Strategies To Alter DNA methylation Patterns In plants

by

Mohd Azinuddin Ahmad Mokhtar

Submitted in accordance with the requirements for the degree of

Doctoral of Philosophy

The University of Leeds

Schools of Biology

February 2017

Intellectual Property and Publication Statement

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to thank my project supervisor Professor Peter Meyer for his support, guidance and encouragement throughout this project and for allowing to work in his lab. I also wish to thanks my co-supervisor Dr Joan Boyes for her help and support.

I am also indebted to Dr Iris Heidmann and Suzan Out for carrying out tomato plant transformations at Enza Zaden. I would also like to thank Dr Fridtjof Weltmeier and his team for carrying out maize plant transformation, selection and growing the plant. His helped made the collaboration with KWS is interesting.

Past and present group members, including Dr Michael Watson, Dr Elham Soliman, Dr Yu-Fei Lin, Dr Valya Vassileva, Dr Hekmat Mustafa, Samuel Brocklehurst, Elizabeth Hollwey, Abirah Zulkifli and Najihah Mohamed for their help and enjoyable working environment.

Finally, my special gratitude to my family, in particular Mum, Dad, Wife and Son. I am indebted to them for their never-ending support and love.

Last but not least to all my friends whose made life at Leeds wonderful, Yusuf, Yati, Hayati, Aisyah, Asraf, Dayah, Khai, Zana, Zul, Nazlah, Suhana, Adam, Is, Laila, Zam, Wana, Arif, Hamdan, Nad, Refiq, Hisham, Aiman, Farzil, Fadzli, Hakam, Razif, Hulubalangs and those I have forgotten to mention. Thank you very much for all your support and fun times.

Abstract

Epigenetic regulation is achieved through cytosine DNA methylation and histone modification. Epigenetic regulation is not only responsible for regulating gene-coding regions, it is also involved in silencing harmful transposable elements and repetitive elements. Naturally, DNA methylation patterns may vary between individual plants of the same species, influenced by difference exposures to environmental stresses. These changes are heritable, as the plants adapt to challenges in their growth environment. The dynamics and heritability of DNA methylation changes makes producing an epi-mutant variety of crop plants interesting. New epi-varieties may potentially carry interesting phenotypes, with high commercial values. Establishment and maintenance of DNA methylation is controlled by DNA methyltransferases, which creates an opportunity for inducing DNA methylation changes by interfering with the expression of DNA methyltransferases in plants. In this study, we used different strategies in various plant species to induce DNA methylation changes. The first strategy used inverted repeats to silence the *MET1* gene, and indicates the importance of having the appropriate level of MET1 expression in maize for plant growth and development. The second strategy employed the TALEN and CRISPR genome editing tools for inducing point mutagenesis in the tomato *MET1* gene. However, high dependency of tomato to *MET*1 gene have inhibited callus regeneration. The third strategy used over-expression of the *CMT2* gene to induce phenotype and methylation pattern changes. In addition to using the available strategies, we developed a novel tool for the proof-of-concept targeted demethylation of stable methylated regions in *Arabidopsis,* which could be extended as epigenome editing tools.

Table of Contents

List of Abbreviation

5mC	5-methyl-cytosine
AGO	Argonaute
BAH	Bromo-Adjacent Homology
BASTA	Glufosinate Ammonium
bp	Base Pair
Cas9	CRISPR Associated Protein 9
CaMV	Cauliflower Mosaic Virus
cDNA	Complementary Deoxyribonucleic Acid
CHROMO	Chromatin Organisation Modifier
CMT1	Chomomethyltransferase 1
CMT2	Chomomethyltransferase 2
CMT3	Chomomethyltransferase 3
COPIA	Transposable Element Copia
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
Dam	DNA Adenine Methyltransferase
DBD	DNA Binding Protein
dCas9	Deactivate Cas9 protein
dCas9-TET1	Deactivate Cas9 fuse with Ten-Eleven Translocase 1
dCas9-TET3	Deactivate Cas9 fuse with Ten-Eleven Translocase 3
DCC	Duplication, Degeneration and Complementation Model
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
DCL	Dicer-Like
Dcm	DNA Cytosine Metlytransferase
DDM1	Decrease in DNA Methylation 1
DEF	Deficiens gene
DME	Demeter
DML	Demeter-Like
DNA	Deoxyribonucleic Acid
DNMT1	DNA Methyltransferase 1
DRM	Domains Re-Arranged Methyltransferase
DSB	Double Stranded Break
dsRNA	Double stranded RNA
DWF	Dwarf
EF1α	Elongation factor 1α
EMS	Ethyl Methanesulfonate
EpiEffector	Epigenetic Effector domain
FLC	Flowering Locus C
fl-cDNA	Full length Complementary Deoxyribonucleic Acid
FPGS	Folylpolyglutamate Synthase
FWA	Flowering Wageningen
G	Guanine
gDNA	Genomic DNA
GNOM	ARF guanine-nucleotide exchange factor
HBB	Beta Globin
HDA6	Histone Deacetylase 6
HR	Homologues Recombination
HSE	Heat Shock Element

IAA	Indole Acetic Acid
IR	Inverted Repeat
KAN	Kanamycin
KIN2	Kinesin-Like Protein 2
LTR	Long Terminal Repeat
MADS69	MADS-Transcription Factor 69
MB	Mega Base
McrBC	Methyl-Cytosine Endonuclease Sensitive Enzyme
MET1	Methyltransferase 1
MLO	Mildew-Resistance Locus
MMR	Mismatch Repair Protein
mRNA	Messenger Ribonucleic Acid
MS	Murashige and Skoog
MSP	Methylation-Sensitive PCR
MTases	Methyltransferases
MYB77	MYB-Transcription Factor 77
NHEJ	Non-Homologous End Joining
NTS	Nuclear Targeting Sequence
ONSEN	Transposon of long terminal repeat family
ORF	Open Reading Frame
OsMET1	Rice Methyltransferase 1
PAM	Protospacer Adjacent Motif
PEV	Position Effect Variegation
POLII	RNA Polymerase II
POLIII	RNA Polymerase III
POLIV	RNA Polymerase IV
PRC2	Polycomb Complex 2
PTGS	Post-Transcriptional Gene Silencing
RD29A	Responsive to Desiccation 29A
RdDM	RNA-directed DNA Methylation
RDR	RNA-Dependent RNA polymerase
RDR2	RNA-Dependent RNA polymerase 2
RDR6	RNA-Dependent RNA polymerase 6
RE	Restriction Enzyme
RFD	Replication Foci Targeting Sequence
RGB	Red/Green/Blue
RISC	RNA-induced Silencing Complex
R-M	Restriction-Modification
RNA	Ribonucleic Acid
RNA-seq	Total Ribonucleic acid Sequencing
ROS1	Repressor of Silencing 1
RPS	Repetitive Petunia Sequence
RTCS	Rootless Concerning Crown and Seminal Lateral Roots
S-AdoMET	S-Adenosyl-L-Methionine
sgRNA	Single guided Ribonucleic Acid
siRNA	Small interfering Ribonucleic Acid
smqRT-PCR	Semi-quantitative Reverse Transcriptase PCR
SNPs	Single Nucleotide Polymorphisms
SPS2	Sucrose phosphate synthase 2
SRA	Set and Ring Associated
ssRNA	Single stranded Ribonucleic Acid

SUVH4	Suppressor of Variegation 3-9 Homologue 4
SUVH5	Suppressor of Variegation 3-9 Homologue 5
SUVH6	Suppressor of Variegation 3-9 Homologue 6
TALE	Transcriptional Activator-Like Effector
RVD	Repeat Variable Di-residues
T	Thymine
TALEN	Transcriptional Activator-Like Effector Nucleases
TALE-TET1	Transcriptional Activator-Like Effector fuse with Ten-Eleven Translocase 1
T-DNA	Transfer DNA
TE	Transposon Element
TET	Ten-Eleven Translocase
TET1CD	Ten-Eleven Translocase 1 Catalytic Domain
TET3	Ten-Eleven Translocase 3
TGS	Transcriptional Gene Silencing
TIR	Terminal Inverted Repeat
tracrRNA	trans-activating CRISPR RNA
TRD	Target Recognition Domain
VIM1	Variable In Methylation 1
WT	Wild Type
ZFN	Zinc Finger Nucleases
ZmMET1	Maize Methyltransferase 1

List of Tables

Chapter 2

Chapter 4

Chapter 5

Chapter 7

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 8

Chapter 1

General Introduction

1.1 Introduction to Epigenetics

The term ‘epigenetic’ was coined by Conrad Waddington in the 1940s, and is an amalgamation of ‘Epi’, meaning upon or over, and ‘genetic’, meaning involving genes, thus, the term reflected the study of events beyond genes. Waddington defined epigenetics as ‘the branch of biology, which studies the causal interactions between genes and their product, which bring the phenotypes into being’ (Waddington, 1968). Over time, with better understanding of the genome structure and gene regulation, epigenetic was defined as ‘the study of changes in gene function that are mitotically and/or meiotically heritable, and that do not entail a change in DNA sequence’ (Wu and Morris 2001). However, since this definition does not include non-mitotic factors, in 2007, the following definition was proposed: ‘the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity’ (Bird, 2007). A year later, during the 2008 Cold Spring Harbour meeting, a consensus definition for epigenetic was made as: ‘reversible and heritable changes in gene expression that occur without any DNA sequence alteration’ (Berger *et al*, 2009).

Epigenetic gene regulation occurs by a series of chemical modifications that occur on DNA or histone proteins (Figure 1.1). Chemical modification of DNA occurs through DNA methylation, a process of adding a methyl group to a cytosine residue. Whereas, histone modification occurs though post-transcriptional modification (PTM), which can involve methylation, acetylation, ubiquitylation, and sumoylation,

and results in chromatin remodelling (Sultan and Day, 2011). In plants, DNA methylation occurs at three different sequence contexts, CG, CHG, and CHH that work redundantly to silence a gene, whereas, histone modification can cause either transcription activation or repression. In general, histone methylation, and ubiquitylation can cause transcription activation or repression, while histone acetylation and phosphorylation are often associated with transcription activation (reviewed in Karim *et al*., (2016); Pfluger and Wagner, (2007). These epigenetic changes can cause several phenomena, such as non-Mendelian inheritance (Hollick and Chandler, 1998), transgene silencing (Meyer *et al.,* 1992), gene imprinting (Kinoshita, 2004), paramutation (Brink *et al.,* 1968), and position-effect variegation (PEV) (Singh *et al.,* 2008b).

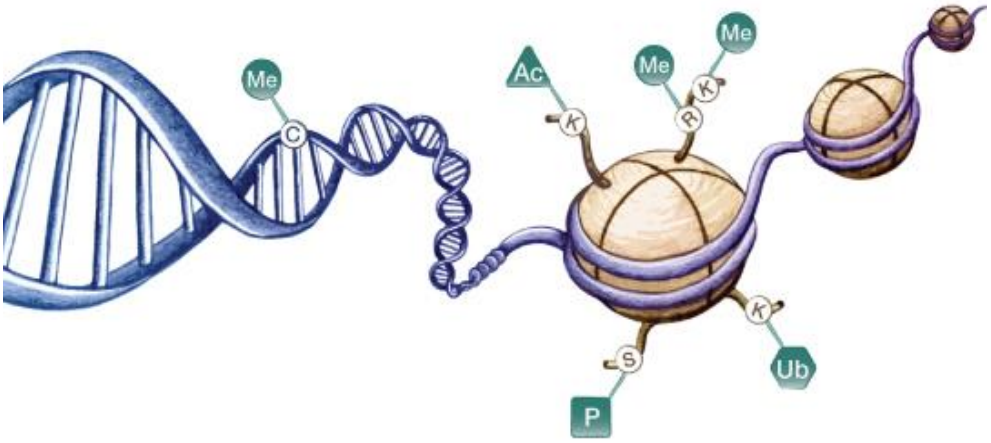


Figure 1.1: Epigenetic gene regulation. Methylation can occur at cytosine residues of DNA, and lysine and arginine residues of histone proteins. Acetylation and ubiquitylation can occur at lysine residues, while phosphorylation occurs at serine residues of histone proteins. Me: methylation; Ac: acetylation; P: phosphorylation; Ub: ubiquitylation; C: cytosine; K: lysine; R: arginine, S: serine. (Taken from Promega, 2017).

1.2 DNA Methylation

DNA methylation is an epigenetic mark that is found in all bacteria, fungi, mammals, and plants studied so far. In bacterial genomes, it is found at a 5-methyl cytosine (m5C), N6-methyl-adenine (m6A), and N4-methy-cytosine (m4C), which is exclusive to bacteria (Jeltsch, 2002). As with eukaryotes, bacterial methylation causes gene regulation, such as control of expression for phase variation, and acts as a signal through a restriction-modification (R-M) system, which protects bacterial genome from its own immune response against infection with foreign DNA (Casadesús and Low, 2006). DNA adenine methyltransferase (Dam) is responsible for the methylation of adenine at GATC sequences, while DNA cytosine methyltransferase (Dcm) is responsible for methylation at the internal C of CCWGG (where W is either A or T) sequences (Casadesús and Low, 2006).

In mammals, DNA methylation is found almost exclusively at CG sequence contexts, which are established during embryotic development (Bestor, 2000). Most mammalian genes possess CG islands, a CG-rich region, at their gene promoter (Bird, 2007), and methylation at these CG islands is known to be responsible for gene repression, such as with inactive X-chromosomes and imprinted genes (Alcaiay and Toniolo, 1988). Even though, CG islands are often found unmethylated, significant numbers of CG islands have been found to be differentially methylated in normal tissues and cell types, indicating gene specific cellular functions (Strathdee *et al.*, 2004).

In plants, DNA methylation occurs at three sequence contexts, CG, CHG, and CHH (where H is either A, T, or G). In general, the distribution of plant DNA methylation

for each sequence context is: > 80%, 20-60%, and < 22%, at CG, CHG, and CHH, respectively, with DNA methylation landscape being shaped by disproportionate distributions of DNA methylation at transposable elements (TE) and repetitive sequence regions (Niederhuth *et al.*, 2016). Unlike in mammals, DNA methylation reprograming in plants occurs partially through sexual reproduction (Jullien *et al.*, 2012), which might explain the transgenerational epigenetic instability seen in plants (Schmitz *et al.*, 2011).

1.3 Plant DNA Methyltransferases

In general, there are two functions of DNA methyltransferases (MTases): The N-terminal domain is responsible for recognising a specific DNA sequence, while the C-terminal domain catalyses the transfer of a methyl group from S-adenosyl-L-methionine (S-AdoMET) to 5-carbon of pyrimidine ring of cytosine (5mC) (Kumar *et al.*, 1994). DNA MTases can be categorised into two distinct but complementary mechanisms based on their targets, the maintenance DNA MTases and the *de novo* DNA methylation (Figure 1.2)

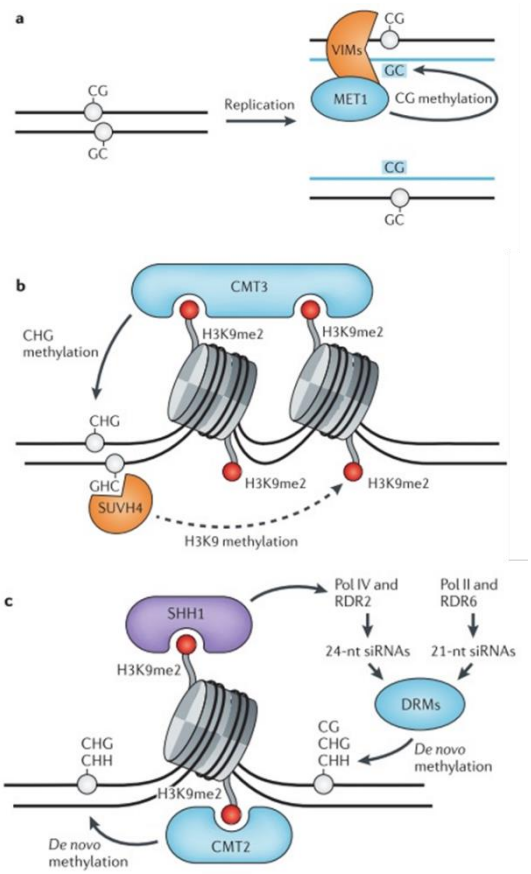


Figure 1.2: Plant DNA methyltransferases and their methylation pathways. A. MET1 mediates CG methylation at a new strand with the help of VIM1 to recognise hemi-methylated patterns. **B**. CHG methylation occurs through a self-reinforced loop, where CMT3 mediates CHG methylation by binding to H3K9me2, and methylated CHG recruits SUVH4 binding for methylating H3K9. **C**. Several pathways are involved in *de novo* methylation, initiated by SHH1 involvement in siRNA biogenesis through POLIV with RDR2, or POLII with RDR6 pathways, followed by recruitment of DRMs complex for methylating all sequence contexts. CMT2 methylates CHH and CHG via *de novo* methylation by binding to H3K9me2. Adapted from Kawashima and Berger (2014).

1.3.1 Maintenance DNA Methyltransferases in Plants

Maintenance DNA MTases are responsible for methylating newly synthesised DNA strands, which are produced from replication events. The MTase uses the sister DNA strands for target identification to ensure the correct methylation pattern is passed to the next generation.

Unlike mammals, which only have one maintenance MTase, DNA methyltransferase 1 (DNMT1), plants possess several distinct MTase families for methylating CG, CHG, and CHH context sequences. In *Arabidopsis*, the plant DNMT1 homologue, DNA METHYLTRANSFERASE 1 (MET1) is responsible for maintaining methylation at the CG context, with the help of VARIABLE IN METHLYLATION 1 (VIM1) for recognising the hemi-methylated CG sequence. MET1 also requires a chromatin remodelling factor, DECREASE IN DNA METHYLATION (DDM1) for methylating heterochromatic regions (Zemach *et al.*, 2013). In some plant species, there is more than one copy of the METs genes, and each have become specialised, to function in different tissues (Bernacchia *et al.*, 1998; Qian *et al.*, 2014; Yamauchi *et al.*, 2014).

CHROMOMETHYLASES (CMTs) are another plant maintenance MTase family, which are uniquely found in plants, and are defined by the presence of a chromodomain at the catalytic domain, which interacts with chromatin protein, directing CMTs to methylate the heterochromatin region (Henikoff and Comai, 1998). The CMT family may have evolved from a specialisation event that methylated the transposon and retrotransposon elements that developed a non-CG promoter (Pavlopoulou and Kossida, 2007). In *Arabidopsis*, the CMT family consists of three genes, CMT1, CMT2, and CMT3. CMT1 is found truncated in most *Arabidopsis* ecotypes (Henikoff and Comai, 1998), while CMT2 mediates methylation via *de novo* DNA methylation mechanisms. CMT2 and CMT3 have redundant function, whereby both can methylate CHG contexts, but CMT2 is also capable of methylating CHH contexts (Stroud *et al.*, 2014). Methylation by CMT3 occurs through a self-reinforcing loop between histone and DNA methylation (Du *et al.*, 2012). The chromodomain of CMT3 binds to H3K9me1 or H3K9me2 and

methylates the nearby CHG context (Du *et al.*, 2012; Stroud *et al.*, 2014). The methylated CHG provides a binding site for the SET- AND RING- ASSOCIATED (SRA) domain of H3K9 methyltransferase, SU(VAR) HOMOLOGUE 4 (SUVH4), which mediates H3K9 methylation (Du *et al.*, 2014).

1.3.2 *De novo* DNA Methyltransferases in Plants

De novo DNA methylation involves the addition of a methyl group to previously unmethylated cytosine residues, and was first observed in tobacco plants infected with viroids (Wassenegger *et al.*, 1994). *De novo* DNA methylation, also known as RNA-directed DNA methylation (RdDM) pathway in plants, methylates all sequence contexts (Matzke and Mosher, 2014; Stroud *et al.*, 2014; Zhong *et al.*, 2014), however CHH methylation is frequently-linked to RdDM because of its asymmetrical sequence. In *Arabidopsis*, the RdDM pathway (reviewed in Kawashima and Berger, (2014)) is initiated by small interfering RNA (siRNA) biogenesis, which involves the coordinated activity of the chromatin binding protein, SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) (Law *et al.*, 2013). This enables the binding of plant RNA POLYMERASE IV (POLIV) to DNA, initiating transcription and producing a single stranded RNA (ssRNA). The ssRNA provides a templates for RNA-DIRECTED RNA POLYMERASE 2 (RDR2) to produce a double stranded RNA (dsRNA), before cleavage by DICER-LIKE 3 (DCL3) into siRNA duplexes of 24-nuclotides long, and is followed by methylation at the 3’ end by HUA ENHANCER 1 (HEN1) (Li et al, 2005).

One strand of siRNA forms a complex with ARGONATE 4 (AGO4) (Zhong *et al.*, 2014), which acts as a target for the RdDM pathway. KOW DOMAIN-

CONTAINING TRANSCRIPTION FACTOR 1 (KTF1) facilitates the binding of the siRNA-AGO4 complex (Rowley *et al.,* 2011) to homologous siRNA sequences of RNA transcripts produced by POLV, and these in turn recruit DOMAIN REARRANGED METHYTRANSFERASE 2 (DRM2) to the DNA, which results in cytosine methylation. The non-canonical RdDM pathway is another methylation pathway, and involves POLII and RDR6 during siRNA biogenesis. This pathway produces 21-nucloetide siRNA, rather than 24-nucleotide siRNA, followed by the formation of a complex with AGO6, before methylating through DRM2 (Nuthikattu *et al.,* 2013).

CMT2 also mediates CHH methylation, but via RdDM-independent mechanisms (Zemach *et al.,* 2013), and causes methylation of regions other than DRM2 and CMT3 targets (Stroud *et al.,* 2014). CMT2 binds to H3K9me2, followed by *de novo* methylation of nearby CHH contexts (Zemach *et al.,* 2013). Recently, the methylation of different CHH regions were explained by the preferential methylation of CAA and CTA sub-contexts by CMT2 (Gouil and Baulcombe, 2016).

1.4 DNA Demethylation in Plants

The reversibility of DNA methylation is important for regulating the gene expression of certain events, such as tissue specialisation or environmental adaptation, and in plants, DNA demethylation can occur either passively or actively. Passive demethylation is caused by inefficient maintenance of DNA methylation, or a shortage of S-AdoMET leading to the inability to supply methyl groups for cytosine methylation, whereas, active demethylation involves the enzymatic removal

of a methyl group from cytosine DNA. In *Arabidopsis* species, DNA glycosylases initiate the removal of 5mC by hydrolysing glycosidic bonds, and subsequently apurinic/apyrimidinic (AP) lyase introduces nicks in the DNA backbone. This triggers DNA repair mechanisms that result in the incorporation of unmethylated cytosine at the gap (Penterman *et al.,* 2007).

Arabidopsis species have four DNA glycosylases, the DEMETER (DME), REPRESSOR OF SILENCING 1 (ROS1), and two DEMETER-LIKE enzymes (DML2 and DML3). Although exact details of the glycosylases are poorly understood, it is known that each one is able to demethylate all sequence contexts (Penterman *et al.,* 2007), and this could involve a targeting mechanism because the triple mutants, *ros1 dml2 dml3* have been shown to cause changes in DNA methylation status at certain loci (Penterman *et al.,* 2007). Furthermore, ROS1 has been reported to form a complex with ROS3, which has RNA-binding capacity (Zheng *et al.,* 2008). ROS1 is also reported to maintain demethylation at several promoters of endogenous genes, while DML2 and DML3 are responsible for preventing hypermethylation in *Arabidopsis* vegetative tissue (Zhu *et al.,* 2007).

1.5 Sequence Context-Independent DNA Methyltransferases

DNA methyltransferases have always linked to the rigid sequence contexts that they methylated. However, there are studies showing that cross-context sequence methylation is undertaken by DNA methyltransferases. For example, MET1 was found to not restrict its target to only CG methylation, and this is because only partial CG methylation was restored when endogenous MET1 was re-introduced into the *met1* mutant background (Watson *et al.,* 2014; Zubko *et al.,* 2012).

Furthermore, a reduction in non-CG methylation at the *REPETITIVE PETUNIA SEQUENCE* (*RPS*) transgene was observed when the transgene was introduced into *met1* via crossing (Singh *et al.,* 2008a). This was further supported by findings that MET1 is required for the methylation of CCG contexts (Gouil and Baulcombe, 2016; Yaari *et al.,* 2015). Interestingly, partially restored CG methylation does not require a passage through germlines, which suggests the possibility of *de novo* methylation mechanisms by MET1 (Zubko *et al.,* 2012). *De novo* activity of MET1 could be region specific, in which an 11 bp stem loop acts as a signal for the targeted recognition of MET1 at the *RPS* transgene (Gentry and Meyer, 2013). However, there are also several unmethylated endogenous regions found in MET1-restored *Arabidopsis* species (Watson *et al.,* 2014). Even though there is growing evidence of MET1 activity at non-CG sites, so far, there is no evidence of other DNA MTases that can methylate CG sites (Stroud *et al.,* 2014), except for a joint activity involving MET1 (Singh *et al.,* 2008a; Gouil and Baulcombe, 2016). However, details surrounding the mechanisms of non-CG MET1-dependent methylation are currently unknown, although it is speculated that MET1 may act as a mediator for quick responses to environmental change (Meyer, 2015).

One study showed that CHG methylation was only partially lost in *cmt3* mutants, indicating that CMT2 and CMT3 both mediate non-CG methylation, and CMT2 mediates CHH and CHG context sequences (Stroud *et al.,* 2014). This was further supported by only a 25% reduction of CHG methylation in *cmt2* mutants (Gouil and Baulcombe, 2016). Furthermore, *in vitro* studies have shown that CMT2 exhibits strong methylation activity at CHG and CHH contexts (Stroud *et al.,* 2014). However, differences in global methylation patterns between *cmt2* and *cmt3* mutants suggest that they might have different sequence preferences (Stroud *et al.,* 2014).

Further studies of sub-context sequences, suggest that there are no sub-context preferences of CMT proteins, but speculate the differences might be caused by the preferences of histone methylases, with SUVH5 and SUVH6 redundantly binding at CCG/CGG sites, while SUVH4 prefers binding to CAG/CTG (Gouil and Baulcombe, 2016).

In addition, context-independent DNA methyltransferases have been reported to work together to methylate certain targets. For example, *RPS* transgenes lose both CG and non-CG methylation when transferred into *drm2/cmt3* mutants, suggesting that MET1, DRM2, and CMT3 may be jointly involved in methylating the *RPS* gene (Singh *et al.,* 2008a). Cooperation between two MTases might be explained from the perspective of sub-context sequences, where MET1 is recruited to methylate mCGG and CmCG (methylation at the first and second cytosine, respectively), triggering the binding of SUVH5/6. This results in the methylation of H3K9, followed by recruitment of CMT3 to methylate the first cytosine of the CCG site (Gouil and Baulcombe, 2016). Furthermore, structural analysis of MET1 has found that the N-terminal domain can directly interact with the C-terminal domain of HISTONE DEACETYLASE 6 (HDA6), which suggests a joint collaboration for maintaining the heterochromatin region (Liu *et al.,* 2012).

1.6 Biological Importance of Plant DNA Methylation

Correct DNA methylation is required for maintaining plant genome stability, especially in maize, where 82% of its genome consists of transposable elements (TE) and repetitive sequences (Haberer *et al.,* 2005; Vicient, 2010). In *Arabidopsis* species, depletion of the DDM1 protein has given rise to several growth

abnormalities, such as *clm* and *bns* phenotypes. The *clm* phenotype is identified by the repression of shoot and petiole development, which are produced by insertion of the CACTA transposon family at the *DWARF 4* (*DWF4*) gene, resulting in a truncated *DWF 4* gene (Miura *et al.,* 2001). Whereas, the *bns* phenotype is identified by short compact inflorescence, and a shorter plant, and is caused by the spread of hypermethylation from long interspersed nuclear element (LINE) retrotransposons at the 3' non-coding region, into the *BONSAI* gene (Saze and Kakutani, 2007). There are also *ddm1*-induced gain-of function phenotypes, such as the late-flowering phenotype produced by the ectopic expression of the *FWA* gene, caused by hypomethylation of repeats at the 5' region of the gene (Soppe *et al.,* 2000). In addition, *bal* phenotypes, characterised by twisted leaves and dwarfed plants, result from the silencing effects of a gene at the *BAL* locus being over-expressed (Stokes *et al.,* 2002).

Genomic imprinting is another phenomenon that is related to DNA methylation, in which one of the alleles originating from the parent is silenced. This parent-of-origin-specific gene expression is primarily found in the endosperm during seed development (Gehring, 2013). Expression of imprinted genes is either controlled by differential DNA methylation, polycomb repressive complex 2 (PRC2) activity, or both (Hsieh *et al.,* 2011). In *Arabidopsis* species, there are several imprinted genes expressed from the maternal genome, such as *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*), *MULTICOPY SUPRESSOR OF IRA 1* (*MSI1*), *FERTILIZATION INDEPENDENT SEED 2* (*FIS2*), and *MEDEA* (*MEA*), while paternal imprinted genes include *FWA*, *MYB THREE REPEAT 2* (*MYB3R2*), and *PHERES 1* (*PHE1*) (reviewed in Gehring, 2013). DME protein is one of the DNA glycosylases, that specifically demethylates maternal imprinted genes, and leads to

gene expression (Gehring *et al.,* 2006). In maize, deep sequencing of endosperm has successfully identified 100 putatively imprinted genes (54 maternal and 46 paternal) (Waters *et al.,* 2011).

DNA methylation is also involved in determining RNA splicing sites. The *DEFICIENS* (*DEF*) gene, an oil palm orthologue of the *Arabidopsis APETALA3* gene, is essential for the formation of flower organs. The discovery of a TE, named *Karma*, which possess a signal sequence for RNA splicing has been identified as the culprit for abnormal flower development and low oil yields. This is because the methylated *Karma* produces a complete DEF protein, while unmethylated *Karma* produced a truncated DEF protein with an early stop codon that is present in the *Karma* sequence (Ong-Abdullah *et al.,* 2015). In maize, high levels of CHG DNA methylation at acceptor splice sites have been found to be less efficiently spliced, compared with low levels CHG methylation, and this inhibits alternative splicing, and production of variant proteins (Regulski *et al.,* 2013).

Transgenerational memory of plant adaptation to biotic and abiotic stresses, are also related to DNA methylation changes. A well-known example in *Arabidopsis* is the *FLOWERING LOCUS C* (*FLC*) gene, in which *FLC* expression is reduced during cold conditions (winter). The signal for vernalisation, and *FLC* levels, remain low even after the temperature increases and warmer conditions return. Differences in *FLC* expression levels might be reflected by changes in H3K27me3 levels at the *FLC* locus (Searle *et al.,* 2006). *Arabidopsis* species have been reported to exhibit ‘memory’ to previous drought conditions because plant adaptations are rapid for the second drought exposure, a phenomenon that H3K4me3 is believed to be responsible for (Ding *et al.,* 2012). Epigenetic stress memory for low humidity

conditions has also been reported in *Arabidopsis* species, as there are DNA methylation changes at various sites of basic helix–loop–helix (bHLH) transcription factor genes, such as *SPEECHLESS* (*SPCH*) and *FAMA* (Tricker *et al.,* 2013). Gene activation has been reported in *Arabidopsis* species, in response to elevated temperatures, in which demethylation of *ONSEN*, a long terminal repeat (LTR) of the retrotransposon family, favours a response to heat stress, but not sufficiently enough to cause activation. This is because it is driven by heat shock elements (HSE), which bind to plant heat shock factor to initiate transcription (Cavrak *et al.*, 2014).

1.7 Strategies to Induced DNA Methylation Pattern Changes in Plants

Reverse genetics have been widely used to study the function of epigenetic genes in plants. The majority of reverse genetic studies induce methylation changes by knocking out genes of interest by introducing a mutation, or foreign DNA insertion, to cause a frame shift mutation or gene disruption. Introduction of a mutation into the genome is relatively straightforward, however randomisation of point mutations or DNA integrations makes studying specific genes difficult. Even though, a large population of mutants are available, there are only limited ones for *Arabidopsis* species, and, for other important crop species, this remains a challenging task. Furthermore, eliminating a gene would certainly give rise to mutant phenotypes because of loss of function effects. In this section, we look into different strategies that could be used to target key methyltransferases, in order to understand gene functions and the use of genome editing technologies for methylation studies.

1.7.1 RNA Interference (RNAi)

RNA interference (RNAi) is a natural gene expression regulatory mechanism, which occurs either by the suppression of transcription (transcriptional gene silencing, TGS), or generation of sequence-specific fragments for activating the mRNA degradation process, post-transcriptional gene silencing (PTGS) (Figure 1.3). In PTGS, POLII synthesises RNA, triggering silencing mechanisms. Single stranded RNA is converted into a hairpin structure and processed by the DCL protein into 22-24 nucleotide RNAs, known as small interference RNA (siRNA). The siRNAs are used to guide the RNA-induced silencing complex (RISC) to homologous mRNA, followed by cleavage of the target mRNA, and RNA degradation (Borges and Martienssen, 2015). During TGS, RDR converts the RNA transcript into dsRNA, before the DCL protein processes it into 21-22 nucleotide siRNAs. The siRNAs are then used to initiate RNA-directed DNA methylation (RdDM) at specific loci, or chromatin remodelling, which causing transcription repression (Borges and Martienssen, 2015).

Understanding RNAi mechanisms has made it one of most important tools for studying functional genomics. Gene silencing using the RNAi approach was first applied to plants in the 1990's, using inverted repeats (IR), which have produced the highest silencing percentage of the *GUS* gene (Waterhouse *et al.,* 1998). Since then, siRNAs have been introduced into plants using various delivery mechanisms such as direct siRNA delivery (Tang *et al.,* 2006), virus infection (Anandalakshmi *et al.,* 1998), carrier peptide (Numata *et al.,* 2014), and the most widely used; transferring a transgene, harbouring hairpin DNA, by *Agrobacterium* mediated transformation (reviewed in McGinnis, (2010).

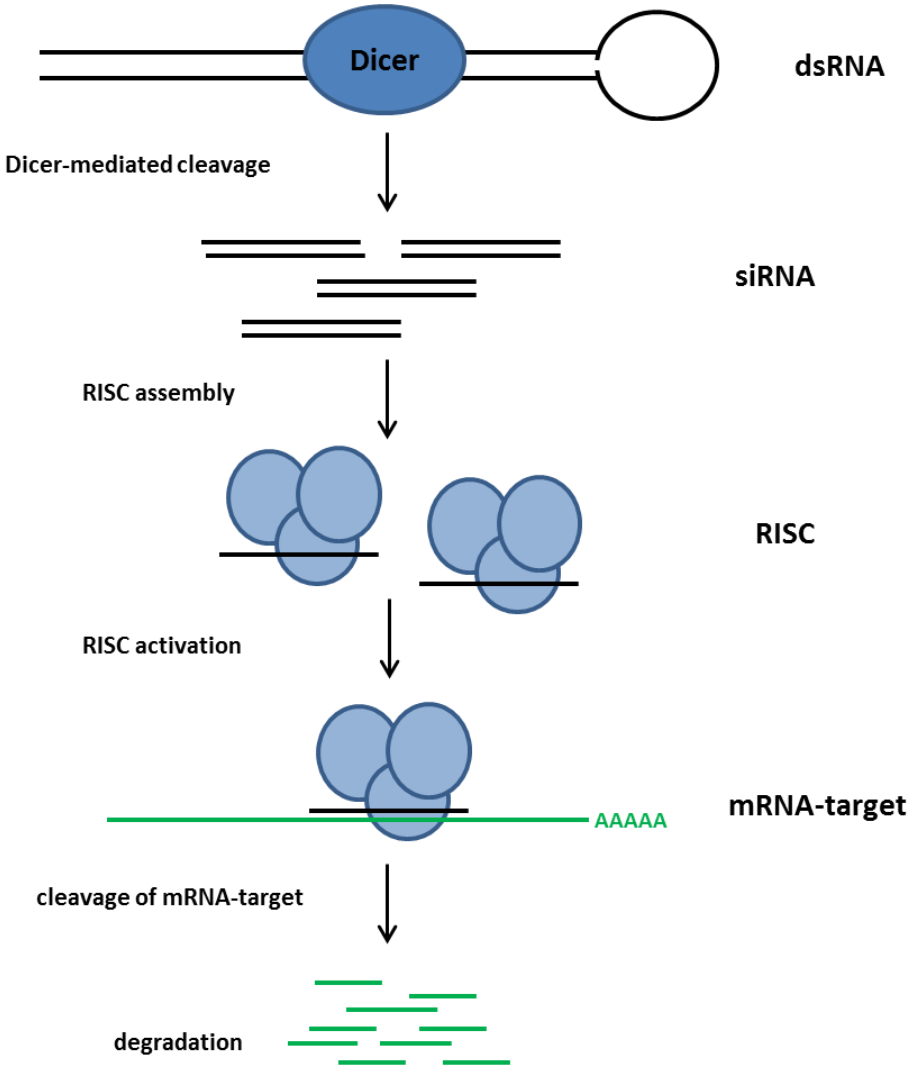


Figure 1.3: Mechanisms of RNAi gene silencing. POLII transcription produces RNA, which folds into dsRNA. dsRNA is processed by Dicer protein, producing siRNAs. siRNAs are used by the RISC complex as a guide for homologous binding with the mRNA target, and the slicer of RISC cleaves the mRNA for degradation.

RNAi technologies themselves are unique, instead of complete protein elimination; a reduction in gene expression level is obtained. Interestingly, variations in gene expression level were also observed in transgenic plants produced. This could be exploited to produce transgenic plants with optimum expression of genes of interest. Furthermore, this characteristic is important, especially for genes that are known to cause negative growth effects when the gene is knocked out or highly expressed.

Since deciphering the mechanisms involved in RNAi, they have been used extensively for crop improvement (reviewed in Angaji *et al.,* (2010) and Senthil-kumar and Mysore, (2010). RNAi has improved seed quality and germination rates in tobacco and maize (Segal *et al.,* 2004). It has helped to produce disease and virus resistant plants, such as in barley (Wang *et al.,* 2000) and banana (Rodoni *et al.,* 1999), and has improved the compound synthesis or protein content in rice (Kusaba *et al.,* 2003) and tomato (Davuluri *et al.,* 2004).

Despites enormous advantages, RNAi gene silencing suffers from a few drawbacks. One of the drawbacks is the variation in the silencing efficiency between mutant plants. Even this might be beneficial, however, selecting a mutant with the required gene expression levels could be difficult, particularly in later generations, where there are transformants that have failed to silence the target gene, even though the transgene was transcribed (Kerschen *et al.,* 2004). Furthermore, variations in gene expression level could result in variations at the phenotypic level, which could be an issue, especially for crops like soy bean (Hayashi *et al.,* 2008) and wheat (Fu *et al.,* 2007) because companies rely on the phenotypic characteristics for commercialisation. Another drawback is that the RNAi-inducing transgene itself could be silenced because it was introduced by *Agrobacterium* infection, in which the plant's defence mechanisms respond by silencing the transgene (reviewed in Matzke and Matzke, (1998).

1.7.2 Transcription Activator-Like Effector Nucleases (TALEN)

Transcriptional activator-like effectors (TALEs) are programmable DNA binding proteins, originating from phytopathogenic *Xanthomonas* spp, which cause disease in rice and cotton (Boch and Bonas, 2010). The N- and C-terminal of TALE proteins are highly conserved, consisting of a translocation signal, nuclear localisation sequence, and transcription activation domain, while the central region consists of tandem repeats that recognise specific DNA sequences (Scholze and Boch, 2011). Naturally, TALE tandem repeats are variable in number, ranging from 1.5 to 33.5, with up to 42 amino acids for each repeat, followed by a single half repeat of 20 amino acids (Boch and Bonas, 2010).

The tandem repeats have identical amino acids, except for position 12 and 13, which are known as repeat variable di-residues (RVD), that recognise different nucleotides (Figure 1.4). The most common four RVDs are HD, NG, NI, and NH, which recognise C, T, A, and G nucleotides, respectively (Yang *et al.*, 2014). As such, a single tandem repeat can recognise a single nucleotide, in which rearrangement of tandem repeat arrays makes targeting regions of interest in the genome possible. However, there are two pre-requisites of TALE proteins prior to DNA binding; the DNA binding site must begin with a thymine and the tandem repeat binding domain must always end with a half-length repeat (0.5 repeat) (Boch and Bonas, 2010).

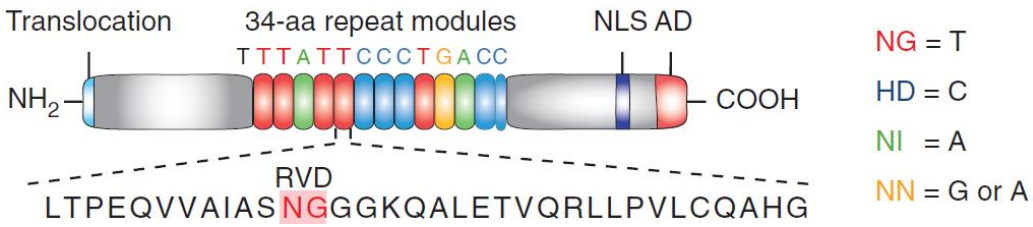


Figure 1.4: Organisation of the TALE protein structure. Tandem repeats of 34 amino acids are shown in red, green, blue, and yellow, which indicate the nucleotide binding specificity of each tandem. Amino acids at position 12 and 13 are known as RVD. The only difference between tandem repeats is highlighted in red. Adapted from Sanjana *et al.,* (2012).

The DNA binding ability of TALEs has been exploited as a tool for genome editing; the TALEs DNA binding domain (TALE DBD) can be customised by combining it with various effector domains to give the TALE DBD various abilities, such as transcription activation and repression, genome modification, and epigenetic modulation. Customisation and versatility of TALE made it the 'Method of the Year' for 2011 (NatureMethods, 2012). TALE transcription activation (TALE-TA) was produced by fusion with the VP64 domain (Sanjana *et al.,* 2012), TALE transcription repressor (TALE-TR) was achieved by fusing it with the EAR repressor domain (Mahfouz *et al.,* 2012), TALE nucleases (TALEN) were achieved by fusing TALE with the Fok1 domain (reviewed in Chen and Gao, (2013), and TALE epigenetic modulation was achieved by fusion of the TALE DBD with an epigenetic modulator domain, such as TET1 for demethylation, (TALE-TET1) (Maeder *et al.,* 2013) and H3K9 tri-methyltransferase Suv39H1 domain for TALE histone modulation enzyme (TALE-HME) (Bieberstein *et al.,* 2016). Therefore, this project focussed on TALENs to induce mutations at methyltransferases.

TALENs can facilitate genome modification by inducing double stranded breaks (DSB) at target regions. Since Fok1, the nuclease domain of TALEN, nucleases the DNA molecule as a dimer, two customised TALEN proteins are required, each designed to bind on opposite strands with inverted orientation, as shown in figure 1.5 (Sanjana *et al.*, 2012). DNA repair mechanisms are recruited to fix the DSB, either by non-homologous end-joining (NHEJ) or homologous recombination (HR). However, NHEJ repairs are error-prone mechanisms, which normally cause frame shift mutations resulting in truncated protein (Figure 1.6) (Chen and Gao, 2013).

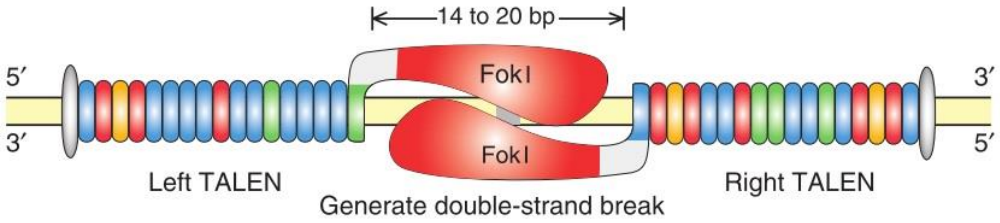


Figure 1.5: Illustration of a pair of TALENs binding to DNA. Left and right TALEN are designed 14 to 20 bp apart, to allow the Fok1 domain to dimerise and break double stand DNA between the TALENs. Adapted from Sanjana *et al.*, (2012).

TALENs have been used extensively in plants, mostly for introducing mutations into the genes of crop plants (reviewed in Chaudhary *et al.*, (2016), and have been used in rice, maize, and wheat, to produce disease or virus resistant plants (Li *et al.*, 2012; Liang *et al.*, 2014; Wang *et al.*, 2014). In barley, TALENs were used to improve the oleic acid content in the seeds (Wendt *et al.*, 2013). Heritable mutations have also been reported in tomato, using TALENs under the control of an inducible promoter (Lor *et al.*, 2014).

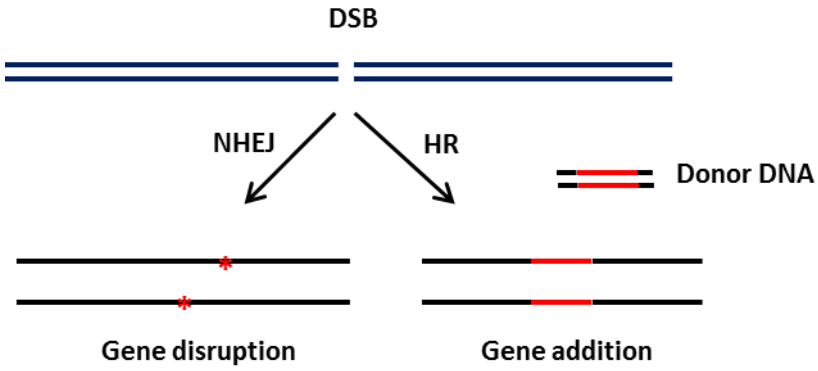


Figure 1.6: DNA repair mechanisms. A pair of TALEN-induce DSB are repaired, either by non-homologous end joining (NHEJ) or homologous recombination (HR). NHEJ repair normally causes a frame shift mutation, which then causes gene disruption. HR mechanisms require donor DNA with homologous flanking sequences for gene replacement or addition.

Although TALEs have enormous potential, there are several limitations with this technology. One of the drawbacks is binding efficiency; even though it is designed to specific TALE DBD targets, the efficiency of introducing the mutation ranges from 100% to 50% (Chen and Gao, 2013). Variations in DNA binding could be due to TALE DBD binding being sensitive to DNA methylation (Valton *et al.*, 2012). Another limitation is that multiplex mutations of several targets are difficult to achieve because a pair of TALE proteins needs to be engineered for each target. Furthermore, cloning of identical repeat sequences of TALE DBD arrays presents high technical challenges, even with 'Golden Gate' cloning strategies.

1.7.3 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

Clustered regularly interspaced short palindromic repeat (CRISPR) is a new targeted genome editing tool discovered in 2012, and is based on the adaptive immune system of bacteria and archaea. It consists of two main components, the CRISPR associated protein 9 (Cas9) from *Streptococcus pyrogenes* and the CRISPR RNA (crRNA), also known as single guided RNA (sgRNA), which contains the

complementary target sequence (Jinek *et al.,* 2012). Instead of engineering a new TALE DBD protein for each new target site, the CRISPR system utilises Watson-Crick base pairing to recognize its target. The system uses sgRNA to bring the Cas9 protein to its target. Furthermore, the Cas9 protein itself contains nuclease domains (the HNH and RuvC-like), which cleaves both strands of the dsDNA at the same time, resulting in a DSB (Jinek *et al.,* 2012). The CRISPR system is simple because sgRNA can be easily modified for a new target, in addition, multiple sgRNAs can work simultaneously with the same Cas9 protein, meaning that the CRISPR system can be multiplexed (Sander and Joung, 2014).

The CRISPR immune system relies on small DNA sequences that are acquired during previous bacteriophage or plasmid infections, and kept as CRISPR arrays. These small foreign DNA sequences are kept interspersed with repeat sequences, and incorporated into the CRISPR array (Figure 1.7) (Makarova *et al.,* 2011). Transcription of the CRISPR array produces pre-crRNA, which is then matured, prior to forming a complex with the Cas protein. There are two class of CRISPR/Cas systems, which are defined by the presence of multiple subunits of the crRNA-Cas complex (Makarova *et al.,* 2015). Class 1 consists of type I, III, and IV, which requires specialised Cas endonucleases for crRNA maturation, while class 2 consists of type II and V, which are simpler, with the crRNA being processed by RNAse III prior to forming a complex with the Cas protein (Makarova *et al.,* 2015). In type II, Cas9 requires crRNA and trans-activating CRISPR RNA (tracrRNA) to stabilise the structure for Cas9 activation (Jinek *et al.,* 2012). Most genome editing tools today, use the type II CRISPR/Cas system.

The 20 nucleotides of crRNA are responsible for recognising the binding site of Cas9 complexes. Naturally, a secondary structure formed by a crRNA:tracrRNA duplex is required for Cas9-crRNA complex stability. In addition, Cas9 requires a protospacer adjacent motif (PAM), an 'NGG' sequence, at the target sequence, which is located next to the 20 nucleotide binding site of sgRNA, that acts as a signal for DNA duplex unwinding and for efficient DNA Cas9-catalyzed cleavage (Figure 1.8) (Jinek *et al.,* 2012). Re-engineering of crRNA:tracrRNA duplexed to a sgRNA has made the system much simpler to use as a genome editing tool (Jinek et al., 2012).

