Heritable epigenetic variation of DNA methylation targets in plants

Michael Robert Watson

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Biology

August 2013

Intellectual Property and Publication Statement.

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

Acknowledgements

Firstly, I would like to thank my project supervisor Professor Peter Meyer for his advice, encouragement and direction. I will always be grateful for the professional qualities and scientific knowledge I have gained under his astute supervision.

I also wish to thank my project supervisor Iris Heidmann and her team, including Suzan Out and Brenda de Lange. Their professional conduct made the collaboration with Enza Zaden possible. In particular, I am grateful to Iris for her patience and training with experimental procedures.

Past and present group members, including Dr Elena Zubko, Matt Gentry, Elham Soliman, Yu-Fei Lin and Mohd Ahmad Mokhtar created a stimulating and enjoyable working environment, which made tough times easier. For this, I am thankful.

Finally, I owe a great deal of gratitude to my family, including Robert, Maria and Rebecca Watson and Sarah Cruise. I am indebted to them for their support and love.

Abstract

DNA methylation marks regulate gene expression and genome structure. Stability and dynamics of DNA methylation patterns are influenced by four major factors including *de novo* methylation, maintenance methylation, passive loss of methylation and active demethylation. Maintenance methylation functions are still well conserved among plants and animals, which separated more than 1.5 billion years ago. In contrast, demethylation mechanisms differ considerably among plants and mammals. Interfering with DNA methylation and demethylation systems could be a source of heritable epigenetic variation if DNA methylation changes are introduced that transcend into stable heritable gene expression changes. The high tolerance of plants to DNA methylation changes makes them an ideal experimental system to exploit DNA methylation and demethylation systems. In this study, four strategies have been developed and tested for their capacity to induce heritable epigenetic variation by interfering with DNA methylation and demethylation systems. These strategies included a chemical treatment with a DNA methylation inhibitor, genetic demethylation using a mutant deficient in the maintenance methyltransferase MET1 and transgenic approaches to over-express MET1 and to express the human TET3 demethylase. While chemical demethylation only generated non-heritable changes, inactivating MET1 induced stable DNA methylation and expression changes at specific loci. Expression of the human TET3 protein also induced locus-specific loss of methylation but the efficiency of demethylation varied in individual transformants independent of TET3 level, which suggests that demethylation is locus-specific but stochastic.

Table of Contents

List of Abbreviations

5-aza	5-azacytidine
A	Adenine
AGO	Argonaute
AI	Alcohol Inducible
AID	Activation-Induced Deaminase
alf	Aberrant lateral root formation
AP	Apurinic/Apyrimidinic
APOBEC	Apolipoprotein B RNA Editing Catalytic Component
aux	Auxin resistant
BAH	Bromo-Adjacent Homology
BER	Base Excision Repair
bp	Base pairs
C	Cytosine
caC	5-carboxyl-cytosine
CaMV	Cauliflower Mosaic Virus
cDNA	Complementary Deoxyribonucleic Acid
CHD	Chromo-Domain
CMT2	Chromomethyltransferase 2
CMT3	Chromomethyltransferase 3
CNR	Colourless Non Ripening
Col	Arabidopsis Columbia
CRP	Cysteine Rich Peptidase
D1	Dwarf1
DAM	DNA Adenine Methyltransferase
DCL	Dicer-Like
DCM	DNA Cytosine Methyltransferase
DDM1	Decrease in DNA Methylation 1
Decitabine	5-aza-2′-Deoxycytidine
DGT	Diageotropica
DME	Demeter
DML	Demeter-Like
DNA	Deoxyribonucleic Acid
DNMT1	DNA Methyltransferase 1
DNMT3A	DNA Methyltransferase 3 A
DNMT3B	DNA Methyltransferase 3 B
DNMT3L	DNA Methyltransferase 3-Like
DRM2	Domains Re-Arranged Methyltransferase 2
ds	Double-stranded
DSBH	Double-Stranded β-Helix
DWF	Dwarf
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic Acid
EMS	Ethyl Methanesulfonate
EpiRILs	Epigenetic Recombinant Inbred Lines
EVD	Evade
fC	5-formyl-Cytosine
FIS2	Fertilization-Independent Seed 2
FT	Flowering Locus T

FWA	Flowering Wageningen
G	Guanine
GK	Glycine-Lysine
GUS	β–Glucuronidase
HDA6	Histone Deacetylase 6
HEK	Human Embryonic Kidney
HI	Heat Inducible
hmC	5-hydroxy-methylcytosine
HOG1	Homology-Dependent Gene Silencing 1
IAA	Isoamyl-Alcohol
KTF1	Kow Domain-Containing Transcription Factor 1
LB	Lysogeny broth
Ler	Arabidopsis Landsberg *erecta*
LINE	Long Interspersed Nuclear Element
LSH1	Lymphoid Specific Helicase 1
MAS	Marker Assisted Selection
MBD	Methyl-CpG Binding Domain
MBD1	Methyl-CpG Binding Domain protein 1
MBD4	Methyl-CpG Binding Domain protein 4
mC	5-methyl-Cytosine
MEA	MEDEA
MEE1	Maternally Expressed in Embryo 1
MET1	Methyltransferase 1
MET1.Rest	MET1 Restored
MORC	Microrchidia
mRNA	Messenger Ribonucleic Acid
MS	Murashige and Skoog
MS-AFLP	Methylation-sensitive-Amplified Fragment Length Polymorphism
MULE	Mutator-Like Element
NASC	Nottingham Arabidopsis Stock Centre
nc	Non-Coding
NCBI	National Center for Biotechnology Information
NLS	Nuclear Localisation Sequence
nt	Nucleotide
NORs	Nucleolus Organiser Regions
NOS	Nopaline synthase
NPT	Neomycin PhosphoTransferase
oxBS-Seq	Oxidative Bisulfite Sequencing
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PEV	Position Effect Variegation
PGC	Primordial Germ Cells
pi	Piwi-Interacting
POLII	RNA Polymerase II
POLIV	RNA Polymerase IV
POLV	RNA Polymerase V
PTGS	Post-Transcriptional Gene Silencing
RD29A	Responsive to Desiccation 29A
RdDM	RNA-directed DNA Methylation
rDNA	Ribosomal DNA

RDR2	RNA-Dependent RNA polymerase 2
RDR6	RNA-Dependent RNA polymerase 6
RFTS	Replication Foci Targeting Sequence
RIN	Ripening Inhibitor
RNA	Ribonucleic acid
ROS	Repressor of Silencing
RPS	Repetitive Petunia Sequence
SAM	S-Adenosyl Methionine
SDS	Sodium Dodecyl Sulphate
si	Small Interfering
SNP	Single Nucleotide Polymorphism
sq	Semi-Quantitative
SRA	Set and Ring Associated
sup	Superman
SUVH4	Suppressor of Variegation 3-9 Homologue 4
T	Thymine
TAE	Tris-acetate-EDTA
TBE	Tris-Borate-EDTA
TDG	Thymine DNA Glycosylase
T-DNA	Transfer-DNA
TE	Transposable Element
TET	Ten-Eleven Translocase
TGS	Transcriptional Gene Silencing
tomMET1	Tomato Methyltransferase 1
VCN	Vegetative Cell Nucleus
WRKY	WRKYGQK
Ws	Arabidopsis Wassilewskija
WT	Wild-type
Zebularine	1-β-D-ribofuranosyl-1,2-dihydropyrimidine-2-one

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 10

1.0. General Introduction

1.1. An Introduction to Epigenetics

In 1942 Conrad Waddington coined the term epigenetics to describe how a genotype gives rise to a phenotype during development (Waddington, 1942). Waddington illustrated his understanding with a visual metaphor called the epigenetic landscape. In this model an undifferentiated totipotent cell is represented by a ball at the top of a hill. The ball will role down the hill following a specific pathway. This pathway is determined by the different peaks and troughs the ball encounters, which represent the developmental commitments of the cell. Over time the definition of epigenetics has broadened to encompass its roles in genome structure and regulation. In 2007 Adrian Bird defined epigenetics as "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird, 2007). This proposal would include processes that are not mitotically or mieotically heritable and is therefore considered wide-ranging by some (Ledford, 2008). A year later, in 2008, a Cold Spring Harbor meeting was hosted to arrive at a consensus definition of epigenetics, where epigenetics was defined as reversible, heritable changes in gene expression without any changes to the underlying DNA sequence (Berger et al, 2009).

Epigenetic gene expression changes have explained phenomena that deviate from 'normal' mendelian genetics and among others include imprinting (Dechiara et al, 1991), position effect variegation (PEV) (Muller, 1930), transgene silencing (Meyer et al, 1992; Napoli et al, 1990) and paramutation (Brink, 1959). Epigenetic gene expression changes are regulated by post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS) mechanisms. PTGS occurs through degradation (Fire et al, 1998) of target mRNA, whereas TGS occurs through changes in chromatin conformation. The nucleosome represents the first level of chromatin, where 146 base pairs (bp) of DNA is wrapped around a histone octomer assembled from 2 copies of histone proteins 2A, 2B, 3 and 4 (Luger et al, 1997). The N-terminal tails of histone proteins can be modified to distinguish between transcriptionally active euchromatin and transcriptionally repressive heterochromatin (Figure 1.1) (Meyer, 2001). Among these modifications are histone methylation, acetylation and phosphorylation. In plants histone modifications often associated with transcriptionally repressive

heterochromatin include mono- and di-methylation of histone 3 lysine 9 (H3K9me1 and H3K9me2), mono- and di-methylation of histone 3 lysine 27 (H3K27me1 and H3K27me2) and mono-methylation of histone 4 lysine 40 (H4K40me1) (Pfluger & Wagner, 2007). Among the histone modifications that correlate with active genes are acetylation of histone 3 lysine 9 (H3K9ac) and acetylation of histone 3 lysine 27 (H3K27ac) (Lauria & Rossi, 2011).

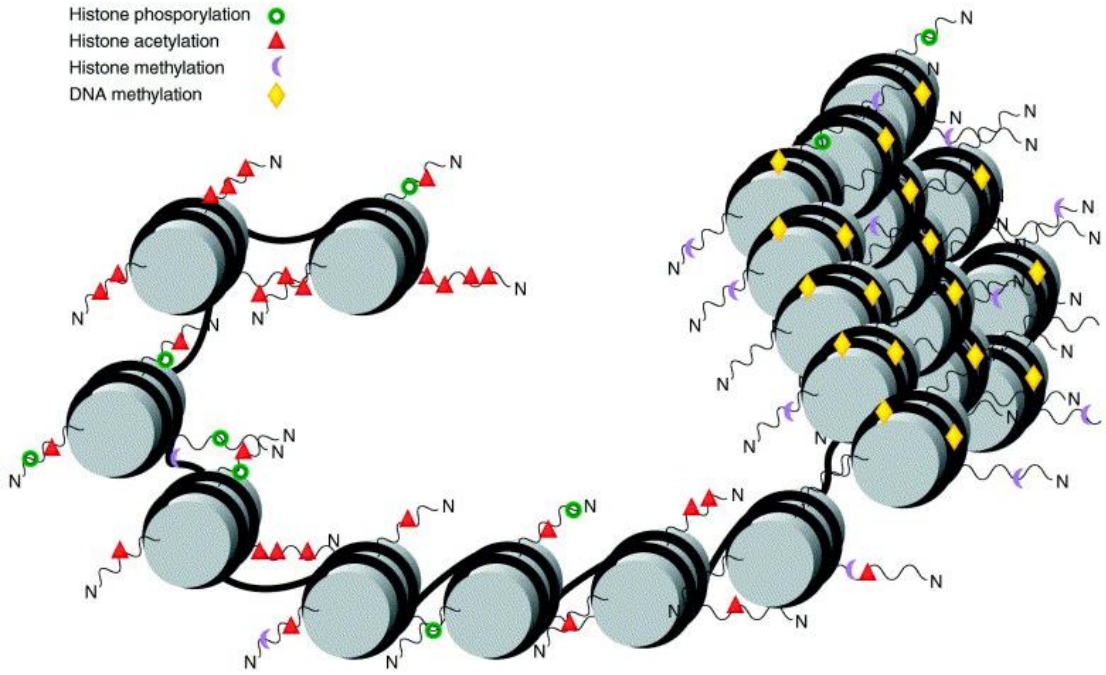


Figure 1.1. Euchromatic and heterochromatic chromatin states. The figure shows two chromatin states and their associated modifications. On the left is transcriptionally active euchromatin and on the right is transcriptionally repressive heterochromatin. Histone octamers are represented as blue cylinders with protruding N-terminal tails of histones labelled with the letter 'N'. DNA, which is rapped around the histone octamer, is represented by the thick black line. Modifications are illustrated as coloured shapes using a key in the top left of the image. Taken from Meyer, (2001).

DNA methylation, a modification often associated with gene repression, offers another layer of epigenetic control. It involves the transfer of a methyl-group ($-CH_3$) from S-adenosyl methionine (SAM) to a DNA base. Histone modifications and DNA methylation cross-talk to co-ordinate chromatin structure and regulate gene expression (Li, 2002). While histone modifications are readily reversible, DNA methylation patterns can vary and are often maintained over generations (Vaughn et

al, 2007). This project therefore focuses on the modification of DNA methylation marks to induce heritable epigenetic variation in plants.

1.2. DNA methylation

DNA methylation has been found in bacteria, fungi, plants and animals but the sequence contexts and epigenetic landscapes can vary among them.

In bacteria, DNA methylation occurs at nitrogen 6 of adenine (^{m6}A), carbon 5 of cytosine (^{m5}C) and in some cases nitrogen 4 of adenine (^{m4}A). In *E. coli*, ^{m6}A is established by the DNA ADENINE METHYLTRANSFERASE (DAM) at the sequence GATC and ^{m5}C is established by the DNA CYTOSINE METHYLTRANSFERASE (DCM) at the internal C of the two sequences CCWGG (where W is an A or a T) (Casadesus & Low, 2006). In bacteria, DNA methylation forms part of a restriction-modification system, which is used as a defence mechanism against foreign DNA. The system requires restriction endonucleases, which are bacterial enzymes that recognise palindromic DNA sequences. Methylated DNA cannot be cleaved by restriction endonucleases, unlike unmethylated invading phage genomes (Casadesus & Low, 2006).

Many studies of DNA methylation in animals have been carried out in mammalian systems where DNA methylation is found exclusively at the carbon 5 of cytosines. Work on individual sequences and genomic DNA digests with methylation-sensitive restriction enzymes provided the first indications that mammalian genomes are globally methylated (Suzuki & Bird, 2008). Long contiguously methylated domains are occasionally interrupted by unmethylated regions called CpG islands (Suzuki & Bird, 2008), which have a higher CG content than the genome average and associate with gene promoters (Larsen et al, 1992) and origins of replication (Antequera & Bird, 1999). DNA methylation in mammals is found at cytosines in a CG sequence context with the exception of embryonic stem cells, where methylation has been found at cytosines in CA and CT sequence contexts (Ramsahoye, 2000).

