warwick.ac.uk/lib-publications

Genomic mechanisms of adaptation to drought-stress in maize

Ryan Robert Morrison MSc, BSc (Hons)

Submitted for the degree of Doctor of Philosophy

The University of Warwick

School of life Sciences

January 2020

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I'd like to thank my supervisors, Dr Logan Kistler and Prof Robin Allaby, for all their expertise, support and for teaching me how to read and write…code. I couldn't have asked for two better people to be my supervisors. I'd also like to thank my advisory panel, Dr Jose Gutierrez-Marcos and Dr Charlotte Allender, for all their help throughout my PhD, and NERC CENTA for funding this project.

I owe a massive thank you to all the people in the Allaby group, past and present: Roselyn Ware, Hsiao-Lei Liu, Fran Allaby, Rosie Everett, Becky Cribdon, Oli Smith, Kiran Johal and Hannah Macdonald. It was amazing to be part of such a warm and friendly bunch. Also, sorry for almost crashing the server that one time, when I accidentally created the largest file on earth.

Other people I would like to thank are Rob Maple and Gary Grant for teaching me how to farm maize at the PBF, Jon Price and Gabriel Erni Cassola Barata for helping me comprehend R, and the University of Leicester's Hollie Marshall for all her help with the mystery that is SeqMonk. I couldn't have finished this PhD without all of your help.

Finally, on a more personal note, I've been told by my mother, Pamela, that I should thank her for being so encouraging and for being a sounding board throughout my PhD. So, thank you Mum…and thank you to my siblings Ross and Amber, to my Dad Terence, and to my lovely Nanny Margaret. You've always been there for me, are infinitely supportive and I wouldn't change any of you for the world.

Declarations

In accordance with the University of Warwick regulations for the degree of Doctor of Philosophy, I certify that this thesis has been written solely by myself. The work contained in this thesis is my own unless otherwise stated. No aspect of this work has been submitted to any other institution for any other degree or award.

Summary

Maize is a globally important crop, that can be devastated by drought stress. Therefore, there is a demand for increasing the adaptability of maize to drought stress. The activity of transposable elements (TEs) can cause long-lasting genetic changes, causing potential adaptation benefits. DNA methylation is an epigenetic process known to regulate the activity of TEs, and can also respond to external stress.

This thesis is an investigation into the link between promoter region methylation, TE methylation and drought stress, in the drought tolerant B76 and drought susceptible B73 maize varieties. There were three major aims: 1) To test if there is a difference in methylation in gene promoter regions caused by drought in both varieties, 2) To ascertain methylation differences found in the TE fraction caused by drought in both varieties, 3) To discover if differentially methylated TEs were found in differentially methylated promoter regions of potentially important drought-tolerance genes.

The primary method of investigation in chapters 3 and 4 involved the generation of bisulfite-treated DNA sequencing. Differential methylation analysis of this data showed that there is a methylation drought stress response found in B76, that is not found in B73. Methylation differences were also found in the majority of TE families in both varieties. However, the *huck* TE sub-family was found to be differentially methylated after drought stress and overrepresented within differentially methylated promoter regions of the B76 variety, suggesting a possible role in drought tolerance through the restriction of expression of particular genes.

This research adds to the knowledge surrounding methylation changes after drought stress in the TE fraction, while also highlighting potential drought tolerance candidate gene, thereby providing a jumping off point for future adaptation studies in maize.

List of common abbreviations

°C	Degrees Celsius
A	Adenine
ABA	Abscisic acid
bp	Base pairs
C	Cytosine
cDNA	Complementary DNA
cm	Centimetres
CG	Cytosine – Guanine methylation site
CTAB	Cetyltrimethylammonium bromide
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
FASTA	FAST-All format
FASTQ	FAST-Quality format
FDR	False discovery rate
FGS	Filtered gene set
g	Gram
g	Gravity or g-force
G	Guanine
GO	Gene ontology
h	Hour
JA	Jasmonic acid
kb	Kilobase
L	Litre
LINE	Long interspersed nuclear element
LTR	Long tandem repeat
L1	LINE-1
M	Molar (mol/L)
MITE	Miniature Inverted-repeat TE
mM	Millimolar

mm	Millimetres
mRNA	Messenger RNA
NA	Not applicable
ng	Nanogram
PCR	Polymerase chain reaction
ROS	Reactive oxygen species
RNA	Ribonucleic acid
RPKM	Reads per kilobase per million
s	Second
SEA	Singular enrichment analysis
SINE	Short interspersed nuclear element
siRNA	Short interfering RNA
sp.	Species (singular)
spp.	Species (plural)
T	Thymine
TE	Transposable element
TIR	Terminal inverted repeat
tRNA	Transfer RNA
U	Uracil
UTE	Unique transposable element
UTR	Untranslated region
µg	Microgram
µl	Microliter

1   Introduction

1.1   Climate change leading to drought

Climate change is a complex phenomenon that has long-term devastating meteorological consequences (IPCC 2014). As such, the effect of an increase in global temperature on many crops cannot be understated, as it affects the predictability of rainfall during pivotal crop growth periods. A global temperature increase of even 1.5°C has the potential to double drought magnitudes for 30% of the global landmass, including the important, major crop growing regions of central USA and North-West China (Naumann et al. 2018). Despite the expectation of global food demand increasing 100-110% by 2050 due to population increases, drought caused by climate change is projected to lead to 5.6%-6.3% yield loss by the end of the 21st century, if the current trajectory is maintained (Godfray et al. 2010; Leng & Hall 2019; Tilman et al. 2011).

Drought has a significant effect on yield in many crop plant species (Aprile et al. 2009; Garg et al. 2015; Lopes et al. 2013), and is the cause of around 70 percent of all yield losses in crop plants worldwide (Forestan et al. 2016). The response of crops to drought stress differs between species, although they consistently involve complex morphological and physiological mechanisms (Lamaoui et al. 2018). Depending on the timing of the drought stress, there are differing reactions from crops, with cereals being particularly affected during the flowering stage (Daryanto, Wang & Jacinthe 2017). Maize (*Zea mays*) production in particular is often affected by drought conditions (Mao et al. 2015; Tan 2010), and with climate change causing increased temperatures and irregular rainfall, more research into crop adaptability in maize is needed.

1.2   The global importance of maize

Maize is a cereal grain crop that started its domestication journey ~9000 years ago, evolving from wild Balsas teosinte (*Zea mays* ssp. *parviglumis*) in lowland Mexico, before being improved upon across different domestication centres, such as the one discovered sometime after ~6500 B.P in the southwestern Amazon region (Kistler et

al. 2018; Piperno et al. 2009). Maize has become one of the most important crops, needing the third largest crop production acreage globally, behind only wheat and rice (Nuccio et al. 2015). It is primarily grown in major economic countries, with the USA harvesting 82.7 million acres with a total corn crop value of $49.1 billion in 2017 (USDA, World Agricultural Supply and Demand Estimates, May 2019), and China using maize as the primary source of feed crop for their expansive meat industry (Ely, Geall & Song 2016). In fact, in fifty years, land for the growth of maize increased from 39.5 million acres across China in 1967 to more than 104.5 million acres in 2017 for both food and feed production (FAO 2019). Therefore, maize is one of the most important crops that will be needed to feed the ever-growing population of the world, estimated to be 13.1 billion people by 2100 (Hoornweg & Pope 2017).

1.3   Drought impact on maize production

The impact of drought on maize production cannot be overstated, and historical drought events have a clear effect on the yield produced in the same years. This is shown in the USA using yearly maize production data from the Food and Agriculture Organization of the United Nations (FAO) spanning 1961 to 2017 and major historical drought events (Figure 1.1.).
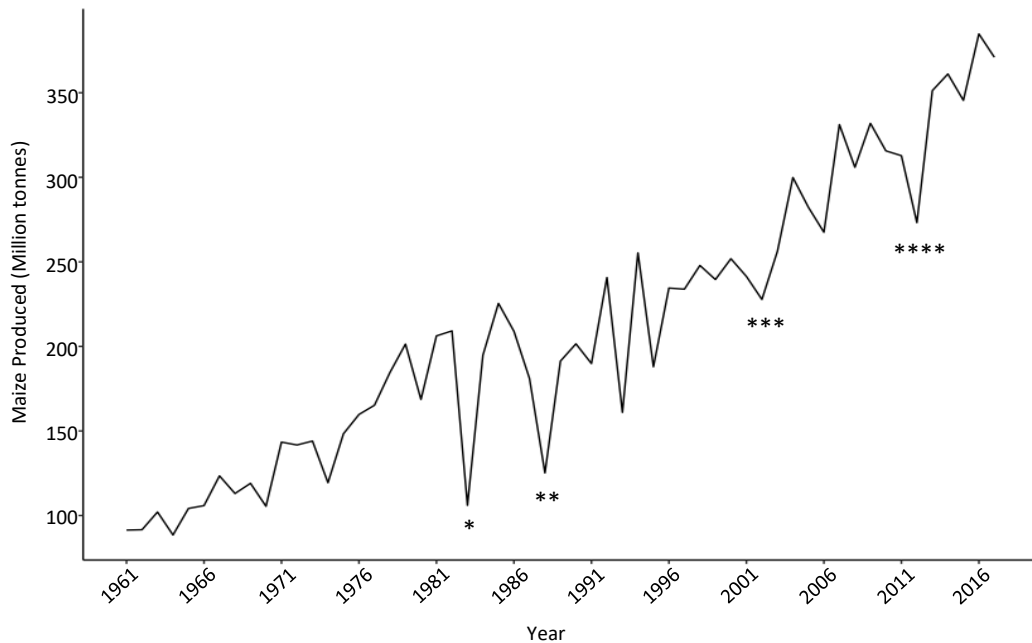
Figure 1.1. A line graph representing the yearly maize production in the USA between 1961 and 2017.

Highlighted are years where there were well-known drought events occurring in the USA maize-growing regions: * 1983 (Fritsch, Kane & Chelius 1986), ** 1988 (Trenberth, Branstator & Arkin 1988), *** 2002 (Cook et al. 2004), *** 2012 (Rippey 2015).

This pattern shows the real-world damage drought stress can have on crop production in one of the largest maize producers in the world, and although the above graph only focuses on the USA, the same is true of other major maize producers. In fact, using data from 1980-2015, it was shown that a 40% reduction of water during growth of maize accounts for a 39.3% reduction in yield (Daryanto, Wang & Jacinthe 2016). Another study shows that variation in global climate patterns accounts for 18% of maize production variability globally (Anderson et al. 2019), and highlights the effect even the slightest change in climate can have on maize production.

There are two specific types of drought stress; terminal drought stress occurs when water in the soil declines gradually, eventually causing plant death, whereas intermittent drought stress is caused by several periods of water scarcity throughout

plant development and is not always lethal. Intermittent drought can cause devastating effects on maize yield depending on which development stage it occurs in, with the pollination and fertilization stages at particular risk. At the silking, pollen shed, and grain filling stages, yield can decrease yield by 3-4%, 8%, and 2-6% per day of drought stress respectively (Jain et al. 2019). In major maize production developed countries, such as China and the USA, there is a need to ensure limited plant development under intermittent drought conditions does not affect the high yields expected for economic growth (Boyer et al. 2013; Neumann 2008; Yu 2011).

Various strategies have been employed in crops, such as maize, to mitigate the effect of drought, through knowledgeable breeding using plant physiology, molecular genetics and molecular biology research (Cattivelli et al. 2008). However, due to the damaging effects of drought, along with its worldwide prevalence, there is a need to increase yield of crops, such as maize, through the adaptability of crops to unpredictable stresses.

1.4   Adaptation to drought

Creating adaptable varieties of maize is important for food security, and has been important since its domestication. During the maize domestication process, it was originally thought that genetic diversity was lost fairly rapidly through domestication bottlenecking, thereby reducing the adaptability of maize plants. However, recent evidence counters this by suggesting that any losses in genetic diversity are due to post-domestication erosion, possibly caused by cropping regimes or serial founder events (Allaby, Ware & Kistler 2019; Smith et al. 2019; Wang et al. 2017), and as a result it is actually current farming practices that are gradually reducing crop diversity.

This reduction in diversity was only tolerated for crop production when technology and weather patterns negated the lack of adaptability of crops to environmental pressures. This has been the case for maize growth in the Midwest of the USA, where the majority of USA maize is grown, and as a result maize varieties used there are actually more sensitive to drought than in the past (Lobell et al. 2014). This is despite

the increase in yield in maize, year on year since the mid-20th century (FAO 2019), which is likely due to the implementation of hybrid maize (Duvick 2001). However, this increase cannot be maintained through climate change without adapting the very plants themselves (Challinor et al. 2016). Thankfully, there are several known drivers of adaptation in the study of DNA, two of which we focus on in this research; the epigenetic process of cytosine methylation, and the activation of TEs.

1.5   Cytosine methylation

Whilst genetics relates to the study of how direct changes to a genome causes heritable expression changes in genes, epigenetics focuses on stable expression changes that are not associated with changes in the genome DNA sequence. Cytosine methylation, hereby referred to as methylation, is one such epigenetic strategy employed by most eukaryotic organisms, used to regulate gene expression in chromosomal DNA (Chan, Henderson & Jacobsen 2005; Law & Jacobsen 2010; Weber et al. 2007). What makes methylation so interesting in terms of adaptability is that increased levels of DNA methylation in promoter regions are linked to the inactivation of corresponding genes, and so can provide a short term response to external stimuli (Candaele et al. 2014).

Methylation is the addition of a methyl group to the DNA base cytosine, and can be split into three different contexts: CG, CHG, and CHH (where H represents either an A, C, or T). Mammal DNA methylation is mainly found as CG, whereas in plants all three contexts are common (Henderson & Jacobsen 2007). The differing symmetries of the methylation contexts changes the heritability of each, and as such they focus on different areas of the genome. CG and CHG methylation in maize are concentrated around the repetitive, intergenic regions, whereas CHH methylation is often found near the genic regions (Gent et al. 2013; Li et al. 2015). As a result of this proximity, CHH methylation is has become integrated into gene regulation responses in the maize genome (Bartels et al., 2018). The difference between where the contexts are likely to be found in the genome also causes a difference in the genome coverage percentage for each context, with CG, CHG and CHH contexts found to methylate

86%, 74% and 5.4% of the genome respectively (Gent et al., 2013). Methylation is also known to be inherited in plants and is a well-maintained and safeguarded pathway, aimed at ensuring that methylation patterns are maintained to the benefit of future offspring (Hofmeister et al. 2017; Law & Jacobsen 2010; Williams & Gehring 2017).

The control of DNA methylation differs between plants and animals. In plants, the three different methylation contexts are regulated using a collaboration of methyltransferases. In *Arabidopsis*, CG methylation is maintained by METHYLTRANSFERASE 1 (MET1) (He, Chen & Zhu 2011; Zhang, Lang & J.-K. Zhu 2018), CHG methylation is maintained by CHROMOMETHYLASE 3 (CMT3) and CHROMOMETHYLASE 2 (CMT2) (Lindroth et al. 2001; Stroud et al. 2014), and CHH methylation is maintained by CMT2 or DOMAINS REARRANGED METHYLASE 2 (DRM2) depending on the genomic region being maintained (Stroud et al. 2014; Zhang, Lang & J. K. Zhu 2018).

1.6   The effects of stress on methylation

Historically, there is evidence linking stress with differing methylation levels in crops. An interesting historical example of this is in barley, where Barley Stripe Mosaic Virus caused an increase in methylation levels from samples dating as far back as 1100-1400 C.E. (Smith et al. 2015). In more modern crops, there is evidence to suggest that salt and cold stresses have caused difference in methylation levels (Konate et al. 2018; Steward et al. 2002; W. Wang et al. 2011). However, there is also evidence in the B73 variety of maize that UV, heat and cold stress do not consistently change DNA methylation patterns (Eichten & Springer 2015), although we are interested in drought stress in more than one variety of maize.

1.7   The effect of drought stress on methylation

There has been plenty of research showing there are directed methylation patterns after drought stress in drought tolerant varieties of rice (W.-S. Wang et al. 2011; Zheng et al. 2013, 2017), tomato (González, Ricardi & Iusem 2013), *Populus* (Liang et

al. 2014), and *Arabidopsis* (Colaneri & Jones 2013), therefore the next logical step is to observe if this is the case in maize. Previous research in other higher plants, such as rice (Li *et al.*, 2019), has also shown that these specific methylation patterns found in drought tolerant varieties may in fact confer drought tolerance after drought stress through changes in gene expression. This was the reasoning behind determining if drought stress in maize varieties with differing tolerances to drought would methylate different gene promoter regions in this experiment.

There is another reason that looking at methylation in drought could be important. Methylation is also known to silence transposable elements (TEs) (Diez, Roessler & Gaut 2014; Okamoto & Hirochika 2001; Slotkin & Martienssen 2007), which make up a large percentage of the maize genome, as mentioned below. Unlike mammals, methylation in plants mostly occurs in the TE and repetitive element fractions of the genome, thereby playing a role in the regulation of these regions, and also potentially changing the DNA sequence itself through TE-induced changes (Zhang et al. 2006). This regulation is important for the genome as methylation protects it against over-expression of potentially problematic DNA, such as TEs (Yoder, Walsh & Bestor 1997). This potential difference in methylation caused by drought stress can affect the regulation of transposable element activation, which is a major driver of adaptation.

Therefore, methylation can change short-term expression of important genes related to external stimuli, and create long-lasting effects on the genome, by the insertion TEs into new genomic regions. These insertions have the potential to change gene functions, and as such this process is of the utmost importance when considering adaptation to environmental stresses.

1.8   Transposable Elements

TEs are repetitive, mobile DNA sequences found in the genomes of all plant and animal species (Muñoz-López & García-Pérez 2010), that can potentially make up >60% of the human genome (de Koning et al. 2011), 20% of the *Drosophila melanogaster* genome (Barrón et al. 2014) and 15% of the *Arabidopsis thaliana* genome (de la Chaux et al. 2012). TE research began with Barbara McClintock's 1950

Nobel Prize winning paper researching maize (McClintock 1950). It is no coincidence that the discovery of TEs was found in maize as they make up around 85% of the entire maize genome (Schnable et al. 2009). Barbara McClintock was also the first person to propose that stress-induced TE reactivations could help an organism adapt to a new environment, through beneficial insertions (McClintock 1984). TEs were initially thought to only have disadvantageous effects on gene function (Orgel & Crick 1980; Slotkin & Martienssen 2007) and were previously labelled as "junk DNA" (Ohno 1972), however TEs are now known to play a crucial role in species-specific genomic adaptation through their responses to environmental stresses (Makarevitch et al. 2015; Miousse et al. 2015). Their ability to self-replicate and jump from one genome position to another can cause changes in the host genome, which can alter gene expression, induce chromosome repositioning and create expanded genomes (Gao et al. 2015). TE insertions can also reshape alleles by adjusting their reading frame or splice pattern, thereby affecting the genomic structure of an organism, again showing their capacity for adaptation (Makarevitch et al. 2015). All this means that TEs have the potential to transform genomic regions either positively or negatively, thereby showing why they are so often linked with adaptation (Daron et al. 2014; Lai et al. 2017; Song & Cao 2017). TE activity in a host genome drives evolution at a higher rate, and with greater phenotypic diversity, than other genetic processes such as point mutation, endosymbiotic gene transfer, polyploidy, endosymbiosis, horizontal gene transfer and short tandem repeat slippage (Oliver & Greene 2009), giving them a greater importance when studying evolution and adaptability.

TEs can be separated into two different, distinct major classes due to their method of transposition, with each class divided into orders and superfamilies (Wicker et al. 2007) (Figure 1.2.)

Classification		Structure	TSD	Code	Occurrence
Order	**Superfamily**				
Class I (retrotransposons)					
LTR	*Copia*	GAG AP INT RT RH	4-6	RLC	P,M,F,O
	Gypsy	GAG AP RT RH INT	4-6	RLG	P,M,F,O
	Bel-Pao	GAG AP RT RH INT	4-6	RLB	M
	Retrovirus	GAG AP RT RH INT ENV	4-6	RLR	M
	ERV	GAG AP RT RH INT ENV	4-6	RLE	M
DIRS	*DIRS*	GAG AP RT RH YR	0	RYD	P,M,F,O
	Ngaro	GAG AP RT RH YR	0	RYN	M,F
	VIPER	GAG AP RT RH YR	0	RYV	O
PLE	*Penelope*	RT EN	Variable	RPP	P,M,F,O
LINE	*R2*	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P,M,F,O
	I	ORF1 APE RT RH	Variable	RII	P,M,F
SINE	*tRNA*		Variable	RST	P,M,F
	7SL		Variable	RST	P,M,F
	5S		Variable	RSS	M,O
Class II (DNA transposons) - Subclass 1					
TIR	*Tc1-Mariner*	Tase*	TA	DTT	P,M,F,O
	hAT	Tase*	8	DTA	P,M,F,O
	Mutator	Tase*	9-11	DTM	P,M,F,O
	Merlin	Tase*	8-9	DTE	M,O
	Transib	Tase*	5	DTR	M,F
	P	Tase	8	DTP	M,F
	PiggyBac	Tase	TTAA	DTB	M,O
	PIF-Harbinger	Tase* ORF2	3	DTH	P,M,F,O
	CACTA	Tase ORF2	2-3	DTC	P,M,F
Crypton	*Crypton*	YR	0	DYC	F
Class II (DNA transposons) - Subclass 2					
Helitron	*Helitron*	RPA Y2 HEL	0	DHH	P,M,F
Maverick	*Maverick*	C-INT ATP CYP POL B	6	DMM	M,F,O

Structural features

→ Long Terminal Repeats ▶ ◀ Terminal Inverted Repeats ▬ Coding Region — Non-coding Region

▬ Diagnostic feature in non-coding Region —/ /— Region that can contain one or more additional ORFs

Protein coding domains

AP Aspartic proteinase	APE Apurinic endonuclease	ATP Packaging ATPase	C-INT C-integrase	CYP Cysteine protease	EN YR Endonuclease
ENV Envelope protein	GAG Capsid protein	HEL Helicase	INT Integrase	ORF Open reading frame of unknown function
POL B DNA polymerase B	RH RNase H	RPA Replication protein A (Only found in plants)	RT Reverse transcriptase
Tase Transposase (* with DDE motif)	YR Tyrosine recombinase	Y2 YR with YY motif

Species groups

P Plants   M Metazoans   F Fungi   O Others

Figure 1.2. Classification system for transposable elements taken from Wicker *et al*. (2007).

Retrotransposons, or Class I elements, use a "copy-and-paste" approach, meaning the original TE is transcribed into an RNA intermediate, which is then reverse transcribed into DNA and pasted into a new genomic position (Fattash et al. 2013; Wessler 2006). DNA transposons, or Class II elements, use a "cut-and-paste" strategy, which means the single or double stranded DNA containing the TE is

excised from the original position and inserted into a new location, without an RNA intermediate (Fattash et al. 2013; Wessler 2006). There are also many subgroups found within each of these two classes, and these are split into autonomous or non-autonomous TE families, depending on if the enzymes required for transposition are produced by the TE itself or if it uses another TEs enzymes to transpose (Fattash et al. 2013).

1.9   The regulation of TEs through methylation

TEs are highly regulated by siRNA, modification of histone tails, chromatin packing alterations and DNA methylation (Slotkin & Martienssen 2007). Methylation plays an important role in regulating TE activation in higher plants (Ito & Kakutani 2014), for example, in maize it has been demonstrated that a TE named *Spm* is regulated through DNA methylation (Schläppi 1994). With this in mind, environmental stressors have proven themselves important by changing methylation states in many plant and mammal genomes, thereby initiating TE expression (Miousse et al. 2015). It is also theorised that osmotic stress of maize can induce the methylation of TEs (Tan 2010), which may affect TE activity throughout the plant. In some cases, these methylation changes are directed in tolerant varieties. An example of this is in drought tolerant rice, where there is an increase in DNA methylation of proximal TEs near drought stress-responsive genes in low water conditions (Garg et al. 2015) possibly due to the upregulation of TE-derived miRNAs (Barrera-Figueroa et al. 2012). In apple plants, water stress also causes widespread DNA methylation responses in TEs (Xu et al. 2017). Therefore, the link between methylation and TE control is of particularly interest for this thesis.

1.10   The effect of drought stress on TE activity

Recent studies have also elucidated some links between stress and TE activity in many higher plants (Horváth, Merenciano & González 2017; Negi, Rai & Suprasanna 2016). This occurs with many types of abiotic stresses, examples include UV light stress inducing *Ty1-copia* retrotransposons in Oat plants (Kimura et al. 2001) and heat stress inducing the *ONSEN-copia* retrotransposon in *Arabidopsis* (Ito et al.

2011). However, this thesis will focus on drought stress. Drought stress on wheat lacking the D genome causes the activation of TEs (Aprile et al. 2009), and TEs in coffee plants (*Coffea* sp.) are also differently activated through changes in drought stress levels due to irrigation conditions (Lopes et al. 2013). Therefore, any information about the possible link between TE activity and environmental stress is crucial when considering the maize genome's evolution and function (Diez et al. 2014). There appears to be a link in maize as TE activity is induced under high salt, cold (5°C), heat (50°C) and UV abiotic stress conditions (Makarevitch et al. 2015). There is further evidence to support this link in maize, as the interposition of particular miniature inverted-repeat transposable element (MITE) into the promoter of the NAC gene, *ZmNAC111*, was associated with maize drought tolerance (Mao et al. 2015).

1.11 Adaptation driven by TE activation and methylation

Of course, as we are discussing adaptability there is a requirement to discuss TEs in relation to genes, and in maize this is especially important as the majority of genes are located within 1 kb of an annotated TE (Baucom et al. 2009; Schnable et al. 2009). In fact, many TE families target areas of the genome near genes, such as the *Mutator* TEs (Dietrich et al. 2002; Sultana et al. 2017): euchromatin appears to be the preferred substrate for most TEs, possibly due to its open state and its affiliation with actively transcribed genes (Liu et al. 2009). This close proximity can potentially cause nearby genes to be affected by TE activity, therefore forcing an adaptable evolutionary change in the genome, saving the plant phenotypically from the environmental stress. In fact, there is evidence showing that the specific regulation role a TE plays in gene regulation differs depending on where in the chromosome the TE is located (Sigman & Slotkin 2016). TEs found near genes, for example, are known to affect the expression of genes. An example of this is ROS1 gene expression which is controlled by a *Helitron* TE found within a promoter region in *Arabidopsis* (Lei et al. 2015; Williams et al. 2015). In other genomic positions, such as pericentromere and knobs, TEs are kept silenced, with the only occasions when this does not occur is when the silencing mechanism is overcome (Sigman & Slotkin 2016). TEs also have a profound effect in creating new genes in a host genome through "exon shuffling"

(Long et al. 2003), as observed in maize, where 60% of *Helitron* TE elements have captured gene fragments within them (Yang & Bennetzen 2009). There have already been studies showing how past TE activation events in maize have created new chimeric functional genes through transduction (Elrouby & Bureau 2010) and how alternative transposition led to a rapid evolution event in the maize species (J. Zhang et al. 2014). Retrotransposons in particular are highly targeted in maize by siRNA signifying that they may still remain highly active (Diez et al. 2014). Different genotypes may show different patterns in response to drought stress, which is why it is imperative that different varieties of maize are studied instead of using the reference B73 genome as a substitute for all maize. In future this knowledge can be used to selectively breed specific traits in maize, as efficient genetic enhancement in a species requires an in-depth understanding of the regulatory mechanisms of gene expression under drought stress (Y. G. Wang et al. 2014).

1.12   Drought tolerance genes in maize

Drought tolerance genes may be affected by TE activity and methylation after drought stress in different maize varieties. These effects may include silencing and expression of important genes for drought tolerance, or the change in sequence caused by a TE insertion. With this in mind, there is a need to highlight potentially important drought tolerant genes that may be targeted by the plant. A genome-wide association study carried out using 367 maize varieties found that there were 42 candidate genes that were significantly associated with drought tolerance at the seedling stage (Xianglan Wang et al. 2016), and so these would be of particular interest for this study when looking at TE insertions or methylation changes caused by drought stress.

One such drought tolerance candidate allele found in the same study was for the gene coding for vacuolar H+ pyrophosphatase, *ZmVPP1*. In maize, a 366-bp insertion in the *ZmVPP1* promoter region was found to be induced by drought, and increased drought tolerance in the plant both naturally and transgenically (Xianglan Wang et al. 2016). Proline accumulation is known to confer drought stress tolerance in plants, acting as a reactive oxygen species (ROS) scavenger, a protein stabilizer, and as a

stress signalling molecule (Hayat et al. 2012; Verbruggen & Hermans 2008; Wei et al. 2009). Studies conducted in transgenic tobacco and rice observed elevated proline levels in each organism, caused by an overexpression of the delta1-pyrroline-5-carboxylate synthetase (P5CS) gene, which consequentially increased the tolerance of the plant to drought stress (Choudhary, Sairam & Tyagi 2005). The P5CS gene from *Arabidopsis thaliana* and rice were separately inserted into petunia plants causing an accumulation of proline, which again improved drought stress tolerance in the transgenic plants (Yamada et al. 2005). In fact, there is evidence showing that proline accumulation may have a positive effect on growth during drought conditions in maize (Ali, Ashraf & Athar 2007). Therefore, the proline pathway could be one potentially interesting area to analyse in the maize genome after drought stress.

Another interesting region of the genome for drought stress effects is the abscisic acid (ABA) pathway, which is overexpressed in a number of plants, such as wheat (Wei et al. 2015), switchgrass (Aimar et al. 2014) and grapevine (Ferrandino & Lovisolo 2014) during drought stress. Previous research shows that ABA has a plenitude of downstream processes that cause antioxidant defence (Ding et al. 2013) and stomatal closure (Lim et al. 2015), both of which protect the plant when experiencing a water deficit. There is already evidence showing that ABA does accumulate in maize during drought stress (Jiang & Zhang 2002), but more work is needed to visualise the genomic differences related to this pathway between drought tolerant and drought susceptible varieties. There are also a number of proposed drought tolerance genes associated with the accumulation of ABA discovered through research into the regulation of maize miRNAs. Downregulated miRNAs that cause accumulation include Phospholipase D, mitogen-activated protein kinase and peroxidase, whereas there is an upregulated miRNA that has been theorised to cause under expression of the proline dehydrogenase (PDH) gene causing an accumulation of proline (Wei et al. 2009). Therefore, there are genomic regions that are of particular interest in this thesis, as well as potentially new regions only found within the varieties chosen for these experiments.

1.13 Scope and Project Aims

From the evidence presented above, we predict that drought stress conditions will have an effect on TE activity and this will in turn affect the adaptability of maize to drought. Therefore, the study of genomic mechanisms in maize due to drought stress and how it links to the evolution of maize is very important for improving future crop yield. With this is mind, there are several aims of this project:

1. To test if there is a difference in methylation in gene promoter regions caused by drought stress in maize.
2. To determine the difference between this response in drought tolerant and drought susceptible varieties of maize, and specific methylation changes found near genes related to drought tolerance.
3. To ascertain methylation differences found in the TE fraction of maize caused by drought stress in drought tolerant and drought susceptible varieties of maize.
4. Using these differences, to deduce if specific TE families show significant differences, and if so, which TE families are also found in the differently methylated promoter regions.

The ultimate aim of this project is to elucidate the role methylation has on the drought tolerance of maize. Information relating to methylation and TE activation does show that they play a role in the production of adaptive varieties of crops, therefore this work will be used to inform future scientists on the role methylation plays in maize, and to show the methylation response of specific drought tolerant varieties to drought stress. This in turn will help future generations understand the maize varieties needed to selectively breeding for adaptive maize, something that is essential for future food security (Watson 2019).

To achieve the aims, we will first compare and contrast methylation levels in promoter regions of drought susceptible and drought tolerant maize after drought stress using bisulfite sequencing data and statistical analysis. Results will elucidate the promoter methylation drought response differences between the varieties. Next,

we will use the same data to observe methylation differences found within the TE fraction of each variety. This can be compared with the results from the previous chapter to allow us to speculate which genes have differentially methylated TEs within their differently methylated promoter regions, giving a list of interesting genes for further study. These methods and results will answer all aims and provide substantial data for future work looking into epigenomic changes caused by drought stress.

2   General Methods

2.1   Maize varieties chosen

Table 2.1. A list of the two maize varieties chosen for the drought stress experiments.

GRIN Accession Number	Drought tolerant variety	Developed at	Pedigree	Drought Susceptibility
PI 550483	B76	Iowa State University	(CI31A x B37^2)B37	Tolerant
PI 550473	B73	Iowa State University	C5	Susceptible

2.2   Maize growth

The phytobiology facility at the University of Warwick provided the environment necessary to grow the maize plants. Plants were grown at 28°C day/ 20°C night in a 16-hour light/8-hour dark cycle with a light intensity of 230 μE m-2 s-1. Seeds were germinated in three-inch diameter pots containing peat-based soil (Costa et al. 2012). Soil water content was measured at regularly occurring intervals during drought stress using a Professional Soil Moisture Meter (Lutron Electronic Enterprise Co., LTD., Taipei, Taiwan) to ensure drought conditions were successfully applied. All plant material was sampled between 11.00 AM and 1.00 PM, to avoid diurnal variation in gene expression that would obscure the effect of stress. Samples were taken from the second leaf once the plant had reached the three-leaf stage. This was chosen as the three-leaf stage is the typical starting point of maize maturity and therefore can be subjected to stress without affecting development. These samples were immediately frozen in liquid nitrogen and stored at -80°C.

2.3   CTAB method

Samples were crushed into a powder using liquid nitrogen and a mortar and pestle. This powder was scraped into an Eppendorf tube. 500 μl of 2% CTAB buffer (2%

cetyltrimethylammonium bromide, 1% polyvinyl pyrrolidone, 100 mM Tris-HCl, 1.4 M NaCl, 20 mM EDTA) was added to the tube and vortexed. This was left in a water bath at 65°C for 1 hour. An equal volume (500 µl) of chloroform was added and again the tube was vortexed. Tubes were centrifuged at 25,000 x g for two minutes with the upper phase being transferred to a new tube. 1.5x volume of AW1 wash buffer (Qiagen, Hilden, Germany) to the sample, this was vortexed and all of the solution was spun through a spin column at 6,000 x g. This spinning through the column was then repeated with 500 µl AW2 wash buffer (Qiagen, Hilden, Germany) and then 300 µl acetone. The column was then placed into a new tube and allowed to dry off any excess acetone. 100 µl of AE buffer (Qiagen, Hilden, Germany) was deposited onto the membrane. The column was spun one last time at 20,000 x g to obtain the AE buffer containing the extracted DNA.

2.4   Qubit

Qubit solution was created using a ratio of 199:1 Qubit dsDNA HS Buffer to Qubit dsDNA HS Reagent. 198 µl of this Qubit solution was added to 2 µl of the AE buffer containing the extracted DNA in a Qubit tube and vortexed. Standards were created by adding 190 µl of the Qubit solution to 10 µl of each corresponding standard solution provided and vortexing for 2-3s. 200 µl of the Qubit solution was used as a negative control. The standard mixture was used to calibrate the Qubit Fluorometer for the 'DNA HS' option. Sample and control mixtures were then measured using the same option, with the DNA concentration noted.

2.5   EZ DNA Methylation Gold

CT Conversion Reagent was prepared by mixing 900 µl water, 300 µl of M-Dilution Buffer, and 50 µl M-Dissolving Buffer to a tube of CT Conversion Reagent. 130 µl of the CT Conversion Reagent was then mixed with 20 µl of the previously extracted DNA in a PCR tube. This mixture was placed in a thermal cycler and run as follows: 98°C for 10 minutes, 64°C for 2.5 hours and held at 4°C. Both the DNA sample and 600 µl of M-Binding Buffer were transferred to a Zymo-Spin IC Column in a collection Tube and spun at 10,000 x g for 30 seconds. Flow-through was discarded, 100 µl of

M-Wash Buffer was added to the column and this was centrifuged as before. 200 μl of M-Desulphonation Buffer was added to the column and this was left to incubate at room temperature for 15 minutes, then spun as before. Another two wash steps using 200 μl of M-Wash Buffer was applied to the column, and spun as before each time. The column was inserted into a 1.5 ml microcentrifuge tube and 10 μl of M-Elution Buffer was pipetted directly onto the column matrix. The column with new tube was then centrifuged for 30 seconds at 10,000 x g to elute the converted DNA.

2.6   TruSeq DNA Methylation Library Prep Kit

50 ng of the bisulfite converted DNA was used for each sample in this protocol. 9 μl of sample DNA was mixed with DNA Synthesis Primer in a PCR tube, and this mixture was run in a thermal cycler at 95°C for 5 minutes to anneal the synthesis primer. 4 μl TruSeq DNA Methyl PreMix, 0.5 μl 100 mM Dithiothreitol and 0.5 μl TruSeq DNA Methyl Pol were added to each sample and this was run on a thermocycler at 25°C for 5 minutes, 42°C for 30 minutes, 37°C for 2 minutes and held at 4°C. 1 μl of Exonuclease I was added to each sample and the samples were run at 37°C for 10 minutes, 95°C for 3 minutes and 25°C for two minutes. Tagging the DNA involved adding 7.5 μl TruSeq DNA Methyl Term Tag PreMix and 0.5 μl DNA Polymerase to each sample and running at 25°C for 30 minutes, 95°C for 3 minutes and held at 4°C. To clean the tagged DNA 40 μl 1.6x AMPure XP beads were mixed with each sample and left to incubate for 5 minutes. The tubes were then placed on a magnet for 5 minutes and all supernatant was removed. Two washes using fresh 80% ethanol were done and all ethanol was removed. 24.5 μl of nuclease-free water was added to each sample and left to incubate for 2 minutes. This was again placed on a magnetic and allowed to sit until the liquid became clear. The resulting supernatant contained the di-tagged DNA. A 10-cycle PCR program was used to amplify the library. This was achieved by adding 25 μl FailSafe PCR PreMix E, 1 μl TruSeq DNA Methyl Forward, 0.5 μl FailSafe PCR Enzyme Mix and 1 μl of the Index PCR Primer to each sample and running at 95°C for 1 minute, then 10 cycles of: 95°C for 30 seconds, 55°C for 30 seconds and 68°C for 3 minutes; after the 10 cycles the samples were run at 68°C for 7 minutes and held at 4°C. The library was cleaned using 50 μl 1x AMPure XP beads as before but this time the libraries were eluted in only 20 μl of nuclease-free water.

3 Promoter regions are differentially methylated after drought stress within drought tolerant maize

3.1 Introduction

Methylation is an epigenetic process that regulates gene expression throughout the living world and is observed in many eukaryotes (Chan, Henderson & Jacobsen 2005). During the physical mechanism of methylation, genomic cytosines gain a methyl group causing a stable, yet reversible, heritable change of state (Elhamamsy 2016). This methylated state allows the genomic regions to be tightly bound around histones, thereby inhibiting transcription through interference with promoter binding (Bernatavichute et al. 2008; Zilberman et al. 2007). In plants, DNA methylation-guided regulation of gene expression is crucial for coordinating organism-wide developmental programs, adapting to environmental stresses and silencing transposable elements (TEs) (Dowen et al. 2012; Zhang, Lang & J.-K. Zhu 2018).