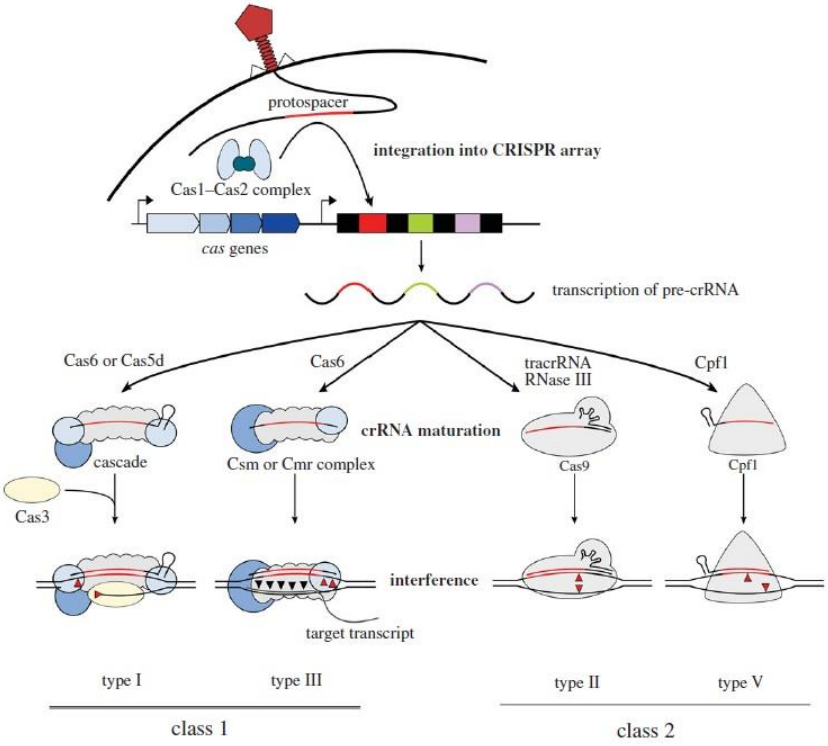


Figure 1.7: Classes and mechanisms of the CRISPR adaptive immune system. The CRISPR adaptive immune system acquires DNA fragments from previous bacteriophage and virus infections, which are kept in a CRISPR array of the cas operon. Transcription of the CRISPR array produces pre-crRNA, which is matured by further processing by the Cas protein in class 1, or RNAse III in class 2. In class I, crRNA requires multiple crRNA-effector subunits prior to acting as a guided endonuclease, while for class 2, crRNA forms a complex with Cas9, prior to acting as a defence mechanism. Adapted from (Hille and Charpentier, 2016)

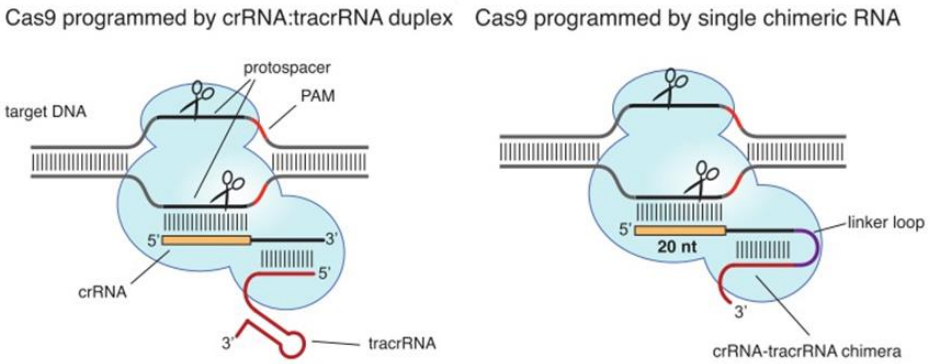


Figure 1.8: Comparison of crRNA:tracrRNA-Cas9 and sgRNA-Cas9 complex. Left. The tracrRNA complementary to the repeat sequence of crRNA forms a tracrRNA:crRNA duplex. Right. A linker loop is introduced to enable transcription of sgRNA in a single RNA molecule. PAM (in red) is essential, and located before the crRNA binding site. Adapted from (Jinek *et al.*, 2012).

The applications of CRISPR are very similar to TALE, but the CRISPR system is much simpler and quicker to modify when for new targets are required. The nuclease activity of the Cas9 protein is responsible for introducing DSB, whereas TALEN require fusion with a nuclease domain. Similar to TALEN, the CRISPR/Cas9 system can be exploited to induce gene knockouts or truncated proteins, which are introduced into the genome by NHEJ, with the aim of introducing new, or modifying current genes via HR. Chromosomal rearrangement, or large chromosomal deletion achieved by introducing two sgRNAs simultaneously. CRISPR can also function beyond the targeted genome editor by deactivating the nuclease activity of the Cas9 protein (dCas9) (Figure 1.9). The dCas9 protein can be engineered to fuse with functional domains to function as a transcription activator or repressor, epigenetic modifier, or DNA labeller (Heidenreich and Zhang, 2015). There are various functional domains that can be used with dCas9 for genome modification effects, which are listed by Laufer and Singh, (2015).

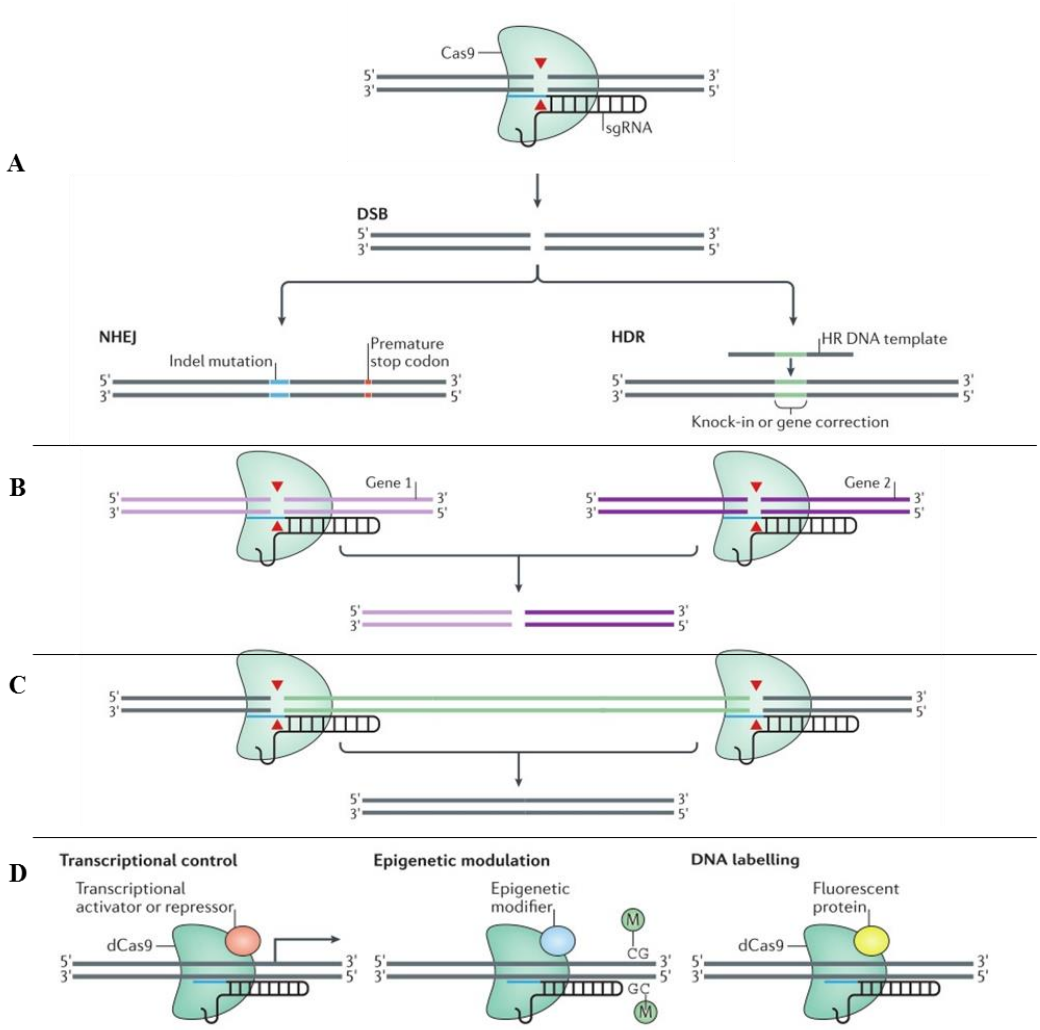


Figure 1.9: Application of CRISR as a genome editor. A. Precise gene editing. The CRISPR/Cas9 system introduces DBS, which are repaired by error-prone mechanisms, the NHEJ, which results in truncated genes, or homology-directed repair (HDR or HR) for gene insertion or modifier. **B**. Chromosomal rearrangement. Two different sgRNAs are introduced for cleavage at two different genes, which results in chromosomal rearrangement. **C**. Large chromosomal deletion. Two almost identical sgRNAs are introduced to induce DNA cleavage at different loci of the same gene, for large DNA deletion. **D**. Inactivated nucleases of Cas9 (dCas9) can be engineered to fuse with functional domains for transcriptional control, epigenetic modulation, or DNA labelling. Adapted from Heidenreich and Zhang, (2015).

Even though CRISPR technologies are relatively new, vast numbers of crop genomes have been successfully edited using the technique, such as rice, wheat, sorghum, maize, and tomato (Shan *et al.*, 2013; Jiang *et al.*, 2013; Feng *et al.*, 2016; Pan *et al.*, 2016; reviewed in Sayre *et al.*, 2016). While most research on CRISPR-edited crop plants is still at the beginning, several beneficial varieties of plants have

been reported. CRISPR has produced plant resistance to phytopathogen for example, wheat has shown an increased resistance to powdery mildew, which was achieved by simultaneously mutating three genes encoding for the MILDEW-RESISTANCE LOCUS (MLO) protein (Wang *et al.*, 2014).

Research on tobacco has found that CRISPR can confer DNA virus resistance from three different germinivirus, beet severe curly top virus, bean yellow dwarf virus, and tomato yellow leaf curly virus (Ali *et al.*, 2015; Baltes *et al.*, 2015; Ji *et al.*, 2015). A new corn variety has also been reported; the waxy corn, which the endogenous waxy gene, *WX1*, was knocked out, producing corn that only contains a single type of starch, the polysaccharide amylopectin. This has a high commercial value due to it being a main component of processed foods, adhesives and, high-gloss paper (Waltz, 2016). Mushrooms with anti-browning characteristics and longer shelf life have also been produced by knocking out one of the six genes that encode polyphenol oxidase (Waltz, 2016). Within just five years of deciphering the CRISPR mechanisms, we have successfully improved several crops, and many more beneficial plant varieties will be produced over the next few years.

As with other technologies, CRISPR technologies suffer from several drawbacks. CRISPR has a high off-target binding rate because even with several mismatches, it is still able to cleave the target DNA (Fu *et al.*, 2013). However, several improvements have been made, using truncated sgRNA (Fu *et al.*, 2014) or Cas9 nickases, which only have one active nuclease domain (Cho *et al.*, 2014). Cas9 also has its limitations because it requires a PAM motif. Even though the NGG sequence it not rare, it still causes problems, especially when precision editing is required. Luckily, this problem has been addressed by re-engineering the Cas9 protein with

altered PAM specificities (Kleinstiver *et al.,* 2015). The size of the large Cas9-sgRNA complex presents a problem for efficient delivery into human cells, however a smaller version of the Cas9 enzyme has been found in *Staphylococcus aureus*, but detailed background information about it is still limited (Ran *et al.,* 2015).

1.7.4 Gene Over-Expression

Gene over-expression, also known as gain-of-function, is an alternative to the loss-of-function approach to studying gene function and mechanisms. The idea of over-expression studies arose from Mongolism or Down's syndrome, a human genetic disease caused by the presence of an extra chromosome (trisomy), and indicates that appropriate gene dosage is required for normal gene function (Prelich, 2012). In plants, the first gene over-expression studies reported, involved over-expression of the *CYTOKININ INDEPENDENT 1* (*CK1*) gene in *Arabidopsis*, which is involved in cytokinin signal transduction (Kakimoto, 1996). Since then, the gain-of-function approach has been used widely by the plant research community in various plant species (reviewed in Zhang, 2003).

There are three methods used for over-expression studies in plants, activation tagging, cDNA over-expression system, and over-expression of open reading frame (ORF). Activation tagging involves a random insertion of transfer DNA (T-DNA) that contains the tetramer of a strong promoter, 35S CaMV, at the T-DNA right border, which results in expression enhancement of a gene near the T-DNA insertion (Kondou *et al.,* 2010). The elevated expression genes can be identified using PCR (Spertini *et al.,* 1999). While for cDNA over-expression, a full-length cDNA (fl-cDNA) is inserted by T-DNA, under the control of the 35S CaMV promoter

(Kondou *et al.*, 2010). The fl-cDNAs used are derived from a cDNA library, which can be focus-specific, by generating a cDNA library from stress plants. This method also eliminates the need of genome sequencing, prior to developing the construct. The third method uses a T-DNA insertion that carries an ORF gene under the 35S CaMV promoter (Kondou *et al.*, 2010). ORFs and fl-cDNAs differ, in that, ORFs do not include the 5' and 3' untranslated region.

Over-expression offers an alternative way of studying gene function, as the loss-of-function approach absolutely, but not conclusively produces mutant phenotypes (Prelich, 2012). Furthermore, many plant genes have redundancy when it comes to their function, as such a single-gene elimination might not be possible to understand gene function, whereas with over-expression, the characterisation of functionally redundant genes is possible (Kondou *et al.*, 2010). Over-expression is also an alternative, particularly for genes with deleterious effects in knockout approaches (Stevenson *et al.*, 2001). Some genes have functions that only show when the gene is over-expressed (Pontier *et al.*, 2001; Van der Graaff *et al.*, 2002). Furthermore, there are also genes that produce phenotypes that are different from the knockout mutant phenotype (Wada *et al.*, 1997). Given these examples, it is clear that over-expression is a powerful tool that differs from knockout or silencing methods. There are several ways that gene over-expression can cause endogenous gene expression and phenotypic changes. For example, over-expressed protein can cause either inhibition or activation of other proteins, changes in protein complexes, or activate different molecular pathways (reviewed by Prelich, (2012).

Gene over-expression has successfully helped to understand gene function and produce mutants with commercial value phenotypes. Several gene functions have

been discovered from over-expression strategies, such as FLAVIN MONO-OXYGENASE (FMO) protein, which was found to be involved in auxin biosynthesis (Woodward *et al.*, 2005), and ENHANCER OF SHOOT REGENERATION 1 (ESR1) protein, which is involved in shoot differentiation (Banno *et al.*, 2001). Some plants have been produced with a mutant phenotype allowing tolerance to different stresses. For example, Chinese Kale has enhanced drought tolerance and osmatic stress, and was produced by over-expressing the *Arabidopsis* HOMEODOMAIN GLABROUS11 (HDG11) protein (Zhu *et al.*, 2016). A rice heat stress transcription factor, OsHsfA2e, was also found by over-expression in *Arabidopsis*, in which the mutant *Arabidopsis* also showed high tolerance to salinity stress (Yokotani *et al.*, 2008).

Gene over-expression does have its limitations. ORF approaches, for example require complete genome sequence information prior to building a construct. This is difficult because the genomes of some crops have not been sequenced yet. However, this issue might be improving as more crops enter the whole genome sequencing project. The identification of genes that are responsible for mutant phenotypes is also difficult, especially with activation tagging method, as the constitutive enhancer is reported to be capable of increasing the expression of genes up to 8 kb away from the insertion site (Ichikawa *et al.*, 2003). Another possible issue is the high level of gene expression by constitutive promoters in the wrong tissues, which could lead to inaccurate conclusions about the protein function. This is because mis-expression in tissues could lead to changes in endogenous gene expression, which could be the reason for mutant phenotypes (Kondou *et al.*, 2010).

1.8 Thesis Objectives

The aims of this study are to induce DNA methylation pattern changes in plants by exploiting genetic engineering tools to interfere with DNA methylation pathways. Four strategies have been developed and tested for the induction of DNA methylation changes in plants. The strategies are gene silencing, gene knockout, target demethylation, and gene overexpression.

1. An siRNA will be used as a tool for the post-transcriptional silencing of the *MET1* gene in maize. The inverted repeat of *MET1*, is introduced into wild-type maize to lower the *MET1* expression level. The reduction of *MET1* should disturb its efficiency, and lead to global DNA methylation pattern changes, that could affect maize growth and development. Heritable methylation patterns and gene expression changes will be analysed in transformants, and plants with a transgene that has been segregated away. In parallel, the characteristic of two copies of the maize *MET1* gene will be investigated.
2. The targeted genome editing technologies, CRISPR and TALEN, will be employed to truncate the MET1 protein of tomato. By knocking-out the *MET1* gene, we hope to produce a *MET1* mutant with a hypomethylated genome, and develop a plant with distinct phenotypes.
3. The target demethylator will be developed by exploit the CRISPR, to demethylates a dense methylated region of *Arabidopsis*, which was found using the epigenome database. *Arabidopsis* was selected as the proof-of-concept for this tool, due to the availability of the epigenome data and the ease with which it can be transformed.

4. And the final strategy is to interfere with the DNA methylation pathway via over-expression of *CMT2*, another important DNA MTase. At the same time, *Arabidopsis* were also transformed with the over-expression of a catalytically inactive *CMT2*, for titrating the effector effects of over-expression. Phenotypes arising from the mutant were observed to narrow down possible targets that had altered DNA methylation status or expression.

Chapter 2

Production and Characterisation of Maize *ZmMET1* RNAi Transformants

2.1 Introduction

Maize is one of the most commercially important crops in the world, with around 177 million hectares of maize planted worldwide (James, 2013). Thus, generating a new variety of maize with high tolerance to an ever-changing environment is crucial for maintaining and enhancing maize production. Traditional plant breeding strategies involve the cross pollination of the desired plant's traits. However, breeding has several limitations, for example, it can only be carried out between compatible plants that are a closely related species or genus, and the hybridised plant expressing the desired traits may also introduce unwanted traits (Rommens *et al.*, 2007). Furthermore, breeding can take up to decades to produce a new variety when woody horticulture crops are involved. This requires an enormous amount of labour and resources, despite the development of marker-assisted selection to accelerate breeding programs (Van Nocker and Gardiner, 2014).

Advances in genetic engineering have helped to overcome the limitations of conventional breeding. Genetic engineering involves the introduction of an expression-cassette; consisting of a promoter and the gene encoding the required phenotype, into plants via bacterial transfer DNAs, known as T-DNA. This process makes stable phenotypic changes that bring about the desirable traits possible. This option has overcome many species barriers and enabled the elimination of breeding unwanted traits (Rommens *et al.*, 2007). Although genetic engineering has a bright future in crop breeding programs, it also raises several issues, such as the

introduction of new classes of allergens that were not previously recognised as being harmful, potential safety risks to the environment, and complex regulatory processes that must be passed in order to classify them as genetically modified foods (GMFs) (Maghari and Ardekani, 2011).

There are also variations produced by the stable changes in gene expression that occur without altering the DNA sequence. These types of changes are called epi-mutations, and involve a modification of the epi-genome. One common plant epi-genome modification is DNA methylation, which regulates genes by transcription silencing (He *et al.,* 2011a). The *FWA* gene in *Arabidopsis* is an example of a natural epi-mutation that causes delays in flowering time (Soppe *et al.,* 2000). The *FWA* gene is normally silenced in mature wild type plants due to the methylation of a repeat sequence in the 5' region of the gene. When the methylation is removed, such as in a *met1* mutated *Arabidopsis* plant, the *FWA* gene is expressed. However, expression of *FWA* remains even after restoring the wild type gene to the plant (Kankel *et al.,* 2003).

Induction of epi-genome changes using a genetic-engineered approach could be beneficial, especially to commercial crops. Maize for example, is an interesting species to study the epi-genome of, since up to 85% of it's genome is comprised of transposable elements (TEs) (Haberer *et al.,* 2005; Vicient, 2010), and DNA methylation of its genome ranges from 65-85% (Gent *et al.,* 2013). Maize TEs are frequently distributed near genes, so they could potentially impact the nearby genes. Therefore, controlling the expression of TEs is necessary for maize. One of the methods used to silence TEs is by DNA methylation. Maize has been shown to have a high dependency on DNA methylation, these is especially true for tissue

differentiation in its leaves (Candaele *et al.*, 2014). Interestingly, maize possesses duplicates of all its functional methyltransferase (MTase) genes (*MET*, *CMT*, and *DRM*), and they all share high levels of homology (Candaele *et al.*, 2014). Duplicate MTase genes in the maize genome are believed to compensate for expansion of the maize genome and its large proportion of repetitive elements (Baucom *et al.*, 2009).

Much research has been carried out on maize DNA methylation, but has mainly focused on natural or stress related changes. Inducing stable methylation and expression changes by interfering with epigenetic mechanisms in maize, could be an interesting strategy for creating new variations to help maize breeders. Therefore, the objective of this chapter is to assess the importance of *DNA METHYLTRANSFERASE 1* (*ZmMET1*) genes in plant development and the methylation states of *ZmMET1* deficient plants.

2.2 Results

2.2.1 Genomic Organisation of *ZmMET1*

Most plants only have a single copy of the *MET1* gene, but maize possesses two copies. So in order, to obtain a better understanding of cytosine methylations in maize, we investigated the differences between these two genes by analysing their DNA sequences. The maize *MET1* genes, known as *ZmMET1a* and *ZmMET1b* (Candaele *et al.*, 2014) and with the gene identifiers: GRMZM2G333916 and GRMZM2G334041, respectively, are located inversely oriented on chromosome 7 and 12,312bp apart from each other.

The length of the nucleotide sequences of *ZmMET1a* and *ZmMET1b* are 5,981bp and 5,595bp, respectively (from start to stop codon), and both have 11 exons. *ZmMET1a* is longer because there is an insertion of a class II TE between exon1 and exon2 (Figure 2.1). Alignment of the promoter region for both genes has shown 81% similarity. Further analysis of both promoters using PlantPAN2.0 has shown the presence of common promoter motifs (-10 motif and -35 motif), indicating that both genes should have no problem being expressed.

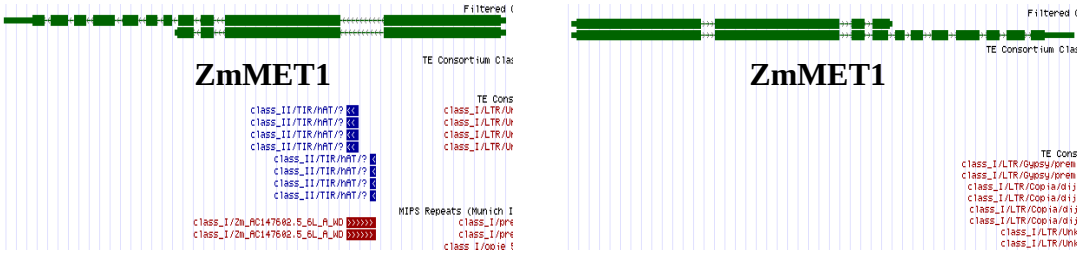


Figure 2.1: Screenshots from Maize database of *MET1* genes. The figure shows both of *ZmMET1* genes consist of 11 exons, but *ZmMET1a* has an insertion of a class II TE (blue) between exon1 and exon2. Both genes can produce two *MET1* variants (indicate by two green lines).

2.2.2 Amino Acid Organisation of the ZmMET1 Protein

The *ZmMET1a* and *ZmMET1b* genes contain an open reading frame (ORF), coding for a protein of 1,536 amino acids with similarity of 99.4%. Both genes can produce two splice variants, with one of them being shorter and a truncated protein (Figure 2.1). The ZmMET1 protein can be divided into two regions: the N-terminus that contains domains required for target identification, and the C-terminus that contains a conserved motif sequence, which is the location of the catalytic domain. The N- and C-terminus of the ZmMET1 protein is linked by a glycine–lysine repeat region, which is highly conserved in DNA MTases (Finnegan and Kovac, 2000). Pfam analysis revealed that both ZmMET1 variants consist of two replication focus

targeting sequence domains (RFD-1, RFD-2), two bromo adjacent domains (BAH-1 and BAH-2), and a methyltransferase domain (Figure 2.2). This suggests that both ZmMET1a and ZmMET1b act as maintenance methyltransferases in maize, as this kind of protein domain architecture has been observed in most of the DNA MTases found in plants (Finnegan and Kovac, 2000).

In brief, the replicating foci domain is involved in binding to replication foci through interaction with VARIANT IN METHYLATION 1 (VIM1), and in recognising hemi-methylated cytosine (Johnson *et al.,* 2007; Woo *et al.,* 2007). The bromo adjacent domain is important for the protein–protein interaction module (Callebaut *et al.,* 1999). The methyltransferase domain contains conserved motifs that are usually found in *MET1* genes. The conserved motifs of ZmMET1 protein was identify by alignment to *Arabidopsis* MET1 protein sequences. There are eight conserved motifs marked by underlining in Figure 2.2. Kumar *et al.,* (1994) have briefly described, the conserved catalytic motifs I and X that are involved in the S-AdoMET binding. Motif IV is the active site of all known cytosine methyltransferases, while motif VI plays an important role in binding to target cytosine. Motif VIII is involved in neutralisation of the DNA strand, and motif IX is required for the organisation of the target recognition domain (TRD). The variable sequences between motifs VIII and IX were proposed to be involved in defining the sequence and base specificity for methylations.

Protein alignment shows there are nine amino acid differences (eight located at the N-terminus, S33P, S59A, R152K, I434V, T671I, I760M, G790D, R921S, and one located at the C-terminus, I1341V) (Figure 2.2). We further investigated the amino acid differences, to determine whether changes in the amino acids could cause

structural or conformational differences between the ZmMET1 proteins, which may also affect their catalytic activities. In general, amino acid mutations can be divided into two categories 1) conservative mutations and 2) non–conservative mutations.

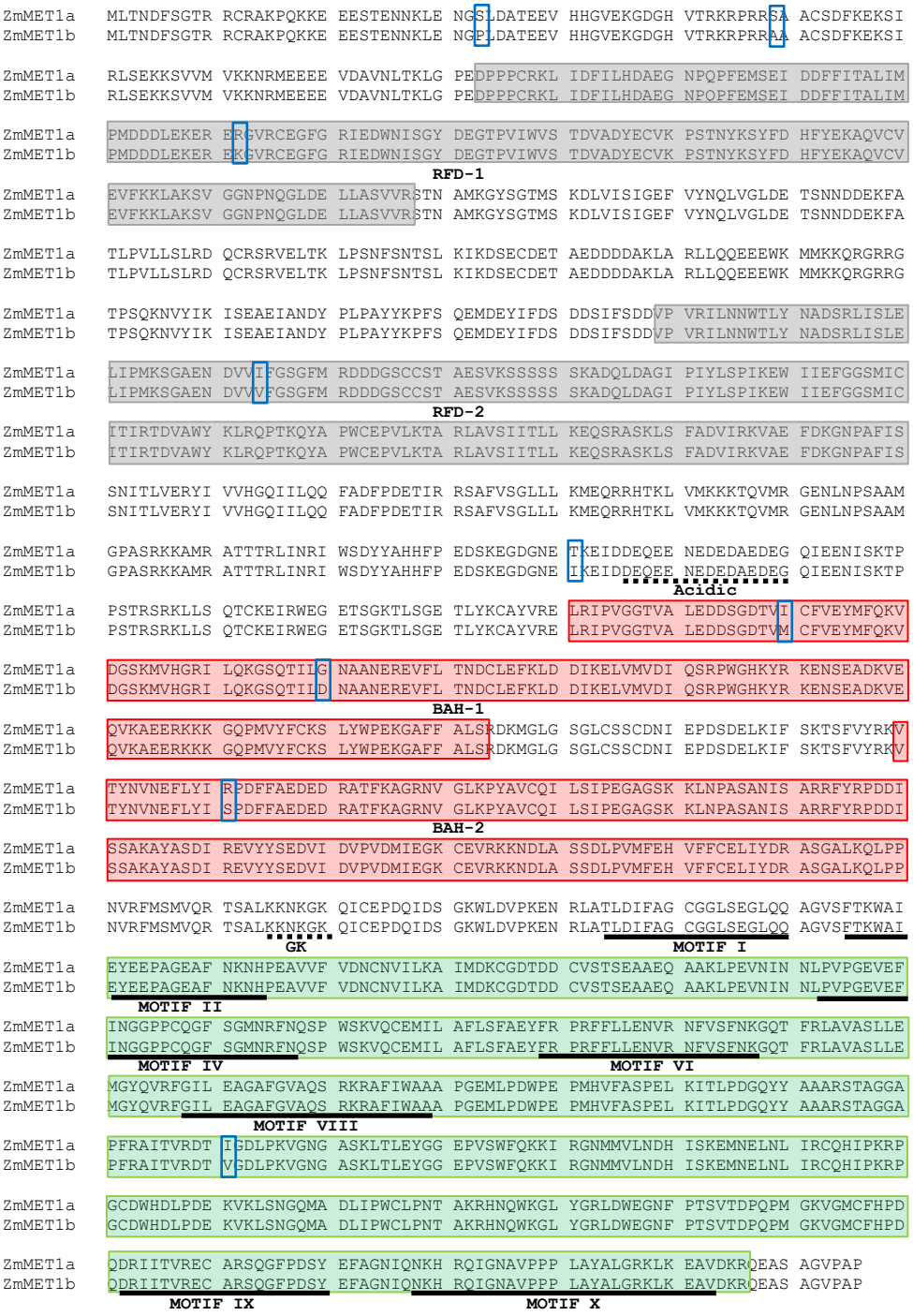


Figure 2.2: The amino acid sequence alignment of ZmMET1a and ZmMET1b. Grey box, red box, blue box and green box indicates replicating focus targeting sequence domain (RFTS), bromo adjacent domain (BAH), amino acid differences and methyltransferase domain, respectively. Bold lines indicate the conserved methyltransferase catalytic motifs I, II, IV, and VI-X. Dotted lines indicate acidic region and glycine-lysine residue region.

A conservative mutation is the change from one amino acid to another amino acid with similar biochemical properties, whereas a non–conservative mutation is the change from one amino acid to another with different properties. The amino acid differences between ZmMET1a and ZmMET1b are listed in Table 2.1. From nine amino acid differences, three were categorised as being non–conservative and the rest were conservative mutations.

Location	ZmMET1a → ZmMET1b	Type of changes
33	S → P	Conservative
59	S → A	Conservative
152	R → K	Conservative
434	I → V	Conservative
671	T → I	Non – conservative
760	I → M	Conservative
790	G → D	Non – conservative
921	R → S	Non – conservative
1341	I → V	Conservative

Table 2.1: Type of amino acid changes between ZmMET1a and ZmMET1b. The amino acids were divided into two types, conservative and non – conservative. Types of amino acid were based on biochemical properties of the amino acids either similar or different. Location, the amino acid number in the genes, which differ between ZmMET1a and ZmMET1b; second column is the amino acid differences.

The amino acid differences between ZmMET1a and ZmMET1b were further analysed by comparing them to other MET1 plants to look for conserved amino acids at the loci of the changes (Figure 2.3). The aligned amino acid sequence of plant MET1 was analysed by using Seq2Logo. Figure 2.3 only shows amino acid changes that were located at the domains. Most of the amino acid changes were conserved among MET1 plants, except for R152K, G790D, and I1341V. The amino acid changes at these three locations are different when compared to other MET1

plants, indicating that the methylation mechanisms in one of the ZmMET1 genes might have a special function, compared to other MET1 plants, or perhaps the changes are random events, specific for maintaining maize genome integrity.

In general, amino acid changes are known to affect protein structure, especially when they involve non–conservative changes. Since it is possible that the structure of ZmMET1a is different to that of ZmMET1b, we built the ZmMET1 protein structure based on the amino acid sequences for both genes. The protein structures were predicted by using I-TASSER, and were then visualised and analysed using YASARA (http://www.yasara.org). Based on the predicted protein structure, there are structural differences between ZmMET1a and ZmMET1b, and these differences occur in the N-terminal region (Figure 2.4). These were expected, as most of amino acid differences were located at the N-terminal region of the proteins. However, a significant structural change was not expected, as the majority of the amino acid differences are conservative mutations. However, these differences do not show that either one of the proteins was inactive, as both proteins have a conserved structure at the C-terminal, where the catalytic domain is located. Perhaps, further protein study on protein activity and site-directed mutagenesis would be appropriate to understand the differences between ZmMETa and ZmMET1b, but in this study, we wanted to focus on the importance of ZmMET1 in plant development.

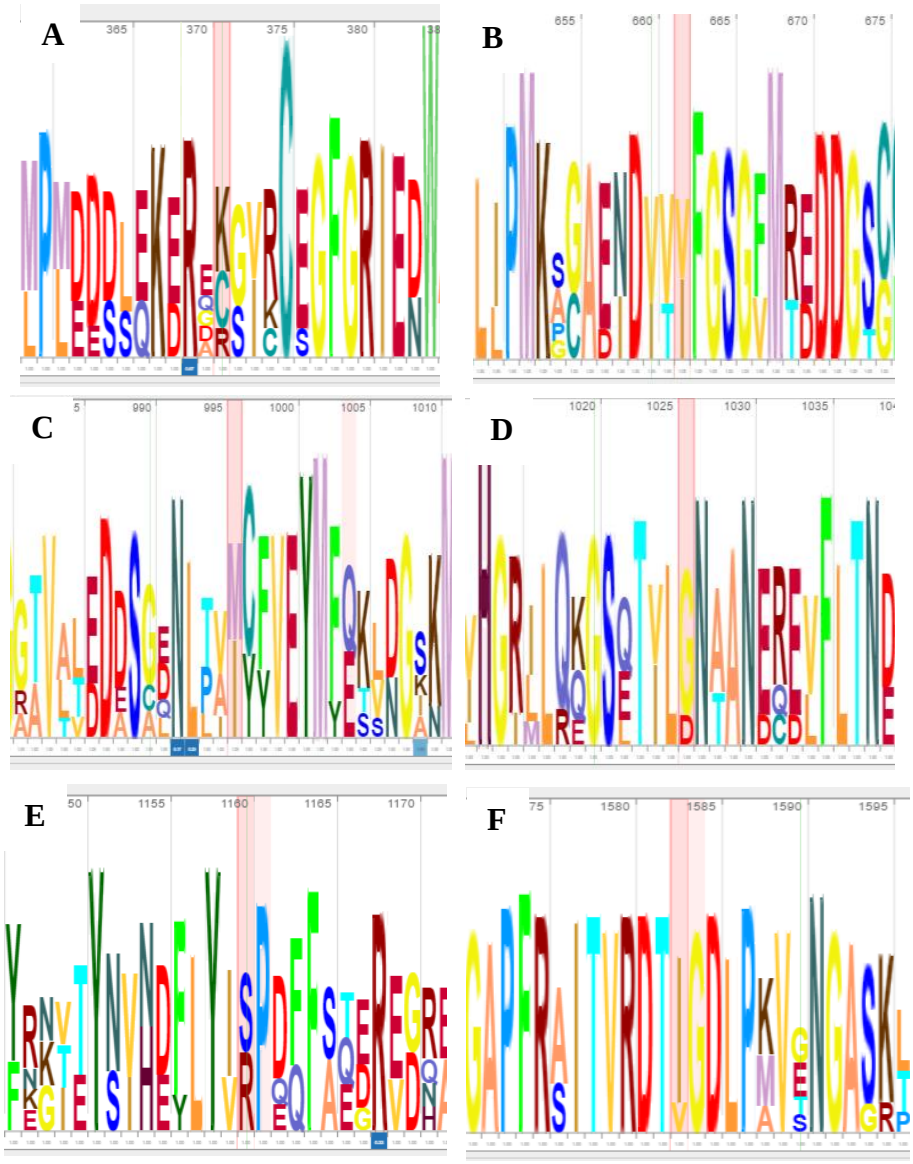


Figure 2.3: Conserved amino acids at domains of ZmMET1 that having amino acids difference. Only amino acid changes that located at domains of ZmMET1 were analysed using Seq2Logo v2.0. The conserved amino acids these locations were identified by comparing ZmMET1 with other plants MET1. A, the amino acid at R152K; B, the amino acid at I434V; C, the amino acid at I760M; D, the amino acid at G790D; E, the amino acid at R921S; F, the amino acid at I1341V.

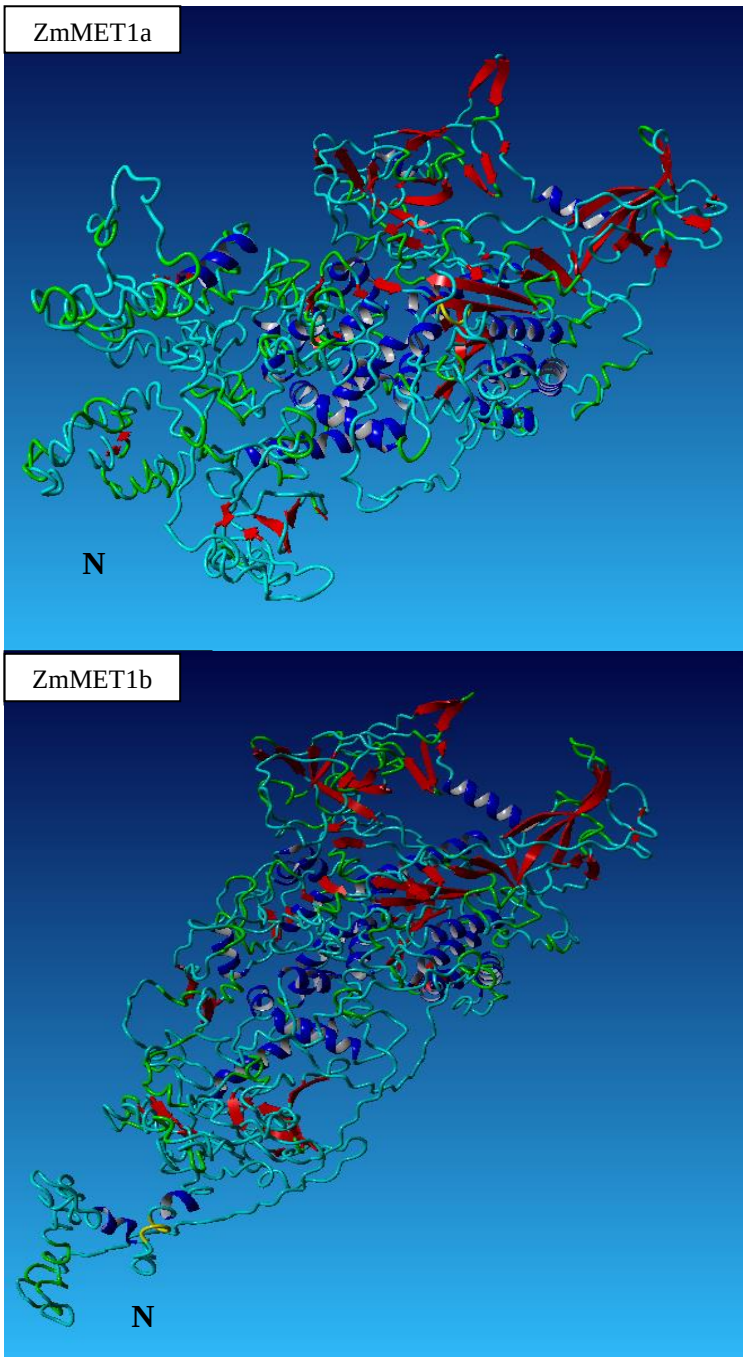


Figure 2.4: Predicted 3D structure of ZmMET1a and ZmMET1b. The 3D structure of ZmMET1a and ZmMET1b was predicted using amino acid sequence with help of iTASSER and was viewed using YASARA. The N terminals (label as N) of both proteins were folded differently showing a different N terminal structure between these two proteins.

2.2.3 Relationship of ZmMET1 With Other Plant DNA Methyltransferases

Phylogenetic trees of *MET1* genes were constructed using MEGA7 to understand the relationship of ZmMET1 protein with other plant methyltransferases. Trees were built using plant MET1 protein sequences downloaded from NCBI (Figure 2.5). The ZmMET1a and ZmMET1b were grouped together as expected because of high sequence similarity. The ZmMET1 proteins were closely related to rice MET1 (OsMET1a and OsMET1b), as they belong to the same clade. This result was expected, as both maize and rice are monocotyledon plants. Interestingly, both ZmMET1a and ZmMET1b were more related to OsMET1b compared to OsMET1a.

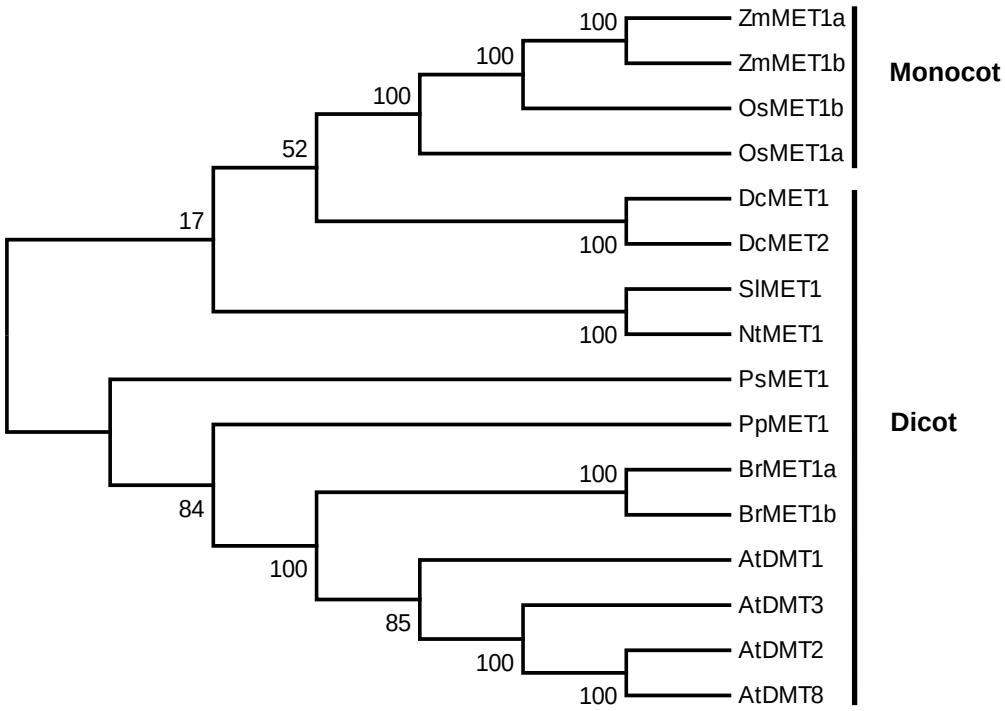


Figure 2.5: Phylogenetic relationship among MET1 genes of maize, rice, carrot, tomato, tobacco, pea, peach, *Brassica rapa* and *Arabidopsis*. Unrooted neighbour-joining (NJ) phylogenetic tree of MET1 proteins with bootstrap value were shown for each clade. The ZmMET1 were in the same grouped with rice OsMET1. Both ZmMET1a and ZmMET1b were more related to OsMET1b. MET1 protein sequences were downloaded from NCBI. Accession numbers and abbreviations are as follows: *Arabidopsis:* AtDMT1 (At5g49160), AtDMT2 (At4g14140), AtDMT3 (At4g13610), AtDMT8 (At4g08990), *Brassica rapa*: BrMET1a (BAF34635), BrMET1b (BAF34636), peach: PpMET1 (AAM96952), pea: PsMET1 (AAC49931), tobacco: NtMET1 (BAF36443), tomato: SlMET1 (CAA05207), carrot: DcMET1 (AAC39355), DcMET2 (AAC39356), rice: OsMET1a (Os03g58400), OsMET1b (Os07g08500), maize: ZmMET1a (GRMZM2G333916) and ZmMET1b (GRMZM2G334041).

2.2.4 The Expression of *ZmMET1a* and *ZmMET1b* Genes

The expression levels of *ZmMET1a* and *ZmMET1b* were evaluated using semi-quantitative reverse transcriptase PCR (smqRT-PCR). To permit discrimination expressions of these two highly similar genes, primers were designed based on the Prim-SNPing technique (Chang *et al.,* 2009) with minor modifications. In order to target the mutation, as described by the Prim-SNPing method, both forward and reserve primers were designed to bind specifically at *ZmMET1a* or *ZmMET1b* by including the base differences at the 3’ end of both primers, instead of only including the base differences at the 3’ end of one of the primers. Two pairs of primers were used to investigate the *ZmMET1a* and *ZmMET1b* gene expression (I and II, respectively), and another pair of primers was used to amplify both genes (III) (Figure 2.6B). The primers used are detailed in Section 7.1.6 of Materials and Methods.

Besides the primer specificity, we also considered looking at the developmental stages of the tissues in order to analyse the *ZmMET1* gene. This is because methylation patterns are known to be different in tissues at different developmental stages, and these reflect the different expression patterns of MTases. This has been

observed in maize leaves, as methylation patterns and expression of MTases varied at different developmental stages (Candaele *et al.,* 2014). This is important for maize leaves because all processes of growth are represented in a single leaf. At the base of the leaf, cells are dividing, while at the more distally located elongation zone, and when the cells stop expanding, they become part of the mature zone (Nelissen *et al.,* 2013).

Thus, to investigate the expression of *ZmMET1a* and *ZmMET1b*, tissue variations were also considered for sampling, with different growth tissues being used for the expression analysis. Three different growth development tissues were used in this study, consisting of tissues from the dividing, elongating, and mature zones (Figure 2.6A). The fifth leaves (V5 growth stage) was used for the expression analysis were from wild type A188 since V5 represent transition between tillering and stem extension development. This leaf was selected since it represent different cell type, the embryonic and non-embryonic cells. Figure 2.6C shows that *ZmMET1a* transcripts were more abundant than *ZmMET1b* in all growth tissue types. However, there were no significant differences between *ZmMET1a* and *ZmMET1b* transcript levels in any of the tissue types. This may have been because the differences in transcript level were too small to be detected by the smqRT-PCR approach.

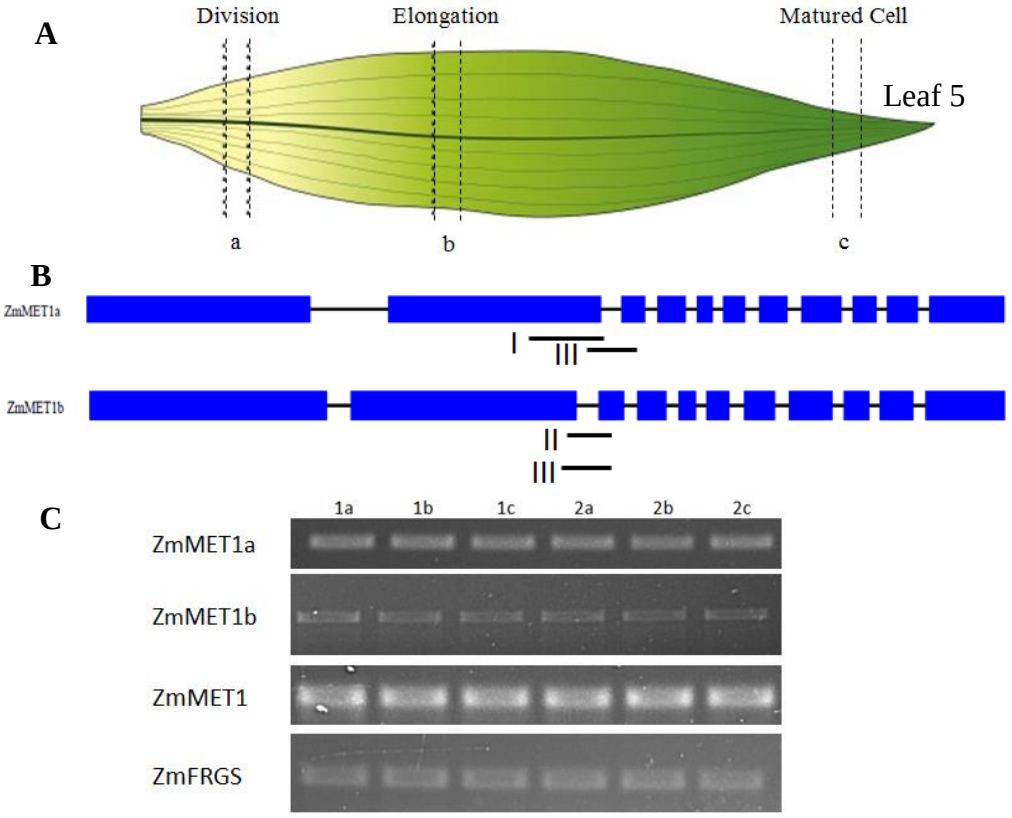


Figure 2.6: The expression of *ZmMET1* at different leaf development. A, A schematic diagram of maize leaf indicating regions used for *ZmMET1* expression analysis. Three zones were subjected to RNA extraction, a; division zone, b; elongation zone, c; matured cells. Leaf no 5 from plants at V5 growth stages was used in this study. **B**, Black lines indicate regions targeted by primers for *ZmMET1* expression analysis. I; primers designed to target this region which only specific to *ZmMET1a*, II; primers designed to target this region which only specific to *ZmMET1b*, III, primers designed to target this region which amplified both *ZmMET1a* and *ZmMET1b*. **C**, Expression of *ZmMET1* from different tissues using smqRT-PCR method. 1 and 2 are replicates which derived from different plant; a, division zone; b, elongation zone; c, matured cells; *ZmFRGS*, housekeeping genes; *ZmMET1*, expression of both ZmMET1a and ZmMET1b.