In comparison to mammals, DNA methylation in plants is also found exclusively at the carbon 5 of cytosines. In contrast, in plants DNA methylation is found at CG,

CHG and CHH (where H represents an A, T or G) sequence contexts. Shotgun bisulfite sequencing revealed that 24% of CG, 6.7% of CHG and 1.7% of CHH sequences are methylated in *Arabidopsis* (Cokus et al, 2008). At transposable elements (TEs) and repetitive sequences all three sequence types of methylation are highly correlated but when methylation is present within the body of protein coding genes it is almost entirely in a CG context (Cokus et al, 2008). The methylation patterns in *Arabidopsis* create a mosaic landscape, where DNA methylation is interspersed throughout the genome (Cokus et al, 2008; Zhang et al, 2006).

1.3. Eukaryotic DNA methyltransferases

1.3.1. Eukaryotic *de novo* DNA methyltransferases

In mammals, reprogramming of DNA methylation patterns occurs at two stages of the life cycle (Monk et al, 1987) and requires *de novo* methyltransferases to establish DNA methylation. DNA METHYLTRANSFERASE 3 A (DNMT3A) and DNA METHYLTRANSFERASE 3 B (DNMT3B) restore DNA methylation lost in the cell cycles before blastulation (Law & Jacobsen, 2010). DNMT3A and DNA METHYLTRANSFERASE 3-LIKE (DNMT3L) establish DNA methylation at imprinted genes and TEs in primordial germ cells (PGCs), which give rise to the germline in mammals (Law & Jacobsen, 2010). DNA methylation is targeted to imprinted genes when DNMT3L associates with unmethylated histone 3 lysine 4 (H3K4) and recruits DNMT3A (Ooi et al, 2007). DNA methylation is targeted to TEs via a class of small interfering (si) RNAs, called piwi-interacting (pi) RNAs, which target *de novo* methyltransferases to homologous sequences during male gametogenesis (Law & Jacobsen, 2010). In plants, DNA methylation patterns appear static across generations and there is no clear evidence for an extensive reprogramming of DNA methylation in the developing embryo. *De novo* methylation in plants is controlled by siRNA dependent (Figure 1.2) and independent pathways.

RNA dependent *de novo* methylation is controlled by the mammalian DNMT3 homologue DOMAINS RE-ARRANGED METHYLTRANSFERASE 2 (DRM2). *De novo* activity of DRM2 was identified when loss of function resulted in maintenance of pre-existing DNA methylation at an endogenous target gene but blockage of *de*

novo methylation at the same target when introduced as a transgene (Cao & Jacobsen, 2002b). The identification of *de novo* methyltransferase activity in plants led to the characterisation of the targeting pathway, RNA-directed DNA methylation (RdDM) (Law & Jacobsen, 2010). In this pathway the chromatin binding protein SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) enables the recruitment of RNA POLYMERASE IV (POLIV) to target loci (Law et al, 2013). Double-stranded (ds) RNA is produced by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) (Chan, 2004), which synthesises the complementary strand of POLIV transcripts. dsRNA is cleaved into 24-nt siRNAs by the DICER-LIKE 3 (DCL3) (Chan, 2004) nuclease, which associates with ARGONAUTE 4 (AGO4) (Chan, 2004). RNA POLYMERASE V (POLV) non-polyadenylated, uncapped transcripts act as scaffold to recruit the AGO4-siRNA complex (Wierzbicki et al, 2008), which is facilitated by SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5-like)/KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1) (He et al, 2009; Law & Jacobsen, 2010). The effector complex directs DRM2 to local DNA (Figure 1.2). Argonaute proteins also possess endoribonuclease activity, which can lead to cleavage of locus-specific POLV transcripts (Qi et al, 2006). dsRNA could be generated from these transcripts by RDR2, leading to the production of secondary siRNAs, resulting in a self-enforcement effect.

Although RdDM influences *de novo* methylation at all three sequence types in plants, recent studies have shown that *de novo* methylation at CHH sites can be initiated independently of siRNAs. This requires CHROMOMETHYLTRANSFERASE 2 (CMT2) (Zemach et al, 2013), which is a member of the CHROMOMETHYLTRANSFERASE family, a unique class of DNA methyltransferases found in plants. It has been hypothesised that targeting of CMT2 occurs through an interaction with di-methylation at lysine 9 on histone 3 (H3K9me2) via its chromo-domain (Pikaard, 2013; Zemach et al, 2013).

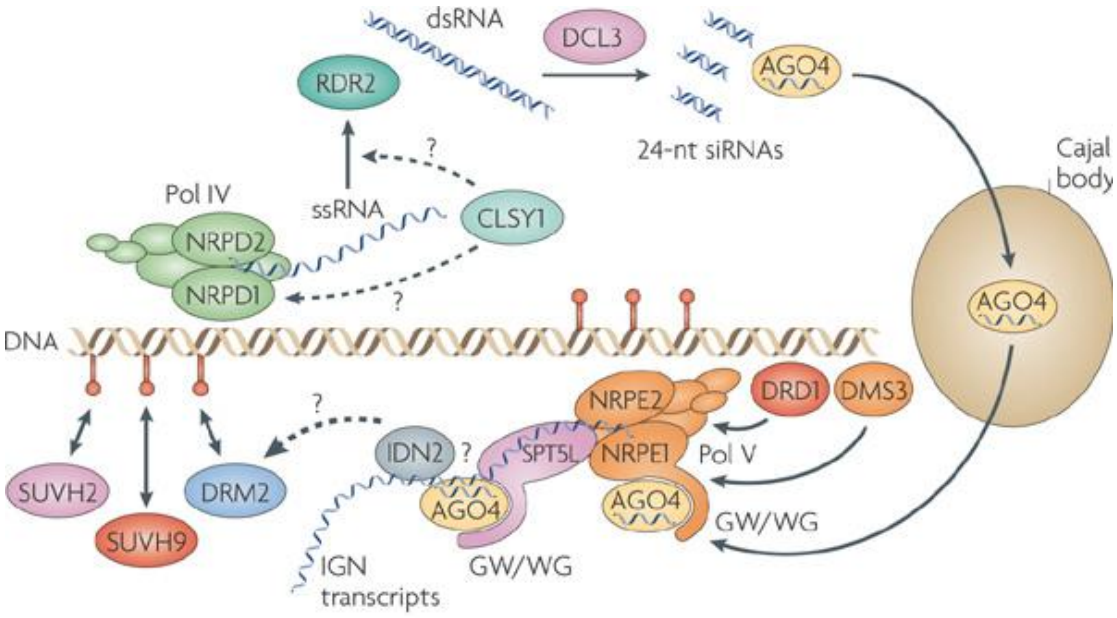


Figure 1.2. RNA-directed DNA methylation in plants. The figure shows key steps in the targeting of DNA methylation via siRNAs in plants. DNA and RNA are illustrated by the brown double helix and blue wavy lines, respectively. Attached methyl groups are represented by the orange pins. dsRNA is produced by RDR2 which synthesizes the complementary strand of POLIV transcripts. dsRNA is then cleaved by the nuclease DCL3 to generate 24-nt siRNAs that associate with AGO4. siRNA-AGO4 complexes are recruited by POLV transcripts, which signals *de novo* methylation by DRM2. Taken from Law & Jacobsen, (2010).

1.3.2. Eukaryotic maintenance DNA methyltransferases

Semi-conservative DNA replication results in each daughter cell inheriting hemi-methylated DNA. To prevent loss of DNA methylation via semi-conservative replication, maintenance methyltransferases recognise hemi-methylated DNA and methylate the symmetrical cytosine. In mammals, DNA METHYLTRANSFERASE 1 (DNMT1) (Crowson & Shull, 1992) maintains DNA methylation at CG sequence contexts. It associates at replication foci via interactions with components of the replication machinery, including PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) (Chuang et al, 1997). The chromatin remodelling factor LYMPHOID SPECIFIC HELICASE 1 (LSH1) is also required for DNMT1 function, but its exact role remains to be elucidated (Dennis et al, 2001).

Maintenance of CG methylation in plants is catalysed by the DNMT1 homologue METHYLTRANSFERASE 1 (MET1) (Kankel et al, 2003). Like DNMT1, MET1 maintenance activity requires a chromatin remodelling factor, DECREASE IN DNA

METHYLATION 1 (DDM1), a homologue of the mammalian LSH1 (Hirochika et al, 2000). However, DNA methylation in a CHH sequence context cannot be maintained in this pathway because of its asymmetry. Instead, CHH methylation is maintained by constant *de novo* activity (Law & Jacobsen, 2010). CHG methylation is maintained by CHROMOMETHYLTRANSFERASE 3 (CMT3) (Lindroth et al, 2001) via a self-enforcing loop between histone and DNA methylation. SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4 (SUVH4), a H3K9 histone methyltransferase, contains a Set and Ring Associated (SRA) domain. Mutation studies and mobility shift assays show that the SRA domain of some histone methyltransferases specifically bind methylated DNA, and that the SRA domain of SUVH4 preferentially binds CHG and CHH sequence types (Johnson et al, 2007; Lindroth et al, 2004). Conversely, the chromodomain of CMT3 has the capacity to bind H3K9 methylated histones, suggesting that histone methylation by SUVH4 recruits CMT3 (Lindroth et al, 2004).

1.3.3. MET1 effects

Recent evidence suggests that the plant DNA methyltransferase MET1 may not be restricted to a CG maintenance function. For example, Zubko et al, (2012) found that methylation lost from the body of an endogenous target gene in a *met1 Arabidopsis* mutant, was partially restored at CG sites when MET1 was re-introduced. Re-methylation did not require passage through the germline, which suggests MET1 may have *de novo* activity at CG sequence contexts (Zubko et al, 2012). MET1 may also influence non-CG methylation as Singh et al (2008) showed a reduction in both CG and non-CG methylation at a *REPETITIVE PETUNIA SEQUENCE* (*RPS*) when introduced into a *met1 Arabidopsis* mutant by a genetic cross. Similarly, both CG and non-CG methylation were eliminated at the *RPS* when transferred into a *drm2/cmt3* mutant. These observations lead to the hypothesis that MET1, DRM2 and CMT3 may establish methylation at the *RPS* by gaining access jointly or the two methylation systems may recruit each other (Singh et al, 2008). For example, CG methylation maintained by MET1 may be required for binding of DRM2 and CMT3 guiding factors (Singh et al, 2008).

The coupling of DNA methylation and histone methylation via the SRA domain of histone methyltransferases (Johnson et al, 2007) suggests that the effects of MET1 maintained CG methylation may extend into the histone modification layer of epigenetic control. This was shown when H3K9 methylation, a mark associated with transcriptionally silent heterochromatin, is lost when CG methylation is completely removed in a *met1 Arabidopsis* mutant (Tariq et al, 2003). H3K9 methylation however, is unaffected when non-CG methylation is lost in a *cmt3* mutant. (Tariq et al, 2003). Therefore, MET1-regulated CG methylation may function in heterochromatin formation by acting as a scaffold to direct H3K9 methylation (Tariq et al, 2003). Interestingly, in an *Arabidopsis met1* mutant, H3K9 methylation appears to accumulate within genes, which has been assigned to down-regulation of the H3K9 demethylase INCREASE IN BONSAI METHYLATION 1 (IBM1) (Rigal et al, 2012). Conveniently down-regulation of IBM1 also provides an explanation for the unusual observation that CHG methylation accumulates within gene bodies in *met1*, as H3K9 methylation is a requisite for CMT3 targetted CHG methylation (Lindroth et al, 2004; Rigal et al, 2012).

To ensure proper regulation of the genome and to defend against mobilisation of TEs, repressive heterochromatin states need to be maintained, which requires DNA methylation and histone modifiers. Liu et al, (2011) showed that the N-terminal domain of MET1 directly interacts with the C-terminal domain of HISTONE DEACETYLASE 6 (HDA6), which suggests MET1 and HDA6 may function co-operatively to maintain heterochromatic gene silencing (Liu et al, 2011).

1.4. DNA demethylation

Establishment and removal of DNA methylation are required by eukaryotes to fine tune gene expression and respond to environmental and developmental signals. DNA demethylation describes the process of passive loss or active removal of DNA methylation.

Passive loss of DNA methylation results from SAM shortages, DNA methyltransferase down-regulation or dysfunction. DNA methyltransferase dysfunction is promoted by conditions such as nucleoprotein blockage (Hsieh, 1999a),

where regulatory nucleoproteins occlude target sequences from methyltransferases and acetylated histone aversion (Wolffe et al, 1999), where acetylated histones repel methyltransferases. Active DNA demethylation describes an enzymatic action where a methylated cytosine is replaced with an unmethylated cytosine and occurs in both plants and animals.

1.4.1. Active DNA demethylation in plants

In *Arabidopsis*, there are four DNA glycosylases including REPRESSOR OF SILENCING 1 (ROS1), DEMETER (DME), DEMETER-LIKE 2 (DML2) and DEMETER-LIKE 3 (DML3). They excise methylated cytosines from DNA by hydrolysing the glycosidic bond between the cytosine and sugar-phosphate backbone. After base removal, DNA glycosylases use apurinic/apyrimidinic (AP) lyase activity to nick the DNA backbone, which induces base excision repair (BER) mechanisms to fill the gap with an unmethylated cytosine (Figure 1.3).

ROS1, DML2 and DML3 are active in vegetative cells and function in sequence specific removal of DNA methylation at the 5' and 3' ends of genes (Penterman et al, 2007). ROS1 counteracts RdDM, possibly to prevent hypermethylation and DNA methylation spreading by self-reinforcement mechanisms that could lead to detrimental gene silencing. The counteracting mechanism is at least partly reliant on a feedback mechanism involving DNA methylation and DNA methyltransferases. For example, when DNA methylation levels are reduced in a *met1* mutant or by treatment with DNA methylation inhibitors, *ROS1* levels are down-regulated (Mathieu et al, 2007). It is still not clear how plant DNA glycosylases are targeted to specific sequences. ROS1 co-localises with REPRESSOR OF SILENCING 3 (ROS3), which possesses RNA binding capacity (Zheng et al, 2008), making an RNA-based targeting system appealing.