Cytosine methylation comes in three different contexts: the symmetrical CG and CHG and the asymmetrical CHH (where H represents A, C or T) (Law & Jacobsen 2010). The symmetry of each methylation context plays a part in how the heritability of each methylation event is maintained, which also means each methylation context is regulated by their own pathways and are primarily found in different genomic regions. In maize specifically, CG and CHG methylation is often localised to the intergenic regions, whereas CHH methylation is enriched in short CHH islands near genes and conserved noncoding sequences (Gent et al. 2013) (Li et al. 2015). This localisation of CHH methylation is likely explained by the integration of TEs into the gene regulation process due to their abundance in the maize genome (Bartels et al. 2018). The focus of symmetrical methylation to intergenic regions, coupled with the large TE fraction of the maize genome, correlates with the higher percentage of methylation overall in maize, as the genomic averages for CG, CHG and CHH methylation are 86%, 74% and 5.4% respectively (Gent et al. 2013).

The activation of methylation events after drought stress is a well-known phenomenon found in many plant species, such as rice (Gayacharan & Joel 2013; W.-S. Wang et al. 2011) and wheat (Kaur, Grewal & Sharma 2018). Specifically, research in higher plants shows that methylation changes within promoter regions cause silencing or activation of corresponding genes (Li et al. 2012; Zilberman et al. 2007). This can thereby have an effect on the tolerance and adaptability of the plant (Baek et al. 2011; Fei et al. 2017; M. Wang et al. 2014). As these methylation changes are affected by the external environment and can change at any time, this makes methylation an interesting area to research when thinking about how maize plants become adapted to drought stress. Recognising this fact, our aim was to observe the effect drought stress has on methylation activity in maize, and therefore elucidate new potential drought tolerance candidate genes. Due to the different proportions and different focuses of each methylation context, we also decided to look at each context individually and as a whole to find patterns throughout this experiment. To achieve this aim, we compared and contrasted methylation levels in the promoter regions of control and drought stressed tissues in drought tolerant and drought susceptible varieties of maize, giving an indication of which genes were potentially being regulated.

The two maize varieties chosen for this project were a drought tolerant variety, B76, and a drought susceptible variety, B73. B73 is used as the reference genome for maize (Schnable et al. 2009), and is known to create a globally important variety of edible maize when crossed with Mo17, used for animal and human consumption across the world (Ranum, Peña-Rosas & Garcia-Casal 2014). B76, on the other hand, is relatively unstudied as a variety, but is known to be drought tolerant from previous growth experiments, and was shown to have a similar growth pattern to that of B73, in terms of time from seed to three-leaf stage. The source of the B76 tolerance is not known, only that it is able to maintain high water content in its leaf after drought stress (Chen et al. 2012). As such, this was deemed an ideal variety to screen for drought tolerance genes controlled by methylation.

Methylation levels are tissue-specific in maize and previous research has recommended that separating tissue data is important when comparing inter-individual methylation variation (Lauria et al. 2017; Lu, Rong & Cao 2008). In particular, heavily methylated upstream CHH islands are known to be associated with tissue-specific up-regulation of nearby genes in maize (Gent et al. 2013; Hsu et al. 2017). In theory, methylation differences of these CHH islands may play a part in this genic regulation, and as a result, it was decided that tissue types should be separated and each tissue response looked at individually. Particularly, in terms of responses to drought, root and leaf were of interest as in other crop plants, such as rice, methylation responses after drought stress, have been observed in both root and in leaf (González, Ricardi & Iusem 2013; W.-S. Wang et al. 2011; W. Wang et al. 2016).

From the literature, it was expected that drought stress would cause a more directed methylation pattern in the drought tolerant variety, as observed in rice (Zheng et al. 2017) (Zheng et al. 2013) (W.-S. Wang et al. 2011), tomato (González, Ricardi & Iusem 2013), *Populus* (Liang et al. 2014), and in *Arabidopsis* (Colaneri & Jones 2013), although this difference in *Arabidopsis* did not change drought responsive gene expression (D. Ganguly et al. 2017). Previous research in other higher plants suggests drought tolerant maize will likely methylate and de-methylate areas of the genome that will help the plant during water scarcity, such as the jasmonic acid (JA) or abscisic acid (ABA) pathway found to be differentially methylated in rice (Li *et al.*, 2019). As such the promoter regions of genes associated with these pathways were of particular interest during this project. Oppositely, the drought susceptible variety was not expected to have this survival 'instinct' against drought stress and as a result would likely methylate and de-methylate haphazardly across the genome using a fight or flight strategy to prevent death. Therefore, the aim was also to determine whether there was a difference in methylation pattern between the two varieties, in leaf and root tissues, across the CG, CHG and CHH methylation contexts.

3.2   Methods

3.2.1   Maize growth

Two varieties of maize, B76 and B73, were chosen specifically for their similar growth patterns (according to data taken from a parallel multigenerational growth experiment) and differing resistance to drought (Chen et al. 2012). Maize plants were grown in the phytobiology facility at the University of Warwick in conditions seen in section 2.2 of the methods chapter.

Figure 3.1. Picture of B73 and B76 maize, grown in three-inch diameter pots at the Phytobiology facility at the University of Warwick.

Three control plants and three plants subjected to drought conditions were grown for both the drought resistant B76 and the susceptible B73 varieties. At the three-leaf stage water was withheld from the chosen drought stressed, experimental plants for six days, with soil water content measurements taken to ensure adequate drought conditions were achieved.

Table 3.1. Table showing the Soil Moisture percentage before and after drought stressing in the control and drought-stressed replicates.

Maize Variety	Condition	Replicate Number	Initial Soil Moisture (%)	Final Soil Moisture (%)
B73	Control	1	22.1	17.8
B73	Control	2	22.8	23.9
B73	Control	3	22.6	20
B73	Stress	1	19.3	0
B73	Stress	2	17.6	0
B73	Stress	3	21.8	0
B76	Control	1	22.8	22.5
B76	Control	2	34.7	22.5
B76	Control	3	50+	25.7
B76	Stress	1	19.5	0
B76	Stress	2	23.5	0
B76	Stress	3	19.8	1

Control plants were watered as normal during this period. Six days was chosen as the limit for drought stressing as at this stage B73 looked noticeably stressed, and additional days of stress caused B73 plants to perish.

Figure 3.2. Picture observing phenotypic differences in control plants (left) and drought stressed plants (right) grown past 3-leaf stage at the phytobiology facility at the University of Warwick.

3.2.2   Sample collection

Samples were taken at the end of the six-day drought period. Leaf tissue was taken from the third leaf of each plant, while the root tissue was taken from one of the major seminal roots. Leaf and root samples were taken from the same plant, immediately flash frozen in liquid nitrogen and stored at -80°C.

3.2.3   DNA extraction, bisulfite treatment and next generation sequencing

Samples were pulverized in liquid nitrogen using a mortar and pestle, and DNA was extracted using the CTAB method (See Chapter 2 - 2.3). DNA concentrations were quantified with a Qubit fluorometer (See Chapter 2 - 2.4). Quantified DNA was sent to Admera Health, for bisulfite conversion, library creation and sequencing. DNA from each sample was bisulfite treated using the EZ DNA Methylation Gold kit (Zymo Research, CA, United States; See Chapter 2 - 2.5), and libraries for sequencing were obtained using the TruSeq DNA Methylation Library Prep Kit (Illumina, CA, United States; See Chapter 2 - 2.6). 2 x 150bp paired-end sequencing using the Illumina HiSeq X10 instrument was performed on bisulfite converted libraries. Read counts

generated are found in Appendices Table 6.1.1.1. Data obtained was used for analysis.

3.2.4   Data analysis

3.2.4.1   Raw data preparation

The initial per base sequence quality of the data was low, therefore sequencing was redone, resulting in two files. A quality check was performed using FastQC (Babraham Institute, Cambridge, UK), to ensure data was above a quality score of 20.

3.2.4.2   Trimming and mapping

Repeat sample FASTQ files were merged and Trim Galore (Babraham Institute, Cambridge, UK) was used to trim adapter sequences and remove reads below a quality Phred score of 20. The options used were --paired for paired end data and --FastQC to visualise quality scores after trimming.

Data was then mapped to the maize reference genome, B73 RefGen_v3, using Bismark (Krueger & Andrews 2011). Bowtie2 options chosen within Bismark were -q --score-min L,0,-0.2 --ignore-quals --no-mixed --no-discordant --dovetail --maxins 500. Mapping efficiency ranged from 31.8% to 47% based on the minimum alignment score of 0 + -0.2 * x, where x is the read length. Resulting files were subjected to deduplicate_bismark to discard all additional reads aligning to the same position at the same orientation. Final coverages for all samples ranged from 0.8X to 2.9X, with and average coverage of ~1.3X.

Bisulfite conversion rate was calculated using the bam files created during the mapping process. Reads that mapped to the chloroplast, with a MAPQ score of 20 or higher were filters from the main bam files for each sample, and the percentage of methylation calls compared with the number of opportunities for methylation was calculated. The use of the unmethylated maize chloroplast to estimate bisulfite conversion rates is a protocol used in other methylation studies (Han et al. 2018; Li et al. 2014). The results were compared against public dataset maize sequencing data

(NCBI BioProject: PRJNA400328) for B97 (SRA run: SRR6201702) and W22 (SRA run: SRR5980380) maize varieties as control samples. The results were as follows:

Table 3.2. Table showing the bisulfite conversion rates of all maize samples.

Maize variety	Tissue	Experimental condition	Replicate number	CpG Conversion Rate	CHG Conversion Rate	CHH Conversion Rate
W22	Leaf	Control	NA	99.6	99.6	99.7
B97	Leaf	Control	NA	99.6	99.6	99.7
B73	Leaf	Stressed	1	95.1	94.5	95.4
B73	Leaf	Stressed	2	97.7	97.1	98.4
B73	Leaf	Stressed	3	85.1	82.8	84
B73	Leaf	Control	1	88.1	85.7	87.4
B73	Leaf	Control	2	93.5	92.9	93.5
B73	Leaf	Control	3	94.2	93.3	94.1
B76	Leaf	Stressed	1	87.5	86.5	86.9
B76	Leaf	Stressed	2	80.3	79	78.8
B76	Leaf	Stressed	3	88.7	87.8	88.5
B76	Leaf	Control	1	86.5	85.2	85.8
B76	Leaf	Control	2	72.7	69.4	70.4
B76	Leaf	Control	3	81.8	80.5	79.9
B73	Root	Stressed	1	70.9	68.7	69.2
B73	Root	Stressed	2	81.3	79.1	80.2
B73	Root	Stressed	3	77.6	76.9	77.9
B73	Root	Control	1	75.5	74.3	76.9
B73	Root	Control	2	87.9	86.7	88.3
B73	Root	Control	3	85	83.4	85.4
B76	Root	Stressed	1	54.9	53.5	54
B76	Root	Stressed	2	33.7	32.2	32.6
B76	Root	Stressed	3	34.7	33.1	33
B76	Root	Control	1	50.9	50.1	49.1
B76	Root	Control	2	55.7	53.3	53.6
B76	Root	Control	3	40.9	39.8	39.6

Bismark's built-in bismark_methylation_extractor module was used to generate the methylation call for every cytosine position analysed. This information was imported into the program SeqMonk (Babraham Institute, Cambridge, UK) for further analysis.

3.2.4.3 SeqMonk analysis

Data was aligned using the annotated AGPv3 maize reference genome downloaded from the SeqMonk servers. Only chromosomes 1-10 were considered for logistic regression as they contain the functional genes that have a known role in drought tolerance. Probes were generated at every read position in the whole dataset, and were then merged into 100 position windows. Average counts per probe were then normalised using SeqMonk's built in Read Count Quantitation program. An outlier cut-off was determined using the formula: (median probe value +/- the interquartile range)*stringency of 10. This stringency was recommended by the Babraham Institute as it matched the dataset distribution. Outliers were removed from each dataset. Replicates were combined into replicate sets representing control or drought stressed tissue and these sets were quantified for methylation by calculating the percentage of methylation for each base and averaging all percentages found in 100-probe windows using the mean value. Probes were then filtered by length and, as per the read length distribution, most statistically significant probes were observed to fall in a normal distribution between 1 and 10 kb.

A logistic regression analysis was carried out between the control and experimental replicate set probes with a Bonferroni-corrected P-value cut-off of 0.05. Promoter regions with significant differences between the two conditions were determined by filtering for probes that were up to 2 kb upstream of a gene. Genes with significant differences in this 2 kb upstream promoter region were noted for further analysis.

Initially SeqMonk 1.40.0 was used to carry out the logistic regression analysis but the updated version (v1.44.0) was used to provide additional methylation difference and false discovery rate information to the annotated probe SeqMonk files. Unlike SeqMonk v1.44.0, v1.40.0 considered absolute differences in methylation percentages and filtered out values with a minimum absolute difference of 5% and below before logistic regression analysis.

As we were interested in the methylation patterns around genes, we visualized methylation in the gene as well as 2 kb regions up- and down-stream of the genes

using the in-built SeqMonk Quantitation trend plot. The quantitation trend plot used differentially methylated region (DMR) probe values to generate a feature profile around all genes in the filtered gene set. Each position in the plot is an average of multiple probes overlapping that position.

Venn diagrams used to visualise the number of genes with DMRs in their promoter region were produced using the *VennDiagram* package in R (Chen, H and Boutros, P, 2013).

3.2.4.4 DMRcaller analysis

Using the R/Bioconductor package *DMRcaller* (Catoni et al. 2018; Zabet & Tsang 2015), DMRs were computed across replicates, within 100bp bin windows, and to a Benjamini and Hochberg adjusted P-value cut-off of 0.05 (Hochberg 1995). This program determined DMRs found in gene promoter regions according to the AGPv3 maize reference. *DMRcaller* performs the Score test for each 100bp bin and determines DMRs based on the statistical significance of the difference in methylation level between the conditions, the threshold value of methylation proportion difference and threshold of the mean number of reads per cytosine. Options selected were; a bin size of 100, a P-value cut-off of 0.05, minimum number of cytosines at 4, minimum proportional difference of 0.4, a minimum gap between DMRs at 150, a minimum size of a DMR at 50 with a minimum number of reads per cytosine at 4. A list of TEs and the filtered gene set was taken from the Gramene database (http://www.gramene.org) and used to define the DMRs located in TEs and genes respectively. 2 kb upstream of each gene was chosen to represent the promoter region of that gene.

3.2.4.5 Gene ontology analysis

Gene ontology (GO) terms for the filtered gene set were obtained via MaizeGDB (http://download.maizegdb.org/maize-GAMER) (Wimalanathan et al. 2018) and GO term names were taken from the Gene Ontology Consortium (http://geneontology.org/). Only genes found in both the logistic regression and

Score test analyses were used for GO analysis. The GO term graph showing the GO terms associated with the "Response to Water Deprivation" GO term was visualised using the Blast2GO program (Götz et al. 2008).

Functional interpretation of the genes with a response to drought stress was achieved using GO term analysis. Genes with promoter regions undergoing hypermethylation were separated from genes undergoing hypomethylation and vice versa. Gene names were converted to maize AGPv4 gene names using personally made Perl scripts and genes that were not present in the updated gene list were removed. Singular Enrichment Analysis (SEA) on significant genes across all methylation contexts found in both analyses was carried out using AgriGO v2.0 (Tian et al. 2017). The reference used for the SEA analysis was Maize v4 (Maize-GAMER), the chi-squared statistical test option was used along with the Hochberg (FDR) multi-test adjustment method. The significance level cut-off was 0.05 and 1 minimum mapping entry was used as the lower cut-off.

Significantly enriched GO terms from the significant genes and their respective SEA p-values were summarised into representative subsets of GO terms using REVIGO (Supek et al. 2011). The options chosen were; medium (0.7) similarity clustering, numbers represent p-values, whole UniProt database and SimRel semantic similarity measure. The resulting semantically similar GO term groups were represented as a treemap based on the absolute $\log^{10}$ of each GO term p-value using the R script created by REVIGO.

3.3   Results

3.3.1   Overall methylation patterns

From quantitation plots focused around the genic regions of all genes, it was observed that in the most contexts, leaf and root tissue showed inconclusive percentages of methylation in the upstream/promoter region under drought conditions when compared to control conditions (Figures 3.1. and 3.2).
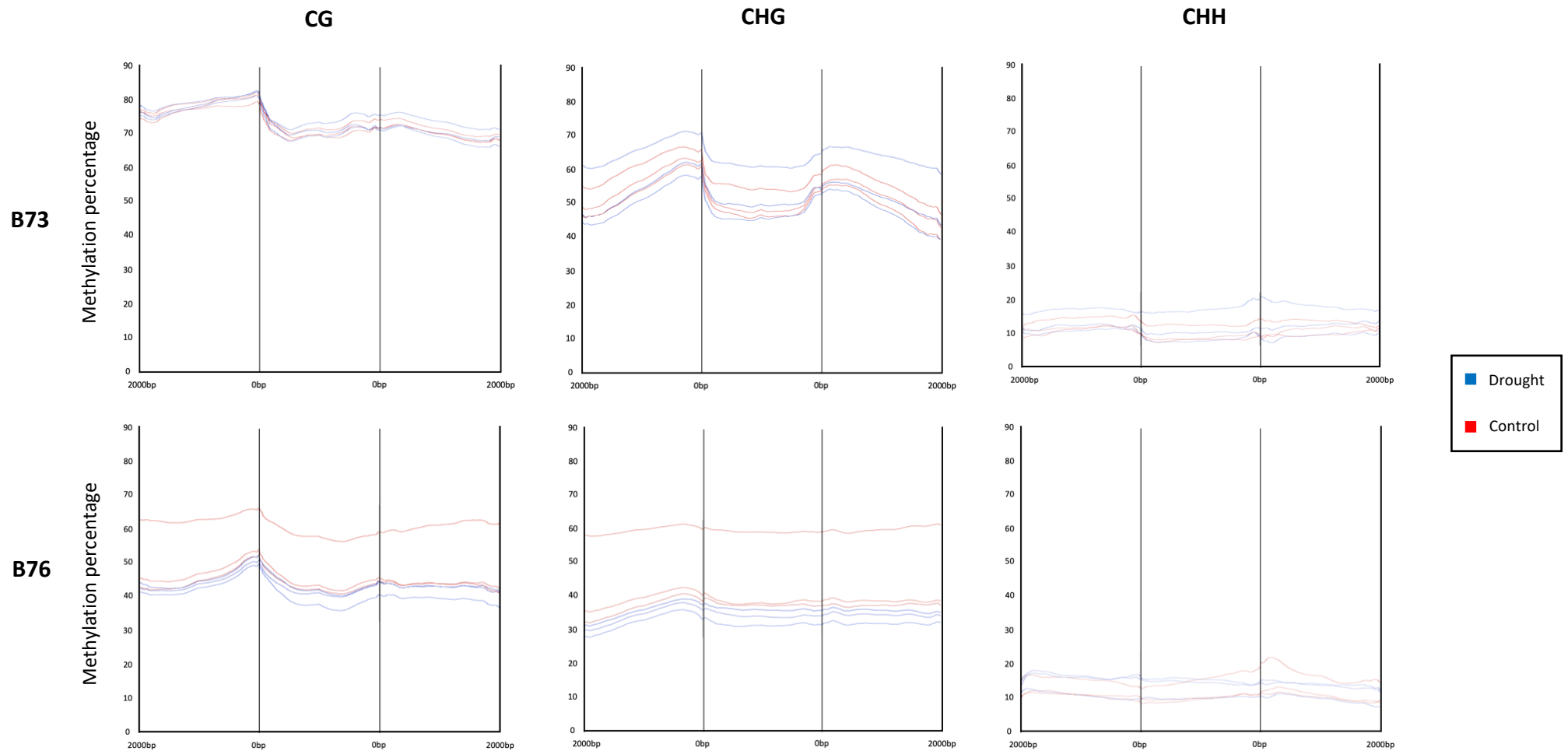
Figure 3.3. Quantitation trend plots depicting whole genome methylation percentages in leaf tissue found around B73 and B76 maize genes.

Methylation contexts were separated into CG, CHG and CHH. The area observed was from 2 kb upstream to 2 kb downstream across all genes. Each line corresponds to a single plant replicate.
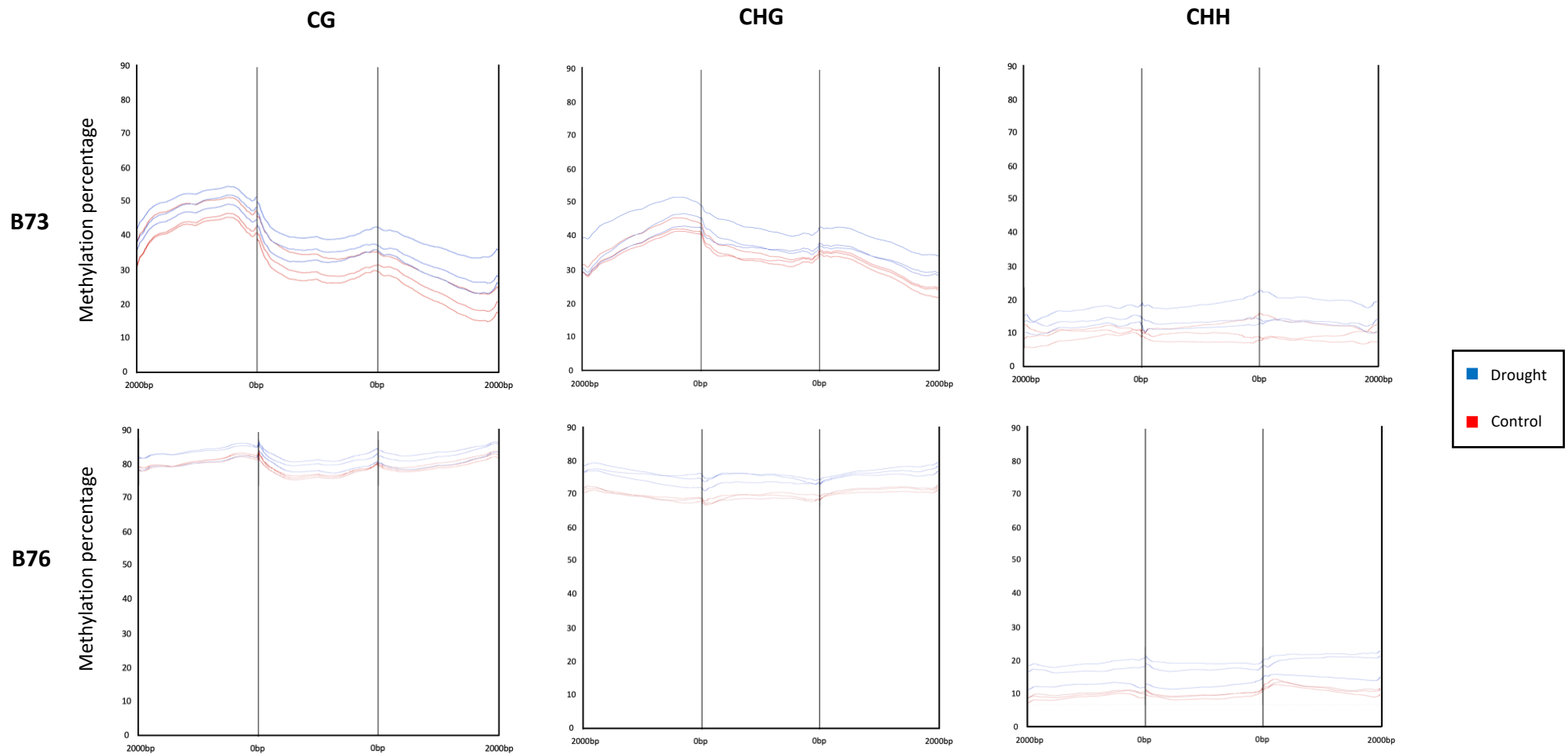
Figure 3.4. Quantitation trend plots depicting whole genome methylation percentages in root tissue found around B73 and B76 maize genes.

Methylation contexts were separated into CG, CHG and CHH. The area observed was from 2 kb upstream to 2 kb downstream across all genes. Each line corresponds to a single plant replicate.

In CG methylation, methylation percentages in the promoter regions of the B73 leaf tissue were higher than the equivalent tissue in the B76. This increase is in spite of a significant outlier in control replicate 2 which showed higher methylation levels than other replicates in that test group. The methylation differences between control and drought stressed tissue were not well pronounced in B73 leaf tissue, whereas in the B76 variety there was a slight decrease in methylation in the drought stressed tissue, with the control replicates all showing greater methylation compared to that of the drought stressed 2 kb promoter regions.

In the root tissue, the CG methylation percentages in the promoter regions were reversed to that of leaf in the same context, with the B73 root showing lower methylation levels than in B76. In B73, there is a slight pattern in the control replicates as two of them showed lower levels of methylation compared to all the drought stressed replicates. This contrasts the B76 variety, where there was a marked increase in methylation in two of three drought stressed replicates compared to the controls.

Tissue specific methylation level differences are found within each variety in the promoter region of genes in the CG context. In B73, leaf methylation levels are higher than root, whereas in B76, leaf levels are lower than root.

In the B73 leaf tissue, CHG methylation levels are greater than in the B76 leaf tissue, although it should be noted that, as with the CG context, control replicate 2 was a significant outlier compared to the other replicates. This is likely due to bisulfite conversion inconsistencies found between samples. Without this replicate there is a clear reduction in methylation of the B76 promoter regions after drought stress. There does not appear to be a difference in methylation patterns between control and drought stressed replicates in B73 leaf, whereas the controls had consistently higher methylation levels than the stressed tissue in B76 leaf.

In root tissue of B73, CHG methylation levels, as with CG methylation, are lower than those in the B76 root. In B73 root tissue, there is a general pattern of increased methylation in the drought stressed replicates, although there was some overlap between control replicates and drought stressed replicates. In the B76 variety, there was a consistent increase in methylation after drought stress, suggesting a multitude of hypermethylation events in gene promoter regions during drought stress events.

As observed in CG, there are tissue specific methylation differences within each variety in the CHG context. Leaf tissue possesses higher methylation levels in B73 when compared to root, whereas in B76, leaf tissue possesses lower methylation levels than root.

In the CHH context, the overall percentage of methylation across upstream promoter regions was lower than in the CG or CHG contexts in corresponding tissues. Within B73 leaf, there did not appear to be a clear and obvious pattern, as individual drought stressed replicates were found at both the upper and lower methylation level extremes in the promoter region. The B76 leaf tissue presented a similar conclusion, as there are two distinct clusters, with each cluster containing at least one replicate from each of the experimental conditions.

In the B73 root tissue, a pattern emerges within the upstream promoter region, with the drought stressed replicates having higher methylation levels generally than the control tissue. This pattern is also observed in the B76 root tissue, where there is also a methylation increase in the promoter regions after drought stress. However, the difference in methylation within the promoter regions of root tissue was more apparent in the B76 variety as the differences were larger.

Methylation differences between tissue types were not as clear within the CHH context, with leaf and root levels remaining similar in both B73 and B76. Even with this revelation, throughout each methylation context, B76 has consistently shown noticeable differences after drought stress in both leaf and root, whereas in B73 these differences are less defined, with replicates overlapping regardless of

experimental condition. This analysis gave the overall patterns of methylation around genes, but further work was needed to find out which genes were significantly differentially methylated in their promoter regions after drought stress.

3.3.2 Genes with differentially methylated regions (DMRs) in promoter regions

3.3.2.1 Logistic regression results

Logistic regression analysis showed significant DMRs found within the promoter region of genes. The list of genes with DMR-containing promoter regions were then separated by maize variety and tissue type to show potential gene expression differences in the CG, CHG and CHH methylation contexts (Figure 3.5.).

| CG | CHG | CHH |
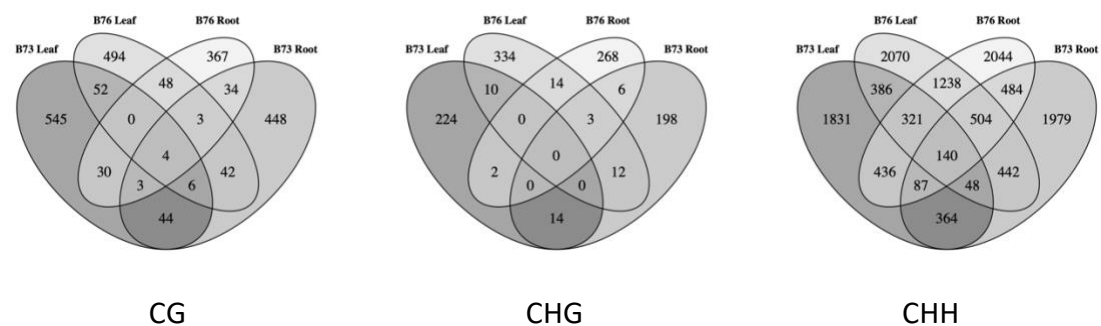
Figure 3.5. Venn diagrams showing the number of genes with promoter regions containing at least one DMR after drought stress in the CG, CHG, and CHH methylation contexts according to logistic regression analysis.

Numbers in overlapping regions represent the number of genes overlapping between different varieties and tissues.

The logistic regression results show that there are more genes with significant promoter DMRs in the CHH context, when compared to CG and CHG. CHH methylation also differs in specificity as it shows a greater number of differentially methylated promoter regions in genes that are found in both varieties and both tissue types. It should be noted that these gene numbers may represent genes that potentially overlap across the three methylation contexts between varieties.

Interestingly, the number of genes with CG and CHG DMRs found in their promoter region remains somewhat consistent across both varieties, although there are differences observed between tissue types. The root tissue of both B76 and B73 contains less promoter regions being differentially methylated than those found in their corresponding leaf tissue.

The major methylation differences caused by drought stress lie in the CHH context, where there are many more genes observed than CG and CHG combined. Even within the CHH context, there are major differences found between the varieties. Both tissues in B76 have a higher number of genes found in the corresponding B73 tissue. Genes found in both B76 leaf and root tissue make up a non-trivial proportion of the overall B76 genes, which is in stark contrast to B73, where there are much less genes shared between the tissues.

3.3.2.2 Score test results

To verify genes discovered using logistic regression, the *DMRcaller* Score test was also applied to the same mapped data. Results gave a competing list of genes with significant DMRs in their promoter region taken from the same data set (Figure 3.6).
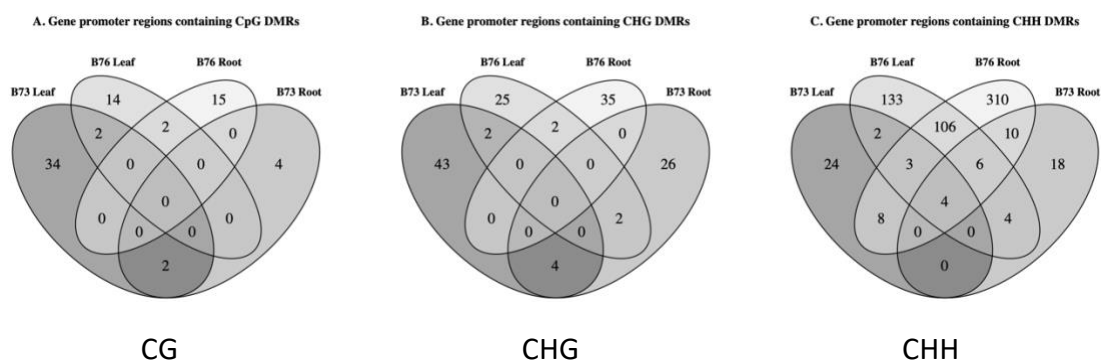
Figure 3.6. Venn diagrams showing the number of genes with promoter regions containing at least one DMR after drought stress in the CG methylation context, CHG methylation context, and CHH methylation context according to *DMRcaller* Score test analysis.

The Score test results show a similar result to that of the logistic regression, with more genes with significant promoter DMRs being found in the in the CHH context. This includes the fact that CHH contained more genes found in different tissues and varieties. Compared to the logistic regression results, there were fewer significant genes observed overall, regardless of tissue type or variety in the CG and CHH contexts, whereas CHG methylation showed more significant differences in the promoter regions compared to the logistic regression analysis.

The most dramatic difference in number between varieties is found in the CHH context again, with B76 root showing the greater number of genes with DMRs in their promoter regions. This Score test analysis corroborates nicely with the logistic regression results in terms of overall pattern, but individual genes relating to drought tolerance were of interest and so genes found to be significant in both analyses were carried forward.

3.3.2.3   Gene matches between analyses

Genes from the logistic regression and Score test analyses were compared to give a definitive list of differently methylated genes (Appendices Table 7.1.1.2). As we were only interested in the differences between B76 and B73, genes that were found in more than one variety were removed. Matching gene numbers found in both analyses were collated in Table 3.3.

Table 3.3 A table showing the number of genes with differentially methylated promoter regions after drought stress in both B76 and B73 maize varieties.

B76	CG	CHG	CHH
Root Only	1	0	111
Leaf Only	0	0	33
Leaf and Root	0	0	21
Total	1	0	165

B73	CG	CHG	CHH
Root Only	0	0	4
Leaf Only	0	2	4
Leaf and Root	0	0	0
Total	0	2	8

Significant DMRs were taken from the matched logistic regression (p-value < 0.05) and Score test (p-value < 0.05) results.

From these results, it is clear to see that a large number of CHH methylation genes were retained in both analyses. On the other hand, there appears to be a limited number of genes that were found in both analyses within CG and CHG methylation contexts.

Interestingly, the number of genes found in the CHH context of B73 is dramatically lower than B76. This is truer in the root tissue, as that is where the majority of the genes are found in the B76 variety. As with the two analyses, there are also a large proportion of genes found in both leaf and root tissue of B76, whereas the same is not true in B73.

The next piece of the puzzle was determining which type of methylation was occurring after drought stress, hypermethylation or hypomethylation. Genes previously found in both leaf and root tissue were treated separately in each tissue

to note tissue specific changes in methylation. Results show that the most of the differently methylated genes were hypermethylated after drought stress (Table 3.4).

Table 3.4. Table showing the number of genes with significantly different methylation levels after drought stress in promoter regions, separated by whether they had undergone hypermethylation or hypomethylation after drought stress.

CG				
	B73 Leaf	B76 Leaf	B73 Root	B76 Root
Hypermethylation	0	0	0	1
Hypomethylation	0	0	0	0

CHG				
	B73 Leaf	B76 Leaf	B73 Root	B76 Root
Hypermethylation	0	0	0	0
Hypomethylation	2	0	0	0

CHH				
	B73 Leaf	B76 Leaf	B73 Root	B76 Root
Hypermethylation	3	54	4	131
Hypomethylation	1	0	0	1

The most noteworthy result is that, despite the large number of genes that show CHH methylation changes after drought stress in B76, almost all these changes are hypermethylation. In fact, there are more instances of hypomethylation after drought stress in B73 genes across all methylation contexts than there are in B76 genes, despite the 20-fold increase in total gene numbers within the B76 variety. It should also be noted that B73 genes were also hypermethylated more than hypomethylated in the CHH context, but as there were far fewer genes this is cannot be considered a conclusive reaction to drought.

Hypomethylation after drought stress is seemingly rare in the genes undergoing differential methylated after drought stress, although there is a limited hypomethylation response in B73 leaf tissue in the CHG context that is not found in B76.

The hypermethylation response in B76 tissue appears to be directed through the CHH pathway, and so the next questions to answer were what role do these significant genes play after drought stress, and are they related to drought tolerance?

3.3.2.4   GO term groups taken from significant genes

To investigate the role of these significant genes, associated GO terms were analysed as a whole, regardless of context within each tissue type. There were overwhelmingly more GO terms found in B76 hypermethylated genes, which is likely explained by the larger number of genes, when compared to the other tissues (Figure 3.7.).
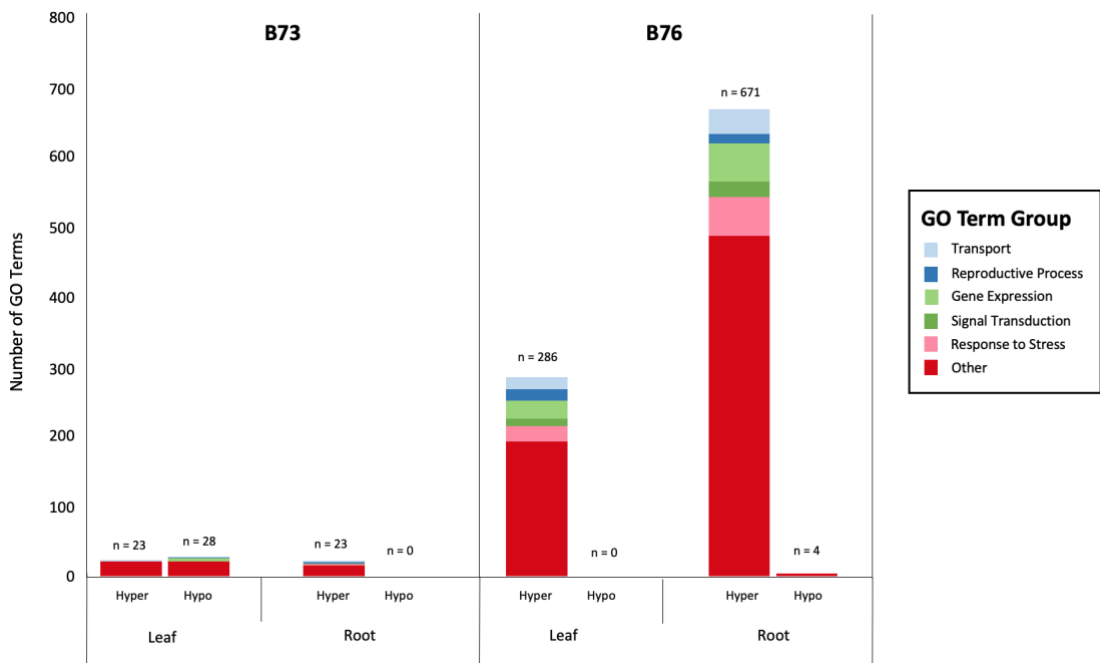
Figure 3.7. Stacked bar graph showing the number of GO terms associated with genes with significantly differentially methylated promoter regions after drought stress.

After grouping GO terms by overlapping parent GO terms, it is clear to see that many significant genes found in the B76 leaf and root tissues are associated with molecule transport, reproductive processes, gene expression, signal transduction and responses to stress. There is also a large percentage of GO terms that could not be so easily grouped, and as a result, the majority of these could not be sufficiently quantified into subgroups.