2.2.5 Knockdown of *ZmMET1* in Maize Using an Inverted Repeat, Driven by 35S CaMV Promoter

To investigate the importance of *ZmMET1* in maize, an inverted repeat construct approach was chosen because a knockout line was not available, and it is difficult to target one for *ZmMET1* due to the high sequence similarity. The backbone vectors used in these experiments were pAM, p7U and pABM (Figure 2.7 (A), (B), (C)). The 35S CaMV promoter and NOS terminator were amplified from another plasmid,

and inserted into pAM. Then, an 840bp region encoding *ZmMET1* was amplified from maize DNA using a different pair of primers, and inserted in the opposite direction between the 35S promoter and terminator. This produced sense and antisense sequences, separated by an intron, and controlled by the 35S promoter. These sequences were used to induce stem-loop structure formation, to favour the production of single stranded RNA fragments. The inverted repeat cassette, which consisted of the 35S promoter, *MET1* sense and antisense sequences separated by an intron, and an NOS terminator, was sub-cloned into p7U, a maize transformation vector. This produced a construct named, p7UZMET1ir (Figure 2.7(D)).

Besides an inverted repeat of *ZmMET1,* driven by the 35S promoter, a construct driven by an alcohol inducible promoter was made because the 35S *ZmMET1* inverted repeat might have been lethal to the transformants, as seen in a previous study by Dr Michael Watson, when transforming a *tomMET1* sense and antisense construct, driven by an 35S promoter into tomato. The inverted repeat cascade was sub-cloned into pABM, which harbours the UAS promoter, an alcohol inducible promoter. The inverted repeat cassette, which consisted of the UAS promoter, *MET1* sense and antisense sequences, NOS terminator, and GAL4 gene, was sub-cloned into p7U. This produced an inducible construct named, p7UUASZMET1ir (Figure 2.7 (E)). The p7UZMET1ir and p7UUASZMET1ir were sent to Dr Fridtjof Weltmeier for the maize transformation. The primers used are detailed in Section 7.1.6 of Materials and Methods. However, only p7UUASZMET1ir was not used in maize transformation because 35S *ZmMET1* inverted repeat successfully producing RNAi transformants and it is not giving a drastic lethal effects as seen in tomato studies.

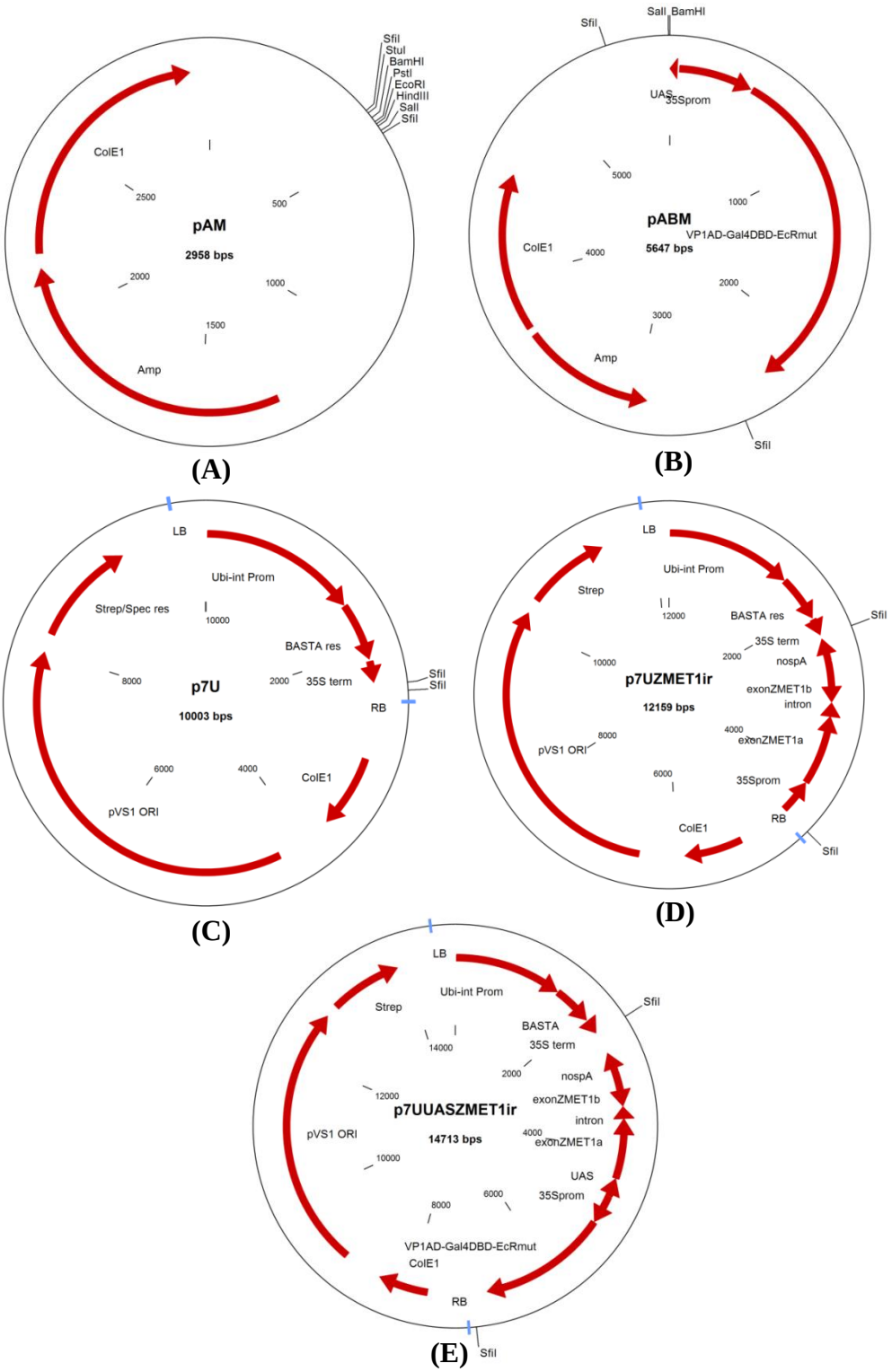


Figure 2.7: Schematic maps of backbones and inverted repeat construct used for knockdown ZmMET1 in maize. A. The figure shows the pAM construct, **B**. The figure shows the pABM construct, **C**. The figure shows the p7U, **D** and **E**. The figure shown the p7UZMET1ir and p7UASZMET1ir constructs produced as described in Section 7.2.1.11 of Materials and Methods. The sequences of the primers used to produce these constructs are provided in Section 7.1.6. LB (left boarder) and RB (right border) mark the T-DNA region. T-DNA region contains the plant selectable marker Glufosinate ammonium (BASTA) and ZmMET1 cassette. ColE1 and pVS1-ORI mark the replication of origin in *E.coli* and *Agrobacterium* respectively. Red arrows indicate gene orientation.

2.2.6 Expression of *ZmMET* in Maize Transformants

Maize A188 was transformed with p7UZMET1ir with the help of *Agrobacterium*. Dr Fridtjof Weltmeier performed the transformation using a method developed by Ishida *et al.,* (2007), and then carried out the growing and analysis of the primary transformants. Ten primary transformants were shown to be harbouring the transgene. The primary transformants were analysed to look for reductions of *ZmMET1* transcripts by Real-Time PCR (RT-PCR) (Figure 2.8). The RT-PCR data was expressed as fold change, relative to the housekeeping gene, *ZmFPGS* gene. The primers used are detailed in Section 7.1.6 of Materials and Methods. All ten of primary transformants show significant silencing of *ZmMET1* genes, indicating that the inverted repeats were successfully targeting the *ZmMET1* genes. Levels of *ZmMET1* in the primary transformants were reduced ranging from 1.5 to 7 fold, compared to the wild type (Figure 2.8). Variations in *ZmMET1* reductions were expected, as the interference in initiation may have happened at different plant development stages. Furthermore, reductions were determined by hairpin copies present in the cells (Hilson *et al.,* 2004; Small, 2007).

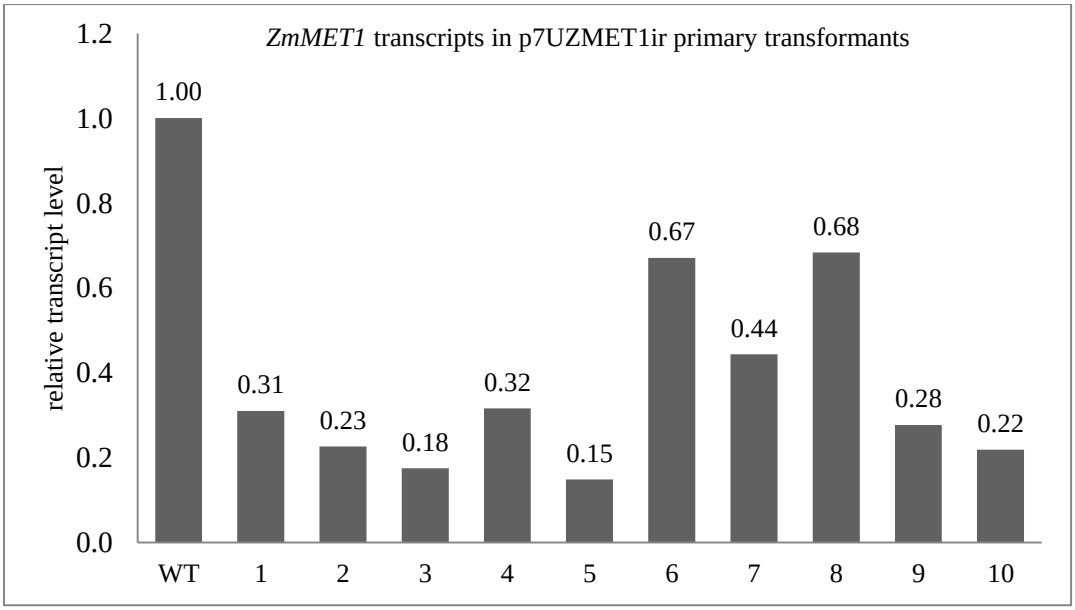


Figure 2.8: Relative expression of *ZmMET1* in ten primary transformants. Various levels of *ZmMET1* silenced transformants obtained, ranging from 1.5 to 7 fold, compared to the wild type WT; wild type A188, 1 – 10; different transformant produce using p7UZMET1ir construct. Numbers on top of bar indicate relative transcript level compared to wild type. The expression analysis was performed in duplicates. The primers used are detailed in Section 7.1.6 of Materials and Methods The RT-PCR was performed by Dr Fridtjof Weltmeier.

2.2.7 Phenotype of transformants with disturbed levels of *ZmMET1* genes

All of the primary transformants were grown under identical conditions in a greenhouse at KWS. Although the *ZmMET1* expressions were repressed in ten transformants, only one of the primary transformants showed growth abnormalities. The transformant 1, with *ZmMET1* expression 3.2 fold lower than the wild type, was dwarfed by, and grew much slower than the others (Figure 2.9A). This abnormality was named “dwarf phenotype”. Interestingly, the growth abnormalities were not observed in other transformants, such as transformant 5, that had even an lower expression of *ZmMET1* (6 fold less), or transformant 8 that had a higher *ZmMET1* expression (4.5 fold greater). This indicates that the abnormalities were not related to the levels of repression, but were probably caused by the randomisation of epigenetic modification, which is non–specific to the induction stimuli. This was

further supported by the formation of abnormal ears by transformant 9, which produced smaller ears with few kernels (Figure 2.9B), although it had a normal growth plant height.

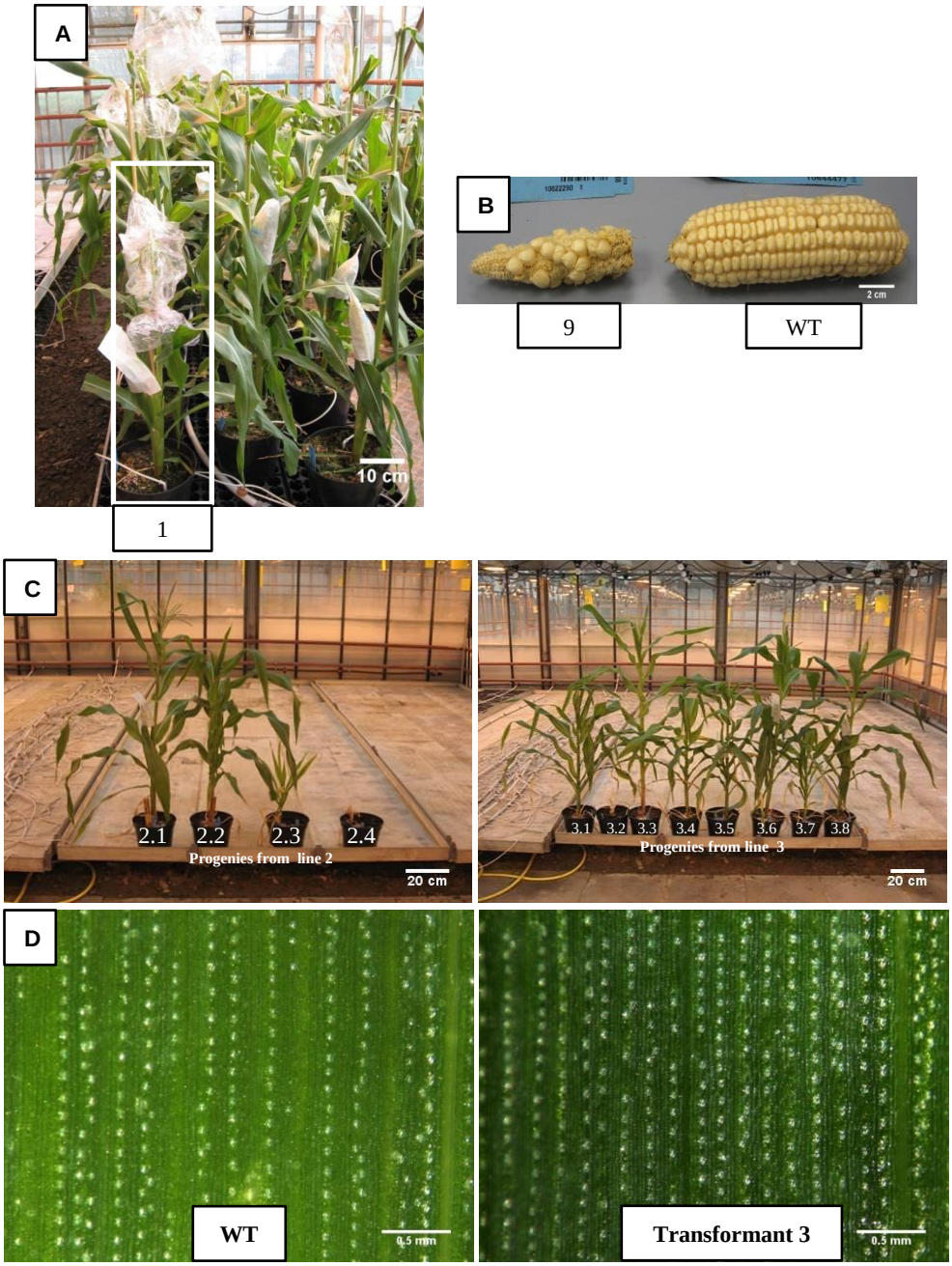


Figure 2.9: Growth abnormalities of maize *ZmMET1* RNAi transformants. **A**: Phenotypes of transformant 1 with dwarf growth (in white box) compared to other transformants (background plants). **B**: Abnormal ears development with lesser kernel from transformant 9. **C**: Pictures of progenies plant from lines 2 and 3, where some plants grow with difference height even their parents came from lines with wild type growth.. More plants with the dwarf phenotype were seen and some plants showed more severe effects such as death. **D**: Stomata distribution of maize leave between wild type (WT) and primary transformant 3. Images are courtesy of Dr Fridtjof Weltmeier.

The kernels produced by the primary transformants were collected and then sown to look for phenotype heritability in the next generation. The kernels that were sown were randomly selected from transformants with both normal and dwarf phenotypes. In the next generation (T1), the effects of repressing *ZmMET1* become more severe, which was demonstrated by an increase in T1 plants showing a dwarf phenotype, even in those that were not progeny of dwarf plants. Out of 38 seeds sown from different primary transformant, 22 (57%) showed various growth abnormalities, such as dwarf phenotype, no germination, and even more severe growth defects by turning yellow, wilting and, even death (Figure 2.9C). Only four of the seeds sown came from line 1 (the dwarfed plant) due to a low grain production. Two of the seeds did not germinate, while the other two died within the first eight weeks after germination. Of the other seeds, which were randomly selected from primary transformants with a normal phenotype, two did not germinate, four died within the first eight weeks after germination, 17 of the seeds grew into plants with a dwarf phenotype, and 11 of the seeds grew into plants with a normal phenotype (Table 2.2). All of these plants harboured the inverted repeat transgenes.

Phenotypes	No of plants
Normal	10
Dwarf	17
Died	6
Not germinate	4

Table 2.2: Number of T1 generations with phenotypes. 38 seeds generated from different transformant lines were selected for studying the phenotype in the next generation. The plants growth abnormalities can be divided into four different phenotypes (normal, dwarf, died and not germinates). Even the plant are originally come from parent that having normal growth, in T1 generation, some of the plant grows with pleotropic phenotypes (dwarf, died and no germination).

The growth abnormalities became more severe as more seeds grew into plants with the dwarf phenotype. These were even worse as some of the seeds did not germinate, and some died during the developmental phase. This could indicate the importance of the *ZmMET1* gene in maintaining the methylation pattern in the maize genomes. The severity of abnormal plant growth in the second generation might have been due to the cumulative effects of ZmMET1 protein deficiency, which led to increased regions where the methylation was not properly maintained.

Besides the plant growth abnormalities, we also investigated stomata distribution of the transformed plants. The stomata developments are related to auxin distribution in the leaves (Le *et al.*, 2014), with a deficiency in plant MET1 protein expression affecting auxin distribution (Xiao *et al.*, 2006). Thus, the stomata distribution changes could be linked to the *MET1* gene. In our study, we observed the stomata distribution in different plant generations (Figure 2.9D). The progenies from primary transformant 3 was used in this studies. The frequencies of stomata found on the leaves of primary transformant plant was higher compared to the wild type (Table 2.3). This was even higher in the next generation that still harboured the transgenes. Surprisingly, the stomata distributions were also high in the T1 generation, even after the transgene had been segregated away.

Plant	Stomata Distribution (mm^3)
Control	62.72±6.32
T0: dwarf phenotype	81.32±14.25
T1: without transgene	116.67±11.7
T1: with transgene	135.40± 22.5

Table 2.3: The stomata distribution from different generation of line 3 of *ZmMET1* RNAi transformant. The stomata distribution was higher in the next generation even without presence of transgene. Stomata distribution was calculated from triplicates image taken from the same leaf development stage.

The *ZmMET1* transformants were also grown in a hydroponic system to look for root phenotype in different transformant lines. Ten seeds, sown from three different lines (Line 2, Line 3, and Line 7) were used in this study, and were selected because in the T1 generation, many of the seeds grew into plants with a dwarf phenotype. Out of three lines, only three progenies of line 7 showed growth abnormalities, plant no 6, 7, and 8, which were not only dwarfed, they also showed abnormalities in their root development (Figure 2.10). The plant no 6 only developed primary and lateral roots, but no crown root, while plant no 7 and plant no 8 developed primary and crown roots, but there was no, or less lateral root development.

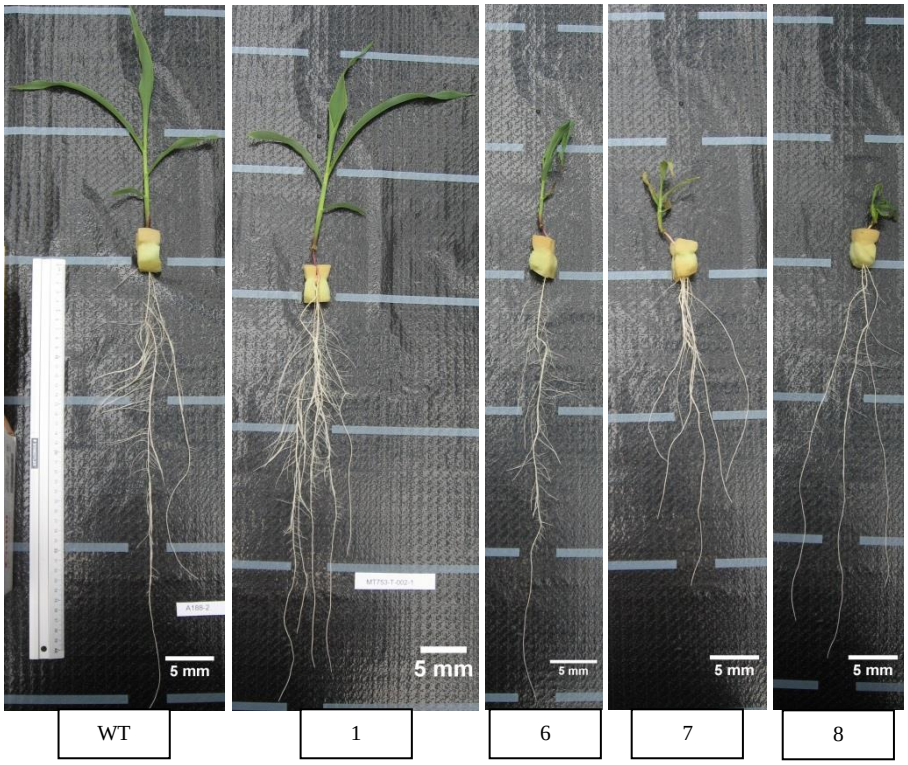


Figure 2.10: Root phenotypes of T1 derived from line 7. Three of line 7 progenies grown with root growth abnormalities, less formation of lateral roots and producing shorter primary root compared to wild type. The images are courtesy of Dr Fridtjof Weltmeier.

Beside looked at the phenotype differences, the root length of plants were measured for quantitative analysis. The root length of ten plants from each lines were measured from images as Figure 2.10 by using ImageJ software. In general, plants from all of the three lines were having shorter roots compared to wild types (Figure 2.11). However, the error bar indicates there might be no significant differences. These could be due to only few T1 plants from each lines showing drastic effect while other plants of the same line grew with normal root length.

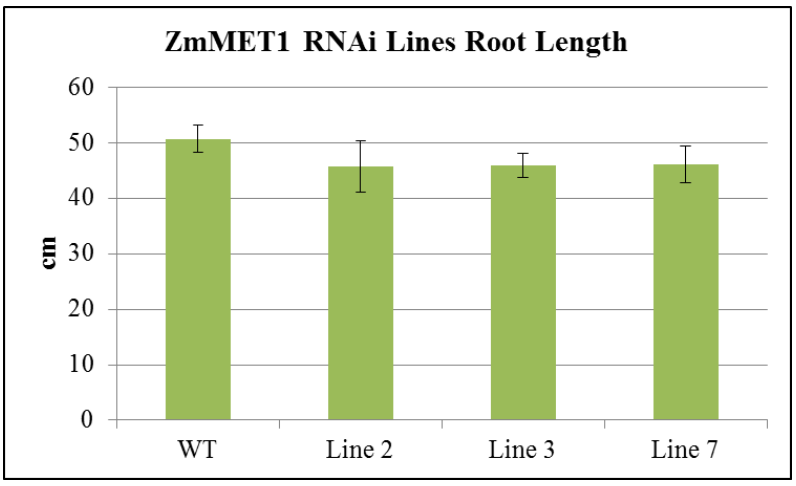


Figure 2.11: Root length average for T1 generation of *ZmMET1* RNAi lines. In general, all lines having shorter root compared to wild type. However the differences is not significant. WT: wild type, Line: refer to lines number of *ZmMET1* RNAi transformants, Error bar indicates standard error with n=10.

This indicates that there were variations in the phenotypes of the next generation and for each line, as the plants did not show the drastic dwarf and root phenotype as observed previously in the T1 generation that grew in soils. The lack of phenotypes of plants with similar MET1 reduction levels is not unusual, as epigenetic changes are chance events, as seen in primary transformants. Perhaps, this also explains the reason for the plants having different types of root development between plant no 6 and 7 in figure 2.10.

Besides looking at the phenotype differences of plants with reduced ZmMET1 protein levels, we were also interested in looking for potential genes that may influence these phenotypes. The potential genes were studied by looking at the expression and methylation status of the promoters and the genes. We focused on genes that were involved in producing the abnormal root phenotypes.

2.2.8 Potential *ZmMET1* Target Genes Related to Phenotype Changes

In order to find the genes that may be responsible for causing the root phenotypes due to low levels of ZmMET1 protein, we looked in the *Arabidopsis* database for maize orthologous genes and in the maize database itself for genes that are known to influence root development. Furthermore, we also looked for genes that are known to be affected by changes in maize MTase mutants. Potential targets are listed in Table 2.4. Expression of these genes was tested in eight progenies from line 7, which showed the root phenotypes in Figure 2.10. The smqRT-PCR method was employed for expression analysis. The primers used are listed in Section 7.1.6 of Materials and Methods.

Out of 13 genes tested, four showed changes in gene expression (Figure 2.11) Genes *IAA14*, *UE1*, and *KIN2* showed a reduction in transcript level, while *GNOM* showed an increase in expression (marked with an asterix). Interestingly, these changes only occurred in several progenies of line 7, plant 1 and 7 for KIN2; plant 6, 7, and 8 for IAA14; and plant 7 for *UE1* and *GNOM*. Surprisingly, most of these changes consistently occurred in plant 7.

Gene identifier (NCBI/ maizedb)	Target genes	References
CO532943	Transposon element copia (NM_001138049)	Makarevitch *et al.*, 2007
GRMZM2G031615	IAA14	Woll *et al.*, 2005
GRMZM2G077147	MYB-transcription factor 77 (MYB77)	Shin *et al.*, 2007
CA827096	uncharacterized LOC103642090	Makarevitch *et al.*, 2007
GRMZM2G171650	MADS-transcription factor 69 (MADS69)	Makarevitch *et al.*, 2007
GRMZM5G877259	Unknown protein	Makarevitch *et al.*, 2007
GRMZM2G002765	Ubiquitin E1 (UE1)	Andorf *et al.*, 2015
GRMZM2G001219	Unknown protein	Andorf *et al.*, 2015
GRMZM2G032198	Unknown protein	Makarevitch *et al.*, 2007
GRMZM2G006474	ARF guanine-nucleotide exchange factor (GNOM)	Okumura *et al.*, 2013
GRMZM2G092542	Rootless concerning crown and seminal lateral roots (RTCS)	Taramino *et al.*, 2007
GRMZM2G136838	Kinesin-like protein 2 (KIN2)	Kreps *et al.*, 2002
GRMZM2G140107	Sucrose phosphate synthase 2 (SPS2)	Hirooka *et al.*, 2005

Table 2.4: Potential target genes found through literature reading. Expressions of these genes were check in progenies of line 7 with short root phenotype.

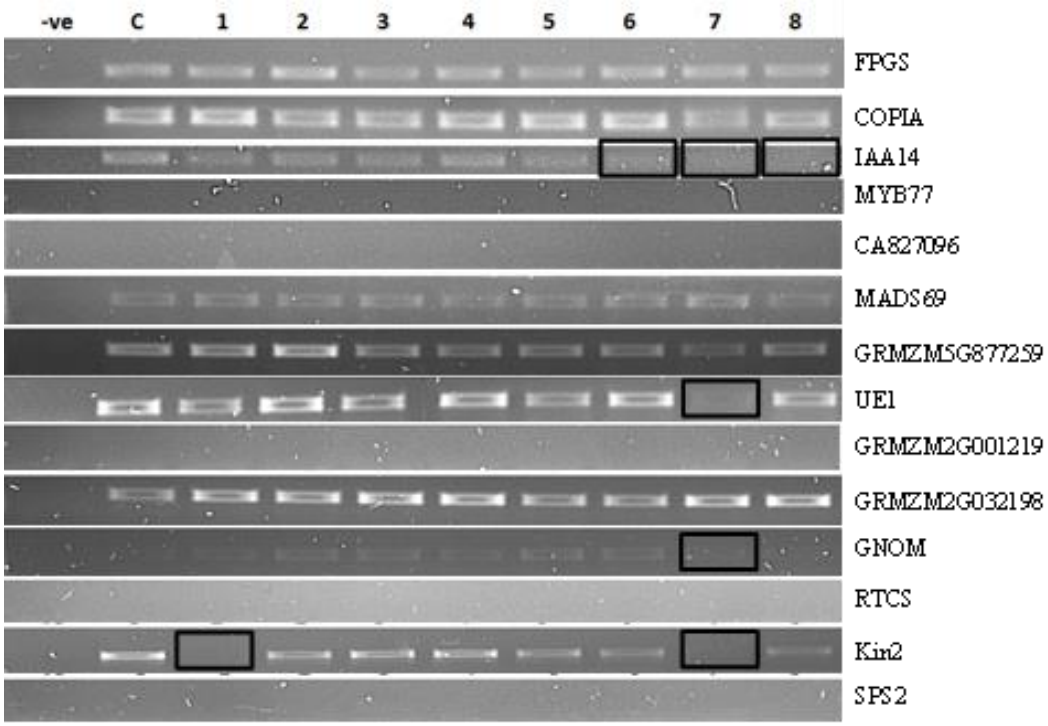


Figure 2.12: Expression analysis of potential ZmMET1 target genes of root samples from progenies of lines 7 using smqRT-PCR. Different in genes expression observed for IAA14, UE1, GNOM and KIN2. The changes only happened at several plants: plant 1 and plant 7 for KIN2, plant 6, 7 and 8 for IAA14, and plant 7 for GRMZM2G002765 and GNOM (marked in the boxes).

Since the transcript differences between progenies were not very clear by smqRT-PCR, we repeated the analysis using RT-PCR (Figure 2.12). Reductions in transcripts were observed in all plants for *KIN2* and *IAA14*. While for *UE1*, reductions were only observed in four samples, and the other four samples showed wild type expression levels. In contrast, *GNOM* transcripts were increased in the same plant except for plant 7 that showed wild type expression. Interestingly, transcripts for four genes were consistently reduced in plant 7.

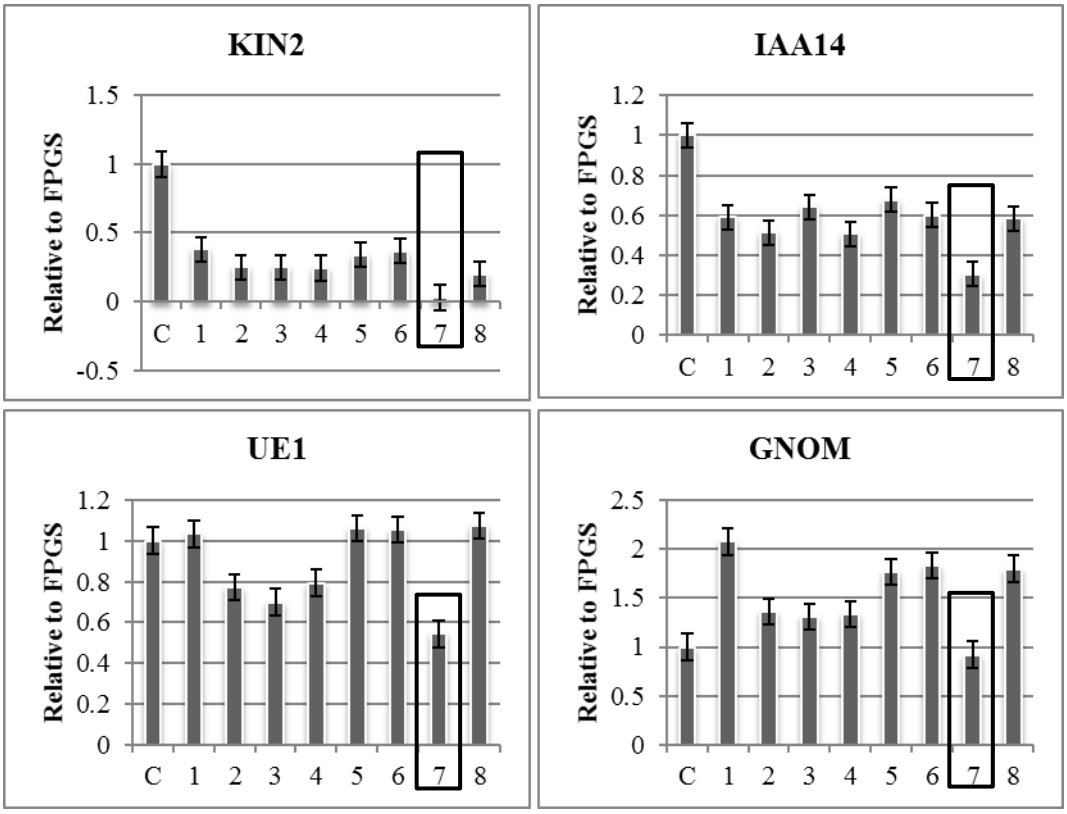


Figure 2.13: Expression analysis by Q-PCR of KIN2, IAA14, UE1 and GNOM to line 7 progenies. Reductions in transcripts were observed for KIN2, IAA14 and GRMZM2G002765 (some plants). However, GNOM transcripts were increased compared to wild type. Interestingly, all transcripts were reduced in plant 7 (marks by boxes). The expression analysis was performed in triplicates. Error bar in the graph indicates standard error.

Epigenetic events were expected to be randomised, because of it being a stochastic event. These randomisations were observed in the root development of the transformants, with only line 7 showing root growth abnormalities. These randomised events were also observed in gene expressions, since the changes only happened to certain plants, even if it came from the same line. There were also consistent changes that happened to plant 7 of line 7, which suggests there might be changes in the epigenetic regulator that are affected by lowering the expression of the *ZmMET1* gene.

Since the reduced expression of *ZmMET1* has an affect on *KIN2*, *IAA14*, UE1, and *GNOM* expression, we investigated this further by analysing the methylation status of these genes, as this could have been the reason for the differential expression pattern. The genes' promoter was targeted for studying the methylation pattern. Due to limited information on the A188 genome sequence, the B73 genome sequence was used. However, none of the primers were able to amplify the promoter region. Thus, we had to change our strategies in order to study the methylation patterns.

2.2.9 Transposable Element Methylation Pattern Changes

In our study, it was planned that the TE regions would be used as methylation controls for investigating the methylation status of the promoters in section 2.2.8 because these regions were known to be methylated, based on the work of Rabinowicz *et al.,* (2003) and the maize database. Surprisingly, we found there were methylation differences in the TE regions during protocol optimisation. Thus, we decided to use this region to study methylation changes in the transformed plants. The maize genome consists predominantly of TEs and repetitive sequences, so it is

important for these regions to be silenced by methylation in order to maintain maize genome integrity. A methylation-sensitive PCR (MSP) method was employed for analysing the methylation pattern at the TE regions. MSP uses the methylation-sensitive restriction enzyme, McrBC, which was used to digest the genomic DNA prior to amplification with specific primers. McrBC is an endonuclease that cleaves DNA containing methyl-cytosine on one or both strands, but it does not act upon un-methylated DNA. The amplicons were later analysed using agarose gel electrophoresis. Any positive bands observed, indicates that there was no, or less methylation, while a negative band indicates there was methylation at the target region (Figure 2.13).

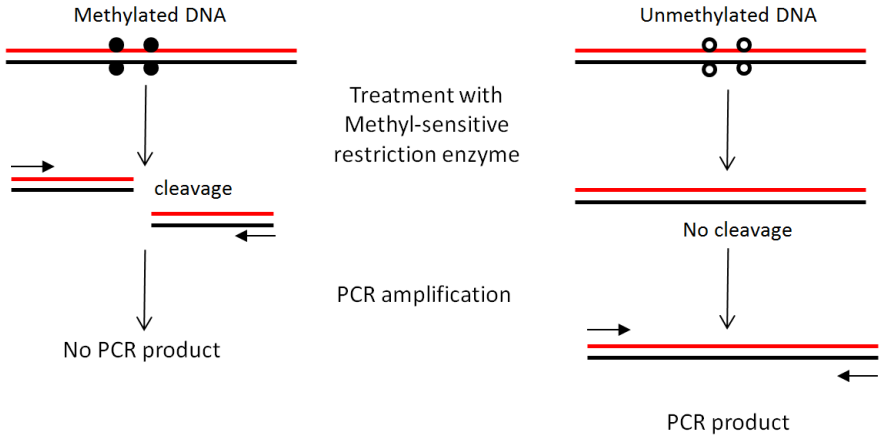


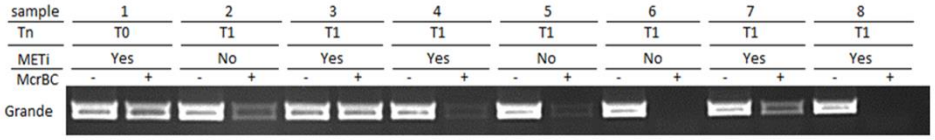
Figure 2.14: Schematic diagram of Methyl-sensitive PCR method. Both Methylated DNA (closed black circle) and un-methylated DNA (open circle) were treated with McrBC prior PCR using specific primer and followed by agarose gel electrophoresis to observe the amplicons.

Plants used for this study were transformants with progenies that had transgenes that were segregated away. These were selected to look for heritability of methylation patterns. The lines used were line 3, 10, and 2. The genomic DNA was extracted from leaves and then the MSP method was applied to the extracted DNA. The primers used are listed on section 7.1.6 of Materials and Methods. Figure 2.14

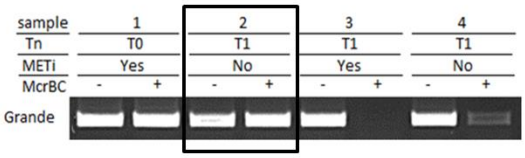
shows changes in DNA methylation at the TEs in transformant progenies, with one plant showing heritable methylation.

One of the TEs, Grande, showed random patterns of methylation changes in all three lines. There were random TE methylation of plants that came from the same lines, as some of the plants lost their methylation at the site of Grande (Line 3: 1, 3, 7; Line 10: 1; Line 2: 1, 3, 4, 5), while for others, the methylation remained (Grande, Line 3: 4 and 8; Line 10: 3; Line 2: 2), even for plants that still harboured the transgenes. These randomised methylation patterns were also observed to other TEs, such as SPM, Prem2, Opie1, and Xilon1. Interestingly, in the same plant, all TEs had the same methylation status. All TEs of the same plant were either methylated or unmethylated. When compared with previous data (gene expressions and phenotypes), randomisation of the same lines was expected. When the transgenes were segregated away, the methylation in these regions was restored. The methylation was restored in all TEs of all lines except for plant Line 10: 2. These consistent changes could be due to changes in key epigenetic regulators as speculated previously.

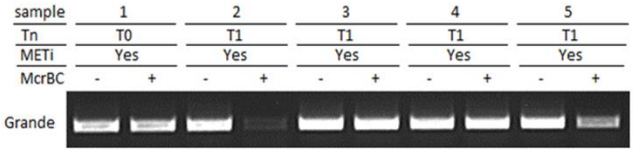
Line 3



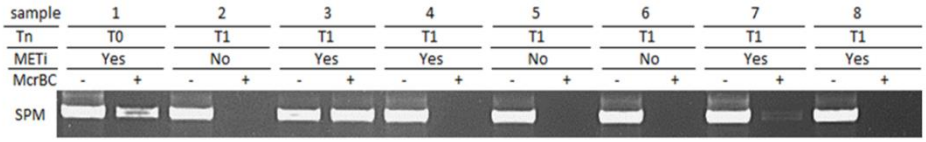
Line 10



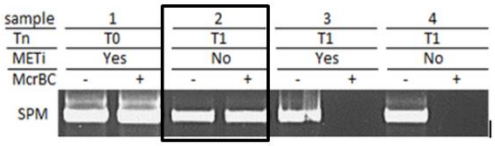
Line 2



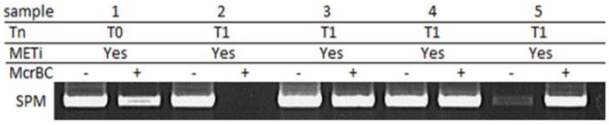
Line 3



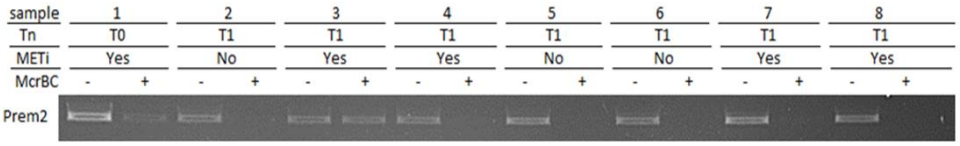
Line 10



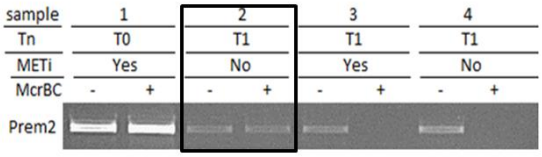
Line 2



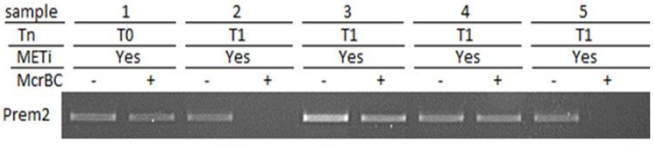
Line 3



Line 10



Line 2



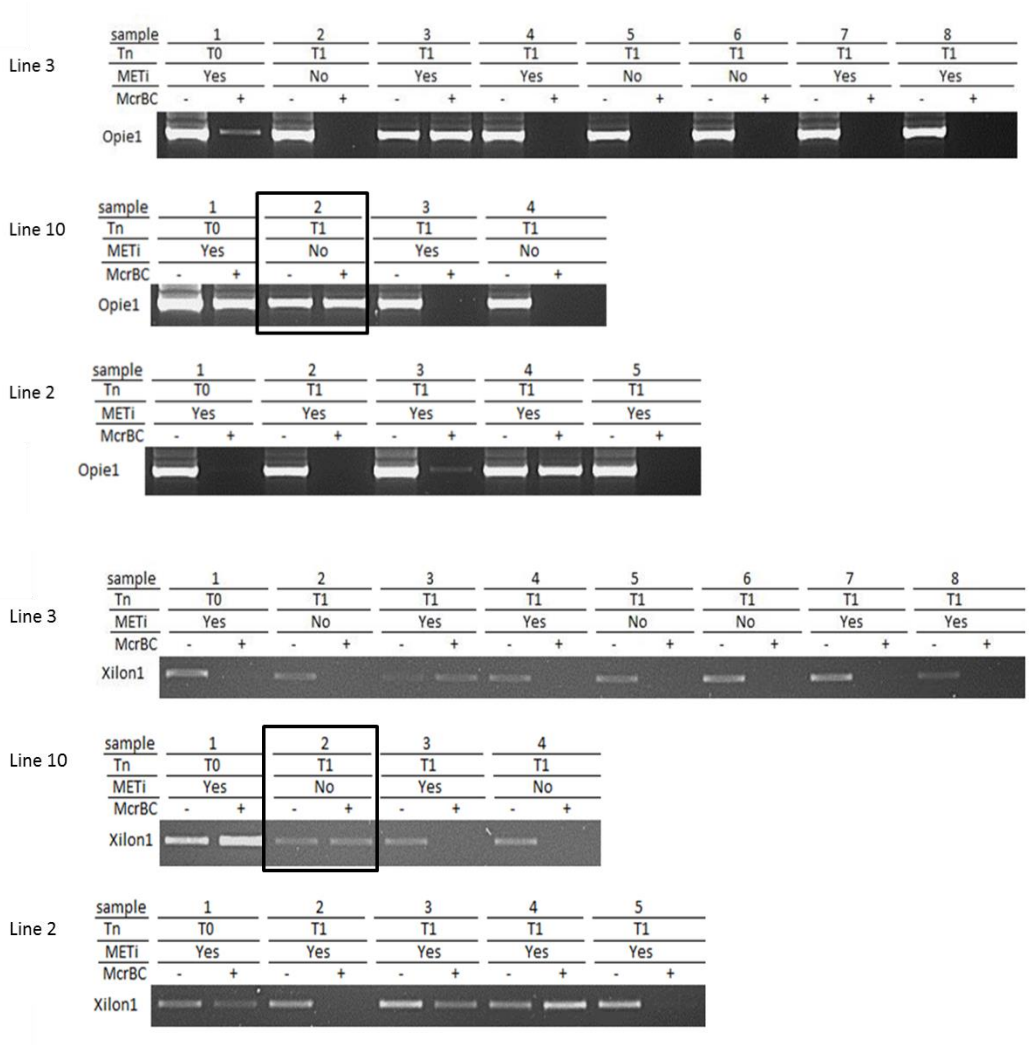


Figure 2.15: Methylation status of studied TEs (Grande, SPM, Prem2, Opie1, Xilon1) from line 3, 10 and 2 of T0 and T1 generation using methyl-sensitive PCR method. The sample from line 3, 10 and 2 was selected because these lines have plants that have lost their transgene in order to look for heritability of the methylation pattern. Out of 16 samples tested, only sample 2 of Line 3 shows demethylation were heritable in all studied TEs (marks by boxes). Sample: numbers is progenies of the lines used in these studies. Tn: transformant generation, METi: presence of transgene, McrBC: presence of methylation-sensitive enzymes.

2.2.10 Expression of TEs in Transgenic Lines

In the classical epigenetic mechanisms, the loss of methylation causes genes to be expressed. Since there are changes in methylation at the TE regions, we employed smqRT-PCR to assess the expression of TEs in plants. The expression of TEs seems to be random, as there are no indications that demethylation caused expression

(Figure 2.15). Only Grande and Opie1 were expressed in the transformants, but expression of Grande and Opie1 were not affected by methylation status. Grande TEs were expressed when they were methylated (Line 3: 5, 6, 8 and Line 10: 2), also when they were not methylated (Line 3: 7 and Line 2: 3, 4, 5). For Opie1, the expression was observed in plants with methylated Opie1 (Line 3: 5, 6, 7, 8; Line 10: 3, 4; and Line 2: 2, 3, 5), and without methylation in plant Line 10: 2 and Line 2: 4.

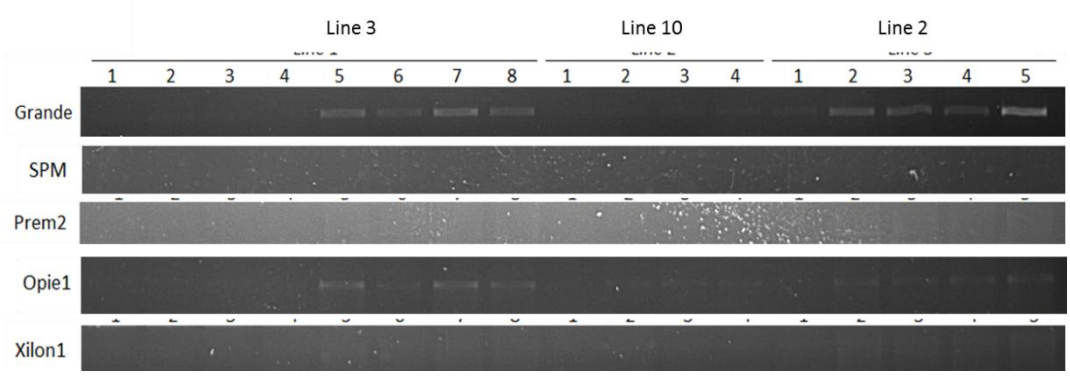


Figure 2.16: Expression of maize TEs in transgenic lines by smqRT-PCR. The similar plant analysed for methylation status of TEs region was used for expression analysis. The RNA and DNA was extracted at the same time in order to minimized the variation. Out of five TEs, only Grande and Opie1 were expressed in several progenies at different transformants lines but there are no correlation between expression and methylation status. Lines indicate independent transformants, Line 3, Line 10, Line 2. Numbers is progenies of the lines used in these studies.