DNA glycosylases likely contribute to the global hypomethylation in the *Arabidopsis* endosperm (Hsieh et al, 2009), a nutrient reservoir, which surrounds the embryo. In *Arabidopsis*, the endosperm and embryo are produced via double fertilisation (Figure 1.4). Each pollen grain contains two sperm cells. These fertilize the egg and diploid central cell of the female gametophyte, giving rise to the embryo and triploid

endosperm, respectively. In the central cell during female gametogenesis *MET1* levels are reduced (Jullien et al, 2008) and *DME* levels are increased (Choi et al, 2002), resulting in the loss of DNA methylation via both passive and active mechanisms. Therefore, global hypomethylation in the endosperm may be an extension of hypomethylation in the central cell.

1.4.2. Active DNA demethylation in mammals

Mammalian DNA glycosylases include METHYL-BINDING PROTEIN 4 (MBD4) and THYMINE DNA GLYCOSYLASE (TDG). Unlike plants, mammalian DNA glycosylases show weak activity against methylated cytosines compared to thymines in T/G mismatches (Zhu et al, 2000). For global epigenetic reprogramming during development and sequence specific demethylation in response to environmental signals in mammals, it is therefore required that cytosines are modified prior to glycosylase intervention.

ACTIVATION-INDUCED DEAMINASE (AID) (Cortellino et al, 2011) and APOLIPOPROTEIN B RNA EDITING CATALYTIC COMPONENT 1 (APOBEC1) (Harris et al, 2003) deaminate methylated cytosine to thymine. This creates a T/G mismatch. The thymine is removed by TDG or MBD4 and replaced in the BER pathway (Figure 1.3). Alternatively, methylated cytosines can be oxidised to either 5-hydroxy-methylcytosine (hmC), 5-formyl-cytosine (fC) or 5-carboxyl-cytosine (caC). This oxidation is catalysed by three TEN-ELEVEN TRANSLOCASE (TET1-3) proteins (Ito et al, 2011). Oxidised bases are removed by glycosylase activity and replaced in the BER pathway (Figure 1.3). TET3 has been found highly expressed (Gu et al, 2011) in the paternal pro-nucleus and therefore likely contributes to the global epigenetic reprogramming during early mammalian developmental stages. TET activity may also facilitate DNA demethylation passively as DNMT1 does not recognise hmC (Valinluck & Sowers, 2007).

Interestingly, recent evidence has emerged from *in vitro* assays that the mammalian *de novo* methyltransferases, DNMT3A and DNMT3B may function as hmC dehydroxymethylases (Chen et al, 2012) (Figure 1.3). A dehydroxymethylation function of DNMT3 proteins would offer an alternative DNA demethylation pathway

and allow the reversion of hmC to mC (Chen et al, 2012). The latter would be useful to correct TET3 errors or hmC produced by natural oxidation.

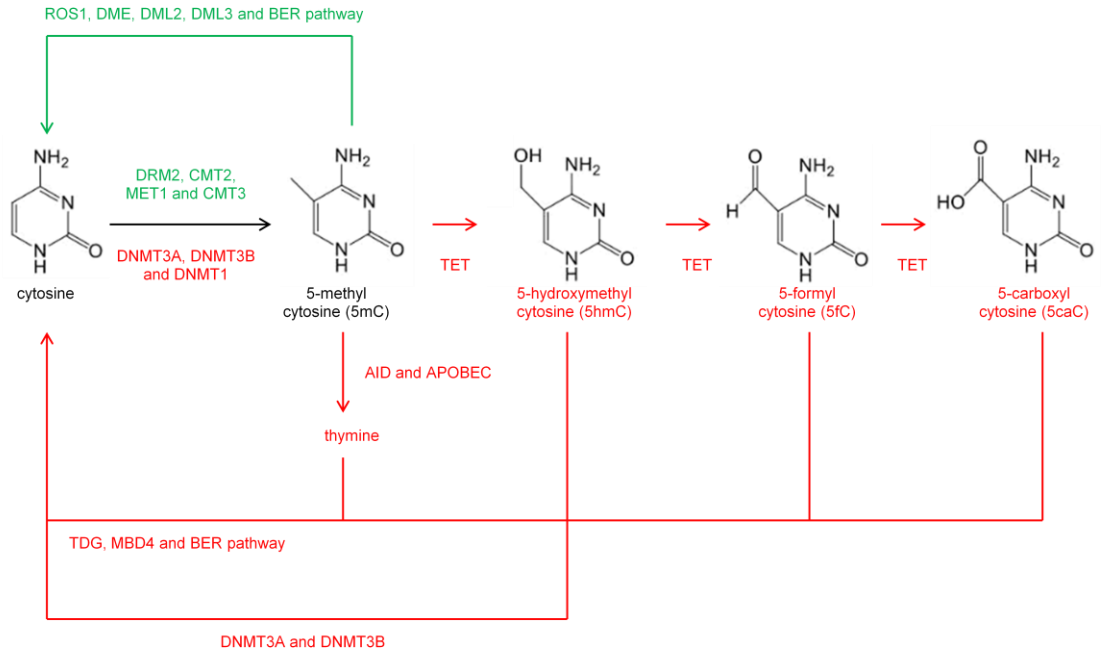


Figure 1.3. Active DNA demethylation pathways in plants and mammals. Green, Red and **Black** labels and arrows indicate proteins and cytosine modifications that occur in plants, mammals or **both** respectively. In plants the DNA glycosylases ROS1, DME, DML2 and DML3 excise methylated bases. Excised bases are replaced with an unmethylated cytosine by BER mechanisms. In mammals DNA demethylation can occur via at least three pathways. A methylated cytosine can be deaminated to thymine by the cytosine deaminases AID and APOBEC resulting in a T/G mismatch. Thymines in a T/G mismatch are excised by the glycosylases TDG and MBD4 and replaced with an unmethylated cytosine by BER mechanisms. In mammals methylated cytosines can also be oxidised by TET proteins. Oxidative products including 5hmC, 5fC and 5caC can be targeted by DNA glycosylases for excision. Alternatively the mammalian *de novo* methyltransferases DNMT3A and DNMT3B may serve as dehydroxymethylases converting hmC to C. Image modified from Ito et al, (2011).

1.5. The biological roles of DNA methylation in plants

DNA methylation within gene promoters is often associated with transcriptional repression, as it can directly obstruct transcription factors and recruit methyl-binding proteins that signal chromatin changes. Gene silencing is required to prevent the activation of TEs, which could threaten genome integrity by inserting into critical genes. The genomes of higher plants contain many TEs that could potentially disrupt genome integrity. In the *Arabidopsis ddm1* mutant, TEs are activated but only mobilise after repeated self-pollinations, which can generate mutations. *ddm1* induced mobilisation of the *CAC1* transposon into the *DWF4* locus produced the defective in

stem and leaf elongation mutant, *clam* (Miura et al, 2001). Similarly, the *EVD* retrotransposon is reverse transcribed in *met1* epigenetic recombinant inbred lines (epiRILs) and inserts within unlinked loci (Mirouze et al, 2009). Despite the link between hypomethylation and TE activation it is also important to acknowledge that in some cases derepression of TEs can occur in the absence of DNA methylation changes. *Arabidopsis* MICRORCHIDIA (MORC) proteins belong to an ATPase family and likely play a role in chromatin superstructure (Moissiard et al, 2012). *Arabidopsis morc* mutants show derepression of methylated genes and TEs without any DNA methylation changes, highlighting a potential role for MORC proteins in DNA methylation-independent gene silencing (Moissiard et al, 2012).

Silencing of repeats via DNA methylation is not restricted to foreign elements. *Ribosomal DNA* (*rDNA*) in *Arabidopsis* is arranged in tandem arrays at two Nucleolus Organiser Regions (NORs). Transcription of *rDNA* is a large energy consuming process due to the demand for rRNA in a cell and is therefore tightly regulated. Part of this regulation is the switching "on" and "off" of *rDNA* repeats, which requires DNA methylation. Interestingly, a positive correlation has been reported between *rDNA* copy number and NOR DNA methylation levels among 41 *Arabidopsis* accessions (Woo & Richards, 2008). This correlation suggests that *rDNA* copy number itself is a determinant of NOR DNA methylation levels and provides insight into a possible mechanism by which *Arabidopsis* accessions silence excess *rDNA* repeats (Woo & Richards, 2008). In this study the inheritance of NOR methylation levels were also analysed in F1 lines derived from crossing accessions with low and high NOR methylation levels (Woo & Richards, 2008). F1 lines could be divided into three classes, including those with intermediate, low or high NOR methylation levels, suggesting that NOR methylation is regulated by faithful inheritance of parental NOR methylation patterns but also reconfiguration of NOR methylation in the hybrids (Woo & Richards, 2008).

In plants, hypermethylated TEs and repeat regions show a low meiotic recombination rate compared to hypomethylated low-copy number genes, which indicates that DNA methylation may influence the rate of recombination (Melamed-Bessudo & Levy, 2012). When recombination was analysed in the hypomethylated mutant *ddm1* the rate of recombination between markers located in euchromatin increased, whereas

rates were unchanged between markers located in heterochromatin. This is surprising considering that heterochromatic regions are most affected by demethylation in the *ddm1* mutant, and would suggest that DNA methylation only has a repressive function in meiotic recombination of euchromatin (Melamed-Bessudo & Levy, 2012).

DNA methylation plays a central role in imprinting, a phenomenon by which genes are expressed in a parent-of-origin specific manner. Imprinting is widespread in plants and found in the endosperm during seed development (Jahnke & Scholten, 2009). Examples of imprinted genes in plants include the *Arabidopsis MEDEA (MEA), FERTILIZATION-INDEPENDENT SEED 2* (*FIS2),* (Luo et al, 2000) and *FLOWERING WAGENINGEN (FWA)* (Kinoshita et al, 2004) genes, which are all expressed from the maternal alleles of the endosperm. This most likely originates from global demethylation in the central cell of the female gametophyte (Choi et al, 2002; Jullien et al, 2008) (Figure 1.4), which gives rise to the endosperm after fertilisation. Global hypomethylation in the endosperm could facilitate silencing of repetitive sequences in the embryo, if siRNAs derived from active TEs in the endosperm migrated into the embryo. Migration of siRNAs has been observed in pollen grains (Slotkin et al, 2009), which are comprised of a vegetative cell nucleus (VCN) and two sperm cells. The genome of the VCN is demethylated due to *DDM1* and *DME* down and up-regulation, respectively (Figure 1.4). Consequently, TEs are activated in the VCN and siRNAs derived from these TEs have been found to silence target genes in sperm cells (Slotkin et al, 2009).

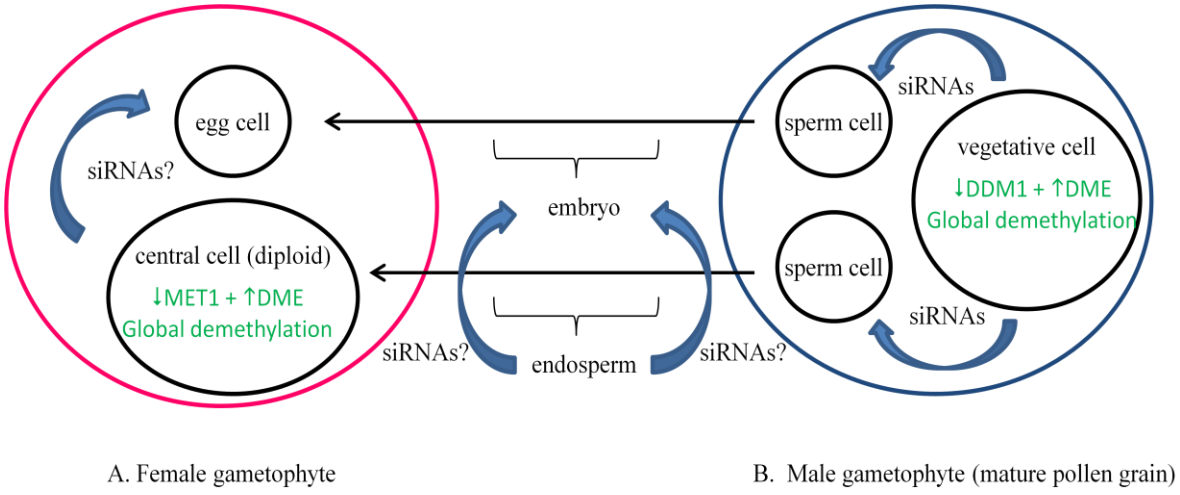


Figure 1.4. Double fertilisation in flowering plants. The figure shows some key epigenetic regulatory networks in plant gametophytes mediated by siRNAs. Two sperm cells fertilise the egg and central cell of the female gametophyte to give rise to the embryo and endosperm. siRNAs have been shown to migrate from the vegetative cell to silence target loci in sperm cells. This leads to the speculations that siRNAs may migrate between the central cell and egg cell of the female gametophyte and between the endosperm and embryo to silence target genes. Speculative pathways are marked with a "?".

Correct DNA methylation patterns are required for regular development. *FWA* alleles, although active in the endosperm, are silent in all other tissues. Silencing of *FWA* relies on DNA methylation of two direct repeats within its 5' coding region. In *fwa* epi-mutants the direct repeats are hypomethylated and *FWA* is expressed. As a consequence *fwa* epi-mutants display a delay in flowering time (Soppe et al, 2000), because FWA inhibits FLOWERING LOCUS T (FT), a protein that acts as a mobile floral signal (Ikeda et al, 2007). This partially explains the late flowering which is observed in the hypomethylated *ddm1* (Kakutani et al, 1995) and *met1* (Kankel et al, 2003) *Arabidopsis* mutants. However, treatment of *Arabidopsis* with the DNA methylation inhibitor azacytidine results in early flowering (Burn et al, 1993). It is therefore likely that multiple genes regulated by DNA methylation are required for correct flowering time in *Arabidopsis*.

The significance of DNA methylation within the body of protein coding genes is less well defined than TEs and gene promoters. Body methylation is predominantly located in exons and more likely to occur within genes that are longer than the average, which supports the hypotheses that body methylation may function in selecting splice regions and prevent aberrant transcription (Takuno & Gaut, 2011). Considering the repressive effects DNA methylation has on single copy genes and

TEs, a role of body methylation in silencing of cryptic promoters in central gene regions would also be an appealing speculation (Zilberman et al, 2007).