Within the response to stress subgroup there was also an interesting minor "response to water deprivation" GO term group (Figure 3.8.) containing five genes within B76 root (See Appendices Table 6.1.1.2). This group of genes was only significantly differentially methylated after drought stress in the drought tolerant B76 variety, which suggests known drought tolerance responses may be regulated by methylation in the B76 variety. However, it should be noted that this "response to water deprivation" subgroup was not significantly enriched (p value cutoff = 0.05) and as such is only anecdoctal.
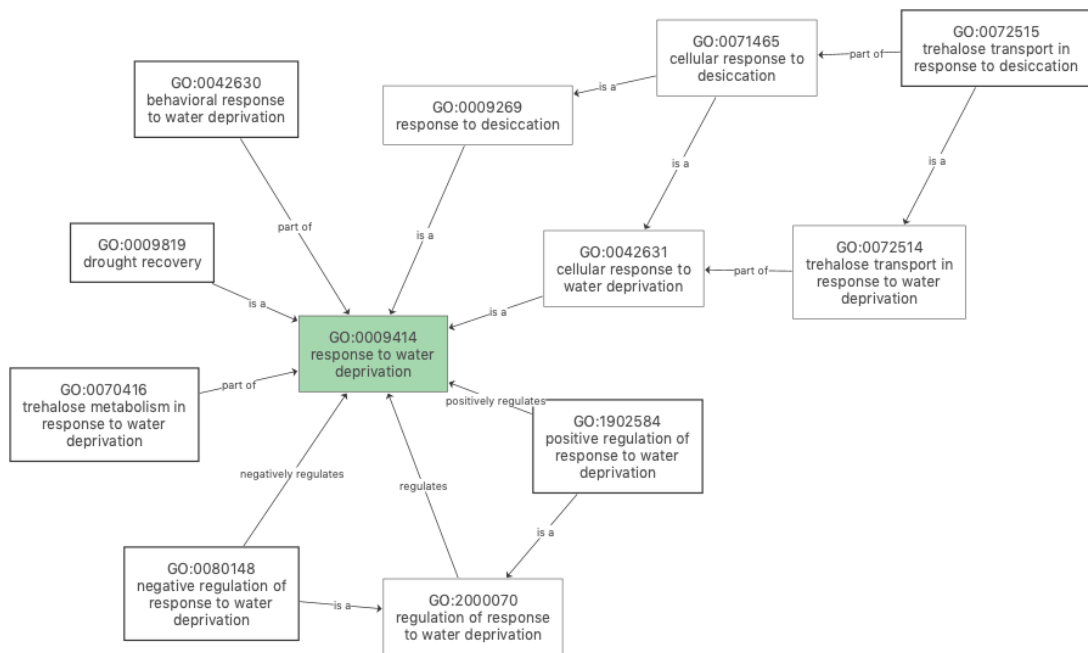
Figure 3.8. A GO graph showing the relationships between GO terms found within the "Response to Water Deprivation" subgroup.

Each connection within this subgroup is represented by an arrow, showing the direction of the relationship, and how they relate. Each box contains a GO term ID and the corresponding GO term name.

However, although there are many more GO terms associated with genes in B76 root, the proportions of these GO terms are not necessarily different to what would be expected if GO terms were chosen randomly from the total maize reference GO terms list. Therefore, there was a need to statistically test whether this was the case or not, and whether the groups observed were indeed significantly enriched within the subset of genes found to be significantly differentially methylated in the promoter region. To do this, singular enrichment analysis (SEA) was used on tissue specific genes found across all methylation contexts.

3.3.2.5 GO term enrichment using SEA

Due to the lack of genes found in all methylation contexts in B73 leaf and root tissues, SEA could not be performed on this variety, as a minimum of ten genes is needed to accurately find enriched GO terms.

As there were over ten genes in both the leaf and the root tissue of B76, SEA was carried out on both. Similar GO terms were collapsed and grouped by semantic and GO term tree associations (Figure 3.9.).

B76 Leaf
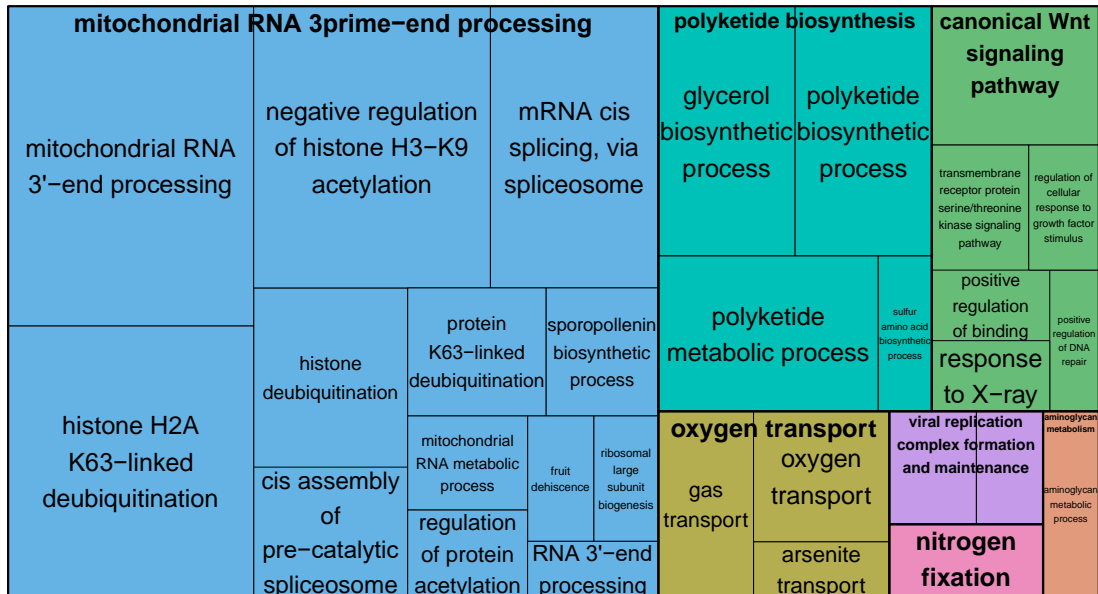
B76 Root
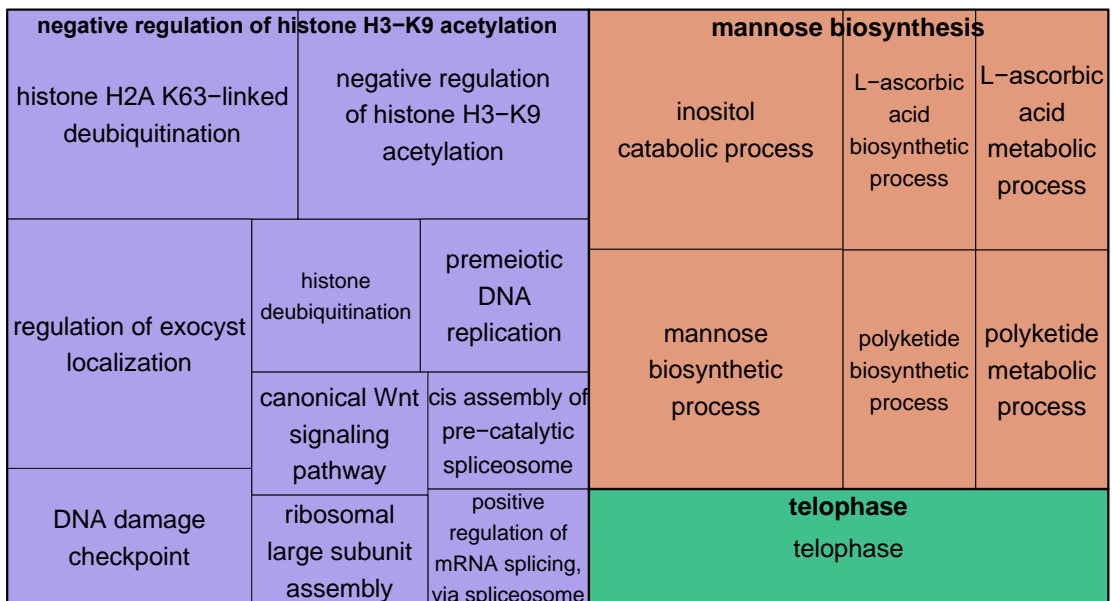
Figure 3.9. REVIGO treemaps summarising GO biological processes over-represented in genes of leaf and root tissue undergoing hypermethylation in their promoter region.

Treemaps were generated using GO terms that were over-represented and their associated p-values (<0.05) according to AgriGOv2 singular enrichment analysis. Semantically similar GO terms were grouped and each colour represents a different functional category. The size of each subgroup is proportional to the associated GO term p-value.

Within the B76 leaf tissue there are seven semantically similar GO term groups that show hypermethylation after drought stress in the promoter regions. The largest semantically similar group is centred around histone regulation and RNA processing. Other major groups are related to metabolite biosynthesis, signalling and transport.

Enriched B76 root tissue GO terms are separated into three similar groups, although there are GO term overlaps with B76 leaf tissue enriched GO terms, likely due to the fact there are a large number of significant genes found in both leaf and root tissue, and have GO term enrichment in both. Like the leaf tissue, the largest group relates to histone regulation, and the second largest group is related to metabolite biosynthesis.

There is a greater number of semantically similar groups in leaf tissue compared to root tissue, and there are also more enriched GO terms in the leaf tissue. After deducing the enriched GO terms in both leaf and root tissues of B76, a list of 24 genes containing these GO terms was collated (Table 3.5.).

Table 3.5. A table showing genes found in leaf and root that are associated with enriched GO terms.

Gene Name	Tissue	Protein Associated	Entrez number	GO term enriched	GO Term Descriptor
GRMZM2G021831	Leaf	Trihelix-transcription factor 15	100274197	GO:0000097	sulfur amino acid biosynthetic process
GRMZM2G461501	Leaf	NA	NA	GO:0000965	mitochondrial RNA 3'-end processing
GRMZM5G883510	Leaf	NA	NA	GO:0000965	mitochondrial RNA 3'-end processing
GRMZM2G062585	Leaf	(DL)-glycerol-3-phosphatase 2	100217236	GO:0006114	glycerol biosynthetic process
GRMZM2G349243	Leaf	L-type lectin-domain containing receptor kinase IV.1	103651308	GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway
GRMZM2G034835	Leaf	Polygalacturonase	100283875	GO:0010047	fruit dehiscence
GRMZM2G168898	Leaf	Hemoglobin 2	732720	GO:0015671	oxygen transport
GRMZM2G060742	Leaf	Citrate transporter family protein	100281495	GO:0015700	arsenite transport
GRMZM5G868062	Leaf	Uncharacterized	100282785	GO:0045292	mRNA cis splicing, via spliceosome
GRMZM2G175513	Leaf	Hypothetical protein	100280332	GO:0051099	positive regulation of binding
GRMZM2G082707	Root	Uncharacterized	100280565	GO:0000027	ribosomal large subunit assembly
GRMZM2G149649	Root	60S ribosomal protein L12	100284082	GO:0000027	ribosomal large subunit assembly
GRMZM2G078954	Root	NAC-transcription factor 129	NA	GO:0000077	DNA damage checkpoint
GRMZM2G300945	Root	Uncharacterized	100284874	GO:0006279	premeiotic DNA replication
GRMZM2G458423	Root	Protein PAIR1	103630570	GO:0007134	meiotic telophase I
GRMZM2G126900	Root	Uncharacterized	100273421	GO:0019853	L-ascorbic acid biosynthetic process
GRMZM2G165535	Root	Phosphomannomutase	100272867	GO:0019853	L-ascorbic acid biosynthetic process
GRMZM2G083783	Root	Glycine-rich RNA-binding protein RZ1C	100381912	GO:0048026	positive regulation of mRNA splicing, via spliceosome
GRMZM2G457003	Root	Uncharacterized	100383401	GO:0060178	regulation of exocyst localization
GRMZM2G018020	Leaf, Root	Uncharacterized	100277950	GO:0000354	cis assembly of pre-catalytic spliceosome
GRMZM2G138429	Leaf, Root	Tobamovirus multiplication protein 2A	100280216	GO:0060070	canonical Wnt signaling pathway

GRMZM2G152436	Leaf, Root	Deubiquitinase pseudogene	100501239	GO:0070537	histone H2A K63-linked deubiquitination
GRMZM2G108894	Leaf	Chalcone synthase	100283134	GO:0080110	sporopollenin biosynthetic process
	Root			GO:0030639	polyketide biosynthetic process
GRMZM2G428933	Leaf, Root	JUMONJI-transcription factor 3	NA	GO:2000616	negative regulation of histone H3-K9 acetylation

Of the 24 genes, there are 9 with uncharacterized or unknown associated proteins, and although there are instances of genes being observed in both leaf and root tissue, therefore potentially having differing GO enrichment patterns, most genes are only associated with an individual enriched GO term.

3.3.3   Genomic regions containing DMRs

Although this chapter was focused on gene promoter regions, there is evidence to suggest that most DMRs are found elsewhere. Each variety and tissue type had its DMR location proportions visualised across the genome, separated by methylation context using *DMRcaller* (Figure 3.10.).
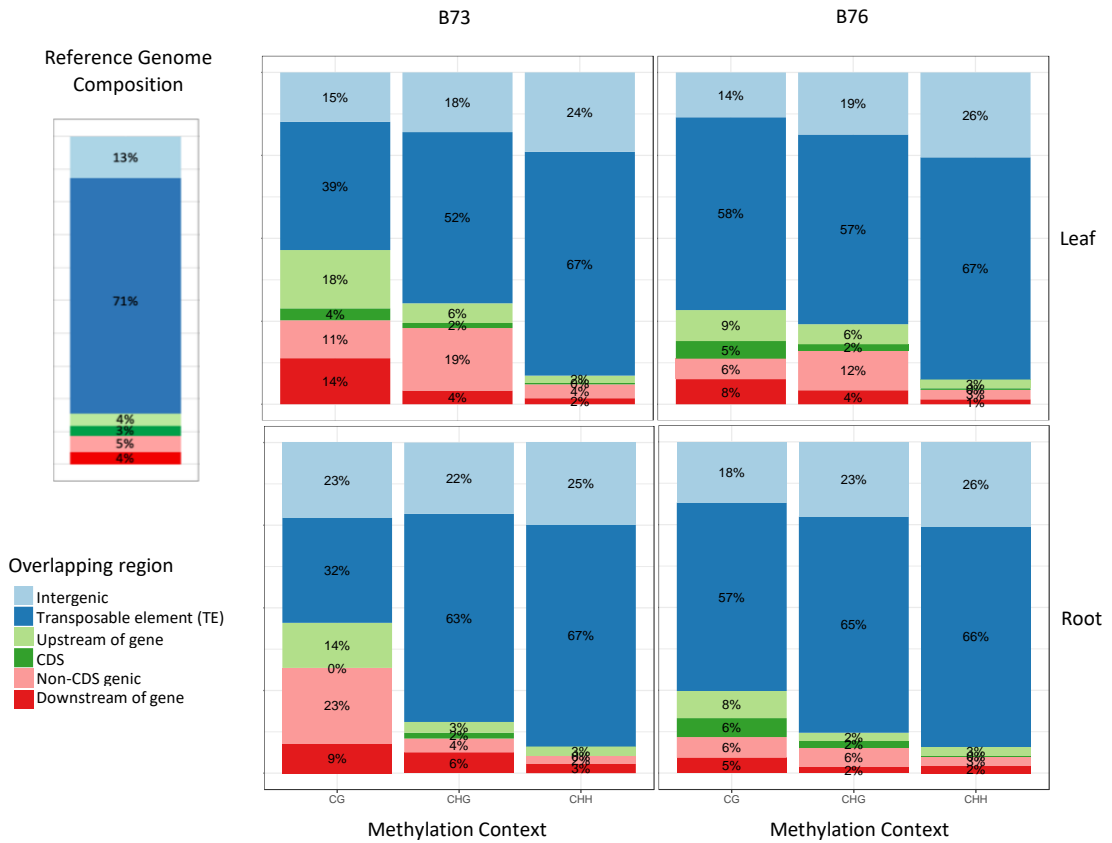
Figure 3.10. Stacked bar graphs representing the percentage of DMRs found across the genome, split by methylation context.

Upstream is considered up to 2 kb upstream of the gene start and downstream is considered up to 2 kb downstream of the gene end. Gene and TE regions were taken from references taken from the Gramene database (http://www.gramene.org).

From Figure 3.10. we can see that there is a higher percentage of DMRs found in the promoter regions of both B73 tissues compared to B76 in the CG context. In general, there is a higher proportion of significant DMRs found in or around genic regions within the CG context compared to CHG and CHH methylation contexts, although this difference is less dramatic in the B76 variety. As a result, the proportion of CG DMRs found in TE fraction of the B76 tissues is higher than that of B73. Lower numbers of DMRs found in the CG context may be the reason for inflated percentages when compared to the other contexts.

Percentages in the CHG context in B73 and B76 tissue are fairly consistent between varieties, with a lower percentage of DMRs found near the genic regions when comparing CG context results, but there are a higher proportion of DMRs found within genes and the TE fraction. There is a difference between leaf and root tissue in both varieties, with a higher proportion of DMRs found upstream/downstream or within the genic regions in leaf and then as a result, a higher proportion found in the TE fraction of root tissue.

The proportion of CHH DMRs in both tissues and both varieties is very consistent and due to the increased number of DMRs found in the CHH context, this appears to be a fairly conclusive result. In each context, tissue type and variety, the largest number of DMRs are also consistently found in the transposable element fraction, thereby hinting at a large reaction within TEs of both drought tolerant and drought susceptible varieties of maize.

3.4   Discussion

3.4.1   Bisulfite conversion rate issues

The bisulfite conversion rates vary substantially between replicates, meaning that the differences between samples observed using this bisulfite converted data may be due biases in the data rather than any biological difference. It is theoretically possible that these conversion rate differences are caused by reads that map similarly to several different locations in the genome. Due to the high percentage of TEs in the maize genome, it could be that we are observing methylated TE reads that also map to the chloroplast, thereby causing artificially increased methylation levels, leading to a decrease in the bisulfite conversion rate. Computational efforts were pursued to resolve this issue, however the bisulfite conversion rates remained similar to the originally calculated rate. Other studies have also used the same bisulfite conversion rate calculation successfully on several different maize varieties and this calculation was successful on the public data sets that were used as controls. Therefore, it must be concluded that these rates are accurate, and as such, results from this thesis can only serve as a pilot study for further investigations into methylation differences caused by drought stress rather than data that can be used for peer-reviewed research papers. However, further investigation into this data did result in potentially interesting results.

3.4.2   Differing tissue- and variety-specific global methylation levels after drought stress in promoter regions

From the quantitation graphs, there are plenty of conclusions that can be theorised about the role methylation plays in both leaf and root tissue of drought tolerant and drought susceptible varieties of maize. The increase in methylation within the promoter regions of B73 leaf compared to B76 leaf in all methylation contexts suggests there is tighter restriction across all genes in B73, as methylation within this region is associated gene transcription silencing. This pattern also matches methylation differences found in rice leaf tissue between drought tolerant and drought susceptible varieties (Gayacharan & Joel 2013). Of course, this does not mean that B76 is less controlled, but rather that there is a greater opportunity for

genes to be potentially expressed in the leaf tissue. This non-restrictive nature in leaf tissue may play some part in the adaptability of B76 but more research would be needed to elucidate this in maize. Unlike B73, there is an overall difference in methylation between control and drought stressed replicates in the promoter region of B76 leaf. This B76 difference suggests there is an overall decrease in methylation response after drought conditions in the CG and CHG contexts and an increased response in the CHH context. CG and CHG methylation are known occur more within gene bodies in maize (West et al. 2014), and as such may not be as abundant when observing promoter regions. In fact, there is evidence to suggest that in maize the majority of CHH methylation is found near TE-associated genes, and may have become important for gene regulation in specific tissues (Bartels et al. 2018; Gent et al. 2013). Therefore, it could be theorised that as the B76 leaf tissue in the CHH context shows a consistent increase in methylation across all drought stressed replicates, there is a controlled silencing of genes across the genome in response to drought.

Interestingly, there was an outlier in one of the B76 replicates and as all plants were grown in the same batch and conditions it is unlikely that growth conditions were the cause of the difference. A possible theory is that the maize originator of the B76 plant may have had a similar methylation pattern to that of B73, and possibly due to the control conditions being different enough to the natural growing conditions, it has caused a reaction in that plant to appear similarly methylated to the B73 control tissue.

In root tissue, the methylation percentages are reversed, with the B73 experiencing lower levels of methylation in the promoter regions than B76. This is counter-intuitive as it was expected that in the drought tolerant variety, there would be a gene expression reaction within root. These results do not exclude this theory, as there are fewer significant methylation events in the promoter regions of the B73 replicates, suggesting that these methylation events are not relevant to a specific genic response to drought. As shown through GO analysis, the change in methylation in B76 is related to a large number of specific genes that could be involved in the

origin of the drought tolerance. This is in keeping with the theory that the B73 response to drought is a limited "jack of all trades, master of none" response to all stress events, whereas the drought tolerant B76 variety has a directed response to drought stress specifically focusing on several tolerance genes. In the CG and CHG contexts there was more methylation after drought stress in the B76 root, which may play a role in drought stress tolerance, although the number of DMRs found in each was significantly lower than CHH, and so may not be as dramatic. Within the CHH context of B76 root tissue, the increase in overall methylation in the drought tissue may suggest a similar reaction to that in the leaf, that there is a controlled silencing of genes after drought stress.

With all this said, there is a statistical limitation when looking at the global methylation patterns of each individual replicate, caused by the slight differences in bisulfite conversion rates. A change of in conversion rate of 1% could dramatically affect downstream analysis due to the size of the genome and the number of reads generated. This limitation is unavoidable and is difficult to account for and therefore results from this data must be corroborated with publicly available expression data before being published.

3.4.3 Methylation pattern of CHH islands in promoter regions

Results from both the logistic regression and the Score test show that there were many more DMRs found within the CHH context, meaning the CHH methylation results are the interesting aspect of this experiment. CHH methylation levels in maize are relatively low compared to plants like *Arabidopsis*, therefore this result is showing a dramatic change in the CHH context above all others. This increase in DMR changes within CHH is probably due to the role they play in the silencing of TEs found near genic regions, such as DNA transposons (Zakrzewski et al. 2017), and due to the fact CHH islands are found upstream of genic regions in maize, rather than CHH being less nuanced in its response to drought stress. These CHH islands are thought to serve as a barrier between the silenced transposable elements near genes and the gene promoter regions (Gent et al. 2013; Li et al. 2015). An explanation for this apparent focus on CHH methylation is that genes near CHH islands are being hypermethylated

to ensure TE regions that are found near these genes are silenced during drought stress. This could be an instance of killing two birds with one stone, with the variety beneficially switching off genes as well as silencing nearby potentially harmful TEs families. Also, as a higher number of genes occur in both leaf and root tissue, there also appears to be a plant-wide reaction to drought stress in the CHH context of B76, and only a lesser response in B73. These combined factors explain why there were more genes with promoter DMRs in the B76 variety, and suggests there is a drought response in the drought tolerant maize variety not found in the drought susceptible variety.

Genes with significant DMRs in their promoters were used as an indicator of a directed reaction to drought stress, and the results show differences in DMR-containing gene promoter numbers between the two varieties in all contexts in both tissues. The number of genes found in all contexts and tissues taken as a whole were very different for both varieties, indicating that the magnitude of genic response within the varieties may be an effective strategy in methylation response to drought stress. This conclusion is logical, as if methylation does play a part in the drought response of drought tolerant plants, then genes relating to the protection of the plant during drought should be differentially methylated, and therefore differentially silenced or expressed. This difference in the number of genes found to be significant also extends to the difference in whether there was hypermethylation or hypomethylation after stress. As mentioned previously, hypermethylation in gene promoter regions is indicative of the silencing of genes after drought stress, whereas hypomethylation in the same region can lead to expression of that gene (Li et al. 2008; Pumplin et al. 2016; Y. Wang et al. 2016; Weinhold, Kallenbach & Baldwin 2013; Liguo Zhang et al. 2012). A hypermethylation pattern emerges across the B76 variety of maize, regardless of tissue type, after drought stress. Unlike the gene numbers found in CHH, the lower number of genes found in the CG and CHG contexts means there are fewer conclusions that can be gathered from results in these two contexts. However, it is interesting to note that even with the 15x fold increase in the overall number of genes from B73 to B76, there are still more instances of hypomethylation after drought stress in B73. This result shows that after drought

stress, there is a motive from the B76 variety to hypermethylate the promoter regions of specific genes.

3.4.4   Enriched GO terms associated with significant genes

From initial analysis of the GO terms associated with significant genes in B76, there appears to be a significant number of transport, reproductive process, gene expression, signal transduction and response to stress GO terms enriched after drought stress. However, even though a large percentage of genes are related to these terms, they were not deemed significantly enriched across the subset of genes when compared to the reference filtered gene set. This may not be surprising as in maize there is a high proportion of genes relating to each of these groups, as the maize v4 GAMER reference contains 39324 genes overall, of which 9455, 4935, 13390, 7006 and 12007 relate to transport, reproductive process, gene expression, signal transduction and response to stress respectively. With this in mind, there is also a known bias around these enriched GO terms, as genes with rarer GO terms are more likely to be significantly enriched, whereas more common GO terms may not appear enriched even with a far more substantial number of genes associated with them. This was the case with these major groups, including the interesting minor "response to water deprivation" GO term family discussed in the results section. Taking all this into account, there are still interesting results outside of the SEA enriched GO terms but these significant responses are more difficult to quantify so SEA was required.

Due to a lack of numbers, enrichment analysis of GO terms associated with the genes found to have differently methylated regions in their promoter regions could not be achieved in any tissue type of the B73 variety. However, these genes are still of interest as they proved significant and as such were still studied. The two genes found hypomethylated within B73 leaf CHG context were associated with uncharacterised proteins, but their GO terms show that these genes relate to the xylan metabolic process and gene silencing pathways. Even though B73 is drought susceptible, this shows that B73 still has a directed reaction to drought, as lowering xylan content in *Arabidopsis* is linked with drought tolerance (Yan et al. 2018). This directed response

is also confirmed in the CHH context of B73 leaf as the two significant genes found are also related to drought tolerance in higher plants, MYB-transcription factors in maize (Wu et al. 2019) and hydroxycinnamoyl transferase in grapevine (Haider et al. 2017). In B73 root, there are four genes found to be significantly hypermethylated after drought stress, all found within the CHH methylation context. Three of these genes are associated with uncharacterised proteins and the fourth is a cell division control protein, which may play a role in drought tolerance although further research would be needed to verify this, as GO terms associated with these genes do not elucidate any particular pathways being controlled. So, although there are no GO terms being enriched in these B73 genes, there are genes with significantly different methylation patterns after drought stress related to drought tolerance.

As mentioned previously, enriched GO terms were only found in the B76 variety, and there were only enough hypermethylated genes after drought stress to perform SEA, so results focused on the function of these potentially silenced B76 genes. In both tissue types of B76, there are GO terms associated with significant genes that are related negative regulation of histone acetylation. These GO terms are consistent with knowledge around response to drought stress in plants, as the regulation of histone deacetylases are generally associated with the repression of genes through chromatin structural changes (Y. Zheng et al. 2016). For example, the histone deacetylase HDA6 is known to regulate the JA signalling pathway in *Arabidopsis* during drought (Kim et al. 2017). JA is well studied as a plant hormone that plays a beneficial role during drought stress in higher plants, such as tobacco (Jin et al. 2011), soybean (Anjum et al. 2011) and *Arabidopsis* (Brossa et al. 2011). There are also histone deacetylases that are found to repress ABA accumulation after drought stress in *Arabidopsis*, therefore negative regulation of this process helps during drought events (Song et al. 2005; Sridha & Wu 2006). The ABA pathway is linked to drought tolerance in maize (Hu et al. 2012; Lu et al. 2013; Lixin Zhang et al. 2012) and so is a good indicator of these genes playing a role in drought tolerance, especially in relation to the negative regulation of ABA (Sah, Reddy & Li 2016; Singh et al. 2015; Wang et al. 2009). There is also an interaction between JA and ABA when regulating plants during drought stress conditions (Brossa et al. 2011; de Ollas, Arbona &

Gómez-Cadenas 2015), suggesting that these pathways may be regulated during drought stress, as previously speculated, using histone deacetylases. Histone deubiquitination was another GO term found in both leaf and root tissues of B76, and this also has a role in the ABA signalling pathway, as in rice, the reversible ubiquitination of histone H2B fine-tunes the ABA pathway and is used to regulate the drought response (Ma et al. 2019). There are also genes relating to polyketide metabolic and biosynthetic processes that have been hypermethylated after drought stress in both tissues in the B76 variety. Polyketides, such as the Chalcone synthase found in this work, are transcriptionally repressed after drought in the maize ovary (Kakumanu et al. 2012), and this gene also relates to the JA-pathway through regulation of maize anthocyanin production (Park, Bae & Ryu 2015). So, as far as global reactions in drought stressed plants are concerned, there appears to be a histone modification focus in both the leaf and root tissues of the B76 variety.

Despite the similarities between the GO terms in both leaf and root, there are some tissue specific responses. In leaf tissue, there are genes associated with the glycerol biosynthetic process, which relates to the production of glycerol. This glycerol production is known to increase grain yield in maize after drought stress (Obata et al. 2015) and as such is beneficial to the maize industry. In root, there are significant genes around the production of ascorbic acid, which plays an important role in the protection of the plant during drought stress, as it protects lipids and proteins against reactive oxygen species (ROS) (Akram, Shafiq & Ashraf 2017; Naz, Akram & Ashraf 2016; Noman et al. 2015). There is also a correlation with the enrichment of the mannose biosynthesis occurring in root tissue, as this metabolite is also involved in the biosynthesis of ascorbic acid (Conklin et al. 1999). ROS are plant signals that induce drought response (Qi et al. 2018), but after prolonged stress they can additionally have an adverse effect on the plant (Cruz de Carvalho 2008). Therefore, the regulation that is occurring in the root of the B76 plant is likely beneficial to the plant after drought through this regulation of ROS. From the collective B76 results, it there appears to be a response within the JA, ABA and ROS pathways, which are crucial drought stress response pathways in many higher plants including maize.

What is less covered is the potential effect ascorbic acid has on the drought tolerance of B76.

3.4.5 Genes relating to enriched GO terms

Of course, to verify the pathways regulated after drought stress, elucidation of specific genes relating to these enriched GO terms was needed. There are five genes found in both B76 tissues that match an enriched GO term, although one of these is uncharacterized. These genes are likely focused around general protection of the plant, and so play a role within B76 regardless of where in the plant they are expressed. As mentioned previously, chalcone synthase is a gene that fits this description, with it being involved in the JA pathway and is known to be repressed after drought, matching the results of this experiment. The JUMONJI-transcription factors, such as the one also differentially methylated in both tissues, regulate transcription through the modification of chromatin structure (Takeuchi et al. 2006; Yu et al. 2008), and confer drought tolerance through their down-regulation in both transgenic tobacco and peanut plants (Govind et al. 2009), although there is a non-tolerance related upregulation observed after drought stress in sea buckthorn (Ye et al. 2018). Another gene found in both tissues, deubiquitinase, removes ubiquitin from targeted proteins, preventing degradation (March & Farrona 2018), and is known to be down-regulated after drought stress in the drought tolerant YE8112 variety of maize (Zenda et al. 2018). This response may point to a regulation of the ubiquitin pathway, which looks to prevent degradation of proteins during stress events. A homologous gene to the tobamovirus multiplication protein 2A, another gene found in both tissues of this experiment, was also found to be differentially expressed in drought tolerant peanut after drought stress (Ding et al. 2014), and the repression of this protein prevents tobamovirus accumulation (Hashimoto et al. 2016). Therefore, although this gene is differentially expressed after drought stress in plants, this may in fact be a defence response to opportunistic biotic stresses during the periods of high stress, such as that caused by drought. The results presented in this study, therefore present four genes with links with responses to stress found in both tissues.

There are also genes found in only one tissue type, which does indicate that looking at both root and leaf during this experiment was justified as there are tissue-specific responses. There are ten leaf-specific genes found to be differentially methylated after drought stress in B76, of which six were characterized. One of these six is a trihelix transcription factor, and these are known to negatively regulate stomatal development and drought tolerance in wheat (X. Zheng et al. 2016), *Arabidopsis* (Yoo et al. 2010) and rice (J. Li et al. 2019). In rice, these transcription factors also respond to ABA and as this gene is differentially methylated after stress, and is in leaf tissue, there is evidence to suggest the same occurs in maize, showing another link to the ABA pathway. The DL-glycerol-phosphatase 2 protein observed in this study dephosphorylates glycerol-3-phosphate into glycerol, thereby protecting cellular structures during osmotic stress and confers drought tolerance in *Arabidopsis* (Caparrós-Martín et al. 2007), although there is overexpression after drought stress in foxtail millet (Pan et al. 2018). Similarly to the conclusion from this project, glycerol-phosphatase is a proposed drought response candidate gene in barley, through QTL mapping after drought (Gudys et al. 2018). L-type lectin-domain containing receptor kinases, like that found to be differently methylated in this study, are both strongly expressed and repressed after drought stress in *Arabidopsis* (Bouwmeester & Govers 2009). However, as of now there is no conclusive evidence to suggest that this gene has a role in drought tolerance and so more research is needed to find out if there is a role. Polygalacturonase is a cell wall hydrolase that degrades cell wall pectin during fruit ripening and was also the protein created through the expression of one of the significantly differentially methylated genes (Liu et al. 2014). Inhibition of polygalacturonase increases in drought tolerant varieties of maize, whereas drought susceptible varieties decrease inhibition (Xiaoli Wang et al. 2016), which would agree with the methylation results presented here. With hypermethylation of the polygalacturonase gene, silencing of this gene is implied, and therefore it is likely methylation plays a part in B76 tolerance. Non-symbiotic hemoglobins in plants, such as hemoglobin 2, can reduce nitrite into nitric oxide (Tiso et al. 2012), and the resulting product is involved in adaptation to drought in higher plants, such as *Arabidopsis* (Kumar et al. 2016; Montilla-Bascón et al. 2017). In rice, there is research showing that non-symbiotic hemoglobins are repressed after

drought acclimation (Marraccini et al. 2012). In fact, there was a similar study to this project looking at drought stress responses in B73 and another drought tolerant variety of maize, Mo18W, and hemoglobin 2 was also highlighted as a candidate gene for drought tolerance (Campbell et al. 2015). Although citrate transporter families are not well studied when relating to drought stress in maize, citrate is the cytosolic substrate of Coenzyme A, an important enzyme in many metabolic processes, and an accumulation of it is linked with drought tolerance (Correia et al. 2018; Gargallo-Garriga et al. 2018). Therefore, the repression of the citrate transporter family protein would be expected to increase accumulation within the leaf tissue, thereby potentially leading to drought tolerance in maize. There are also four genes that have a reaction to drought stress in this tolerant variety, yet have uncharacterized associated proteins. These would likely be interesting candidate genes to study in future to gauge their role in drought tolerance if any. In the leaf there is not a collective and clear mechanism when responding to drought stress in B76, rather, there are responses that collectively aid drought tolerance, although more research on the characterized and uncharacterized genes is needed to discover how important they are to the drought tolerance of B76.

There were nine genes that were found to be significantly differently expressed in root, and as such were potential candidate genes for drought tolerance in the B76 maize variety. Only five of these genes were characterized and therefore had relevant research around them. One such gene was 60S ribosomal protein L12, which is part of the ribosomal protein family known to be a part of the translational machinery of the cell, and there are indications that they also respond to abiotic stresses in maize (Casati & Walbot 2003; Wu et al. 2016). There are unclear links between drought and ribosomal protein expression, as in wheat there is a decrease in ribosomal protein abundance after drought stress (Faghani et al. 2015), whereas there are mixed expression levels across the different ribosomal protein families within the *Stipa purpurea* plant (Li et al. 2016). NAC transcription factors, such as NAC-transcription factor 129, play an important role in the response to abiotic stresses, such as drought, and are known to be induced by dehydration and ABA (Nakashima et al. 2012). In *Arabidopsis*, the NAC transcription factor encoded by the

RD26 gene responds to ABA, JA and dehydration (Fujita et al. 2004), and in rice, overexpression of the *SNAC1* transcription factor gene during drought conditions causes drought tolerance in field conditions (Hu et al. 2006). As stated previously, the JA pathway plays a role in antagonising ABA biosynthesis during initial drought stress and is linked to drought tolerance in plants (Gupta et al. 2017; Harb et al. 2010; Seo et al. 2011). Even in maize there is evidence of this action, as the transcription factor associated with the *ZmNAC111* gene is regulated by ABA, and is involved in maize drought tolerance (Mao et al. 2015), and *ZmSNAC1* is also induced by ABA and dehydration in transgenic Arabidopsis plants, again causing drought tolerance (Lu et al. 2012). The *PAIR1* gene shown in this study, is essential for homologous chromosome pairing and cytokinesis during meiosis in rice (Nonomura 2004). Therefore, repression of the *PAIR1* gene through hypermethylation of drought in maize is likely linked to reproduction of the plant, suggesting a limiting of reproduction processes during drought to ensure survival of the plant. Phosphomannomutase, on the other hand, catalyses the interconversion of mannose-6-phosphate and mannose-1-phosphate, and regulates ascorbic acid production in plants (Qian et al. 2007; Wheeler, Jones & Smirnoff 1998; Yu et al. 2010). As mentioned previously, this ascorbic acid pathway helps protect against long term ROS damage through regulation of this signalling molecule. Glycine-rich RNA-binding proteins, such as RZ1C, were first discovered in maize, in relation to their response to drought stress due to fact they are induced by the ABA hormone (Gómez et al. 1988). Interestingly, the glycine-rich RNA-binding protein 2 is also found to be differentially expressed after drought stress in maize leaf tissue (Zhao et al. 2016). In *Arabidopsis*, the expression of the Glycine-rich RNA-binding proteins RZ1A and protein 4, has a negative impact on the plant during the seed and seedling stages throughout dehydration (Kim et al. 2007; Kwak, Kim & Kang 2005). Therefore, the silencing of this gene using through hypermethylation of the promoter region is likely beneficial after drought stress. From the root genes found to contain significantly different methylation within their promoter regions after drought stress, there is some commonality between the responses, and the GO terms found to be related also show a consensus. ABA regulation of the NAC transcription factor and glycine-rich RNA-binding protein show that this is an important response within B76 root,

but as this is such a widespread and known response to stress, it is not a surprise. However, these, plus the uncharacterized genes, would make interesting candidate genes in future, and more research is needed to elucidate their role in the drought tolerance of B76.

3.4.6   Conclusions

To summarise, across both leaf and root in the B76 variety, 24 candidate genes with differently methylated promoter regions after drought stress were discovered, 9 of which were uncharacterized. As the drought susceptible B73 plants only had limited number of genes that were significantly differentially methylated in both analyses after drought stress and no enriched GO terms, it can be inferred that the response is not as concentrated as B76. This is likely part of why the B73 variety is not as adaptable to drought stress as B76. From the group of 15 genes with known functions found in the B76 variety, it can be postulated that drought tolerant B76 is regulating genes using methylation within the ABA pathway, through JA and ROS signalling after drought stress. ABA accumulation is a well-known drought tolerance technique employed by higher plants, but signalling around it can have undesired damaging consequences for the plant, and as such there is a need for tight regulation. There is also a methylation prevalence around DMRs found in CHH islands 2 kb upstream of genes in the B76 variety of maize, which could indicate that this variety has adapted to using the TE regulation machinery as a method of controlling both TE activity and the negative regulation of important drought stress genes. The total list of candidate genes is also a great leaping off point for future gene expression experiments, and hopefully future gene expression results can be used to corroborate the expected expression inferences afforded by hypermethylation in this experiment. As of the writing of this thesis, there is no public dataset RNA expression data for B76 maize that relates to drought stress, although there are B76 control datasets. Therefore, a future experiment is needed to create the drought stressed dataset to establish that RNA expression differences occur in the same genes with DMRs in their promoter regions theorised in this chapter.