2.3 Discussion

2.3.1 Features and Comparison of the DNA and Amino Acid Sequences of *ZmMET1a* and *ZmMET1b*

The DNA sequence and translated proteins of both *ZmMET1a* and *ZmMET1b* were identified as DNA MTases based on their homology to other eukaryotic MTases (Figure 2.5). The DNA lengths varied as *ZmMET1a* possesses a class II TE between exon1 and exon2 (Figure 2.1). The TE that has been incorporated is a class II transposon that carries a terminal inverted repeat (TIR), known as hAT. Analysis of the chromosomal distribution shows that *ZmMET1* undergoes tandem duplication gene events (Qian *et al.,* 2014), which is an event that normally happens in plant genomes. For example, in *Arabidopsis*, nearly 16% of the genes occur as tandem duplications (Rizzon *et al.,* 2006). Tandem duplication can happen either by unequal crossing, or intra-chromosomal recombination. A large fraction of *Arabidopsis* and rice genes are reported to arise from unequal crossing, with 80% and 88% being incorporated in direct orientation, respectively (Rizzon *et al.,* 2006). While for intra-chromosomal recombinations, these mechanisms usually produce tandem duplicates with opposite orientation (Schuermann *et al.,* 2005). Thus, as *ZmMET1a* is in inversed orientation of *ZmMET1b*, *ZmMET1* might have undergone intra-chromosomal recombination.

Furthermore, we can speculate that repetitive elements might play a role in the *ZmMET1* duplication process because of the presence of hAT at intron 1. The *Rider* long terminal repeat retrotransposon, for example, have caused the duplication of the *SUN* gene in *Solanum lycopersicum*, which led to increased *SUN* expression, and

significantly altered the phenotypes of the fruit shape (Xiao *et al.,* 2008). Another example is the *ULP*-like gene in *Arabidopsis,* which duplicates with the help of *mutator-like* elements (MULEs) (Hoen *et al.,* 2006). Considering the high frequency of repetitive elements in the maize genome, their contributions to both gene and gene expression evolution is significant. This is further supported by Woodhouse *et al.,* (2010) who found evidence of flanking repeats, associated with many of the transposed genes that arose by tandem array mechanisms in *Arabidopsis*. As such, we can speculate that *ZmMET1a* was duplicated with the help of hAT, and *ZmMET1a* is the newly synthesised copy of the *ZmMET1* gene.

ZmMET1a and *ZmMET1b* have high similarity at the amino acid level, particularly at the C-terminus domain, where the catalytic domain is located. This indicates that both ZmMET1a and ZmMET1b were potentially active as the catalytic domains were preserved. A total of nine amino acid substitutions were scattered along the ORFs., and the conserved acidic region of plant MET1 sequence (Finnegan and Kovac, 2000) was present in both forms of ZmMET1.

Two long RFD domains were identified upstream of the acidic regions, which are commonly found in plant MTases. In comparison, only one RFD domain is found in mammalian MTases. The RFD domain of DNMT1, a mammalian homolog of MET1, has been implicated in subcellular localisation, protein association, and catalytic function (Syeda *et al.,* 2011). This domain contains the binding site for: Ubiquitin-like containing PHD and RING fingers domains 1 (UHRF1), which is a plant homolog of VIM1 (Hashimoto *et al.,* 2008). Furthermore, the RFD domain of MET1 and DNMT1 also acts as a safety mechanism for protecting the genome from replication-independent DNA methylation, by interacting with its own catalytic

motif (Garvilles *et al.,* 2015; Bashtrykov *et al.,* 2014; Garg *et al.,* 2014). Two of the amino acid substitutions were located at RFD domains, with one substitution for each domain (Figure 2.2).

ZmMET1 contains two BAH domains, which are located downstream of an acidic region, and are involved in protein–protein interactions in gene silencing mechanisms (Callebaut *et al.,* 1999). This indicates that involvement of ZmMET1 facilitates DNA methylation, since BAH domains have also been observed in rice MTases (Teerawanichpan *et al.,* 2004). Each BAH domain has a different function, for example, the BAH-1 domain recognises methylated–lysine marks based on the formation of an aromatic cage as its protein structure, and this cage formation has also been found in the mammalian DNMT1 (Wang *et al.,* 2016). However, the BAH-2 domain, found in peanut and legumes, has a significant sequence similarity with the polybromo BAH domain, suggesting an involvement in specific interactions with histones and other chromatin proteins (Wang *et al.,* 2016; Garg *et al.,* 2014). In rice, the BAH-2 domain of OsMET1 has also been reported to bear a putative nuclear localisation sequence (Teerawanichpan *et al.,* 2004), indicating that its MET1 activities occur within the nucleus. However, all of these suggestions have come from sequence analysis or protein structure predictions, therefore, appropriate and substantial protein experiments need to be carried out.

As speculated, *ZmMET1a* might be the new copy of the *ZmMET1* gene. The newly duplicated gene would normally be facing several possible fates, and mostly the newly formed gene pair will be loss-of-function genes (Walsh, 1995), arising due to the introduction of deleterious mutations, especially in the duplicated genes with redundant roles. The mutations may be accumulated over time and undergo

pseudogenes (Ohta, 1987; Walsh, 1995), supporting our suggestion that *ZmMET1a* is the new gene. Amino acid substitutions would be more likely to occur in ZmMET1a compared to ZmMET1b, based on the alignment with other plant MTases (Figure 2.3). The *ZmMET1a* gene might fit with neo-functionalisation, where the duplicated genes are expected to have redundant functions, so one copy is free to accumulate mutations without affecting the gene's ancestral role.

However, if we consider the requirement of maize in maintaining its large genome size, integrity from TEs, the differences in predicted protein structure (Figure 2.4), and the variation in expression level of ZmMET1a and ZmMET1b proteins (Figure 2.6), perhaps, the duplication of *ZmMET1a* fits better into the subfunctional duplication, degeneration and complementation (DDC) model. The subfunctional DCC model is where both gene copies will accumulate the natural mutations, leading to impairment of function, but complementing each other to cover the full spectrum of their ancestral expression pattern. The DCC model, where reported to have happened to a pair of MADS-box transcription factor genes (*ZAG1* and *ZMM2*) in maize, showed varied expression patterns in different organ types. For example, *ZAG1* is highly expressed in the carpel, but weakly expressed in the stamen, and *ZMM2* is highly expressed in the stamen (Mena *et al.,* 1996; Force *et al.,* 1999). This is in contrast to their single orthologous gene, *AGAMOUS,* found in *Arabidopsis*, which is highly expressed in both tissues (Coen and Meyerowitz, 1991). This model was further supported by spatial expression patterns of MTases that are closely related to *ZmMET1,* the *OsMET1a* and *OsMET1b* genes (Figure 2.5). The expression of *OsMET1b* is higher in the callus, root, node, sheath, immature panicle, and inflorescence compared to *OsMET1a*. However, expression of *OsMET1b* was not detected in differentiated tissue, matured leaves, and matured panicle

(Teerawanichpan *et al.*, 2004; Ahmad *et al.*, 2014). Differences in expression levels between *MET1* copies were also observed in maize leaves (Figure 2.6). Perhaps we can speculate that *ZmMET1* might also be expressed differently in different organs or tissue types because of the differences in protein structure prediction between ZmMET1a and ZmMET1b (Figure 2.4), which might lead to tissue specific functions. However, these speculations need to be substantiated experimentally as suggested in section 2.2.2.

2.3.2 Different Expressions of ZmMET1a and ZmMET1b Protein

The specificity of primers designed using the Prim-SNPing technique (Chang *et al.*, 2009) has enabled us to differentiate the expression of *ZmMET1a* and *ZmMET1b* (Figure 2.6B). An assessment of ZmMET1 expression in leaf tissues has found an abundance of *ZmMET1* transcripts in all types of growth stage (Figure 2.6C). These suggest that ZmMET1 is involved in methylation activities in all tissues. However, as mentioned in section 2.2.4, a further study is required, especially by RT-PCR to verify this data. As the maintenance methylation protein, MET1 is known to be highly expressed in proliferating or growth tissues of various plants (Bernacchia *et al.*, 1998; Teerawanichpan *et al.*, 2004; Ahmad *et al.*, 2014), it could suggest that ZmMET1 might behave spatially in plant tissues.

In plants that possess two copies of the *MET1* gene, one of the genes is reported to encode the key protein for methyltransferases activities of the plant. In rice, OsMET1b is more abundant and plays an important role as DNA MTases in rice, compared to OsMET1a (Yamauchi *et al.*, 2008; Hu *et al.*, 2014). OsMET1a is also

said to act as a genetic backup mechanism to maintain the methylation (Yamauchi *et al.*, 2014). While for carrot, DcMET1a and DcMET1b differ in their relative abundance (Bernacchia *et al.*, 1998). Perhaps we can speculate that there is a high possibility of the disproportion of responsibility being held by one of the *ZmMET1* genes, with high levels of *ZmMET1a* transcripts compared to *ZmMET1b* (Figure 2.6C), especially with close relationships of *ZmMET1* with *OsMET1a* (Figure 2.5).

2.3.3 Consequences of Disturbing *ZmMET1* Expression Levels

Given the high sequence similarity of *ZmMET1a* and *ZmMET1b*, the inverted repeat approach used in these experiments was expected to cause the degradation of mRNA in both *MET1* genes, thereby disrupting the methylation mechanisms or pathways, leading to changes in methylation patterns in the maize transformants. The phenotypic changes in the T0 generation that caused transformants with debilitating maintenance of methylation was insignificant, with only one transformant showing the dwarf phenotype (Figure 2.9A). However, the lack of phenotype in plants with reduced *ZmMET1* expression levels in the T0 generation is not unusual, as epigenetic changes are chance events. The lack of phenotypes have been reported in the reduced MET1 *Arabidopsis* transformants also, by using an antisense approach, and only the normal phenotype is observed in the primary transformants (Finnegan *et al.*, 1996).

The similar chance event was also seen in rice heterozygote *OsMET1b* mutants, where no overt phenotype was observed, even though their methylation was decreased (Yamauchi *et al.*, 2014). Furthermore, the dwarf or retarded phenotype is a typical growth phenotype that occurred in *met1* mutants plants, such as

Arabidopsis (Kankel *et al.,* 2003), tobacco (Oh *et al.,* 2009), and rice (Hu *et al.,* 2014). Thus, we can speculate that a reduction of ZmMET1 may have increased the probability for an epigenetic switch, but this is not guaranteed. In this respect, a reduction in ZmMET1 protein expression is necessary, but not always sufficient to cause phenotypic changes. In contrast to *Arabidopsis*, where only one copy of the *MET1* gene is available in the genome, the *met1* homozygous mutants have produced plants with consistent delays in flowering time (Saze *et al.,* 2003), which indicates that the effect of deficient MET1 is severe compared to in rice and maize. Perhaps, the presence of a second copy of the *MET1* gene in rice and maize should not be overlooked, especially when it is known that *OsMET1a* can act as a back-up mechanism in maintaining methylation in rice (Yamauchi *et al.,* 2014).

The abnormal growth of maize was passed on and propagated in the next generation, where more plants grew into a dwarf plant, despite their ancestors having a normal phenotype (Figure 2.9C). This indicates the importance of having the correct level of ZmMET1 expression during plant development for a proper methylation pattern, especially when it is known that expressions of ZmMET1 vary at different developmental stages (Candaele *et al.,* 2014). Similar severity in plant growth abnormalities in the next generation was also observed in *Arabidopsis*, in plants with reduced MET1. Half of the primary transformants give rise to progenies with a range of pleotropic phenotypes (i.e affecting the floral structure and flowering time), and at the same time have a reduction in the methylation levels (Finnegan *et al.,* 1996), while for *met1* homozygous plants, the progenies also have severe phenotypes, manifested by getting smaller with each generation (Mathieu *et al.,* 2007). There are several genes that are known to meiotically inherit epigenetic changes, and cause plant development abnormalities, such as *SUPERMAN (SUP*),

and an allele named *clark kent* (*clk*), which alters the floral pattern formation (Jacobsen and Meyerowitz, 1997).

The cumulative effects of ZmMET1 protein deficiency have led to more regions where methylation is not properly maintained. Perhaps, there are small windows of opportunity for inducing methylation pattern changes during the developmental stage where epigenetic switching occurs. After the existing epigenetic patterns have established, the patterns are simply propagated. Male gametogenesis is a potential stage for epigenetic switching to occur as CHH methylation is lost from retrotransposon in microspore and sperm cells during DNA reprogramming (Calarco *et al.,* 2012). These might have led to genome instability due to failure to retain silencing of retrotransposon in the embryo, which resulted in the severe phenotypes in the second generation (Table 2.2).

However, these drastic dwarf phenotypes were not observed in the hydroponic cultivation, as only three plants, from line 7 showed the root growth abnormalities (Figure 2.10). The lack of drastic phenotypes may be linked to a different growing season, growth conditions (hydroponic compared to soil), different plant age, or simply because of the chance event that was explained previously.

Interestingly, the phenotype that developed from deficient ZmMET1 was retained, even after the transgene had been segregated away (Figure 2.9D). These were seen in the stomata distribution of the transformants, where the stomata in primary and secondary transformant's leaves were 1.3 and 1.8 times more abundant than in controls, respectively. The stomata density was increased to 2.2 times, even after the transgene had been removed. A similar phenomenon was observed in *Arabidopsis*

met1-1 mutants, where, stomata density increased to three times that of the Col-0 wild type (Vassileva *et al.,* 2016). Further tests have been done, by growing the *Arabidopsis met1* homozygous plant under low humidity conditions, which have shown that there are no significant effects, confirming that the increase in stomata density is caused by a MET1 deficiency (Tricker *et al.,* 2012).

Perhaps, deficiencies in MET1, have led to changes in the methylation pattern of genes involved in stomata development, which could be an epialleles or a MET1 direct target under an RNA-independent pathway, especially when the phenotype is retained in MET1 restored wild type expression level plants. Discoveries of new epialleles or heritable phenotypes in plants that have been previously exposed to defective MTases are common, especially in *Arabidopsis*. Examples of epialleles found by methyltransferase depletion studies are non-coding RNA (ncRNA) (*AT4G15242*) (Watson *et al.,* 2014), and *FWA* (*FLOWERING WAGENINGEN)* (Reinders *et al.,* 2009).

The phenotype caused by a ZmMET1 deficiency that has influenced certain genes to be differentially expressed, possibly occurred by changing the methylation pattern of the gene promoter region, or the secondary effects thereof. As expected, differential expression changes occurred in plants with dwarf phenotypes (Figure 2.10), but one dwarf plant showed consistent changes in all four genes. However, we were unable to link the gene expression with methylation status of the gene promoter because no genomic sequence information was available for the A188 variety. Hence, methylated TE regions were used to look for changes in methylation patterns of the plants. Interestingly, the methylation status for most of the TEs is similar in the same plant and these patterns are consistent in most of the studied plants (Figure 2.14).

This indicates that low levels of ZmMET1 might have activated an epigenetic regulator responsible for methylating TEs in the plants. A common regulator that is known to silence TEs by methylation is 24nt siRNA through RNA-directed DNA methylation (RdDM). The 24nt siRNA is known to function as a guide in targeting CHH islands of TEs that are located near genes (Gent *et al.,* 2013). However, production of 24nt siRNA is also subject to MET1 expression, in which MET1 is responsible for repressing the formation of double stranded RNA precursors (Blevins *et al.,* 2009). Besides RdDM, DNA methylation can also occur by a MET1 RNA-independent pathway, since 63% of the methylated regions found in *Arabidopsis* do not have matching siRNAs (Zhang *et al.,* 2006). Even though, RNA-independent DNA methylation pathways were found in *Arabidopsis*, there is a high possibility this pathway may also be present in maize, since differences in ZmMET1 protein structure observed previously could lead to specific function or target recognition. However, we were unable to link the demethylation of TE with causing TE expression (Figure 2.15). Perhaps, the regions selected for MSP were not suitable targets to relate with TE expression profiles, especially if the targeted regions were not densely methylated.

In this chapter, there was evidence of the importance of ZmMET1 protein in regulating maize plant growth and development. The disruptions in the steady-state of ZmMET1 protein led to severe plant growth abnormalities in the next generation and caused changes in TE methylation patterns.

Chapter 3

Mutation of *MET1* using Targeted Genome Editing Technologies

3.1 Introduction

Gene knockout and induced mutations are common strategies used to investigate and understand the functions of genes in plants. A conventional approach to introducing mutations in plants is to expose them to X-ray radiation. X-rays have been used as a mutagen since 1927 because they are easily accessible to researchers, and have successfully introduced phenotypic variations in barley seedlings and sterility in maize tassels (Stadler, 1928a; Stadler, 1928b). Over time more sophisticated mutagens, such as gamma-rays and neutron radiation were used (Auerbach and Robson, 1946; Auerbach, 1949), before the discovery of chemical mutagens, such as ethyl methane-sulphonates (EMS) (Westergaard, 1957), which are still in use today even though both methods have their own limitations. Radiation-based methods tend to cause cell and genetic damage on a larger scale, and may severely reduce cell viability (Wu *et al.*, 2005). Furthermore, chemical-based methods lead to single nucleotide polymorphisms (SNPs) during DNA replication, which are random point mutations that occur throughout the genome. This makes the study of specific genes difficult because the phenotypic effects seen might be caused by a mutation in other regions of the genome.

With the advance of genetic engineering, different ways of disrupting gene function have been explored, such as insertional mutagenesis, which is when foreign DNA is used to disrupt the gene of interest, and involves the use of TE or transfer DNA (T-DNA). This type of approach has been widely applied to *Arabidopsis* species

because the introns of this plant genus are small; a single T-DNA insertion is usually enough to cause gene disruption. Although this technique is still based on random integration, there is a good chance that the T-DNA would be successfully inserted within the gene of interest, because of its small genome and the large population of T-DNA-transformed lines available. However, the population of T-DNA-transformed lines is limited to *Arabidopsis*, so studying the genes of important species, such as crop plants, remains a challenging task. The problems with random insertion have been addressed by using post-transcriptional silencing techniques to target genes of interest with short RNA interference (siRNA). However, the silencing levels are variable and incomplete, which could allow certain genes to function correctly, thereby showing little or no phenotypic effects.

Recently, a technique that induces double-strand breaks (DSB) at specific DNA regions using customised engineered endonucleases, such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeat (CRISPR) RNA-guided nucleases, has given researchers the opportunity to study gene functions in crop plants. ZFN DNA binding domains (DBDs) recognise three bases (Wolfe *et al.*, 2000), while TALEN DBDs, recognise a single nucleotide DNA (Cermak *et al.*, 2011). The latest and simplest targeted genome-editing technology, is CRISPR, which utilises the Watson-Crick RNA-DNA binding principle to find a particular target, with the help of the Cas9 protein (Makarova *et al.*, 2011). All of these endonuclease systems rely on host DNA-repair mechanisms for re-ligating the DSBs. The DSBs are either repaired by non-homologous end joining (NHEJ) pathways, an error-prone mechanism that often results in frame-shift mutations, or homologous recombination (HR) pathways, which require the presence of complementary DNA fragments.

ZFNs, TALENs and CRISPRs have been widely used for editing various plant genomes. For example, ZFNs have been used to introduce mutations in the *ABSISCIC ACID INSENSITIVE 4* (*ABI4*) gene and the GUS reporter in *Arabidopsis* (Osakabe *et al.,* 2010; Tovkach *et al.,* 2009). In maize, ZFNs have helped to integrate additional copies of a herbicide-resistant marker gene (Ainley *et al.,* 2013). However, because of the high toxicity and limitations with triple nucleotide targets, the ZFN approach was less favoured by the research community (Pruett-Miller *et al.,* 2009). Despite this, it has been reported to have been applied in the agricultural industry.

TALENs have also been used to target several genes in various plant species. For example, the *ALCOHOL DEHYROGENEASE 1* (*ADH1*) gene in *Arabidopsis* (Cermak *et al.,* 2011); the *OsSWEET14* gene, a disease-susceptibility gene, in rice (Li *et al.,* 2012); the GUS reporter, and the *ACETOLACTATE SYNTHASE* (*ALS*) gene in tobacco (Mahfouz *et al.,* 2011; Zhang *et al.,* 2013). In addition, barley (Wendt *et al.,* 2013) and maize (Liang *et al.,* 2014) have been reported to have used TALENs.

The first application of CRISPR in plants was reported in August 2013 from three independent groups (Nekrasov *et al.,* 2013; Shan *et al.,* 2013; Li *et al.,* 2013), and these were followed by more reports shortly after. The simplicity of the CRISPR tool makes it convenient, and is preferred over other genome-editing tools. CRISPR has since been used to target various genes in many plant species, such as rice, wheat, sorghum, maize, and tomato (Shan *et al.,* 2013; Jiang *et al.,* 2013; Feng *et al.,* 2016; Pan *et al.,* 2016).

MET1 is one of the key DNA methyltransferases (MTases) involved in maintaining DNA methylation in plants. Knocking out the *MET1* gene has resulted in hypomethylation throughout the genome in *A. thaliana Col* (Cokus *et al.*, 2008). The *MET1* gene is an interesting target to study due to the effects of global DNA methylation pattern changes upon disruption. With the successful induction of DNA methylation changes in *Arabidopsis* by targeting MET1, we would like to further study the *MET1* gene in crop plants. Tomato was selected as one of the model crop plants, along with tomato lack of available *met1* (*tomMET1*) mutants. We have previously experienced difficulties in producing *tommet1* mutants using the IR approach, with no plants being successfully regenerated from callus transformation (Watson, 2013). Thus, in this study we decided to employ the targeted genome editor for producing *tommet1* mutants in a high precision manner. Furthermore, this also offers the chance of producing a heterozygous mutant, if the homozygous mutants are detrimental. As such, the objective of this chapter is to employ the TALEN and CRISPR systems to induce mutations in *MET1,* knocking out the gene in tomato. The plant transformations with these constructs were executed with the help of Enza Zaden.

3.2 Results

3.2.1 MET1 Targeted Mutagenesis using TALEN and CRISPR

Targeted genome-editing was chosen as an alternative way to study the tomato *MET1* gene. This is a continuation of studies by Dr. Michael Watson, who was unable to produce tomato with the silenced *MET1* gene using the inverted repeat

approach (Watson, 2013). In this study, both the TALEN and CRISPR systems were employed to increase the chances of producing viable mutants, especially as the main reason for the failure to produce *tommet1* mutants with inverted repeat techniques is not known.

One of the possible factors might be that in tomato, MET1 is so crucial for the maintenance of DNA methylation during explant regeneration, even a small reduction in MET1 levels leads to lethality. Or perhaps a very high rate of transgene transcription is counterproductive to explant regeneration. Thus, a cleaner and more specific tool for producing mutants is required. Furthermore, we aimed to produce a *tommet1* mutant with only one of the alleles mutated in order to reduce the detrimental effects of MET1 depletion. This can be achieved using a weak promoter to drive the expression of the TALEN and CRISPR systems. The genome editor was employed to target the exon2 region of the *tomMET1* gene, and a region with SfaNI and Bsr1 restriction enzyme digestion sites was selected to help with mutant identification in later analyses (Figure 3.1).

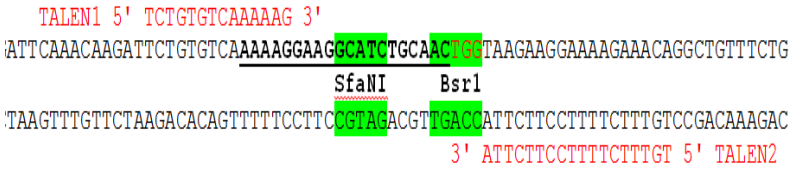


Figure 3.1: Regions of MET1 targeted by TALENs and CRISPRs for tomato. The TALEN and CRISPR binding region for Tom*MET1*. The TALEN binding site is marked in red. CRISPR binding site is underlined and bold. The protospacer adjacent motif (PAM) is marked in red.

3.2.2 Development of TALEN Construct for Targeting the *MET1* Gene

The TALEN construct was constructed based on a protocol described by Sanjana *et al.*, (2012). Monomers that encoded a DBD for recognising specific nucleotides were amplified from template plasmids. Each of the monomers were amplified using specific primers as described by Sanjana *et al.*, (2012). Naturally, the TALEN protein recognised thymine (T) as the first base of the DNA binding site, thus, this parameter was taken into consideration when finding a target region (Boch *et al.*, 2009). The TAL Effector-Nucleotide Targeter 2.0 (TALE-NT) software (https://tale-nt.cac.cornell.edu/node/add/talen) was used to assist in finding the binding site. The primers included a BsmBI restriction site, which produces a 5'-end overhang, with a sequence that is complementary the next monomer digested, allowing ligation of the monomers. The monomers were kept in an organised manner to ease with monomer mixing and production of the tandem hexamers.

As the basic mechanisms for producing TALEN constructs are the same, the construction of TALEN1 for targeting the *TomMET1* gene is discussed here (Figure 3.). The targeted sequence was divided into sub-sequences of six base pairs in length, (T) CTGTGT CAAAAA (G), where the initial nucleotide must be a thymine (T) because it is TALE's natural target, and the final nucleotide is a guanine (G), which is indicated by a bracket, and not included in the design of the hexamers. G is recognised by half of the monomers that are already included in the destination plasmid, as described by Sanjana *et al.*, (2012). Thus, for *TomMET1* TALEN1, the hexamers were, 1 = HD-NG-NN-NG-NN-NG, and 2 = HD-NI-NI-NI-NI-NI, and were built simultaneously. Because G is the final base, the NN backbone, pTALEN_v2 (NN) was used. Six monomers were subjected to the golden-gate

reaction, a sequential digestion-ligation activity involving BsmBI, and DNA ligase. The 5'- and 3'- end of the hexamers were designed to be complementary to each other, thus the ligation favoured the formation of circular DNA. The completed golden-gate reaction was treated with PlasmidSafe endonuclease, an enzyme that only digests the non-circularised DNA, preventing the carryover of false binding monomers.

Each of the circular DNA hexamers were amplified using HEX primers, and the amplicons subjected to a second golden-gate reaction using BsaI, in the presence of pTALEN_v2 (NN). This formed two ligated hexamers in the backbone of the pTALEN_v2 (NN) plasmid, and the TALEN construct was sent for sequence verification. Since the nuclease domain, Fok1, only works in dimers to introduce DSB, two TALENs were needed for a single target. Thus, two TALENs were constructed for each *MET1* gene target. For targeting the *TomMET1* genes, the DBD was assembled in an arrangement as shown in Figure 3.2. The complete TALEN cassette consisted of two sets of 'promoter-TALEN-terminator' sub-cloned into the plant transformation vectors. The TALEN expression was controlled by a bidirectional mannopine promoter for the expression of TomMET1TALEN (Figure 3.4). The TomMET1TALEN construct was then sent to Enza Zaden for tomato transformation.

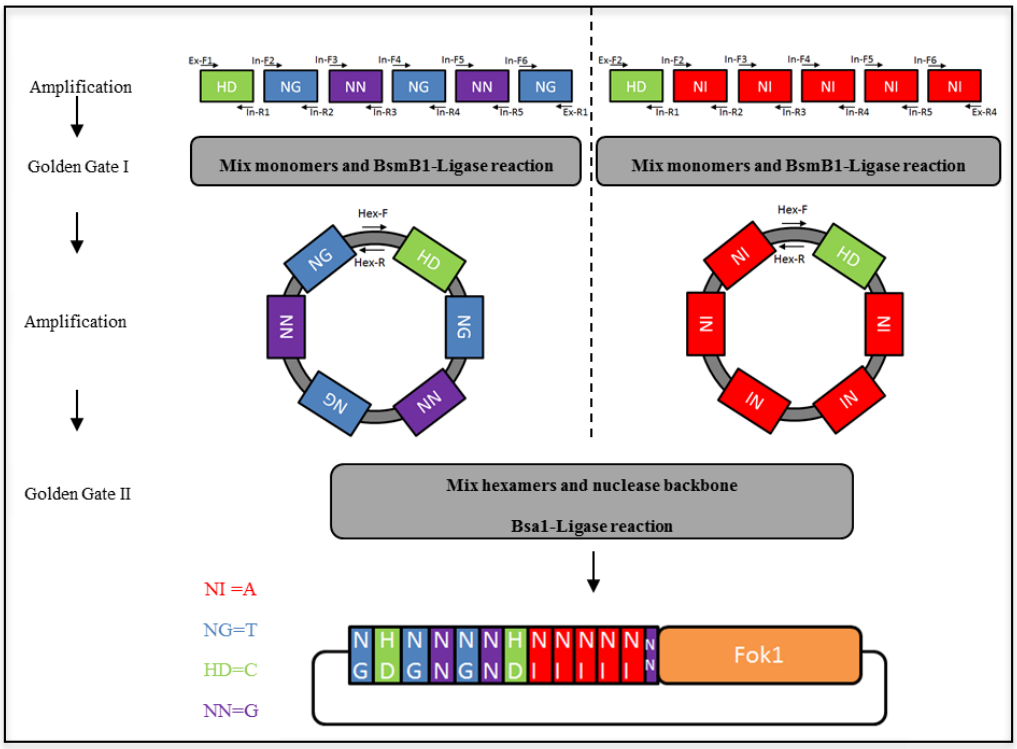


Figure 3.2: Steps for making the TomMET1TALEN1 protein-binding DNA fused with nuclease domain (Fok1). Monomers were amplified using specific primers from template plasmids, followed by golden-gate reaction, and a digestion-ligation reaction with BsmB1-DNA ligase enzymes. Monomers formed circular hexamers. The two hexamers were amplified and underwent another golden-gate reaction. The backbone plasmid harboured half of the RVD domain (NN) (indicated by smaller NN before the Fok1 domain) and the nuclease domain (pTALEN_v2 (NN), to form a complete TALEN construct. The NI domain recognised Adenine (A) (Red), NG domain recognised thymine (T) (Blue), HD domain recognised cytosine (C) (Green), and the NN domain recognised guanine (G) (Purple).

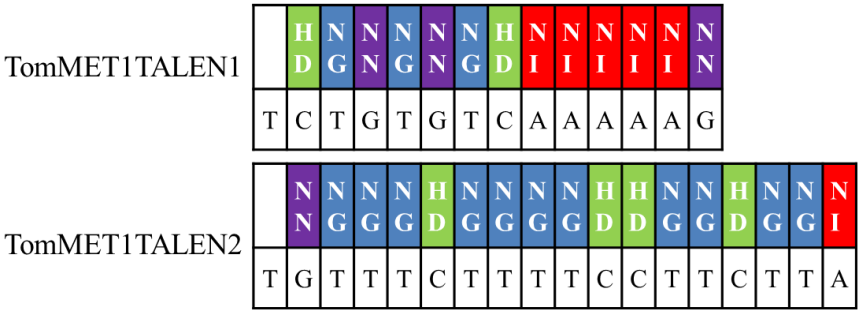


Figure 3.3: Arrangement of TALEN monomers for targeting TomMET1. One target region required two TALEN s because the nuclease domain, Fok1, works in dimer structures. Each TALEN construct was designed to be specific to the target region only. The NI domain recognised adenine (A) (Red), NG domain recognised thymine (T) (Blue), HD domain recognised cytosine (C) (Green), and NN domain recognised guanine (G) (Purple).

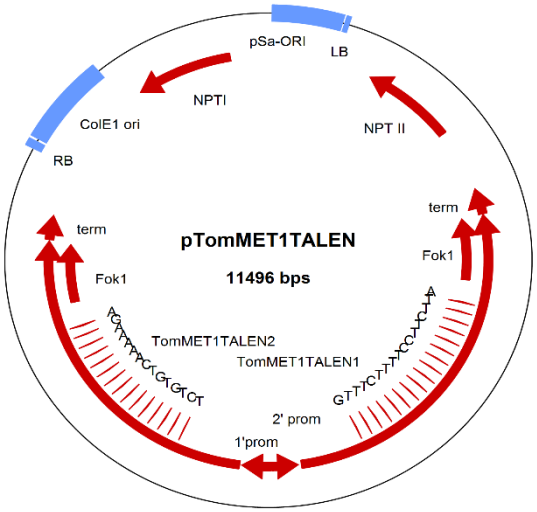


Figure 3.4: A schematic representation of pTomMET1TALEN in the plant transformation vector. The TomMET1TALEN construct was driven by the bidirectional mannopine promoter, with the plant selectable marker for Kanamycin (NPTI). The constructs were produced as described in Section 7.2.1.11 of Materials and Methods. The sequences of the primers used to produce these constructs are provided in Section 7.1.6. LB (left boarder) and RB (right border) mark the T-DNA region. The T-DNA region contains the plant selectable marker and TALEN cassette. ColE1 and pVS1-ORI/pSa-ORI mark the replication of origin in *E.coli* and *Agrobacterium*, respectively. Red arrows indicate gene orientation**.**

3.2.3 Development of CRISPR Constructs for Targeting the *MET1* Gene

The CRISPR targeting system requires two main components to work: the Cas9 protein, which possesses nuclease activity for introducing the DSB, and the sgRNA for target recognition. The *Cas9* gene was amplified from the *Streptococcus pyogenes* genome, and cloned in a modified pGreenII0029, containing the 35 CaMV promoter and a nuclear targeting sequence (NTS). For the sgRNA, the cassette architecture was designed based on the work of Xing *et al.*, (2014). The promoter of tomato U6 small nuclear RNA (Accession number: X51447.1) was selected for transcription of the sgRNA. The primers used to amplify the tomato U6 promoter were designed to include: (1) a stretch of T sequence, which is the RNA Polymerase III terminating signal (Nielsen, *et al.*, 2013); (2) the tracrRNA sequence (Jinek *et al.*, 2012), which is recognised by the CAS9 protein to form a CRISPR complex; and (3)

one BsaI digestion site. The amplicons produced, were cloned into pGreenII0029 to produce the sgRNA cassette architecture as shown in Figure 3.5A. Two oligonucleotides were designed, consisting of a complementary tomato *MET1* target sequence. The oligonucleotides favoured the formation of overhangs at each end, which were annealed followed by ligation at the Bsa1 site between the TomU6 promoter and tracrRNA sequence (Figure 3.5B). The complete sgRNA cassette, consisting of the TomU6 promoter, target sequence, tracrRNA, and terminator signal, was sub-cloned into a plasmid containing the *CAS9* gene. This produced pCas9_tomMET1 (Figure 3.5C), which consisted of the expression cassette for both CRISPR components. The cassette was then sent to Enza Zaden for transformation into tomato.

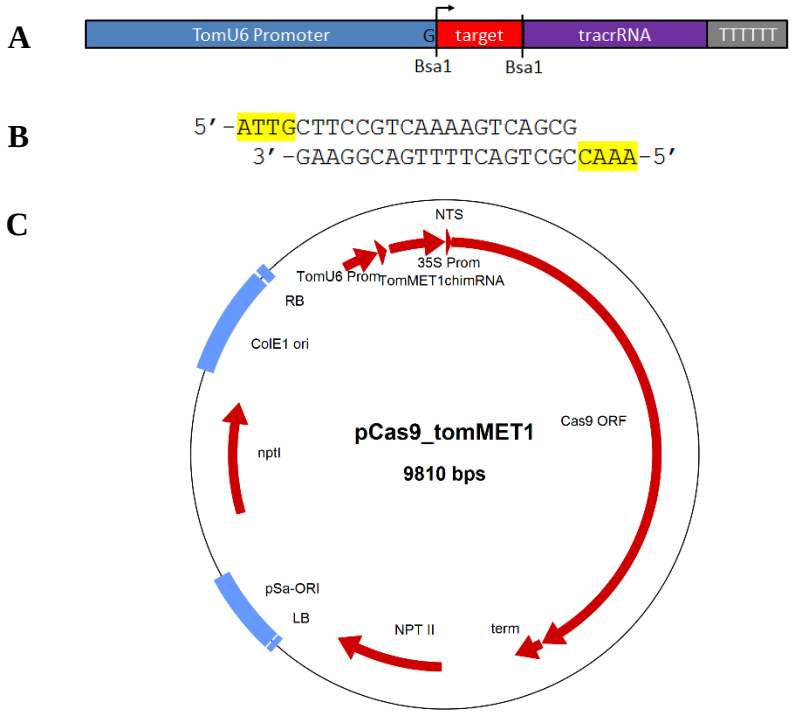


Figure 3.5: A schematic representation of the CRISPR construct for targeting tomato *MET1*. A. sgRNA construct for transcribing sgRNA under the control of the TomU6 promoter. The construct contains a stretch of thymine (T) as a signal for the Pol III terminator, and guanine (G) for a transcription signal where the transcription process starts. **B.** Complementary oligos of tomMET1 sequence, with overhangs at both ends that are complementry for ligation between the TomU6 promoter and tracrRNA. **C.** A schematic representation of the complete CRISPR construct, targeting tomMET1.

3.2.4 Production of the *MET1* Tomato Mutant

3.2.4.1 Production of the Tomato Mutant using TALEN Construct

The tomato transformations were carried out by Dr Iris Heidmann, at Enza Zaden. A total of 1402 tomato leaf explants from EZCBT1 variety were used for transforming the TomMET1TALEN construct. Seven transformed explants were grown into callus, however they failed to regenerate into the next growing stage. From this work, it was shown that the transformation was successful, but the mutation of the *MET1* gene was detrimental for regeneration of the callus. Perhaps, we can speculate that the TALEN successfully mutated both *MET1* alleles, and produced a complete *MET1* knockout. Furthermore, there is also the possibility that the TALEN itself was toxic, due to being off-target, and led to unwanted DSBs at a different location that might have been detrimental for tomato plants (Christian *et al.,* 2013).

3.2.4.2 Production of Tomato Mutants using a CRISPR Construct

Failure to produce a tomato TALEN-transformant, the CRISPR construct was used as an alternative method. Dr Iris Heidmann used a total of 1550 tomato leaf explants to transform tomato with the pCas9_tomMET1 cassette. The transformation successfully produced six stable transformants (Figure 3.6A). However, analysis using restriction enzyme digestion of the target region amplicons indicated the absence of any point mutations (Figure 3.6B), indicates by absence of DNA fragment resistance to RE digestion. The fragment was further verified by sequencing (Figure 3.6C).

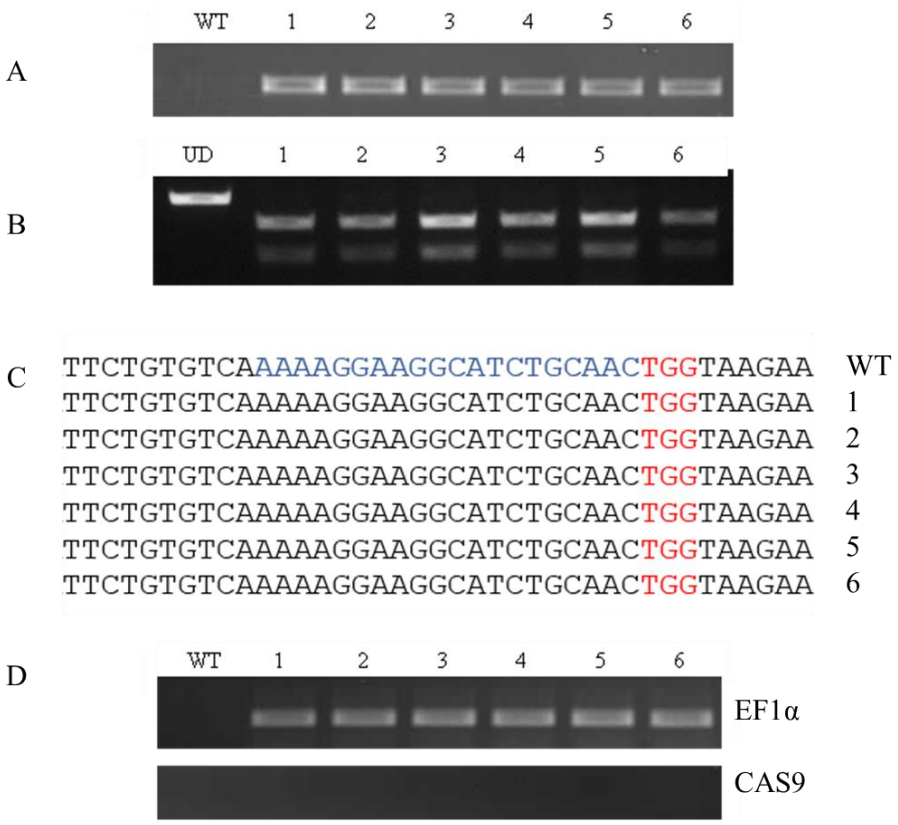


Figure 3.6: Analysis of tomato CRISPR transformants. A. Genotyping of transformants using primers that amplified the *CAS9* gene. The CRISPR construct was successfully transformed into all six lines. **B.** Digestion of *MET1* amplicons, targeted by CRISPR using SfaNi and Bsr1 for identifying mutations in the target region. Amplicons from all six samples showed no resistance to digestion, indicated by the presence of two DNA bands smaller that the undigested sample. **C.** Sequencing of amplicons from different CRISPR transformants. The sequence showing no mutagenesis was introduced by *CAS9* gene. **D.** Expression analysis of tomato transformants using smqRT-PCR. Tomato *EF1α* was used as internal expression control. The *CAS9* gene was not expressed in all six lines. WT: wild-type, UD: undigested, 1,2,3,4,5,6 independent lines of transformants. Blue sequence: CRISPR target, Red Sequence: PAM motif.

In order to investigate the cause of failure, total RNA was extracted from the transformant plants for analysis of the *CAS9* gene expression. Expression analysis via smqRT-PCR indicates that the *CAS9* transgene was silenced (Figure 3.6D). This was quite surprising, as all six of the transformants showed silencing of the transgene. Transgene silencing has been reported before (Meyer *et al.*, 1992), however, it is usually cause by DNA methylation and only occurs after several generations (Meyer *et al.*, 1992). Furthermore, DNA methylation is a randomisation event, as such silencing of six transformants in T1 generation has never been seen

before. As a re-transformation is a time-consuming process, and there were also chances that we would not produce a tomato *MET1* mutant using the TALEN strategy, we did not proceed further with this study.

3.3 Discussion

Difficulties in producing *tomMET1* mutants are not new. Our group has experienced these difficulties previously by using the IR approach (Watson 2013). Even though the possibility of off-target effects of siRNA have been reported before (Lu *et al.*, 2015), the inverted repeat sequences introduced into the tomato explants have shown high homology to the *tomMET1* gene (Watson, 2013). However, this does not eliminate the possibility of off-target mutations that might develop from siRNA biosynthesis (Senthil-Kumar and Mysore, 2011). The reasons for the lethality were not known, therefore, precision genome editing was employed for the continuation of the studies by Watson, (2013).

TALEN is known to have high specificity for its targets; it requires two sets of protein to induce mutagenesis at the selected region because in order to introduce DSBs, Fok1 must be dimerised. As an extra precautionary step, TALEN was placed under the control of the bidirectional mannopine promoter to promote low expression of the TALEN protein, and avoid, or reduce, the possibility of inducing off-target DSBs in the plant, and increase the chance of producing a heterozygous mutant. This favours mutagenesis at one allele if lethality is caused by homozygous *met1* mutant. However, even with all of these precautions, similar results were obtained as in the Watson, (2013) study; no explant was successfully regenerated. As such the CRISPR method was employed as an alternative, however, this also

failed to produce *TomMET1* transformants with a mutated *MET1* gene. Even though CRISPR is a robust method, it has been reported to be capable of introducing DSBs at non-target regions. This suggested that the lethality might be due to the depletion of MET1, rather than the effects that derived from the genome editor itself.

Even though both of the editing strategies have been reported to have high specificity and robust, there are still a lot of unknown information on the capabilities of these mechanisms especially for the CRISPR method. CRISPR relies on two main components to function both of which are synthesized in the target organism where the genome is to be edited. This has opened up a lot of questions on (i) the folding mechanisms of Cas9 protein, (ii) the efficiency of sgRNA transcription (iii) mechanisms in forming sgRNA-Cas9 complexes and (iv) the relatively low numbers of mutants produced by the techniques which requires high throughput approaches especially for identifying mutants with desired mutations.

The lack of transformants from two different gene-targeting methods, suggest that inactivating or silencing the *TomMET1* gene might have an extreme effect on viability, at least during explant regeneration. The extreme effects of *MET1* gene knockouts have not been seen in other plant species before. The *MET1* gene has been successfully silenced or knocked out in *Arabidopsis* (Kankel *et al.,* 2003; Finnegan *et al.,* 1996), tobacco (Oh *et al.,* 2009), rice (Yamauchi *et al.,* 2009), and maize (as described in Chapter 2). This suggests that disrupting the *MET1* gene expression in tomato has detrimental effects that are not seen in other plant species studied so far. Interestingly, the other plants mentioned, aside from *A. thaliana,* have undergone similar gene transfer methods, as in tomato, i.e., transformation via explant inoculation and regeneration, which suggests other plants can tolerate the

depletion of the *MET1* gene during explant regeneration. However, reduction or depletion of MET1 in plants will certainly cause plant growth abnormalities. The *met1* mutants of *A. thaliana*, tobacco, maize, and rice display abnormal growth development, produce dwarf plants, and changes in global DNA methylation patterns (Finnegan *et al.*, 1996; Oh *et al.*, 2009; Yamauchi *et al.*, 2009; Yamauchi *et al.*, 2014), which indicates that changes in plant DNA methylation can result in growth developmental consequences.

The lethal effect of disrupting *DNMT1*, a mammalian homolog of plant *MET1*, has been reported. The *dnmt1, dnmt3a, and dnmt3b* mutants indicate the importance of MTases in embryonic stem cell viability (Okano *et al.*, 1999; Li *et al.*, 1992). Furthermore, global disruption of DNA methylation patterns are lethal to mammals (Li *et al.*, 1992), even changes in focal demethylation, or hypermethylation, at imprinted loci can cause developmental abnormalities (Tycko, 2000). Therefore, we can speculate that changes in global DNA methylation patterns in tomato are detrimental for tomato growth, at least during cell differentiation.

Based on previous studies (Watson, 2014), as well as this study, the way forward in producing a tomato *MET1* mutant is to use an inducible system with high mutagenesis precision to mutate the *TomMET1* gene in a healthy transformants. This would address the uncertainty regarding the lethality of tomato transformants during explant regeneration. However, we should not ignore the recent reports about mutating RdDM components in tomato, producing a mutant with severe growth abnormalities that led to plant death (Gouil and Baulcombe, 2016).

In this chapter, we successfully developed two tools for mutating the *TomMET1* gene in tomato and transformed into tomato explant. However, due to high dependency on MET1 protein for the correct methylation pattern, callus regeneration was repressed.

Chapter 4

Plant Targeted DNA Demethylation Using the CRISPR System

4.1 Introduction

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system has been a mainstream method of targeted genome editing used by many biological researchers since it was first introduced in 2012. The ease, low cost, and speed of designing a complete system to edit a DNA region are the key factors that make it the favourite genome editing tool (Baker, 2014). CRISPR requires only two main components to function: the RNA-guided-nuclease protein, CRISPR associated protein 9 (CAS9), for induction of DNA double stranded breaks (DSB) at the DNA region of interest, and the single-guided RNA (sgRNA) to bring the CAS9 protein to the target sequence. This tool relies on the endogenous DNA repair mechanisms to repair the DSB, one of which is non-homologues end joining (NHEJ), an error-prone repair pathway that leads to a frame-shift mutation, resulting in a loss of function protein. CRISPR was said to have removed the barrier of reverse genetics and gene function studies, because there is, literally, no gene that cannot be targeted, and no organism whose genome cannot be edited. This has been demonstrated with endogenous genes in various cell types and organisms that been efficiently tweaked by using CRISPR (reviewed in Bortesi and Fischer, 2015).

The versatility of Zinc-Finger (ZF) and Transcription Activator-Like Effector (TALE), two genome editing tools before CRISPR, function as more than just sequence editors, they have influenced ideas for potential applications of CRISPR, other than as a mutation inducer. ZF and TALE were engineered to be

fused with any protein domain for a specific task. For example, in plants, Zinc-Finger Nucleases (ZFN) have been produced, by fusing nuclease with DNA binding domains to successfully cleave DNA at targeted sites (reviewed in Petolino, 2015). In mammals, ZF have also been reported to function as transcription activators, transcription repressors, ligand dependent activators, and methyltransferases (reviewed in Davis and Stokoe, 2010). Likewise, TALE is reported to have different functions when linked to different domains. In plants, TALE can function as nucleases (Wendt *et al.,* 2013), transcription activators (Morbitzer *et al.,* 2010), transcription repressors (Mahfouz *et al.,* 2012), and target demethylation domains (Maeder *et al.,* 2013). TALE also has the potential to act as targeted DNA methyltransferases, targeted chromatin modifiers, and drug screening proteins, etc. (Sun and Zhao, 2013).

The same versatility is seen with CRISPR, as suggested in a review by Barrangou and Doudna, (2016). The CAS9 protein, can have different functions when fused with different functional domains, as long as its nuclease activity is neutralised. With ZF and TALE, the nuclease domain is linked to a DNA-binding domain, whereas with CRISPR, the CAS9 protein itself contains the nuclease domain. When fused with different proteins or domains, CRISPR can function beyond a genome-editing tool. For example, it can be used as: (1) a gene activator (fused with a transcription activator), (2) a gene repressor (fused with a transcription repressor), (3) cargo delivery (fused with a DNA-binding protein to deliver diverse cargos to specific genomic locations), and (4) as an epigenome editor (fused with a demethylation domain or chromatin modifier domain) (Bortesi and Fischer, 2015).