1.6. Thesis objective

The repressive functions of DNA methylation provide an opportunity to exploit this epigenetic mark to induce phenotypic variation in plants. This requires the development of a strategy that creates stable changes in DNA methylation that alter gene expression. In this study four strategies have been developed and tested for their capacity to induce DNA methylation changes. This included a chemical treatment to inhibit DNA methylation, which was applied directly to a commercial crop. Target loci were analysed in the treated lines by DNA methylation-sensitive detection techniques. Somatic and trans-generational stability of the chemically induced DNA methylation changes were analysed using the same detection techniques after treatment withdrawal and in subsequent generations, respectively. In parallel, genetic demethylation approaches were trialled in a model organism, including the inactivation of the epigenetic modifier MET1. DNA methylation and expression changes of target genes were predicted using an epigenome browser and analysed by methylation-sensitive and gene expression detection techniques. To analyse the stability of target gene expression changes in the epigenetic modifier mutant the wild-type alleles were restored via a genetic cross. The target genes identified in the mutant were analysed in subsequent generations with the wild-type alleles restored. The third strategy interfered with DNA methylation pathways via over-expression of MET1 and a catalytically inactive MET1. Targets that altered their DNA methylation or expression profiles upon MET1 inactivation were analysed in the over-expression lines. Additional target genes were identified by screening an epigenome browser for loci that accumulate DNA methylation in a glycosylase mutant, since one possible result of MET1 over-expression is an increase in DNA methylation. Finally, a DNA demethylation strategy was tested expressing the mammalian demethylase TET3. Target genes for the mammalian demethylase in plants were selected based on their homology with TET3 targets in mammals. To analyse the stability of TET3-induced changes DNA methylation patterns at target genes were analysed over multiple generations. The most successful strategies were extended into commercially viable crops with the support of an industrial partner.

2.0. Inducing epigenetic variation in tomato using the DNA methylation inhibitor zebularine

2.1. Introduction

DNA methylation is an epigenetic mark and involves the transfer of a methyl-group (-CH_3) to the carbon 5 of cytosines within DNA. It is established and maintained by DNA methyltransferases (DNMTs) and actively removed by DNA glycosylases (Gehring & Henikoff, 2007). A core feature of DNA methylation is transcriptional repression, either by direct obstruction of transcriptional proteins or by serving as a target for specific proteins, which signal chromatin condensation (Klose & Bird, 2006). Mammals, plants and fungi all have DNA methylation systems but the regulatory proteins, DNA methylation levels, locations and sequence types vary. In plants, DNA is methylated at three sequence types, CG, CHG and CHH (where H is A, T or G) (Law & Jacobsen, 2010). Interfering with such systems by either DNMT knockout (Kankel et al, 2003) or knockdown approaches (Kim et al, 2008) have shown that establishment and maintenance of DNA methylation are required for normal vigour, morphology and gene expression.

DNA methylation inhibitors including 5-Azacytidine (azacytidine), 5-aza-2′-deoxycytidine (decitabine) (Baylin, 2005) and 1-β-D-ribofuranosyl-1,2-dihydropyrimidine-2-one (zebularine) (Baubec et al, 2009) (Figure 2.1 (Ewald et al, 2008)) have been used to interfere with DNA methylation systems. These are cytosine analogues that undergo cellular uptake by nucleotide transporters and subsequent phosphorylation allows their incorporation at cytosine positions into replicating DNA. DNMTs are unable to methylate incorporated analogues due to their structural differences at the carbon 4 or 5 position (Galmarini et al, 2001). Indirect hypomethylation occurs because DNMTs have a reduced dissociation rate from these cytosine analogues compared to native cytosines, due to the formation of covalent adducts (Baubec et al, 2009; Santi et al, 1983). Their analogy with cytosines means that DNA methylation inhibitors can also be incorporated into RNA during transcription, which can lead to aberrant protein synthesis during translation (Baylin, 2005).

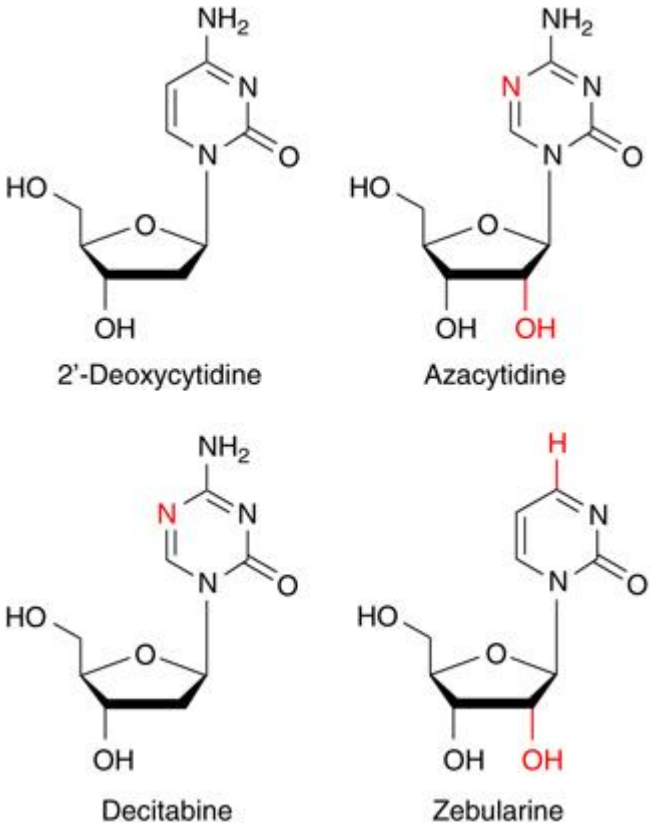


Figure 2.1. Chemical Structures of cytosine, azacytidine, decitabine and zebularine. The figure shows the chemical structures for cytosine (top left), azacytidine (top right), decitabine (bottom left) and zebularine (bottom right). Red symbols indicate structural deviations from cytosine. Taken from Ewald et al, (2008).

In mammals, DNA methylation inhibitors have shown both *in vivo* and *in vitro* to release silencing of some tumour suppressor genes, which occurs as a result of altered epigenetic modifier activity during malignancy. Therefore, azacytidine and decitabine are currently undergoing clinical trails in potential cancer therapies (Raj & Mufti, 2006), (Das & Singal, 2004).

In plants, DNA methylation inhibitors induce multiple phenotypes. Azacytidine transiently inhibits shoot induction in *Petunia* tissue culture (Prakash et al, 2003), causes segregating dwarfism in rice (Sano et al, 1990), increases total protein content in wheat seeds (Vanyushin et al, 1990) and reduces flower number in sugar beet (Iudanova et al, 2012). The direct mechanisms responsible for these phenotypes have not been identified and have been assigned to global hypomethylation and

transcriptional activation of silent genes. Azacytidine induces fruit ripening in the tomato *colourless non ripening* (*cnr*) epi-mutant (Manning et al, 2006; Martel et al, 2011), as demethylation at the *cnr* gene promoter allows the binding of the transcription factor RIN (Ripening INhibitor) and subsequent gene expression. Interestingly, the offspring from azacytidine treated tomato plants do not germinate (Zhong et al, 2013). Such observations highlight the potential for DNA methylation inhibitors to induce variation but also raise questions about direct heritable changes and indirect cytotoxic effects in plants.

If DNA methylation inhibitors can be used to induce trans-generational heritable changes, then they could become a fast, non-transgenic approach to induce variation. This hypothesis is investigated in this chapter by analysing DNA methylation changes and their heritability induced by zebularine in tomato.

2.2. Results

2.2.1. Zebularine induces transient growth inhibition of tomato

It was first necessary to determine a concentration at which zebularine caused detectable DNA methylation changes in tomato. This was achieved by growing tomato on MS30 medium with increasing concentrations of zebularine. At 80 μM, phenotypic changes were detected after 10 days that included inhibition of epicotyl and lateral root growth (Figure 2.2). To analyse if this observation was due to indirect cytotoxic effects, genomic DNA from treated and control plants was digested with the methylation-sensitive restriction isoschizomers *Msp*I and *Hpa*II. Both restriction enzymes cleave the sequence CCGG but *Hpa*II only cuts the sequence when it is unmethylated and *Msp*I cuts the sequence when it is unmethylated and when the internal C is methylated ($C^{m}CGG$) (Waalwijk & Flavell, 1978). An 80 μM zebularine treatment caused loss of DNA methylation in tomato, as some of the high-molecular-weight DNA was digested in the *Hpa*II lane of treated plants (white box, Figure 2.3) but not in the *Hpa*II lane of untreated control plants (black box, Figure 2.3). To favour seed development, which would allow the analysis of subsequent generations,

zebularine treatment was withdrawn. 7 days after the treatment was withdrawn epicotyl and lateral root growth reverted (Figure 2.2).

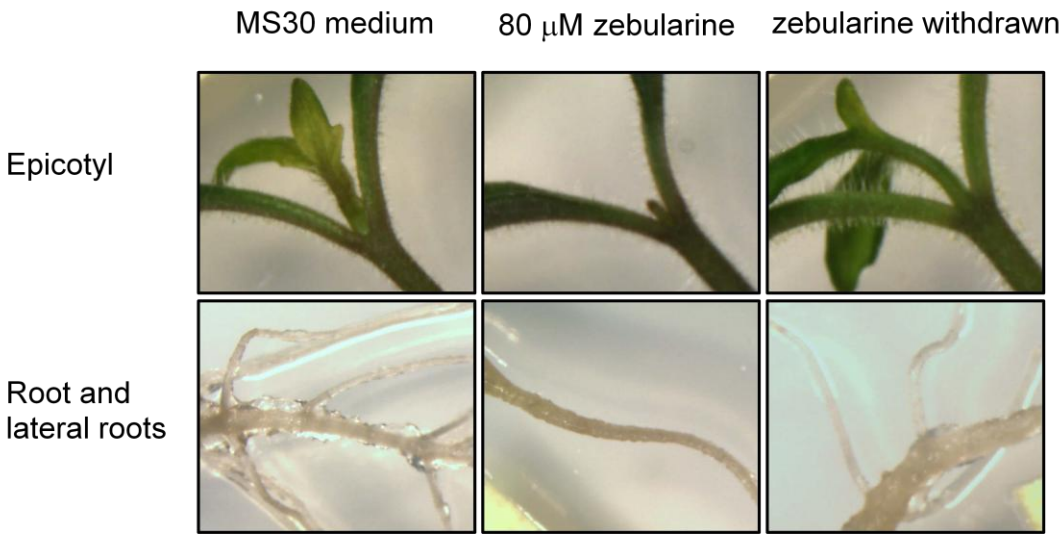


Figure 2.2. Zebularine treatment of tomato. Epicotyl (top) and root growth (bottom) are shown for tomato grown for 10 days on MS30 medium, MS30 medium with 80 μM zebularine and MS30 medium with 80 μM zebularine, which was then withdrawn by transplanting tomato onto MS30 medium for 7 days. The growth conditions used are as described in Section 8.2.3.3 of Materials and Methods.

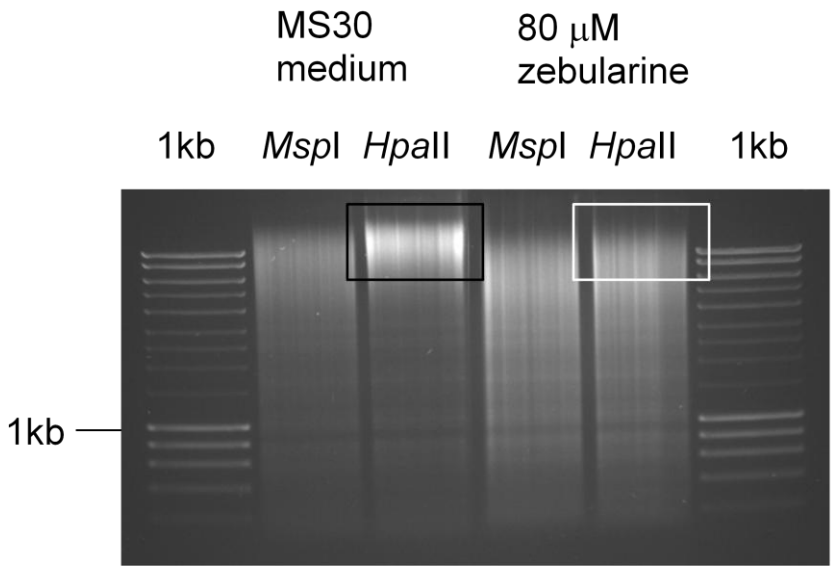


Figure 2.3. *Msp*I and *Hpa*II restriction enzyme digests of genomic DNA from zebularine treated tomato. The figure shows an ethidium bromide stained agarose gel of genomic DNA isolated from tomato grown on MS30 medium (left) and MS30 medium with 80 μM zebularine (right) digested with restriction isoschizomers *Msp*I and *Hpa*II according to Sections 8.2.1.1 and 8.2.1.13 of Materials and Methods. *Msp*I and *Hpa*II recognise and cleave the sequence CCGG. *Hpa*II will cleave the sequence when unmethylated and *Msp*I will cleave the sequence when unmethylated and when the internal C is methylated (Waalwijk & Flavell, 1978). White and black boxes are used to highlight differences in the digestion pattern between zebularine treated and non-treated samples, respectively. 1kb (Bioline) was used as a DNA marker that was loaded on the far left and right of the gel, and the position of the 1kb band is indicated on the left of the image.

2.2.2. Zebularine induces somatic DNA methylation changes in tomato

After confirming zebularine-induced hypomethylation in tomato it was necessary to determine if these changes were heritable. First, somatic heritability was analysed using methylation-sensitive amplified fragment length polymorphism (MS-AFLP) analysis. MS-AFLP is a standard AFLP analysis where adaptors are ligated to DNA digested with *Eco*RI (Vos et al, 1995) and either of the methylation-sensitive restriction isoschizomers *Msp*I or *Hpa*II. PCR amplification with adaptor specific primers yields a DNA fingerprint and differences between the fingerprint of controls and treated samples is evident of DNA methylation changes at both CG and CHG sequence types (Figure 2.4) (Paun et al, 2010; Portis et al, 2004).