It should also be noted that the conclusions of this experiment only covered the gene

promoter regions. As observed in Figure 3.10., it was observed that most drought-stress related DMRs are found within the TE fraction, and as the maize genome is uniquely dominated by this TE fraction, further research is also needed to observe how methylation changes after drought stress within this region.

However, with the limitations caused by the inconsistent bisulfite rates, these conclusions need to be corroborated using successfully converted data.

4 Drought-stress specific methylation changes across maize Transposable Element (TE) subfamilies

4.1 Introduction

Transposable element (TE) regulation through DNA methylation is a well known phenomonom in plants (Diez, Roessler & Gaut 2014; Slotkin & Martienssen 2007), with TE silencing arising through hypermethylation across the length of each TE (Zilberman et al. 2007). Along with siRNAs, methylation provides the safety net needed to protect against potentially catastrophic TE activation events, especially in TE-rich plants such as maize (Li et al. 2014). TE silencing caused by methylation can also play an integral part of genic expression in the plant, an example of which is the role methylation regulation plays in the creation and survival of genes derived from *Mutator*-like TEs in rice, which ultimately drive adaptation in the plant (J. Wang et al. 2016). This driving of adaptation is of the utmost importance to the future potential use of maize as a food crop during changes in climate, and as such understanding the regulation of the large TE fraction in maize provides a method through which adaptation could be directed. As the TE fraction is highly regulated through methylation, there is an interest in whether a hypomethylation change can cause a reactivation of the TE fraction in maize. This is observed in other higher plants, such as *Arabidopsis*, where methylation is found to be particularly important for silencing TEs and hypomethylation does cause massive transcriptional TE reactivation (Tsukahara et al. 2009; Zhang et al. 2006).

Drought stress is known to change methylation patterns near genes (See Chapter 3), and it is also theorised that this may be the case within the TE fraction of maize. There are 13 TE superfamilies in maize made up of 1526 families (Baucom et al. 2009; Schnable et al. 2009; Tenaillon et al. 2011), each with it's own activity and genetic structure. As such, during this study we were interested in which TE superfamilies were differentially methylated, possibly leading to reactivation during drought stress. A hypomethylation pattern observed in a TE may be an indicator of a reactivation of that particular TE after drought stress, and this is known to be a driver

of adaptation if the reactivated TE is inserted into the right genomic area, such as promoters.

Interestingly, there is evidence outside the plant kingdom positing that TE integration into the promoter regions of stress response genes during activation events enhances gene expression, thereby driving evolution (Feng, Leem & Levin 2013), and this is also observed in maize after heat, salt and UV stress (Makarevitch et al. 2015). It is not only through promoter enhancements that TEs can affect gene expression as plants, as there is evidence of insertions within exons creating novel genes (Hirsch & Springer 2017). However, as methylation changes found in promoter regions after drought stress in drought tolerant maize was already shown, this study was more interested in if TEs were being used as a methylation off/on switch within drought-response gene promoter regions in maize.

It should also be noted that drought tolerant varieties of maize, such as the B76 variety used in this chapter, have already undergone some kind of adaptation, and therefore may not show dramatic methylation changes caused by drought stress. However, it would be expected that B73 should show methylation changes after stress, as a survival method during potentially deadly stresses. This is evidenced in other higher plants, such as the drought tolerant varieties of rice, that are known to keep methylation levels consistent in the TE fraction after drought stress when compared to susceptible varieties (Zheng et al. 2017).

Although the relaxation of TE silencing through hypomethylation after stress may benefit the plant, this benefit in fitness may not be in tune with what humans need from this crop. For example, hypomethylation in a oil palm LINE (long interspersed nuclear element) retrotransposon causes an abnormality in the plant that causes a reduction in yield (Ong-Abdullah et al. 2015). This kind of adaptation would obviously not be considered useful when considering it as a crop for human and animal consumption. Therefore, there is a definite need for the plant to not only survive periods of drought, but also reduce any negative impact on grain yield.

Three main aims were chosen to answer the question of whether TE methylation plays a role in the control of drought tolerance genes in maize: 1) To determine if there are global methylation differences in the TE fraction after drought stress, and if so which TEs are affected. 2) To determine if these differentially methylated TEs are found within promoter regions of important genes. 3) To determine if the genes with differentially methylated TEs in their promoter regions also have differentially methylated promoter regions, therefore playing a potentially important role in their regulation.

4.2   Methods

4.2.1   Data information

Bisulfite data from Chapter 3 was reused for the analysis of this chapter. Therefore, maize growth, sample collection, DNA extraction, bisulfite treatment, next generation sequencing and raw data preparation were all the same from Chapter 3 (See Chapter 3 - 3.2.1, 3.2.2, 3.2.3, 3.2.4.1).

4.2.2   Data analysis

4.2.2.1   Genomic mapping

Whole-genome B73 and B76 leaf and root methylation data, generated in the previous chapter, was mapped to three reference databases using SMALT (https://www.sanger.ac.uk/science/tools/smalt-0), using a modified version of the Diez *et al.* (2014) protocol. The first reference was the filtered gene set (FGS) taken from the B73 RefGen_v2 maize genome (Schnable et al. 2009). The second reference was a custom-made knob and centromeric sequence database, as described in Diez *et al.* (2014). The last and most relevant reference for this study was the unique transposable element database (UTE) developed by Tenaillon *et al.* (2011). This mapping allows for the resulting UTE bam files to be split up into 35 bp kmers, which are matched to TE families, thereby giving a good indication of methylation differences in each TE family.

4.2.2.2   Kmer analysis

The resulting bam files were converted to FASTQ files using the bedtools *bamtofastq* program. To measure methylation in each TE family, mapped reads were divided into 35 bp kmers and each kmer was counted throughout each sample genome using *Jellyfish* (Marçais & Kingsford 2011), as long as the kmer was found in the reference UTE reads. Options for *Jellyfish* count were *-m 35 -s 30000000000 -c 7 -C -L 2*. Kmers were assigned to TE subfamilies if they were found at least once in an associated TE family. To reduce data complexity, any kmer found in more than one TE was removed from the dataset. Using personally developed Perl scripts, kmers that occurred in a minimum of seven reads and contained at least four cytosine opportunities were

matched to TE families. For each cytosine within each kmer the ratio of methylated cytosines to total methylated cytosine opportunities was calculated. This step allowed the isolation of kmers that are informative about the TE in which the kmer is found, even using the degenerative nature of whole-genome bisulfite sequencing data.

4.2.2.3 Statistical analysis

Only TE families with ratios found in all replicates were analysed in R. Using the Shapiro-Wilk normality test in R, all data sets were determined to be non-normal and as a result two-tailed, independent, non-parametric tests were used. All kmers and related ratios found within each TE were collectively used to perform Wilcoxon rank sum tests on each TE, comparing methylation ratios in control tissue and drought stressed tissue. Plots were created in R, using the log of Wilcoxon rank sum tests p-values as the y-axis, and the ratio difference between the mean TE ratios of drought stressed and control replicates as the x-axis. Bonferroni correction was applied to account for the number of tests performed. TEs with a corrected p-value < 0.01 were considered significant and continued for further analysis. TEs that were found to have significant differences in methylation after drought stress were separated into TE superfamilies using the unified classification system for TEs (Wicker et al. 2007).

4.2.2.4 Promoter region analysis

BED files were generated detailing 2 kb regions upstream of all genes found in the maize reference B73 RefGen_v3 using personally developed bash and Perl scripts. Gene information was taken from Gramene (http://www.gramene.org, Data Release 61). The bedtools *getfasta* program was used with these BED files and the reference genome FASTA file to separate CDS and promoter regions (http://www.gramene.org/). Perl scripts were used to compare significant differentially methylated TEs and maize promoter regions, whilst removing genes that contained significant TEs in both the genic and promoter regions.

4.3   Results

4.3.1   Significantly differentially methylated TEs after drought stress

To perform this experiment, drought stressed and control leaf and root tissue were taken from B73 and B76 maize plants as described in section 3.2.1 and 3.2.2 of chapter 3. From the extracted bisulfite sequencing data, methylation differences between the control and drought stressed replicates were calculated for each TE family. From the scatter plots generated using this data, there appears to be a contrast in significant TE numbers and methylation differences between leaf and root tissues in both varieties (Figure 4.1.).
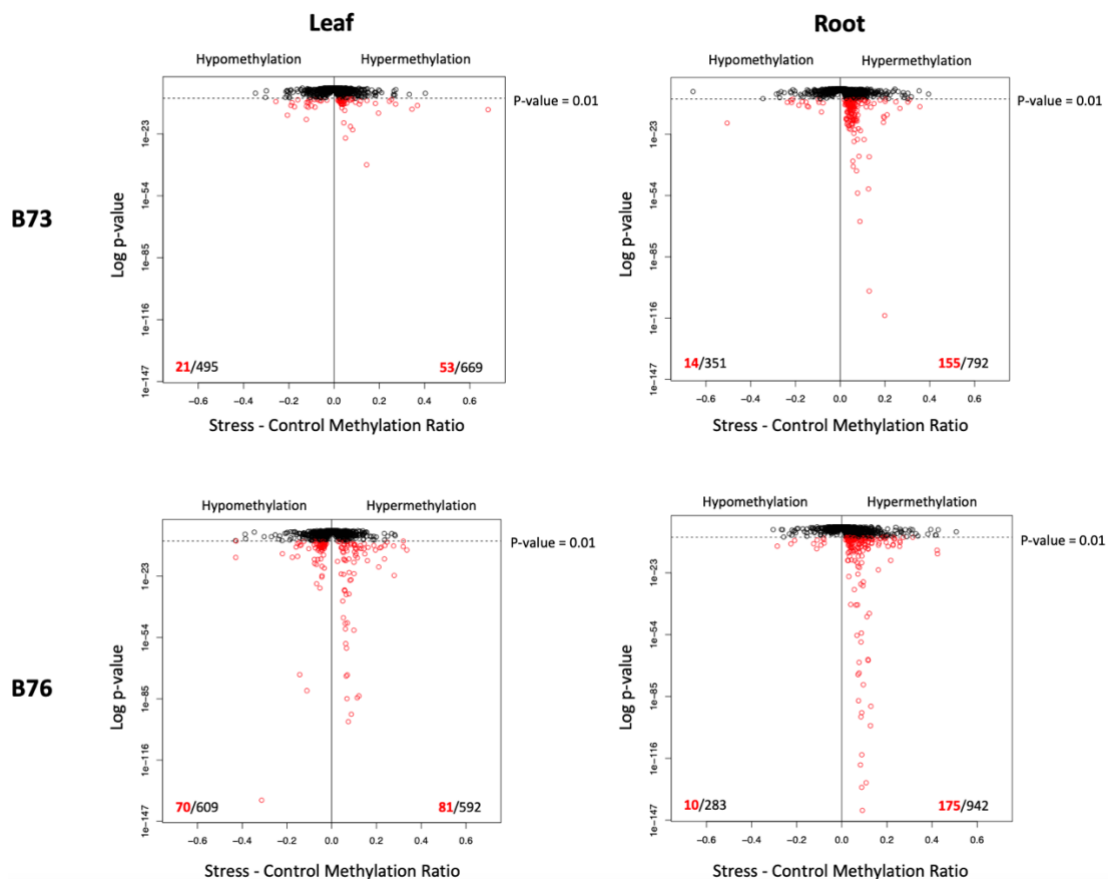
Figure 4.1. Scatter plots showing Wilcoxon test log p-values against differences in methylation ratios between control and stressed replicates in all transposable element (TE) families.

Each TE family is represented by either a black (not significant) or red (significant) circle. A cutoff of Bonferroni-corrected p < 0.01 was used to represent a significant difference between the two conditions. Positive x-axis values indicate more methylation in the stressed than control tissue after drought stress hence

68

hypomethylation in test tissue. Negative values indicate hypomethylation in test tissue after drought stress. Numbers at the bottom of each plot represent number of significant TE families (Red) in the total number of TE families tested (Black) that have undergone hypo- or hypermethylation. TE families that experienced no change are not included in these numbers.

Of the 1526 TE families observed in the TE database, examples of 1165, 1148, 1377 and 1388 families are found in all replicates of the B73 leaf, B73 root, B76 leaf and B76 root tissue respectively. This shows that significant TE differences found between tissue types are not due to differences in number of TEs analysed. However, there is a slight difference in number of TEs covered between varieties, as in B73 leaf and root tissues there are ~200 less TEs found in all replicates, when compared to the same tissues in B76. However, important TE families that are mentioned later, such as the *huck* TE family, were similarly represented in both.

Although TE numbers are similar across the board, there are differences between tissue types in terms of hypermethylation or hypomethylation. In root tissue, there are more TEs undergoing hypermethylation and fewer undergoing hypomethylation after drought stress in both varieties. Therefore, there are tissue specific differences in methylation after drought stress, regardless of variety, with root undergoing more hypermethylation events than leaf tissue in the TE fraction. The general trends are that B76 has more TE family kmer coverage than B73 and more root TE families undergo hypermethylation after drought stress than leaf.

4.3.2 Superfamily breakdown of significant TEs

Using Wilcoxon tests to compare TE methylation ratios in control and drought stressed replicates, a number of TEs are found to be significantly differentially methylated (Bonferroni corrected $p < 0.01$) after drought stress. In the B73 variety, 74 significantly different TEs are observed in leaf tissue, and in root this number was 169. In the B76 variety, there are 151 differentially methylated TEs after drought stress in leaf, and 175 in root. These significant TE raw numbers, along with reference

TE raw numbers are found in Appendices Table 6.2.1.1. The visualization of the breakdown of significant TE family numbers is represented in Figure 4.2.
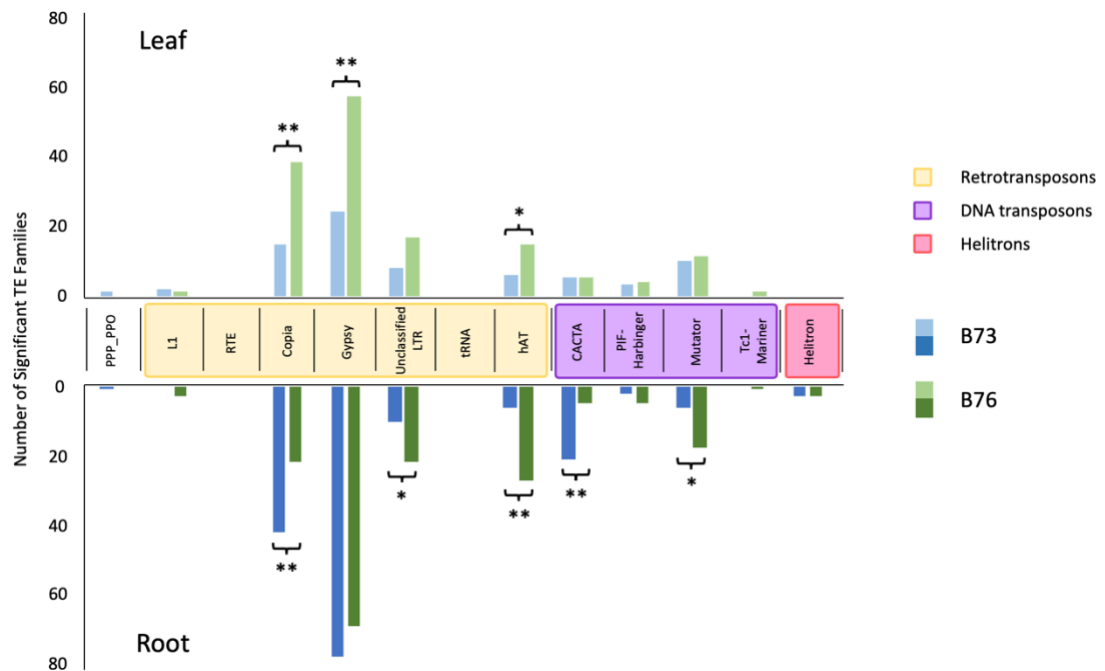
Figure 4.2. A bar graph showing TE representation of significantly differentially methylated TEs across each maize TE superfamily in leaf and root tissue.

TE superfamily names are taken from Wicker *et al.* (2007). The maize variety B73 is in blue, and the B76 variety is in green. Differences determined by using a Z-test: *Significant ($p < 0.05$), **Significant ($p < 0.01$).

There is representation across most TE superfamilies of significantly differentially methylated TEs after drought stress. In the maize TE reference, there is one TE that does not fit any higher classification, *PPP_PPO*. In B73, the difference in the genomic methylation of the *PPP_PPO* TE is significant in both leaf and root tissue, whereas this is not the case in B76.

As these numbers do not consider the number of families expected within the reference genome, these reference TE family numbers were visualized in Figure 4.3.
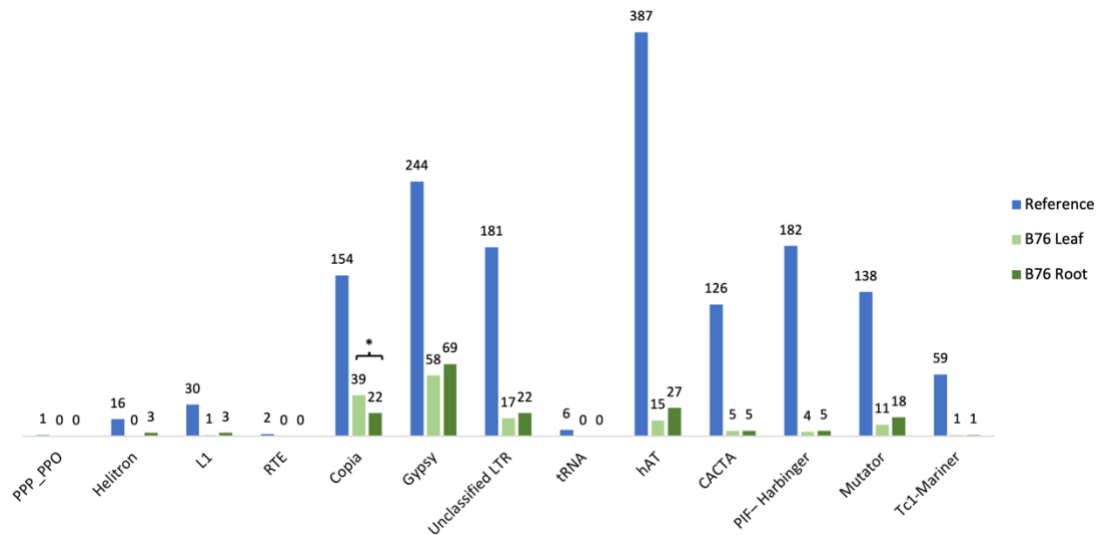
Figure 4.3. A Bar graph showing the number of TE families within each TE superfamily in the B73 maize reference genome alongside the number of significantly different families found in B76 after drought stress.

Differences determined by using a Z-test: *Significant ($p < 0.05$).

4.3.2.1   Class I TEs - Retrotransposons

L1 is the first TE superfamily described to have tissue specific differences in this research, with B73 only having *L1* TE families with significant differential methylation after drought stress in the leaf tissue. B76, on the other hand, contains more *L1* TE families within its root, but both tissue types still have *L1* representation. *L1* TE families are not well represented in the TE database with only 2% coming from this superfamily. *RTE* TEs are not found to be differentially methylated in any tissue or variety as there were no significant *RTE* TE families found in either tissue or variety, although this is not a surprise as *RTE* TE families represent less than 1% of all TE families in the maize TE database. These first four TEs, although containing some interesting differences, do not represent the majority of the TE families found in the reference TE database, therefore minor changes may not be significant.

The *Copia* superfamily is the first of the major maize TE superfamilies, and represents 10% of the total TE families found in the maize TE database. Within each maize variety there are differing numbers of significant TE families across the different tissues (Bonferroni-corrected p-value < 0.01). In B73, there are fewer TE families in the leaf tissue compared to root, whereas in B76 this trend across significant TEs is

71

reversed between the tissues. When comparing between the two varieties within leaf tissue, there are more TE families found after drought stress in B76, and in the root tissue there are more TE families in B73.

Another major TE superfamily within maize is the *Gypsy* family, which represents 16% of all the TE families found in maize. This representation, along with the large number of significant TE families found in this family, means that this family is noteworthy. Within B73, there are dramatically more *Gypsy* TE families in the root tissue when compared to the leaf tissue, and this is also the case in B76, although to a much lesser degree. The leaf tissue also shows a similarly dramatic difference between B73 and B76, with more significant TE families found in B76, however B73 root contains more significant TEs than any other TE family across all tissues.

The class I retrotransposons are made up of *L1*, *RTE*, *Copia* and *Gypsy* TEs, which have a higher number of families found in the B76 variety, regardless of tissue type, but particularly in B76 root.

4.3.2.2   Class II TEs - DNA transposons

Unclassified *LTR* TEs make up 12% of the total TE families in the maize TE reference, and therefore they are also considered a major TE superfamily. Compared to the *Copia* and *Gypsy* superfamilies, the significantly differentially methylated unclassified *LTR* TE families are underrepresented in both leaf and root tissues of both varieties. There are consistently lower numbers of significant unclassified *LTR* TE families in both the leaf and root tissue of B73, when compared to the same tissues in B76. This difference is not observed within each variety, as unclassified *LTR* superfamily numbers remain consistent regardless of tissue type.

Like that of the *RTE* superfamily, there are no significant *tRNA* TE superfamily found in either variety, regardless of tissue type, and this is likely due to them representing less than 1% of all the TE families found in the TE reference.

The *hAT* TE superfamily are the largest maize TE superfamily, representing 25% of all the TE families found in the TE database. Surprisingly, this family is also proportionally underrepresented across all tissues and varieties in this study, although it must be remembered even though *hAT* TEs represent a large majority of all TE families, this does not equate to actual TE copy numbers across the genome. In terms of pure *hAT* family numbers, the B73 variety is particularly underrepresented across both leaf and root tissues, when compared to the equivalent tissue in B76, although there are consistent numbers of significant *hAT* TE families regardless of tissue within B73. From these results, it is also evident that significant *hAT* TE families are more abundant in root than in leaf, in the B76 variety.

The *CACTA* TE superfamily represents 8% of all TE families found in the TE database, and most tissues found in both varieties show similar numbers of significantly differentially methylated TE families found in this superfamily. The only outlier is that of B73 root tissue, where there is a 4-fold increase in the number of significant TE families when compared to all the other tissues.

The *PIF-Harbinger* superfamily is also a major TE superfamily as it represents 12% of all the TE families observed in the TE reference database. Within each variety, the number of significant *PIF-Harbinger* families remain consistent across both tissues. However, this superfamily is underrepresented, with B76 only showing a slightly higher number of families when compared to B73 in both leaf and root tissue types, although this difference is potentially negligible due to the low number of TE families in each tissue.

The *Mutator* superfamily makes up 9% of all the families in maize, and in this study, there are more significant *Mutator* families in the leaf tissue of B73, when compared to the root tissue, which is the opposite of the tissue difference in B76. The main difference between the two varieties occurs in the root tissue, where B76 shows noticeably more *Mutator* TE families than the equivalent tissue in B73.

The *Tc1-Mariner* superfamily represents 4% of the TE families found in the TE reference database, which may explain the absence of this superfamily in B73, and why there is only one *Tc1-Mariner* family that is significant in each of the leaf and root tissues in B76.

The *Helitron* TE superfamily is only found to be differentially methylated after drought stress in the root tissues, in both B73 and B76. The same number of TEs within the *Helitron* superfamily are found in the root tissue of both varieties. The number of TEs found to be differentially methylated after drought stress corresponds with the low representation of the *Helitron* superfamily in the TE database, which only represents 1% of all TE families.

The class II subclass 1 DNA transposons consist of unclassified *LTR*, *tRNA*, *hAT*, *CACTA*, *PIF-Harbinger*, *Mutator*, *Tc1-Mariner* TEs. The overall pattern of these class II families is that they are more prevalent in the leaf tissue of B76 than the leaf tissue in B73. *Helitrons* are the only class II subclass 2 DNA transposon superfamily found in maize and therefore there are not many found in maize from this class, regardless of variety.

4.3.3 *Differentially methylated TE families found in differentially methylated promoter regions*

Following up on the previous chapter's work, there was an interest in how many globally differentially methylated TEs actually correlate with differentially methylated promoter regions after drought stress, and whether or not this correlation could potentially imply a link between the two. To test this, the total number of TEs found at least once in unique promoter regions across the genome was formulated and separated into each corresponding superfamily by percentage of overall promoter region TE count (Figure 4.4.).
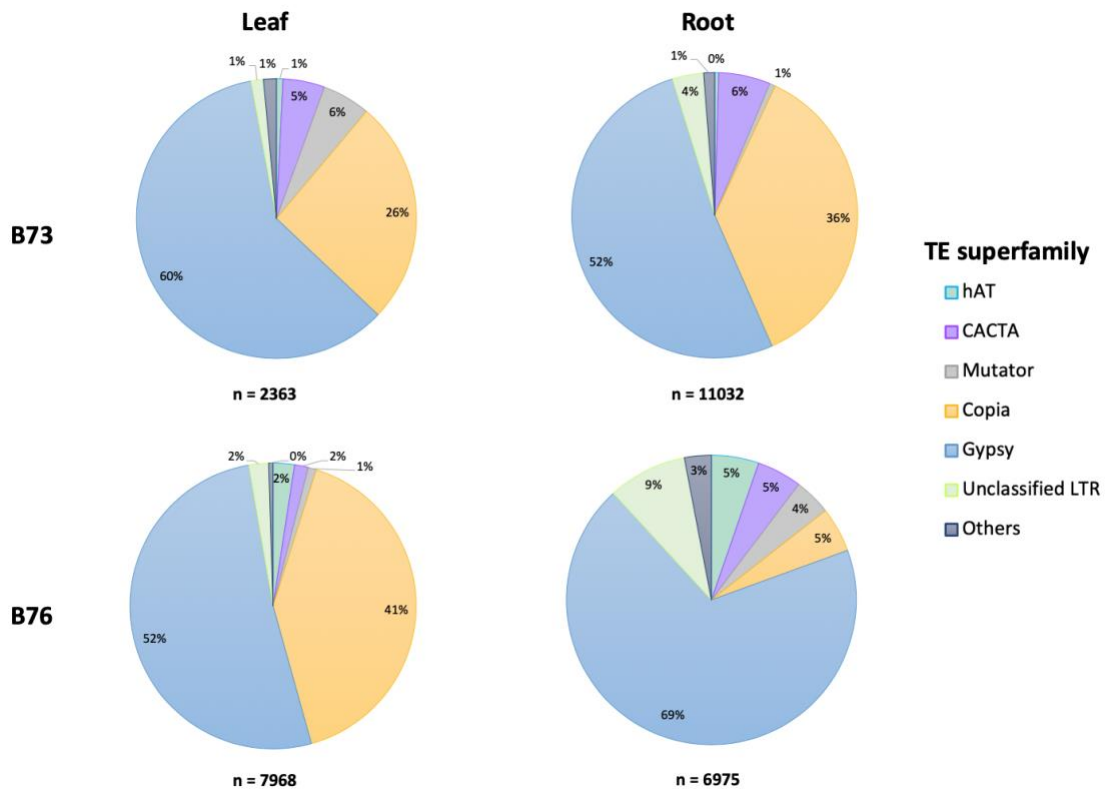
Figure 4.4. Pie charts showing TE superfamily breakdown of instances of significant TEs overlapping promoter regions of leaf and root tissue of B73 and B76 maize.

Each number represents a TE found in a gene promoter region. Multiple TEs found in the same gene promoter region are treated as separate numbers as long as they are not from the same TE family. Each percentage compares the number of significant TEs from each superfamily found in a promoter region against the total number of significant TEs found in promoter regions.

In terms of the number of TE families found in promoter regions at least once, B73 root tissue contains the most across the genome, whereas B73 leaf tissue contains the least. This juxtaposition in B73 is not the case in B76, where TE family numbers remain somewhat consistent between the tissues, with B76 leaf tissue containing only slightly more TE families in promoter regions than B76 root. Between the varieties, in the leaf tissue there are much fewer significant TE families within the promoter regions of the B73 variety when compared to the B76, whereas the opposite is true in the root tissue, with B73 containing many more than B76.

The major differences in superfamily proportions occur in the B76 root tissue, where there is a larger proportion of promoter TEs found within the *Gypsy* TE superfamily than any other tissue type, which could also account for the lower proportion of *Copia* TEs in that tissue compared to other tissues. There is a pattern in the two tissues with the largest TE proportion being found in promoter regions, B73 root and B76 leaf, as they also show the highest proportion of significant *Copia* TEs. B76 root has similar proportions of unclassified *LTR* TE, *hAT*, *CACTA*, *Mutator* and *Copia* TEs, and as such has an equally varied superfamily spread across the genome. Clearly the two biggest TE groups represented in promoter regions are *Gypsy* and *Copia*, and if they are removed from the analysis then there are only 14%, 12%, 7% and 26% of minor TEs in B73 leaf, B73 root, B76 leaf and B76 root respectively. This suggests a more nuanced response to stress in B76 root.

The number of individual TE families found in the promoter region parallels the number of genes with differentially methylated TEs in their promoter regions. In fact, a fairly large proportion of all genes in the maize filtered gene set contains at least one significantly differentially methylated TE (Figure 4.5.).
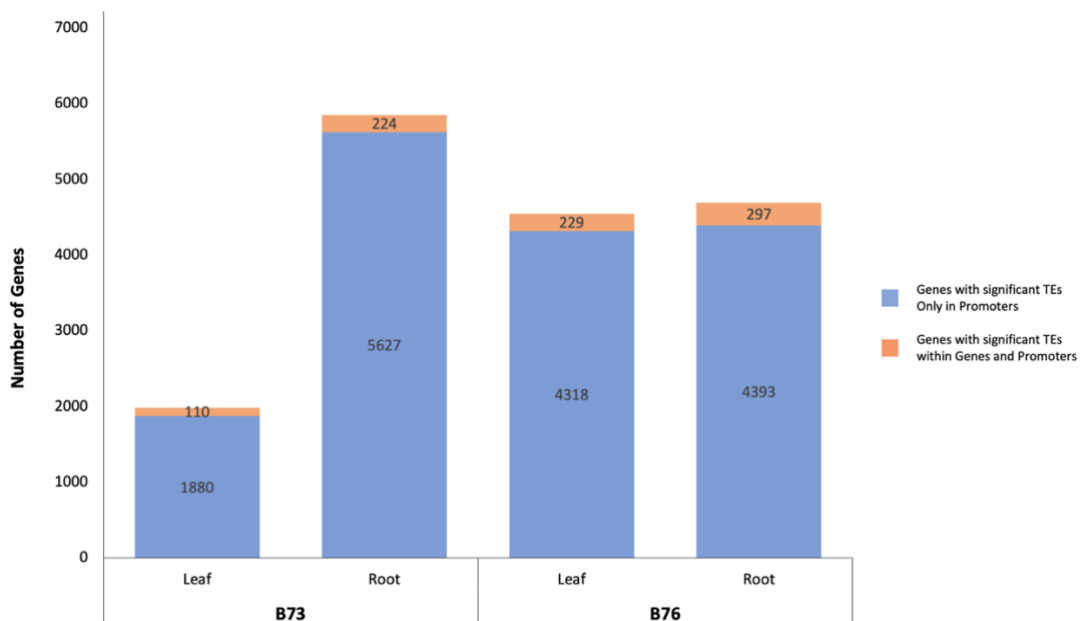
Figure 4.5. A bar graph showing the total number of genes, in B76 and B73 leaf and root tissue, with significantly differentially methylated TEs in their promoter regions.

Each tissue type also includes genes that were removed from analysis as they were also found to be significantly differentially methylated within genic regions.

From the results, there is evidence that TEs undergoing significantly different genome-wide methylation reactions after drought stress are found in a large number of gene promoter regions in maize. Of the ~32k genes found in the filtered gene set, the percentage of genes with TEs that are significantly globally differentially methylated in their promoter region ranges between 6% and 18%. Between B73 and B76 there are differences in how many gene promoter regions are represented by these globally differentially methylated TE families. Within B73 leaf tissue, there are less than half the number of genes found in the same tissue of B76. The opposite is the case in the root, where B73 has more genes with significant TEs in the promoter region than B76. Similar differences are also observed within the B73 variety, with leaf tissue having much less promoter regions represented than that of root, whereas in B76 the number of gene promoter regions remains consistent, regardless of tissue type. This graph also shows the number of genes that were removed from the analysis as they also contained differentially methylated kmers within their genic regions, which remains a fairly consistent proportion of the total number of genes with significant TEs in their promoter region.

A problem with Figure 4.5 is that it does not consider the number of overlapping genes between the two varieties and the two tissue types, therefore a Venn diagram was used to visualise this (Figure 4.6.).
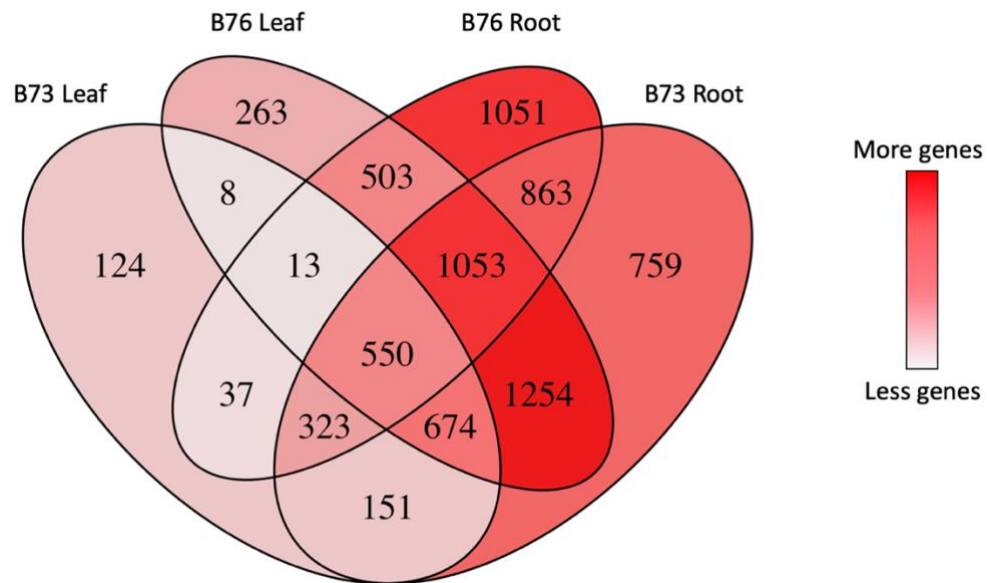
Figure 4.6. Venn diagram showing the number of genes with significantly differentially methylated TEs in their promoter region shared across leaf and root tissues in both B76 and B73 maize varieties.

The interesting aspect of the breakdown of genes with significant TEs in their promoter region, is that there is a lot of overlap between tissue types and maize varieties. What becomes apparent is that there are many genes that are found in both the root tissue and leaf tissue across both varieties. B76 leaf in particular, shares most of the same genes to that of B73 and B76 root tissue, perhaps suggesting some shared genealogy between the varieties. It is also evident that the number of genes that are uniquely found in leaf tissues are much lower than root, and despite a large number of genes found in the B73 root, B76 root has the most unique genes of any tissue.

4.3.4   Link between promoter DMRs and differentially methylated TEs

Although there are many genes that have globally differentially methylated TEs in their promoter region, this does not automatically mean that the promoter region itself is differentially methylated after drought stress. This is where the results from the previous chapter come into play, giving us a list of genes that have significant

methylation differences in their promoter regions after drought stress. Therefore, the next logical step was determining how many genes in each tissue type have differentially methylated TEs found within their differentially methylated promoter regions.

A comparison of all 166 genes with significant drought-response DMRs within their promoter region, taken from Chapter 3 (See Appendices Table 6.1.1.3), was made with genes containing significant differentially methylated TEs in their promoter region observed after drought stress. There are 10 genes that are significant in both chapters (Table 4.1.). These genes are found to be significant in the same tissue, same variety and underwent the same methylation pattern in both the TEs and DMRs found within the promoter region. Only genes from the B76 variety of maize were deemed significant in both studies, which was intuitive as the majority of differentially methylated gene promoter regions found in chapter 3 were also only significantly different in B76.

Table 4.1. A list of genes that have significant differentially methylated TEs within their significant differentially methylated promoter regions.

Gene Name	Tissue	Protein Associated	Entrez number	TE superfamily	Significant TEs in Promoter Region	Function
GRMZM5G806108	Leaf	probable receptor-like protein kinase At1g49730	103646449	Gypsy	RLG_huck_AC199418-6452 RLG_huck_AC210079-10574	Regulates seed germination and dormancy (Sharma et al. 2011)
GRMZM2G445296	Leaf	hydroxyproline-rich glycoprotein family protein	103644471	Unclassified LTR	RLX_small_AC217574-13522	Structural protein in cell wall (Kavi Kishor 2015)
GRMZM2G152436	Leaf	deubiquitinase pseudogene	100501239	Gypsy	RLG_huck_AC186656-1609 RLG_huck_AC194973-4393 RLG_huck_AC195575-4652 RLG_huck_AC199418-6452 RLG_huck_AC199444-6460 RLG_huck_AC216048-13250	Cell homeostasis, signal transduction, transcriptional gene regulation, protein degradation and endocytosis (Hershko, Avram and Ciechanover 1998)
GRMZM2G428933	Leaf	JUMONJI-transcription factor 3	NA	Gypsy	RLG_huck_AC186656-1609 RLG_huck_AC190900-2713 RLG_huck_AC186656-1609 RLG_huck_AC199418-6452 RLG_huck_AC199444-6460 RLG_huck_AC208546-9913 RLG_huck_AC216048-13250	Involved in histone demethylation and maintaining TE silencing in rice (Cui et al. 2013; Saze et al. 2008; Sun & Zhou 2008)
GRMZM2G158831	Root	QWRF motif-containing protein 3	103643626	L1	RIL_totyru_AC203014-0	Involved in microtubule reorientation (Farquharson 2013; Pignocchi et al. 2009)
GRMZM2G416622	Root	probable membrane-associated kinase regulator 2	103634998	hAT	DTA_ZM00030_consensus	Transcription factor regulating lateral root initiation and signalling (Jaillais et al. 2011; Kang & Hardtke 2016; Xuan et al. 2015)
GRMZM2G133972	Root	wox9a - WUSCHEL related homeobox 9a	103636156	Unclassified LTR	RLX_loukuv_AC197842-5799	Positively regulates primary root growth and lateral root initiation (Wu, Dabi & Weigel 2005)
GRMZM2G177792	Root	prx35 - peroxidase35	100281950	Gypsy	RLG_bygum_AC188125-2053	Antioxidant defence (Laxa et al. 2019)
GRMZM2G085974	Root	Uncharacterized	100382861	Gypsy	RLG_dagaf_AC208646-9966	Unknown
GRMZM2G029407	Root	putative leucine-rich repeat protein kinase family protein	100383609	Gypsy	RLG_prem1_AC186287-1362	Regulates cell proliferation, stem cell maintenance, hormone perception, defence response, wounding response, and symbiosis (Torii 2004)

All of these genes are only found to be significant in the drought-tolerant B76 maize variety and are all hypermethylated after drought stress.