The success of TALE as a targeted demethylator in mammalian cells (Maeder *et al.*, 2013) has given us inspiration for exploiting the potential of CRISPR for use as a plant-targeted epigenome editor. In previous chapters, DNA methylation was induced by silencing and targeting the key plant MTase, the *MET1* gene. The effects of mutating and silencing the *MET1* gene resulted in demethylation of plant genomes. In this chapter, we developed new, targeted demethylation CRISPR tools to demethylate specific regions in the plant genome, without interfering with endogenous methyltransferases genes. The nuclease domains of CAS9 were mutated to produce the deactivated CAS9 (dCAS9) protein and the mammalian 5-methyl-cytosine (5mC) oxidation protein, Ten-Eleven Translocation methyl-cytosine dioxygenase (TET), was fused with dCAS9, to enable the transport of the TET protein to targeted regions for oxidisation of the epigenetic marks. Regions with stable and densely methylated in plants are selected as targets for proof-of-concept for demethylation activity.

4.2 Results

4.2.1 *ncRNA 15242* (AT4G15242) and *PI4Kγ3* (AT5G24240) are ideal targets for CRISPR demethylation.

The construction of plant targeted CRISPR demethylation systems is possible because of the availability of an epigenome database for *Arabidopsis*. Regions with a dense and stable methylation pattern are an ideal target to test the demethylation efficacy of the CRISPR system. Suitable targets were screened from the 1001 *Arabidopsis* genomes, methylomes, transcriptomes, and the physical maps database (http://neomorph.salk.edu/1001.aj.php) for dense methylation patterns and the

presence of mRNA, as indicators of gene expression. A second database, the Transgenerational Inheritance of Methylation Variants database (http://neomorph.salk.edu/30_generations/browser.html), was used to confirm the stability of the methylation patterns at the regions, across 30 generations. Two target region were identified, (1) the *ncRNA 15242*, which were also identified based on previous findings by our group, which found a stable epiallele (Watson *et al.*, 2014), (2) the *PI4Kγ3* gene, which was found during genetic screening of RNA-seq data of MET1 over-expression in *Arabidopsis* lines, by the Meyer group (unpublished).

The *ncRNA 15242* is an epiallele with unknown function that is methylated and silenced in *A. thaliana Columbia* (*Col*), but un-methylated and expressed in *A. thaliana Wassilewskija* (*Ws*) (Watson *et al.*, 2014). This indicates that the expression of the *ncRNA 15242* gene is repressed by the presence of methylation, and this is in agreement with data from the *Arabidopsis* epigenome database, which reports dense methylation at the *ncRNA 15242* gene, with no evidence of the presence of mRNA (Figure 4.1A). Dense methylation patterns were stabled and inherited in up to 30 generations, as described in the database (Figure 4.1B).

Figure 4.1: Methylation and expression patterns of the *ncRNA 15242* gene as reported in the *Arabidopsis* epigenome database. A: Dense methylation pattern and expression profiles of the *ncRNA 15242* gene (red box). The *ncRNA 15242* gene is shown as a black arrow, which indicates the gene location and transcription direction. Dense methylation of the *ncRNA 15242* gene is shown by the coloured bar pointing upwards and downwards from the black line. There is no mRNA detected from *ncRNA 15242*, indicated by the absence of a coloured bar at the lower black line. **B**: Methylation of *ncRNA 15242* gene for 30 generations. The coloured bar pointing upwards and downwards from the black lines indicates the presence of methylation at the region (red box). The methylation patterns of the *ncRNA 15242* gene are consistently seen in Col at 30 generations. Image modified from *Arabidopsis* Epigenome database (http://signal.salk.edu).

The second target identified was AT5G24240, which encodes the PI4Kɣ3 protein, is localised in the nucleus, and is involved in autophosphorylation activity. An over-expression study of *PI4Kɣ3* gene produced plants with late-flowering phenotypes that had a high tolerance for growth in salty conditions (Akhter *et al.*, 2016). These genes were found later than the *ncRNA 15242* gene, which explain why there are more target sites was selected for this gene. *PI4Kɣ3* was found during screening of the RNA-seq data from *MET1* over-expression lines, where methylation of this gene was speculated to be controlled by an RNA-independent methylation pathway, and the methylation was introduced by MET1 (unpublished). From the *Arabidopsis* epigenome database, AT5G24240 is densely methylated and no mRNA has been detected, indicating that the gene is in a repressed state (Figure 4.2A). The

methylation pattern of this gene is consistent in almost all generations, except in line 59 of the 30th generation (Figure 4.2B).

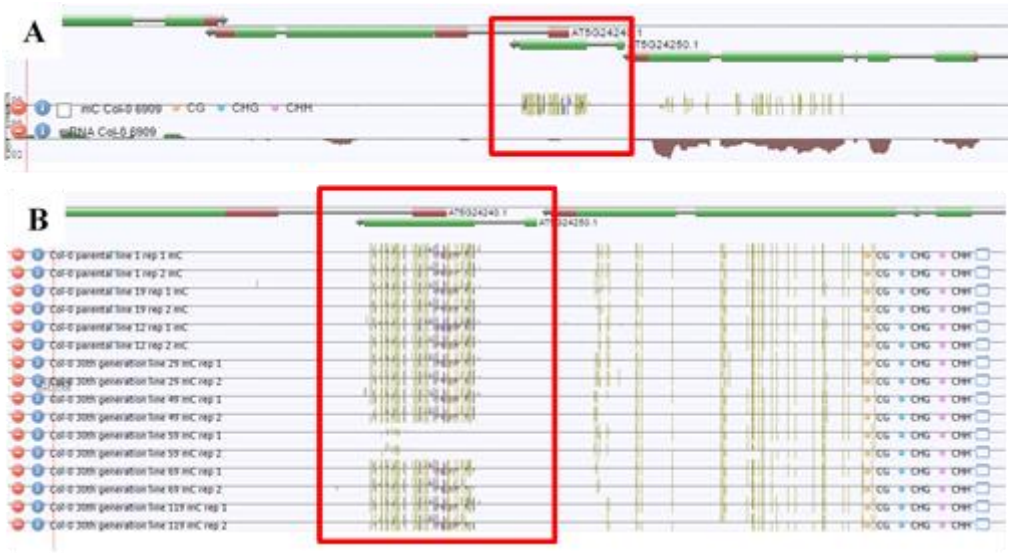


Figure 4.2: Methylation and expression patterns of the *PI4Kγ3* (AT5G24240) gene as reported in the *Arabidopsis* epigenome database. A: Dense methylation and expression patterns of *PI4Kγ3* (red box). The *PI4Kγ3* gene is shown as a green arrow, indicating the gene location and transcription direction. Dense methylation is indicated by the coloured bar pointing upwards or downwards from the black line. There is no mRNA detected from the *PI4Kγ3* gene, shown by the absence of a coloured bar at the lower black line. **B**: Methylation at the *PI4Kγ3* gene for 30 generations. The coloured bar pointing upwards or downwards from the black lines indicates the presence of methylation at the region (red box). The methylation patterns of the *PI4Kγ3* gene are consistently seen in *Col* for 30 generations, except for 30th generation of line 59 replicates 1 and 2. Image modified from *Arabidopsis* Epigenome database (http://signal.salk.edu).

Thus, these two genes were selected because their expression has been proven to be controlled by DNA methylation, and are consistently densely methylated in many generations. Stable methylation of these regions are important factors to consider for studying these genes, especially when DNA methylation patterns can to be influenced by stress (reviewed in Peng and Zhang, 2009).

4.2.2 Development of plant targeted genome demethylation tools; CRISPR-based systems

The TALE-based targeted demethylation tool was developed by producing a fusion protein, consisting of the TALE array as a DNA-binding domain (DBD) with a TET1 catalytic domain (TET1CD) (Maeder *et al.,* 2013). Inspired by success of the TALE strategy for targeted demethylation in mammalian cells, the targeted demethylation CRISPR-based system was developed, with the motivation that fusion of dCAS9 with TET1CD would also deliver the same results as the TALE-TET1 protein. Furthermore, research by the Meyer group has proven that the TET3 catalytic domain (TET3CD), a member of the TET protein family, is fully functional in plants to induce DNA demethylation (Hollwey *et al.,* 2016). Based on this foundation, we decided to use TET1CD and TET3CD as the demethylator domains. The demethylator was fused with the CAS9 protein as a tool for bringing the demethylator protein to a specific target to induce DNA demethylation. Both the TET1CD and TET3CD domains were selected to increase the chances of successful demethylation of the targets.

The CRISPR demethylation tool was developed by terminating the nuclease activity of the CAS9 protein to produce dCAS9. The CAS9 protein consists of six domains, two of which, RuvC-1 and HNH, have nuclease catalytic activity (Figure 4.3). Thus, to produce the dCAS9 protein, point mutations were introduced D10A and H840A at the domains, RuvC-1 and HNH, respectively, as described by Bikard *et al.,* (2013). The TET1CD (as described by Tan and Shi, 2012) and the TET3CD (as described by Hollwey *et al.,* 2016) gene sequences were fused with the *dCAS9* gene, after removal of the *dCAS9* gene stop codon to allow continuous transcription for

producing the dCAS9-TET1 and dCAS9-TET3 fusion proteins. Mutating the nuclease domain and removing the stop codon was performed using appropriate primers to produce fragments containing mutations, followed by sub-cloning into the same constructs to replace the original sequence. The point mutations and stop codon removal are shown in a simplified diagram in Figure 4.4. The construct developed in Chapter 3, the modified pgreenII0029 with *CAS9* gene, was used for making the *dCAS9* gene. The *dCAS9* gene was sequenced for verification. The fusion genes were place under the control of the 35S CaMV promoter with a nuclear targeting sequence (NLS) located before the *dCAS9* gene to ensure the fusion proteins, dCAS9-TET1 and dCAS9-TET3, could enter the nucleus. Several control constructs were also made as demethylation controls (Table 4.1).

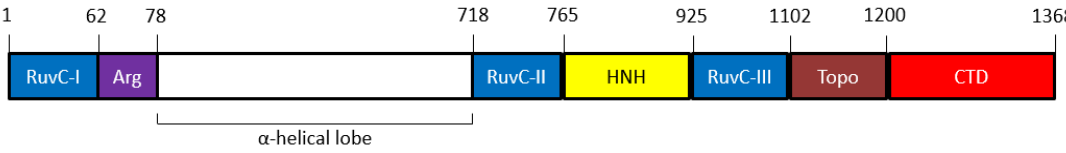


Figure 4.3: A schematic representation of the CAS9 domains architecture. CAS9 consists of six domains with two of the domains, RuC-1 and HNH, having nuclease catalytic activity to induced DNA DSB. Image extracted from Jinek *et al.*,, (2014).

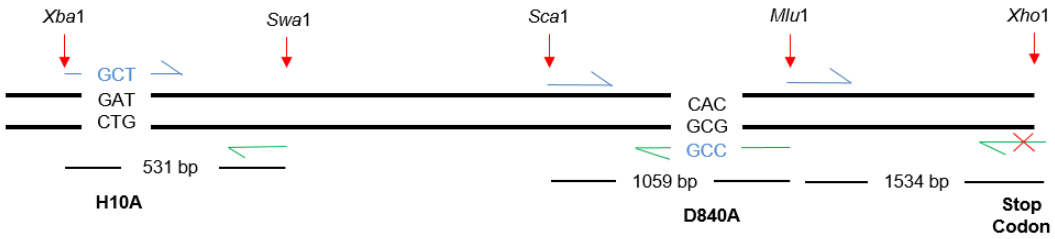


Figure 4.4: Schematic representation of primer locations for mutating the nuclease domains and removing the CAS9 gene stop codon. Three pairs of primers were used for introducing mutations at H10A, D840A, for and removing the stop codon. For the H10A mutation, the forward primer was designed with a mutated sequence and the amplicons produced were used to replace the 531bp region of the CAS9 gene, using Xba1 and Swa1 restriction enzymes. For the D840A mutation, the reverse primer was designed for the mutated sequence, and the amplicons produced were used to replace the 1059bp region of the CAS9 gene using Sca1 and Mlu1 restriction enzymes. For removal of the stop codon, the reverse primer was designed without the stop codon sequence, and the amplicons produced were used to replace the 1534bp region of the CAS9 gene using Mlu1 and Xho1 restriction enzymes. The stop codon was removed to enable the production of the fusion proteins, dCAS9-TET1 and dCAS9-TET3. All of the primers were designed to include restriction enzyme digestion sites. The blue arrow indicates the forward primers; the green arrow indicates the reverse primers.

For the expression of sgRNA, the same expression cassette developed using the tomato U6 promoter as in chapter 3 was used, but because *A. thaliana* was the target genome, in this chapter the *A. thaliana* U6 promoter (Accession nu: X52528.1) was used for expression of the sgRNA. The target sequences are sequence that targeting the *ncRNA15242* and *PI4Kγ3* genes (Figure 4.5D). The location and orientation of the targets are indicated by arrows in Figure 4.6. The target sequence was assembled as previously described in section 3.2.3 of Chapter 3. One target was selected for the *ncRNA15242* gene, while four targets were selected for the *PI4Kγ3* gene (Figure 4.6). The sgRNA expression cassettes were sub-cloned into the same plasmid carrying the dCAS9-TET1 and dCAS9-TET3 expression cassettes to produce plasmids that harbouring a complete CRISPR demethylation component. The constructs produced for this study are listed in Table 4.1, and the primers used are detailed in section 7.1.6 of Materials and Methods.

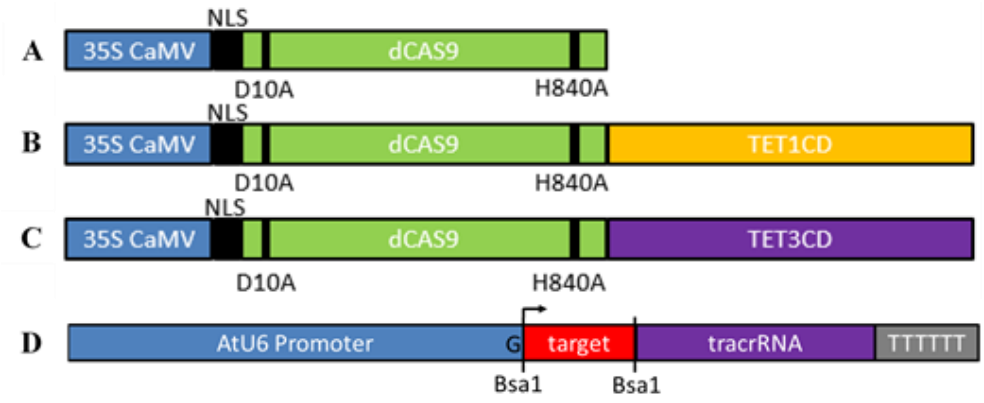


Figure 4.5: A Schematic representation of demethylation CRISPR constructs. All the constructs made were under the control of the 35S CaMV promoter. A. The control construct, only expressing the dCAS9 gene. B. The fusion dCAS9-TET1 construct. C. The fusion dCAS9-TET3 construct. D. The sgRNA expression cassette. All constructs harbouring CAS9 gene were mutated at D10A and H840A to produce a deactivated CAS9 (dCAS9) protein. Nuclear localisation sequence (NLS) was placed between the 35S CaMV promoter and dCAS9 gene. TET1CD and TET3CD were fused with the dCAS9 gene by removing the stop codon of the dCAS9 gene.

Figure 4.6: Locations of sgRNA at target genes. The red arrow indicates the location and orientation of sgRNA at *ncRNA 15242* (AT4G15242) and *PI4Kγ3* (AT5G24250) genes. The coloured bar pointing upwards and downwards from the black line indicates the presence of 5mC, at upper and lower strand, respectively. The three DNA methylation sequence types, CG, CHG and CHH are colour coded. **A**. The *ncRNA 15242* (AT4G15242) gene. **B**. The *PI4Kγ3* (AT5G24250) gene. Image modified from *Arabidopsis* Epigenome database (http://signal.salk.edu).

Constructs	Label
35S-dCAS9	d9
35S-dCAS9-TET1	d9T1
35S-dCAS9-TET3	d9T3
35S-dCAS9-TET1-AtU6ncRNA	d9T1nc
35S-dCAS9-TET3-AtU6ncRNA	d9T3nc
35S-dCAS9-TET1-AtU6PI4Kɣ3-1	d9T1PIsg1
35S-dCAS9-TET1-AtU6PI4Kɣ3-2	d9T1PIsg2
35S-dCAS9-TET1-AtU6PI4Kɣ3-3	d9T1PIsg3
35S-dCAS9-TET1-AtU6PI4Kɣ3-4	d9T1PIsg4
35S-dCAS9-TET3-AtU6PI4Kɣ3-1	d9T3PIsg1
35S-dCAS9-TET3-AtU6PI4Kɣ3-2	d9T3PIsg2
35S-dCAS9-TET3-AtU6PI4Kɣ3-3	d9T3PIsg3
35S-dCAS9-TET3-AtU6PI4Kɣ3-4	d9T3PIsg4

Table 4.1: List of constructs made, with their simplified name used in this study. All the constructs were under the control of the 35S CaMV promoter. The demethylation CRISPR used two TET genes catalytic domains, TET1CD and TET3CD, which were fused with the dCAS9 gene (d9T1 and d9T3). All sgRNA were under the control of *A. thaliana* U6 promoter (AtU6). Two genes used in this study were ncRNA 15242 (nc) and PI4Kɣ3 (PI). Eight of the constructs bind at PI4Kɣ3, four with d9T1, and another four with d9T3. Each of the four constructs contained different binding sites.

4.2.3 Production of Demethylation-CRISPR *A. thaliana Col* transformants

Four-week-old *A. thaliana Col* were transformed with the constructs produced in section 4.2.2, via the floral dip *Agrobacterium*-mediated transformation method. The seeds produced from floral dip plants were harvested and sown onto MS media plates, supplemented with kanamycin for positive transformants selection. The two-week-old positive transformants were transferred to soil for seed collection. Due to this being a preliminary study, only one positive transformant for each construct was used for the analysis. Leaves of four-week-old transformants were collected for expression and methylation analysis. DNA and RNA were extracted simultaneously from the same leaf samples to ensure expression was linked to the methylation status of the genes. Each sample was genotyped for the presence of the construct by amplifying the *dCAS9* gene. The total RNA were converted to cDNA prior to

expression analysis, using smqRT-PCR, with the *ELONGATION FACTOR 1α* (*EF1α*) gene as the expression control (Figure 4.7).

The expression level of dCAS9 varied as expected, and this might not have affected analysis of the target genes because as long as the CRISPR-demethylation was expressed, methylation should be removed by oxidation of the TET domain.

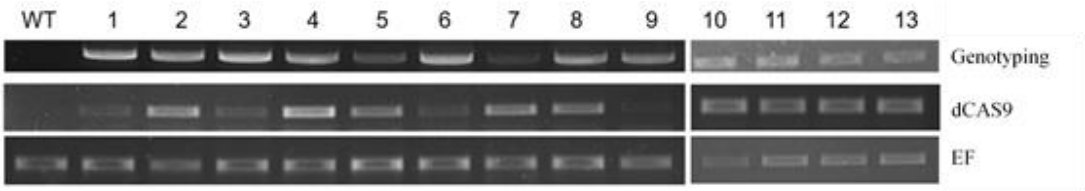


Figure 4.7: Analysis of targeted demethylation-CRISPR transformants. The positive transformants were genotyped for the presence of the dCAS9 gene and expression of dCAS9 was analysed using smqRT-PCR with EF1α expression used as the control for comparing gene expression levels. WT: wild-type A. thaliana Col; 1: d9; 2: d9T1; 3: d9T3; 4: d9T1nc; 5: d9T3nc, 6: d9T1PIsg1; 7: d9T1PIsg2, 8: d9T1PIsg3; 9: d9T1P1sg4; 10: d9T3PIsg1; 11: d9T3PIsg2; 12: d9T3PIsg3; 13: d9T3P1sg4.

4.2.4 Expression and DNA methylation pattern analysis of *ncRNA 15242* and *PI4Kγ3*

Epigenetic studies have so far relied on correlations between epigenetic marks and gene expression. As mentioned previously, expression of the *ncRNA 15242* and *PI4Kγ3* genes, were repressed by the presence of DNA methylation. If DNA methylation of these genes was oxidised by the TET domain of the dCAS9-TET fusion protein, the transcripts of these genes should be detected. Thus, based on this principle, expression of the *ncRNA15242* and *PI4Kγ3* genes were examined using the smqRT-PCR method.

Expression analysis found no evidence of either *ncRNA 15242* or *PI4Kγ3* transcripts in the positive transformants, which suggests that both genes were still repressed by the presence of methylation (Figure 4.8). Absence of transcription does not demonstrate that the targeted demethylation failed. For some genes, whole gene sequences need to be freed from DNA methylation, or at least the key methylation repressor needs to be removed before the transcription machinery can bind. Most of the key methylation that controls gene expression is located at the promoter region, and near the start codon. This kind of repressor has been seen in mammals at the Estrogen Receptor Alpha (*ESR1*) gene, where a single CpG is responsible for gene repression, and further studies have revealed that the CpG is a binding site for a transcriptional repressor that attracts the Histone Deacetylase 1 (HDAC1), which leads to chromatin condensation and repression (Fürst *et al.,* 2012). For the *ncRNA 15242* and *PI4Kγ3* genes, the methylation was not located at a promoter region, and both genes were densely methylated throughout the gene body, making it difficult to demethylate the whole gene, and trigger transcription. It is even harder to recognise and target the key methylation or methylated region that is responsible for gene repression.

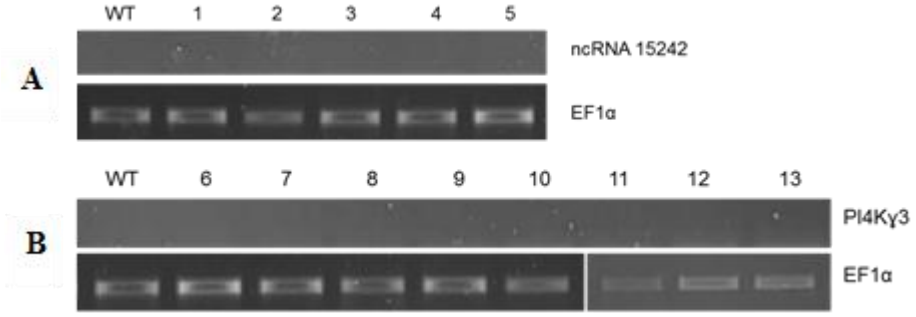


Figure 4.8: Expression analysis of ncRNA 15242 and PI4Kγ3 in demethylation-CRISPR transformants by using smqRT-PCR. No ncRNA 15242 and PI4Kγ3 transcripts were detected in any of the transformants. EF1α was used as an internal expression control. A. Expression of ncRNA 15242 in transformants. B. Expression of PI4Kγ3 in transformants. WT: wild-type Col; 1: d9; 2: d9T1; 3: d9T3; 4: d9T1nc; 5: d9T3nc, 6: d9T1PIsg1; 7: d9T1PIsg2, 8: d9T1PIsg3; 9: d9T1P1sg4; 10: d9T3PIsg1; 11: d9T3PIsg2, 12: d9T3PIsg3; 13: d9T3P1sg4.

Targeted demethylation tools are known to have limitations in the length of region that they are capable of demethylating. The TALE-based demethylation system, TALE-TET1CD, has been shown to demethylate regions up to 200bp from the TALE binding site (Maeder *et al.*, 2013). It could be speculated that the same demethylation efficacy may be applied to dCAS9-TET for demethylating *ncRNA15242* and *PI4Kγ3*. genes. If we assume that a region with a maximum length of 200bp was demethylated, this could explain the absence of *ncRNA 15242* and *PI4Kγ3* transcripts in the positive transformants, since *ncRNA15242* and *PI4Kγ3* can only be expressed if the methylation was completely removed (Watson *et al.*, 2014).

To assess the methylation status of the *ncRNA 15242* and *PI4Kγ3* genes, the methylation-sensitive PCR (MSP) method was employed. MSP was selected as a method for screening the transformants for changes in methylation before further analysis were carried out. MSP involved digestion of genomic DNA with the methyl-sensitive restriction enzyme, *Mcr*BC, prior amplification with specific

primers targeting the *ncRNA15242* and *PI4Kγ3* genes. Details of this method are described in section 2.2.9 of chapter 2.

The MSP shows that the demethylation-CRISPR successfully demethylated the *PI4Kγ3* gene in one of the transformants, d9T3P1sg4, while for other transformants, methylation was still present at the amplified regions (Figure 4.9). Even though only one of the transformants showed differences in methylation status at the target region, it does not necessarily mean the methylation status of other transformants were unchanged. This is because the MSP method requires the target fragment to be freed from methylation throughout the amplified regions. If the target region is large, there is a high probability that the DNA is digested by McrBC, as it only requires one 5-mC between the primer binding site to cause DNA digestion, and failed amplification. We speculate the presence of methylation at the *ncRNA 15242* and *PI4Kγ3* genes because of the limitations of the target demethylation tools, which can only demethylate up to 200bp regions (Maeder *et al.,* 2013). In this study, the amplified regions were large (501bp for the *ncRNA15242* gene and 378bp for the *PI4Kγ3* gene), as such, data generated using the MSP method might not reflect the true methylation status of the *ncRNA15242* and *PI4Kγ3* genes.

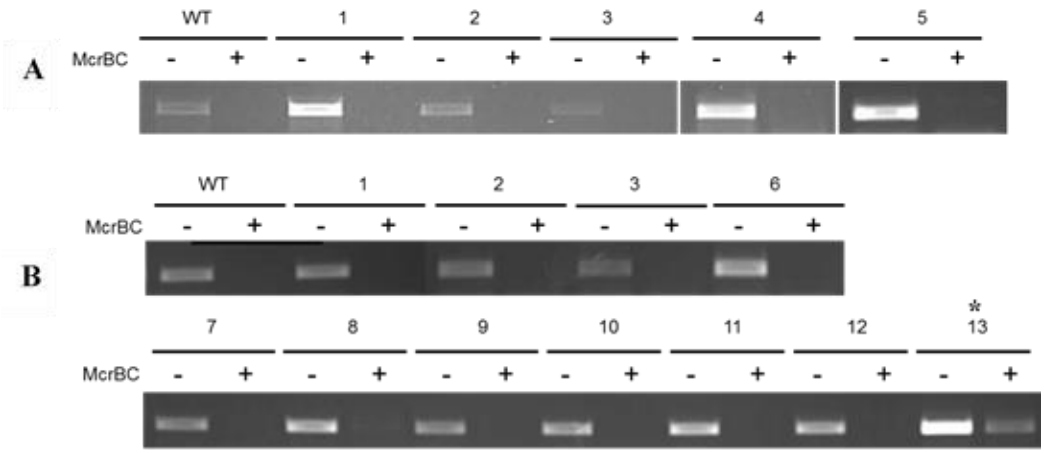


Figure 4.9: Methylation status of ncRNA15242 and PI4Kγ3 genes in demethylation-CRISPR transformants. Only sample 13 (d9T3P1sg4) shows loss of methylation in the PI4Kγ3 gene (marks by an asterisk). A. Methylation status of ncRN15242 gene. B. Methylation status of PI4Kγ3 gene. McrBC: presence of methylation-sensitive enzymes. WT: wild-type Col; 1: d9; 2: d9T1; 3: d9T3; 4: d9T1nc; 5: d9T3nc, 6: d9T1PIsg1; 7: d9T1PIsg2, 8: d9T1PIsg3; 9: d9T1P1sg4; 10: d9T3PIsg1; 11: d9T3PIsg2, 12: d9T3PIsg3; 13: d9T3P1sg4.

To confirm the loss of methylation in the *PI4Kγ3* gene, bisulfite sequencing was used. The genomic DNA of transformants that harboured the d9T3P1sg4 construct was treated with bisulfite prior to amplifying the *PI4Kγ3* gene. In addition to d9T3P1sg4, genomic DNA from wild-type (WT) and d9T1nc was also selected for bisulfite sequencing, as a methylation-control. WT and d9T1nc was used to act as the wild-type and positive transformant-control, respectively, for comparing the methylation pattern at the *PI4Kγ3* gene region. The amplicons produced were cloned into the pGEM-T Easy vector (Promega, UK) prior to sequencing. Sequences were aligned using BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html), and followed by methylation analysis using the Cytosine Methylation Analysis Tool for Everyone (CyMATE) (http://www.cymate.org).

CyMATE analysis illustrates that methylation was lost in almost all context sequence in d9T3P1sg4 (Figure 4.10). In some fragments, it was almost free of

methylation, which explains the positive amplification observed during the MSP analysis. This is interesting as it is proof that demethylation can occur in regions up to 378bp in length. Methylation patterns in the control transformant (d9T1nc) appear consistent with the WT, indicating the loss of methylation as the cause of the fusion protein localization to the *PI4Kγ3* gene.

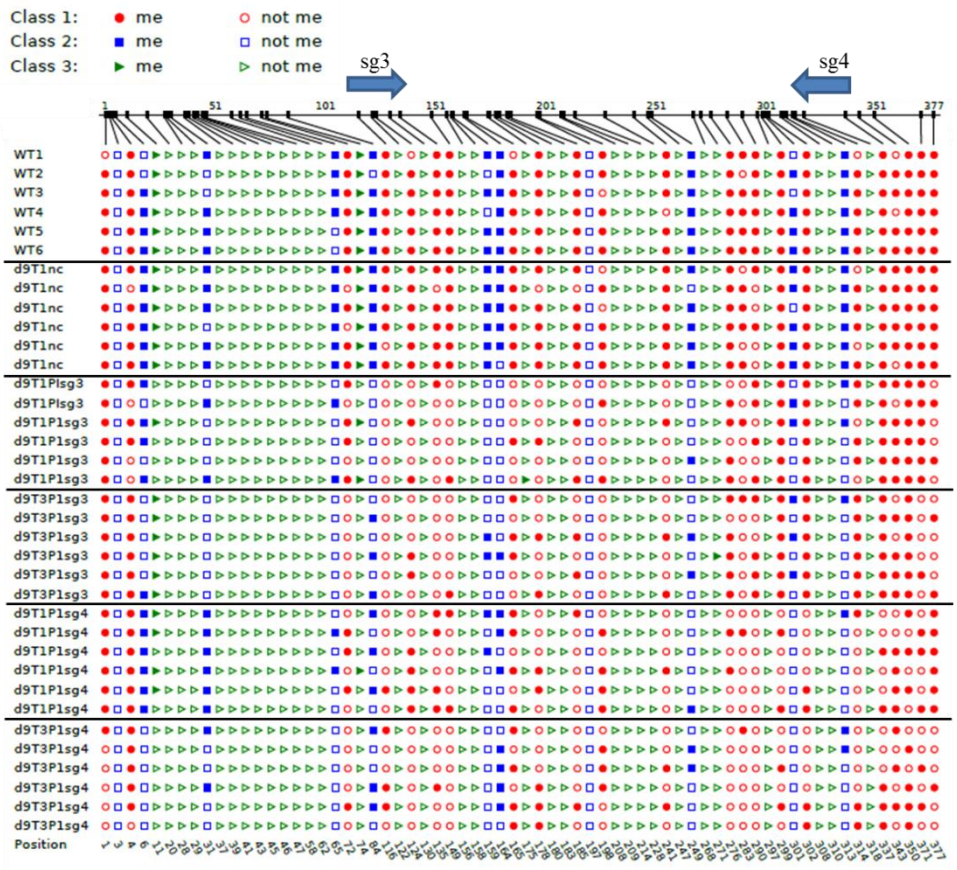


Figure 4.10: Bisulphite sequencing analysis of the *PI4Kγ3* gene in the *Arabidopsis* line expressing the d9T1nc, d9T1P1sg3, d9T3P1sg3, d9T1P1sg4, and d9T3P1sg4. The figure shows a region of the *PI4Kγ3* gene in T1 Arabidopsis lines expressing d9T1nc, d9T1P1sg3, d9T3P1sg3, d9T1P1sg4, and d9T3P1sg4 analysed by bisulphite sequencing, as described in Section 7.2.1.14 of Materials and Methods, using DNA from four-week-old seedlings. The black lines separating sequencing data for the wild-type (WT), d9T1nc, d9T1P1sg3, d9T3P1sg3, d9T1P1sg4, and d9T3P1sg4, which were analysed using CyMATE (Hetzl *et al.*, 2007). The three DNA methylation sequence types, CG, CHG and CHH are colour coded. Methylated and un-methylated bases are represented as filled and empty shapes, respectively. Cytosine positions are provided at the bottom of the figure.

As speculated earlier, the negative result generated from the MSP analysis does not necessarily demonstrate that methylation patterns at the *ncRNA15242* and *PI4Kγ3* genes were unchanged. Thus, to confirm this, several transformants harbouring the demethylation-CRISPR constructs were selected for bisulphite sequencing. The methylation pattern at the *PI4Kγ3* gene in d9T1PIsg3, d9T3PIsg3, and d9T1PIsg4 transformants were analysed. CyMATE analysis shows loss of methylation was observed in the *PI4Kγ3* gene, in all three transformants (Figure 4.10). However, regions covered by demethylation-CRISPR might be shorter than regions seen in d9T3PIsg4, which could explain the negative result obtained during the MSP analysis. While for the *ncRNA15242* gene, because no further study was carried out on this region, we can speculate that demethylation may occur at this region, as observed in other transformants. Perhaps the negative result generated by MSP, indicates the presence of methylation at the *ncRNA15242* gene, which could be due to limited coverage of CRISPR to demethylate the *ncRNA15242* gene region, which is also observed at the *PI4Kγ3* gene in some transformants. Comparison between the activity of constructs harbouring TET1 and TET3 domains, shows that they are capable of demethylating the target region (Figure 4.11), however, it is difficult to determine which domain is more efficient in demethylating its targets because of the small sample size that was used in this study.

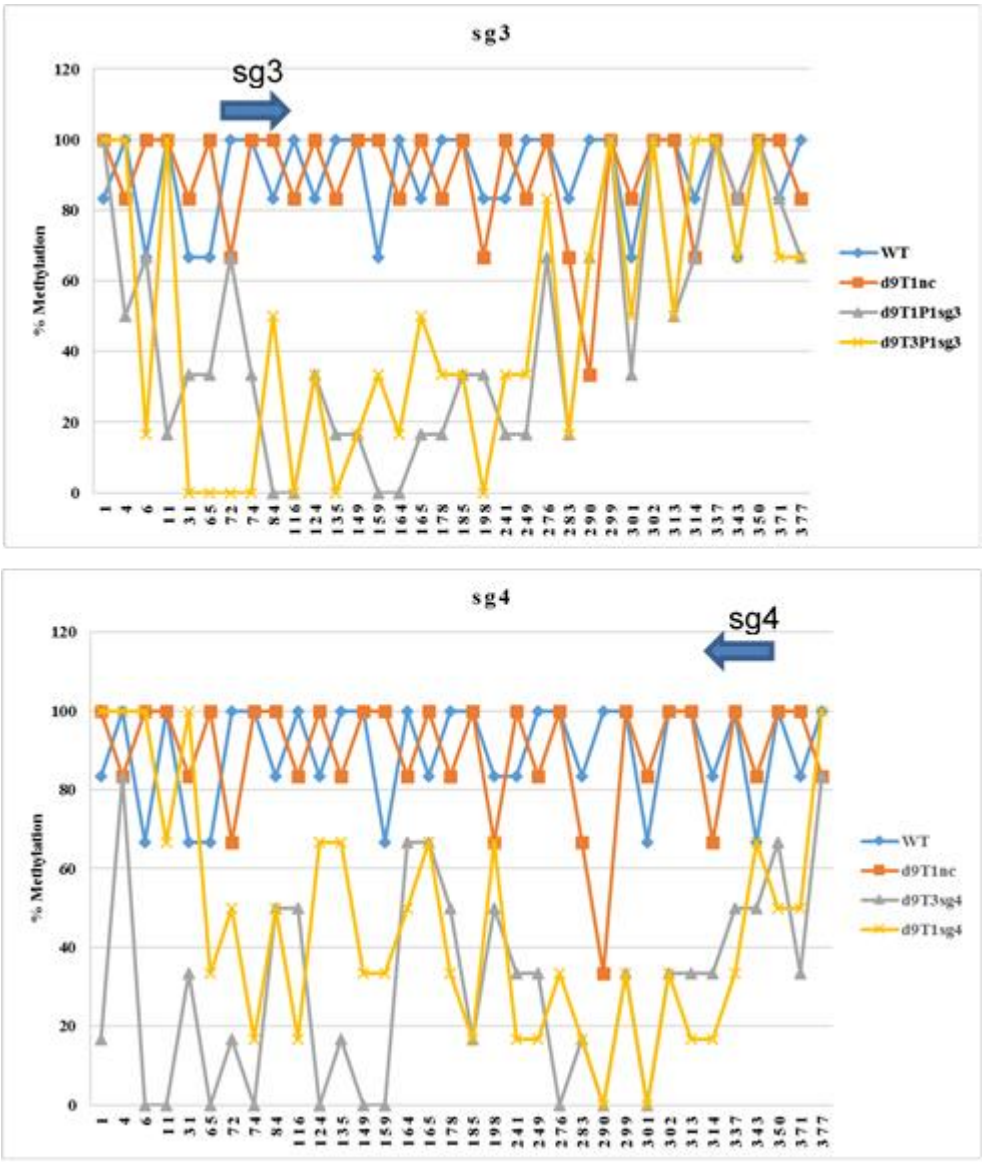


Figure 4.11: Percentages of cytosine methylation of the *PI4Kγ3* gene region in *Arabidopsis* transformant lines. Percentage of cytosine methylation of the *PI4Kγ3* gene region generated from bisulfite sequencing. Both TET1 and TET3 show they are capable of demethylating specific locations. When compared to WT and d9T1nc, the methylation status is obviously reduced at regions near the sgRNA binding site. Transformants and the control are colour coded, WT, d9T1nc, d9T3, d9T1; Cytosine positions are provided at the bottom of the figure, arrows indicate the location and orientation of sgRNA.

4.3 Discussion

4.3.1 TET1 and TET3 as plant demethylation domains

Successfully removing methylation at the *PI4Kγ3* gene, as observed in Figure 4.10, indicates that the mammalian TET gene family is active in plants. Interestingly, both TET1 and TET3 are active in plants for demethylating the 5-methylcytosine (5-mC). This supports findings by Hollwey *et al.,* (2016), which have detected catalytic activity of TET3 in demethylating rDNA loci in *A. thaliana* TET3 over-expression lines. However, the catalytic activity of the TET1 gene in plants has not previously been studied. Thus, this could be an interesting finding, as it gives researchers more options when selecting demethylase domains for plant epigenetic studies. Even though it is a new finding, the ability of TET1 to demethylate 5-mC in plants was expected because all TET gene family members share high homology at their C-terminal catalytic domain (Tan and Shi, 2012).

Naturally, demethylation of 5-methylcytosine (5-mC) in plants was carried out through a base excision repair process, which involved the DNA glycosylase domain-containing protein gene family like DEMETER (DME), REPRESSOR OF SILENCING 1 (ROS1), and DEMETER-LIKE proteins (DML, DML2, and DML3) (Ponferrada-Marín *et al.,* 2009). In mammalian cells, the demethylation process, by sequential oxidation, involves all three TET gene family members (TET1, TET2, and TET3), oxidation of 5-mC into 5-hydroxymethylcytosine (5-hmC), then further oxidation into 5-formylcytosine (5-fC), followed by oxidation into 5-carboxycytosine (5-caC) (He *et al.,* 2011b). Interestingly, in *A. thaliana*, there is no evidence of 5-hmC existing in the plant genome (Erdmann *et al.,* 2014). Thus, we

can speculate that 5-hmC, 5-fC, or 5-caC marks could be present in the *ncRNA15242* and *PI4Kγ3* genes, as found in the *Arabidopsis* TET3 over-expression lines (Hollwey *et al.,* 2016). As such, further analysis should be carried out to investigate the presence of cytosine intermediates in these genes to ensure the TET domains are responsible for the loss of methylation.

Besides the TET1 and TET3 domains, several other domains have been reported as successfully being used as epigenome editors. These domains can be categorised by their effects, such as domains that cause gene activation (e.g. VP64 and p65), repression (e.g. KRAB and SID), DNA methylation (e.g. DNMT3A and DNMT3B), DNA demethylation (e.g. TET2 and TDG), and histone modification (e.g. G9a, LSD1 and JMJD3) (reviewed in Laufer and Singh, 2015). Interestingly, for targeted demethylation, so far only three domains have been reported, the TDG, TET1, and TET2 (Chen *et al.,* 2014; Gregory *et al.,* 2013; Maeder *et al.,* 2013). Thus, with the success of demethylating the *PI4Kγ3* gene, the TET3 domain is available for researchers as an alternative domain for the targeted demethylation-CRISPR system. However, TET3 might not be the most favourable domain for making constructs because it has the largest catalytic domain of all the TET gene family members (Tan and Shi, 2012).

4.3.2 Plant Targeted Epigenetic Editor, as a way of mediating genes

The development of dCAS9-TET constructs, and the successful demethylation of the *PI4Kγ3* gene (Figure 4.11) means that the constructs developed in this study are the first plant demethylation CRISPR-based systems ever reported. During this study, at least three independent groups have reported success in developing a

CRISPR-based demethylation system, but none of them have resulted in the demethylation of plant genes; all CRISPR-based demethylation systems were developed for targeting mammalian genes (Choudhury *et al.,* 2016; Liu *et al.*, 2016; Xu *et al.,* 2016).

The concept of epigenome editing is not new; research into the editing of epigenetic marks has been progressing since the discovery of DBDs. For example, the yeast Gal4-UAS system, the NFκB DBD for affecting NFκB targets, and more recently, ZF, TALE, and CRISPR, the DBDs that can be engineered to specifically bind to almost any gene, (reviewed in De Groote *et al.*, (2012) and Thakore *et al.,* (2016). Each of the systems have their own limitations, especially when it comes to ways for assembling the DBD to target the gene of interest, which not all laboratories are capable of undertaking successfully.

The efficiency of target demethylation systems might be difficult to compare because the inability of some fusion proteins to cover substantial areas for demethylation, might be caused by locus-dependent effects, such as chromatin structure, nucleosome, DNA methylation, or other parameters that could impact on DNA binding (Maeder *et al.,* 2013). However, it has been reported that when ZF-TET1 and TALE-TET1 were used to target the same site, the *KLF4* and *HBB* genes, both managed to induce demethylation with comparable efficiency (Maeder *et al.,* 2013). While TALE-based systems have been a favourable choice for epigenetic modification compared to ZF, TALE-based targeting suffered for being sensitive to CpG methylation, which is difficult to overcome (Valton *et al.,* 2012). This made CRISPR-based technologies the choice for genome editing because it is insensitive to CpG methylation, and easier to assemble for targeting multiple sites.

Development of the CRISPR-based demethylation system also relied on the TET1 domain to demethylate the targets (Choudhury *et al.,* 2016; Liu *et al.,* 2016; Xu *et al.,* 2016). Compared to TALE-based demethylation, CRISPR-based demethylation has demonstrated high efficacy in methylation editing (Liu *et al.,* 2016). In general, the CRISPR-based system is able to demethylate CpG methylation in regions up to 300bp in length. This is interesting because almost the same length of demethylation was observed when targeting the *PI4Kγ3* gene (378bp) (Figure 4.9). However, as mentioned earlier, variations in the demethylation length can be due to locus-dependent factors.

Another important factor that determines the efficiency of demethylation is the length of the linker used to hold the TET1 domain. In TALE-TET1, lengthening the linker between the TALE array DBD and TET1 did not significantly increase the demethylation efficiency (Maeder *et al.,* 2013), whereas for the CRISPR-based system, dCAS9-TET1 fusion worked better when no linker was used to induce CpG demethylation at the target site. With the presence of a long linker, the efficiency was lower, but TET1 was still able to extend demethylation at the region where the fusion protein without the linker could not reach (Choudhury *et al.,* 2016).

Loss of DNA methylation is not necessarily followed by gene transcription. This was observed for the *PI4Kγ3* gene, where no transcripts were detected by smqRT-PCR. As speculated, the genes were still repressed because this region does not play a critical role in gene regulation. A similar reason was also speculated for the BRCA1 promoter (Choudhury *et al.,* 2016). Another potential reason could be that the long linker between the dCAS9 and TET1 masks the DNA sequence, making it unrecognisable to the transcription factor (Choudhury *et al.,* 2016). However, as

seen in Figure 4.8, in this study even without using a linker, the *PI4Kγ3* gene was still not expressed. Another potential explanation is that binding of the CRISPR complex to the genes could also result in transcription blocking, as reported in CRISPR interference (Larson *et al.*, 2013).

The losses of DNA methylation at the *PI4Kγ3* gene illustrates that developing new plant target demethylation CRISPR systems was successful. We have developed two CRISPR demethylation systems, dCAS9 fusion with TET1 and dCAS9 fusion with TET3, both of which are capable of demethylating their targets in plants. However, to acquire broader knowledge regarding the efficacy of methylation resolution, further investigation needs to be carried out by using more transformant lines as replicates. It would also be beneficial if more sgRNA could be used to target the *ncRNA 15242* and *PI4Kγ3* genes, with the aim of completely demethylating these genes, leading to gene activation.

Chapter 5

Investigating *CMT2* Over-expression and Methylation-independent Function in *Arabidopsis*

5.1 Introduction

In *Arabidopsis*, 5-mC occurs at three sequence contexts, CG, CHG, and CHH (where H represents either A, T, or C). The CG and CHG sites are symmetrical, with the opposite strand having a mirrored cytosine, whereas the CHH site is asymmetrical. DNA methylation at symmetrical and asymmetrical sites is performed by a distinct group of maintenance or *de novo* DNA methyltransferase, respectively. CHROMOMETHYLTRANSFERASES (CMTs) are an important group of methylation maintenance enzymes that are plant-specific. They were a new class of methyltransferase in land plants that arose from a duplication event in correlation with high levels of CHG methylation (Lee *et al.*, 2010). *A. thaliana* has three genes belonging to the CMT gene family, *CMT1*, *CMT2*, and *CMT3* (Pavlopoulou and Kossida, 2007), which are characterized by the presence of a chromatin organisation modifier (CHROMO) domain between the cytosine methyltransferase catalytic motifs I and IV (Figure 5.1) (Papa *et al.*, 2001).

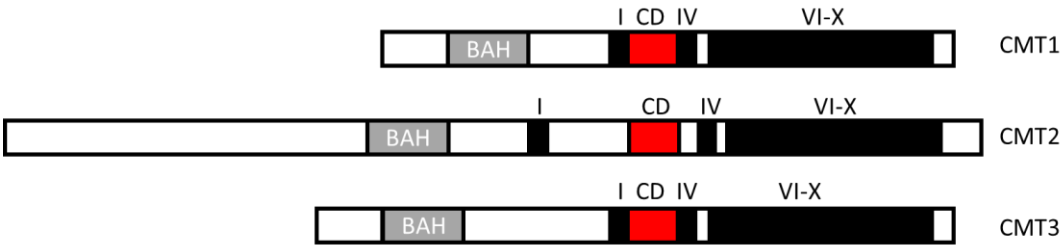


Figure 5.1: Domain Architecture of *Arabidopsis* CHROMOMETHYLTRANSFERASE (CMT) Genes. Each of the CMT genes possess one bromo adjacent homology (BAH) domain, and one chromodomain (CD), with CD being located between motifs I and IV of the methyltransferase domain. Taken from (Papa *et al.*, 2001).

Even though CMTs often methylate CHG sites, they are also reported to be responsible for methylation of CHH sites (Stroud *et al.*, 2014), for example the *CMT2* gene. Methylation is maintained by *CMT3* through a positive feedback loop involving histone H3 lysine 9 di-methylation (H3K9me2), which is catalysed by the KRYPTONITE (KYP)/ Su(var)3-9 homologue 4 (SUVH4) of lysine methyltransferases (Du *et al.*, 2012). *CMT2* has been found to be responsible for methylation of long transposable elements at the pericentromeric region (Stroud *et al.*, 2014), while *CMT1* is thought to be a nonessential gene, as it was found to be truncated by the insertion of an *Evelknievel* retroelement at exon 13, and the presence of base substitutions result in early stop codons in the majority of *A. thaliana* ecotypes (Henikoff and Comai, 1998).

The *CMT2* gene has been extensively studied in the model plant, *A. thaliana,* using a knockout *cmt2* mutant (Jackson *et al.*, 2002; Stroud *et al.*, 2013; Zemach *et al.*, 2013). A genome-wide analysis of natural variation of the *CMT2* gene, showed altered genomic CHH methylation, which was said to be involved in the adaptation of *Arabidopsis* ecotypes at higher temperate climates (Dubin *et al.*, 2015; Shen *et al.*, 2014). The *CMT2* gene has been suggested to have arose from a duplication

event within a common ancestor of all flowering plant species, to allow specialisation of the CMT2 protein for methylating CHH sites (Noy-Malka *et al.*, 2014). Recently, expression of the *CMT2* gene was found to be moderately expressed, and approximately in equal levels, in most tissues except developing seeds and endosperm (Ashapkin *et al.*, 2016).

In this chapter, we investigate the effects of the *CMT2* gene (*ac*) using an over-expression approach, rather than a knockout *cmt2* mutant. In parallel, a catalytically inactive *CMT2* gene (*in*) was made, and over-expressed, to eliminate methylation-independent effects of over-expression. The phenotypes and possible target gene expressions were analysed in the transformant lines.