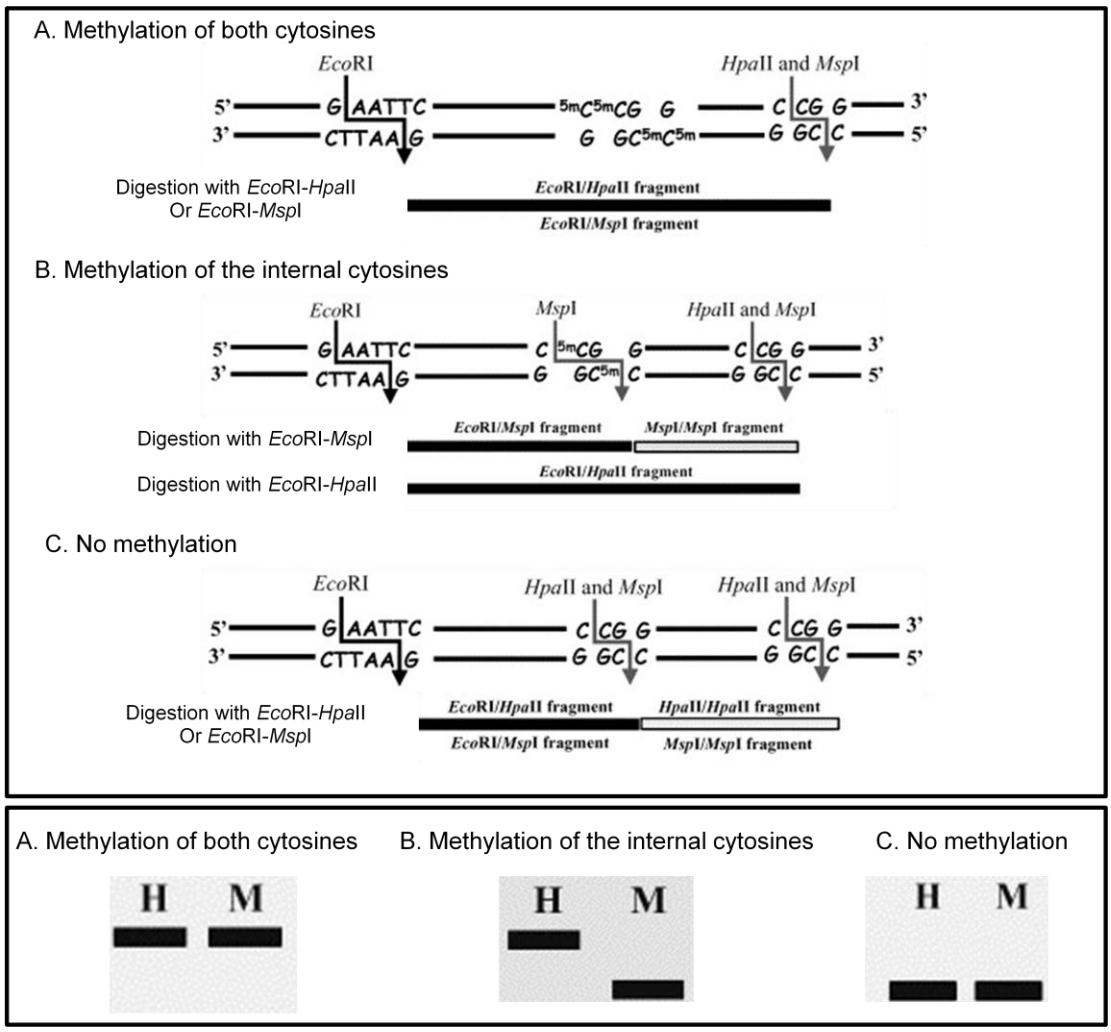


Figure 2.4. Possible outcomes of a MS-AFLP analysis. The figure shows a schematic diagram of all the possible outcomes of a MS-AFLP analysis using the restriction enzymes *Eco*RI, *Msp*I and *Hpa*II. In the top panel DNA is represented by the horizontal thin black line with 5' and 3' labelled ends. Cleavage by restriction enzymes is indicated is by staggered arrows and methylation is represented by "5m". Restriction enzyme cleavage products are provided below the DNA schematic. The bottom panel shows a schematic outcome of the restriction enzyme cleavage products in the top panel if they were analysed using a polyarcylamide gel stained with ethidium bromide after adaptor ligation and PCR amplification. **Top panel, A**. Methylation at both cytosines in a CCGG sequence prevents digestion by *Msp*I and *Hpa*II. As a consequence the next unmethylated CCGG site is cleaved. This creates a larger amplicon after adapter ligation and PCR amplification with adaptor specific primers (**bottom panel, A**). **Top panel, B**. Loss of methylation from the external cytosine in the sequence CCGG allows *Msp*I digestion but not *Hpa*II. *Eco*RI and *Msp*I digests result in smaller amplicons after adaptor ligation and PCR amplification, while *Eco*RI and *Hpa*II digests result in one larger amplicon (**bottom panel, B**). **Top panel, C**. Both *Msp*I and *Hpa*II cleave an unmethylated CCGG site, which results in smaller amplicons after adaptor ligation and PCR amplification (**bottom panel, C**). Modified from Portis et al, (2004).

For the MS-AFLP analysis genomic DNA was used in duplicate from three whole tomato seedlings grown for 17 days on MS30 medium, 17 days on MS30 medium with 80 μM zebularine and 10 days on MS30 medium with 80 μM zebularine, which was then withdrawn by transplanting tomato onto MS30 medium for 7 days. The 7 day withdrawal period allowed the formation of the epicotyl to analyse somatic changes and only indisputable changes were selected for subsquence analysis.

It is worth noticing the somatic change (Figure 2.5). An amplicon is visible when DNA from tomato grown on 80 μM zebularine is analysed by the *Hpa*II MS-AFLP. The signal presence and intensity are comparable when analysing DNA from tomato where treatment had been withdrawn. There is a weak detection of the target when analysing the DNA from tomato grown on MS30 medium using the *Msp*I MS-AFLP, but the signal strength increases with treatment and when treatment is withdrawn. The occurrence of bands with treatment in the *Hpa*II MS-AFLP and the increase in intensity of bands with treatment in the *Msp*I MS-AFLP, which are maintained after treatment is withdrawn, is indicative of a somatically heritable DNA methylation change at both CG and CHG sequence types.

Interestingly, very few regions responded to treatment. One locus, labelled unaffected (Figure 2.5), maintained internal methylation at a $C^{m}CGG$ site. This can be concluded when a signal is detected throughout the *Msp*I MS-AFLP analysis, which is insensitive to internal C methylation, but not in the *Hpa*II MS-AFLP analysis, which in sensitive to internal C methylation, and there is no change in signal presence or strength with treatment in either MS-AFLP analysis.

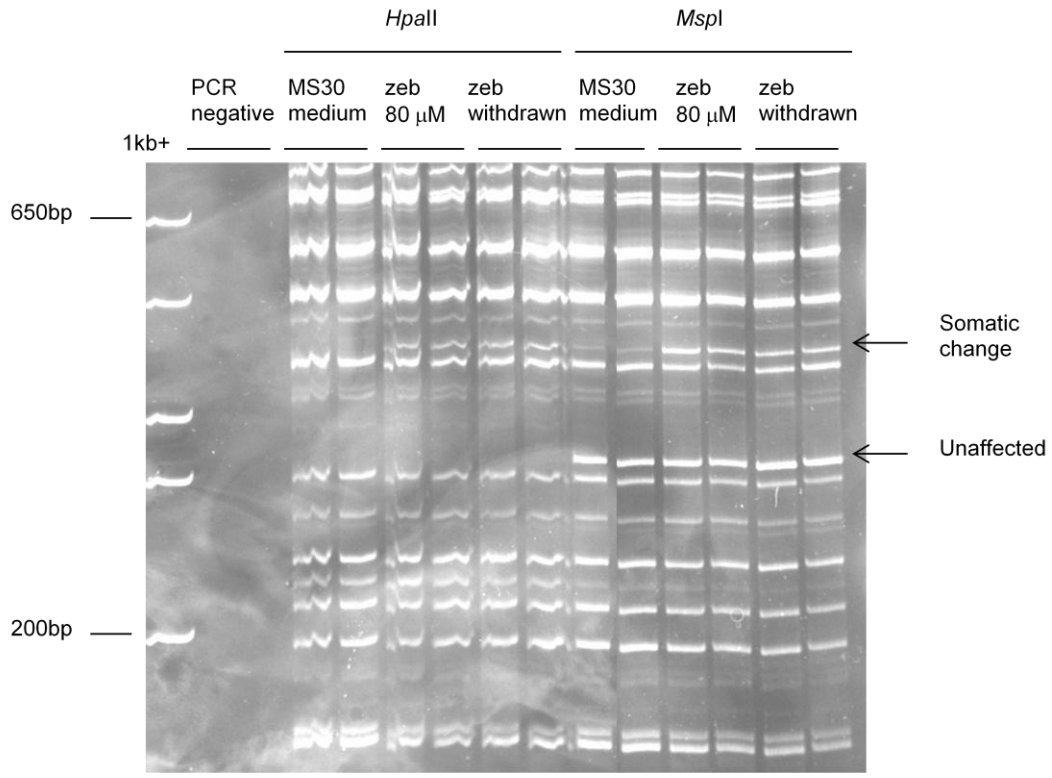


Figure 2.5. MS-AFLP screen to analyse somatic DNA methylation changes in zebularine treated tomato. The figure shows a MS-AFLP screen analysed using a polyacrylamide gel stained with ethidium bromide performed as described in Section 8.2.1.12 with primers from Section 8.1.4.5 of Materials and Methods. All samples were analysed in duplicate including, from left to right, a water only PCR negative control, DNA from tomato grown for 17 days on MS30 medium, 17 days on MS30 medium with 80 μM zebularine and 10 days on MS30 medium with 80 μM zebularine, which was then withdrawn by transplanting tomato onto MS30 medium for 7 days. The three DNA samples were analysed by a MS-AFLP analysis using *Hpa*II (left) and *Msp*I (right). *Msp*I and *Hpa*II methylation sensitivities are as described in Figure 2.4. 1kb+ (Invitrogen) was used as a DNA marker with some sizes provided on the left of the figure. A somatic change is highlighted on the right (top). Methylation has been lost from both cytosines in the sequence CCGG with zebularine treatment, enabling both *Hpa*II and *Msp*I to cut the sequence, resulting in PCR amplification after adaptor ligation (zeb 80μM). The signal changes are detected using DNA from tomato where treatment was withdrawn for 7 days (zeb withdrawn). An unaffected locus is highlighted on the right (bottom). There is no change in signal detection with zebularine treatment (zeb 80μM).

To test if the somatic DNA methylation changes caused by zebularine were genetically stable the analysis was extended into the next generation. DNA for both untreated and treated tomato was prepared again in parallel with the DNA from offspring of treated parental plants. The same regions responsive and unresponsive to treatment were identified (zeb 80μM, Figure 2.6). This indicates, for the highlighted loci, that the DNA methylation changes induced by zebularine that were identified in this study, are reproducible. The somatic change reverted in the offspring to that of untreated tomato grown on MS30 medium (mC reversion, offspring, Figure 2.6), suggesting that DNA methylation changes induced by zebularine at this locus are not stable across generations. The signal for the unaffected locus has been lost in the

offspring (unaffected*, offspring, Figure 2.6). This could be explained if the locus existed as an epi-allele in the parent plant, where only one allele is methylated at both cytosines and the other allele is methylated at the internal cytosine of the sequence CCGG. The offspring from this parent used in this analysis could inherit two fully methylated alleles preventing *Msp*I digestion.

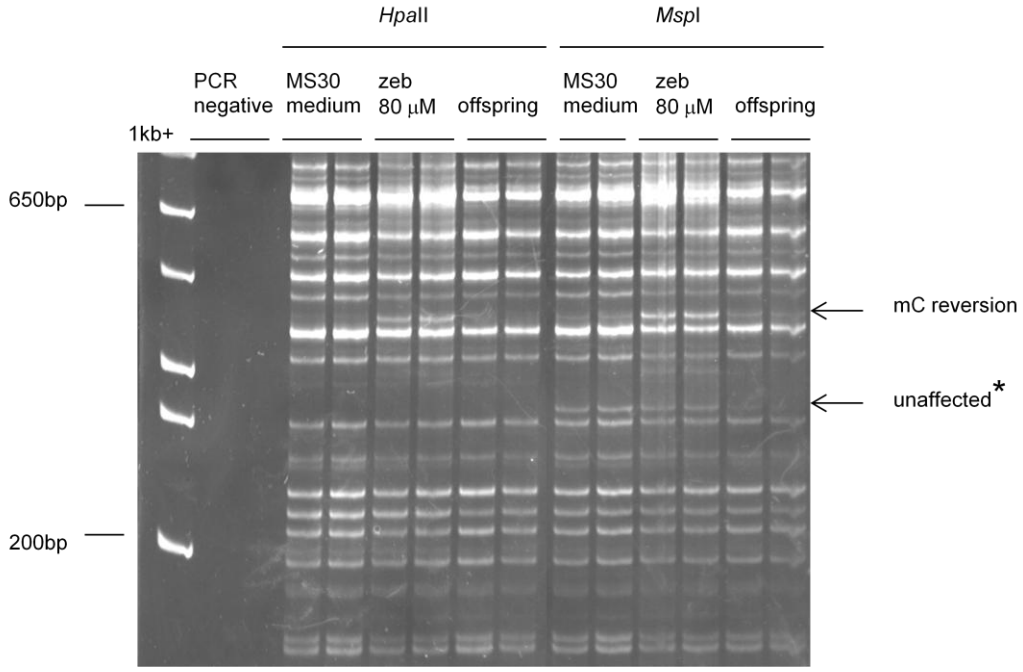
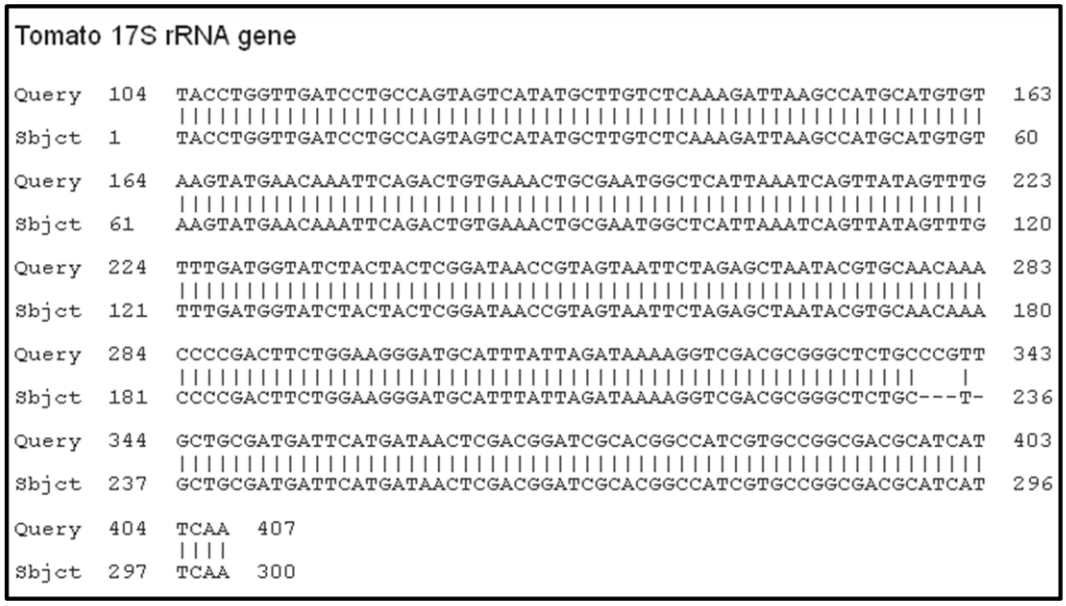


Figure 2.6. MS-AFLP screen to analyse trans-generational DNA methylation changes in zebularine treated tomato. The figure shows a MS-AFLP screen performed as described in Figure 2.5, however, DNA isolated from offspring of zebularine treated parent plants was analysed to determine the trans-generational stability of induced DNA methylation changes. The DNA methylation change that reverts in the offspring of treated parental plants is indicated on the right hand side of the image using the label "mC reversion". The locus that is unaffected by zebularine treatment and escapes detection in the offspring is indicated on the right hand side of the image using the label "unaffected *".