Two of the genes in this list overlap the genes with significantly enriched GO terms found in chapter 3, GRMZM2G152436 and GRMZM2G428933. Each of these 10 significant genes were only associated with one TE superfamily, however three genes contained more than one TE family within the promoter region, and in each case these families were part of the *Gypsy* superfamily. As *Gypsy* TEs are the most frequent significant TEs found in leaf and root tissue of B76, there was an initial expectation that more *Gypsy* TEs would be observed than any other TE superfamily, and this is the case. Interestingly, within the leaf tissue, there is a prevalence of TEs found in the *huck* TE sub-family, which only accounts for 20 TE families in the maize TE reference. A *L1* TE also was found in the promoter region of GRMZM2G158831, which is surprising as *L1* TEs make up a low proportion of the significant TEs found in promoter regions, as seen in Figure 4.3.

4.3.4.1 *Permutation test on significant genes to determine observation likelihood*

To determine whether or not the number of TEs observed in each superfamily was a significant result, rather than occurring by chance, permutation tests were carried out on each TE superfamily found in the significant genes (Table 4.2). To perform this test, the same number of genes as the number found to be significant (10) were randomly selected over 1,000,000 iterations using R, and the number of TEs found within TE superfamilies was counted. If the number of observations for a TE superfamily was higher in the random subset than the observed count of that superfamily found in the significant genes, it was recorded. The total number of observations that were higher than the 10 significant genes was divided by the total number of iterations, to determine the p-value of each TE/superfamily. The lower the p-value, the less likely the TE breakdown found in the significant genes occurred by chance and therefore the higher the chance it is overrepresented for the data set.

Table 4.2. Leaf and root permutation test results for each superfamily found in the significant genes.

TE superfamily	Overall Count in tissue	Subset Count in tissue	Permutation test p-value	Tissue
Gypsy	4112	15	0.1894307	Leaf
Unclassified LTR	178	1	0.7698682	Leaf
hAT	370	1	0.5633445	Root
L1	21	1	0.046509*	Root
Gypsy	4794	3	0.992932	Root
Unclassified LTR	616	1	0.7564815	Root

Overall gene numbers were taken from the number of genes with corresponding significant superfamilies found in promoter regions. Subsets are taken from the significant gene information taken from Table 4.1. *Significant ($p < 0.05$).

Results show that the number of TEs found in each TE superfamily, taken from the subset of ten genes with significantly differentially methylated TEs found within differentially methylated promoter regions, can occur by chance if ten genes are chosen at random (p-value < 0.05). This was the case for all TE superfamilies in leaf tissue, however, in root the *L1* superfamily is an outlier, as it is unlikely to occur by chance. This significance may be explained as an artifact of the low number of *L1* TEs observed in promoter regions of genes across the genome, with a total of only 21. Permutation tests were also performed on individual TEs, to find out which TEs, rather than TE superfamilies, are occurring by chance in this list of genes (Table 4.3.).

Table 4.3. Leaf and root permutation test results for each TE found in B76 significant genes.

TE family name	Overall Count in tissue	Subset Count in tissue	Permutation test p-value	Tissue
RLG_huck_AC186656-1609	89	2	0.017494*	Leaf
RLG_huck_AC190900-2713	94	1	0.189738	Leaf
RLG_huck_AC194973-4393	209	1	0.376245	Leaf
RLG_huck_AC195575-4652	128	2	0.032658*	Leaf
RLG_huck_AC199418-6452	189	3	0.008985*	Leaf
RLG_huck_AC199444-6460	136	2	0.036257*	Leaf
RLG_huck_AC208546-9913	116	1	0.229049	Leaf
RLG_huck_AC210079-10574	206	1	0.371903	Leaf
RLG_huck_AC216048-13250	104	2	0.022755*	Leaf
RLX_small_AC217574-13522	30	1	0.0658	Leaf
DTA_ZM00030_consensus	10	1	0.02242*	Root
RIL_totyru_AC203014-0	8	1	0.019914*	Root
RLG_bygum_AC188125-2053	73	1	0.155484	Root
RLG_dagaf_AC208646-9966	46	1	0.101258	Root
RLG_prem1_AC186287-1362	70	1	0.149952	Root
RLX_loukuv_AC197842-5799	98	1	0.203138	Root

Overall gene numbers were taken from the number of genes with corresponding significant TE families found in promoter regions. Subsets are taken from the significant gene information taken from Table 4.1. *Significant ($p < 0.05$).

The results show that there is a total of 7 overrepresented TEs (p-value < 0.05) within the leaf and root tissue gene subsets, 5 and 2 in each tissue type respectively. Within the root tissue, the 2 overrepresented TEs are not linked to any other TE through sub-families, and these two TEs are not well represented in the B76 dataset overall. This lack of representation may present an inherent bias in the permutation test as only one occurrence is required to become significant, however, the permutation test indicates that these are overrepresented in the 10 gene subset observed. Interestingly, in the leaf, there are several overrepresented TEs from the *huck* sub-family, suggesting a possible directed *huck* response within important gene promoter regions. With this being the case, there was an interest in if this *huck* sub-family group, found within the *Gypsy* superfamily, may also be overrepresented as a whole. The total number of significant *huck* TEs found in promoter regions of all maize genes in B76 was 2468, and therefore they constitute the majority of *Gypsy* TEs in B76 maize. Even so, *huck* TEs were significantly overrepresented (permutation test p-value = 0.040914) in the subset of genes analysed, thereby suggesting they may play a role in the regulation of leaf response to drought stress in drought tolerant B76 maize through methylation.

4.4 Discussion

4.4.1 Bisulfite conversion rate issues

As before, there are inherent biases caused by inconsistent bisulfite conversion rates across all the samples, thereby limiting the impact of these results. However, further analysis was performed.

4.4.2 Significant differentially methylated TEs after drought stress

The kmer-based technique used during this chapter's research allowed for the majority of TE families to be analysed. Unfortunately, not all TEs were covered, and this is likely due to the fact that some TEs in the reference are shorter than the 35 bp kmer length used to perform this analysis. There are 16 TEs in the reference that are shorter than 35 bp and 52 shorter than 100 bp, which likely affected the number of TEs that could be theoretically observed within each replicate, and was a known bias when deciding to choose the 35 bp kmer size. The advantage of specificity afforded by the 35 bp length was deemed more valuable than the disadvantage caused by this inherent bias. Only exact kmer matches were used when comparing against the TE references using a self-made Perl script, thereby there was minimal mapping bias introduced, even with the smaller TEs.

Of the analysed TEs, we observed a proclivity for TEs to undergo hypermethylation after drought stress in the root tissue of both varieties, thereby also showing a tissue-specific methylation response in maize TEs. This methylation result is in contrast to that found within TEs in the root tissue of soybean after heat stress, in which hypomethylation occurred (Hossain et al. 2017). This may suggest that there are specific TE methylation responses that change depending on the type of organism or type of stress subjected. The hypermethylation reaction within the root likely creates stability within that tissue, as there is research showing that hypermethylation of TEs found in promoter regions is known to cause gene silencing in *Arabidopsis* (Le et al. 2014). This stability in the root should help the plant search for new sources of water, and prevents a catastrophic change in this tissue. The surprising result of the initial analysis is that B76 leaf tissue shows a large percentage of significant TEs undergoing

85

hypomethylation after drought stress, suggesting a switching on of TEs. This is unexpected as stability would make sense when relating to a response to stress in a drought tolerant variety, however, this may also indicate an adaptation response through TE activation in the leaf.

In general, hypermethylation occurs to a greater degree after drought stress across both tissues in the drought tolerant B76 variety when compared to the drought susceptible B73 variety, which was also the result found in rice (Garg et al. 2015). Overall, this does match with the theory that the cause of B76 drought tolerance is the ability of the plant to keep root tissue stable during stress events, and a potential mechanism to adapt through TE hypomethylation in the leaf, with a caveat that depends on the activity of these hypomethylated TE. This hypermethylation response may also explain promoter region methylation changes near drought tolerance important genes observed in the previous chapter, indicating a silencing process across the genome to also protect the plant. Alternatively, promoter differences could also be explained by regional methylation changes, of which TEs are caught up in, therefore more research is needed to clarify this.

4.4.3   Breakdown of TE superfamilies that are differentially methylated

TEs that are differentially methylated after drought stress in B76 and B73 show patterns when they are separated into TE superfamilies, as observed in Figure 4.2.

The majority of TEs undergoing methylation changes after drought stress reside in the *Gypsy* superfamily, and many more *Gypsy* TEs are found in drought tolerant leaf tissue, when compared to the drought susceptible. This makes the *Gypsy* TE superfamily disproportionally higher than it should be, as the *hAT* superfamily has the most TE families within it. There is also an increase in the proportion of the differentially methylated *Gypsy* superfamily TEs within the promoter region of B76 root genes, when compared to B73 root tissue. This increase in the drought tolerant variety is plausibly because *Gypsy* TEs carry an insulator element within their body and as such they are able to obstruct gene interactions with transcription factors and through the blocking of upstream genetic enhancers (Singer, Liu & Cox 2012; Slotkin

& Martienssen 2007). This obstruction opportunity may indicate that these *Gypsy* TEs are playing a role in the repression of gene transcription after drought stress in drought tolerant maize through the obstruction of promoter/transcription factor interactions.

Helitron TEs are only found in the root tissue of each variety genome, and are known to have DNA helicase and rolling-circle replication initiator domains, that perform DNA cleavage and ligation (Kapitonov & Jurka 2001, 2007). *Helitron* TEs can occur in different tissue types, and as they are only significantly differentially methylated in root, this suggests that *Helitron* TEs have different roles in root than they do in leaf, similar to how they are differentially spliced in shoot and root in maize (Barbaglia et al. 2012). Interestingly, *Helitron* TEs have shown recent activity in maize and can capture flanking exon regions (Lai et al. 2005), meaning that it is possible that an active *Helitron,* such as the *Helitron* TEs found in root promoter regions in this study, may have contributed to the tolerance of B76 through promoter region insertion of important drought tolerance genes.

The number of *Copia* TE families that are significantly differentially methylated is higher in drought tolerant leaf and lower in drought tolerant root when compared to the drought susceptible tissues, and there is a massive reduction in the proportion of *Copia* TEs in the drought tolerant root promoter regions compared to any other tissue type. This is unexpected as in higher plants, *Copia* TEs respond to abiotic stress, such as heat in rice (Ito et al. 2011), causing insertions that affect the stress responsiveness of nearby genes, and so we would expect insertions near important tolerance genes in root. However, in the leaf tissue of the drought tolerant B76 variety, the largest proportion of promoter *Copia* TEs are found, which may match the hypothesis that insertions are selected against in B76 root, but selected for within the leaf, although this would need more research.

It is interesting that there are more TEs with unclassified differentially methylated *LTR*s in the promoter region of both B76 tissues, as there is evidence that unclassified *LTR* TEs, such as the mPING in rice, do not respond to drought stress (Casacuberta &

González 2013). Due to the large proportion of the unclassified *LTR* TEs in promoter regions of genes across the B76 genome, there was likely one or more activation events that caused this expansion, and as they are differentially methylated, they could play a role in the silencing of genes, although more research is needed to elucidate this.

 The *hAT* TEs make up a large percentage of maize TE families in the maize TE reference, but are underrepresented in the significantly differentially methylated TEs observed in the genome of both B73 and B76. There is evidence that *hAT* TEs can be reactivated through stress of maize tissue culture (Smith, Hansey & Kaeppler 2012), but it is apparent that they may not be activated in a response to drought stress from the results found here.

There is a similar reaction to that of the *hAT* TEs in the *Mutator* TEs within root tissues, which may also indicate a drought tolerance reaction within root. *Mutator* TEs are known to be regulated by DNA methylation (Singer, Yordan & Martienssen 2001), and are shown to be reactivated after abiotic stress in maize (Qian et al. 2010), and the activation of *Mutator* TEs can cause something akin to chronic stress in maize (Skibbe et al. 2009). Therefore, the difference in the number of TEs differentially methylated may suggest that B76 actively inhibits the reactivation of *Mutator* TEs after drought stress through a major hypermethylation response, thereby stabilizing the tissue during drought events.

Initially, it appears that B73 root has the highest total number of genes with at least one differentially methylated TE in their promoter regions, but further analysis shows that B76 root tissue has more root specific genes than that of B73 root. There are also differences between the tissue types regardless of variety. This shows that if the drought stress methylation reaction is caused by differentially methylated TEs in the promoter regions of genes, then it is tissue specific, with reaction in leaf possibly being less nuanced than root, or that TE methylation is not widely used for drought-responses in leaf. The fact that B76 contains more unique genes than that of B73,

even with the increased number of genes in B73 root, is important as it shows that the B76 response is larger in the TE fraction of promoter regions.

This sort of experiment also highlights the importance of looking at separate tissues when working with methylation, as the difference between leaf and root is dramatic regardless of drought tolerance.

4.4.4 Genes with significant DMRs and significantly differentially methylated TEs within promoter region

There were significant genes found in both this chapter and the previous chapter, which indicates at least some overlap between differentially methylated TEs and differentially methylated promoter regions. These overlapping genes, listed in Table 4.1, also have links to drought stress responses in higher plants.

The probable receptor-like protein kinase At1g49730 for example, which regulates seed germination and dormancy in maize through its interaction with the abscisic acid response element binding factor 1, is known to be induced by different abiotic stresses (Sharma et al. 2011). The overexpression of another receptor-like protein kinase, LRK2, provokes drought tolerance in rice (Kang et al. 2017), and in maize a similar protein, ZmSIRK1, is downregulated after drought stress in leaf tissue (Sekhon et al. 2011; Stelpflug et al. 2016). Therefore, there is evidence to suggest that the downregulation of this probable receptor-like protein kinase gene, possibly through hypermethylation of TEs in the promoter region, has an effect on the drought tolerance of B76 through the regulation of seed production, thereby improving the fitness of the variety.

Another significant gene, the hydroxyproline-rich glycoprotein, is differentially expressed after maize dwarf mosaic virus inoculation in susceptible maize (Cassone et al. 2014), and therefore may play a part in initial responses to both biotic and abiotic stresses. It is also one of a group of major structural proteins found in the plant cell walls, that experience downregulation after drought stress in other crops,

such as potato, through the upregulation of miRNA (Kavi Kishor 2015; Kieliszewski et al. 2010; N. Zhang et al. 2014).

As mentioned in the previous chapter, deubiquitinase removes the polypeptide ubiquitin from a target protein post-translationally, and this deubiquitination is known to play a role in stopping unnecessary protein degradation after drought stress in maize (Zenda et al. 2018). This gene could play a role in the drought tolerance of B76 through histone H2A K63-linked deubiquitination, as this GO term was found to be enriched in the previous chapter, although more research is needed to confirm this.

Another gene that is significant in both chapters, and also has an enriched GO term, is the JUMONJI-transcription factor 3. This is an interesting gene as it relates to TE silencing and demethylation, as it maintains TE silencing near genes in rice through the demethylation of histones (Cui et al. 2013; Saze et al. 2008; Sun & Zhou 2008). Not only that, but there is also research indicating that this protein also responds to drought stress in rice, with some drought tolerance being conferred when a JUMONJI-transcription factor, *OsJMJ703*, is knocked down (Qian et al. 2015; Song et al. 2018). Therefore, there is evidence to suggest that the downregulation of JUMONJI-transcription factor 3 through hypermethylation of promoter region *huck* TEs likely plays a role in drought tolerance through the negative regulation of histone H3-K9 acetylation.

QWRF motif-containing proteins, such as those associated with significant genes presented here, are a group of microtubule associated proteins found in plants such as *Arabidopsis* (Farquharson 2013; Pignocchi et al. 2009). Therefore, it is theorised that drought stress of the B76 variety may affect microtubule organisation and in turn confer drought tolerance. There is some evidence to suggest this may be the case through studies relating to Protein Phosphatase 2Cs in *Arabidopsis* (Bhaskara et al. 2017), although nothing of note in maize has been published to date.

Membrane-associated kinase regulators, like those made through the expression of the probable membrane-associated kinase regulator 2 gene found in this research, are signaling proteins known to regulate lateral root initiation in *Arabidopsis* (Xuan et al. 2015). Interestingly, this gene has tissue-specific expression in plants (Jaillais et al. 2011; Kang & Hardtke 2016), and as this gene is only significant in root tissue in this study, then it is likely this gene is only regulating root growth after drought stress. This appears counterintuitive as a search for water may be imperative during drought events, although this search from the plant may come at a cost in the form of molecule limitation due to cellular expansion, and so B76 may be conserving rather than expanding during stress periods. This theory matches the overall silencing in root, and therefore conservation, of TEs using hypermethylation across the B76 genome.

Peroxidases, like peroxidase35, are proteins involved in antioxidant defense, and in maize there are previous studies detailing how peroxidases are initially upregulated after drought stress (Laxa et al. 2019). The interesting fact about maize peroxidases is that this initial upregulation is maintained in drought susceptible varieties, whereas in drought tolerant varieties there is eventual downregulation after prolonged drought exposure to reduce damage caused by extended peroxidase exposure (Anjum et al. 2017; Chugh et al. 2013; Ge et al. 2006). This is in line with what is occurring in this experiment through TE methylation differences, and may point to how this initial upregulation and eventual downregulation is controlled after prolonged drought stress.

WUSCHEL related homeobox 9 is a transcription factor that also positively regulates primary root growth and lateral root initiation in *Arabidopsis* (Wu, Dabi & Weigel 2005). So, there is a link with that of the membrane-associated kinase regulator also found in this study, as both play a role in regulating root growth. WUSCHEL related homeobox 9B enhances drought tolerance in rice through triggering flowering earlier (Minh-Thu et al. 2018), although, as this gene is theoretically downregulated in the root tissue, this may not be the case in B76. Another study shows that WUSCHEL related homeobox 9 is also involved in panicle and endosperm development in rice,

and may regulate the process of embryogenesis in response to drought (Cheng et al. 2014). Therefore, it can be summated that regulation of this gene through methylation of TEs possibly regulates root growth, while at the same time possibly signaling to other areas of the plant to modify seed production during drought.

Leucine-rich repeat protein kinases regulate many developmental and defense-related processes, including cell proliferation, stem cell maintenance, hormone perception, host-specific and non-host-specific defence responses, wounding responses, and symbiosis (Torii 2004). Previous work by Kakumanu *et al.* (2012) discovered that there is no significant difference in expression of this gene after drought stress in leaf tissue of B73 maize, which matches the results of this study even though that study did not focus on root tissue. In rice, the leucine-rich repeat protein kinase, OsSIK1, confers drought tolerance when overexpressed in leaf and stem (Ouyang et al. 2010), which differs to what was observed in B76 maize which is theoretically downregulating gene expression after drought stress, although previous work was also unconfirmed in root. Conversely, the receptor-like protein Leaf Panicle 2, also found in rice, is downregulated after drought stress, as it is involved in the opening of stomata cells in leaves (Wu et al. 2015). So, it appears that these kinases regulate a wide variety of processes across higher plants, but there is a link to drought tolerance after differing levels of expressions, which fits in well with the results found here, however, the role they play in root tissue is not known and needs further research.

There is also an uncharacterised gene that was also downregulated in B73 leaf tissue after 6 days of drought stress in a previous maize study (Zhang et al. 2018). This leaf downregulation does not appear to be caused by the differential methylation found in TEs as this was not observed in the significantly differentially methylated TEs in leaf tissue of either B76 or B73. Therefore, there may be a difference in control responses after drought stress depending on tissue, which has been observed throughout this research. It is possible that this gene is downregulated in both tissue types of B73 and B76, and yet it may only be regulated by promoter region TE methylation in the root of B76.

4.4.5 Overrepresented TEs in significant genes

From the significant genes' associated TEs we can see that genes in the root tissue were more varied in terms of TE families. This suggests that the response in root, by way of TE methylation as a control mechanism, is less focused than the leaf of B76, as responses in the leaf are seemingly through the control of certain members of the *huck* TE family.

Some of the significantly differentially methylated TEs within significantly differentially methylated promoter regions of these ten genes were also unlikely to occur by chance. Initially there was evidence to suggest that only the *L1* TE superfamily contained overrepresented numbers, through the use of permutation tests, but this was proven incorrect when testing individual families, and individual TEs. It is true that the *L1* superfamily was the only superfamily overrepresented in the subset of interesting genes, but many individual TE families were also overrepresented.

Interestingly, the *huck* TE family found within the *Gypsy* superfamily were also overrepresented, and a cursory glance at the significant genes provided, and their associated TE families, also show this. The *huck* TEs are one of the major TE families in maize and represent 15-21% of the maize genome (Estep, DeBarry & Bennetzen 2013; Meyers, Tingey & Morgante 2001; Vicient 2010), however this family group does not have any recent activity in maize (Estep, DeBarry & Bennetzen 2013). Therefore, it is unusual that there is a difference in the number of *huck* TEs between the two varieties in focus here. As *huck* TEs are known to be fairly inactive, one could speculate that the real reason that these *huck* TEs are overrepresented in B76 and not in B73, or the reference, is due to the fact that non-*Huck* TE reactivation and insertion during the selective breeding process could have disrupted the promoter regions of B73 genes. As such, this disruption could have potentially removed a previous *Huck*-related drought tolerance response in important genes, causing susceptibility of B73 to drought stress. As this *huck* significance is only seen in leaf tissue, this is likely where TE reactivation occurred in B73, however, this does not

completely explain the differences also found in root of both varieties and so TE reactivation analysis is required.

4.4.6   Conclusions

To summarise, the findings here support the notion that the majority of TEs undergo hypermethylation after drought stress, especially in the root tissue. Although both drought tolerant and drought susceptible maize follow this overall pattern, there is more hypermethylation in the drought tolerant B76 variety. Each individual TE superfamily has a different number of families that are significantly differentially methylated after drought stress across both varieties, and these differences also appear to be tissue specific. Significantly differentially methylated TEs are found in a large percentage of gene promoter regions in maize, and a large percentage of these TEs are shared across promoter regions in both varieties and tissue types, although root has more unique genes with TEs in the promoter regions than leaf. There are ten genes that contain globally differentially methylated TEs within their differentially methylated promoter region, suggesting a possible link between TE methylation and the control of drought-response specific gene expression. Future work would include performing bisulfite PCR on these ten gene promoter regions, to confirm that these differences in methylation are occurring within the TEs found within the promoter regions. These ten genes contain overrepresented TE families including the *huck* sub-family, usually inactive, in leaf tissue of the drought tolerant variety. Therefore, it is theorised here that this overrepresentation of *huck* family TEs may be related to drought-response control in leaf tissue, and there is a suggestion that the drought susceptible variety has lost this *huck* methylation control of important drought-response genes possibly through genetic recombination. There is also a need to elucidate if this overrepresentation is related to differing TE activities between the varieties caused by drought stress. As in the previous chapter, the next logical step would be to look at RNA expression data after drought stress in these ten important genes in the B76 variety, however, the data needed is not in the public datasets yet.

5 General Discussion

5.1 Final summary

The purpose of this research was:

1. To show if there is a difference in gene promoter region methylation caused by drought stress.
2. To determine if there is consistency in the methylation response in promoter regions between drought susceptible and drought tolerant varieties of maize.
3. To ascertain the methylation differences found in the TE fraction of maize caused by drought stress.

DNA methylation changes are well documented after drought stress in other higher plants (Banerjee & Roychoudhury 2017), and as such it was expected that drought stress would have a similar effect in maize therefore answering the first aim stated above. However, there is contrary evidence suggesting that stress-caused methylation changes in maize are a result of stochastic changes, and are therefore not consistent (Eichten & Springer 2015). The work stating this as fact only used the drought susceptible B73 as an example, and as this thesis shows, B73 does not have a consistent methylation response near promoters. What the work in this thesis has shown is that there is a consistent methylation response in the drought tolerant B76 variety. Therefore, it may be that specific methylation responses in maize may only occur in response to stress in the tolerant varieties, thereby regulating the response in a beneficial way.

From analysis of B76 and B73 leaf and root tissue, 24 candidate genes were discovered to be differentially methylated in their promoter regions associated with enriched GO terms, all of which were present in the B76 variety. This implies that the methylation response in the drought tolerant B76 variety of maize is directed, focusing on specifically enriched pathways after drought stress. It was also clear that there are tissue specific responses in each variety, and so all future work on maize regarding methylation needs to take this into account. There was also a total of 166 genes that have differently methylated promoter regions after drought stress across

both B73 or B76 regardless of GO enrichment, with B76 containing significantly more. This shows that there are differences in methylation happening within both drought tolerant and drought susceptible maize following drought stress, however, in B73 there are less genes with significantly differently methylated promoter regions. This lesser methylation response in B73 may be in part due to stochasticity, rather than a directed response in the variety, as observed in Eichten and Springer (2015). There is a definite response in the drought tolerant variety B76, relating to drought tolerance, as the methylation changes are found in promoter regions of genes associated with the ABA, JA and ROS signalling pathways (Ahmad et al. 2016; Cruz de Carvalho 2008; Sah, Reddy & Li 2016). This response is likely contained within each generation of plant growth as there is evidence in *Arabidopsis* that the methylome remains stable after transgenerational drought stress (Van Dooren et al. 2018; D. R. Ganguly et al. 2017). As such, stochasticity in the methylation response after drought stress is not occurring in all varieties of maize, rather it may be enhancing drought tolerance in some maize varieties.

It is known that TE activity changes after environmental stresses in plants (Negi, Rai & Suprasanna 2016), so this was expected to be similar in maize, although less was known about the link between methylation of TEs in the promoter regions and the effect on TE activation. Therefore, we looked at methylation differences found in the in the TE fraction, which revealed that the majority of TE families undergo hypermethylation after drought stress regardless of drought tolerance. We also deduced that many TE families were differentially methylated after drought stress in both varieties of maize, with obvious differences found between leaf and root tissues. Comparing differently methylated TE families with the 166 genes with significantly differentially methylated promoter regions, we were also able to find ten genes that contained differentially methylated TEs within their differentially methylated promoter regions. This gave a list of genes that were potentially being regulated through hypermethylation changes in specific TE families after drought stress. An interesting result was that the TE subfamily *huck* was found to be overrepresented in promoter regions of these ten genes after drought stress in root and leaf tissues, meaning they are possibly being used as regulators for drought

tolerance responses. This information will hopefully be used to guide future experiments aiming to test whether hypermethylation caused by drought stress actually silences the activity of TE families in the drought tolerant varieties but not in drought susceptible varieties.

The overall conclusion of this thesis is that drought stress does cause differences in methylation in the promoter regions of drought-tolerant varieties of maize, like that of B76. These differently methylated promoter regions are associated with potential drought tolerance genes, and also contain TE families that are differentially methylated after drought stress too. Certain TE families are overrepresented in differentially methylated promoter regions, such as the *huck* sub-family of TEs in B76. It may be the case that the genomic mechanisms that silence B73 *huck* TEs are being disrupted during drought stress in similar ways, like that found for other abiotic stresses (Bucher, Reinders & Mirouze 2012; Casacuberta & González 2013; Ito et al. 2016), hence why they are not observed, however more research is needed to confirm this. The hypermethylation of these TE families causes their silencing within promoter regions, reducing their numbers over successive drought stressed generations Therefore, using the results gathered from this thesis, I would speculate that the drought tolerance of B76, and possibly other maize varieties, is in some way caused by the insertions of particular TE families into the promoter regions of particular genes, relating to ABA, JA and ROS signalling in B76 for example. Due to the natural silencing of these TE families through hypermethylation during drought stress, the promoter region was also hypermethylated, causing the associated gene to be silenced in response to drought. This silencing of genes proved beneficial to the plant, thereby improving future fitness of the variety and making B76 drought tolerant. This work shows that drought stress related methylation changes, and TE regulation plays an important part in the drought tolerance of maize.
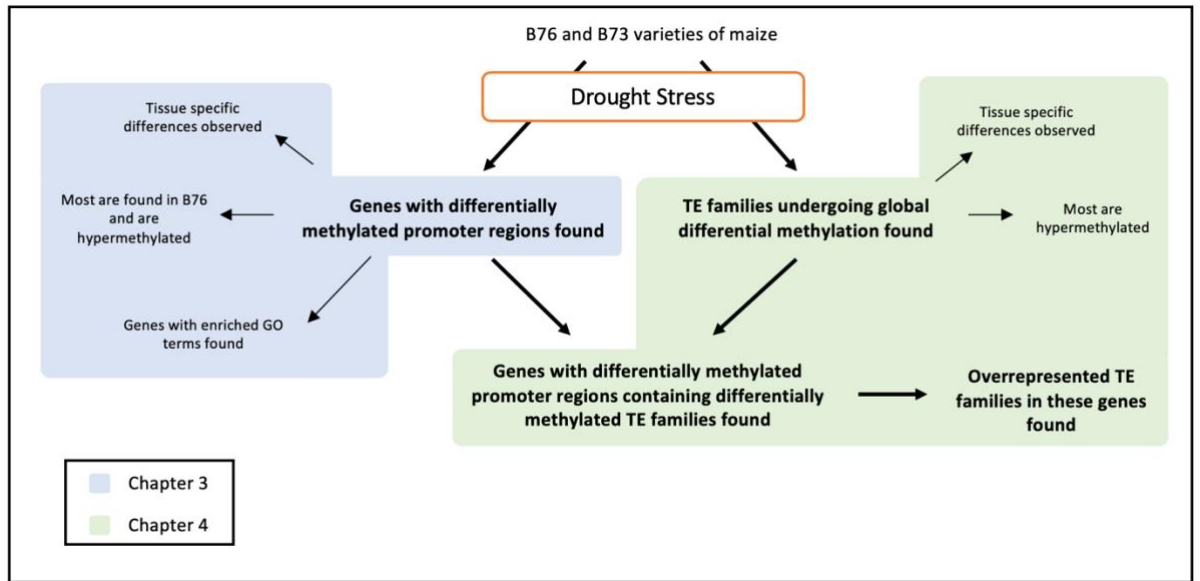
Figure 5.1. Diagram showing the major findings from each chapter.

5.2 Future work

The results presented in this thesis provide a start for further research around the control of TE transposition during drought stress. Similar to the work being done in other plants, there is much more research needed to show these results are meaningful.

With the data generated during this work, it is clear that the TE methylation response in drought tolerant varieties of maize is inconsistent across the different varieties, and as such there is a possibility that other TE families, separate from the *huck* TE sub-family, are possibly differentially methylated in other varieties of maize. Therefore, an interesting next step would be to repeat the experiments of chapter 3 and 4 using other drought tolerant varieties of maize. This way, a methylation profile could be set up as before, allowing us to observe which TE families play an important part in their drought tolerance response, possibly showing structural patterns found in the TEs used for drought response. Of course, there is also the chance that there is no universal TE methylation response across all varieties after drought stress, which would not surprising as it expected that a species has variation in drought tolerance methods (Fang & Xiong 2015), but this needs to be confirmed by further research.

There is further research needed that is related to the differentially methylated TEs within promoter regions observed in this thesis. Importantly, bisulfite PCR is the next step required to confirm that TEs within promoter regions are actually undergoing differential methylation after drought stress. This amplification of the differentially methylated promoter regions will prove that it is the TEs found within those regions of interest that are differently methylated. This would allow us to have definite answers about certain TE families, such as the *huck*s, and could allow us to attempt to manipulate their effect on genes through the addition or removal of methylation, thereby determining the role TEs play on the regulation of associated genes.

Apart from the methylation side of things, performing an expression analysis on the genes that have differently methylated TE families within their differently methylated promoter regions would prove that methylation changes are having an effect. Although there is plenty of evidence suggesting that hypermethylation in promoter regions causes the silencing of the associated gene (Zhang et al. 2006), this is still yet to be confirmed in B76. Another interesting aspect not covered in this thesis is the methylation changes found within the gene bodies. There is evidence to suggest that methylation changes found within genes can change expression of the gene it is within (Bewick & Schmitz 2017). The work carried out in this thesis excluded any genes that had intron and promoter region methylation differences and so this would provide another avenue for future research.

The silencing of TEs through hypermethylation can inhibit their reactivation within the genome (Diez, Roessler & Gaut 2014; Okamoto & Hirochika 2001; Slotkin & Martienssen 2007). Therefore, an important next step is to test whether the hypermethylation of B76 TEs caused by drought stress corresponds with changes in activity over several generations. Within this framework, it would also be crucial to observe whether changes in activity caused by methylation differences is a common response found in other drought tolerant varieties of maize. Previous research has indicated that abiotic stress does cause the reactivation of TEs in higher plants (Horváth, Merenciano & González 2017; Negi, Rai & Suprasanna 2016). For example,

UV light stress induces *Ty1-copia* retrotransposons in Oat plants (Kimura et al. 2001) and heat stress induces the *ONSEN-Copia* retrotransposon in *Arabidopsis* (Ito et al. 2011). Therefore, looking for drought-specific TE activation changes in drought tolerant and susceptible maize could be a genomic mechanism for adaptation.

There are some studies showing that there are active TEs in plants, including those found in the *hAT* superfamily, such as the *Ac* and *TCUP* in maize (McClintock 1950; Smith, Hansey & Kaeppler 2012) and nDART1 in rice (Tsugane et al. 2006). This is interesting though as they only make up ~1.5% of all TEs in maize (Stitzer et al. 2019). Some are also found to be activated after environmental stresses. In oat for example, there is evidence of *Copia*-like TE activation after biotic and abiotic stresses (Kimura et al. 2001) and in Arabidopsis the *Copia*-like ONSEN TE is activated by heat stress (Ito et al. 2016). It is possible there is an activation and removal balance found within the maize genome, as it has been predicted that the retention of TEs in a genome aids TE silencing efficiency (Roessler et al. 2018). There are also genetic mechanisms, such as non-homologous recombination (Devos, Brown & Bennetzen 2002) and genetic recombination (Kent, Uzunović & Wright 2017), that reduce TE numbers in plants, especially after several generations of selfing (Roessler et al. 2019). Methylation could be one such mechanism, as it is used to silence the activity of *CACTA* TEs over several generations in *Arabidopsis* (Kato, Takashima & Kakutani 2004) and the hypomethylation of *CACTA*-like Pack-TYPE elements does cause mobilisation (Catoni et al. 2019). It is known that hypermethylation causes the silencing of TEs in plants, limiting their ability to activate and accumulate (Le et al. 2014).

Of course, there is an inherent danger when reactivating TEs within an organism, as TEs have the potential to insert themselves into genomic areas important for survival. With this danger comes the potential for this insertion to cause beneficial effects through the creation of new genes, which has been observed in other higher plants. Different TE insertions in the FATTY ACID ELONGATION1 gene of yellow mustard (*Sinapis alba*) for example, have resulted in four different alleles, each causing a difference in erucic acid content as a result (Zeng & Cheng 2014). Allelic differences

caused by TE insertion have been observed in maize, where *Mutator* TE insertions have resulted in differential expression of the hormonal transcription factor *knotted1* gene (Bolduc et al. 2012; Greene, Walko & Hake 1994). Another *Copia*-like TE insertion upstream of the *Ruby* gene causes the distinctive colour of blood oranges when the organism undergoes cold stress (Butelli et al. 2012). Therefore, there is future interest in observing the activity of TEs after abiotic stresses, such as drought, as there is potential for the creation of new genes. This sort of research could be achieved by first studying the genes highlighted during this PhD.

It is not only the creation of new genes that TEs play a part in, it is also the fact that many genes that are responsive to stress contain TEs within their promoter regions. An example of this was observed in *Arabidopsis*, where the up-regulation of stress response genes was facilitated by the targeted demethylation of promoter TEs in *Arabidopsis* (Le et al. 2014). In rice, hypermethylation events in promoter region TEs occur after phosphate starvation near stress response genes, indicating another adaptation link between the TEs and stress response genes (Secco et al. 2015). Therefore, there is indication that TE insertions near genes can affect their expression to stress and may contribute to genomic adaptation in organisms.

Variable stress responses in important genes, along with the creation of new genes through TE insertions, reinforces the idea that TEs are one of the drivers of adaptation in plants. In fact, the coupling of TE activation during abiotic stress events and TE-derived stress-response gene expression means that TE activation can have a lasting effect over several generations. This is evidenced in *Arabidopsis*, where two TEs are thought to have contributed to the adaptation of the plant by facilitating its spread outside of Europe (Li et al. 2018). In fact, stress in one plant generation can even affect TE activation in the next generation, potentially causing a helpful stress-response in progeny. This was observed in *Arabidopsis*, where the progeny of heat stressed plants contained new *Copia*-like *ONSEN* insertions in siRNA-deficient mutants, conferring a heat response in nearby genes (Ito et al. 2011). This suggests that multigenerational stress of a plant lineage may create potential stress responses through TE insertions in all descendants of an initial plant.

This multigenerational stress adaptation may also play out in maize, where an increase in non-redundant TE insertion frequency and TEs found near stress-related genes suggests that TEs have played a role in the adaptation process during domestication too (Lai et al. 2017). In fact, TEs are known to act as local enhancers of gene expression in response to cold, heat, UV and high salt stress in maize (Makarevitch et al. 2015). These enhancements can have real world consequences in plants, and their ancestors, allowing them to thrive in new environments. For example, selective breeding in maize has caused the insertion of a *cis*-regulatory Harbinger-like TE, prompting repression of the *ZmCCT9* gene, leading to longer flowering periods in the prolonged daytime of higher altitude climates (Huang et al. 2018). Another TE insertion, a MITE, was also found to be associated with the *Vgt1* locus in maize, which is linked with early flowering time (Castelletti et al. 2014). These two examples show the power TE insertions have on the phenotypic response of an organism and its progeny.

Therefore, the future research to follow up on this research should show that activation of TEs can occur after successive generations of drought stress events and to determine which TEs are activated after drought stress. If this is the case then does drought-responsive hypermethylation of B76 promoter region TE families causes TE activation differences over several generations. Lastly, this activation difference would need to compared against the phenotypic responses that drought stress has on the drought tolerant varieties of maize. The next step in confirming that the TE methylation changes found in this thesis relate to expression changes is to perform a largescale multigenerational drought stress experiment. This would require each individual maize plant to be subjected to drought stress over its lifespan, and then self-breeding to ensure the limiting of genomic changes caused by cross-pollination. Each plant lineage should be DNA sequenced before drought stress in the first generation and before drought stress in the last generation for an accurate comparison. Unfortunately, due to the complexity of the maize genome, this process could prove difficult in deducing any conclusions by itself. Therefore, expression data would also be needed for each lineage to confirm that changes in the genome are

being caused by drought stress, rather than just stochasticity caused by other factors. Ideally, this would require several drought tolerant and drought susceptible varieties of maize to conclude if the TEs observed in this experiment are found in all drought tolerant varieties of maize, or if they are specific to B76. Additionally, other drought susceptible varieties of maize would determine if the lack of methylation regulation in B73 TEs is specific to that variety, or in all susceptible maize. Even with this said, it is unlikely that much could be found with the current technology available, therefore this is an idea that likely needs to be explored further in the future. As this work relates to the real-world consequences of drought stress, it would also be interesting to observe the phenotypic changes that would be associated with changes in TE caused by methylation.

