L., Huttly, A. K., Toscano-Underwood, C., Feeney, K., Welham, S., ... & Mitchell, R. A. (2008). Transcriptome analysis of grain development in hexaploid wheat. BMC genomics, 9(1), 121.

106. Wang, Wangxia, Basia Vinocur, and Arie Altman. "Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance." Planta 218.1 (2003): 1-14.

107. Wang, W. S., Pan, Y. J., Zhao, X. Q., Dwivedi, D., Zhu, L. H., Ali, J., ... & Li, Z. K. (2011). Drought-induced site-specific DNA methylation and its association with drought tolerance in rice (Oryza sativa L.). Journal of experimental botany, 62(6), 1951-1960.

108. Wang, Zhong, Mark Gerstein, and Michael Snyder. "RNA-Seq: a revolutionary tool for transcriptomics." Nature reviews genetics 10.1 (2009): 57-63.

109. Wu, J., Zhou, W., Gong, X., & Cheng, B. (2016). Expression of ZmHDZ4, a maize homeodomain-leucine zipper I gene, confers tolerance to drought stress in transgenic rice. Plant Molecular Biology Reporter, 34(4), 845-853.

110. Xiong, H., Yu, J., Miao, J., Li, J., Zhang, H., Wang, X., ... & Li, Y. (2018). Natural variation in OsLG3 increases drought tolerance in rice by inducing ROS scavenging. Plant Physiology, 178(1), 451-467.

111. Yebra, M. J., & Bhagwat, A. S. (1995). A cytosine methyltransferase converts 5-methylcytosine in DNA to thymine. Biochemistry, 34(45), 14752-14757.

112. Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science, 328(5980), 916-919.

113. Zhao, Chuang, et al. "Temperature increase reduces global yields of major crops in four independent estimates." Proceedings of the National Academy of Sciences 114.35 (2017): 9326-9331.

114. Zhang, C., Bai, M.Y. and K. Chong. 2014. Brassinosteriod-mediated regulation of agronomic traits in rice. Plant cell reports 33;683-696.

115. Zhang, X., Liu, X., Zhang, D., Tang, H., Sun, B., Li, C., ... & Xie, X. (2017). Genome-wide identification of gene expression in contrasting maize inbred lines under field drought conditions reveals the significance of transcription factors in drought tolerance. PLoS One, 12(7), e0179477.

116. Zheng, X., Chen, L., Li, M., Lou, Q., Xia, H., Wang, P., ... & Luo, L. (2013). Transgenerational variations in DNA methylation induced by drought stress in two rice varieties with distinguished difference to drought resistance. PloS one, 8(11), e80253.

117. Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. Cell, 167(2), 313-324.

118. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., & Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature genetics, 39(1), 61-69.

VITA

Alexander Lim

Candidate for the Degree of

Master of Science

Thesis: DROUGHT-INDUCED EPIGENETIC MODULATION AND
TRANSCRIPTIONAL VARIATION OF WINTER WHEAT

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:

Completed the requirements for the Master of Science in your major at
Oklahoma State University, Stillwater, Oklahoma in December, 2020.

Completed the requirements for the Bachelor of Science in your major at
Oklahoma State University, Stillwater, Oklahoma in December, 2015.

Experience:

Professional Memberships:
2015-2020: Biochemistry and Molecular Biology Graduate Students'
Association (Member, Faculty Representative)