2.2.3. Zebularine causes DNA methylation changes at *rDNA*

In order to elucidate the target specificity of zebularine the DNA sequences of both affected and unaffected loci were isolated, sequenced and the sequencing results were aligned with the tomato genome. To ensure the correct amplicons had been isolated they were analysed using a polyacrylamide gel after isolation in parallel with the MS-AFLP analysis. The region that twice lost DNA methylation with treatment was located within *rDNA* and the region which was twice unaffected by treatment was located within the jinling2 retroelement (Figure 2.7, Appendix 10.1).

A.



B.

Figure 2.7. Identifying the target specificity of zebularine. The figure shows sequencing reads from targets identified using the MS-AFLP analysis (Figures 2.5 and 2.6) aligned against the tomato genome. Alignments were carried out using the NCBI basic nucleotide blast search tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The Query represents the DNA sequence of the target and the Subject (Sbjct) represents a region of the tomato genome with the highest homology to the Query. Nucleotide positions of the Query and Sbjct are provided on the left and right side of the DNA sequence. **A**. The somatic change is within *rDNA*. **B**. The unaffected locus is within the jinling2 retroelement.

2.3. Discussion

In this chapter, the use of zebularine as a tool to introduce trans-generational DNA methylation changes in tomato has been assessed.

It was found that, at 80 μM, zebularine causes transient inhibition of epicotyl and lateral root growth in tomato. This suggests that the effects of zebularine, which are responsible for these phenotypes are reversible. Recently, it has been shown that 20, 40 and 80 μM zebularine transiently inhibits growth of *Arabidopsis* (Baubec et al, 2009) and 10 μM inhibits root elongation in wheat. This was attributed to a reduced mitotic index found in the root tips of treated wheat when compared to control plants (Cho et al, 2011). Analysing the mitotic index in the affected tomato organs would not be possible due to their complete inhibition with treatment. The tomato *DIAGEOTROPICA (DGT)* gene encodes a cyclophilin protein and when mutated prevents the growth of lateral roots. It is believed to have a role in the auxin signalling pathway (Oh et al, 2006). There are multiple auxin-related *Arabidopsis* mutants that prevent the initiation and maturation of lateral roots. They include *alf1*, *alf3*, *alf4*, (Celenza et al, 1995) *aux1*, *axr1*, *axr4* (Hobbie & Estelle, 1995) and *sur* (Boerjan et al, 1995). In the future, analysing expression and DNA methylation changes at both the endogenous *DGT* gene and *Arabidopsis* auxin-related homologues could provide some insight into the actions of zebularine responsible for the lateral root phenotype in tomato.

A viable seed set was obtained from tomato treated with zebularine that enabled the analysis to be extended into the next generation. Previous studies show that treating tomato fruit with azacytidine did not yield viable seeds, preventing trans-generational analysis (Zhong et al, 2013). Treating tomato fruit may be more damaging to seed production than the approach used in this study or 5-aza may have higher cytotoxic or demethylating capacities than zebularine in tomato. To analyse this, the approach used in this studied could be carried out with 5-aza instead of zebularine.

In two independent repetitions it was shown that zebularine caused DNA methylation changes within *rDNA* but not within the jinling2 retroelement in tomato. This

indicates that cytosines at different loci differ in their sensitivity to zebularine and that *rDNA* appears to be particularly susceptible. In *Arabidopsis*, zebularine causes global hypomethylation resulting in the activation of silent single copy genes such as *FWA* and transposable elements including *CACTA* and *MULE* (Baubec et al, 2009). DNA methylation changes at single copy genes and transposable elements were not found in this study. The possibility of such an event cannot be eliminated because the MS-AFLP analysis relies on the use of selective primers, the requirement for the presence of a restriction enzyme site and multi-fragment amplification of templates. To identify all targets large scale bisulfite sequencing would be required.

The detected DNA methylation changes introduced by zebularine were somatically heritable after treatment was withdrawn. When treatment is withdrawn from *Arabidopsis* DNA methylation changes are restored to wild-type levels within 8 weeks (Baubec et al, 2009). Therefore, the somatic changes in tomato may require a longer period for restoration, which could be analysed if the experiment was repeated by sampling tissue long after the withdrawal from zebularine. Alternatively, DNA methylation changes in tomato may be more stable than in *Arabidopsis*. The reversion of somatic changes in tomato was found in the offspring of treated parental plants, indicating the potential for a resetting mechanism in the germline or during early embryogenesis. Evidence for epigenetic reprogramming during early embryogenesis in plants is also observed in maize studies, where the *MATERNALLY EXPRESSED IN EMBRYO 1 (MEE1)* maternal allele is demethylated on fertilisation and methylated later in embryogenesis (Jahnke & Scholten, 2009).

It is important to note that the method used to screen for DNA methylation changes only screens a limited number of target regions. The *Msp*I and *Hpa*II MS-AFLP analysis can detect zebularine-induced changes at both CG and CHG sequence types but is unable to detect CHH methylation changes. The MS-AFLP approach was chosen not to identify all changes but to detect some changes that could be analysed for their transmission stability. To detect all DNA methylation changes would require large scale bisulfite sequencing.

The evidence presented in this chapter supports the use of inhibitors as a tool to study DNA methylation in plants. No evidence was found that indicates DNA methylation

changes induced by zebularine are transmitted across generations but the approach does not allow the elimination of this theory. Alternative approaches with better prospects to generate and identify trans-generational epigenetic changes are required, which are discussed in subsequent chapters.

3.0. Inducing epigenetic variation in *Arabidopsis* by inactivating DNA METHYLTRANSFERASE 1

3.1. Introduction

Chemical interference with DNA methylation using the DNA methylation inhibitor zebularine was analysed in Chapter 2. Using this strategy it was not possible to detect trans-generational DNA methylation changes. An alternative strategy to induce detectable heritable DNA methylation changes is to exploit *Arabidopsis* DNA methyltransferase mutants.

DNA methyltransferases are required to catalyse methyl-group transfer from S-adenosyl methionine to the carbon 5 of cytosine within DNA. In plants, this occurs at three sequence types including CG, CHG and CHH (where H is A, T or G). In *Arabidopsis*, DNA methyltransferases function in two DNA methylation systems. One system is *de novo* methylation, which is required for the establishment of DNA methylation at all three sequence types. *De novo* methylation is achieved by an RNA directed DNA methylation (RdDM) pathway. In this pathway 24-nucleotide (24-nt) small interfering RNAs (siRNAs) align with homologous regions of the genome to recruit DOMAINS RE-ARRANGED METHYLTRANSFERASE 2 (DRM2) (Law & Jacobsen, 2010). *De novo* methylation at CHH sites can also occur independently of RdDM and requires CHROMOMETHYLTRANSFERASE 2 (CMT2) (Zemach et al, 2013). It has been hypothesised that targeting of CMT2 occurs through its ability to bind di-methylation at lysine 9 on histone 3 (H3K9me2) via its chromo-domain (Pikaard, 2013). The second system in *Arabidopsis* is maintenance methylation, which is required to preserve DNA methylation after every cellular DNA replication cycle (Goll & Bestor, 2005). METHYLTRANSFERASE 1 (MET1) recognises hemi-methylated DNA in the daughter cells to maintain methylation at CG sites (Goll & Bestor, 2005). CHROMOMETHYLTRANSFERASE 3 (CMT3) recognises H3K9 methylation to maintain DNA methylation at CHG sequences (Lindroth et al, 2004). Methylation at CHH sites is maintained by constant *de novo* activity of DRM2 (Law & Jacobsen, 2010).

Despite the characterisation of DNA methylation targeting and maintenance systems in *Arabidopsis*, it is still unclear why there is a large variation in DNA methylation profiles among wild-type accessions. This variation was apparent when the methylation status of 18 genomic loci was analysed among 96 *Arabidopsis* accessions and only one locus, *FWA* remained uniformly methylated (Vaughn et al, 2007). It has been suggested that such epigenetic variation could arise from genetic polymorphisms of cytosine residues (Vaughn et al, 2007). This does not, however, explain differences in DNA methylation among accessions at genetically identical sequences. It is unclear if epigenetic variation is controlled in *cis* or if *trans*-acting factors exists among *Arabidopsis* accessions, which contribute to the variation in epialleles.

To analyse if epigenetic modifier mutants can be used to introduce detectable heritable DNA methylation changes, the *Arabidopsis met1* mutant *met1-1* (Kankel et al, 2003) was analysed after restoring *MET1* wild-type alleles. The *met1-1* mutant contains a C to T missense mutation, which replaces a proline with a serine within the MET1 catalytic domain at position 1300 (Kankel et al, 2003). *met1-1* is not a null mutant, unlike the *Arabidopsis met1-3* mutant that contains a Transfer-DNA (T-DNA) insertion, which disrupts the catalytic domain (Saze et al, 2003). The *Arabidopsis met1-1* mutant was selected because MET1 is responsible for maintaining methylation at CG di-nucleotides (Finnegan et al, 1996), and null activity of *met1-3* would likely prevent the analysis of DNA methylation quantity on gene expression. The analysis identified a novel epigenetically regulated target gene coding for a *ncRNA*. The characterisation of this gene highlighted the natural epigenetic variation between *Arabidopsis* accessions. Therefore, the *ncRNA locus* became an ideal target to analyse if *trans*-acting factors contributed to epigenetic variation between *Arabidopsis* accessions.

3.2. Results

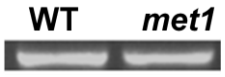
3.2.1. Analysing novel epigenetically regulated target genes in an *Arabidopsis met1* mutant

To analyse if heritable DNA methylation changes were induced by inactivating *MET1*, it was first necessary to identify novel epigenetically regulated target genes. To identify such targets the epigenome browser (http://neomorph.salk.edu/epigenome/epigenome.html) was screened for loci that atypically change their expression or DNA methylation patterns in the *met1* mutant. Atypical changes, for example, would include down-regulated or up-regulated genes without direct DNA methylation changes or reduced DNA methylation at non-CG sequence types. Of ten targets identified, expression changes were confirmed for five by semi-quantitative RT-PCR (sqRT-PCR), including *AT4G15242, AT5G15360, AT4G10850, AT1G19070* and *AT2G41380* (Figure 3.1, Appendix 10.2). Two MET1-regulated targets, previously characterised in the literature, *AT4G25530 (FWA)* (Saze et al, 2003) and *AT2G12210 (CACTA)* (Kato et al, 2003) were included as controls. Changes in expression of two novel target genes, *AT4G15242* and *AT5G15360* correlated with a direct loss in DNA methylation (Figure 3.1A). Interestingly, changes in gene expression also occurred without a direct change in DNA methylation and these changes included both an increase in expression of *AT4G10850* and a decrease in expression of *AT1G19070* and *AT2G41380* (Figures 3.1 B and C). It should be noticed that down-regulation of *AT1G19070* contridicts the epigenome browser result that indicates up-regulation of *AT1G19070* in *met1-3* (Appendix 10.2).

Changes in target gene expression without direct changes in DNA methylation may be a result of indirect effects of *met1* inactivation. Indirect changes in target gene expression could for example occur by transcriptional interference if local over-lapping or antisense genes were activated in the *met1* mutant (Shearwin et al, 2005). Local over-lapping and antisense genes were identified for each target gene using the epigenome browser (http://neomorph.salk.edu/epigenome/epigenome.html), and were subsequently analysed for expression changes by sqRT-PCR in the *met1* mutant. No changes in local over-lapping or antisense gene expression could be detected for any target with an indirect change.

Surprisingly, for some targets the epigenome browser (http://neomorph.salk.edu/epigenome/epigenome.html) indicated a direct loss in DNA methylation at all three sequence types, despite MET1 being the classical CG maintenance methyltransferase (Figure 3.1A). These changes were confirmed by analysing the methylation patterns within the promoter and transcribed region of *AT4G15242*, which codes for a *ncRNA* with an unknown function (http://www.arabidopsis.org/) (Figure 3.2 A and B). Methylation patterns were analysed using bisulfite sequencing. In this technique DNA is treated with sodium bisulfite, which deaminates only unmethylated cytosines to uracil. Uracil base pairs with thymine when a target locus is amplified by PCR from bisulfite treated DNA. Individual clones are sequenced to determine a single-nucleotide methylation profile.

A.

Gene ID	Expression	mC types present in WT	mC types lost in *met1*
AT1G07940 (EF)		None	None
AT4G25530 (FWA)		CG, CHG, CHH	CG, CHG, CHH
AT2G12210 (CACTA)		CG, CHG, CHH	CG
AT4G15242 (ncRNA locus)		CG, CHG, CHH	CG, CHG, CHH
AT5G15360 (unkown)		CG, CHH	CG, CHH

B.

Gene ID	Expression
AT4G10850 (Nodulin MtN3 protein)	

C.