5.2 Results

5.2.1 The Development of Catalytically Active and Inactive 35S:CMT2 Constructs

To investigate the roles of the *CMT2* gene in *A. thaliana*, a catalytically active *CMT2* over-expression construct was developed. In parallel, a catalytically inactive *CMT2* over-expression construct was also made for titrating the effects of methylation caused by over-expression of the *CMT2* gene.

The cDNA of *A. thaliana Col* was used to isolate a full-length *CMT2* gene. The *CMT2* gene was isolated using three pairs of primers because of the large gene size. However, several attempts at amplifying the 5’ end of the *CMT2* gene failed, possibly due to RNA degradation of the 5’ end (Yu et al, 2016). To overcome this,

genomic DNA was used as a template, which resulted in the introduction of intron one and two into the *CMT2* gene (Figure 5.2). The fragmented gene was cloned into pGreenIII 0029 with the aid of appropriate restriction enzymes, and the complete *CMT2* gene was sequenced for verification. The complete *CMT2* gene was sub-cloned between the 35S CaMV promoter and terminator in the plant transformation vector, pGreen III 0029 to produce the p35At*CMT2* construct.

To develop the *CMT2* over-expression methylation-independent construct, the DNA methylase motif IV of the *CMT2* gene was mutated to produce a catalytically inactive CMT2 protein. The point mutation at C915S was introduced as described by Hsieh, (1999) using a site-directed mutagenesis kit. The primers were designed, as suggested by NEBaseChanger (http://nebasechanger.neb.com), with one of the primers carrying the mutation. The mutated *CMT2* gene produced a vector named p35At*CMT2mut*, and the mutated region was sequenced for verification.

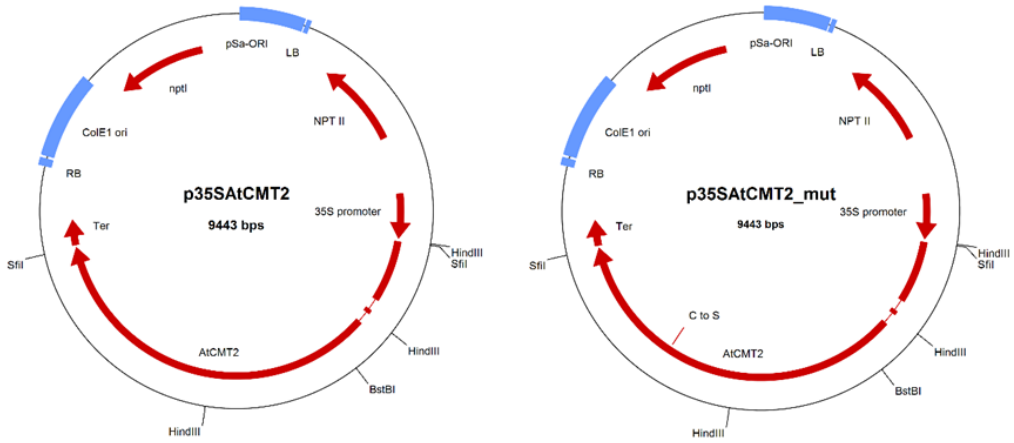


Figure 5.2 : A schematic representation of p35SAt*CMT2* and p35SAt*CMT2mut* constructs in plant transformation vector pGreen III 0029. The sequences of the primers used to produce these constructs are provided in Section 7.4.1 of Materials and Methods. LB (left boarder) and RB (right border) mark the T-DNA region. T-DNA region contains the plant selectable marker kanamycin (NPTII) and At*CMT2* expression cassette. ColE1 and pSa-ORI mark the replication of origin in *E.coli* and *Agrobacterium,* respectively. Red arrows indicate gene orientation.

5.2.2 The Production of Active *CMT2* and Inactive *CMT2 Arabidopsis* Transformant Lines

Four-week-old *A. thaliana Col* were transformed with p35SAt*CMT2* and p35SAt*CMT2mut* via floral dip *Agrobacterium*-mediated transformation. The seeds, developed from floral dip plants, were harvested and sown onto MS, supplemented with kanamycin for selection of positive transformants. Two-week-old kanamycin resistant *Arabidopsis* plants were transferred to soil for seed collection. The T0 transformants were genotyped for the presence of the construct using primers flanking the 35S CaMV promoter and *CMT2* gene. Three transformants from each construct were selected for further analysis. Total RNA was extracted from the leaves of four-week-old positive transformants to assess the over-expression of the *CMT2* gene. Expression analysis of the transformants showed that the *CMT2* gene was more highly expressed in transformant lines compared to wild-type (WT), indicating that the transgene was expressed in the transformants (Figure 5.3). Low

expression of the *CMT2* gene was detected in the WT, and was derived from expression of the endogenous *CMT2* gene (Figure 5.3). Seeds produced from the transformant lines were collected for analysis of transgenerational effects by CMT2 over-expression.

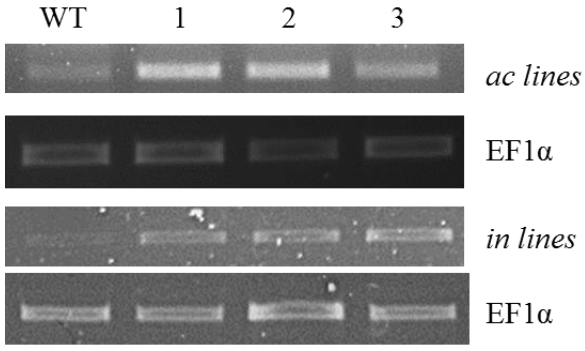


Figure 5.3: Expression analysis of *CMT2* gene in *Arabidopsis* transformant lines. Expression of *CMT2* gene was analysed using smqRT-PCR. Expression of the endogenous *CMT2* gene is lower in the wild-type (WT) compared to expression of *CMT2* transgenes in transformant lines, indicating that the transgene was expressed in the transformant lines. Three transformants were used for each construct in this analysis. *ac*: contains active CMT2 construct; *in*: contains inactive CMT2 construct. *EF1α* was used as an internal expression control.

5.2.3 Phenotype Analysis of the *CMT2* Over-expression Transformants

The growth of the transformants' shoots and roots were assessed to look for the effects of *CMT2* gene over-expression. Shoot abnormalities of four-week-old T0 transformants, which have undergone kanamycin selection were observed. To ease discussion in this chapter, the 35S*AtCMT2* lines are referred to as active CMT2 (*ac*), while the 35*AtCMT2mut* lines are referred to as inactive CMT2 (*in*). All of the phenotypes were analysed in four-week-old plants.

For *in* transformants, the shoot growth rates varied between lines (Figure 5.4), with *in*2 and *in*3 showing slower growth compared to *in*1. Furthermore, *in*1 produced

more leaves compared to the WT, and none of the *in* lines produced stems. The growth abnormality for *in*2 and *in*3 was given the name ‘dwarf phenotype’, while *in*1 was named ‘leafy phenotype’. Interestingly, this overt phenotype was observed for *in* lines, but not *ac* lines, as all three of the *ac* lines grew similar to the WT.

Figure 5.4: Shoot phenotypes of primary transformants for *in* lines (T0) compared to WT. Phenotypes for three independent lines of *in* compared to *A. thaliana Col* (WT) at four-week-old. Line *in*1 showing leafy phenotype while lines *in*2 and *in*3 showing dwarf phenotype. WT: wild type *Col; in*: three independent lines of inactive At*CMT2* transformants.

In addition to shoot abnormalities, we observed differences in seed coat colour for seeds produced by the T0 *in* lines, for example pale brown seeds were produced by *in*2 and *in*3 plants (Figure 5.5I). The seed coat colours were quantified quantitatively using ImageJ (https://imagej.nih.gov/ij/), based on Red/Green/Blue (RGB) values for each seed image, which were taken based on triplicate images of seeds. The RGB value showed that there were changes in the seed coat colour (Figure 5.5I *in*2 and *in*3). Interestingly, seed coat colour for line *in*1 was similar to that of the WT (Figure 5.5I *in*1), while for *ac* lines, the seed coat colour produced are the same as WT (Figure 5.5I).

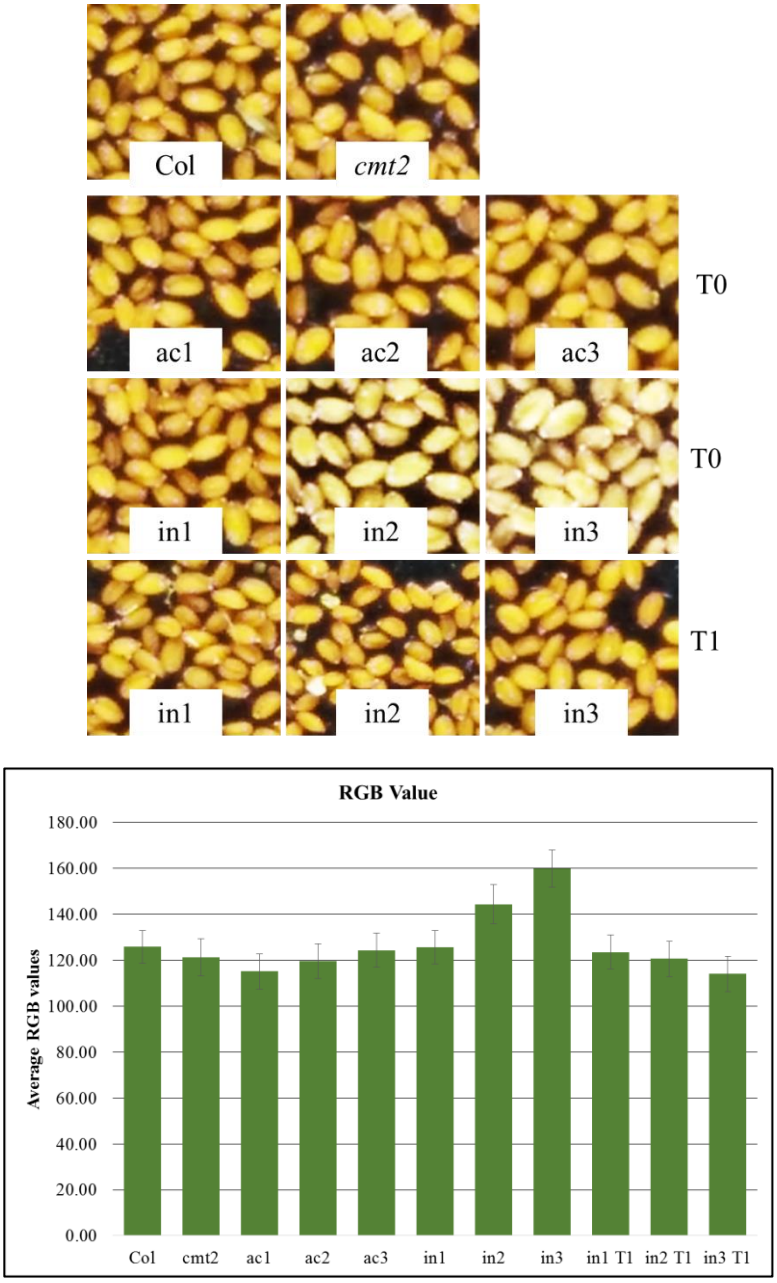


Figure 5.5: Differences in seed coat colour.

I. Seeds produced by T0 and T1 generation. Seed coat colour for *in*2 and *in*3 was pale brown compared to others. The red brown colour was restored in the next generation, even if it still possessed the transgene.

II. Quantification of seed coat colour using ImageJ based on RGB values. Higher values produce by *in*2 and *in*3 indicates a lack of red brown colour in the images. Triplicate images were used for RGB quantification. The error bars represent standard error values. A: *Col* wild-type; B: *cmt2* knockout mutant (WISCDSLOX7E02); *ac*: active *CMT2*; *in*: inactive *CMT2*; T0:T0 generation; T1:T1 generation.

Interestingly, the dwarf phenotype and pale brown seed coats observed occurred in the same lines (*in*2 and *in*3), suggesting there were changes in the gene expression of these transformants that could be involve similar genes. This could indicate changes in DNA methylation status of the genes. Furthermore, differences in the phenotypes produced by *in* lines, with *in*1 producing a leafy phenotype, but *in*2 and *in*3 producing dwarf phenotypes; the pale brown seed coat colour for *in*2 and *in*3, but WT seed coat colour for *in*1, are expected to occur in epigenetic studies. In fact, this type of variation was also seen in section 2.2.7 of chapter 2, and is caused by epigenetic chance events, normally found in epigenetic mutants. As epigenetics is about trans-generation changes, seeds from primary transformants were collected and sown to study the heritability of the growth abnormalities in the next generation.

Since there were no growth abnormalities observed for the *ac* lines in primary transformants, the seeds produced for each line were harvested and sown in parallel with *in* lines, to look for effects in the next generation. Seeds from *A. thaliana Col* and *cmt2* knock-out mutants (WISCDSLOX7E02) were sown as controls. However, the dwarf and pale brown seed coat colour that was found in T0 was lost in the T1 *in* lines (Figure 5.5 and Figure 5.6). Of the 24 seeds sown for line 3. 10 and 2, only a few plants grew with abnormalities that were different than those seen in T0, and other plants grew normally like the WT, even though they still expressed the transgene. The growth abnormalities in T1 seemed to be milder compared to T0, because they were able to produce stems (plant 1 for *in*1, *in*2, *in*3 lines) (Figure 5.6I). Moreover, *in*1 lost their leafy phenotype and grew faster than their ancestor, having almost a similar height as the WT. Interestingly, for the *ac* lines, even though their ancestor grew normally, each of the *ac* lines produced a few plants with slow growth (plant 1 for *ac*1, *ac*2, *ac*3), while other plants grew the same as the WT

(plants other than plant 1 for *in*). The *cmt2* mutants did not show any obvious phenotypes compared to the WT. The number of plants with growth abnormalities are described in Figure 5.6II.

Phenotypes that were either lost, or were milder, for T1 of the *in* lines indicate that the gene expression changes, which could be controlled by DNA methylation, are semi-stable and reversible. This is because, the dwarf phenotype of the ancestor plants has been partially or fully corrected in their progenies indicate by producing a milder dwarf phenotype or grown as WT phenotype. From this data, we could speculate that the changes in DNA methylation pattern that occurred in their ancestors might have been partially, or fully restored. Perhaps, other MTases have taken the responsible for methylating *CMT2* target region, because the *cmt2* knock-out mutant grew normally. The *cmt2* knock-out mutant included in this study could be from a later generation, in which the DNA methylation had been restored and corrected after several generations. The reversible methylation status was further supported by the T2 *in* lines, in which the transgene had been segregated away, and all of the plants grew with the same phenotype as the WT (Figure 5.6I).

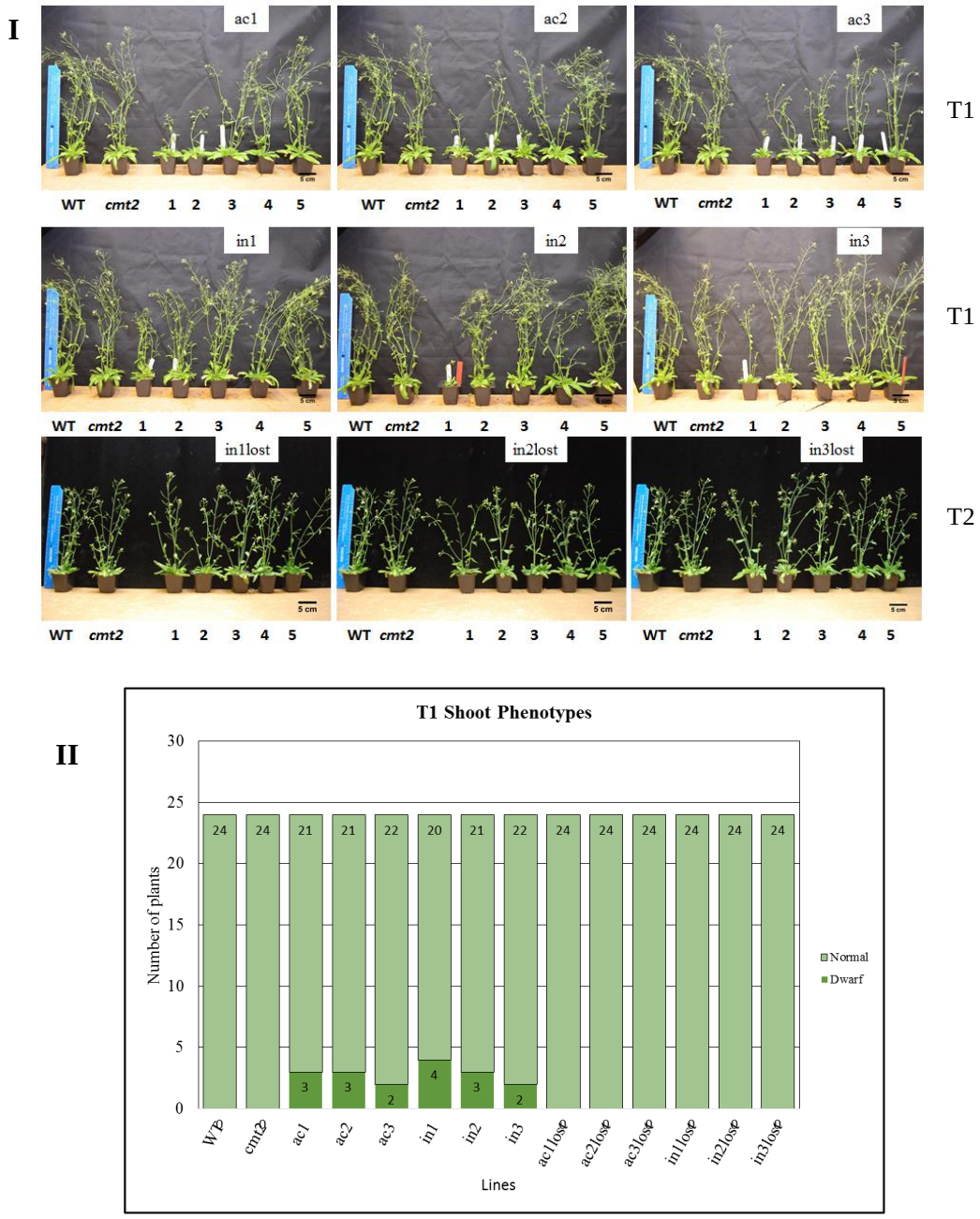


Figure 5.6: Shoot phenotypes of T1 *CMT2* over-expression lines.

I. Shoot phenotypes images of *CMT2* lines for T1 generation. Out of 24 seeds sown for *ac* and *in* lines, only a few plants grew with milder versions of the dwarf phenotype as their parent for *in* lines, while most of the other T1 plants grew like the WT. For *ac* lines, some T1 plants grew into dwarf plants that were not seen in the T0 *ac* generation. Compared to T2 of in lines that have lost transgene by segregation, the shoot phenotypes grew like the WT.

II. Numbers of plants with dwarf and normal shoot phenotypes for every line of the T1 generation (n = 24). A few plants grew with dwarf phenotypes for *ac* and *in* lines, with the presence of the transgene, while normal growth was observed for lines that had lost the transgene. WT: *Col; cmt2*: knockout *cmt2* mutant; *ac*: active *CMT2* lines, *in*: inactive *CMT2* lines, *ac*lost: active *CMT2* lines that have lost transgene; *in*lost: inactive *CMT2* lines that have lost transgene.

Interestingly, for the T1 *ac* lines, some plants grew with abnormalities that were not observed in their ancestors. The growth abnormalities in T1 could have resulted from the cumulative effects of high concentrations of CMT2 protein, which could resulted in changes of DNA methylation over time. The severity plant growth of in next generation of epigenetic mutants are common and have been described before in section 2.2.7 of chapter 2. However, because the growth abnormalities of the mutant plants is less than the plants with WT growth, the cumulative effects seen in the previous section might not be of the same level as in these transformants. Perhaps, the growth abnormalities occurring in this generation resulted from an increase in methylation by randomisation events of the *CMT2* gene over-expression.

In addition to observing the shoot growth, we also investigated the effects of *CMT2* over-expression on root growth. To study the root growth abnormalities, seeds produced by T0 lines were germinated onto horizontal MS plates, followed by transfer of the 7-day old seedlings to vertical MS plates. At least 25 plants from each line were used to score the root phenotype. Plate images were taken every week and the root lengths were measured with the help of ImageJ.

The CMT2 over-expression lines show variation in the primary root length (Figure 5.7). In general, the primary root growth for *ac* line plants were suppressed, as shown by a shorter primary root. However, some of the plants did produce primary roots with WT length, such as those found in *ac*3 (Figure 5.7). Interestingly, antagonistic root growth were observed for *in* lines, with the primary root being longer compared to WT plants (Figure 5.7II), however, variation in root length among individual plants was not seen for *in* lines. For ease of discussion, the shorter root is named ‘short phenotype’, whereas the longer root is named ‘long phenotype’.

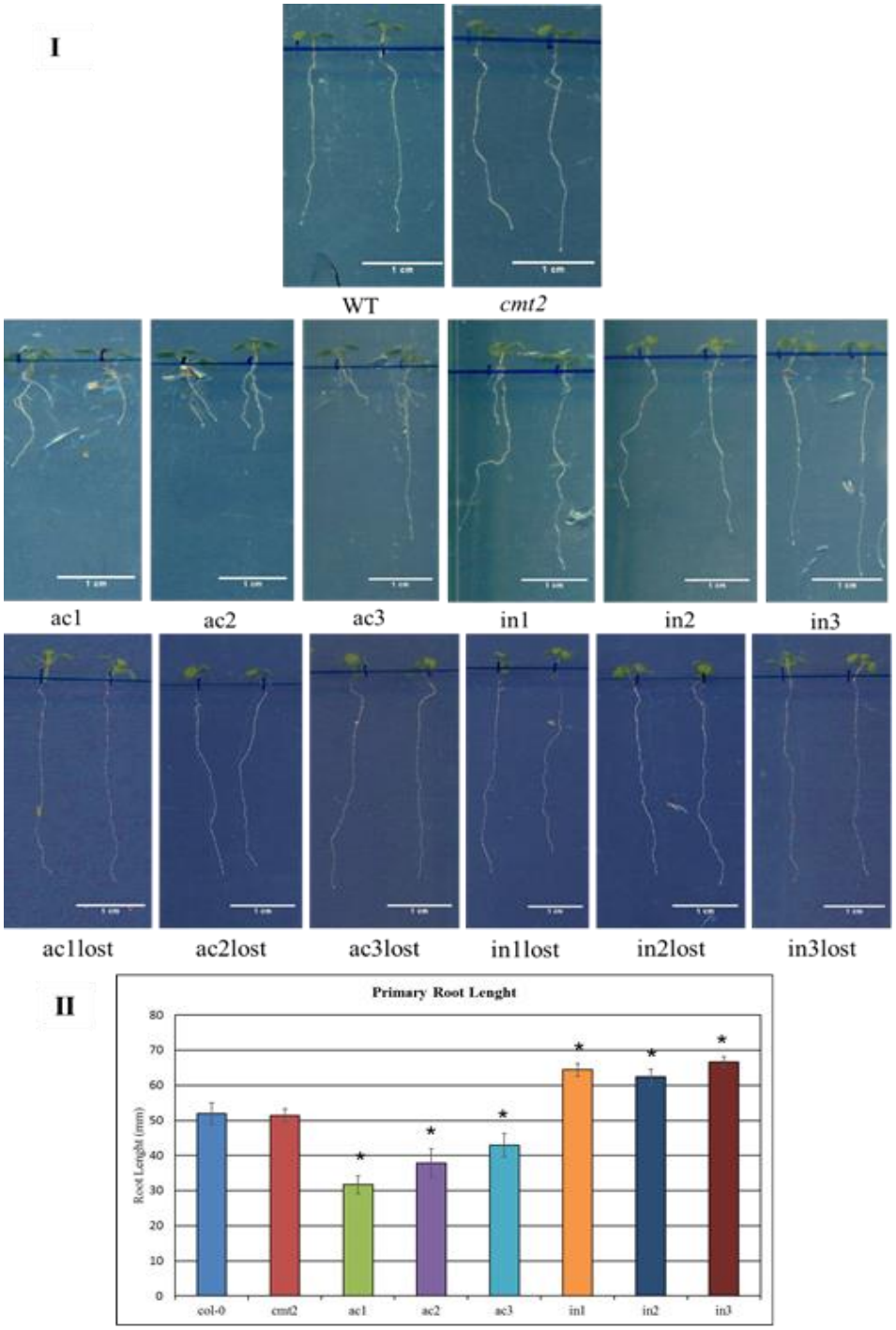


Figure 5.7: Root phenotypes for CMT2 over-expression lines.

I. Root phenotype images of T1 generation of *CMT2* lines. Only plants with phenotypes are shown in the images. Primary root growth was supressed for *ac*1 and *ac*2, whereas mixed growth was observed for *ac*3. The opposite effect was observed for *in* lines, and only normal root growth for lines that have lost the transgenes.

II. Length of primary root of T1 generation of each line was compared to WT and *cmt2*. In general, *ac* lines produced shorter primary roots than *in* lines. WT: *Col*; *cmt2*: knockout cmt2 mutant; ac: active *CMT2* lines, in: inactive *CMT2* lines, aclost: active *CMT2* lines that have lost the transgene; inlost: inactive *CMT2* lines that have lost the transgene. *n*=30. The error bars indicate standard errors. Asterisks denote significant differences between independence lines by using T-test with $p < 0.05$.

Interestingly, for plants that have lost the transgene, since this is a heterozygous line, the short and long phenotypes were lost and they grew as WT (Figure 5.7I). Interestingly, similar to the findings in the shoot studies, as the transgene is lost, the plant loses its phenotype.

The root growth abnormalities found consistently for *ac* and *in* lines, indicate that there may be a similar gene effects, as speculated for the shoot studies. This might reflect the state of the CMT2 protein, which has both active and inactive states. Active CMT2 protein might have increased methylation, whereas the inactive form may cause a reduction in methylation. Furthermore, there might be a natural CMT2 protein target that only affects the root growth genes. Even though the opposite effect between *ac* and *in* lines were unexpected, perhaps it can be explained by the active and inactive states of CMT2 protein, and the randomisation events of it binding to H3K9me2. This is supported by the long phenotypes produce by some individual plants from the *ac*3 line. Nonetheless, the opposite growth abnormalities of the roots are an interesting finding, especially when the roots of the *cmt2* mutant grew normally. Furthermore, these changes were restored when the CMT2 protein was restored back to its wild type expression.

In order to study the heritability of root phenotypes, plants with a short and long phenotype were transferred to soil for harvesting their seeds. Seeds produced were sown as T1 root analysis. Interestingly, the root phenotypes observed in T1 were lost in the T2 generation, with the *ac* and *in* lines, growing with the WT phenotype, thereby losing their ancestral phenotypes, despite still expressing the transgene.

Based on phenotypic studies, the over-expression of the inactive CMT2 protein caused severe effects on plant growth, compared to the over-expression of the active CMT2 protein. The over-expression of methylation-independent CMT2 might have induced the DNA methylation pattern by (i): limiting the availability of S-AdoMET in cells, by forming complexes with the CMT2 protein, and (ii): methylation of CHH context sequences were blocked by the CMT2 protein binding to H3K9me2. However, since the first hypothesis were not observed in the over-expression of active *CMT2* lines, we can speculate that S-AdoMET is present in high enough quantities in the cell, and there is a high probability that the phenotypes arose by the blocking of CHH methylation.

Milder and reversible phenotypes observed in the *CMT2* lines makes identification of the target genes responsible for producing the shoot and root phenotypes a more challenging task. The changes were not only semi-stable, they could also be restored. We speculate that in the later generations, the phenotype becomes milder than that seen in T1, and could be entirely lost in subsequent progenies.

5.2.4 Expression analysis of potential *CMT2* target genes

The shoot and root growth abnormalities observed in the *CMT2* over-expression lines could reflect changes in gene expression. However, finding a potential CMT2 target gene would be a difficult task, as the phenotypic changes to the plants are reversible. In order to narrow down the potential target gene, the phenotypes were used as a guideline for the gene search. Furthermore, information on methylation status of the genes from the *Arabidopsis* epigenome database of epigenetic mutant (http://genomes.mcdb.ucla.edu/AthBSseq/) was used to determine the probability

that the gene is a potential target of the CMT2 protein. In addition, a literature search and RNA-seq data from *MET1* over-expression studies, were used to find any potential CMT2 target genes, which are listed in Table 5.1

The *FWA* gene was included in this study as a control because the methylation is known to be controlled by the *MET1* gene. Two genes (AT3G01345 and AT4G15242) were selected from the RNA-sequencing (RNA-seq) because CHH methylation of these genes seems to be controlled by two methyltransferases, MET1 and CMT2. Despite none of the genes involved in the anthocyanin pathway being methylated (Stroud *et al.,* 2013), this pathway are known to be responsible for the pale brown seed coat colour (Abrahams *et al.,* 2002), as such, only one of the genes was selected based on this phenotype. For the genes selected from Groth *et al.,* (2016), there is a high chance that they were methylated by the CMT2 protein because the dense methylation of CHH context is lost in *cmt2*.

Gene identifier (TAIR)	Gene	DNA Methylation Status (Stroud et al, 2013)	Reference
AT4G25530	*FLOWERING WAGENINGEN (FWA)*	WT: Methylation at promoter *met1*: Loss methylation *cmt2*: Reduce CHH methylation	Kankel *et al.*, 2003
AT3G30720	*QUA-QUINE STARCH, (QQS)*	WT: Densely methylated *met1*: Loss CG, CHG methylation *cmt2*: Loss CHH methylation	Based on RNA-seq data
AT1G40390	*DNASE 1-LIKE SUPERFAMILY PROTEIN*	WT: Densely methylated *met1*: Loss CG, CHG methylation *cmt2*: Loss CHH methylation	Based on RNA-seq data
AT3G01345	*HYDROLASE*	WT: Densely methylated *met1*: Loss CG, CHG and CHH methylation *cmt2*: Loss CHH methylation	Based on RNA-seq data
AT4G15242	Unknown protein	WT: Densely methylated *met1*: Loss CG, CHG and CHH methylation *cmt2*: Loss CHH methylation	Based on RNA-seq data
AT3G51240	*FLAVANONE 3-HYDROXYLASE,TRANSPARENT TESTA 6 (F3H)*	WT: No methylation *met1*: No methylation *cmt2*: No methylation	Selected based on seed coat colour phenotypes
AT5G34795	*PSEUDOGENE*	WT: Densely methylated *met1*: Loss CG methylation *cmt2*: Loss CHH methylation	Groth *et al.*, 2016
AT5G33393	Unknown protein	WT: Densely methylated *met1*: Loss CG methylation *cmt2*: Loss CHH methylation	Groth *et al.*, 2016
AT3G28915	Non-LTR retrotransposon family (LINE)	WT: Densely methylated *met1*: Loss CG methylation *cmt2*: Loss CHH methylation	Groth *et al.*, 2016
AT5G35935	Copia-like retrotransposon family	WT: Densely methylated *met1*: Loss CG methylation *cmt2*: Loss CHH methylation	Groth *et al.*, 2016

Table 5.1: List of potential *CMT2* target genes. 10 genes were selected, based on RNA-seq data, literature search, and phenotypes observed in over expression of *CMT2* lines. The methylation status of each gene was checked against bisulphite sequencing of 86 silencing epigenetic mutants (http://genomes.mcdb.ucla.edu/AthBSseq/). WT: Wild-type; *met1*: *MET1* knockout mutant; *cmt2*: *CMT2* knockout mutant.

The smqRT-PCR method was employed for gene expression analysis, which was only tested for one each of the *ac* and *in* lines because they show similar growth abnormalities. Plants from lines which had lost the transgene were also selected to look for heritability effects. In order to obtained reliable data, a pooled sample from 10 shoot plants of each line were used in this study. The plants were grown in a tissue culture room at 20 ^{o}C with a 16 hr photo period. *EF1α* was used as an

expression control. The primers used are listed in section 7.1.6 of Materials and Methods.

The expression analysis showed that there were no significant differences in the expression of the genes studied (Figure 5.8). Most of the genes remained silenced, or had almost or similar expression levels compared to the WT and *cmt2*. Interestingly, even though the methylation status of this gene was expected to be different (Stroud *et al.*, 2013), it was not enough to cause changes in gene expression. Perhaps, differences in the approach used in this study, compared to Stroud *et al.*, (2013), i.e. over-expression versus silenced mutant, respectively, could be the main reason for these discrepancies. Moreover, this result was expected because the phenotypes are reversible, and it would be quite difficult to find potential CMT2 target genes, unless transcript profiling analysis was employed.

The genes that remained silenced or had similar gene expression as the WT and *cmt2*, as observed in Figure 5.8, might have been caused by the loss of CHH methylation, preventing changes to gene expression. The level of CG methylation is almost always higher than CHG and CHH methylation (typically >80%, 20-60% and <20%, respectively) (Niederhuth *et al.*, 2016). With less than 20% of CHH methylation, it is quite difficult for the CMT2 protein to be the sole protein responsible for repressing a certain gene. This is explained by several genes found by Groth *et al.*, (2016), where the genes have lost their methylation at the CHH site in *cmt2* mutants, but CG and CHG methylation were still present. Perhaps, the gene expression was controlled by CG and CHG methylation, especially when CG and CHG sites are the major methylated sequence in the genome (Niederhuth *et al.*, 2016).

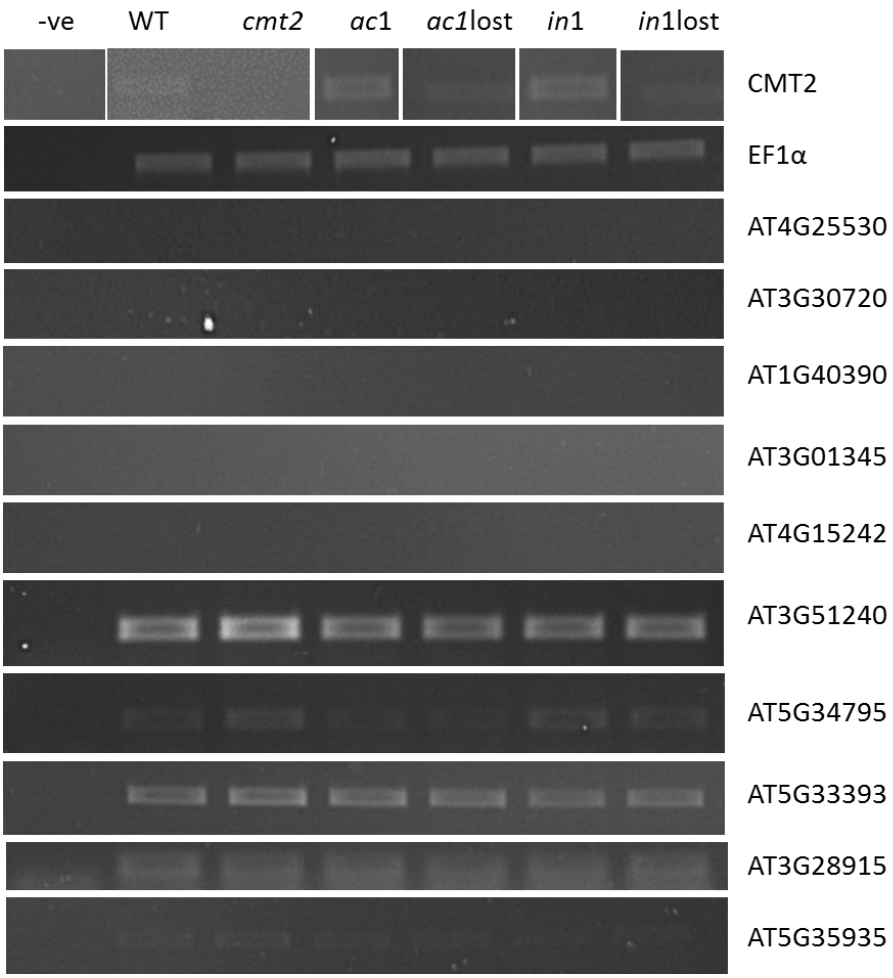


Figure 5.8: Expression analysis of potential CMT2 target genes in T1 active and inactive *CMT2* over-expression line 1 of and line 1 that have lost their transgene. The expression analysis of ten genes shows that genes either remain silenced or have similar expression to WT and *cmt2* mutant. All the genes studied have a similar expression to WT WT: wildtype; *cmt2*: knockout *cmt2* mutant; EF1α was used as an expression control.

5.3 Discussion

Gene over-expression, using a strong promoter, has become an important tool for reverse genetics, in addition to gene silencing that uses RNAi or mutation by EMS. Gene elimination would be certain to cause a mutant phenotype for some genes, whereas gene over-expression offers various phenotypes arising from multiple effects, other than loss of function (Prelich, 2012). This might explain the phenotype variations seen in the root growth development of the *ac* lines (Figure 5.7). Phenotype variations have been observed in *Arabidopsis* over-expression

transformants that express endogenous micro RNA *JAW* (Palatnik *et al.,* 2003). However, as mentioned in section 5.2.3, there is also the possibility of randomisation events, since this study involved key epigenetic proteins.

Motif I and IV are known to occur in all DNA methyltransferases, such as the CMT family (Figure 5.1) (Papa et al, 2001), which is responsible for the methylation properties of CMTs. CMT2, the most important member of the CMT family after the CMT3 protein, is responsible for methylating most of the CHH sites in *Arabidopsis, in vivo,* and have been proven to be responsible for methylating CHG and CHH sites, *in vitro* (Stroud *et al.,* 2014). Mutating motif IV, as suggested by Hsieh, (1999), has proven successful in producing a deactivated DNA methyltransferase protein, which is still able to carry other functions. A similar approach was also used by Watson, (2013), when studying MET1 over-expression effects. Even though methyltransferase activity of inactive CMT2 has not been tested, based on two studies mentioned previously, we can speculate that the ability of CMT2 to methylate cytosine has been deactivated (Hsieh, 1999; Watson, 2013). Perhaps, an *in vivo* study would be appropriate to confirm the methylase activity.

The shoot growth abnormalities observed in T0 of the *in* lines (Figure 5.4), indicate there are changes in gene expression, which could be caused by high concentrations of the CMT2 protein. Changes in gene expression could also be related to changes in methylation, since CMT2 is a methyltransferase. Based on these phenotype changes, we can speculate that the accumulation of inactive CMT2 protein may have blocked CHH methylation because the *ac* lines grew as WT, with the assumption that (1) the CMT2 concentration was at similar levels, (2) there were enough S-AdoMET supplies for the CMT2 protein, and (3) similar or same growth abnormality genes

were affected. The CHH hypomethylation states might have occurred in *in* lines by inactive CMT2 protein binding to H3K9me2, and blocking nearby CHH sites from being methylated by endogenous CMT2 protein. A protein blocking effect by gene over-expression might not have been reported previously, since it is a distinctive property of the CMT2 protein, which requires H3K9me2 binding prior to methylation. However, an almost similar mechanism has been proposed and is referred to as a dominant negative mutation where the mutant protein has a higher capacity for competing for binding partners, with loss of its second function (Herskowitz, 1987; Prelich, 2012).

Lack of shoot phenotype for the T0 generation of active CMT2 lines (Figure 5.4) does not demonstrate that there are no molecular level changes in the plants, especially when there are root growth abnormalities in the T1 generation (Figure 5.7). Perhaps, it is the randomisation effect of chance events that is commonly found in epigenetic mutants. These randomisation events were also found in root growth of the *ac*3 line, the seed coat colour of *in* lines, and the shoot growth of T1 generation for both *ac* and *in* lines. Lack of phenotypes with methylation pattern changes are common for epigenetic mutants, especially for *met1* mutants, which have been reported in rice (Yamauchi *et al.*, 2014), *Arabidopsis* (Finnegan *et al.*, 1996), and maize, as described in Chapter 2. Similar variation in phenotypes were also observed in *ddm1*, one of strong epigenetic mutant, which produced several growth abnormalities such as *clm*, *bns* and delay in flowering phenotypes (Miura *et al.*, 2001; Saze and Kakutani, 2007; Soppe *et al.*, 2000). We can speculate that the effects of CHH methylation changes to the *Arabidopsis* genome are minimal, since less than 20% of CHH sites are methylated compared to a higher methylation percentage for CG and CHG sites (Stroud *et al.*, 2014). After all, CG methylation is

known as the central coordinator for many epigenetic mechanisms (Mathieu *et al.*, 2007). This were further supported by the normal growth of *cmt2* mutants (Figure 5.6). From these data, changes in CHH methylation are required, but are not necessary to cause growth abnormalities in *Arabidopsis*.

Interestingly, growth abnormalities found in the *CMT2* lines of earlier generation were lost or became milder in the next generation, even though the plants still expressed the transgenes (Figure 5.6 and Figure 5.7). The ability of DNA methylation to be reversed in plants is not new. DNA methylation has been found to be dynamic and reversible in several plants when environmental stresses are introduced (reviewed in Meyer, 2015), such as *Arabidopsis* (Lopez *et al.*, 2015), Grapevine (Baránek *et al.*, 2015) and *Brassica* (Cao *et al.*, 2016). As for *CMT2* over-expression lines, even though plants did not undergo any stress treatment, we can speculate that the methylation pattern is reversible. Perhaps, the stress could be come from over-expression of CMT2 itself or the responsibilities for methylating CHH sites have been taken on by another MTase, such as DRM1, DRM2, or CMT3 which are all known to methylate CHH sites in the *Arabidopsis* genome (Stroud *et al.*, 2014). However, reversal of the methylation pattern does not necessarily cause a complete change, as some methylation sites are semi-stable (Melquist and Bender, 2004; Vaughn *et al.*, 2007),which can explain some of the milder phenotypes produced by the *CMT2* lines. Furthermore, the percentage of CHH methylation in the genome is low, thus, the methylation changes caused by over-expression of *CMT2* could be easily fixed, by changes to the methylation at CG and CHG sites, which might also result in gene expression changes.

However, database searches and identification of potential target genes from previous epigenetic studies, failed to identify genes with significant expression changes. This might because CMT2 was involved, but the responsibility for methylating non-CG at near gene regions were upheld by four proteins, DRM1, DRM2, CM2, and CMT3, however substantial up-regulated genes were found in the triple mutants *drm1, drm2, cmt3* instead of the quadruple mutants *drm1, drm2, cmt2, cmt3* (Stroud *et al.,* 2014). Moreover, repression of TE also requires a collective role of CMT2 and CMT3 proteins for methylating CHH and CHG, respectively (Stroud *et al.,* 2014). This suggests, that CMT2 is an important protein for CHH methylation, but demethylation at the CHH site is enough, but not significantly so, to cause differential gene expression. However, differences in the study approach (knockout versus over-expression) should not be overlooked. Perhaps, there are different roles played by CMT2 when in high concentration, especially when the CMT2 over-expression lines give rise to overt and mild phenotypes.

In this chapter, there is evidence supporting the importance of the CMT2 protein in regulating *Arabidopsis* plant growth and development. Different growth abnormalities were produced by different types of CMT2 protein. Interestingly, the growth abnormalities were semi-stable or were restored in the next generation.

Chapter 6

General Discussion

DNA methylation patterns correlate with DNA MTase activity, and disturbing the concentration of DNA MTases induces DNA methylation changes in the genome. Several strategies have been developed from this study to disturb MTase expression, which resulted in DNA methylation changes. These strategies provide an alternative method of studying MTase function in plants. In addition to interfering with DNA MTase expression, a novel strategy for inducting specific DNA demethylation changes was also developed, which could be used as a tool for studying methylated gene function, or crop improvement. In this chapter, we discuss the importance and effects of disturbing *MET1* and *CMT2* gene expression across different species, and the potential uses of the targeted demethylation tool developed from this study.

6.1 MET1, an Important DNA Methyltransferase

MET1, a key MTase that is found in all plants studied so far, has a conserved sequence in the C-terminal region, where the methyltransferase domain is located. Although *Arabidopsis* species only have one copy of the *MET1* gene for maintaining DNA methylation in its 135 megabases (MB) genome, some plants carry more than one copy. For example, maize (2500 MB), soy bean (1150 MB), sorghum (730 MB), rice (430 MB), carrot (473 MB), brassica (485 MB), and *Brachypodium* (355 MB) have been reported to possess two copies of the *MET1* gene (Bernacchia *et al.,* 1998; Candaele *et al.,* 2014; Fujimoto *et al.,* 2006; Liew *et al.,* 2013; Teerawanichpan *et al.,* 2004; Thomas *et al.,* 2014). The presence of two copies of the *MET1* gene correlates with their larger genome, which is highly abundant with TE and repetitive

sequences, indicates that the *MET1* gene is important for maintaining genome stability. However, tomato, with a genome of 781.6 MB has not yet been reported to possess two copies of the *MET1* gene (Cao *et al.,* 2014). Wheat, has a much larger genome of 2125 MB, and has been reported to have nine *MET1*-like genes (Thomas *et al.,* 2014).

Plants with two copies of the *MET1* gene were found to have specialised function, which are non-redundant to the other copies. This study found that in maize, *ZmMET1a* is more abundant than *ZmMET1b* in all tissue types tested. While in rice, *OsMET1b* was found to be highly expressed in callus, root, node, sheath, immature panicle, and inflorescence, compared to *OsMET1a,* but was not detected in differentiated tissue, mature leaves or mature panicle (Teerawanichpan *et al.,* 2004; Ahmad *et al.,* 2014). In carrot, *DcMET1a* and *DcMET1b* show differences in their relative abundance (Bernacchia *et al.,* 1998), while in barley, *HvMET1a* and *HvMET1b* have different expression levels on different days during a short day treatment (Liew *et al.,* 2013). This potentially specialised function of the MET1 protein is further supported by differences in the *ZmMET1* protein structure at the N-terminal region, which acts as a target recognition site for the MET1 protein (Pavlopoulou and Kossida, 2007).

Disturbing MET1 protein production in plants is detrimental to plant growth and development. *Arabidopsis met1-1* mutants have produced various growth abnormalities, such as late flowering, thick inflorescence stem, production of aerial rosette, and delay in senescence (Kankel *et al.,* 2003). Other epigenetic mutants also show variations in their phenotype, such as the *ddm1* mutant (Miura *et al.,* 2001; Saze and Kakutani, 2007; Soppe *et al.,* 2000), *ZmMET1* RNAi lines as described in

Chapter 2 and *CMT2* over-expression mutant as described in Chapter 5. These Phenotype variations were caused by changes to global methylation, which affected different genes, suggesting that plantsis capable of adapting to global DNA methylation changes. In this study, however, attempts to knockout the *MET1* gene in tomato using TALEN, a precision editing tool, failed to produce a *met1* mutant. Although the transformation was successful, the callus was unable to grow to the next stages. Tomato, with a genome consisting of approximately 60% repetitive elements (Mehra *et al.,* 2015), shows a high dependency for the *MET1* gene. This is not the first attempt to disturb the tomato *MET1* gene; an attempt using an inverted repeat approached has also producing callus with suppression of regeneration (Watson, 2013). These extreme effects of MET1 depletion have not been observed before in other plants. In *Arabidopsis*, the *MET1* gene has been successfully studied by various approaches, for example, the knockout mutant (Kankel *et al.,* 2003), has been silenced using inverted repeats (Chen *et al.,* 2008; Finnegan *et al.,* 1996), and over-expression (Watson, 2013). In other plants, MET1 expression has been reduced using RNAi (tobacco) (Oh *et al.,* 2009), and DNA insertion to produce a null mutant of both *OsMET1* (while in rice) (Yamauchi *et al.,* 2014). The lethality effects of disturbing MET1 expression have been documented in mammals when targeting *DNMT1* (a homologue of plant *MET1*), *DNMT3a,* and *DMT3b;* the mutant was unable to survive embryogenesis (Li *et al.,* 1992; Okano *et al.,* 1999). Interestingly mutating tomato *NUCLEAR RNA POLYMERASE D 1A* (*SIRPD1*) and *DNA-DIRECTED RNA POLYMERASE V SUBUNIT 1* (S*INRPE1*) genes encode major subunits of POLIV and POLV, respectively, a main component for RdDM mechanisms, both mutants survive the transformation stages, but the *slrpd1* mutant either dies at a later growth stage or has extreme growth abnormalities, such as epinasty, abnormal flowers, and sterility. The *slnrpe1* mutant also grows with

similar phenotype to *slrpd1* (Gouil and Baulcombe, 2016), showing that losing DNA methylation, at least at CG and CHH sequence contexts, is lethal or can be detrimental to tomato growth.

Another strategy used in this study was silencing the *MET1* gene in maize, using inverted repeats. This approach was selected for maize because maize has two copies of the *MET1* gene. As reported for *Arabidopsis* RNAi mutants, lack of phenotypes was observed in the first generation of *Zm*MET1 transformant lines, but the growth abnormalities became severe in later generations (Chen *et al.,* 2008; Finnegan *et al.,* 1996). A similar level of severity was also seen after self-fertilisation of the *ddm1* mutant (Stoke et al, 2002), and in heterozygous *OsMET1b* mutants (Yamauchi *et al.,* 2014).