Therefore, in this largescale growth experiment, it would also be useful to match TE changes with several phenotypic responses in maize. Plant height in maize, for example, is known to initially increase in response to drought stress during the development stage, before maturing into a shorter than usual plants eventually, all within a single growth cycle (Su et al. 2019). This could help the understanding of drought tolerance mechanisms in maize as a lower overall plant size is advantageous during drought conditions, as it helps the movement of water in a plant (Olson et al. 2018). The reduction of stem width found in maize and tomato trials, regardless of resistance to drought, also help with the retention of water (Aslam et al. 2015; Meng et al. 2017), and as such may also be an avenue of interest. Grain weight is clearly the most important phenotypic measurement during future experiments as field studies in maize show that kernel weight per ear can decrease to 24.1% of that grown in favourable conditions (Ion et al. 2013). Due to B73 being one of the most important varieties for the maize production industry, the decrease in grain weight shows why this crop needs to become more adaptable to drought stress. Inbreeding depression also needs to be accounted for during a self-pollinated multigenerational experiment as it can reduce yield as much as 59.2% in maize, and has been observed as far back as Charles Darwin's time, (Darwin 1876; Pacheco et al. 2002). As a logical next step, this sort of long-term experiment is unfortunately very resource intensive and would

require everything from genomic data to phenotypic data to confirm a result that is rarely observed in maize.

Methylation of TEs are regulated by small interfering RNAs (siRNAs) (Law & Jacobsen 2010; Simon & Meyers 2011), so another logical next step would be to analyse the expression levels of siRNAs during drought stress. As siRNA molecules act as a regulator of methylation and TE activation, it would be interesting to observe what part they play in this process, and more importantly which siRNAs are involved in the regulation of the TEs observed to be overrepresented in this thesis. Interesting, siRNAs maintain genomic stability through the suppression of transgenerational transposition of TEs in plants (Ito 2012), and so focusing on a similar transgenerational experiment, but measuring siRNA expression after each drought stress event, rather than TE activation, would be a great method for determining if they play a part in the TE drought tolerance response.

Another area of future research interest would be to analyse the individual TE families themselves. There is no current research relating to the two TE families that are active in all varieties of maize, DTA_ZM00036_consensus and DTA_ZM00067_consensus. These provide an exciting avenue for experimentation as it would be interesting to discover if these TE families are found to insert themselves into certain genomic areas. Another role they may play is in the creation of new genes through shifting reading frames or advantageous insertions, and so discovering genes with artefacts relating to these TE families may elucidate how important their activity is to the survival of the species. Therefore, if insertions are found near genes it is important to understand the structural changes these new proteins may have, so protein expression analysis could be used to determine the function of new proteins.

The nature of science has changed over the last decade due to the influx of genetic sequencing techniques, and this PhD thesis represents that change. The quality and quantity of data gathered during this PhD will present the chance for myself and future scientists to explore the effect of drought stress on maize. This data can also

be used to perform future metagenomic analyses to make informed decisions about crop safety during drought events. As this data and accompanying thesis will be published into the public domain, the purpose of this data will be to increase the knowledge around drought tolerance and adaptability, hopefully contributing to important crop research and having real-world consequences. Despite the link between DNA methylation and TE activation being well known previously, this research provides the first example of how drought affects varieties of maize with differing tolerances to drought. It also proves that work involving TEs and methylation are no longer as difficult as they once were, instead, advancements in computing power and the introduction of useful computer programs, such as *Jellyfish* and *DMRcaller*, are helping to provide exciting focal points for adaptation studies in crops.

6 Appendix

6.1 Chapter 3

6.1.1 Tables

Appendices Table 6.1.1.1. Read count of maize whole genome samples after two runs of next generation bisulfite sequencing.

Maize Variety	Tissue	Experimental condition	Replicate number	1st run read count	2nd run read count	Total read count
B73	Leaf	Stressed	1	15761847	13748221	29510068
B73	Leaf	Stressed	2	35849513	11752682	47602195
B73	Leaf	Stressed	3	23958134	13429667	37387801
B73	Leaf	Control	1	14670544	14169477	28840021
B73	Leaf	Control	2	45448292	24079597	69527889
B73	Leaf	Control	3	15213548	12547848	27761396
B76	Leaf	Stressed	1	29867993	15695601	45563594
B76	Leaf	Stressed	2	23849039	25355948	49204987
B76	Leaf	Stressed	3	27203603	27812369	55015972
B76	Leaf	Control	1	24284264	27028491	51312755
B76	Leaf	Control	2	21765743	13930858	35696601
B76	Leaf	Control	3	20663873	27689491	48353364
B73	Root	Stressed	1	17912292	14666913	32579205
B73	Root	Stressed	2	15307401	14403746	29711147
B73	Root	Stressed	3	14268402	11993377	26261779
B73	Root	Control	1	13905125	15099967	29005092
B73	Root	Control	2	12480285	13524568	26004853
B73	Root	Control	3	16503819	12561699	29065518
B76	Root	Stressed	1	21571665	28275841	49847506
B76	Root	Stressed	2	19624755	28050690	47675445
B76	Root	Stressed	3	23267860	28373075	51640935
B76	Root	Control	1	18669793	25706378	44376171
B76	Root	Control	2	26396420	23923350	50319770
B76	Root	Control	3	21598285	28602347	50200632

Appendices Table 6.1.1.2. A list of genes found within B76 root tissue that are associated with the "response to water deprivation" GO term group.

Gene name	Protein associated	Entrez number	GO term associated
GRMZM2G064426	CCAAT-DR1-transcription factor 15	100273424	GO:0009414
GRMZM2G083783	uncharacterized	100381912	GO:0009414
GRMZM2G038783	C2C2-CO-like-transcription factor 13	100281837	GO:0042631
GRMZM2G432926	Aquaporin PIP2-2	107648857	GO:0009269
GRMZM2G071659	E3 ubiquitin-protein ligase RGLG2	103631276	GO:0080148

Appendices Table 6.1.1.3 List of 166 genes with significantly differentially methylated promoter regions.

Gene Name	Variety	Tissue	Context
GRMZM2G069758	B73	Leaf	CG
GRMZM2G346455	B73	Leaf	CG
GRMZM2G048313	B73	Root	CG
GRMZM2G429972	B76	Leaf	CG
GRMZM2G016561	B76	Root	CG, CHG
GRMZM5G895799	B76	Root	CG
GRMZM2G070639	B73	Leaf	CHG
GRMZM2G132227	B73	Leaf	CHG
GRMZM5G858983	B73	Leaf	CHG
GRMZM2G144273	B73	Leaf, Root	CHG
AC203909.3_FG006	B76	Leaf	CHG
GRMZM2G314954	B76	Leaf	CHG
GRMZM2G034360	B73	Leaf	CHH
GRMZM2G139583	B73	Leaf	CHH
GRMZM2G177901	B73	Leaf	CHH
GRMZM2G335564	B73	Leaf	CHH
GRMZM2G354711	B73	Leaf	CHH
GRMZM2G381195	B73	Leaf	CHH
GRMZM5G803874	B73	Leaf	CHH
GRMZM2G081603	B73	Root	CHH
GRMZM2G118834	B73	Root	CHH
GRMZM2G349570	B73	Root	CHH
AC205331.3_FG004	B76	Leaf	CHH
AC207619.3_FG001	B76	Leaf	CHH

GRMZM2G019965	B76	Leaf	CHH
GRMZM2G021831	B76	Leaf	CHH
GRMZM2G022777	B76	Leaf	CHH
GRMZM2G025646	B76	Leaf	CHH
GRMZM2G025680	B76	Leaf	CHH
GRMZM2G034835	B76	Leaf	CHH
GRMZM2G040467	B76	Leaf	CHH
GRMZM2G042662	B76	Leaf	CHH
GRMZM2G044100	B76	Leaf	CHH
GRMZM2G053690	B76	Leaf	CHH
GRMZM2G053766	B76	Leaf	CHH
GRMZM2G060742	B76	Leaf	CHH
GRMZM2G080079	B76	Leaf	CHH
GRMZM2G083195	B76	Leaf	CHH
GRMZM2G085932	B76	Leaf	CHH
GRMZM2G100084	B76	Leaf	CHH
GRMZM2G129243	B76	Leaf	CHH
GRMZM2G162505	B76	Leaf	CHH
GRMZM2G163830	B76	Leaf	CHH
GRMZM2G166759	B76	Leaf	CHH
GRMZM2G349243	B76	Leaf	CHH
GRMZM2G360821	B76	Leaf	CHH
GRMZM2G369803	B76	Leaf	CHH
GRMZM2G466517	B76	Leaf	CHH
GRMZM2G511318	B76	Leaf	CHH
GRMZM2G542753	B76	Leaf	CHH
GRMZM5G810246	B76	Leaf	CHH
GRMZM5G822313	B76	Leaf	CHH
GRMZM5G868062	B76	Leaf	CHH
GRMZM5G883510	B76	Leaf	CHH
AC187037.4_FG008	B76	Root	CHH
AC199068.2_FG017	B76	Root	CHH
AC202000.4_FG006	B76	Root	CHH
AC203535.4_FG001	B76	Root	CHH
AC209819.3_FG003	B76	Root	CHH
AC211669.4_FG003	B76	Root	CHH
AC213050.3_FG002	B76	Root	CHH
AC213769.3_FG001	B76	Root	CHH
AC233956.1_FG002	B76	Root	CHH
GRMZM2G003489	B76	Root	CHH
GRMZM2G005865	B76	Root	CHH
GRMZM2G016281	B76	Root	CHH

GRMZM2G017388	B76	Root	CHH
GRMZM2G017789	B76	Root	CHH
GRMZM2G018760	B76	Root	CHH
GRMZM2G019819	B76	Root	CHH
GRMZM2G025793	B76	Root	CHH
GRMZM2G028763	B76	Root	CHH
GRMZM2G029407	B76	Root	CHH
GRMZM2G031398	B76	Root	CHH
GRMZM2G032024	B76	Root	CHH
GRMZM2G035785	B76	Root	CHH
GRMZM2G039089	B76	Root	CHH
GRMZM2G039954	B76	Root	CHH
GRMZM2G055450	B76	Root	CHH
GRMZM2G059015	B76	Root	CHH
GRMZM2G059020	B76	Root	CHH
GRMZM2G060876	B76	Root	CHH
GRMZM2G064426	B76	Root	CHH
GRMZM2G066997	B76	Root	CHH
GRMZM2G071196	B76	Root	CHH
GRMZM2G071659	B76	Root	CHH
GRMZM2G072939	B76	Root	CHH
GRMZM2G074356	B76	Root	CHH
GRMZM2G078954	B76	Root	CHH
GRMZM2G080556	B76	Root	CHH
GRMZM2G080851	B76	Root	CHH
GRMZM2G085974	B76	Root	CHH
GRMZM2G091191	B76	Root	CHH
GRMZM2G092123	B76	Root	CHH
GRMZM2G107302	B76	Root	CHH
GRMZM2G108309	B76	Root	CHH
GRMZM2G109624	B76	Root	CHH
GRMZM2G114315	B76	Root	CHH
GRMZM2G114650	B76	Root	CHH
GRMZM2G122941	B76	Root	CHH
GRMZM2G124759	B76	Root	CHH
GRMZM2G126083	B76	Root	CHH
GRMZM2G126900	B76	Root	CHH
GRMZM2G129288	B76	Root	CHH
GRMZM2G130634	B76	Root	CHH
GRMZM2G133563	B76	Root	CHH
GRMZM2G133972	B76	Root	CHH
GRMZM2G141799	B76	Root	CHH

GRMZM2G144671	B76	Root	CHH
GRMZM2G145978	B76	Root	CHH
GRMZM2G149649	B76	Root	CHH
GRMZM2G157061	B76	Root	CHH
GRMZM2G163184	B76	Root	CHH
GRMZM2G165535	B76	Root	CHH
GRMZM2G166459	B76	Root	CHH
GRMZM2G169973	B76	Root	CHH
GRMZM2G301148	B76	Root	CHH
GRMZM2G321033	B76	Root	CHH
GRMZM2G323413	B76	Root	CHH
GRMZM2G331533	B76	Root	CHH
GRMZM2G352926	B76	Root	CHH
GRMZM2G359298	B76	Root	CHH
GRMZM2G367026	B76	Root	CHH
GRMZM2G371033	B76	Root	CHH
GRMZM2G374313	B76	Root	CHH
GRMZM2G375015	B76	Root	CHH
GRMZM2G376619	B76	Root	CHH
GRMZM2G414475	B76	Root	CHH
GRMZM2G423518	B76	Root	CHH
GRMZM2G425545	B76	Root	CHH
GRMZM2G429442	B76	Root	CHH
GRMZM2G432926	B76	Root	CHH
GRMZM2G450833	B76	Root	CHH
GRMZM2G458423	B76	Root	CHH
GRMZM2G460090	B76	Root	CHH
GRMZM2G460988	B76	Root	CHH
GRMZM2G465226	B76	Root	CHH
GRMZM2G471253	B76	Root	CHH
GRMZM2G501086	B76	Root	CHH
GRMZM2G557158	B76	Root	CHH
GRMZM2G702386	B76	Root	CHH
GRMZM5G897944	B76	Root	CHH
AC189771.3_FG001	B76	Leaf, Root	CHH
AC203366.4_FG001	B76	Leaf, Root	CHH
GRMZM2G018020	B76	Leaf, Root	CHH
GRMZM2G062585	B76	Leaf, Root	CHH
GRMZM2G075229	B76	Leaf, Root	CHH
GRMZM2G082707	B76	Leaf, Root	CHH
GRMZM2G105657	B76	Leaf, Root	CHH
GRMZM2G108894	B76	Leaf, Root	CHH

GRMZM2G127780	B76	Leaf, Root	CHH
GRMZM2G145996	B76	Leaf, Root	CHH
GRMZM2G151015	B76	Leaf, Root	CHH
GRMZM2G152436	B76	Leaf, Root	CHH
GRMZM2G158831	B76	Leaf, Root	CHH
GRMZM2G175513	B76	Leaf, Root	CHH
GRMZM2G379804	B76	Leaf, Root	CHH
GRMZM2G428933	B76	Leaf, Root	CHH
GRMZM2G459642	B76	Leaf, Root	CHH
GRMZM2G704475	B76	Leaf, Root	CHH
GRMZM5G878070	B76	Leaf, Root	CHH
GRMZM5G899855	B76	Leaf, Root	CHH
AC191264.3_FG002	B76	Leaf, Root	CHH
GRMZM2G000729	B76	Leaf, Root	CHH
GRMZM2G079538	B76	Leaf, Root	CHH
GRMZM2G477533	B76	Leaf, Root	CHH

6.1.2   SeqMonk analysis
1. Create new project -> Genomes -> Zea Mays -> AGPv3
2. File -> import Data -> Text (Generic) -> chose the merged sample files
    a. Start at row 0
    b. Chr col = 3
    c. Start and End = 4
    d. Strand = 2
3. Data -> Define Probes -> Read position probe generator
    a. Select all data sets
    b. All reads
    c. Min count = 1
    d. Valid positions = 100
    e. Ignore strand = Yes
4. Read Count Quant
    a. All reads
    b. Correct for total read count
    c. Largest data store
    d. Don't log transform
5. Filtering -> by statistical test -> Outlier stats -> Box Whisker
    a. Stringency > 10 Above median
    b. At least 1 of selected data stores
    c. Right click new probe list created and convert to annotation track
6. File -> import data -> visible data stores
    a. By excluding the newly created annotation track
7. Data -> Edit data sets
    a. Delete the old data sets
8. Data -> Edit replicate sets

       a. Add new replicate sets called control and drought
       b. Put the correct samples in the correct set
9. Data -> Define Probes -> Read position probe generator
       a. Select control and drought data sets
       b. All reads
       c. Min count = 1
       d. Valid positions = 100
       e. Ignore strand = Yes
       f. Close out of pop up window
10. Data -> quantitation pipelines
       a. Bisulphite methylation over features
       b. Over existing probes
       c. Min read count = 4
       d. Apply minimum count over all
       e. Minimum obs = 1
11. Filtering -> by probe length
       a. 1 to 10000
12. Highlight New probe list
13. Filter by statistical test -> Proportion based -> replicated -> Logistic regression
       a. Highlight Control and drought
       b. P-value = 0.05 (Bonferroni Correction = 0.05/Probe number)
       c. Min obs = 10 (For SeqMonk 1.40.0 this value is not included and replaced by abs diff cut-off of 5)
       d. Apply multiple testing
14. Highlight new probe list created
15. Filtering -> by features
       a. Feature to design around = gene
       b. Upstream of feature from –2000 to 0
16. While highlighting Logistic regression
       a. Reports -> Annotated Probe Report
          i. 2000bp Upstream of gene
          ii. Exclude unannotated probes
          iii. Save as B76_CHG_leaf_PromGene

6.1.3   Finding genes containing promoter DMRs

Input file is the SeqMonk output file (e.g. B76_CHG_leaf_PromGene.txt).

1. Remove double tabs from these files in BBEdit
       a. Open the files in BBEdit
       b. Cmd+F -> replace all \t\t with \t
2. Extract just the gene names
       a. for i in *PromGene.txt ; do f=`basename $i .txt` ; cut -f7 $i | grep -v "ID" > $f.GeneNames.txt ; done
3. Remove repeats
       a. for i in *GeneNames.txt ; do f=`basename $i .GeneNames.txt` ; sort -u $i > $f.Uniq.GeneNames.txt ; done
4. Delete non-unique files

5. Separate the genes into a unique list for each variety combination
   a. perl seperateuniquegenes2.pl
      B76_CG_leaf_PromGene.Uniq.GeneNames.txt B76_CG_root_
      PromGene.Uniq.GeneNames.txt B73_CG_leaf_
      PromGene.Uniq.GeneNames.txt B73_CG_root_
      PromGene.Uniq.GeneNames.txt
6. Compare this list with *DMRcaller* list to find matching genes

6.1.4   DMRcaller Analysis
1. Created DMRcaller files from .cov files (From Bismark) for use for DMRcaller
   in R
   a. for i in *cov.gz; do f=`basename $i` ; nohup coverage2cytosine --
      genome_folder DMR/ -CX -o ${f%.cov.gz}.DMRCaller ${i} & done
   b. split files by chromosome

2. Created the large split script to turn DMRcaller files into RData files for
   DMRcaller.R
   a. for i in {1..10} ; do for m in *chr$i.gz ; do gunzip $m ; done ; for j in
      B73 B76 ; do for k in leaf root ; do Rscript DMRCalling.05.R $i $j $k" ;
      done ; done ; for n in *chr$i ; do gzip $n ; done ; done
3. Ran Genome analysis R script to determine DMR numbers, heatmaps, and
   DMR intercept numbers.
   a. for i in B73 B76 ; do for f in root leaf ; do Rscript
      GenomeDMRAnalysisScripts2.R $i $f ; done ; done
4. Then ran an R code to get gene names that have DMRs in their promoter
   regions
   a. for i in B73 B76 ; do for j in leaf root ; do Rscript DMRGenes2.R $i $j ;
      done ; done
5. Compared the DMRcaller gene list with the SeqMonk gene list
   a. for i in B76 B73 ; do for j in CG CHG CHH ; do for k in leaf root ; do
      s="$(tr '[:lower:]' '[:upper:]' <<< ${k:0:1})${k:1}" ; do perl
      matchingGenes.pl ${j}_${i}_${k}_Prom_Genes.txt {j}.${i}${s}Only.txt ;
      done ; done ; done

6.1.5   Analysis of matching genes
1. Run geneID2GOTerms.pl on gene list to get function counts and GO terms
   for each gene
   a. for i in *Only.txt ; do perl geneID2GOTermsv2.pl maize_v3.agg.gaf $i
      ; done ; mkdir GOterms ; mv *Gene+GOT* GOterms/ ; mkdir
      GOfunctionFiles ; mv *GOfunctionCount* GOfunctionFiles/
2. Count how many times each Biological_Process GO term shows up in these
   genes
   a. for i in * ; do perl GO_TermCount.pl $i GOTerms.txt ; done ; mkdir
      GOtermCounts ; mv *GOcount.txt GOtermCounts/

6.1.6   Scripts used
Viewable at:

https://drive.google.com/drive/folders/1qWJz0mPV0MLuBxn-uCL_v5RMbzw-dJlp?usp=sharing

seperateuniquegenes2.pl
DMRCalling.05.R
GenomeDMRAnalysisScripts2.R
DMRGenes2.R
matchingGenes.pl
geneID2GOTermsv2.pl
GO_TermCount.p

6.2 Chapter 4

6.2.1 Tables

Appendices Table 6.2.1.1. A table showing the number of significantly differentially methylated families after drought stress, found within each Class, Order and Superfamily. The total number of families found in each Class, Order and Superfamily are also included for comparison.

	Class I - Retrotransposons					Class II – DNA Transposons – Subclass 1						Class II – DNA Transposons – Subclass 2	PPP_PPO
	TIRs					LINEs		LTRs			SINES	Helitrons	
	hAT	CACTA	PIF–Harbinger	Mutator	Tc1-Mariner	L1	RTE	Copia	Gypsy	Unclassified LTR	tRNA	Helitron	
B73 Leaf	6	5	3	10	0	2	0	15	24	8	0	0	1
B76 Leaf	15	5	4	11	1	1	0	39	58	17	0	0	0
B73 Root	6	21	2	6	0	0	0	42	78	10	0	3	1
B76 Root	27	5	5	18	1	3	0	22	69	22	0	3	0

	hAT	CACTA	PIF–Harbinger	Mutator	Tc1-Mariner	L1	RTE	Copia	Gypsy	Unclassified LTR	tRNA	Helitron	PPP_PPO
Total in Reference	387	126	182	138	59	30	2	154	244	181	6	16	1

6.2.2   UTE file preparation

1. Ran SMALT on adapter trimmed files
    a. perl SMALT_pipe.pl knobC.index UTE.index FGS.index
       RM02_S0_L007_R1_001.truncated RM02_S0_L007_R1_001
2. Coverted UTE bams to fq files:
    a. bedtools bamtofastq -i RM02_S0_L007_R1_001.smalt.UTE.bam -fq
       RM02_S0_L007.R1.fq
3. Merged R1's and R2's using cat
    a. cat RM02_S0_L007.R1.fq RM02_S0_L007.R2.fq >
       RM02_L007.merged.fq
4. Run *Jellyfish* on fq's to create hash of 35mers, any count above 2
    a. zcat RM02_S0_L007.UTE.merged.fq.gz | jellyfish count -m 35 -s
       3000000000 -t 10 -o RM05_L006.hash -c 7 -C -L 2
5. Grab the kmer count info from the hash
    a. nohup jellyfish dump -o RM02_L007.kmerCount.fa -L 2
       RM02_L007.hash &
6. Converted jellyfish .fa counts into degenerative files and v2.txt files that are
   more manageable than the .fa
    a. perl jellyFa2degen.pl RM02_L007.kmerCount.fa.gz
7. Add annotation info into these files for this and reverse complemented
   version
    a. perl AddAnno2degen+revCompl.pl UTE.sorted.txt
       revCompl.UTE.sorted.txt degen.RM02_L007.kmerCount.txt.gz &
8. Counted number of C's and T's
    a. perl masterBam2Finalv2.pl
       anno.modified.degen.mergedSample1.kmerCount.txt.gz
       mergedSample1.kmerCount.txt.gz.v2.txt.gz
9. Got rid of the non-T kmers
    a. gunzip -c RM02_L007.kmerCount.fa.v2.txt.gz.compared.txt.gz | awk
       '$1 ~ /T/ {print}' >
       modified.RM02_L007.kmerCount.fa.v2.txt.gz.compared.txt
10. Analyse C:T ratio
    a. perl C%Analysis.pl
       modified.RM02_L007.kmerCount.fa.v2.txt.gz.compared.txt.gz
11. Create threshold for samples to be compared against
    a. zcat
       modified.RM02_L007.kmerCount.v2.txt.gz.compared.txt.gz.analysis.t
       xt.gz | perl -e 'while ($line = <>) {chomp $line ; ($kmer, $locs, $n,
       $calls) = split /\s+/, $line; @c = split /\|\|/, $calls ; undef @goodSites
       ; foreach $pos (@c) {($p, $lev) = split /\;/, $pos ; if ($lev >= 0.9) {push
       @goodSites, $p}} @goodSites = sort {$a <=> $b} @goodSites ; next if
       (@goodSites < 4) ; $newSites = join ";", @goodSites ; print
       $kmer."\t".$newSites."\n"}' >
       modified.RM02_L007.kmerCount.v2.min90percC.min4sites.txt

12. This creates a modified.RM02_L007.kmerCount.v2.min90percC.min4sites.txt file that can be compared to the samples using sampleKmerCull.pl

6.2.3   Sample file preparation

1. Merge pair1 and pair2 using cat
    a. cat sample1.pair1.truncated.gz sample1.pair2.truncated.gz > sample1.combined.gz
2. Repeated steps 4-10 from UTE preparation section on samples
3. Compared samples to UTE references
    a. perl sampleKmerCull.pl modified.RM05_L006.kmerCount.v2.min90percC.min4sites.txt sample1.kmerCount.analysis.txt.gz > sample1.min90percC.min4sites.txt
4. Converted the sample min90 files into Rinput files
    a. cat sample1.min90percC.min4sites.txt | awk '{print $2"\t"$3"\t"$5}' | sed 's/|.//' > sample1.Rinput.txt
5. Ran the resulting files in R using the MethRatioWilcoxonTEv2.R script

6.2.4   Matching kmers to promoter regions

1. Compare significant TEs with analysis files and grab the kmers associated with each TE, store in a big file, one for stressed and one for control
    a. perl sigTEList2kmers.pl B76_Leaf_0.01.txt sample1.min90percC.min4sites.txt
2. get rid of new lines in .fa files for easier comparison for the cds
    a. awk '!/^>/ { printf "%s", $0; n = "\n" } /^>/ { print n $0; n = "" } END { printf "%s", n }' Zea_mays.AGPv4.cds.all.fa | awk '/^>/ { a=$0 } !/^>/ { print a"\t"$0}' > Zea_mays.AGPv4.collapsed.cds.txt
3. Create the bed files using the cds files collapsed files, sorted them
    a. cut -d " " -f3 Zea_mays.AGPv4.collapsed.cds.txt | sed 's/:/   /g' | cut -f3,4,5,6
    b. sort -nk 1,1 -nk 2,2 Zea_mays.AGPv4.collapsed.cds.bed > sorted.Zea_mays.AGPv4.collapsed.cds.bed
    c. perl sortedbed2promoter.pl sorted.Zea_mays.AGPv4.collapsed.cds.bed
4. Remove contigs
    a. grep -v "B73" sorted.Zea_mays.AGPv4.cds.promoters.bed > contigless.Zea_mays.AGPv4.cds.promoters.bed
5. Get the promoter regions FASTA files
    a. bedtools getfasta -s -fi Zea_mays.AGPv4.dna.toplevel.fa -bed contigless.Zea_mays.AGPv4.cds.promoters.bed -fo promoters.fa -name
6. collapse this FASTA file to make it easier to work with

a. awk '/^>/ { a=$0 } !/^>/ { print a"\t"$0}' promoters.fa > collapsed.promoters.txt

7. Made a reverse complemented version too
   a. perl revProm.pl collapsed.promoters.txt

8. Match the kmers found in significant TEs to the promoters
   a. perl kmerPromoterMatch.pl collapsed.promoters.txt revCompl.collapsed.promoters.txt sample1 .sigTEs.txt

9. Grab the genes with sig diff methylated TEs in promoter regions and their associated TEs
   a. cat sample1.sigTEinProm.txt sample2.sigTEinProm.txt sample3.sigTEinProm.txt | cut -f1,2 | sed 's/\(_\).*\(\)/   /' | sed 's/|.*//' > B76_Root_Stress.txt
   b. cut -f1 B73_Root_Stress.txt | sort -u > uniq.B73_Root_Stress.txt

10. Searched through these gene lists for genes found in chapter 3
    a. grep "genename" genelist.txt

11. Did the same with the the genes including cds/introns/exons/UTRs using bed file
    a. grep -v "B73" sorted.Zea_mays.AGPv4.collapsed.gene.bed > contigless.Zea_mays.AGPv4.cds.collapsed.gene.bed
    b. awk '{if($5 == "1") $5="+";}1' OFS=\\t contigless.Zea_mays.AGPv4.cds.collapsed.gene.bed | awk '{if($5 == "-1") $5="-";}1' OFS=\\t > contigless.Zea_mays.AGPv4.gene.collapsed.bed
    c. bedtools getfasta -s -fi Zea_mays.AGPv4.dna.toplevel.fa -bed contigless.Zea_mays.AGPv4.gene.collapsed.bed -fo cds.fa -name
    d. awk '/^>/ { a=$0 } !/^>/ { print a"\t"$0}' genes.fa > collapsed.genes.txt
    e. perl revProm.pl collapsed.genes.txt
    f. perl kmerAllGenicMatch.pl collapsed.genes.txt revCompl.collapsed.genes.txt sample10.min90percC.min4sites.sigTEs.txt

12. Find the Genes with the same sig TEs in the Genic regions, to remove them from promoter list, then count the number of individual genes removed

6.2.5   Scripts used
Viewable at:
https://drive.google.com/drive/folders/1qWJz0mPV0MLuBxn-uCL_v5RMbzw-dJlp?usp=sharing

SMALT_pipe.pl
jellyFa2degen.pl
AddAnno2degen+revCompl.pl
masterBam2Finalv2.pl
C%Analysis.pl
sampleKmerCull.pl
MethRatioWilcoxonTEv3.R
sigTEList2kmers.pl

kmerPromoterMatch.pl
revProm.pl
kmerAllGenicMatch.pl

7   Bibliography

Ahmad, P., Rasool, S., Gul, A., Sheikh, S.A., Akram, N.A., Ashraf, M., Kazi, A.M. & Gucel, S. 2016, 'Jasmonates: Multifunctional Roles in Stress Tolerance', *Frontiers in Plant Science*.

Aimar, D., Calafat, M., Andrade, A.M., Carassay, L., Bouteau, F., Abdala, G. & Molas, M.L. 2014, 'Drought effects on the early development stages of Panicum virgatum L.: Cultivar differences', *Biomass and Bioenergy*, vol. 66, pp. 49–59.

Akram, N.A., Shafiq, F. & Ashraf, M. 2017, 'Ascorbic Acid-A Potential Oxidant Scavenger and Its Role in Plant Development and Abiotic Stress Tolerance.', *Frontiers in plant science*, vol. 8, p. 613.

Ali, Q., Ashraf, M. & Athar, H.-U.-R. 2007, 'Exogenously applied proline at different growth stages enhances growth of two maize cultivars grown under water deficit conditions', *Pak. J. Bot*, vol. 39, no. 4, pp. 1133–44.

Allaby, R.G., Ware, R.L. & Kistler, L. 2019, 'A re-evaluation of the domestication bottleneck from archaeogenomic evidence', *Evolutionary Applications*, vol. 12, no. 1, pp. 29–37.

Anderson, W.B., Seager, R., Baethgen, W., Cane, M. & You, L. 2019, 'Synchronous crop failures and climate-forced production variability', *Science Advances*, vol. 5, no. 7, p. eaaw1976.

Anjum, S.A., Ashraf, U., Tanveer, M., Khan, I., Hussain, S., Shahzad, B., Zohaib, A., Abbas, F., Saleem, M.F., Ali, I. & Wang, L.C. 2017, 'Drought Induced Changes in Growth, Osmolyte Accumulation and Antioxidant Metabolism of Three Maize Hybrids', *Frontiers in Plant Science*, vol. 08, p. 69.

Anjum, S.A., Wang, L., Farooq, M., Khan, I. & Xue, L. 2011, 'Methyl Jasmonate-Induced Alteration in Lipid Peroxidation, Antioxidative Defence System and Yield in Soybean Under Drought', *Journal of Agronomy and Crop Science*, vol. 197, no. 4, pp. 296–301.

Aprile, A., Mastrangelo, A.M., De Leonardis, A.M., Galiba, G., Roncaglia, E., Ferrari, F., De Bellis, L., Turchi, L., Giuliano, G. & Cattivelli, L. 2009, 'Transcriptional profiling in response to terminal drought stress reveals differential responses along the wheat genome.', *BMC genomics*, vol. 10, p. 279.

Aslam, M., Zamir, M.S.I., Anjum, S.A., Khan, I. & Tanveer, M. 2015, ' An investigation into morphological and physiological approaches to screen maize (Zea mays L.) hybrids for drought tolerance ', *Cereal Research Communications*, vol. 43, no. 1, pp. 41–51.

Baek, D., Jiang, J., Chung, J.-S., Wang, B., Chen, J., Xin, Z. & Shi, H. 2011, 'Regulated AtHKT1 Gene Expression by a Distal Enhancer Element and DNA Methylation in the Promoter Plays an Important Role in Salt Tolerance', *Plant and Cell Physiology*, vol. 52, no. 1, pp. 149–61.

Banerjee, A. & Roychoudhury, A. 2017, 'Epigenetic regulation during salinity and drought stress in plants: Histone modifications and DNA methylation', *Plant Gene*.

Barbaglia, A.M., Klusman, K.M., Higgins, J., Shaw, J.R., Hannah, L.C. & Lal, S.K. 2012, 'Gene Capture by Helitron Transposons Reshuffles the Transcriptome of Maize', *Genetics*, vol. 190, no. 3, p. 965.

Barrera-Figueroa, B.E., Gao, L., Wu, Z., Zhou, X., Zhu, J., Jin, H., Liu, R. & Zhu, J.-K.

2012, 'High throughput sequencing reveals novel and abiotic stress-regulated microRNAs in the inflorescences of rice.', *BMC plant biology*, vol. 12, p. 132.

Barrón, M.G., Fiston-Lavier, A.-S., Petrov, D.A. & González, J. 2014, ' Population Genomics of Transposable Elements in Drosophila ', *Annual Review of Genetics*, vol. 48, no. 1, pp. 561–81.

Bartels, A., Han, Q., Nair, P., Stacey, L., Gaynier, H., Mosley, M., Huang, Q., Pearson, J., Hsieh, T.-F., An, Y.-Q. & Xiao, W. 2018, 'Dynamic DNA Methylation in Plant Growth and Development', *International Journal of Molecular Sciences*, vol. 19, no. 7, p. 2144.

Baucom, R.S., Estill, J.C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J.-M., Westerman, R.P., Sanmiguel, P.J. & Bennetzen, J.L. 2009, 'Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome.', *PLoS genetics*, vol. 5, no. 11, p. e1000732.

Bernatavichute, Y. V., Zhang, X., Cokus, S., Pellegrini, M. & Jacobsen, S.E. 2008, 'Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in Arabidopsis thaliana', B.P. Dilkes (ed.),*PLoS ONE*, vol. 3, no. 9, p. e3156.

Bewick, A.J. & Schmitz, R.J. 2017, 'Gene body DNA methylation in plants', *Current Opinion in Plant Biology*.

Bhaskara, G.B., Wen, T.-N., Nguyen, T.T. & Verslues, P.E. 2017, 'Protein Phosphatase 2Cs and Microtubule-Associated Stress Protein 1 Control Microtubule Stability, Plant Growth, and Drought Response.', *The Plant cell*, vol. 29, no. 1, pp. 169–91.

Bolduc, N., Yilmaz, A., Mejia-Guerra, M.K., Morohashi, K., O'Connor, D., Grotewold, E. & Hake, S. 2012, 'Unraveling the KNOTTED1 regulatory network in maize meristems.', *Genes & development*, vol. 26, no. 15, pp. 1685–90.

Bouwmeester, K. & Govers, F. 2009, 'Arabidopsis L-type lectin receptor kinases: phylogeny, classification, and expression profiles', *Journal of Experimental Botany*, vol. 60, no. 15, pp. 4383–96.

Boyer, J.S., Byrne, P., Cassman, K.G., Cooper, M., Delmer, D., Greene, T., Gruis, F., Habben, J., Hausmann, N., Kenny, N., Lafitte, R., Paszkiewicz, S., Porter, D., Schlegel, A., Schussler, J., Setter, T., Shanahan, J., Sharp, R.E., Vyn, T.J., Warner, D. & Gaffney, J. 2013, 'The U.S. drought of 2012 in perspective: A call to action', *Global Food Security*, vol. 2, no. 3, pp. 139–43.

Brossa, R., López-Carbonell, M., Jubany-Marí, T. & Alegre, L. 2011, 'Interplay Between Abscisic Acid and Jasmonic Acid and its Role in Water-oxidative Stress in Wild-type, ABA-deficient, JA-deficient, and Ascorbate-deficient Arabidopsis Plants', *Journal of Plant Growth Regulation*, vol. 30, no. 3, pp. 322–33.

Bucher, E., Reinders, J. & Mirouze, M. 2012, 'Epigenetic control of transposon transcription and mobility in Arabidopsis', *Current Opinion in Plant Biology*, vol. 15, no. 5, pp. 503–10.

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G. & Martin, C. 2012, 'Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges.', *The Plant cell*, vol. 24, no. 3, pp. 1242–55.

Campbell, M.T., Proctor, C.A., Dou, Y., Schmitz, A.J., Phansak, P., Kruger, G.R., Zhang, C. & Walia, H. 2015, 'Genetic and molecular characterization of submergence response identifies Subtol6 as a major submergence tolerance locus in maize.',

PloS one, vol. 10, no. 3, p. e0120385.

Candaele, J., Demuynck, K., Mosoti, D., Beemster, G.T.S., Inzé, D. & Nelissen, H. 2014, 'Differential methylation during maize leaf growth targets developmentally regulated genes.', *Plant physiology*, vol. 164, no. 3, pp. 1350–64.

Caparrós-Martín, J.A., Reiland, S., Köchert, K., Cutanda, M.C. & Culiáñez-Macià, F.A. 2007, 'Arabidopsis thaliana AtGpp1 and AtGpp2: two novel low molecular weight phosphatases involved in plant glycerol metabolism', *Plant Molecular Biology*, vol. 63, no. 4, pp. 505–17.

Casacuberta, E. & González, J. 2013, 'The impact of transposable elements in environmental adaptation', *Molecular Ecology*, vol. 22, no. 6, pp. 1503–17.

Casati, P. & Walbot, V. 2003, 'Gene expression profiling in response to ultraviolet radiation in maize genotypes with varying flavonoid content.', *Plant physiology*, vol. 132, no. 4, pp. 1739–54.

Cassone, B.J., Chen, Z., Chiera, J., Stewart, L.R. & Redinbaugh, M.G. 2014, 'Responses of highly resistant and susceptible maize to vascular puncture inoculation with Maize dwarf mosaic virus', *Physiological and Molecular Plant Pathology*, vol. 86, pp. 19–27.

Castelletti, S., Tuberosa, R., Pindo, M. & Salvi, S. 2014, 'A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL Vgt1.', *G3 (Bethesda, Md.)*, vol. 4, no. 5, pp. 805–12.

Catoni, M., Jonesman, T., Cerruti, E. & Paszkowski, J. 2019, 'Mobilization of Pack-CACTA transposons in Arabidopsis suggests the mechanism of gene shuffling', *Nucleic Acids Research*, vol. 47, no. 3, pp. 1311–20.

Catoni, M., Tsang, J.M.F., Greco, A.P. & Zabet, N.R. 2018, 'DMRcaller: A versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts', *Nucleic Acids Research*.