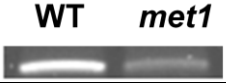
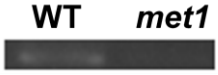
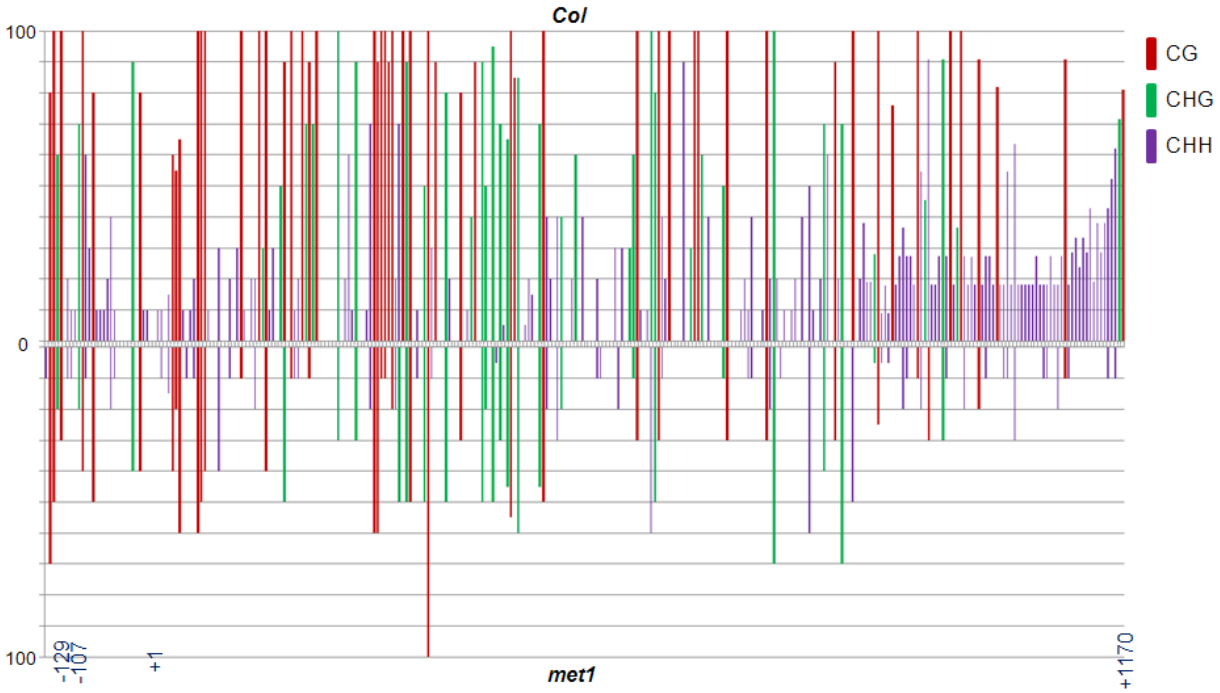
Gene ID	Expression
AT1G19070 (F-Box protein)	
AT2G41380 (Protein methyltransferase)	

Figure 3.1. Novel target gene expression changes in an *Arabidopsis met1* mutant. The figure shows target gene expression and DNA methylation changes in a *met1* mutant. Target gene ID and name is provided in the left hand column of each table. The column second from the left shows a sqRT-PCR analysis using RNA from 2 week old seedlings and an ethidium bromide stained agarose gel for the respective gene in wild-type (WT) and *met1*. *AT1G07940* (*ELONGATION FACTOR 1a, EF*), which contains no methylation in wild-type or *met1* was used as an internal control for expression levels. *FWA* and *CACTA* were used as controls to mark *met1* induced gene expression changes. The sequences of the primers used for the sqRT-PCR analysis are provided in Section 8.1.4.2 of the Materials and Methods. If DNA methylation is present within WT at the target locus then the methylation types are provided in the third column from the left and colour coded. If DNA methylation is lost in *met1* from the target locus then the methylation types lost are provided in the final column on the right. **A**. Target genes with an increase in expression and direct changes in DNA methylation in the *Arabidopsis met1* mutant. **B**. Target genes with an increase in expression and no direct DNA methylation changes in the *Arabidopsis met1* mutant. **C**. Target genes with a reduction in expression and no direct DNA methylation changes in the *Arabidopsis met1* mutant.

A.



B.

AT4G15242 DNA methylation levels			
	Col	***met1***	**% reduction**
% mC	35.8	10.4	70.9 %
% mCG	93.1	26.1	71.9 %
% mCHG	66.0	26.4	60.0 %
% mCHH	15.5	3.4	78.1 %

Figure 3.2. DNA methylation levels at the *ncRNA locus* in the *Arabidopsis Columbia* wild-type and *met1* mutant. **A**. The figure shows the percentage of methylation at individual cytosines determine by bisulfite sequencing ten clones for five overlapping regions of the *ncRNA* promoter (up to nucleotide position -129) and transcribed region (+1 to +1170) in *Arabidopsis Columbia* wild-type (top) and *met1* mutant (bottom). A percentage scale is on the Y-axis and nucleotide positions are provided below the profile. The sequences of the primers used for PCR amplification from bisulfite converted DNA isolated from 2 week old seedlings, and the softwares for compiling bisulfite sequence reads are provided in Sections 8.1.4.3 and 8.2.1.11 of Materials and Methods, respectively. Due to the low levels of CHH methylation deamination of cytosines in a CHH sequence context was used as an indicator for sodium bisulfite conversion and DNA methylation types are colour coded according to the key in the top right. **B**. The figure shows total DNA methylation levels for each methylation type within the *ncRNA locus* in *Arabidopsis Columbia* wild-type (*Col*) and *met1*. Total methylation (%mC) and the methylation sequence types are indicated in the left hand column. The columns second and third from the left provide the total methylation levels for each sequence type in *Arabidopsis Columbia* wild-type (*Col*) and *met1*, respectively. The final column on the right shows the percentage of DNA methylation reduction at each sequence type caused by the *met1* mutation.

3.2.2. Analysing trans-generational epigenetic changes induced by inactivating MET1

The induction of heritable epigenetic changes could be a source of phenotypic variation. To test if the epigenetic changes induced by inactivating MET1 were heritable, it was necessary to restore the wild-type alleles in the *met1* mutant. The wild-type alleles were restored with the help of Dr Elena Zubko and Dr Andrea Kunova, P. Meyer lab, at the University of Leeds. The *met1* mutant was crossed with a wild-type plant to produce a heterozygous line. This heterozygous line was self fertilised to produce a segregating seed population. From this segregating seed population lines with the *MET1* wild-type alleles were selected and self fertilised (Figure 3.3). Tracking the alleles was possible because the *met1* mutant allele has a deletion within a *Hae*III restriction enzyme site.

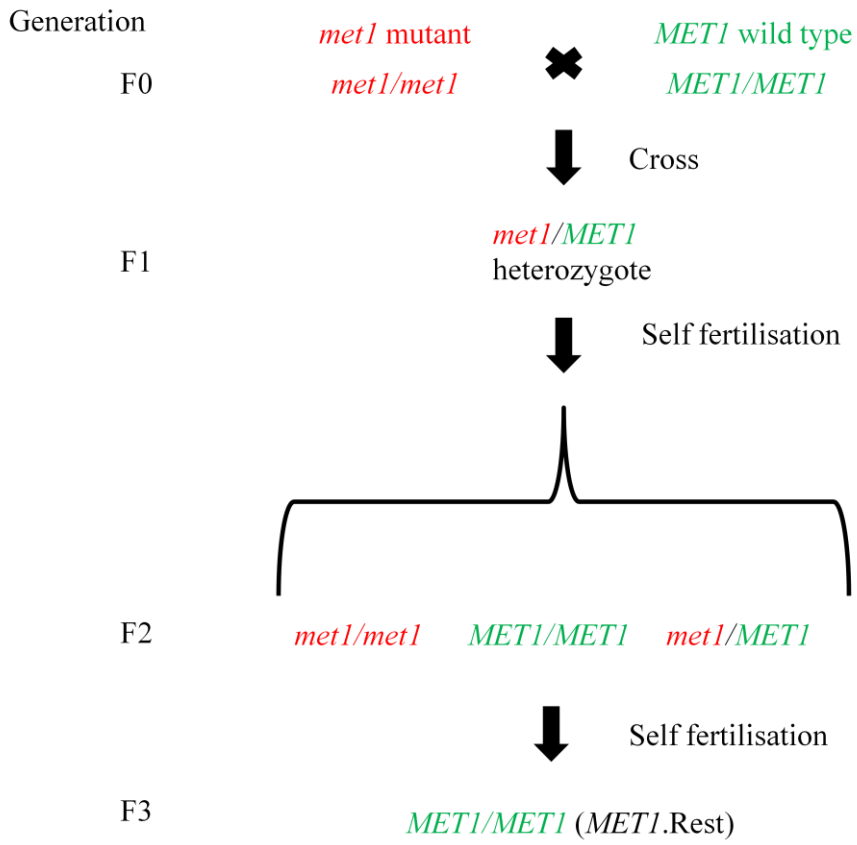


Figure 3.3. A diagram to illustrate how the *MET1* wild-type alleles were restored in the *met1* mutant. The figure shows a schematic diagram of crosses between wild-type and *met1* and subsequent self pollinations of the resulting hybrids. Crosses and genotyping were carried out as described in Section 8.1.1 of Materials and Methods. Red and green labels indicate the *met1* mutant and *MET1* wild-type alleles, respectively. The generation is provided on the left of the figure. Beginning from the top of the figure, an *Arabidopsis* F0 *met1* mutant and a wild-type line were crossed to produce an F1 heterozygote. The F1 heterozygote was self fertilised to generate a segregating F2 population. *Arabidopsis* lines with restored wild-type alleles (*MET1*.Rest) were selected and self fertilised to produce an F3 seed population.

After restoring the *MET1* wild-type alleles in the *met1* mutant, the MET1-regulated targets with direct changes in DNA methylation and expression (Figure 3.1A) were analysed for their stability across generations. This was achieved by analysing target gene expression in two F1 *met1/MET1* heterozygote lines and two F3 *MET1/MET1* restored lines. As a control for restored *MET1* activity *AT4G25530* was analysed, which encodes the homeo-domain containing transcription factor FWA that controls flowering (Koornneef et al, 1991). Two F1 *met1/MET1* heterozygous lines were analysed for *FWA* expression and in both *FWA* expression are reset and no longer detected (Figure 3.4A). In contrast, the *ncRNA locus* was stable in its expression up to the F3 generation with the *MET1* wild-type alleles restored (Figure 3.4B). This data suggests that some *met1* induced expression changes are stably inherited over

multiple generations, whereas others are reset upon restoration of one wild-type allele. Although not applicable to the example shown in Figure 3.4, it is important to acknowledge that in the F2 generation and subsequent generations with the *MET1* wild-type alleles restored, there is no control over the origin of the target alleles. If expression of the target locus was not detected in the F2 generation or subsequent generations, it would not be possible to determine if this was due to the inheritance of the wild-type alleles or epigenetic resetting using an expression analysis.

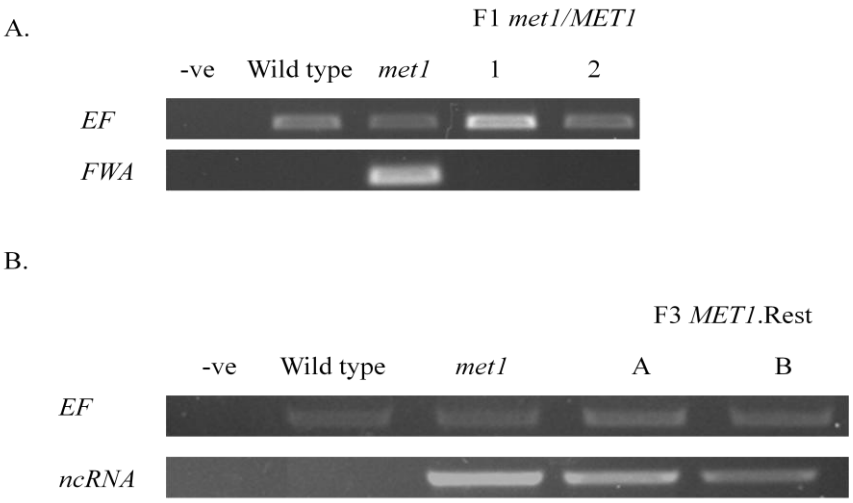


Figure 3.4. Analysing trans-generational stability of *met1* induced expression changes. The figure shows a sqRT-PCR analysis performed as described in Figure 3.1. Water only was used as a negative PCR control (-ve) and *ELONGATION FACTOR* (*EF*) gene expression was used as an internal control to compare target gene expression levels. **A**. *FWA* expression in wild-type, *met1* and two F1 *met1/MET1* heterozygous lines (F1 *met1/MET1* 1 and 2) where one *MET1* wild-type allele has been restored. **B**. *ncRNA locus* expression in wild-type, *met1* and two F3 *MET1/MET1* lines (F3 *MET1*.Rest A and B) where the *MET1* wild-type alleles have been restored.

3.2.3. Characterising epigenetic regulation of the *ncRNA locus*

The reduction in DNA methylation at CHG and CHH sequence types in addition to CG in the *met1* mutant questioned the involvement of the CHG- and CHH-specific methyltransferases, CMT3 and DRM2 in the regulation of the *ncRNA locus*. To analyse if DRM2 or CMT3 influenced the expression of the *ncRNA locus*, its expression was analysed in the *drm2* single and *drm1,drm2,cmt3* (*ddc*) triple mutants.

The expression of the *ncRNA locus* was not detected in the *drm2* single mutant (Figure 3.5A). While this suggests that inactivating DRM2 is not sufficient to activate the *ncRNA locus*, it does not eliminate the possibility that DNA methylation and silencing are still maintained by MET1 after DRM2 loss. This has been reported for some targets including *FWA*, where DRM2 is required for *de novo* but not maintenance of gene silencing (Cao & Jacobsen, 2002b).

The *ddc* triple mutant is a hybrid of *Arabidopsis Landsberg erecta (Ler)* and *Wassilewskija (Ws)* backgrounds. Therefore, *Arabidopsis Ler* and *Ws* wild-type accessions were analysed for the expression of the *ncRNA locus* in parallel with the *ddc* mutant. The expression of the *ncRNA locus* was not detected in *Arabidopsis Ler*. Interestingly, the *ncRNA locus* was active in *Arabidopsis Wassilewskija* wild-type (Figure 3.5B). Therefore, no conclusions could be drawn with regards to the regulation of the *ncRNA locus* by CMT3 specific CHG methylation because the *ddc* mutant included a *Ws* background.

DEFICIENT IN DNA METHYLATION 1 (DDM1) is a member of the SW12/SNF2 family of proteins and is required for MET1 activity (Brzeski & Jerzmanowski, 2003). To analyse if the expression status of the *ncRNA locus* required DDM1, its expression was analysed in a *ddm1* mutant. The *ncRNA locus* was active in the *ddm1* mutant (Figure 3.5C).