6.2 Novel Plant Targeted Epigenome Editor

The idea of epigenome editing is not new. Induced changes in DNA methylation and chromatin modification have been studied since the advancement of molecular biology and understanding of epigenetic machinery in regulating gene expression. One of simplest ways of inducing DNA methylation changes for crop improvement, is to treat rapeseeds (*Brassica napus*) with 5-azacytidine, which successfully improves the protein content of the seeds (Amoah *et al.,* 2012). However, this involves a random change in global genome methylation, and it would be too laborious to screen an epi-mutant for commercial value phenotypes.

In mammals, the first targeted epigenome editing tool was developed using TALE binding protein fused with the TET1 domain to catalyse demethylation of three

endogenous genes; *Kruppel-like factor 4* (*KLF4), Rhox Homeobox Family Member 2* (*RHOXF2),* and *Beta globin (HBB*) (Maeder *et al.,* 2013). Since then, the capabilities of CRISPR as a demethylation tool has been explored. However in plants, TALE and CRISPR have only been used as a gene editor, activator, and repressor (reviewed in Bortesi and Fischer, 2015). Attempts to make plant targeted demethylation tools have not been made before, probably because of difficulties in identifying a stable target region. In this study, we successfully carried out a proof-of-concept targeted demethylation in plants, by fusing dCas9 with the mammalian demethylation proteins TET1 and TET3. This provides a new tool for gene activation based on removing DNA methylation, instead of using a transcription factor. Even though the plant-targeted demethylation tools developed, failed to induced gene activation, it is known that in mammals, multiple sgRNAs are required for a single target gene activation (Konermann *et al.,* 2014). Furthermore, there are a significant number of reports detailing the successful initiation of gene activation when removing methylation, using CRISPR-based methods in mammals (Chen *et al.,* 2014; Gregory *et al.,* 2013; Maeder *et al.,* 2013). This shows that targeted demethylation in plants can initiate gene activation and this finding has opened up the possibility of studying a wide range of EpiEffector domains (listed in Kungulovski and Jeltsch, (2016)) in plants. Recently, mammalian transcription was activated by targeting acetylation of H3K27 at the promoter of the endogenous genes, *Interleukin 1 Receptor Antagonist* (*IL1RN), Myogenic differentiation 1* (*MYOD,)* and *Octamer-binding transcription factor 4*(*OCT4*). Interestingly, the transcription activity was even higher than when using the CRISPR-based transcription factor domain (Hilton *et al.,* 2015).

6.3 *CMT2* Over-Expression Effects

Gene over-expression is another strategy used in this study for inducing epigenetic changes. Although, gene-over-expression is not new for studying genes, it is relatively uncommon for epigenetic genes, especially in plants. Agius *et al.* (2006), have over-expressed the *ROS1* gene in *Arabidopsis*, a DNA glycosylase that catalyses the removal of methyl from cytosine residues. They found a reduction in DNA methylation at the endogenous *RD29A* promoter, and activation of luciferase gene expression, driven by the *RD29A* promoter. The regional specificity of the ROS1 protein was also observed in tobacco when the *Arabidopsis ROS1* gene was over-expressed, causing demethylation at several promoters of genes that respond to conditions with high salt concentrations (Bharti *et al.*, 2015). Interestingly, Bharti *et al.*, (2015) have shown that demethylation only, is not enough to cause expression of salt-response genes, but their expression is higher than observed in the wild type, in response to high salt conditions, which suggest that the demethylation has improved the plant's response to high salt enviroment. Faster responses to stress have also been seen for TE *ONSEN*, in *ddm1* mutants, in which demethylation has favoured the *ONSEN* gene's response to heat (Cavrak *et al.*, 2014). Perhaps, demethylation provides easier access for plant stress-related transcription factors to induce gene expression. This could explain the lack of phenotypes and similar gene expression levels in the *CMT2* over-expression lines, as the effects could only be seen when induced by stress conditions.

In addition to over-expression of *ROS1*, over-expression studies of the *MET1* gene in *Arabidopsis* have also been reported (Watson, 2013). Two-types of MET1 protein were over-expressed, the endogenous *MET1* gene and the mutated *MET1* gene

(inactivated methyltransferase activity), which resulted in the reduction of *RD29A* expression, however, there was no correlation with DNA methylation patterns. *MET1* over-expression lines also produced variations in the phenotypes, with some of the plants showing late flowering time. Variations in the phenotypes that are common for epigenetic mutants were also observed in the *CMT2* over-expression lines.

Recently in *Arabidopsis*, methylation at TE regions was reported to be under the control of CMT2-dependent RdDM-independent mechanisms, which frequently methylate CAA and CTA sequence contexts (Gouil and Baulcombe, 2016), and could be used to narrow down the genes that are direct targets of CMT2 protein.

6.4 Outlook and Open Question

6.4.1 MET1 Depletion in Maize

MET1 gene expression disturbances in maize become more severe in the following generation, indicating the importance that maintenance methyltransferases having steady expression levels in maize, especially when undergoing gametogenesis, which is when epigenetic reprogramming occurs. However, since the inverted repeat targets both copies of the *MET1* genes, the dominant *MET1* in maize is still unknown. Therefore, producing maize with mutations at each copy of the MET1 gene individually, along with both copies at the same time, would certainly provide a better understanding of the unique features of the *MET1* gene in maize. Knockout mutants would also provide information on tissue specificity of the *MET1* gene, as seen in rice (Yamauchi *et al.*, 2014). High sequence similarity between the two

copies of the maize *MET1* gene could be a problem, thus, employing a targeted genome-editing tool such as TALEN or the CRISPR/Cas9 nuclease system would be the best way of introducing specific mutagenesis.

Furthermore, line 7 of ZmMET1 inverted repeat transformants would be an interesting subject for further investigation, as the abnormalities observed in the root growth have shown relation to gene expression changes. There might be a correlation of methylation with expression amongst the root development genes. The application of genetics in generating stable lines should be considered for further study of this line in order to find differences in methylation status and it is recommended to study stable lines of later generations (T3 or T4) to avoid memory stresses induced during transgene transformation.

6.4.2 Induced DNA Methylation Pattern Changes in Tomato

Understanding DNA methylation through loss-of-function seem to be impossible in tomato, because knocking out the subunit of POLIV and POLV, which is responsible for RdDM, leads to lethality (Gouil and Baulcombe, 2016). This shows the importance of DNA methylation in tomato, similarly to observations in mammals. Thus, one method to study MTases in tomato would be to use a targeted genome editing control with an inducible promoter, such as Estrogen. Estrogen-induction TALEN systems have successfully produced a heritable mutation in the *PROCERA* gene in tomato (Lor *et al.*, 2014). Furthermore, we have an option to studying heterozygous lines if homozygous mutagenesis is lethal. Producing a mutant using an induction system will also show which developmental stages the presence of *MET1* is crucial for. Alternatively, we can introduce mutagenesis to co-

factors or epigenetic genes with multiple copies, which may be less detrimental or harmful than targeting the key DNA MTases. Examples of potential target genes are one of the VIM family genes, *SIDRM6* or *SIDRM7* (homologues of the *Arabidopsis DRM1* gene) (Gallusci *et al.,* 2016; Woo *et al.,* 2008). Besides mutagenesis, an over-expression strategy would be a reliable approach for studying the lethal MTases in tomato. However, due to the nature of over-expression (which could result in gain-of-function or loss-of-function), interpreting over-expression data would be more complicated.

6.4.3 Testing and Exploiting Targeted Epigenome Editor Tools

Further testing is required for the plant targeted demethylation tools developed in this study. Several factors that need to be addressed are: (1) the length of DNA region that the tools can demethylate, (2) the efficiency of demethylating at other DNA regions, (3) the heritability and stability of demethylated regions, (4) other construct architectures such as linkers between the CAS9 and TET domains, (5) a smaller TET protein, and (6) other sgRNA structures that can improve demethylation activity.

As mentioned earlier, this opens the possibility of producing a plant epigenome editor, simply by changing the EpiEffector domain. As such, we are able to control the natural epigenetic re-setting issues at regions of interest in crop plants, such as the methylation status of *Karma* in oil palm that resulted in low yields (Ong-Abdullah *et al.,* 2015).

6.4.4 Further Studies on *CMT2* over-Expression Lines

The *CMT2* over-expression lines produced in this study need to be further tested by stress treatment. Exposing the mutants to stress will reveal more information about the CMT2 protein. Perhaps it produces plants with a faster response to certain stresses (due to methylation that was already removed at the stress-response genes), or slower response (due to heavy methylation, by over-expressing *CMT2*). Both of the effects would identify a new target gene, which has the potential to be applied in other plant species, especially to crop plants. It is also recommended to use stable transformants to further study these lines to reduce the effects of memory stresses generated from transgene transformation and antibiotic selection.

Chapter 7

Materials and Methods

7.1. Materials

7.1.1. Plant Materials

The maize (A188) seeds used in the present study were courtesy of Dr Fridtjof Weltmeier. The leaf, root, DNA, and RNA samples were prepared by Dr Fridtjof Weltmeier at KWS, before being delivered to the P. Meyer Lab, University of Leeds, in a box containing dry ice. The tomato (EZCBT1) transformations were performed by Dr Iris Heidmann at Enza Zaden, Holland. For the *Arabidopsis* studies, the mutants and wild-types used in this study were originally from the *Columbia* background. The *Arabidopsis cmt2* T-DNA mutant (WISCDSLOX7E02) was obtained from NASC, and genotyped using primers in section 7.1.6.

7.1.2. Bacterial Genomic DNA

The genomic DNA of *Streptococcus pyogenes,* strain *SF370; M1 GAS [ATCC® 700294™]* was obtained from ATCC.

7.1.3. Vectors

The pAM, p7U, and alcohol inducible pABM plasmids were provided by Dr Fridtjof Weltmeier (KWS SAAT AG, Einbeck, Germany), pGreenII0029, containing the 35S CaMV promoter and NOS terminator, the pACN and pSRNACNBin alcohol

inducible system, the modified pPIBT7, p35SPGFP, the pGreenII-TET1 and pGreenII-TET3 were provided by Dr Michael Watson (P. Meyer lab, Centre for Plant Sciences, University of Leeds, Leeds, LS2 9JT) (Caddick *et al.*, 1998; Roslan *et al.*, 2001). The TALE monomer (pNI_v2, pNG_v2, pNN_v2, pHD_v2), TALE nuclease (TALEN) backbone plasmid (pTALEN_v2 (NI), pTALEN_v2 (NG), pTALEN_v2 (NN), pTALEN_v2 (HD) were provided by Dr Adam Kupinski (Joan Boyes Lab, School of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT) (Cermak *et al.*, 2011).

7.1.4. Bacterial Strain

Plasmid cloning was carried out using *Escherichia coli* DH5α (New England Biolabs). Plant transformations were carried out using *Agrobacterium tumefaciens* GV3101::pMP90 (Koncz and Schell, 1986).

7.1.5. URLs Used for DNA and Protein Sequence Analysis

Programme Name	URL
Maize database	http://www.maizegdb.org
Arabidopsis database	https://www.arabidopsis.org
Epigenome database	https://www.plant-epigenome.org/links
Tomato database	https://solgenomics.net
PlantPAN2.0	http://PlantPAN2.itps.ncku.edu.tw/
Pfam	http://pfam.xfam.org
Bioedit	http://www.mbio.ncsu.edu/bioedit/bioedit.html
Clone Manager	http://www.scied.com/pr_cmbas.htm
Blast	http://blast.ncbi.nlm.nih.gov/Blast.cgi
Seq2Logo	http://www.cbs.dtu.dk/biotools/Seq2Logo/
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/
YASARA	http://www.yasara.org
NCBI	http://www.ncbi.nlm.nih.gov
MEGA	http://www.megasoftware.net

Table 7.1: URLs and programs used for DNA and protein sequence analysis.

7.1.6. Primer Sequences

7.1.6.1. Maize Studies

Primer Name	Primer Sequence (5' - 3') Forward	Reverse
ZmMET1 expression		
ZmMET1a	TAGAATTCAAATTAGATGACATCAAG	GCACATCAATTACATCCTCACTA
ZmMET1b	TAGAATTCAAATTAGATGACATCAAA	ATCCAGCCCCTTCAGGGATGGAT
ZmMET1	CGTGGAACACCATCCCAGAA	AGCTTGTACCAGGCCACATC
ZmFPGS	GCACAGTCCTGAAAGCATGG	GACTGCTCCAAAGAAGATGGT

Creation of 35S ZmMET1 IR constructs		
35S	AGGCCTCGTACCCCTACTCCAAAA ATGTC	GGATCCGGGCTGTCCTCTCCA AATGAAATG
ZmMET1a	GGATCCATCAGTTGGTGGGAATCC TAACCAGGG	CTGCAGTGCAGGTACCTGAAA TAAAACTC
ZmMET1b	CTGCAGGGCCACATCAGTCCGAAT GGTTATAC	GAATTCGAAGAATGGAAAATG ATGAAGAAACAG
NOS	AAGCTTAGCTCGAATTTCCCCGAT CGTTCA	GTCGACTCCCGATCTAGTAAC ATAGATGACA
Potential *ZmMET1* Target gene expression		
Copia	CGGAGGATGCATGTGGTAATAA	GGGATGTGCTCAACTCTGAAT
IAA14	TGGCAACCAACTTCTTCTGC	GTCGTTGGTTCTTCGGCTTT
MYB77	AACGAAGACGCAAACAAATCAGT	TCAGAATAATTTTGCACATGT
CA827096	CTATGAGCCCACCCCAGAAG	CCGTTGCACCCTTGATTATTA
MADS69	GAGGAACTCCTGACAAATGCTT	AGGTAAGAGGAGCGTGAGACC
GRMZM5G877259	GATGCAACCAAGGGATCGTC	TCGATTACTTCCAGCTGCATG
UE1	TCGACAGTGAGATTGACGAGG	AGACTTGGGTACTCGTTGGG
GRMZM2G001219	TCCACTTTTGACAGGACCCT	CCTGGTCCAGCAGTATGACA
GRMZM2G032198	CTTCTCCAGCTGATGGTAGCAC	CACCACCGTCTTTCTGTTTCT
GNOM	CTGTCAAGCCAGAGATGGGTG	ACGAGGTTGTGCAAGCCTTA
RTCS	CGCCGTCACCATCTCCTAC	GGGTCATAACCTGCTGCTGT
KIN2	CCTCGGTACTGTCTCTCTCC	TCCACAATGGTTCCTTTCTCA
SPS2	AAGGTCACAGGCACTCAGGTCT	GAGCCAGCAGTAGAATTAGCA
Methylation status of transposable elements		
Grande	GCCCCATAATTCGTAAGGTC	ATGCTTTATGCGATGGGTTA
SPM	TCCGAAGGGGCAGGACTAAATGAG	CAACTCTGTGAGGATTGGTGT
Prem2	GAGTCTCGAGCCGTAATCGG	TACCATGCTCTTGGGGCTTG
Opie1	GCTGCTCTCGGGAAGTGATT	AATGCCAAGTGGCTAGAGAAG
Xilon1	CTGCAATTGTCTCGCTACG	CCTTTGAAACTGCTGATGCT

Table 7.2: Oligonucleotides used for PCR reaction in maize studies.

7.1.6.2. Targeted Genome Modification Studies

Primer Name	Primer Sequence (5'-3') Forward	Reverse
Creation of TomMET1TALEN constructs		
Primers for assembling of TALE DBD as described in (Sanjana *et al.*, 2012)		
TomMETTLN1	GGGGGGATCCATAATATATCAATG GACTATAAGGACCACGAC	GGGGCTCGAGTTATGAGCGGA AATTGATCTC
TomMETTLN2	GGGGTTCGAAATAATATATCAATG GACTATAAGGACCACGACGGAGAC	GGGGATCGATTTATGAGCGGA AATTGATCTC
Creation of TomMET1 CRISPR constructs		
Cas9	TCTAGAATGGATAAGAAATACTC	CTCGAGTCAGTCACCTCCTAG CTGAC
TomU6	TCGACGCGGCCGCATAAATCTTTT TAATTTATAG	CCAAACTACACTGTTAGATTT CGCAGAGACCGAATTCGGTCT CTTTTAGAGCTAGAAATAGCA AGTTAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGC ACCGAGTCGGTGCTTTTTTTT GGATTC
TomMET1 Target	ATTGCTTCCGTCAAAAGTCAGCG	GAAGGCAGTTTTCAGTCGGCC AA
Analysis of CRISPR Targeted MET1 Region		
TomCRISPR Targeted region	TATACATCAAGATCAATGA	AGGCCATATCAGTCCGAATT
CAS9exp	TCTAGAATGGATAAGAAATACTC	TATCAGGATTTAAATCTCCCT CAATCA
TomEF1α	GAGCGATGGATGGTGAATCT	TTGTACGTGCGTCCAGAAAG

Table 7.3: Oligonucleotides used for PCR reaction in targeted genome modification studies.

7.1.6.3. Demethylation CRISPR Studies

Primer Name	Primer Sequence (5'-3')	
	Forward	Reverse
Creating of dCas9 Construct		
D10A	TCTAGAATGGATAAGGAAATACTC AATAGGCTTAGCTATCGGCACAAA TAGCGTC	TATCAGGATTTAAATCTCCCT CAATCA
H840A	TGATTATGATGTCGATGCCATTGT TCCACAAAGTTT	AAACTTTGTGGAACAATGGCA TCGACATCATAATCA
Remove Cas9 stop codon	TCTAGAATGGATAAGAAATACTC	CTCGAGACTAGTTGTCACCTC CTAGCTGACTCAA
TET3	ACTAGTAATGGAGGAGCGGTATGG AGA	CTCGAGAAGCGTAATCTGGAA CATCGTA
TET	GTCGACATGTCTCGATCCCG	GTCGACCCTAGACCCAATGGT
Creating of sgRNA Chimeric		
AtU6	GCGGCCGCCTTCGTTGAACAACGG AAACTCG	ATCACTACTTCGACTCTAGCT GAGACCGAATTCGGTCTCTTT TAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCACCG AGTCGGTGCTTTTTTTTGGAT TC
ncRNA15242	ATTGAGGTGTAGCTTGTAGTGCTG	TCCACATCGAACATCACGACC AAA
PI4Kγ3-1	ATTGACTTCCGTCAAAAGTCAGCG	TGAAGGCAGTTTTCAGTCGCC AAA
PI4Kγ3-2	ATTGGGAACGCTAAAATTTTCTGG	CCTTGCGATTTTAAAAGACCC AAA
PI4Kγ3-3	ATTGGAACAGCGGCGAATTTTGTG	CTTGTCGCCGCTTAAAACACC AAA
PI4Kγ3-4	ATTGTCCATGATCGGAAGAGACAC	AGGTACTAGCCTTCTCTGTGC AAA

Table 7.4 : Oligonucleotides used for PCR reaction in demethylation CRISPR studies

7.1.6.4. *Arabidopsis* CMT2 Studies

Primer Name	Primer Sequence (5'-3') Forward	Reverse
Creating of CMT2 constructs		
AtCMT2-1	TTTTGGATCCTCTCTGATGTT ATCGCCGGCCAAATGTGAG	TTTTCTGCAGTTCGAAATG AGCAAGATGCCAGAGACAT
AtCMT2-2	TTTTTCGAAGATCTACAAGGC TATCTGGAA	TTTTAAGCTTGTTTTCGGA CGTTTTAGGATTTC
dAtCMT2	GACCTCCATctCAAGGAATTA GTG	CCCACATATCACACCAAC
***cmt2* mutant Genotyping**		
LP and RP	LP:TGGGTTTCTTCTTCTTCACCC	RP:AGGAAAACCCAGATCTTC TGG
Potential CMT2 Target gene expression		
AtCMT2	ATCTTGCCACTTCCTGGTCG	ACAAGACGGCTCAAAGCGTA
EF1A	GCGTGTCATTGAGAGGTTCG	GTCAAGAGCCTCAAGGAGAG
AT4G25530	ATGGACACAGGCAAATGGGT	AGGCTGGTAGAGTTGGTGGA
AT3G30720	ACAAGACTCACACGGTCAGC	GTAGAACTGAAGCCCGACCC
AT1G40390	CTGAGATCCCTAAGGTGGCG	TTCTCTCAGCCTCAGTATTCA TTT
AT3G01345	CGAGGCCAAAGCTTCCAAAC	ATTGACTTCAAGGGGAGCCG
AT4G15242	CGATCTGTGCGCTTTACTCCC	GGCTTGGGAAATGGAAAGAGG
AT3G51240	CGACCTCTTCGTTCGTCAGT	CTTCTCCCTGGAGGTGACTA
AT3G51240	CGACCTCTTCGTTCGTCAGT	CTTCTCCCTGGAGGTGACTA
AT5G34795	GCCAAAACAAGCAGGAGGTG	TGAAGAAACGGACGGTCAGG
AT5G33393	AGCCTTACCGACAACCAGTG	AAAGTGTGCAGAGTTTTCAGC
AT3G28915	CAAGGGAGGACGTCTTGGTC	CAGTGGAGTACCGGGAGAGA
AT5G35935	TGTTCCATGGAGGAGGTTGC	GCCTAACCATGTCACCCCAA

Table 7.5: Oligonucleotides used for PCR reaction in *CMT2* over-expression studies.

7.2. Methods

7.2.1. DNA Analysis and Cloning Techniques

7.2.1.1. Restriction Digests

The digestion reactions were carried out using the appropriate restriction enzymes, in a final volume of 20 µl. Approximately 1 µg of DNA was digested according to the manufacturer's instructions.

7.2.1.2. Agarose Gel Electrophoresis of DNA

DNA and RNA molecules were separated electrophoretically using horizontal agarose gels (0.7-2.0%), containing 0.1 µg/ml ethidium bromide in TAE (Sambrook *et al.*, 1989). DNA and RNA was visualised on a UV trans-illuminator and captured by a digital imaging system (Syngene Bio-imager and GeneSnap).

7.2.1.3. Annealing of DNA Oligonucleotides

Two DNA oligonucleotides were mixed at equimolar concentration in annealing buffer (10 mM Tris, pH 7.5–8.0, 50 mM NaCl, 1 mM EDTA). The oligonucleotide mixtures were heated to 95 ^{o}C for 5 min. before being allowed to cool to room temperature on the work-bench.

7.2.1.4. De-Phosphorylation of DNA Fragments

The purified digested DNA fragments or plasmids were de-phosphorylated prior to ligation, to reduce the background colonies. The de-phosphorylation reaction was performed using alkaline phosphatase (CIAP) (Promega, M2825), by following the manufacturer's instructions.

7.2.1.5. Klenow Treatment of DNA Fragments

Klenow treatment was employed for the removal of the 3' overhang, and filling the 5' overhang. This was done using DNA polymerase I (NEB, M0210S) according to the manufacturer's instructions.

7.2.1.6. Ligation of Vectors and Insertion of DNA

The ligation reaction was carried out using T4 DNA Ligase (Promega, M180A) according to the manufacturer's instructions. An insert to vector ratio of 3:1 was used, and the reaction was incubated at 4 ^{o}C overnight.

7.2.1.7. Isolation of Genomic DNA from Plants

Isolation of plant genomic DNA was carried out using the modified (Vejlupkova and Fowler, 2003) method. 560 µl of extraction buffer (2 M NaCl; 200 mM Tris-HCl, pH 8.0; 7 0mM EDTA, pH 8.0; and 20 mM Sodium Bisulphite) and 50 µl of 5% Sarkosyl was added to 0.5 g of plant tissue ground in liquid nitrogen. 2 phenol:chloroform:isoamylalcohol (25:24:1) and 1 chloroform extraction was

performed after 2 hr incubation at 60 ^{o}C. The DNA was precipitated with 100% isopropanol and followed by washing with 70% ethanol. The DNA pellet was allowed to air-dry. Re-suspension was carried out using 100 µl of sterile distilled H_2O. The DNA was stored at -20 ^{o}C.

7.2.1.8. Isolation of Plasmid DNA from *E.coli*

Isolation of plasmid DNA was carried out using a modified alkaline lysis method (Sambrook *et al.,* 1989). Cells were harvested from overnight culture grown at 37 ^{o}C in liquid lysogeny broth (LB) medium (10 g/L bacto-tryptone; 5 g/L bacto-yeast extract; 10 g/L NaCl) supplemented with the required antibiotics. 100 ul of solution 1 (I (50 mM glucose; 25 mM Tris-HCl, pH 8; and 10 mM EDTA, pH 8), 200 µl of freshly prepared solution II, (0.2 M NaOH and 1% SDS), and solution III (5 M Potassium Acetate; pH 8) were added to pelleted cells. The lysate was centrifuged at maximum speed for 5 min to removed cell debris. Plasmid DNA was precipitated and washed with 100% isopropanol and 70% ethanol, respectively. Plasmid DNA was re-suspended in 100 µl of sterile distilled H_2O.

7.2.1.9. Chemically Competent Cells and Plasmid Transformation

Chemically competent *E.coli* were prepared according to (Sambrook *et al.,* 1989). *E.coli* was grown in 500 ml of LB medium (10 g/L bacto-tryptone; 5 g/L bacto-yeast extract; 10 g/L NaCl) in a shaking incubator until OD_{600} 0.4. The culture was chilled at 4 ^{o}C for 30 min followed by centrifugation at 5000 g to harvest the cells. Cells were re-suspended in chilled TSS (85% LB medium; 10% PEG 8000; 5% DMSO; 50 mM $MgCl_2$). The re-suspension was aliquoted, frozen using liquid nitrogen, and

stored at -80 ^{o}C. For heat shock transformation, 10–50 ng of DNA plasmid, or DNA from ligation reactions was added to thawed competent cells. The mixture was incubated for 10 min. followed by incubation at 42^{o}C for 90 seconds and the immediate transfer to ice for 2 min. 900 µl of SOC medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) was added to the cells, followed by incubation in a 37 ^{o}C shaking incubator for 1 hr. Positive bacteria was selected on LB medium with antibiotics, by growing overnight at 37 ^{o}C.

7.2.1.10. Polymerase Chain Reaction

PCR for genotyping, semi-quantitative PCR, and methyl-sensitive analysis was carried out using My Taq DNA polymerase (Bioline), according to the manufacturer's instructions. PCR for plasmid construction was carried out using the proof reading polymerase Q5 (NEB), according to the manufacturer's instructions.

7.2.1.11. Creation of Constructs

7.2.1.11.1. 35S ZmMET1 Inverted Repeat

The 35S promoter was amplified using PCR from pGreenII0029-35S using the 35S primers and then ligated into the *Stu*1-*Bam*H1 sites of pAM (Figure 2.7(A)). The two fragments of the *ZmMET*1 genes were amplified using the ZmMET1aF and ZmMET1aR primers for the *ZmMET1a* fragment, and ZmMET1bF and ZmMET1bR primers for the *ZmMET1b* fragment, followed by ligation into *Bam*H1-*Pst*1 and *Pst*1-*EcoR*1 sites of the pAM plasmid. The Nos terminator was amplified

from pGreenII0029-NOS, using the NOS primers and ligated into *Hind*III-*Sal*I site of pAM. For plant transformations, the 35S-*ZmMET*-inverted-repeat-Nos cascade from the pAM vector was sub-cloned to p7U with the help of the *Sfi*I restriction enzyme. This produced the construct named, p7UZMET1ir (Figure 2.7(D)).

7.2.1.11.2. Alcohol Induction *ZmMET1* Inverted Repeat

For the alcohol inducible construct, the inverted repeat from pAMZMET1ir was sub-cloned into pABM (Figure 2.7(B)) using *Bam*HI and *Sal*I. The inverted repeat cassette, which consisted of the UAS promoter, MET1 sense and antisense sequence, NOS terminator, and *GAL4* gene from pABM were sub-cloned into p7U using the *Sfi*I restriction enzyme. This produced the construct named p7UUASZMET1ir (Figure 2.7(E)). Both constructs were sent to Dr Fridtjof Weltmeier for maize transformation. The primers used are detailed in Section 7.1.6 of Materials and Methods.

7.2.1.11.3. Transcription Activator-Like Effector Nuclease (TALEN)

The constructs for targeting the *MET1* gene in maize and tomato were created using protocols described previously (Sanjana *et al.*, 2012). All of the monomers required for making the constructs were amplified from template plasmids using sets of primers (Ex-F, In-F, Ex-R and In-R primer series). The DNA fragments produced were digested with *BsmB*1 to produce compatible ends. Constructs were made using two-steps of the Golden Gate digestion-ligation reactions. In the first Golden Gate step, monomers required to target the *MET1* gene were digested with *BsmB*1 and ligated with T7 DNA ligase, to form tandem repeats of hexamers in a group of three

hexamers (1–6, 7–12 and 13–18), which were then amplified with Hex-F and Hex-R primers. Fragments of the three hexamers were subjected to the second Golden Gate reaction; *Bsa*I and T7 DNA ligase were used to ligate in-frame with the nuclease domain in the TALEN backbone plasmid. The constructs were sent for sequencing to ensure the monomers were ligated in the correct sequence. The regions targeted by the TALENs are shown in Figure 3.1. Each target required two TALEN proteins, because the nuclease operates in dimers, thus, two TALENs were produced for each target. Two TomMET1-TALENs were sub-cloned into modified pPIBT7 containing the bi-directional promoter, using *BamH*1-*Xho*1 and *Cla*1-*BstB*1. The complete TomMET1-TALEN cassette, consisting of bi-directional promoters, two TALEN genes, and terminators, was sub-cloned into pGreenII0029 using *Hind*III and *EcoR*1. This produced pGreenTomMET1TALENMET (Figure 3.3), which was sent to Enza Zaden for tomato transformation. Details of the primers used are provided in section 7.1.6.

7.2.1.11.4. Clustered Regularly Interspaced Short Palindromic Repeats *(*CRISPR)

The CRISPR/Cas system was designed, based on work done by Xing *et al.,* (2014). The *Arabidopsis* U6-26 (AT3G13855) promoter was amplified using AtU6pro-F and AtU6pro-R. The reverse primer was designed to includes the cRNA, tracrRNA and terminator sequences. The fragments were ligated into pGreenII0029, using *Xho*1 to form pGreenAtU6. The target sequence fragments were made by annealing two DNA oligonucleotides,which produced a compatible end when pGreenAtU6 was linearized by Sfi1. The target sequence fragment was fused between AtU6 promoter and cRNA sequence. . Other CRISPR targets were also subjected to the

same protocols. The *CAS9* gene was amplified from *Streptococcus pyogenes* genomic DNA using Cas9-F and Cas9-R, and replaced the GFP gene in p35SPGFP, to form p35SCas9. The p35SPGFP plasmid was selected because it has a nuclear localisation sequence (NLS) after the 35S promoter, to drive the CAS9 protein entering plant nuclei. The complete Cas9 cassette consisting of the 35S, NLS, *CAS9* gene, and Nos terminator, was sub-cloned into pGreenAtU6 with the appropriate target chimeric RNA for plant transformations (Figure 3.9). Details of the primers used are provided in section 7.1.6.

7.2.1.11.5. Demethylation CRISPR

The Q5® Site-Directed mutagenesis kit (NEB) was used to mutate the RuvC- and HNH- nuclease domains, D10A and H840A, respectively of the *CAS9* gene in p35SCas9. Mutations were performed to remove the nuclease capability of the CAS9 protein. Mutation was performed according to the manufacturer's instructions. The stop codons of *CAS9* were removed using site-directed mutagenesis. The *TET1* and *TET3* genes were amplified from the pGreenII-TET1 plasmid using TET1-F and TET1-R, while TET3 from the pGreenII-TET3 plasmid, using TET3-F and TET3-R. *TET1* and *TET3* were fused in-frame to the *Cas9* gene using *Xho*1 and *Spe*1. This produced two different constructs, p35SCas9-TET1 and p35SCas9-TET3. At the same time, guide chimeric RNA for genes: *ncRNA* (AT4G15242) and *PI4Kγ3* (AT5G24240) were made, as described previously in section 7.2.1.11.4. The chimeric RNA with AtU6 promoter was sub-cloned into p35SCas9-TET1 and p35SCas9-TET3 using *Not*1. *Arabidopsis Col-0* was transformed with all of the constructs produced (Figure 5.10). Details of the primers used are provided in section 7.1.6.

7.2.1.11.6. Catalytically Active AtCMT2 Over-expression

Total RNA was extracted from *Arabidopsis Col-0*, followed by cDNA synthesis. The *CMT2* (AT4G19020) gene was amplified from the cDNA, followed by cloning into modified pGreenII0029, between the 35S promoter and Nos terminator. Since the *CMT2* gene was too long, the genes were amplified into two fragments (one from genomic DNA, the other one from cDNA), using two pairs of primers, producing two amplicons the AtCMT2-1 and AtCMT2-2, followed by cloning to produced p35SAtCMT2 (Figure 5.11). *Arabidopsis Col-0* was transformed with this construct using the floral dip method. Details of the primers used are provided in section 7.1.6.

7.2.1.11.7. Catalytically Inactive AtCMT2 Over-expression

The Q5® Site-Directed mutagenesis kit (NEB) was used to mutate the methyltransferase motif of the *CMT2* gene. To remove the catalytic function from CMT2, the strategy described by Hsieh (1999) was exploited. Primers for dAtCMT2 were used to produce the p35AtCMT2_mutant (Figure 5.12). *Arabidopsis Col-0* was transformed with this construct, using the floral dip method. Details of the primers used are provided in section 7.1.6.

7.2.1.12. DNA Sequencing

Extracted plasmids were sent for sequencing. At least 100 ng/µl of plasmid DNA was sent for one sequencing reaction.

7.2.1.13. Methyl-Sensitive PCR Method

Methyl-sensitive PCR was carried out using the McrBC enzyme (NEB), according to the manufacturer's instructions. The gDNA was treated with McrBC prior to amplification with specific primers. Details of the primers used are provided in section 7.1.6.

7.2.1.14. Bisulphite Sequencing Method

Bisulphite treatment was performed using the EZ DNA Methylation-Lightening Kit (ZYMO Research), according to the manufacturer's instructions. Target regions were subsequently amplified by PCR, using primers specifically designed for bisulphite treated DNA. All amplicons were cloned into pGEM-T Easy (Promega), according to the manufacturer's instructions. Clones containing the insert were sent for sequencing using universal primers. Sequencing reads were aligned sing BioEdit (Hall, 1999), and analysed using CyMATE (Hetzl *et al.*, 2007).

7.2.2. RNA Work

7.2.2.1. Isolation of RNA from Plants

Total plant RNA was isolated using a standard procedure (Stam *et al.,* 2000). 0.5 g of plant tissue was ground in liquid nitrogen, followed by addition of 750 µl of RNA extraction buffer (100 mM Tris-HCl, pH 8.5; 100 mM NaCl; 20 mM EDTA; 1% sarcosyl). 2 phenol:chlorophorm:isoamyl-alcohol (24:24:1) extractions were performed, followed by precipitation with 100% isopropanol, 4 M LiCl and 3 M

NaAc pH 7.0. The RNAs were treated with the TURBO DNase kit (Ambion applied biosystems), according to the manufacturer's instructions to remove any contaminating DNA.

7.2.2.2. cDNA Synthesis

Synthesis of cDNA was performed using the Superscript II Reverse transcriptase kit (Invitrogen), according to the manufacturer's instructions.

7.2.2.3. Real Time-PCR

Real-time PCR was carried out using the BioRad CFX96 real-time C1000 thermal cycler, and using SsoFast EvaGreen supermix (BioRad) as per the manufacturer's instructions. Target gene expression was calculated as an average of 3 technical replicates, normalised to the housekeeping gene, Folylpolyglutamate synthase (Manoli *et al.*, 2012). Details of the primers used are provided in section 7.1.6.

7.2.3. Plant Work

7.2.3.1. Growing Plants

The maize seeds were soaked in sterile distilled water for overnight followed by sowing into the soil. The seeds were germinated and grown to maturity at KWS and University of Leeds greenhouse at 20 ^{o}C with 18 h/6 h (light/dark) photoperiod, unless stated otherwise. *Arabidopsis* seeds were subjected to vapour-phase sterilisation using chlorine gas (Clough and Bent, 1998). The seeds were then

planted by either sprinkling them carefully over the surface of soil, for greenhouse conditions, or over ½ MS medium, for tissue culture conditions. They were stratification at 4 ^{o}C in the dark for 2 days, before being transferred to normal growth conditions.

7.2.3.2. Plant Phenotypes and Stomata Distribution

Root and Shoot: The 7-day old *Arabidopsis* seedlings that germinated in tissue culture plates were transferred to square plates with MS medium, regularly spaced at approximately 1.0 cm apart, arranged in lines (for root phenotype), and transferred to soil for growth in the greenhouse. The plate was placed horizontally in normal growth conditions, and the phenotypes were observed every week. At least 30 plants for each line were used.

Stomata: Approximately 12 weeks of maize leaves were observed under light microscope. The same leaf growth stages were used for observing the stomata distribution. At least six images of the leaf epidermis were captured.

Seed Coat Color: Images of the seeds were analysed using ImageJ to measure the Red/Green/Blue (RGB) values, and assess the colour of the seed coats. Three images were used to increase consistency.

7.2.4. *Agrobacterium* Work

7.2.4.1. Preparation of Electro-Competent *Agrobacterium*

The electro-competent *Agrobacterium* was prepared by following the protocol described by Shen and Forde (1989). The fresh overnight cultures were inoculated in 500 ml of LB medium with antibiotic, and grown overnight at 28 ^{o}C in a shaking incubator, to OD_{600} = 0.4. The cells were harvested and washed three times with ice-cold sterile distilled H_2O. The cell pellets were re-suspended in a 5 ml solution of ice-cold sterile distilled H_2O and 10% (v/v) sterile glycerol, aliquoted, and stored at -80 ^{o}C.

7.2.4.2. *Agrobacterium* Plasmid Electroporation

The constructs prepared in section 7.2.1.11 were co-transferred with pSoup into *Agrobacterium,* using the method described by Mersereau *et al.,* (1990). Briefly, 1 mm gap cuvettes were used to electroporate the plasmid constructs, pSoup, and *Agrobacterium* electro-competent cells. Electroporation was carried out using the BioRAD Gene pulser cell-porator, with the following parameters: C = 25 µF, R = 400 Ω, 8–9 ms delay, and pulsed at V = 1.8 KV. LB medium (950 µl) was added and incubated in 1.5 ml tubes, followed by growth in a shaking incubator at 28 ^{o}C, for four hr. Positive cells were selected on LB plates with antibiotics, and incubated at 28 ^{o}C for three days.

7.2.4.3. Plasmid Extraction from *Agrobacterium* Cultures

Isolation of plasmids from *Agrobacterium* was carried out using a modified alkaline lysis method (Wang, 2006). After overnight LB medium culture with antibiotics, cells were harvested. The cell pellets were lysed with 100 µl of lysis buffer (50 mM glucose; 25 mM Tris-HCl, pH 8; 10 mM EDTA, pH 8; and 4 mg/ml Lysozyme), 200 µl of a freshly prepared solution II (0.2M NaOH and 1% SDS), and 150 µl of 3 M sodium acetate (pH 4.8). Cell debris was pelleted, and the supernatant was extracted with 1x phenol:chloroform:isoamylalcohol (25:24:1) and 1x chloroform. The plasmid DNA was precipitated from the upper layer with isopropanol, and washed with 70% ethanol. The DNA was re-suspended in 50 µl of sterile distilled water and kept at -20 ^{o}C.

7.2.4.4. Maize Transformations and Selections

The *Agrobacterium*-mediated maize transformation was carried out for the p7UZMET1ir construct using the method described by (Ishida *et al.,* 2007). The transformants were grown under selection. Maize transformation and selection was performed by Dr Fridtjof Weltmeier, KWS.

7.2.4.5. *Arabidopsis* Floral-Dip *Agrobacterium*-Mediated Transformation

The floral dip transformation was performed by following methods described by Clough and Bent (1998). The *Arabidopsis* plants were grown until they were flowering. The *Agrobacterium* colony, which contained the construct, was inoculated in 100 ml of LB medium, and incubated overnight at 28 ^{o}C with vigorous

agitation. The *Agrobacterium* cells were harvested and re-suspended in 5% sucrose and 0.05% SilwetL-77 surfactant to a final OD_{600} = 0.8. The plants were inverted, and all the above-ground parts were dipped into the solution for 10 sec. with gentle agitation. The dipped plants were covered, and placed in the dark overnight. The plants were watered and grown normally. Dipping was repeated after one week. The plants were grown until seeds could be collected. Primary transformants were selected with MS medium containing selection medium.

References

Angaji SA, Hedayati SS, Poor RH, Poor SS, Shiravi S, Madani S (2010) Application of RNA interference in plants. *Plant OMICS* **3**(3): 77–84.

Abrahams S, Tanner GJ, Larkin PJ, Ashton AR (2002) Identification and biochemical characterization of mutants in the proanthocyanidin pathway in *Arabidopsis*. *Plant Physiol* **130**(2): 561–576.

Agius F, Kapoor A, Zhu JK (2006) Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc Nat Acad Sci USA* **103**(31): 11796–11801.

Ahmad F, Huang X, Lan HX, Huma T, Bao YM, Huang J, Zhang HS (2014) Comprehensive gene expression analysis of the DNA (cytosine-5) methyltransferase family in rice (*Oryza sativa L.*). *Genet Mol Res* **13**(3): 5159–72.

Ainley WM, Sastry-Dent L, Welter ME, Murray MG, Zeitler B, Amora R, Corbin DR, Miles RR, Arnold NL, Strange TL, Simpson MA, Cao Z, Carroll C, Pawelczak KS, Blue R, West K, Rowland LM, Perkins D, Samuel P, Dewes CM, Shen L, Sriram S, Evans SL, Rebar EJ, Zhang L, Gregory PD, Urnov FD, Webb SR, Petolino JF (2013) Trait stacking via targeted genome editing. *Plant Biotechnol J* **11**(9): 1126-34.

Akhter S, Uddin MN, Jeong IS, Kim DW, Liu XM, Bahk JD (2016) Role of *Arabidopsis* AtPI4Kϒ3, a type II phosphoinositide 4-kinase, in abiotic stress responses and floral transition. *Plant Biotechnol J* **14**(1): 215–230.

Alcaiay M, Toniolo D (1988) CpG islands of the X chromosome are gene associated. *Nucleic Acids Res* **16**(20): 9527–9543.

Ali Z, Abulfaraj A, Idris A, Ali S, Tashkandi M, Mahfouz MM (2015) CRISPR/Cas9-mediated viral interference in plants. *Genome Biol* **16**(1): 238.

Amoah, S, Kurup S, Rodriguez-Lopez CM, Welham SJ, Powers SJ, Hopkins CJ, Wilkinson MJ, King GJ (2012). A Hypomethylated population of *Brassica rapa* for forward and reverse Epi-genetics. *BMC Plant Biol* **12**(1): 193.

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998) A viral suppressor of gene silencing in plants. *Proc Nat Acad Sci USA*, **95**(22): 13079–13084.

Andorf CM, Cannon EK, Portwood JL, Gariner JM, Harper LC, Schaeffer ML, Braun BL, Campbell DA, Vinnakota AG, Sribalusu VV, Huerta M, Cho KT, Wimalanathan K, Richter JD, Mauch ED, Rao BS, Birkett SM, Sen TZ, Lawrence-Dill CJ (2016) MaizeGDB update: new tools, data and interface for maize model organism database. *Nucleic acids res* **44**(D1): D1195-D1201.

Ashapkin VV, Kutueva LI, Vanyushin BF (2016) Plant DNA Methyltransferase Genes : Multiplicity, Expression, Methylation Patterns. *Biochem (Mosc)* **81**(2): 141–151.

Auerbach C (1949). Chemical mutagenesis. *Genetics and Biology of Drosophila*, (January): 1299–1341.

Auerbach C, Robson JM (1946) Chemical production of mutations. *Nat* **157**(18): 302.

Baker M (2014) Gene editing at CRISPR speed. *Nat Biotechnol* **32**(4): 309–12.

Baltes NJ, Hummel AW, Konecna E, Cegan R, Bruns AN, Bisaro DM, Voytas DF (2015) Conferring resistance to geminiviruses with the CRISPR–Cas prokaryotic immune system. *Nat Plants* **1**(10): 15145.

Banno H, Ikeda Y, Niu QW, Chua NH(2001) Overexpression of *Arabidopsis* ESR1 induces initiation of shoot regeneration. *Plant Cell* **13**(12): 2609–18.

Baránek M, Čechová J, Raddová J, Holleinová V, Ondrušíková E, Pidra M (2015) Dynamics and reversibility of the DNA methylation landscape of grapevine plants (*Vitis vinifera*) stressed by in vitro cultivation and thermotherapy. *PLoS ONE* **10**(5): 1–16.

Barrangou R, Doudna JA (2016) Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* **34**(9): 933–941.

Bashtrykov P, Rajavelu A, Hackner B, Ragozin S, Carell T, Jeltsch A (2014) Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. *Chem Bio Chem* **15**(5): 743–748.

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, SanMiguel PJ, Bennetzen, JL (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genet* **5**(11): e1000732.

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. *Genes Dev* **23**(7): 781–783.

Bernacchia G, Primo A, Giorgetti L, Pitto L, Cella R (1998) Carrot DNA-methyltransferase is encoded by two classes of genes with differing patterns of expression. *Plant J* **13**(3): 317–329.

Bestor TH (2000) The DNA methyltransferases of mammals. *Hum Mol Gen* **9**(16): 2395–2402.

Bharti P, Mahajan M, Vishwakarma AK, Bhardwaj J, Yadav SK (2015) AtROS1 overexpression provides evidence for epigenetic regulation of genes encoding enzymes of flavonoid biosynthesis and antioxidant pathways during salt stress in transgenic tobacco. *J Exp Bot* **66**(19): 5959–5969.

Bieberstein NI, Kozáková E, Huranová M, Thakur PK, Krchňáková Z, Krausová M, Oesterreich FC, Staněk D (2016) TALE-directed local modulation of H3K9 methylation shapes exon recognition. *Sci Rep* **6**: 29961.

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini L (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**(15): 7429–7437.

Bird A (2007) Perceptions of epigenetics. *Nat* **447**(7143): 396–398.

Blevins T, Pontes O, Pikaard CS, Meins F (2009) Heterochromatic siRNAs and *DDM1* independently silence aberrant 5S rDNA transcripts in *Arabidopsis*. *PLoS ONE* **4**(6): e5932.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors.*Sci* **326** (5959): 1509-1512.

Boch J, Bonas U (2010) *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**: 419–436.

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* **16**(12): 1–15.

Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol Adv* **33**(1): 41–52.

Brink A, Styles ED, Axtell JD (1968) Paramutation : Directed Genetic Change. *Science* **159**(3811): 161-170.

Calarco JP, Borges F, Donoghue MTA, Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijó JA, Becker JD, Martienssen RA (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* 151(1): 194–205.

Callebaut I, Courvalin JC, Mornon JP (1999) The BAH (bromo-adjacent homology) domain: A link between DNA methylation, replication and transcriptional regulation. *FEBS Lett* **446**(1): 189–193.

Candaele J, Demuynck K, Mosoti D, Beemster GTS, Inzé D, Nelissen H (2014) Differential methylation during maize leaf growth targets developmentally regulated genes. *Plant Physiol* **164**(3): 1350–64.

Cao D, Ju Z, Gao C, Mei X, Fu D, Zhu H, Luo Y, Zhu B (2014) Genome-wide identification of cytosine-5 DNA methyltransferases and demethylases in *Solanum lycopersicum*. *Gene* **550**(2): 230–237.

Cao L, Yu N, Li J, Qi Z, Wang D, Chen L (2016) Heritability and reversibility of DNA methylation induced by in vitro grafting between *Brassica juncea* and *B. oleracea*. *Sci Rep* **6**: 27233.

Casadesús J, Low D (2006) Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol R* **70**(3): 830–56.

Cavrak VV, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Scheid OM (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genetics* **10**(1): e1004115.

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting *Nucleic Acids Res*, **39**(12): e82.

Chang HW, Chuang LY, Cheng YH, Hung YC, Wen CH, Gu DL, Yang CH (2009) Prim-SNPing: A primer designer for cost-effective SNP genotyping. *BioTechniques* **46**(6): 421–431.

Chaudhary K, Pratap D, Sharma PK (2016) Transcription activator-like effector nucleases (TALENs): An efficient tool for plant genome editing. *Eng Life Sci* **16**(4): 330–337.

Chen H, Kazemier HG, De Groote ML, Ruiters MHJ, Xu GL, Rots MG (2014) Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. *Nucleic Acids Res* **42**(3): 1563–1574.

Chen K, Gao C (2013) TALENs: Customizable Molecular DNA Scissors for Genome Engineering of Plants. *J Genet Genomics* **40**(6): 271–279.

Chen M, Ha M, Lackey E, Wang J, Chen ZJ (2008) RNAi of met1 Reduces DNA methylation and induces genome-specific changes in gene expression and centromeric small RNA accumulation in *Arabidopsis* allopolyploids. *Genetics* **178**(4): 1845–1858.

Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**(1): 132–41.

Choudhury SR, Cui Y, Lubecka K, Stefanska B, Irudayaraj J (2016) CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **7**(29): 46545-46556.