Cattivelli, L., Rizza, F., Badeck, F.W., Mazzucotelli, E., Mastrangelo, A.M., Francia, E., Marè, C., Tondelli, A. & Stanca, A.M. 2008, 'Drought tolerance improvement in crop plants: An integrated view from breeding to genomics', *Field Crops Research*, vol. 105, no. 1–2, pp. 1–14.

Challinor, A.J., Koehler, A.-K., Ramirez-Villegas, J., Whitfield, S. & Das, B. 2016, 'Current warming will reduce yields unless maize breeding and seed systems adapt immediately', *Nature Climate Change*, vol. 6, no. 10, pp. 954–8.

Chan, S.W.-L., Henderson, I.R. & Jacobsen, S.E. 2005, 'Gardening the genome: DNA methylation in Arabidopsis thaliana', *Nature Reviews Genetics*, vol. 6, no. 5, pp. 351–60.

Chen, J., Xu, W., Velten, J., Xin, Z. & Stout, J. 2012, 'Characterization of maize inbred lines for drought and heat tolerance', *Journal of Soil and Water Conservation*, vol. 67, no. 5, pp. 354–64.

Cheng, S., Huang, Y., Zhu, N. & Zhao, Y. 2014, 'The rice WUSCHEL-related homeobox genes are involved in reproductive organ development, hormone signaling and abiotic stress response', *Gene*, vol. 549, no. 2, pp. 266–74.

Choudhary, N.L., Sairam, R.K. & Tyagi, A. 2005, 'Expression of delta1-pyrroline-5-carboxylate synthetase gene during drought in rice (Oryza sativa L.).', *Indian journal of biochemistry & biophysics*, vol. 42, no. 6, pp. 366–70.

Chugh, V., Kaur, N., Grewal, M.S. & Gupta, A.K. 2013, 'Differential antioxidative response of tolerant and sensitive maize (Zea mays L.) genotypes to drought

stress at reproductive stage.', *Indian journal of biochemistry & biophysics*, vol. 50, no. 2, pp. 150–8.

Colaneri, A.C. & Jones, A.M. 2013, 'Genome-Wide Quantitative Identification of DNA Differentially Methylated Sites in Arabidopsis Seedlings Growing at Different Water Potential', *PLoS ONE*, vol. 8, no. 4, p. e59878.

Conklin, P.L., Norris, S.R., Wheeler, G.L., Williams, E.H., Smirnoff, N. & Last, R.L. 1999, 'Genetic evidence for the role of GDP-mannose in plant ascorbic acid (vitamin C) biosynthesis.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 7, pp. 4198–203.

Cook, E.R., Woodhouse, C.A., Eakin, C.M., Meko, D.H. & Stahle, D.W. 2004, 'Long-term aridity changes in the western United States', *Science*.

Correia, B., Hancock, R.D., Amaral, J., Gomez-Cadenas, A., Valledor, L. & Pinto, G. 2018, 'Combined Drought and Heat Activates Protective Responses in Eucalyptus globulus That Are Not Activated When Subjected to Drought or Heat Stress Alone', *Frontiers in Plant Science*, vol. 9.

Costa, L.M., Yuan, J., Rouster, J., Paul, W., Dickinson, H. & Gutierrez-Marcos, J.F. 2012, 'Maternal control of nutrient allocation in plant seeds by genomic imprinting.', *Current biology : CB*, vol. 22, no. 2, pp. 160–5.

Cruz de Carvalho, M.H. 2008, 'Drought stress and reactive oxygen species: Production, scavenging and signaling.', *Plant signaling & behavior*, vol. 3, no. 3, pp. 156–65.

Cui, Xiekui, Jin, P., Cui, Xia, Gu, L., Lu, Z., Xue, Y., Wei, L., Qi, J., Song, X., Luo, M., An, G. & Cao, X. 2013, 'Control of transposon activity by a histone H3K4 demethylase in rice', *Proceedings of the National Academy of Sciences*, vol. 110, no. 5, pp. 1953–8.

Daron, Josquin, Glover, Natasha, Pingault, Lise, Theil, Sébastien, Jamilloux, Véronique, Paux, Etienne, Barbe, Valérie, Mangenot, Sophie, Alberti, Adriana, Wincker, Patrick, Quesneville, Hadi, Feuillet, Catherine, Choulet, Frédéric, McClintock, B., Feschotte, C., Jiang, N., Wessler, S., Arabidopsis, G., Schnable, P., Ware, D., Fulton, R., Stein, J., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T., Devos, K., Bennetzen, J., Ma, J., Devos, K., Tenaillon, M., Hollister, J., Gaut, B., Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C., Richardson, A., Okumoto, Y., Tanisaka, T., Wessler, S., Li, B., Choulet, F, Heng, Y., Hao, W., Paux, E, Liu, Z., Yue, W., Jin, W., Feuillet, C, Zhang, X., Vitte, C., Bennetzen, J., Matsumoto, T., Wu, J., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., Mizuno, H., Yamamoto, K., Antonio, B., Baba, T., Paterson, A., Bowers, J., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Gao, X., Hou, Y., Ebina, H., Levin, H., Voytas, D., Marín, I., Lloréns, C., Chatterjee, A., Leem, Y., Kelly, F., Levin, H., Devos, K., Brown, J., Bennetzen, J., Ma, J., Devos, K., Bennetzen, J., Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J., Jackson, SA, Gaut, B., Ma, J., Duret, L., Marais, G., Biémont, C., Long, M., Betrán, E., Thornton, K., Wang, W., Yang, L., Bennetzen, J., Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., Rafalski, A., Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R., Lisch, D., Meyers, B., Shiu, S.-H., Jiang, N., Jiang, N., Wessler, S., Juretic, N., Bureau, T., Bruskiewich, R., Jiang, N., Feschotte, C., Zhang, X., Wessler, S., Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., Kuang, H., Chen, J., Hu, Q., Zhang, Y., Lu, C., Kuang, H., Bennetzen, J., Coleman, C., Liu, R., Ma, J.,

Ramakrishna, W., Lerat, E., Choulet, F, Wicker, T., Rustenholz, C., Paux, E, Salse, J., Leroy, P., Schlub, S., Paslier, M.-C. Le, Magdelenat, G., Gonthier, C., SanMiguel, PJ, Ramakrishna, W., Bennetzen, J., Busso, C., Dubcovsky, J., Li, W., Zhang, P., Fellers, J., Friebe, B., Gill, B., Sabot, F., Guyot, R., Wicker, T., Chantret, N., Laubin, B., Chalhoub, B., Leroy, P., Sourdille, P., Bernard, M., Paux, E, Roger, D., Badaeva, E., Gay, G., Bernard, M., Sourdille, P., Feuillet, C, Charles, M., Belcram, H., Just, J., Huneau, C., Viollet, A., Couloux, A., Segurens, B., Carter, M., Huteau, V., Coriton, O., Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G., D'Amore, R., Allen, A., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Ling, H., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., Fan, H., Li, D., Dong, L., Tao, Y., Gao, C., Wu, H., Li, Y., Cui, Y., Guo, X., Zheng, S., Wang, B., Yu, K., Liang, Q., Yang, W., Lou, X., Chen, J., Feng, M., Jian, J., Zhang, X., Luo, G., Jiang, Y., Liu, J., Wang, Z., Sha, Y., Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., Appels, R., Pfeifer, M., Tao, Y., Zhang, X., Jing, R., Zhang, C., Ma, Y., Gao, L., Gao, C., Spannagl, M., Mayer, KF, Li, D., Pan, S., Zheng, F., Hu, Q., Xia, X., Li, J., Liang, Q., Chen, J., Wicker, T., Gou, C., Kuang, H., He, G., Luo, Y., Choulet, F, Alberti, A, Theil, S, Glover, N, Barbe, V, Daron, J, Pingault, L, Sourdille, P., Couloux, A., Paux, E, Leroy, P., Mangenot, S, Guilhot, N., Gouis, J. Le, Balfourier, F., Alaux, M., Jamilloux, V, Poulain, J., Durand, C., Bellec, A., Gaspin, C., Safar, J., Dolezel, J., Rogers, J., Vandepoele, K., Aury, J., Mayer, K, Berges, H., Quesneville, H, Wincker, P, Baidouri, M. El, Panaud, O., Flutre, T., Duprat, E., Feuillet, C, Quesneville, H, Vogel, J., Garvin, D., Mockler, T., Schmutz, J., Rokhsar, D., Bevan, M., Barry, K., Lucas, S., Harmon-Smith, M., Lail, K., Baucom, R., Estill, J., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J.-M., Westerman, R., SanMiguel, PJ, Bennetzen, J., Rustenholz, C., Choulet, F, Laugier, C., Šafář, J., Šimková, H., Doležel, J., Magni, F., Scalabrin, S., Cattonaro, F., Vautrin, S., Wicker, T., Guyot, R., Yahiaoui, N., Keller, B., Liu, R., Vitte, C., Ma, J., Mahama, A., Dhliwayo, T., Lee, M., Bennetzen, J., Rees, H., Walters, M., Kashkush, K., Feldman, M., Levy, A., Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., Ainouche, M., Chalhoub, B., Grandbastien, M., Wright, S., Agrawal, N., Bureau, T., Yang, G., Lee, Y., Jiang, Y., Shi, X., Kertbundit, S., Hall, T., Han, Y., Qin, S., Wessler, S., Kuang, H., Padmanabhan, C., Li, F., Kamei, A., Bhaskar, P., Ouyang, S., Jiang, J., Buell, C., Baker, B., Consortium, I., Batzer, M., Deininger, P., Hawkins, J., Kim, H., Nason, J., Wing, R., Wendel, J., Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D., Jackson, S, Wing, R., Vitte, C., Panaud, O., Hawkins, J., Proulx, S., Rapp, R., Wendel, J., Zabala, G., Vodkin, L., Takahashi, S., Inagaki, Y., Satoh, H., Hoshino, A., Iida, S., Wicker, T., Buchmann, J., Keller, B., Jiang, N., Bao, Z., Zhang, X., Eddy, S., Wessler, S., Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., Enright, A., Dongen, S. Van, Ouzounis, C., Katoh, K., Kuma, K., Toh, H., Miyata, T., Waterhouse, A., Procter, J., Martin, D., Clamp, M., Barton, G., Stajich, J., Block, D., Boulez, K., Brenner, S., Chervitz, S., Dagdigian, C., Fuellen, G., Gilbert, J., Korf, I., Lapp, H., Edgar, R., SanMiguel, P, Gaut, B., Tikhonov, A., Nakajima, Y., Bennetzen, J., Price, M., Dehal, P., Arkin, A., Suzuki, R. & Shimodaira, H. 2014, 'Organization and evolution of transposable elements along the bread wheat chromosome 3B', *Genome Biology*, vol. 15, no. 12, p. 546.

Darwin, C. 1876, *The effects of self and cross fertilization in the vegetable kingdom*,

John Murray, London.

Daryanto, S., Wang, L. & Jacinthe, P.-A. 2016, 'Global Synthesis of Drought Effects on Maize and Wheat Production.', *PloS one*, vol. 11, no. 5, p. e0156362.

Daryanto, S., Wang, L. & Jacinthe, P.A. 2017, 'Global synthesis of drought effects on cereal, legume, tuber and root crops production: A review', *Agricultural Water Management*.

Devos, K.M., Brown, J.K.M. & Bennetzen, J.L. 2002, 'Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis.', *Genome research*, vol. 12, no. 7, pp. 1075–9.

Dietrich, C.R., Cui, F., Packila, M.L., Li, J., Ashlock, D.A., Nikolau, B.J. & Schnable, P.S. 2002, 'Maize Mu transposons are targeted to the 5' untranslated region of the gl8 gene and sequences flanking Mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome.', *Genetics*, vol. 160, no. 2, pp. 697–716.

Diez, C.M., Meca, E., Tenaillon, M.I. & Gaut, B.S. 2014, 'Three groups of transposable elements with contrasting copy number dynamics and host responses in the maize (Zea mays ssp. mays) genome', *PLoS Genet*, vol. 10, no. 4, p. e1004298.

Diez, C.M., Roessler, K. & Gaut, B.S. 2014, 'Epigenetics and plant genome evolution', *Current Opinion in Plant Biology*, vol. 18, no. 1, pp. 1–8.

Ding, H., Zhang, Z.M., Qin, F.F., Dai, L.X., Li, C.J., Ci, D.W. & Song, W.W. 2014, 'Isolation and characterization of drought-responsive genes from peanut roots by suppression subtractive hybridization', *Electronic Journal of Biotechnology*, vol. 17, no. 6, pp. 304–10.

Ding, Y., Cao, J., Ni, L., Zhu, Y., Zhang, A., Tan, M. & Jiang, M. 2013, 'ZmCPK11 is involved in abscisic acid-induced antioxidant defence and functions upstream of ZmMPK5 in abscisic acid signalling in maize.', *Journal of experimental botany*, vol. 64, no. 4, pp. 871–84.

Van Dooren, T., Silveira, A., Gilbaut, E., Jimenez-Gomez, J.M., Martin, A., Bach, L., Tisne, S., Quadrana, L., Loudet, O. & Colot, V. 2018, 'Mild drought induces phenotypic and DNA methylation plasticity but no transgenerational effects in Arabidopsis', *bioRxiv*.

Dowen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Dowen, J.M., Nery, J.R., Dixon, J.E. & Ecker, J.R. 2012, 'Widespread dynamic DNA methylation in response to biotic stress.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 32, pp. E2183-91.

Duvick, D.N. 2001, 'Biotechnology in the 1930s: The development of hybrid maize', *Nature Reviews Genetics*.

Eichten, S.R. & Springer, N.M. 2015, 'Minimal evidence for consistent changes in maize DNA methylation patterns following environmental stress', *Frontiers in Plant Science*, vol. 6, p. 308.

Elhamamsy, A.R. 2016, 'DNA methylation dynamics in plants and mammals: overview of regulation and dysregulation', *Cell Biochemistry and Function*, vol. 34, no. 5, pp. 289–98.

Elrouby, N. & Bureau, T.E. 2010, 'Bs1, a new chimeric gene formed by retrotransposon-mediated exon shuffling in maize.', *Plant physiology*, vol. 153, no. 3, pp. 1413–24.

Ely, A., Geall, S. & Song, Y. 2016, 'Sustainable maize production and consumption in

China: practices and politics in transition', *Journal of Cleaner Production*, vol. 134, pp. 259–68.

Estep, M.C., DeBarry, J.D. & Bennetzen, J.L. 2013, 'The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution', *Heredity*, vol. 110, no. 2, p. 194.

Faghani, E., Gharechahi, J., Komatsu, S., Mirzaei, M., Ali Khavarinejad, R., Najafi, F., Karimi Farsad, L. & Hosseini Salekdeh, G. 2015, 'Data in support of comparative physiology and proteomic analysis of two wheat genotypes contrasting in drought tolerance', *Data in Brief*, vol. 2, pp. 26–8.

Fang, Y. & Xiong, L. 2015, 'General mechanisms of drought response and their application in drought resistance improvement in plants', *Cellular and Molecular Life Sciences*, vol. 72, no. 4, pp. 673–89.

FAO 2019, *FAOSTAT database collections*, *Food and Agriculture Organization of the United Nations*, Rome.

Farquharson, K.L. 2013, 'AUGMIN Subunit8 Participates in Microtubule Reorientation in Arabidopsis Hypocotyls', *The Plant Cell*, vol. 25, no. 6, pp. 1913–1913.

Fattash, I., Rooke, R., Wong, A., Hui, C., Luu, T., Bhardwaj, P., Yang, G. & Bainard, J. 2013, 'Miniature inverted-repeat transposable elements: discovery, distribution, and activity [1]', *Genome*, vol. 56, no. 9, pp. 475–86.

Fei, Y., Xue, Y., Du, P., Yang, S. & Deng, X. 2017, 'Expression analysis and promoter methylation under osmotic and salinity stress of TaGAPC1 in wheat (Triticum aestivum L)', *Protoplasma*, vol. 254, no. 2, pp. 987–96.

Feng, G., Leem, Y.-E. & Levin, H.L. 2013, 'Transposon integration enhances expression of stress response genes.', *Nucleic acids research*, vol. 41, no. 2, pp. 775–89.

Ferrandino, A. & Lovisolo, C. 2014, 'Abiotic stress effects on grapevine (Vitis vinifera L.): Focus on abscisic acid-mediated consequences on secondary metabolism and berry quality', *Environmental and Experimental Botany*, vol. 103, pp. 138–47.

Forestan, C., Cigliano, R.A., Farinati, S., Lunardon, A., Sanseverino, W. & Varotto, S. 2016, 'Stress-induced and epigenetic- mediated maize transcriptome regulation study by means of transcriptome reannotation and differential expression analysis', *Nature Publishing Group*.

Fritsch, J.M., Kane, R.J. & Chelius, C.R. 1986, 'The Contribution of Mesoscale Convective Weather Systems to the Warm-Season Precipitation in the United States', *Journal of Climate and Applied Meteorology*, vol. 25, no. 10, pp. 1333–45.

Fujita, M., Fujita, Y., Maruyama, K., Seki, M., Hiratsu, K., Ohme-Takagi, M., Tran, L.-S.P., Yamaguchi-Shinozaki, K. & Shinozaki, K. 2004, 'A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway', *The Plant Journal*, vol. 39, no. 6, pp. 863–76.

Ganguly, D., Crisp, P.A., Eichten, S.R. & Pogson, B.J. 2017, 'The Arabidopsis DNA methylome is stable under transgenerational drought stress', *Plant Physiology*, p. pp.00744.2017.

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*.

Gao, D., Jiang, N., Wing, R. a, Jiang, J. & Jackson, S. a 2015, 'Transposons play an important role in the evolution and diversification of centromeres among

closely related species.', *Frontiers in plant science*, vol. 6, no. April, p. 216.

Garg, R., Narayana Chevala, V., 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*, vol. 5, p. 14922.

Gargallo-Garriga, A., Preece, C., Sardans, J., Oravec, M., Urban, O. & Peñuelas, J. 2018, 'Root exudate metabolomes change under drought and show limited capacity for recovery', *Scientific Reports*, vol. 8, no. 1, p. 12696.

Gayacharan & Joel, A.J. 2013, 'Epigenetic responses to drought stress in rice (Oryza sativa L.).', *Physiology and molecular biology of plants : an international journal of functional plant biology*, vol. 19, no. 3, pp. 379–87.

Ge, T. da, Sui, F. gong, Bai, L. ping, Lu, Y. yan, Zhou, G. sheng & Zhang, Y. min 2006, 'Effects of Water Stress on the Protective Enzyme Activities and Lipid Peroxidation in Roots and Leaves of Summer Maize', *Agricultural Sciences in China*, vol. 5, no. 4, pp. 291–8.

Gent, J.I., Ellis, N.A., Guo, L., Harkess, A.E., Yao, Y., Zhang, X. & Dawe, R.K. 2013, 'CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize.', *Genome research*, vol. 23, no. 4, pp. 628–37.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M. & Toulmin, C. 2010, 'Food security: the challenge of feeding 9 billion people.', *Science (New York, N.Y.)*, vol. 327, no. 5967, pp. 812–8.

Gómez, J., Sánchez-Martínez, D., Stiefel, V., Rigau, J., Puigdomènech, P. & Pagès, M. 1988, 'A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein', *Nature*, vol. 334, no. 6179, pp. 262–4.

González, R.M., Ricardi, M.M. & Iusem, N.D. 2013, 'Epigenetic marks in an adaptive water stress-responsive gene in tomato roots under normal and drought conditions.', *Epigenetics*, vol. 8, no. 8, pp. 864–72.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. 2008, 'High-throughput functional annotation and data mining with the Blast2GO suite', *Nucleic Acids Research*.

Govind, G., Harshavardhan, V.T., ThammeGowda, H.V., Patricia, J.K., Kalaiarasi, P.J., Dhanalakshmi, R., Iyer, D.R., Senthil Kumar, M., Muthappa, S.K., Sreenivasulu, N., Nese, S., Udayakumar, M. & Makarla, U.K. 2009, 'Identification and functional validation of a unique set of drought induced genes preferentially expressed in response to gradual water stress in peanut.', *Molecular genetics and genomics : MGG*, vol. 281, no. 6, pp. 591–605.

Greene, B., Walko, R. & Hake, S. 1994, 'Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations.', *Genetics*, vol. 138, no. 4, pp. 1275–85.

Gudys, K., Guzy-Wrobelska, J., Janiak, A., Dziurka, M.A., Ostrowska, A., Hura, K., Jurczyk, B., Żmuda, K., Grzybkowska, D., Śróbka, J., Urban, W., Biesaga-Koscielniak, J., Filek, M., Koscielniak, J., Mikołajczak, K., Ogrodowicz, P., Krystkowiak, K., Kuczyńska, A., Krajewski, P. & Szarejko, I. 2018, 'Prioritization of Candidate Genes in QTL Regions for Physiological and Biochemical Traits Underlying Drought Response in Barley (Hordeum vulgare L.)', *Frontiers in Plant Science*, vol. 9, p. 769.

Gupta, A., Hisano, H., Hojo, Y., Matsuura, T., Ikeda, Y., Mori, I.C. & Senthil-Kumar, M. 2017, 'Global profiling of phytohormone dynamics during combined drought and pathogen stress in Arabidopsis thaliana reveals ABA and JA as major regulators.', *Scientific reports*, vol. 7, no. 1, p. 4017.

Haider, M.S., Zhang, Cheng, Kurjogi, M.M., Pervaiz, T., Zheng, T., Zhang, Chaobo, Lide, C., Shangguan, L. & Fang, J. 2017, 'Insights into grapevine defense response against drought as revealed by biochemical, physiological and RNA-Seq analysis', *Scientific Reports*, vol. 7, no. 1, p. 13134.

Han, Z., Crisp, P., Stelpflug, S., Kaeppler, S., Li, Q. & Springer, N. 2018, 'Targeted epigenomic changes to the maize methylome resulting from tissue culture', *bioRxiv*, p. 242081.

Harb, A., Krishnan, A., Ambavaram, M.M.R. & Pereira, A. 2010, 'Molecular and physiological analysis of drought stress in Arabidopsis reveals early responses leading to acclimation in plant growth.', *Plant physiology*, vol. 154, no. 3, pp. 1254–71.

Hashimoto, M., Neriya, Y., Yamaji, Y. & Namba, S. 2016, 'Recessive Resistance to Plant Viruses: Potential Resistance Genes Beyond Translation Initiation Factors', *Frontiers in Microbiology*, vol. 7.

Hayat, S., Hayat, Q., Alyemeni, M.N., Wani, A.S., Pichtel, J. & Ahmad, A. 2012, 'Role of proline under changing environments: a review.', *Plant signaling & behavior*, vol. 7, no. 11, pp. 1456–66.

He, X.J., Chen, T. & Zhu, J.K. 2011, 'Regulation and function of DNA methylation in plants and animals', *Cell Research*.

Henderson, I.R. & Jacobsen, S.E. 2007, 'Epigenetic inheritance in plants', *Nature*, vol. 447, no. 7143, pp. 418–24.

Hershko, Avram and Ciechanover, A. 1998, 'The ubiquitin system', *Trends in Biochemical Sciences*.

Hirsch, C.D. & Springer, N.M. 2017, 'Transposable element influences on gene expression in plants', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1860, no. 1, pp. 157–65.

Hochberg, B. 1995, 'Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society*.

Hofmeister, B.T., Lee, K., Rohr, N.A., Hall, D.W. & Schmitz, R.J. 2017, 'Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation', *Genome Biology*, vol. 18, no. 1, p. 155.

Hoornweg, D. & Pope, K. 2017, 'Population predictions for the world's largest cities in the 21st century', *Environment and Urbanization*.

Horváth, V., Merenciano, M. & González, J. 2017, 'Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response', *Trends in Genetics*, vol. 33, no. 11, pp. 832–41.

Hossain, M.S., Kawakatsu, T., Kim, K. Do, Zhang, N., Nguyen, C.T., Khan, S.M., Batek, J.M., Joshi, T., Schmutz, J., Grimwood, J., Schmitz, R.J., Xu, D., Jackson, S.A., Ecker, J.R. & Stacey, G. 2017, 'Divergent cytosine DNA methylation patterns in single-cell, soybean root hairs', *New Phytologist*, vol. 214, no. 2, pp. 808–19.

Hsu, F.-M., Yen, M.-R., Wang, C.-T., Lin, C.-Y., Wang, C.-J.R. & Chen, P.-Y. 2017, 'Optimized reduced representation bisulfite sequencing reveals tissue-specific

mCHH islands in maize', *Epigenetics & Chromatin*, vol. 10, no. 1, p. 42.

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q. & Xiong, L. 2006, 'Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 35, pp. 12987–92.

Hu, X., Wu, X., Li, C., Lu, M., Liu, T., Wang, Y. & Wang, W. 2012, 'Abscisic acid refines the synthesis of chloroplast proteins in maize (Zea mays) in response to drought and light.', *PloS one*, vol. 7, no. 11, p. e49500.

Huang, C., Sun, H., Xu, D., Chen, Q., Liang, Y., Wang, X., Xu, G., Tian, J., Wang, C., Li, D., Wu, L., Yang, X., Jin, W., Doebley, J.F. & Tian, F. 2018, '*ZmCCT9* enhances maize adaptation to higher latitudes', *Proceedings of the National Academy of Sciences*, vol. 115, no. 2, pp. E334–41.

Ion, V., Dicu, G., State, D., Fintineru, G., Iuliana Epure, L., Basa, A. & Toader, M. 2013, 'Yield Components of different hybrids of maize (Zea mays L.) cultivated in South Romania under drought conditions', *Scientific Papers. Series A. Agronomy*, vol. 56, pp. 276–83.

IPCC 2014, *Climate Change 2014*, *Climate Change 2014: Synthesis Report*.

Ito, H. 2012, 'Small RNAs and transposon silencing in plants', *Development, Growth & Differentiation*, vol. 54, no. 1, pp. 100–7.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. & Paszkowski, J. 2011, 'An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress', *Nature*, vol. 472, no. 7341, pp. 115–20.

Ito, H. & Kakutani, T. 2014, 'Control of transposable elements in Arabidopsis thaliana', *Chromosome Research*, vol. 22, no. 2, pp. 217–23.

Ito, H., Kim, J.-M., Matsunaga, W., Saze, H., Matsui, A., Endo, T.A., Harukawa, Y., Takagi, H., Yaegashi, H., Masuta, Y., Masuda, S., Ishida, J., Tanaka, M., Takahashi, S., Morosawa, T., Toyoda, T., Kakutani, T., Kato, A. & Seki, M. 2016, 'A Stress-Activated Transposon in Arabidopsis Induces Transgenerational Abscisic Acid Insensitivity', *Scientific Reports*, vol. 6, no. 1, p. 23181.

Jaillais, Y., Hothorn, M., Belkhadir, Y., Dabi, T., Nimchuk, Z.L., Meyerowitz, E.M. & Chory, J. 2011, 'Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor.', *Genes & development*, vol. 25, no. 3, pp. 232–7.

Jain, M., Kataria, S., Hirve, M. & Prajapati, R. 2019, 'Water Deficit Stress Effects and Responses in Maize', *Plant Abiotic Stress Tolerance*, Springer International Publishing, Cham, pp. 129–51.

Jiang, M. & Zhang, J. 2002, 'Water stress-induced abscisic acid accumulation triggers the increased generation of reactive oxygen species and up-regulates the activities of antioxidant enzymes in maize leaves', *Journal of Experimental Botany*, vol. 53, no. 379, pp. 2401–10.

Jin, W.-W., Wang, Y., Zhang, H.-H., Jiao, Z.-L., Wang, P., Li, X., Yue, B.-B. & Sun, G.-Y. 2011, 'Effects of foliar spraying methyl jasmonate on leaf chlorophyll fluorescence characteristics of flue-cured tobacco seedlings under drought and re-watering.', *Ying yong sheng tai xue bao = The journal of applied ecology*, vol. 22, no. 12, pp. 3157–62.

Kakumanu, A., Ambavaram, M.M.R., Klumas, C., Krishnan, A., Batlang, U., Myers, E., Grene, R. & Pereira, A. 2012, 'Effects of drought on gene expression in maize

reproductive and leaf meristem tissue revealed by RNA-Seq.', *Plant physiology*, vol. 160, no. 2, pp. 846–67.

Kang, J., Li, J., Gao, S., Tian, C. & Zha, X. 2017, 'Overexpression of the leucine-rich receptor-like kinase gene *LRK2* increases drought tolerance and tiller number in rice', *Plant Biotechnology Journal*, vol. 15, no. 9, pp. 1175–85.

Kang, Y.H. & Hardtke, C.S. 2016, 'Arabidopsis MAKR5 is a positive effector of BAM3-dependent CLE45 signaling.', *EMBO reports*, vol. 17, no. 8, pp. 1145–54.

Kapitonov, V. V. & Jurka, J. 2001, 'Rolling-circle transposons in eukaryotes', *Proceedings of the National Academy of Sciences*, vol. 98, no. 15, pp. 8714–9.

Kapitonov, V. V. & Jurka, J. 2007, 'Helitrons on a roll: eukaryotic rolling-circle transposons', *Trends in Genetics*, vol. 23, no. 10, pp. 521–9.

Kato, M., Takashima, K. & Kakutani, T. 2004, 'Epigenetic control of CACTA transposon mobility in Arabidopsis thaliana', *Genetics*, vol. 168, no. 2, pp. 961–9.

Kaur, A., Grewal, A. & Sharma, P. 2018, 'Comparative analysis of DNA methylation changes in two contrasting wheat genotypes under water deficit', *Biologia Plantarum*, vol. 62, no. 3, pp. 471–8.

Kavi Kishor, P.B. 2015, 'Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny', *Frontiers in Plant Science*, vol. 6.

Kent, T. V., Uzunović, J. & Wright, S.I. 2017, 'Coevolution between transposable elements and recombination', *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1736, p. 20160458.

Kieliszewski, M.J., Lamport, D.T.A., Tan, L. & Cannon, M.C. 2010, 'Hydroxyproline-Rich Glycoproteins: Form and Function', *Annual Plant Reviews*, vol. 41, pp. 321–42.

Kim, J.-M., To, T.K., Matsui, A., Tanoi, K., Kobayashi, N.I., Matsuda, F., Habu, Y., Ogawa, D., Sakamoto, T., Matsunaga, S., Bashir, K., Rasheed, S., Ando, M., Takeda, H., Kawaura, K., Kusano, M., Fukushima, A., Endo, T.A., Kuromori, T., Ishida, J., Morosawa, T., Tanaka, M., Torii, C., Takebayashi, Y., Sakakibara, H., Ogihara, Y., Saito, K., Shinozaki, K., Devoto, A. & Seki, M. 2017, 'Acetate-mediated novel survival strategy against drought in plants', *Nature Plants 2017 3:7*, vol. 3, no. 7, p. 17097.

Kim, Y.-O., Pan, S., Jung, C.-H. & Kang, H. 2007, 'A Zinc Finger-Containing Glycine-Rich RNA-Binding Protein, atRZ-1a, Has a Negative Impact on Seed Germination and Seedling Growth of Arabidopsis thaliana Under Salt or Drought Stress Conditions', *Plant and Cell Physiology*, vol. 48, no. 8, pp. 1170–81.

Kimura, Y., Tosa, Y., Shimada, S., Sogo, R., Kusaba, M., Sunaga, T., Betsuyaku, S., Eto, Y., Nakayashiki, H. & Mayama, S. 2001, 'OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses', *Plant and Cell Physiology*, vol. 42, no. 12, pp. 1345–54.

Kistler, L., Yoshi Maezumi, S., De Souza, J.G., Przelomska, N.A.S., Costa, F.M., Smith, O., Loiselle, H., Ramos-Madrigal, J., Wales, N., Ribeiro, E.R., Morrison, R.R., Grimaldo, C., Prous, A.P., Arriaza, B., Thomas P Gilbert, M., De Oliveira Freitas, F. & Allaby, R.G. 2018, 'Multiproxy evidence highlights a complex evolutionary legacy of maize in South America', *Science*.

Konate, M., Wilkinson, M., Mayne, B., Pederson, S., Scott, E., Berger, B. & Rodriguez Lopez, C. 2018, 'Salt Stress Induces Non-CG Methylation in Coding Regions of Barley Seedlings (Hordeum vulgare)', *Epigenomes*.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A. & Pollock, D.D. 2011, 'Repetitive elements may comprise over two-thirds of the human genome.', *PLoS genetics*, vol. 7, no. 12, p. e1002384.

Krueger, F. & Andrews, S.R. 2011, 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.', *Bioinformatics (Oxford, England)*, vol. 27, no. 11, pp. 1571–2.

Kumar, N., Astegno, A., Chen, J., Giorgetti, A. & Dominici, P. 2016, 'Residues in the Distal Heme Pocket of Arabidopsis Non-Symbiotic Hemoglobins: Implication for Nitrite Reductase Activity.', *International journal of molecular sciences*, vol. 17, no. 5.

Kwak, K.J., Kim, Y.O. & Kang, H. 2005, 'Characterization of transgenic Arabidopsis plants overexpressing GR-RBP4 under high salinity, dehydration, or cold stress', *Journal of Experimental Botany*, vol. 56, no. 421, pp. 3007–16.

de la Chaux, N., Tsuchimatsu, T., Shimizu, K.K. & Wagner, A. 2012, 'The predominantly selfing plant Arabidopsis thaliana experienced a recent reduction in transposable element abundance compared to its outcrossing relative Arabidopsis lyrata.', *Mobile DNA*, vol. 3, no. 1, p. 2.

Lai, J., Li, Y., Messing, J. & Dooner, H.K. 2005, 'Gene movement by Helitron transposons contributes to the haplotype variability of maize', *Proceedings of the National Academy of Sciences*.

Lai, X., Schnable, J.C., Liao, Z., Xu, J., Zhang, G., Li, C., Hu, E., Rong, T., Xu, Y. & Lu, Y. 2017, 'Genome-wide characterization of non-reference transposable element insertion polymorphisms reveals genetic diversity in tropical and temperate maize', *BMC Genomics*, vol. 18, no. 1, p. 702.

Lamaoui, M., Jemo, M., Datla, R. & Bekkaoui, F. 2018, 'Heat and Drought Stresses in Crops and Approaches for Their Mitigation.', *Frontiers in chemistry*, vol. 6, p. 26.

Lauria, M., Echegoyen-Nava, R.A., Rodríguez-Ríos, D., Zaina, S. & Lund, G. 2017, 'Inter-individual variation in DNA methylation is largely restricted to tissue-specific differentially methylated regions in maize', *BMC Plant Biology*, vol. 17, no. 1, p. 52.

Law, J.A. & Jacobsen, S.E. 2010, 'Establishing, maintaining and modifying DNA methylation patterns in plants and animals', *Nature Reviews Genetics*, vol. 11, no. 3, pp. 204–20.

Laxa, M., Liebthal, M., Telman, W., Chibani, K. & Dietz, K.-J. 2019, 'The Role of the Plant Antioxidant System in Drought Tolerance.', *Antioxidants (Basel, Switzerland)*, vol. 8, no. 4.

Le, T.-N., Schumann, U., Smith, N.A., Tiwari, S., Au, P.C.K., Zhu, Q.-H., Taylor, J.M., Kazan, K., Llewellyn, D.J., Zhang, R., Dennis, E.S. & Wang, M.-B. 2014, 'DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis', *Genome Biology*, vol. 15, no. 9, p. 458.

Lei, M., Zhang, H., Julian, R., Tang, K., Xie, S. & Zhu, J.-K. 2015, ' Regulatory link between DNA methylation and active demethylation in Arabidopsis ', *Proceedings of the National Academy of Sciences*.

Leng, G. & Hall, J. 2019, 'Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future', *The Science of the Total Environment*, vol. 654, p. 811.

Li, J., Zhang, M., Sun, J., Mao, X., Wang, Jing, Wang, Jingguo, Liu, H., Zheng, H., Zhen,

Z., Zhao, H., Zou, D., Li, J., Zhang, M., Sun, J., Mao, X., Wang, Jing, Wang, Jingguo, Liu, H., Zheng, H., Zhen, Z., Zhao, H. & Zou, D. 2019, 'Genome-Wide Characterization and Identification of Trihelix Transcription Factor and Expression Profiling in Response to Abiotic Stresses in Rice (Oryza sativa L.)', *International Journal of Molecular Sciences*, vol. 20, no. 2, p. 251.

Li, P., Yang, H., Wang, L., Liu, H., Huo, H., Zhang, C., Liu, A., Zhu, A., Hu, J., Lin, Y. & Liu, L. 2019, 'Physiological and Transcriptome Analyses Reveal Short-Term Responses and Formation of Memory Under Drought Stress in Rice', *Frontiers in Genetics*, vol. 10, p. 55.

Li, Q., Eichten, S.R., Hermanson, P.J., Zaunbrecher, V.M., Song, J., Wendt, J., Rosenbaum, H., Madzima, T.F., Sloan, A.E., Huang, J., Burgess, D.L., Richmond, T.A., McGinnis, K.M., Meeley, R.B., Danilevskaya, O.N., Vaughn, M.W., Kaeppler, S.M., Jeddeloh, J.A. & Springer, N.M. 2014, 'Genetic Perturbation of the Maize Methylome', *The Plant Cell*, vol. 26, no. 12, pp. 4602–16.

Li, Q., Gent, J.I., Zynda, G., Song, J., Makarevitch, I., Hirsch, C.D., Hirsch, C.N., Dawe, R.K., Madzima, T.F., McGinnis, K.M., Lisch, D., Schmitz, R.J., Vaughn, M.W. & Springer, N.M. 2015, 'RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 47, pp. 14728–33.

Li, X., Wang, X., He, K., Ma, Y., Su, N., He, H., Stolc, V., Tongprasit, W., Jin, W., Jiang, J., Terzaghi, W., Li, S. & Deng, X.W. 2008, 'High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression.', *The Plant cell*, vol. 20, no. 2, pp. 259–76.

Li, X., Yang, Yunqiang, Yang, S., Sun, X., Yin, X., Zhao, Y. & Yang, Yongping 2016, 'Comparative proteomics analyses of intraspecific differences in the response of Stipa purpurea to drought.', *Plant diversity*, vol. 38, no. 2, pp. 101–17.

Li, X., Zhu, J., Hu, F., Ge, S., Ye, M., Xiang, H., Zhang, G., Zheng, X., Zhang, H., Zhang, S., Li, Q., Luo, R., Yu, C., Yu, J., Sun, J., Zou, X., Cao, X., Xie, X., Wang, J. & Wang, W. 2012, 'Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression', *BMC Genomics*, vol. 13, no. 1, p. 300.

Li, Z.-W., Hou, X.-H., Chen, J.-F., Xu, Y.-C., Wu, Q., González, J. & Guo, Y.-L. 2018, 'Transposable Elements Contribute to the Adaptation of Arabidopsis thaliana.', *Genome biology and evolution*, vol. 10, no. 8, pp. 2140–50.

Liang, D., Zhang, Z., Wu, H., Huang, C., Shuai, P., Ye, C.-Y., Tang, S., Wang, Y., Yang, L., Wang, J., Yin, W. & Xia, X. 2014, 'Single-base-resolution methylomes of Populus trichocarpa reveal the association between DNA methylation and drought stress.', *BMC genetics*, vol. 15 Suppl 1, no. Suppl 1, p. S9.