Christian M, Qi Y, Zhang Y, Voytas DF (2013) Targeted mutagenesis of *Arabidopsis* thaliana using engineered TAL effector nucleases. *G3* **3**(10): 1697–705.

Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**(6): 735–743.

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nat* **353**(6339): 31–37.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nat* **452**(7184): 215–219.

Davis D, Stokoe D (2010) Zinc finger nucleases as tools to understand and treat human diseases. *BMC Med* **8**: 42.

Davuluri GR, Van Tuinen A, Mustilli AC, Manfredonia A, Newman R, Burgess D, Brummell DA, King SR, Palys J, Uhlig J, Pennings HMJ, Bowler C (2004) Manipulation of DET1 expression in tomato results in photomorphogenic phenotypes caused by post-transcriptional gene silencing. *Plant J* **40**(3): 344–354.

De Groote ML, Verschure PJ, Rots MG (2012) Epigenetic Editing: Targeted rewriting of epigenetic marks to modulate expression of selected target genes. *Nucleic Acids Res* **40**(21): 10596–10613.

Ding Y, Fromm M, Avramova Z (2012) Multiple exposures to drought "train" transcriptional responses in *Arabidopsis*. *Nat. Commun* **3**: 740.

Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S (2014) Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol Cell* **55**(3): 495-504.

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E,Vashisht AA, Terragni J, Chin HG, Tu A, Hetzel J, Wohlschlegel JA, Pradhan S, Patel DJ, Jacobsen SE (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**(1): 167–180.

Dubin MJ, Zhang P, Meng D, Remigereau MSS, Osborne EJ, Casale PF, Drewe P, Kahles A, Jean G, Vilhjálmsson B, Jagoda J, Irez S, Voronin V, Song Q, Long Q, Rätsch G, Stegle O, Clark RM, Nordborg M (2015) DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *eLife* **4**: e05255.

Erdmann RM, Souza AL, Clish CB, Gehring M (2014) 5-hydroxymethylcytosine is not present in appreciable quantities in *Arabidopsis* DNA. *G3* **5**(1): 1–8.

Feng C, Yuan J, Wang R, Liu Y, Birchler JA, Han F (2016) Efficient Targeted Genome Modification in Maize Using CRISPR/Cas9 System. *J Genet Genomics* **43**(1): 37–43.

Finnegan EJ, Kovacn K (2000) Plant DNA methyltransferases. *Plant Mol Biol* **43**(23): 189–201.

Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc Nat Acad Sci USA* **93**(16): 8449–8454.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**(4): 1531–1545.

Fu D, Uauy C, Blechl A, Dubcovsky J (2007) RNA interference for wheat functional gene analysis. *Transgenic Res* **16**(6): 689–701.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**(9): 822–6.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**: 279–284.

Fujimoto R, Sasaki T, Nishio T (2006) Characterization of DNA methyltransferase genes in *Brassica rapa*. *Genes Genet Sys* **81**(4): 235–42.

Fürst RW, Kliem H, Meyer HHD, Ulbrich SE (2012) A differentially methylated single CpG-site is correlated with estrogen receptor alpha transcription. *J. Steroid Biochem. Mol. Biol.* **130(**1–2): 96–104.

Gallusci P, Hodgman TC, Teyssier E, Seymour G (2016) DNA methylation and chromatin regulation during fleshy fruit development and ripening. *Front Plant Sci* **7**: 807.

Garg R, Kumari R, Tiwari S, Goyal S (2014) Genomic survey, gene expression analysis and structural modeling suggest diverse roles of DNA methyltransferases in legumes. *PLoS ONE* **9**(2): e88947.

Garvilles RG, Hasegawa T, Kimura H, Sharif J, Muto M, Koseki H, Takahashi S, Suetake I, Tajima S (2015) Dual functions of the RFTS domain of dnmt1 in replication-coupled DNA methylation and in protection of the genome from aberrant methylation. *PLoS ONE*, **10**(9): 1–19.

Gehring M (2013) Genomic imprinting: insights from plants. *Annu Rev Genet* **47**: 187–208.

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation *Cell* **124**(3): 495–506.

Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK (2013) CHH islands: De novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res* **23**(4): 628–637.

Gentry M, Meyer P (2013) An 11bp region with stem formation potential is essential for *de novo* DNA methylation of the RPS element. *PLoS ONE* **8**(5): e63652.

Gouil Q, Baulcombe DC (2016) DNA methylation signatures of the plant Chromomethyltransferases. *PLOS Genetics* **12**(12): e1006526.

Gregory DJ, Zhang Y, Kobzik L, Fedulov AV (2013) Specific transcriptional enhancement of inducible nitric oxide synthase by targeted promoter demethylation. *Epigenetics* **8**(11): 1205–1212.

Groth M, Moissiard G, Wirtz M, Wang H, Garcia-Salinas C, Ramos-Parra PA, Bischof S, Feng S, Cokus SJ, John A, Smith DC, Zhai J, Hale CJ, Long AJ, Hell R, Díaz de la Garza RI, Jacobsen SE (2016) MTHFD1 controls DNA methylation in *Arabidopsis*. *Nat Commun* **7**: 11640.

Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, Fuks G, Butler E, Wing RA, Rounsley S, Birren B, Nusbaum C, Mayer KFX, Messing J (2005) Structure and architecture of the maize genome. *Plant Physiol* **139**(4): 1612–24.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95-98.

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nat* **455**(7214): 826–829.

Hayashi S, Gresshoff PM, Kinkema M (2008) Molecular analysis of lipoxygenases associated with nodule development in soybean. *Mol Plant Microbe Interact :* **21**(6): 843–853.

He XJ, Chen T, Zhu JK (2011a) Regulation and function of DNA methylation in plants and animals. *Cell Res* **54**(3): 442–465.

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL (2011b) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**(6047): 1303–7.

Heidenreich M, Zhang F (2015) Applications of CRISPR–Cas systems in neuroscience. *Nat Rev Neurosci* **17**(1): 36–44.

Henikoff S, Comai L (1998) A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. *Genetics* **149**(1): 307–318.

Herskowitz I (1987) Functional inactivation of genes by dominant negative mutations. *Nat* **329**(6136): 219–222.

Hetzl J, Foerster AM, Raidl G, Scheid OM (2007) CyMATE: A new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J* **51**(3): 526–536.

Hille F, Charpentier E (2016) CRISPR-Cas : biology , mechanisms and relevance. *Phil. Trans. R. Soc. B* **371**: 20150496.

Hilson P, Allemeersch J, Altmann T, Aubourg S, Avon A, Beynon J, Bhalerao RP, Bitton F, Caboche M, Cannoot B, Chardakov V, Cognet-Holliger C, Colot V, Crowe M, Darimont C, Durinck S, Eickhoff H, de Longevialle AF, Farmer EE, Grant M, Kuiper MT, Lehrach H, Léon C, Leyva A, Lundeberg J, Lurin C, Moreau Y, Nietfeld W, Paz-Ares J, Reymond P, Rouzé P, Sandberg G, Segura MD, Serizet C, Tabrett A, Taconnat L, Thareau V, Van Hummelen P, Vercruysse S, Vuylsteke M, Weingartner M, Weisbeek PJ, Wirta V, Wittink FR, Zabeau M, Small I (2004) Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: Transcript profiling and reverse genetics applications. *Genome Res* **14**(10B): 2176–2189.

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**: 510–517.

Hirooka K, Izumi Y, An CL, Nakazawa Y, Fukusaki E, Kobasyashi A (2005) Functional analysis of two solanesyl diphosphate synthases from *Arabidopsis thaliana*. *Biosci Biotechnol Biochem* **69** (3): 592-601.

Hoen DR, Kyong CP, Elrouby N, Yu Z, Mohabir N, Cowan RK, Bureau TE (2006) Transposon-mediated expansion and diversification of a family of ULP-like genes. *Mol Biol Evol* **23**(6): 1254–1268.

Hollick JB, Chandler VL (1998) Epigenetic allelic states of a maize transcriptional regulatory locus exhibit overdominant gene action. *Genetics* **150**(2): 891–897.

Hollwey E, Watson M, Meyer P (2016) Expression of the C-Terminal domain of mammalian TET 3 DNA dioxygenase in *Arabidopsis thaliana* induces heritable methylation changes at rDNA loci, *Adv Biosci Biotechnol* **7**: 243-250.

Hsieh CL (1999) In vivo activity of murine *de novo* methyltransferases, Dnmt3a and Dnmt3b. *Mol Cell Biol* **19**(12): 8211–8.

Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, Fischer RL (2011) Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc Nat Acad Sci USA* **108**(5): 1755–62.

Hu L, Li N, Zhong S, Lin X, Yang J, Zhou T, Yuliang A, Wu Y, Chen Y-R, Cao X, Zemach A, Rustgi S, Wettstein DV, Liu B (2014) Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression and seedling lethality. *Proc Nat Acad Sci USA* **111** (29) 10642-7.

Ichikawa T, Nakazawa M, Kawashima M, Muto S, Gohda K, Suzuki K, Ishikawa A, Kobayashi H, Yoshizumi T, Tsumoto Y, Tsuhara Y, Iizumi H, Goto Y, Matsui M (2003) Sequence database of 1172 T-DNA insertion sites in *Arabidopsis* activation-tagging lines that showed phenotypes in T1 generation. *Plant J* **36**(3): 421–429.

Ishida Y, Hiei Y, Komari T (2007) *Agrobacterium*-mediated transformation of maize. *Nat Protoc* **2**(7): 1614–1621.

James, C (2013) Global Status of Commercialized Biotech/GM Crops :2013. *ISAAA Brief* No 46. ISAAA: Ithaca. NY.

Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nat* **416**(6880): 556–60.

Jacobsen SE, Meyerowitz EM (1997) Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* **277**(5329): 1100–1103.

Jeltsch A (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem* **3**(4): 274–93.

Ji X, Zhang H, Zhang Y, Wang Y, Gao C (2015) Establishing a CRISPR–Cas-like immune system conferring DNA virus resistance in plants. *Nat Plants*, **1**: 15144.

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Res* **41**(20): 1–12.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A Programmable dual-rNA – guided DNA endonuclease in adaptice bacterial immunity. *Science* **337**: 816–822.

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**(6176): 1247997.

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* **17**(4): 379–384.

Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F (2012) DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*. *CurrBiol* **22**(19): 1825–1830.

Kakimoto T (1996) CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* **274**(6): 982–985.

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* **163**(3): 1109–1122.

Karim R, Nuruzzaman M, Khalid N, Harikrishna JA (2016) Importance of DNA and histone methylation in in vitro plant propagation for crop improvement: a review. *Ann Appl Biol* **169**(1): 1–16.

Kawashima T, Berger F (2014) Epigenetic reprogramming in plant sexual reproduction. *Nat Rev Genet* **15**(9): 613–624.

Kerschen A, Napoli CA, Jorgensen RA, Müller AE (2004) Effectiveness of RNA interference in transgenic plants. *FEBS Lett* **566**(1–3): 223–228.

Kinoshita T (2004) One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA Methylation. *Science* **303**(5657): 521–523.

Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales APW, Li Z, Peterson RT, Yeh JRJ, Aryee MJ, Joung JK (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nat* **523**(7561): 481–485.

Koncz C, Schell J (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol Gen Genet* **204**(3): 383–396.

Kondou Y, Higuchi M, Matsui M (2010) High-throughput characterization of plant gene functions by using gain-of-function technology. *Annu Rev Plant Biol* **61**: 373–393.

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2014) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nat* **517**: 583–588.

Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG (1994) The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res* **22**(1): 1–10.

Kungulovski G, Jeltsch A (2016) Epigenome editing: state of the art, concepts, and perspectives. *Trends Genet* **32**(2): 101–113.

Kusaba M, Miyahara K, Iida S, Fukuoka H, Takano T, Sassa H, Nishimura M, Nishio, T (2003) Low glutelin content1: a dominant mutation that suppresses the glutelin multigene family via RNA silencing in rice. *Plant Cell* **15**:1455–1467.

Kreps JA, Wu Y, Chang H-S, Zhu T, Wang X, Harper JF (2002) Transciptome changes for *Arabidopsis* in response to salt, osmotic and cold stress. *Plant physiol* **130**: 2129-2141.

Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**(11): 2180–96.

Laufer BI, Singh SM (2015) Strategies for precision modulation of gene expression by epigenome editing: an overview. *Epigenet Chromatin* **8**(1): 34.

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nat* **498**(7454): 385–389.

Le J, Liu XG, Yang KZ, Chen XL, Zou JJ, Wang HZ, Wang, Vanneste S, Morita, Tasaka MM, Ding ZJ, Frim J, Beeckman T, Sack F (2014) Auxin transport and activity regulate stomatal patterning and development. *Nat Commun* **5**: 3090.

Lee TF, Zhai J, Meyers BC (2010) Conservation and divergence in eukaryotic DNA methylation. *Proc Nat Acad Sci USA* **107**(19): 9027–9028.

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**(6): 915–926.

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013) Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* **31**(8): 688–691.

Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr Biol* **15**(16): 1501–1507.

Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol* **30**(5): 390–392.

Liang Z, Zhang K, Chen K, Gao C (2014) Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *J Genet Genomics* **41**(2): 63–68.

Liew LC, Singh MB, Bhalla PL (2013) An RNA-Seq Transcriptome Analysis of Histone Modifiers and RNA Silencing Genes in Soybean during Floral Initiation Process. *PLoS ONE* **8**(10), 1–23.

Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna, S. Shu J, Dadon D, Young RA, Jaenisch R (2016) Editing DNA methylation in the mammalian genome. *Cell* **167**(1): 233–247.

Liu X, Yu CW, Duan J, Luo M, Wang K, Tian G, Cui Y, Wu K. (2012) HDA6 Directly Interacts with DNA Methyltransferase MET1 and Maintains Transposable Element Silencing in *Arabidopsis*. *Plant Physiol* **158**(1): 119–129.

Lopez A, Stassen J, Furci L, Smith L, Ton J (2015) The role of DNA (de)methylation in immune responsiveness of *Arabidopsis*. *Plant J* **88**(3): 361-374.

Lor VS, Starker CG, Voytas DF, Weiss D, Olszewski N (2014) Targeted mutagenesis of the tomato PROCERA gene using transcription activator-like effector nucleases. *Plant Physiol* **166**: 1288–1291.

Lu S, Yin X, Spollen W, Zhang N, Xu D, Schoelz J, Bilyeu K, Zhang ZJ (2015) Analysis of the siRNA-mediated gene silencing process targeting three homologous genes controlling soybean seed oil quality. *PLOS One 10*(6): e0129010.

Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE, Costello JF, Wilkinson MF, Joung JK (2013) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**(12): 1137–1142.

Maghari BM, Ardekani AM (2011) Genetically Modified Foods and Social Concerns. *Avicenna Journal of Medical Biotechnology* **3**(3): 109–117.

Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK (2011) *De novo*-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl Acad. Sci. USA* **108**(6): 2623–2628.

Mahfouz MM, Li L, Piatek M, Fang X, Mansour H, Bangarusamy DK, Zhu JK (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol Biol* **78**(3): 311–321.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and Classification of the CRISPR–Cas Systems. *Nat Rev Microbiol* **9**(6): 467–477.

Makarova KS, Wolf Y, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**(11): 722–736.

Makarevitch I, Stupar RM, Iniguez AL,Haun WJ, Barbazuk WB, Kaeppler SM, Springer NM (2007) Natural variation for alleles under epigenetic control by the maize chromomethylase *Zmet2*. *Genetics* **177** (2): 749-760.

Mathieu O, Reinders J, Čaikovski M, Smathajitt C, Paszkowski J (2007) Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG Methylation. *Cell* **130**(5): 851–862.

Matzke AJ, Matzke MA (1998) Position effects and epigenetic silencing of plant transgenes. *Curr Opin Plant Biol* **1**: 142–148.

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* **15**(6):394–408.

Manoli A, Sturaro A, Trevisan S, Quaggiotti S, Nonis A (2012) Evaluation of candidate reference genes for qPCR in maize. *Plant Physiol* **169** (8): 807-815.

McGinnis KM (2010) RNAi for functional genomics in plants. *Brief Func Genomics Proteomics* **9**(2): 111–117.

Mehra M, Gangwar I, Shankar R (2015) A deluge of complex repeats: The *Solanum* genome. *PLoS ONE***10**(8): 1–38.

Melquist S, Bender J (2004) An internal rearrangement in an *Arabidopsis* inverted repeat locus impairs DNA Methylation triggered by the locus. *Genetics* **166**(1): 437–448.

Mena M, Ambrose BA, Meeley RB, Briggs SP, Yanofsky MF, Schmidt RJ (1996) Diversification of C-Function activity in maize flower development. *Science* **274**(5292): 1537–1540.

Mersereau M, Pazour GJ, Das A (1990) Efficient transformation of *Agrobacterium tumefaciens* by electroporation. *Gene* **90** 149–151.

Meyer P (2015) Epigenetic variation and environmental change. *J Exp Bot* **66**(12): 3541–3548.

Meyer P, Linn F, Heidmann I, Meyer H, Niedenhof I, Saedler H (1992) Endogenous and environmental factors influence 35S promoter methylation of a maize A1 gene construct in transgenic petunia and its colour phenotype. *Mol Gen Genet* **231**(3): 345–352.

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nat* **411**:212–214.

Morbitzer R, Römer P, Boch J, Lahaye T (2010) Regulation of selected genome loci using *de novo*-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Nat Acad Sci USA*, **107**(50): 21617–21622.

NatureMethods (2012) Method of the Year 2011. *Nature Methods* **9**(1): 1.

Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**(8): 691–693.

Nelissen H, Rymen B, Coppens F, Dhondt S, Fiorani F, Beemster GTS (2013) Kinematic analysis of cell division in leaves of mono- and dicotyledonous species: a basis for understanding growth and developing refined molecular sampling strategies. In I. De Smet (Ed.), *Plant Organogenesis: Methods and Protocols* (pp. 247–264). Totowa, NJ: Humana Press.

Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim K, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, Egesi C, Schmutz J, Grimwood J, Jackson SA, Springer NM, Schmitz RJ (2016) Widespread natural variation of DNA methylation within angiosperms. *Genome Biol* **17(**1): 194.

Nielsen S, Yuzenkova Y, Zenkin N (2013) Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* **340**(6140): 1577–80.

Noy-Malka C, Yaari R, Itzhaki R, Mosquna A, Auerbach Gershovitz N, Katz A, Ohad N (2014) A single CMT methyltransferase homolog is involved in CHG DNA methylation and development of *Physcomitrella patens*. *Plant Mol Biol* **84**(6): 719–735.

Numata K, Ohtani M, Yoshizumi T, Demura T, Kodama Y (2014) Local gene silencing in plants via synthetic dsRNA and carrier peptide. *Plant Biotechnol J* **12**(8): 1027–1034.

Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol* **162**(1): 116–31.

Oh YJ, Chung H, Yu JG, Park YD (2009) Newly developed MSAP analysis reveals the different polymorphism patterns in transgenic tobacco plants with the dsRNA MET1 gene. *Plant Biotechnol Rep*, **3**(2): 139–145.

Ohta T (1987) Simulating evolution by gene duplication. *Genetics* **115**(1): 207–213.

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**(3): 247–257.

Okumura K-I, Goh T, Toyokura K, Kasahara H, Takebayashi Y, Mimura T, Kamiya Y, Fukaki H (2013) GNOM/FEWER ROOTS is required for the establishment of an auxin response maximum for *Arabidopsis* lateral roots initiation. *Plant cell physiol* **54** (3): 406-417.

Ong-Abdullah M, Ordway JM, Jiang N, Ooi SE, Kok S, Sarpan e al. (2015) Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **525**(7570): 533–537.

Osakabe K, Osakabe Y, Toki S (2010) Site-directed mutagenesis in *Arabidopsis* using custom-designed zinc finger nucleases. *Proc Nat Acad Sci USA* **107**(26): 12034–12039.

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control of leaf morphogenesis by microRNAs. *Nat* **425**: 257–263.

Pan C, Ye L, Qin L, Liu X, He Y, Wang, J, CHen L, Lu G (2016). CRISPR / Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. *Sci Rep* **6:**24765.

Papa CM, Springer NM, Muszynski MG, Meeley R, Kaeppler SM (2001) Maize chromomethylase Zea methyltransferase2 is required for CpNpG methylation. *Plant Cell* **13**(8): 1919–28.

Pavlopoulou A, Kossida S (2007) Plant cytosine-5 DNA methyltransferases: Structure, function, and molecular evolution. *Genomics* **90**(4): 530–541.

Peng H, Zhang J (2009) Plant genomic DNA methylation in response to stresses: Potential applications and challenges in plant breeding. *Prog Nat Sci* **19**(9): 1037–1045.

Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL (2007) DNA demethylation in the *Arabidopsis* genome. *Proc Nat Acad Sci USA* **104**(16): 6752–7.

Petolino JF (2015) Genome editing in plants via designed zinc finger nucleases. *In Vitro Cell Dev Biol Plant* **51**(1): 1–8.

Petrova NS, Zenkova MA, Chernolovskaya EL (2013) Structure - Functions relations in small interfering RNAs, practical applications in biomedical engineering, Dr. Adriano Andrade (Ed.), InTech, DOI: 10.5772/53945.

Pfluger J, Wagner D (2007) Histone modifications and dynamic regulation of genome accessibility in plants. *Curr Opin Plant Biol* **10**(6): 645–652.

Ponferrada-Marín MI, Roldán-Arjona T, Ariza RR (2009) ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. *Nucleic Acids Res* **37**(13): 4264–4274.

Pontier D, Miao ZH, Lam E (2001) Trans-dominant suppression of plant TGA factors reveals their negative and positive roles in plant defense responses. *Plant J* **27**(6): 529–538.

Prelich G (2012) Gene overexpression: Uses, mechanisms, and interpretation. *Genetics* **190**(3): 841–854.

Pruett-Miller SM, Reading DW, Porter SN, Porteus MH (2009) Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. *PLoS Genet* **5**(2): e1000376.

PROMEGA *Epigenetic* [Online] [Viewed on 14 February 2017] Available from https://www.promega.co.uk/resources/product-guides-andselectors/protocols-and-applications-guide/epigenetics/.

Qian Y, Xi Y, Cheng B, Zhu S (2014) Genome-wide identification and expression profiling of DNA methyltransferase gene family in maize. *Plant Cell Rep*, **33**(10): 1661–1672.

Rabinowicz PD, Palmer LE, May BP, Hemann MT, Lowe SW, McCombie WR, Martienssen RA(2003) Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Res* **13**(12): 2658–2664.

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F (2015) In vivo genome editing using *Staphylococcus aureus* Cas9. *Nat* **520**(7546): 186–190.

Regulski M, Lu Z, Kendall J, Donoghue MTA, Reinders J, Llaca V, Deschamps S, Smith A, Levy D, McCombie WR, Tingey S, Rafalski A, Hicks J, Ware D, Martienssen RA (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Res* **23**(10): 1651–1662.

Reinders J, Wulff BB, Mirouze M, Marí-Ordóñez A, Dapp M, Rozhon W, Bucher E, Theiler G, Paszkowski J (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev* **23**(8): 939–950.

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Comput Biol* **2**(9): 989–1000.

Rodoni BC, Dale JL, Harding RM (1999) Characterization and expression of the coat protein-coding region of banana bract mosaic potyvirus, development of diagnostic assays and detection of the virus in banana plants from five countries in southeast Asia. *Arch Virol* **144**(9): 1725–1737.

Rommens CM, Haring MA, Swords K, Davies HV, Belknap WR (2007) The intragenic approach as a new extension to traditional plant breeding. *Trends in Plant Sci* **12**(9): 397–403.

Rowley MJ, Avrutsky MI, Sifuentes CJ, Pereira L, Wierzbicki AT (2011) Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. *PLoS Genet* **7**(6): 1–9.

Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**(4): 347–55.

Sanjana NE, Cong L, Zhou Y, Cunniff MM, Feng G, Zhang F (2012) A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* **7**(1): 171–192.

Sambrook J, Fritsch E, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*: Cold Spring Harbour Laboratory Press.

Sayre R, Kumar V, Tuteja N, Khatodia S, Bhatotia K, Passricha N, Khurana SMP (2016) The CRISPR/Cas genome-editing tool: Application in improvement of crops. *Fron Plant Sci* **7**: 506.

Saze H, Kakutani T (2007) Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J* **26**(15): 3641–52.

Saze H, Scheid MO, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* **34**(1): 65–9.

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**(6054):369-73.

Scholze H, Boch J (2011) TAL effectors are remote controls for gene activation. *Curr Opin Microbiol***14**(1): 47–53.

Schuermann D, Molinier J, Fritsch O, Hohn B (2005) The dual nature of homologous recombination in plants. *Trends Genet* **21**(3): 172–181.

Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, Amasino RA, Coupland G (2006) The transcription factor *FLC* confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes Dev* **20**(7): 898–912.

Segal G, Song R, Messing J (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. *Genetics* **165**(1): 387–397.

Senthil-kumar M, Mysore KS (2010) RNAi in plants : recent developments and applications in agriculture. In *Gene Silencing Theory, Techniques and Applications*, Nova Science Publsisher (pp 183–199).

Senthil-kumar M, Mysore KS (2011) Caveat of RNAi in Plants: The Off-Target Effect In *RNAi and Plant Gene Function Analysis: Methods and Protocols*, Humana Press (pp 13-25).

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu JL, Gao C(2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* **31**(8): 686–688.

Shen X, De Jonge J, Forsberg SKG, Pettersson ME, Sheng Z, Hennig L, Orjan C (2014) Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genet* **10**(12): e1004842.

Shen WJ Forde BG (1989) Efficient transformation of *Agrobactenum* spp. by high voltage electroporation. *Nucleic Acids Res* **17**(20): 8385.

Shin R, Burch AY, Huppert KA, Tiwari SB, Murphy AS, Guilfoyle TJ, Schachtman, DP (2007) The *Arabidopsis* transcription factor MYB77 modulates auxin signal transduction. *Plant cell* **19**: 2440-2453.

Singh A, Zubko E, Meyer P (2008a) Cooperative activity of DNA methyltransferases for maintenance of symmetrical and non-symmetrical cytosine methylation in *Arabidopsis thaliana*. *Plant J* **56**(5): 814–823.

Singh J, Freeling M, Lisch D (2008b) A position effect on the heritability of epigenetic silencing. *PLoS Genet* **4**(10): e1000216.

Small I (2007) RNAi for revealing and engineering plant gene functions. *Curr Opin Biotechnol* **18**(2): 148–153.

Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJ (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene.*Mol Cell* **6**(4): 791–802.

Spertini D, Béliveau C, Bellemare G (1999) Screening of transgenic plants by amplification of unknown genomic DNA flanking T-DNA. *BioTechniques*, **27**(2): 308–314.

Stadler LJ(1928a). Genetic effects of X-rays in maize. *Proc Nat Acad Sci USA* **14**: 69–75.

Stadler LJ (1928b). Mutation in Barlye induced by X-Rays and Radium. *Science* **68**(1756): 186–7.

Stam M, De Bruin R, Van Blokland R, Van Der Hoorn RAL, Mol JNM, Kooter JM (2000). Distinct features of post-transcriptional gene silencing by antisense transgenes in single copy and inverted T-DNA repeat loci. *Plant J* **21**(1): 27–42.

Stevenson LF, Kennedy BK, Harlow E (2001) A large-scale overexpression screen in *Saccharomyces cerevisiae* identifies previously uncharacterized cell cycle genes. *Proc Nat Acad Sci USA* **98**(7): 3946–3951.

Stokes TL, Kunkel BN, Richards EJ (2002) Epigenetic variation in *Arabidopsis* disease resistance epigenetic variation in *Arabidopsis* disease resistance. *Genes Dev* **16**(2): 171–182.

Strathdee G, Sim A, Brown R (2004) Control of gene expression by CpG island methylation in normal cells. *Biochem Soc Trans* **32**(Pt 6): 913–5.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE (2014) Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat Struct Mol Biol* **21**(1): 64–72.

Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**(1–2): 352–364.

Sultan FA, Day JJ (2011) Epigenetic mechanisms in memory and synaptic function. *Epigenomics* **3**(2): 157–81.

Sun N, Zhao H (2013) Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnol Bioeng* **110**(7): 1811–1821.

Syeda F, Fagan RL, Wean M, Avvakumov GV, Walker JR, Xue S, Dhe-Paganon S, Brenner C (2011) The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J BiolChem* **286**(17): 15344–15351.

Tan L, Shi YG (2012) Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Dev* **139**(11): 1895–902.

Tang W, Weidner DA, Hu BY, Newton RJ, Hu XH (2006) Efficient delivery of small interfering RNA to plant cells by a nanosecond pulsed laser-induced stress wave for posttranscriptional gene silencing. *Plant Sci* **171**(3): 375–381.

Taramino G, Sauer M, Stauffer JL Jr, Multani D, Niu X, Sakai H, Hochholdinger F (2007) The maize (*Zea mays* L.) *RTCS* gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. *Plant J* **50** (4): 649-659.

Teerawanichpan P, Chandrasekharan MB, Jiang Y, Narangajavana J, Hall TC (2004) Characterization of two rice DNA methyltransferase genes and RNAi-mediated reactivation of a silenced transgene in rice callus. *Planta* **218**(3): 337–349.

Thakore PI, Black JB, Hilton IB, Gersbach CA (2016) Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods* **13**(2): 127–137.

Thomas M, Pingault L, Poulet A, Duarte J, Throude M, Faure S, Pichon JP, Paux E, Probst AV, Tatout C (2014) Evolutionary history of Methyltransferase 1 genes in hexaploid wheat. *BMC Genomics* **15**: 922.

Tovkach A, Zeevi V, Tzfira T (2009) A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells. *Plant J* **57**(4): 747–757.

Tricker PJ, Gibbings JG, López CMR, Hadley P, Wilkinson MJ (2012) Low relative humidity triggers RNA-directed de novo DNA methylation and suppression of genes controlling stomatal development. *J Exp Bot* **63**(10): 3799–3814.

Tricker PJ, López CMR, Gibbings G, Hadley P, Wilkinson MJ (2013) Transgenerational, dynamic methylation of stomata genes in response to low relative humidity. *Int J Mol Sci* **14**(4): 6674–6689.

Tycko B (2000) Epigenetic gene silencing in cancer. *J Clin Invest* **105**(4): 401–407.

Valton J, Dupuy A, Daboussi F, Thomas S, Maréchal A, Macmaster R, Melliand K, Juillerat A, Duchateau P (2012) Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J Biol Chem* **287**(46): 38427–38432.

Van der Graaff E, Hooykaas PJJ, Keller B (2002) Activation tagging of the two closely linked genes *LEP* and *VAS* independently affects vascular cell number. *Plant J* **32**(5): 819–830.

Van Nocker S, Gardiner SE (2014) Breeding better cultivars, faster: applications of new technologies for the rapid deployment of superior horticultural tree crops. *Hortic Res* **1**:14022.

Vassileva V, Hollwey E, Todorov D, Meyer P (2016) Leaf epidermal profiling as a phenotyping tool for DNA methylation mutants. *Genet Plant Physiol* ISSN**:** 1314-5770 (In Press).

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD (2007) Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol* **5**(7): e174.

Vejlupkova Z, Fowler JE (2003) Maize DNA preps for undergraduate students: A robust method for PCR genotyping. *Maize Genet Coop News lett* (7): 24-25.

Vicient CM (2010) Transcriptional activity of transposable elements in maize. *BMC Genomics* **11**(1): 1–10.

Wada T, Tachibana T, Shimura Y, Okada K (1997) Epidermal cell differentiation in *Arabidopsis* determined by a Myb homolog, CPC. *Science* **277**(5329): 1113–1116.

Waddington CH (1968) Towards a Theoretical Biology. *Nat* **2**(218): 2–4.

Walsh JB (1995) How often do duplicated genes evolve new functions? *Genetics* **139**(1): 421–428.

Waltz E (2016) CRISPR-edited crops free to enter market, skip regulation. *Nat Biotechnol* **34**(6): 582–582.

Wang MB, Abbott DC, Waterhouse PM (2000). A single copy of a virus-derived transgene encoding hairpin RNA gives immunity to barley yellow dwarf virus. *Mol Plant Pathol* **1**: 347–356.

Wang K (2006) *Agrobacterium Protocols*, Vol. 1: Humana Press.

Wang P, Gao C, Bian X, Zhao S, Zhao C, Xi H, Song H, Hou L, Was S, Wang X (2016) Genome-wide identification and comparative analysis of cytosine-5 DNA Methyltransferase and demethylase families in wild and cultivated Peanut. *Front Plant Sci* **7**(7): 1–14.

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu JL (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* **32**: 947-951.

Wassenegger M, Heimes S, Riedel L, Sanger HL (1994). RNA-directed *De-Novo* methylation of genomic sequences in plants. *Cell* **76**(3): 567–576.

Waterhouse PM, Graham MW, Wang MB (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Nat Acad Sci USA* **95**(23): 13959–13964.

Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh CT, Xu W, Schnable PS, Vaughn MW, Gehring M, Springer NM (2011) Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* **23**(12): 4221–4233.

Watson M (2013) Heritable epigenetic variation of DNA methylation targets in plants. *Ph.D. thesis, University of Leed*s.

Watson M, Hawkes E, Meyer P (2014). Transmission of epi-alleles with MET1-dependent dense methylation in *Arabidopsis thaliana*. *PloS One* **9**(8): e105338.

Wendt T, Holm PB, Starker CG, Christian M, Voytas DF, Brinch-Pedersen H, Holme IB (2013) TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. *Plant Mol Biol* **83**(3): 279–285.

Westergaard M (1957) Chemical mutagenesis in relation to the concept of the gene. *Experientia* **13**(6): 224–234.

Wolfe SA, Nekludova L Pabo CO (2000) DNA recognition by Cys2His2 Zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**: 183–212.

Woll K, Borsuk LA, Stransky H, Nettleton D, Schnable PS, Hochholdinger F (2005) Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. Plant physiol 139: 1255-1267.

Woo HR, Dittmer TA, Richards EJ (2008) Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genet* **4**(8): e1000156.

Woo HR, Pontes O, Pikaard CS, Richards EJ (2007) VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev* **21**(3): 267–277.

Woodhouse MR, Pedersen B, Freeling M (2010) Transposed genes in *Arabidopsis* are often associated with flanking repeats. *PLoS Genet* **6**(5): e1000949.

Woodward C, Bemis SM, Hill EJ, Sawa S, Koshiba T, Torii KU (2005) Interaction of auxin and *ERECTA* in elaborating *Arabidopsis* inflorescence architecture revealed by the activation tagging of a new member of the YUCCA family putative flavin monooxygenases. *Plant Physiol* **139**(1): 192–203.

Wu JL, Wu C, Lei C, Baraoidan M, Bordeos A, Madamba MR, Ramos-Pamplona M, Mauleon R, Portugal A, Ulat VJ, Bruskiewich R, Wang G, Leach J, Khush G, Leung H (2005) Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol Biol* **59**(1): 85–97.

Wu CT, Morris JR (2001) Genes, genetics, and epigenetics: a correspondence. *Science* **293**(5532): 1103–5.

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A Retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**(2008): 1527–1530.

Xiao W, Custard KD, Brown RC, Lemmon BE, Harada JJ, Goldberg RB (2006) DNA methylation is critical for *Arabidopsis* embryogenesis and seed viability. *Plant Cell* **18**(4): 805–814.

Xing HL, Dong L, Wang ZP, Zhang HY, Han CY, Liu B, Wang XC, Chen QJ (2014) A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol*, **14**(1): 327.

Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, Ruan K, Wang F, Xu GL, Hu R (2016) A CRISPR-based approach for targeted DNA demethylation. *Cell Dis* **2**: 16009.

Yaari R, Noy-Malka C, Wiedemann G, Gershovitz NA, Reski R, Katz A, Ohad N (2015) DNA METHYLTRANSFERASE 1 is involved in mCG and mCCG DNA methylation and is essential for sporophyte development in *Physcomitrella patens*. *Plant Mol Biol* **88**(4–5): 387–400.

Yamauchi T, Johzuka-Hisatomi Y, Fukada-Tanaka S, Terada R, Nakamura I, Iida S (2009) Homologous recombination-mediated knock-in targeting of the *MET1*a gene for a maintenance DNA methyltransferase reproducibly reveals dosage-dependent spatiotemporal gene expression in rice. *Plant J* **60**(2): 386–396.

Yamauchi T, Johzuka-Hisatomi Y, Terada R, Nakamura I, Iida S (2014) The *MET1b* gene encoding a maintenance DNA methyltransferase is indispensable for normal development in rice. *Plant Mol Biol* **85**(3): 219–232.

Yamauchi T, Moritoh S, Johzuka-Hisatomi Y, Ono A, Terada R, Nakamura I Iida S (2008) Alternative splicing of the rice *OsMET1* genes encoding maintenance DNA methyltransferase. *J Plant Physiol* **165**(17): 1774–1782.

Yang J, Zhang Y, Yuan P, Zhou Y, Cai C, Ren Q, Wen D, Chu C, Qi H, Wei W (2014) Complete decoding of TAL effectors for DNA recognition. *Cell Res* **24**(5): 628–31.

Yokotani N, Ichikawa T, Kondou Y, Matsui M, Hirochika H, Iwabuchi M, Oda K (2008) Expression of rice heat stress transcription factor OsHsfA2e enhances tolerance to environmental stresses in transgenic *Arabidopsis*. *Planta* **227**(5): 957–967.

Yu X, Willmann MR, Anderson SJ, Gregory BD (2016) Genome-wide mapping of uncapped and cleaved transcripts reveals a role for the nuclear mRNA cap-binding complex in co-translational RNA Decay in *Arabidopsis*. *Plant Cell* **28**(10): 2385-2397.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**(1) 193-205.

Zhang JZ (2003) Overexpression analysis of plant transcription factors. *Curr Opini Plant Biol* **6**(5): 430–440.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**(6): 1189–1201.

Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, Bogdanove AJ, Voytas DF (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol* **161**(1): 20–27.

Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, Iida K, KapoorA, Pikaard CS Zhu JK (2008) ROS3 is an RNA-binding protein required for DNA demethylation in *Arabidopsis*. *Nat* **455**(7217): 1259–1262.

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA, Patel DJ, Jacobsen SE (2014) Molecular mechanism of action of plant *DRM de novo* DNA methyltransferases. *Cell* **157**(5): 1050–1060.

Zhu J, Kapoor A, Sridhar VV, Agius F, Zhu JK (2007) The DNA Glycosylase/Lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. *Curr Biol* **17**(1): 54–59.

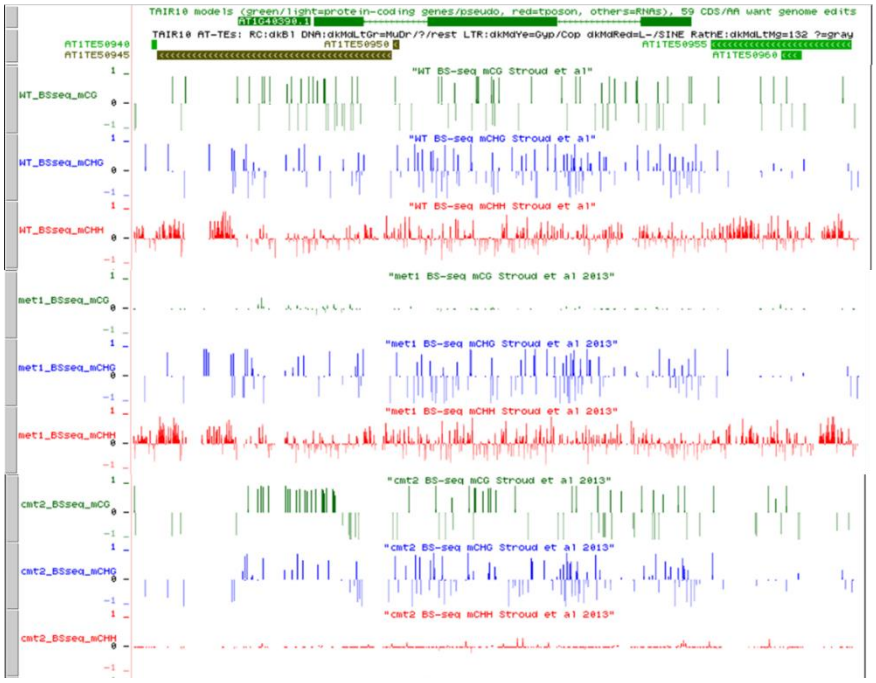
Zhu Z, Sun B, Xu X, Chen H, Zou L, Chen G, Cao B, Chen C, Lei J (2016) Overexpression of *AtEDT1/HDG11* in Chinese Kale (*Brassica oleracea var. alboglabra*) enhances drought and osmotic stress tolerance. *Front Plant Sci* **7**: 1285.

Zubko E, Gentry M, Kunova A Meyer P (2012) *De novo* DNA methylation activity of METHYLTRANSFERASE 1 (MET1) partially restores body methylation in *Arabidopsis thaliana*. *Plant J* **71**(6): 1029–1037.

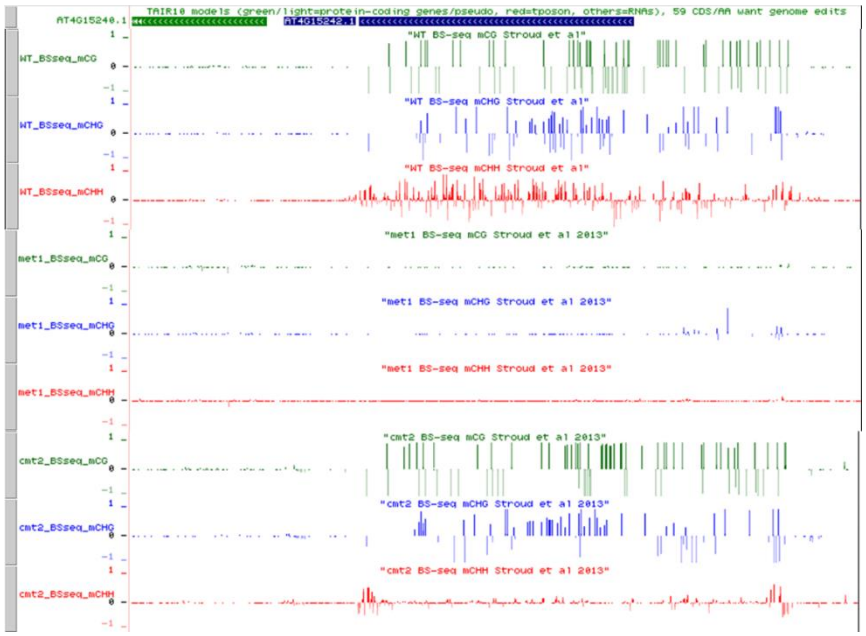
Appendix

Figure 8.1 : Epigenome browser screen shots of several potential CMT2 targets gene as described in Table 5.1. The screen shots were obtained from (http://genomes.mcdb.ucla.edu/AthBSseq/). Methylation types including CG, CHG and CHH are colour coded and represent by different line. The top are TAIR locus number with gene orientation. The first three colour lines are methylation at WT, the second three colour lines are methylation at met1 and the last three colour are methylation at cmt2. Methylation pattern are observed in *Arabidopsis Columbia* (*Col*-0) background

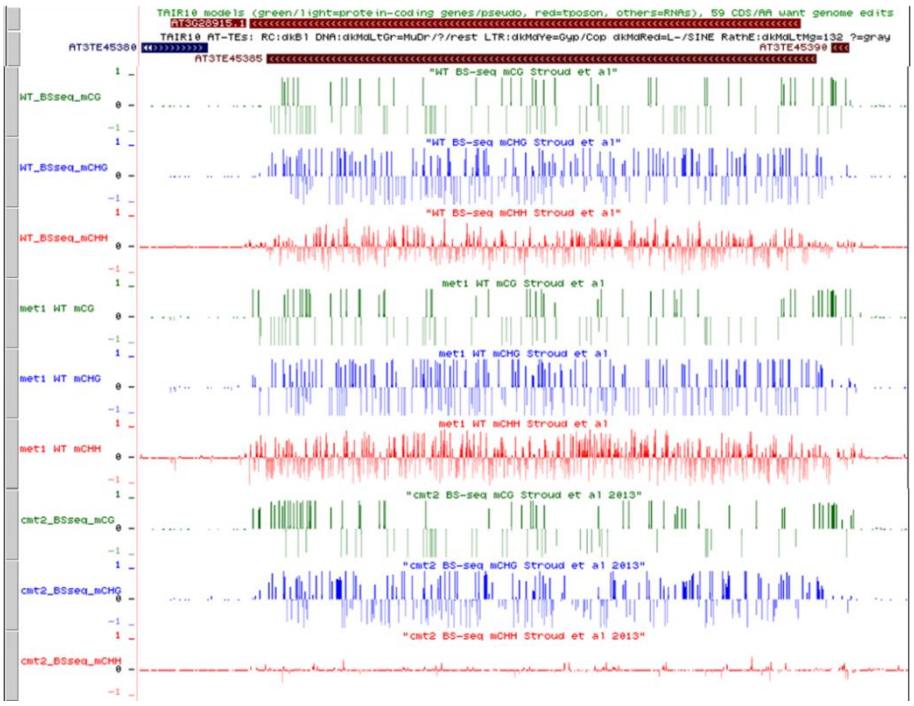
AT1G40390



AT4G15242



AT3G28915



AT5G35935

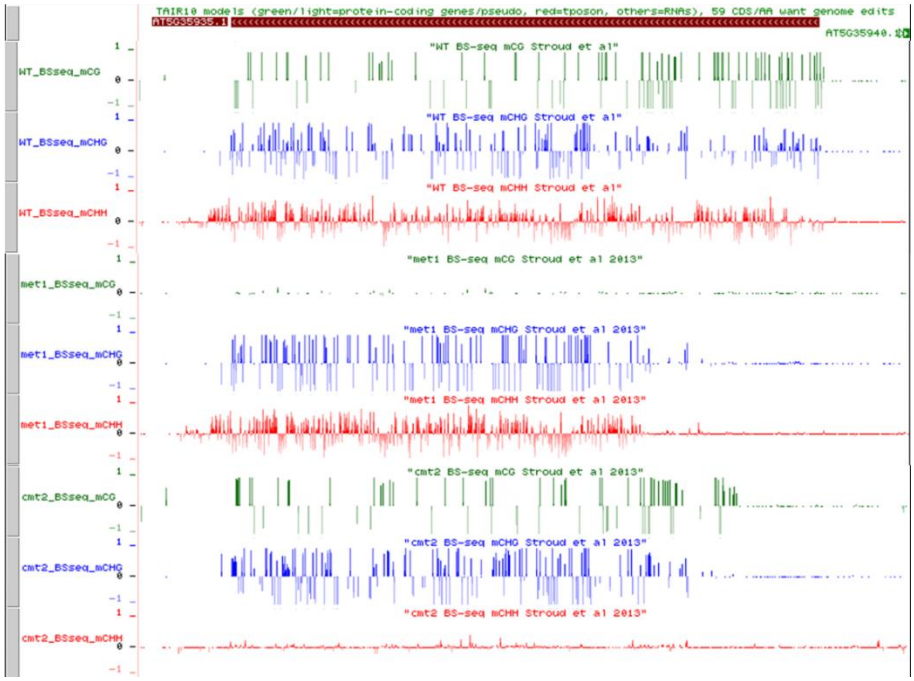


Figure 8.2 : Epigenome browser screen shots and phenotype for AT3G51240 gene. The screen shots were obtained from (http://genomes.mcdb.ucla.edu/AthBSseq/). Methylation types including CG, CHG and CHH are colour coded and represent by different line. The top are TAIR locus number with gene orientation. The first three colour lines are methylation at WT, the second three colour lines are methylation at met1 and the last three colour are methylation at cmt2. Methylation pattern are observed in *Arabidopsis Columbia* (*Col*-0) background. Phenotypes characteristic screen short from TAIR show all mutant at this gene show pale brown seeds coat.

AT3G51240

Locus: AT3G51240 What's new on this page Add a Comment

Representative Gene Model	AT3G51240.1
Gene Model Type	protein_coding
Other names:	F3'H, F3H, FLAVANONE 3-HYDROXYLASE, TRANSPARENT TESTA 6, TT6
Description	Encodes flavanone 3-hydroxylase that is coordinately expressed with chalcone synthase and chalcone isomerases. Regulates flavonoid biosynthesis. Not responsive to auxin or ethylene stimulus (qRT-PCR).
Other Gene Models	AT3G51240.2 (splice variant)

Germplasm

Showing 15 of 29 entries (see all)

name	polymorphisms	background	stock name	select
CS8600	fah1-7; tt6-1		CS8600	not available

images: None available

phenotypes:

- double mutant; altered phenylpropanoid metabolism; disruption of the synthesis of brown pigment in seed coat (testa), pale brown seeds. Phenotype curated by ABRC.

CS8594	tt3-1; tt6-1	CS8594	☐

images: None available

phenotypes:

- double mutant; disruption of the synthesis of brown pigment in seed coat (testa); pale brown seeds. Phenotype curated by ABRC.

CS8593	tt5-1; tt6-1	CS8593	☐

images: None available

phenotypes:

- double mutant; disruption of the synthesis of brown pigment in seed coat (testa); yellow seeds. Phenotype curated by ABRC.