Lim, C.W., Baek, W., Jung, J., Kim, J.-H. & Lee, S.C. 2015, 'Function of ABA in Stomatal Defense against Biotic and Drought Stresses.', *International journal of molecular sciences*, vol. 16, no. 7, pp. 15251–70.

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S. & Jacobsen, S.E. 2001, 'Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation', *Science*.

Liu, Huanhuan, Ma, Y., Chen, N., Guo, S., Liu, Huili, Guo, X., Chong, K. & Xu, Y. 2014,

'Overexpression of stress-inducible OsBURP16, the β subunit of polygalacturonase 1, decreases pectin content and cell adhesion and increases abiotic stress sensitivity in rice.', *Plant, cell & environment*, vol. 37, no. 5, pp. 1144–58.

Liu, S., Yeh, C.-T., Ji, T., Ying, K., Wu, H., Tang, H.M., Fu, Y., Nettleton, D. & Schnable, P.S. 2009, 'Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome.', H.S. Malik (ed.),*PLoS genetics*, vol. 5, no. 11, p. e1000733.

Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M. & Hammer, G.L. 2014, 'Greater Sensitivity to Drought Accompanies Maize Yield Increase in the U.S. Midwest', *Science*, vol. 344, no. 6183, pp. 516–9.

Long, M., Betrán, E., Thornton, K. & Wang, W. 2003, 'The origin of new genes: glimpses from the young and old', *Nature Reviews Genetics*, vol. 4, no. 11, pp. 865–75.

Lopes, F.R., Jjingo, D., da Silva, C.R.M., Andrade, A.C., Marraccini, P., Teixeira, J.B., Carazzolle, M.F., Pereira, G.A.G., Pereira, L.F.P., Vanzela, A.L.L., Wang, L., Jordan, I.K. & Carareto, C.M.A. 2013, 'Transcriptional activity, chromosomal distribution and expression effects of transposable elements in Coffea genomes.', *PloS one*, vol. 8, no. 11, p. e78931.

Lu, M., Ying, S., Zhang, D.-F., Shi, Y.-S., Song, Y.-C., Wang, T.-Y. & Li, Y. 2012, 'A maize stress-responsive NAC transcription factor, ZmSNAC1, confers enhanced tolerance to dehydration in transgenic Arabidopsis', *Plant Cell Reports*, vol. 31, no. 9, pp. 1701–11.

Lu, Y., Li, Y., Zhang, J., Xiao, Y., Yue, Y., Duan, L., Zhang, M. & Li, Z. 2013, 'Overexpression of Arabidopsis molybdenum cofactor sulfurase gene confers drought tolerance in maize (Zea mays L.).', *PloS one*, vol. 8, no. 1, p. e52126.

Lu, Y., Rong, T. & Cao, M. 2008, 'Analysis of DNA methylation in different maize tissues', *Journal of Genetics and Genomics*, vol. 35, no. 1, pp. 41–8.

Ma, S., Tang, N., Li, Xu, Xie, Y., Xiang, D., Fu, J., Shen, J., Yang, J., Tu, H., Li, Xianghua, Hu, H. & Xiong, L. 2019, 'Reversible Histone H2B Monoubiquitination Fine-Tunes Abscisic Acid Signaling and Drought Response in Rice', *Molecular Plant*, vol. 12, no. 2, pp. 263–77.

Makarevitch, I., Waters, A.J., West, P.T., Stitzer, M., Hirsch, C.N., Ross-Ibarra, J. & Springer, N.M. 2015, 'Transposable elements contribute to activation of maize genes in response to abiotic stress', *PLoS Genet*, vol. 11, no. 1, p. e1004915.

Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Tran, L.S. & Qin, F. 2015, 'A transposable element in a NAC gene is associated with drought tolerance in maize seedlings', *Nat Commun*, vol. 6, p. 8326.

Marçais, G. & Kingsford, C. 2011, 'A fast, lock-free approach for efficient parallel counting of occurrences of k-mers', *Bioinformatics*, vol. 27, no. 6, pp. 764–70.

March, E. & Farrona, S. 2018, 'Plant Deubiquitinases and Their Role in the Control of Gene Expression Through Modification of Histones', *Frontiers in Plant Science*, vol. 8, p. 2274.

Marraccini, P., Vinecky, F., Alves, G.S.C., Ramos, H.J.O., Elbelt, S., Vieira, N.G., Carneiro, F.A., Sujii, P.S., Alekcevetch, J.C., Silva, V.A., DaMatta, F.M., Ferrão, M.A.G., Leroy, T., Pot, D., Vieira, L.G.E., Silva, F.R. da & Andrade, A.C. 2012, 'Differentially expressed genes and proteins upon drought acclimation in

tolerant and sensitive genotypes of Coffea canephora', *Journal of Experimental Botany*, vol. 63, no. 11, p. 4191.

McClintock, B. 1950, 'The origin and behavior of mutable loci in maize.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 36, no. 6, pp. 344–55.

McClintock, B. 1984, 'The significance of responses of the genome to challenge', *Science*, American Association for the Advancement of Science, pp. 792–801.

Meng, Z., Duan, A., Chen, D., Dassanayake, K.B., Wang, X., Liu, Z., Liu, H. & Gao, S. 2017, 'Suitable indicators using stem diameter variation-derived indices to monitor the water status of greenhouse tomato plants', P. Pardha-Saradhi (ed.),*PLOS ONE*, vol. 12, no. 2, p. e0171423.

Meyers, B.C., Tingey, S. V. & Morgante, M. 2001, 'Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome', *Genome Research*, vol. 11, no. 10, pp. 1660–76.

Minh-Thu, P.T., Kim, J.S., Chae, S., Jun, K.M., Lee, G.S., Kim, D.E., Cheong, J.J., Song, S.I., Nahm, B.H. & Kim, Y.K. 2018, 'A WUSCHEL Homeobox Transcription Factor, OsWOX13, Enhances Drought Tolerance and Triggers Early Flowering in Rice', *Mol Cells*, vol. 41, no. 8, pp. 781–98.

Miousse, I.R., Chalbot, M.-C.G., Lumen, A., Ferguson, A., Kavouras, I.G. & Koturbash, I. 2015, 'Response of transposable elements to environmental stressors', *Mutation Research/Reviews in Mutation Research*, vol. 765, pp. 19–39.

Montilla-Bascón, G., Rubiales, D., Hebelstrup, K.H., Mandon, J., Harren, F.J.M., Cristescu, S.M., Mur, L.A.J. & Prats, E. 2017, 'Reduced nitric oxide levels during drought stress promote drought tolerance in barley and is associated with elevated polyamine biosynthesis', *Scientific Reports*, vol. 7, no. 1, p. 13311.

Muñoz-López, M. & García-Pérez, J.L. 2010, 'DNA transposons: nature and applications in genomics.', *Current genomics*, vol. 11, no. 2, pp. 115–28.

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*, vol. 1819, no. 2, pp. 97–103.

Naumann, G., Alfieri, L., Wyser, K., Mentaschi, L., Betts, R.A., Carrao, H., Spinoni, J., Vogt, J. & Feyen, L. 2018, 'Global Changes in Drought Conditions Under Different Levels of Warming', *Geophysical Research Letters*, vol. 45, no. 7, pp. 3285–96.

Naz, H., Akram, N.A. & Ashraf, M. 2016, *Impact of ascorbic acid on growth and some physiological attributes of cucumber (Cucumis Sativus) plants under water-deficit conditions*, *Pakistan Journal of Botany*.

Negi, P., Rai, A.N. & Suprasanna, P. 2016, 'Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response', *Frontiers in Plant Science*, vol. 7, p. 1448.

Neumann, P.M. 2008, 'Coping Mechanisms for Crop Plants in Drought-prone Environments', *Annals of Botany*, vol. 101, no. 7, pp. 901–7.

Noman, A., Ali, S., Naheed, F., Ali, Q., Farid, M., Rizwan, M. & Irshad, M.K. 2015, 'Foliar application of ascorbate enhances the physiological and biochemical attributes of maize (*Zea mays* L.) cultivars under drought stress', *Archives of Agronomy and Soil Science*, vol. 61, no. 12, pp. 1659–72.

Nonomura, K.-I. 2004, 'The Novel Gene HOMOLOGOUS PAIRING ABERRATION IN

RICE MEIOSIS1 of Rice Encodes a Putative Coiled-Coil Protein Required for Homologous Chromosome Pairing in Meiosis', *the Plant Cell Online*, vol. 16, no. 4, pp. 1008–20.

Nuccio, M.L., Wu, J., Mowers, R., Zhou, H.-P., Meghji, M., Primavesi, L.F., Paul, M.J., Chen, X., Gao, Y., Haque, E., Basu, S.S. & Lagrimini, L.M. 2015, 'Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions', *Nature Biotechnology*, vol. 33, no. 8, pp. 862–9.

Obata, T., Witt, S., Lisec, J., Palacios-Rojas, N., Florez-Sarasa, I., Yousfi, S., Araus, J.L., Cairns, J.E. & Fernie, A.R. 2015, 'Metabolite Profiles of Maize Leaves in Drought, Heat, and Combined Stress Field Trials Reveal the Relationship between Metabolism and Grain Yield.', *Plant physiology*, vol. 169, no. 4, pp. 2665–83.

Ohno, S. 1972, 'So much "junk" DNA in our genome.', *Brookhaven symposia in biology*, vol. 23, pp. 366–70.

Okamoto, H. & Hirochika, H. 2001, 'Silencing of transposable elements in plants', *Trends in Plant Science*, vol. 6, no. 11, pp. 527–34.

Oliver, K.R. & Greene, W.K. 2009, 'Transposable elements: powerful facilitators of evolution', *Bioessays*, vol. 31, no. 7, pp. 703–14.

de Ollas, C., Arbona, V. & Gómez-Cadenas, A. 2015, 'Jasmonic acid interacts with abscisic acid to regulate plant responses to water stress conditions', *Plant Signaling and Behavior*.

Olson, M.E., Soriano, D., Rosell, J.A., Anfodillo, T., Donoghue, M.J., Edwards, E.J., León-Gómez, C., Dawson, T., Camarero Martínez, J.J., Castorena, M., Echeverría, A., Espinosa, C.I., Fajardo, A., Gazol, A., Isnard, S., Lima, R.S., Marcati, C.R. & Méndez-Alonzo, R. 2018, 'Plant height and hydraulic vulnerability to drought and cold.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 29, pp. 7551–6.

Ong-Abdullah, M., Ordway, J.M., Jiang, N., Ooi, S.-E., Kok, S.-Y., Sarpan, N., Azimi, N., Hashim, A.T., Ishak, Z., Rosli, S.K., Malike, F.A., Bakar, N.A.A., Marjuni, M., Abdullah, N., Yaakub, Z., Amiruddin, M.D., Nookiah, R., Singh, R., Low, E.-T.L., Chan, K.-L., Azizi, N., Smith, S.W., Bacher, B., Budiman, M.A., Van Brunt, A., Wischmeyer, C., Beil, M., Hogan, M., Lakey, N., Lim, C.-C., Arulandoo, X., Wong, C.-K., Choo, C.-N., Wong, W.-C., Kwan, Y.-Y., Alwee, S.S.R.S., Sambanthamurthi, R. & Martienssen, R.A. 2015, 'Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm', *Nature*, vol. 525, no. 7570, pp. 533–7.

Orgel, L.E. & Crick, F.H. 1980, 'Selfish DNA: the ultimate parasite.', *Nature*, vol. 284, no. 5757, pp. 604–7.

Ouyang, S.-Q., Liu, Y.-F., Liu, P., Lei, G., He, S.-J., Ma, B., Zhang, W.-K., Zhang, J.-S. & Chen, S.-Y. 2010, 'Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (Oryza sativa) plants', *The Plant Journal*, vol. 62, no. 2, pp. 316–29.

Pacheco, C.A.P., Santos, M.X. dos, Cruz, C.D., Parentoni, S.N., Guimarães, P.E. de O., Gama, E.E.G. e, Silva, Á.E. da, Carvalho, H.W.L. de & Vieira Júnior, P.A. 2002, 'Inbreeding depression of 28 maize elite open pollinated varieties', *Genetics and Molecular Biology*, vol. 25, no. 4, pp. 441–8.

Pan, J., Li, Z., Wang, Q., Garrell, A.K., Liu, M., Guan, Y., Zhou, W. & Liu, W. 2018,

'Comparative proteomic investigation of drought responses in foxtail millet', *BMC Plant Biology*, vol. 18, no. 1, p. 315.

Park, Y.-S., Bae, D.-W. & Ryu, C.-M. 2015, 'Aboveground Whitefly Infestation Modulates Transcriptional Levels of Anthocyanin Biosynthesis and Jasmonic Acid Signaling-Related Genes and Augments the Cope with Drought Stress of Maize.', *PloS one*, vol. 10, no. 12, p. e0143879.

Pignocchi, C., Minns, G.E., Nesi, N., Koumproglou, R., Kitsios, G., Benning, C., Lloyd, C.W., Doonan, J.H. & Hills, M.J. 2009, 'ENDOSPERM DEFECTIVE1 Is a Novel Microtubule-Associated Protein Essential for Seed Development in Arabidopsis', *the Plant Cell Online*, vol. 21, no. 1, pp. 90–105.

Piperno, D.R., Ranere, A.J., Holst, I., Iriarte, J. & Dickau, R. 2009, 'Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 13, pp. 5019–24.

Pumplin, N., Sarazin, A., Jullien, P.E., Bologna, N.G., Oberlin, S. & Voinnet, O. 2016, 'DNA Methylation Influences the Expression of DICER-LIKE4 Isoforms, Which Encode Proteins of Alternative Localization and Function.', *The Plant cell*, vol. 28, no. 11, pp. 2786–804.

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*, vol. 60, no. 9, pp. 805–26.

Qian, S., Wang, Y., Ma, H. & Zhang, L. 2015, 'Expansion and Functional Divergence of Jumonji C-Containing Histone Demethylases: Significance of Duplications in Ancestral Angiosperms and Vertebrates', *Plant Physiology*, vol. 168, no. 4, pp. 1321–37.

Qian, W., Yu, C., Qin, H., Liu, X., Zhang, A., Johansen, I.E. & Wang, D. 2007, 'Molecular and functional analysis of phosphomannomutase (PMM) from higher plants and genetic evidence for the involvement of PMM in ascorbic acid biosynthesis in Arabidopsis and Nicotiana benthamiana', *The Plant Journal*, vol. 49, no. 3, pp. 399–413.

Qian, Y., Cheng, X., Liu, Y., Jiang, H., Zhu, S. & Cheng, B. 2010, 'Reactivation of a silenced minimal Mutator transposable element system following low-energy nitrogen ion implantation in maize', *Plant Cell Reports*, vol. 29, no. 12, pp. 1365–76.

Ranum, P., Peña-Rosas, J.P. & Garcia-Casal, M.N. 2014, 'Global maize production, utilization, and consumption', *Annals of the New York Academy of Sciences*, vol. 1312, no. 1, pp. 105–12.

Rippey, B.R. 2015, 'The U.S. drought of 2012', *Weather and Climate Extremes*, vol. 10, pp. 57–64.

Roessler, K., Bousios, A., Meca, E. & Gaut, B.S. 2018, 'Modeling Interactions between Transposable Elements and the Plant Epigenetic Response: A Surprising Reliance on Element Retention.', *Genome biology and evolution*, vol. 10, no. 3, pp. 803–15.

Roessler, K., Muyle, A., Diez, C.M., Gaut, G.R.J., Bousios, A., Stitzer, M.C., Seymour, D.K., Doebley, J.F., Liu, Q. & Gaut, B.S. 2019, 'The Genomics of Selfing in Maize (Zea mays ssp. mays): Catching Purging in the Act', *bioRxiv*, p. 594812.

Sah, S.K., Reddy, K.R. & Li, J. 2016, 'Abscisic Acid and Abiotic Stress Tolerance in Crop Plants.', *Frontiers in plant science*, vol. 7, p. 571.

Saze, H., Shiraishi, A., Miura, A. & Kakutani, T. 2008, 'Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana', *Science*.

Schläppi, M. 1994, 'Epigenetic regulation of the maize Spm transposable element: Novel activation of a methylated promoter by TnpA', *Cell*, vol. 77, no. 3, pp. 427–37.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.-T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.-P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.-M., Deragon, J.-M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A. & Wilson, R.K. 2009, 'The B73 maize genome: complexity, diversity, and dynamics.', *Science (New York, N.Y.)*, vol. 326, no. 5956, pp. 1112–5.

Secco, D., Wang, C., Shou, H., Schultz, M.D., Chiarenza, S., Nussaume, L., Ecker, J.R., Whelan, J. & Lister, R. 2015, 'Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements', *eLife*.

Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Buell, C.R., de Leon, N. & Kaeppler, S.M. 2011, 'Genome-wide atlas of transcription during maize development', *The Plant Journal*, vol. 66, no. 4, pp. 553–63.

Seo, J.-S., Joo, J., Kim, M.-J., Kim, Y.-K., Nahm, B.H., Song, S.I., Cheong, J.-J., Lee, J.S., Kim, J.-K. & Choi, Y. Do 2011, 'OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice', *The Plant Journal*, vol. 65, no. 6, pp. 907–21.

Sharma, P.D., Singh, N., Ahuja, P.S. & Reddy, T.V. 2011, 'Abscisic acid response element binding factor 1 is required for establishment of Arabidopsis seedlings during winter', *Molecular Biology Reports*, vol. 38, no. 8, pp. 5147–59.

Sigman, M.J. & Slotkin, R.K. 2016, 'The First Rule of Plant Transposable Element Silencing: Location, Location, Location', *The Plant Cell*, vol. 28, no. 2, p. 304.

Simon, S.A. & Meyers, B.C. 2011, 'Small RNA-mediated epigenetic modifications in plants', *Current Opinion in Plant Biology*.

Singer, S.D., Liu, Z. & Cox, K.D. 2012, 'Minimizing the unpredictability of transgene expression in plants: The role of genetic insulators', *Plant Cell Reports*, pp. 13–25.

Singer, T., Yordan, C. & Martienssen, R.A. 2001, 'Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1)', *Genes & Development*, vol. 15, no. 5, pp. 591–602.

Singh, A., Jha, S.K., Bagri, J. & Pandey, G.K. 2015, 'ABA inducible rice protein phosphatase 2C confers ABA insensitivity and abiotic stress tolerance in Arabidopsis.', *PloS one*, vol. 10, no. 4, p. e0125168.

Skibbe, D.S., Fernandes, J.F., Medzihradszky, K.F., Burlingame, A.L. & Walbot, V. 2009, 'Mutator transposon activity reprograms the transcriptomes and proteomes of developing maize anthers', *The Plant Journal*, vol. 59, no. 4, pp. 622–33.

Slotkin, R.K. & Martienssen, R. 2007, 'Transposable elements and the epigenetic regulation of the genome.', *Nature reviews. Genetics*, vol. 8, no. 4, pp. 272–85.

Smith, A.M., Hansey, C.N. & Kaeppler, S.M. 2012, 'TCUP: A Novel hAT Transposon Active in Maize Tissue Culture', *Frontiers in Plant Science*, vol. 3, p. 6.

Smith, O., Clapham, A.J., Rose, P., Liu, Y., Wang, J. & Allaby, R.G. 2015, 'Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process', *Scientific Reports*, vol. 4, no. 1, p. 5559.

Smith, O., Nicholson, W. V., Kistler, L., Mace, E., Clapham, A., Rose, P., Stevens, C., Ware, R., Samavedam, S., Barker, G., Jordan, D., Fuller, D.Q. & Allaby, R.G. 2019, 'A domestication history of dynamic adaptation and genomic deterioration in Sorghum', *Nature Plants*, vol. 5, no. 4, pp. 369–79.

Song, C.-P., Agarwal, M., Ohta, M., Guo, Y., Halfter, U., Wang, P. & Zhu, J.-K. 2005, 'Role of an Arabidopsis AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses.', *The Plant cell*, vol. 17, no. 8, pp. 2384–96.

Song, T., Zhang, Q., Wang, H., Han, J., Xu, Z., Yan, S. & Zhu, Z. 2018, 'OsJMJ703, a rice histone demethylase gene, plays key roles in plant development and responds to drought stress', *Plant Physiology and Biochemistry*, vol. 132, pp. 183–8.

Song, X. & Cao, X. 2017, 'Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice', *Current Opinion in Plant Biology*.

Sridha, S. & Wu, K. 2006, 'Identification of AtHD2C as a novel regulator of abscisic acid responses in Arabidopsis', *Plant Journal*, vol. 46, no. 1, pp. 124–33.

Stelpflug, S.C., Sekhon, R.S., Vaillancourt, B., Hirsch, C.N., Buell, C.R., de Leon, N. & Kaeppler, S.M. 2016, 'An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development', *The Plant Genome*, vol. 9, no. 1, p. 0.

Steward, N., Ito, M., Yamaguchi, Y., Koizumi, N. & Sano, H. 2002, 'Periodic DNA Methylation in Maize Nucleosomes and Demethylation by Environmental Stress', *Journal of Biological Chemistry*, vol. 277, no. 40, pp. 37741–6.

Stitzer, M.C., Anderson, S.N., Springer, N.M. & Ross-Ibarra, J. 2019, 'The Genomic

Ecosystem of Transposable Elements in Maize', *bioRxiv*, p. 559922.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J. & Jacobsen, S.E. 2014, 'Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis', *Nature Structural and Molecular Biology*.

Su, Y., Wu, F., Ao, Z., Jin, S., Qin, F., Liu, B., Pang, S., Liu, L. & Guo, Q. 2019, 'Evaluating maize phenotype dynamics under drought stress using terrestrial lidar', *Plant Methods*, vol. 15, no. 1, p. 11.

Sultana, T., Zamborlini, A., Cristofari, G. & Lesage, P. 2017, 'Integration site selection by retroviruses and transposable elements in eukaryotes', *Nature Reviews Genetics*, vol. 18, no. 5, pp. 292–308.

Sun, Q. & Zhou, D.-X. 2008, 'Rice jmjC domain-containing gene JMJ706 encodes H3K9 demethylase required for floral organ development.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 36, pp. 13679–84.

Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. 2011, 'Revigo summarizes and visualizes long lists of gene ontology terms', *PLoS ONE*.

Takeuchi, T., Watanabe, Y., Takano-Shimizu, T. & Kondo, S. 2006, 'Roles ofjumonji andjumonji family genes in chromatin regulation and development', *Developmental Dynamics*, vol. 235, no. 9, pp. 2449–59.

Tan, M. 2010, 'Analysis of DNA methylation of maize in response to osmotic and salt stress based on methylation-sensitive amplified polymorphism.', *Plant physiology and biochemistry : PPB / Société française de physiologie végétale*, vol. 48, no. 1, pp. 21–6.

Tenaillon, M.I., Hufford, M.B., Gaut, B.S. & Ross-Ibarra, J. 2011, 'Genome size and transposable element content as determined by high-throughput sequencing in maize and Zea luxurians.', *Genome biology and evolution*, vol. 3, pp. 219–29.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W. & Su, Z. 2017, 'agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update.', *Nucleic acids research*, vol. 45, no. W1, pp. W122–9.

Tilman, D., Balzer, C., Hill, J. & Befort, B.L. 2011, 'Global food demand and the sustainable intensification of agriculture', *Proceedings of the National Academy of Sciences*, vol. 108, no. 50, pp. 20260–4.

Tiso, M., Tejero, J., Kenney, C., Frizzell, S. & Gladwin, M.T. 2012, 'Nitrite reductase activity of nonsymbiotic hemoglobins from Arabidopsis thaliana', *Biochemistry*, vol. 51, no. 26, pp. 5285–92.

Torii, K.U. 2004, 'Leucine-Rich Repeat Receptor Kinases in Plants: Structure, Function, and Signal Transduction Pathways', *International Review of Cytology*, vol. 234, pp. 1–46.

Trenberth, K.E., Branstator, G.W. & Arkin, P.A. 1988, 'Origins of the 1988 North American drought', *Science*.

Tsugane, K., Maekawa, M., Takagi, K., Takahara, H., Qian, Q., Eun, C.H. & Iida, S. 2006, 'An active DNA transposon nDart causing leaf variegation and mutable dwarfism and its related elements in rice', *Plant Journal*.

Tsukahara, S., Kobayashi, A., Kawabe, A., Mathieu, O., Miura, A. & Kakutani, T. 2009, 'Bursts of retrotransposition reproduced in Arabidopsis', *Nature*, vol. 461, no. 7262, pp. 423–6.

Verbruggen, N. & Hermans, C. 2008, 'Proline accumulation in plants: a review', *Amino*

Acids, vol. 35, no. 4, pp. 753–9.

Vicient, C.M. 2010, 'Transcriptional activity of transposable elements in maize.', *BMC genomics*, vol. 11, p. 601.

Wang, J., Yu, Y., Tao, F., Zhang, J., Copetti, D., Kudrna, D., Talag, J., Lee, S., Wing, R.A. & Fan, C. 2016, 'DNA methylation changes facilitated evolution of genes derived from Mutator-like transposable elements.', *Genome biology*, vol. 17, no. 1, p. 92.

Wang, L., Beissinger, T.M., Lorant, A., Ross-Ibarra, C., Ross-Ibarra, J. & Hufford, M.B. 2017, 'The interplay of demography and selection during maize domestication and expansion', *Genome Biology*, vol. 18, no. 1.

Wang, M., Qin, L., Xie, C., Li, W., Yuan, J., Kong, L., Yu, W., Xia, G. & Liu, S. 2014, 'Induced and Constitutive DNA Methylation in a Salinity-Tolerant Wheat Introgression Line', *Plant and Cell Physiology*, vol. 55, no. 7, pp. 1354–65.

Wang, W.-S., Pan, Y.-J., Zhao, X.-Q., Dwivedi, D., Zhu, L.-H., Ali, J., Fu, B.-Y. & 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*, vol. 62, no. 6, pp. 1951–60.

Wang, W., Qin, Q., Sun, F., Wang, Y., Xu, D., Li, Z. & Fu, B. 2016, 'Genome-Wide Differences in DNA Methylation Changes in Two Contrasting Rice Genotypes in Response to Drought Conditions.', *Frontiers in plant science*, vol. 7, p. 1675.

Wang, W., Zhao, X., Pan, Y., Zhu, L., Fu, B. & Li, Z. 2011, 'DNA methylation changes detected by methylation-sensitive amplified polymorphism in two contrasting rice genotypes under salt stress', *Journal of Genetics and Genomics*, vol. 38, no. 9, pp. 419–24.

Drought-responsive mechanisms in plant leaves revealed by proteomics 2016 (Multidisciplinary Digital Publishing Institute (MDPI)).

Wang, Xianglan, Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., Yang, X. & Qin, F. 2016, 'Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings', *Nature Genetics*.

Wang, Y., Beaith, M., Chalifoux, M., Ying, J., Uchacz, T., Sarvas, C., Griffiths, R., Kuzma, M., Wan, J. & Huang, Y. 2009, 'Shoot-specific down-regulation of protein farnesyltransferase (alpha-subunit) for yield protection against drought in canola.', *Molecular plant*, vol. 2, no. 1, pp. 191–200.

Wang, Y., Xue, X., Zhu, J.-K. & Dong, J. 2016, 'Demethylation of ERECTA receptor genes by IBM1 histone demethylase affects stomatal development', *Development*, vol. 143, no. 23, pp. 4452–61.

Wang, Y.G., An, M., Zhou, S.F., She, Y.H., Li, W.C. & Fu, F.L. 2014, 'Expression Profile of Maize MicroRNAs Corresponding to Their Target Genes Under Drought Stress', *Biochemical Genetics*, vol. 52, no. 11–12, pp. 474–93.

Watson, D. 2019, 'Adaption to Climate Change: Climate Adaptive Breeding of Maize, Wheat and Rice', *Sustainable Solutions for Food Security*.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M. & Schübeler, D. 2007, 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nature Genetics*, vol. 39, no. 4, pp. 457–66.

Wei, L., Wang, L., Yang, Y., Wang, P., Guo, T. & Kang, G. 2015, 'Abscisic acid enhances tolerance of wheat seedlings to drought and regulates transcript levels of genes

encoding ascorbate-glutathione biosynthesis.', *Frontiers in plant science*, vol. 6, p. 458.

Wei, L., Zhang, D., Xiang, F. & Zhang, Z. 2009, 'Differentially Expressed miRNAs Potentially Involved in the Regulation of Defense Mechanism to Drought Stress in Maize Seedlings', *International Journal of Plant Sciences*, vol. 170, no. 8, pp. 979–89.

Weinhold, A., Kallenbach, M. & Baldwin, I.T. 2013, 'Progressive 35S promoter methylation increases rapidly during vegetative development in transgenic Nicotiana attenuata plants.', *BMC plant biology*, vol. 13, p. 99.

Wessler, S.R. 2006, 'Transposable elements and the evolution of eukaryotic genomes.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17600–1.

West, P.T., Li, Q., Ji, L., Eichten, S.R., Song, J., Vaughn, M.W., Schmitz, R.J. & Springer, N.M. 2014, 'Genomic distribution of H3K9me2 and DNA methylation in a maize genome.', *PLoS one*, vol. 9, no. 8, p. e105267.

Wheeler, G.L., Jones, M.A. & Smirnoff, N. 1998, 'The biosynthetic pathway of vitamin C in higher plants', *Nature*, vol. 393, no. 6683, pp. 365–9.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P. & Schulman, A.H. 2007, 'A unified classification system for eukaryotic transposable elements', *Nature Reviews Genetics*, vol. 8, no. 12, pp. 973–82.

Williams, B.P. & Gehring, M. 2017, 'Stable transgenerational epigenetic inheritance requires a DNA methylation-sensing circuit', *Nature Communications*, vol. 8, no. 1, p. 2124.

Williams, B.P., Pignatta, D., Henikoff, S. & Gehring, M. 2015, 'Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat', *PLoS Genetics*.

Wimalanathan, K., Friedberg, I., Andorf, C.M. & Lawrence-Dill, C.J. 2018, 'Maize GO Annotation-Methods, Evaluation, and Review (maize-GAMER)', *Plant Direct*, vol. 2, no. 4, p. e00052.

Wu, F., Sheng, P., Tan, J., Chen, X., Lu, G., Ma, W., Heng, Y., Lin, Q., Zhu, S., Wang, Jiulin, Wang, Jie, Guo, X., Zhang, X., Lei, C. & Wan, J. 2015, 'Plasma membrane receptor-like kinase leaf panicle 2 acts downstream of the DROUGHT AND SALT TOLERANCE transcription factor to regulate drought sensitivity in rice.', *Journal of experimental botany*, vol. 66, no. 1, pp. 271–81.

Wu, J., Jiang, Y., Liang, Y., Chen, L., Chen, W. & Cheng, B. 2019, 'Expression of the maize MYB transcription factor ZmMYB3R enhances drought and salt stress tolerance in transgenic plants', *Plant Physiology and Biochemistry*, vol. 137, pp. 179–88.

Wu, L., Tian, L., Wang, S., Zhang, J., Liu, P., Tian, Z., Zhang, H., Liu, H. & Chen, Y. 2016, 'Comparative Proteomic Analysis of the Response of Maize (Zea mays L.) Leaves to Long Photoperiod Condition.', *Frontiers in plant science*, vol. 7, p. 752.

Wu, X., Dabi, T. & Weigel, D. 2005, 'Requirement of homeobox gene STIMPY/WOX9 for Arabidopsis meristem growth and maintenance.', *Current biology : CB*, vol. 15, no. 5, pp. 436–40.

Xu, J., Zhou, S., Gong, X., Song, Y., van Nocker, S., Ma, F. & Guan, Q. 2017, 'Single-Base Methylome Analysis Reveals Dynamic Epigenomic Differences Associated

with Water Deficit in Apple', *Plant Biotechnology Journal*.

Xuan, W., Audenaert, D., Parizot, B., Möller, B.K., Njo, M.F., De Rybel, B., De Rop, G., Van Isterdael, G., Mähönen, A.P., Vanneste, S. & Beeckman, T. 2015, 'Root Cap-Derived Auxin Pre-patterns the Longitudinal Axis of the Arabidopsis Root.', *Current biology : CB*, vol. 25, no. 10, pp. 1381–8.

Yamada, M., Morishita, H., Urano, K., Shiozaki, N., Yamaguchi-Shinozaki, K., Shinozaki, K. & Yoshiba, Y. 2005, 'Effects of free proline accumulation in petunias under drought stress', *Journal of Experimental Botany*, vol. 56, no. 417, pp. 1975–81.

Yan, J., Aznar, A., Chalvin, C., Birdseye, D.S., Baidoo, E.E.K., Eudes, A., Shih, P.M., Loqué, D., Zhang, A. & Scheller, H. V. 2018, 'Increased drought tolerance in plants engineered for low lignin and low xylan content', *Biotechnology for Biofuels*, vol. 11, no. 1, p. 195.

Yang, L. & Bennetzen, J.L. 2009, 'Distribution, diversity, evolution, and survival of Helitrons in the maize genome.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 47, pp. 19922–7.

Ye, G., Ma, Y., Feng, Z. & Zhang, X. 2018, 'Transcriptomic analysis of drought stress responses of sea buckthorn (Hippophae rhamnoidessubsp. sinensis) by RNA-Seq.', *PloS one*, vol. 13, no. 8, p. e0202213.

Yoder, J.A., Walsh, C.P. & Bestor, T.H. 1997, 'Cytosine methylation and the ecology of intragenomic parasites.', *Trends in genetics : TIG*, vol. 13, no. 8, pp. 335–40.

Yoo, C.Y., Pence, H.E., Jin, J.B., Miura, K., Gosney, M.J., Hasegawa, P.M. & Mickelbart, M. V 2010, 'The Arabidopsis GTL1 transcription factor regulates water use efficiency and drought tolerance by modulating stomatal density via transrepression of SDD1.', *The Plant cell*, vol. 22, no. 12, pp. 4128–41.

Yu, C. 2011, 'China's water crisis needs more than words.', *Nature*, vol. 470, no. 7334, p. 307.

Yu, C., Li, Y., Li, B., Liu, X., Hao, L., Chen, J., Qian, W., Li, S., Wang, G., Bai, S., Ye, H., Qin, H., Shen, Q., Chen, L., Zhang, A. & Wang, D. 2010, 'Molecular analysis of phosphomannomutase (PMM) genes reveals a unique PMM duplication event in diverse Triticeae species and the main PMM isozymes in bread wheat tissues.', *BMC plant biology*, vol. 10, p. 214.

Yu, X., Li, Li, Li, Lei, Guo, M., Chory, J. & Yin, Y. 2008, 'Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins ELF6 and REF6 in Arabidopsis.', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 21, pp. 7618–23.

Zabet, N. & Tsang, J. 2015, 'DMRcaller: Differentially Methylated Regions caller', *R package version 1.8.0.*

Zakrzewski, F., Schmidt, M., Van Lijsebettens, M. & Schmidt, T. 2017, 'DNA methylation of retrotransposons, DNA transposons and genes in sugar beet (*Beta vulgaris* L.)', *The Plant Journal*, vol. 90, no. 6, pp. 1156–75.

Zenda, T., Liu, S., Wang, X., Jin, H., Liu, G. & Duan, H. 2018, 'Comparative Proteomic and Physiological Analyses of Two Divergent Maize Inbred Lines Provide More Insights into Drought-Stress Tolerance Mechanisms.', *International journal of molecular sciences*, vol. 19, no. 10.

Zeng, F. & Cheng, B. 2014, 'Transposable Element Insertion and Epigenetic Modification Cause the Multiallelic Variation in the Expression of FAE1 in Sinapis

alba', *The Plant Cell*, vol. 26, no. 6, pp. 2648–59.

Zhang, H., Lang, Z. & Zhu, J.-K. 2018, 'Dynamics and function of DNA methylation in plants', *Nature Reviews Molecular Cell Biology*, vol. 19, no. 8, pp. 489–506.

Zhang, H., Lang, Z. & Zhu, J.K. 2018, 'Dynamics and function of DNA methylation in plants', *Nature Reviews Molecular Cell Biology*.

Zhang, J., Zuo, T., Wang, D. & Peterson, T. 2014, 'Transposition-mediated DNA re-replication in maize.', *eLife*, vol. 3, p. e03724.

Zhang, Liguo, Cheng, Z., Qin, R., Qiu, Y., Wang, J.-L., Cui, X., Gu, L., Zhang, X., Guo, X., Wang, D., Jiang, L., Wu, C., Wang, H., Cao, X. & Wan, J. 2012, 'Identification and Characterization of an Epi-Allele of FIE1 Reveals a Regulatory Linkage between Two Epigenetic Marks in Rice', *The Plant Cell*, vol. 24, no. 11, p. 4407.

Zhang, Lixin, Gao, M., Hu, J., Zhang, X., Wang, K. & Ashraf, M. 2012, 'Modulation Role of abscisic acid (ABA) on growth, water relations and glycinebetaine metabolism in two maize (Zea mays L.) cultivars under drought stress.', *International journal of molecular sciences*, vol. 13, no. 3, pp. 3189–202.

Zhang, N., Yang, J., Wang, Z., Wen, Y., Wang, J., He, W., Liu, B., Si, H. & Wang, D. 2014, 'Identification of Novel and Conserved MicroRNAs Related to Drought Stress in Potato by Deep Sequencing', R. Margis (ed.),*PLoS ONE*, vol. 9, no. 4, p. e95489.

Zhang, X., Lei, L., Lai, J., Zhao, H. & Song, W. 2018, 'Effects of drought stress and water recovery on physiological responses and gene expression in maize seedlings', *BMC Plant Biology*, vol. 18, no. 1, p. 68.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W.-L., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E. & Ecker, J.R. 2006, 'Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis.', *Cell*, vol. 126, no. 6, pp. 1189–201.

Zhao, F., Zhang, D., Zhao, Y., Wang, W., Yang, H., Tai, F., Li, C. & Hu, X. 2016, 'The Difference of Physiological and Proteomic Changes in Maize Leaves Adaptation to Drought, Heat, and Combined Both Stresses.', *Frontiers in plant science*, vol. 7, p. 1471.

Zheng, X., Chen, L., Li, M., Lou, Q., Xia, H., Wang, P., Li, T., Liu, H. & Luo, L. 2013, 'Transgenerational variations in DNA methylation induced by drought stress in two rice varieties with distinguished difference to drought resistance', *PLoS ONE*, vol. 8, no. 11, p. e80253.

Zheng, X., Chen, L., Xia, H., Wei, H., Lou, Q., Li, M., Li, T. & Luo, L. 2017, 'Transgenerational epimutations induced by multi-generation drought imposition mediate rice plant's adaptation to drought condition.', *Scientific reports*, vol. 7, p. 39843.

Zheng, X., Liu, H., Ji, H., Wang, Y., Dong, B., Qiao, Y., Liu, M. & Li, X. 2016, 'The Wheat GT Factor TaGT2L1D Negatively Regulates Drought Tolerance and Plant Development', *Scientific Reports*, vol. 6, no. 1, p. 27042.

Zheng, Y., Ding, Y., Sun, X., Xie, S., Wang, D., Liu, X., Su, L., Wei, W., Pan, L. & Zhou, D.-X. 2016, 'Histone deacetylase HDA9 negatively regulates salt and drought stress responsiveness in Arabidopsis', *Journal of Experimental Botany*, vol. 67, no. 6, pp. 1703–13.

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*, vol.

39, no. 1, pp. 61–9.