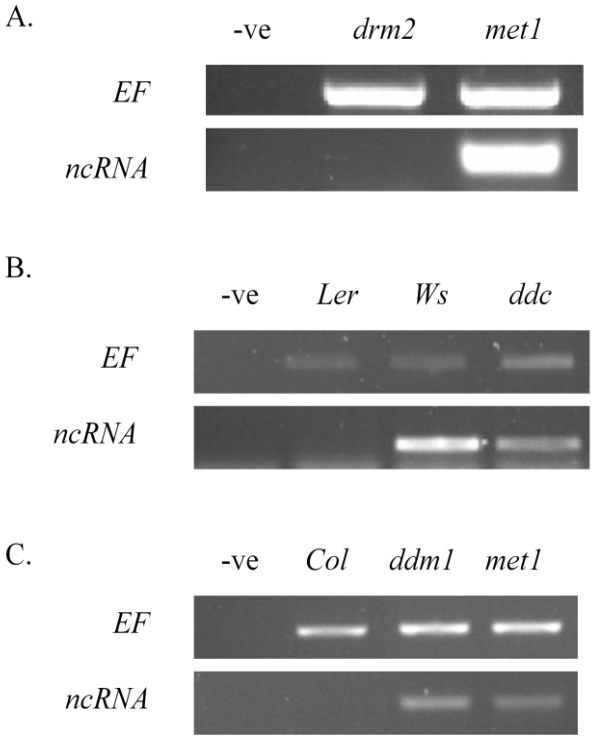


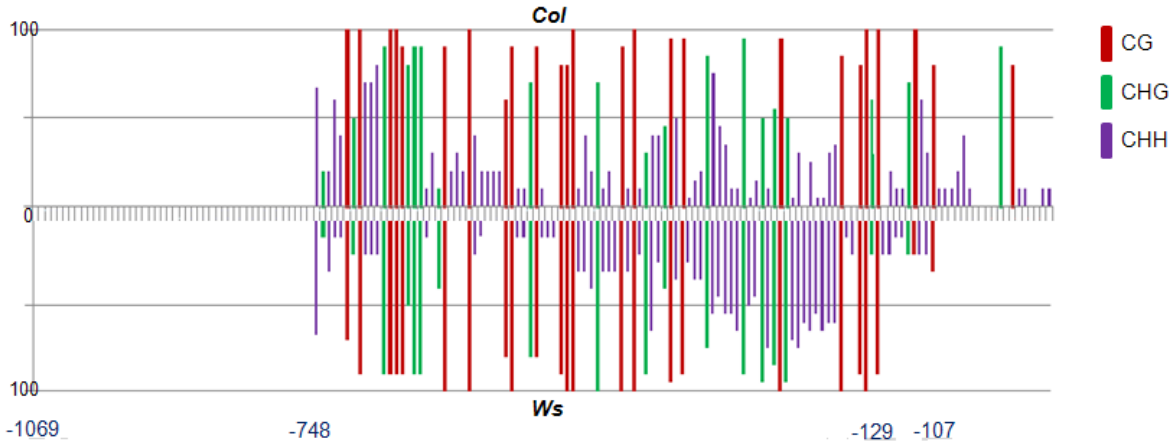
Figure 3.5. Characterising epigenetic regulation of the *ncRNA locus*. The figure shows a sqRT-PCR analysis performed as described in Figure 3.1. Labelling on the top indicates the respective wild-type or mutant. Labelling on the left indicates the respective gene analysed for expression. Water only was used as a negative PCR control (-ve) and *ELONGATION FACTOR* (*EF*) gene expression was used as an internal control to compare target gene expression levels. **A**. Expression of the *ncRNA locus* in the *de novo* methyltransferase mutant *drm2* and *met1*. **B**. Expression of the *ncRNA locus* in *Arabidopsis Landsberg Erecta (Ler)*, *Arabidopsis Wassilewskija (Ws)* and the *drm1/drm2/cmt3* (*ddc*) triple mutant. **C**. Expression of the *ncRNA locus* in *Arabidopsis Columbia (Col), ddm1,* and *met1* mutants.

3.2.4. Analysing DNA methylation and expression levels of the *ncRNA locus* in *Arabidopsis Columbia* and *Wassilewskija* wild-type accessions

The difference in expression of the *ncRNA locus* between *Arabidopsis Columbia* and *Wassilewskija* highlighted the potential for natural epigenetic variation between the accessions. This offered an opportunity to test the link between expression and DNA methylation, and the involvement of *cis* and *trans*-acting factors in DNA methylation variants. To analyse if the expression differences of the *ncRNA locus* between the two accessions were due to different DNA methylation profiles, the promoter and transcribed region of the *ncRNA locus* were analysed in the two accessions by bisulfite sequencing (Figure 3.6). This involved sequencing ten clones for seven overlapping regions from each accession. In *Arabidopsis Columbia (Col),* where expression of the *ncRNA locus* is not detected, DNA methylation begins at the 5'

region of the promoter (top, Figure 3.6A) and persists throughout the transcribed region (top, Figure 3.6B). However, in *Arabidopsis Wassilewskija (Ws),* DNA methylation was found at the 5' promoter region, but it ceased at position -107 (bottom, Figure 3.6 A and B). This indicates that both epialleles show a similar methylation profile in the upstream promoter region between positions -748 to -129 (Figure 3.6), while in *Ws* methylation is almost completely eliminated -107nt upstream of the transcription start site and within the transcribed region (Figure 3.6).

A.



B.

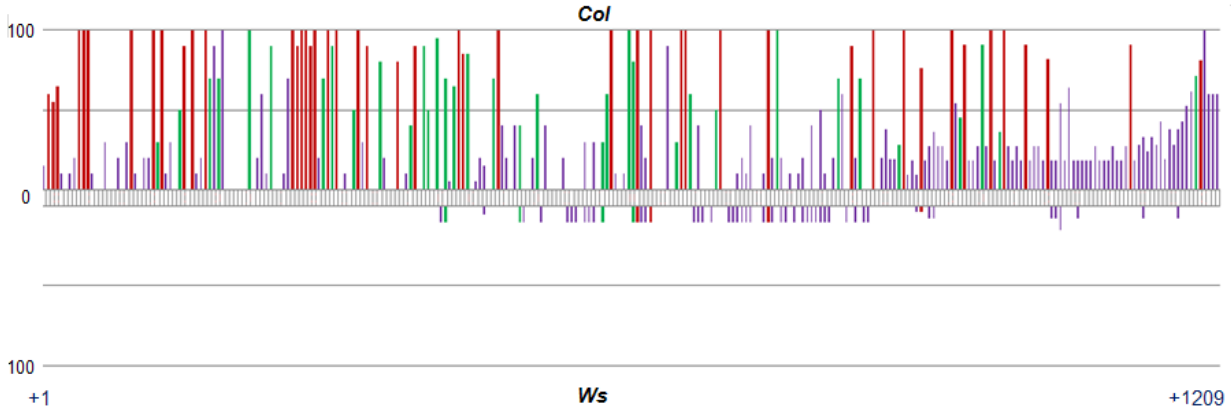


Figure 3.6. The DNA methylation profile of the *ncRNA locus* in *Arabidopsis Wassilewskija* and *Columbia*. The figure shows the percentage of methylation at individual cytosines within the promoter (**A**) and transcribed region (**B**) of the *ncRNA locus* in *Arabidopsis Columbia (Col)* (top) and *Wassilewskija (Ws)* (bottom). DNA methylation was analysed by bisulfite sequencing as described in Figure 3.2. A percentage scale is on the Y-axis and nucleotide positions are provided below each profile. DNA methylation types are colour coded according to the key in the top right.

3.2.5. Quantifying expression levels of the *ncRNA locus* in *Arabidopsis Columbia*, *Wassilewskija* and in the *met1* mutant

The differences in DNA methylation and expression levels among *Arabidopsis Col*, *met1* and *Ws* highlighted a possible link between the level of DNA methylation and the level of expression of the *ncRNA locus*. To analyse this, the expression levels of the *ncRNA locus* were determine by quantitative Real Time-PCR (qRT-PCR). Interestingly, the low, moderate and high methylation levels found in *Arabidopsis Ws*, *met1* and *Col* (Figure s 3.2a and 3.6) correlated with highest, moderate and absent expression levels of the *ncRNA locus*, respectively (Figure 3.7).

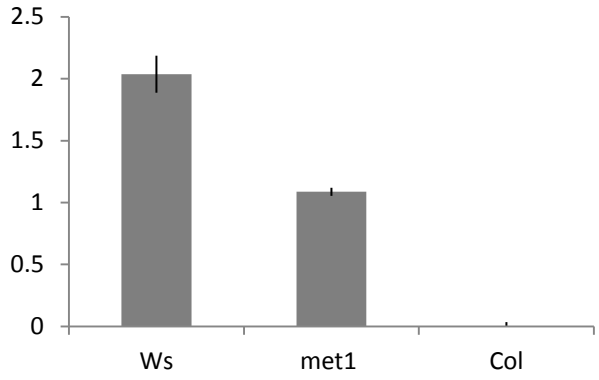


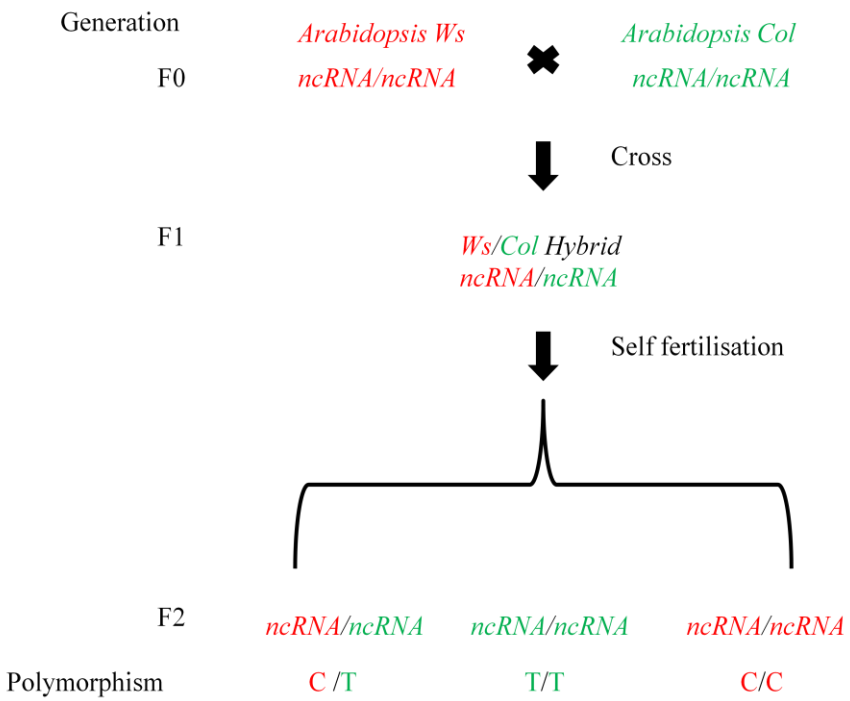
Figure 3.7. Quantitative expression analysis of the *ncRNA locus* in *Arabidopsis Wassilewskija, met1* and *Columbia.* The figure shows a qRT-PCR analysis performed as described in Section 8.2.2.3 of Materials and Methods using RNA from 2 week old seedlings and primers described in Section 8.1.4.2. The X-axis indicates the respective wild types and mutant, including *Arabidopsis Wassilewskija (Ws), met1* and *Columbia (Col)*. The Y-axis indicates fold expression levels of the *ncRNA locus* normalised to *ELONGATION FACTOR* gene expression.

3.2.6. Exploiting the accession specific epigenetic variation of the *ncRNA locus* to analyse epiallele stability

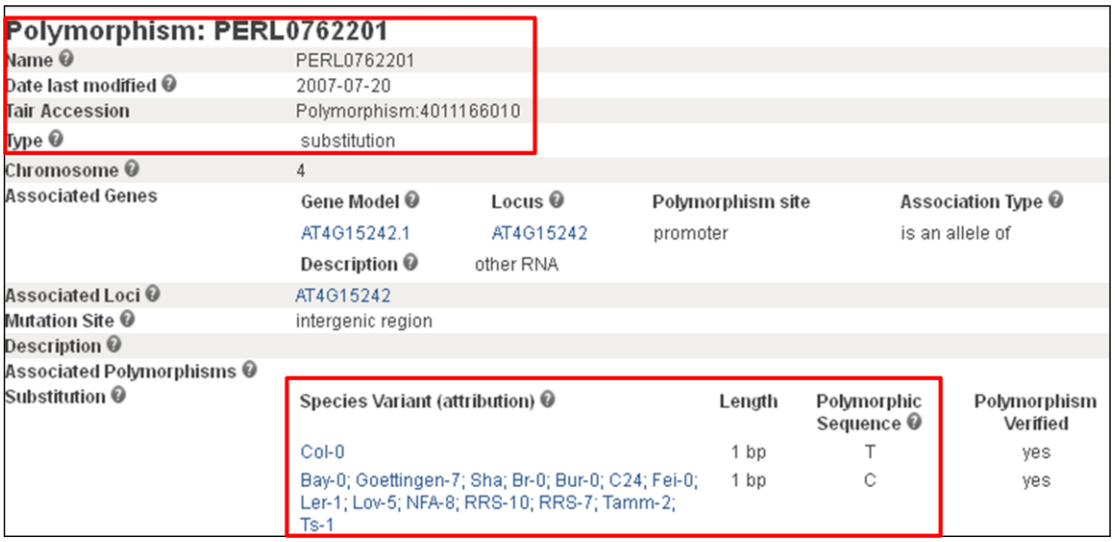
The differences in DNA methylation and expression profiles of the *ncRNA locus* between *Arabidopsis Col* and *Ws* allowed the locus to be used to study epiallele stability between the two accessions. To analyse epiallele stability *Arabidopsis Ws* and *Col* were crossed to produce a *Ws/Col* F1 hybrid. The resulting hybrid was self-fertilised to produce a segregating F2 population (Figure 3.8A). The *ncRNA* alleles were followed using a *Col* specific polymorphism upstream of the transcribed region (Figure 3.8B), which was identified using a polymorphism search tool (http://www.arabidopsis.org/). Individual F2 lines that contained the three possible combinations of *ncRNA* alleles were selected from the segregating population and the DNA methylation and expression levels of the *ncRNA locus* were determined (Figure 3.9 A and B). Region -129 to +1 (Figure 3.6A) where there is a sharp transition between methylated and un-methylated cytosines in *Arabidopsis Ws*, was used as a target region to analyse DNA methylation levels in the F2 lines (Figure 3.9A). The methylation levels in the F2 lines are almost an exact replica of the methylation levels in their respective parent. In addition, expression of the *ncRNA locus* was dependent on the parent of origin alleles. For example, the F2 line with two *Ws ncRNA* alleles has the highest *ncRNA locus* expression levels. However, expression of the *ncRNA*

locus could not be detected in the F2 line with two *Col ncRNA* alleles. The F2 line with *Col* and *Ws ncRNA* alleles showed intermediate expression levels (Figure 3.9B). This suggests that the *Col* and *Ws ncRNA* epialleles are stable when transmitted through a *Col/Ws* hybrid.

A.



B.



Polymorphism: PERL0762201

Name	PERL0762201
Date last modified	2007-07-20
Tair Accession	Polymorphism:4011166010
Type	substitution
Chromosome	4

Associated Genes

Gene Model	Locus	Polymorphism site	Association Type
AT4G15242.1	AT4G15242	promoter	is an allele of

Description: other RNA

Associated Loci: AT4G15242

Mutation Site: intergenic region

Description:

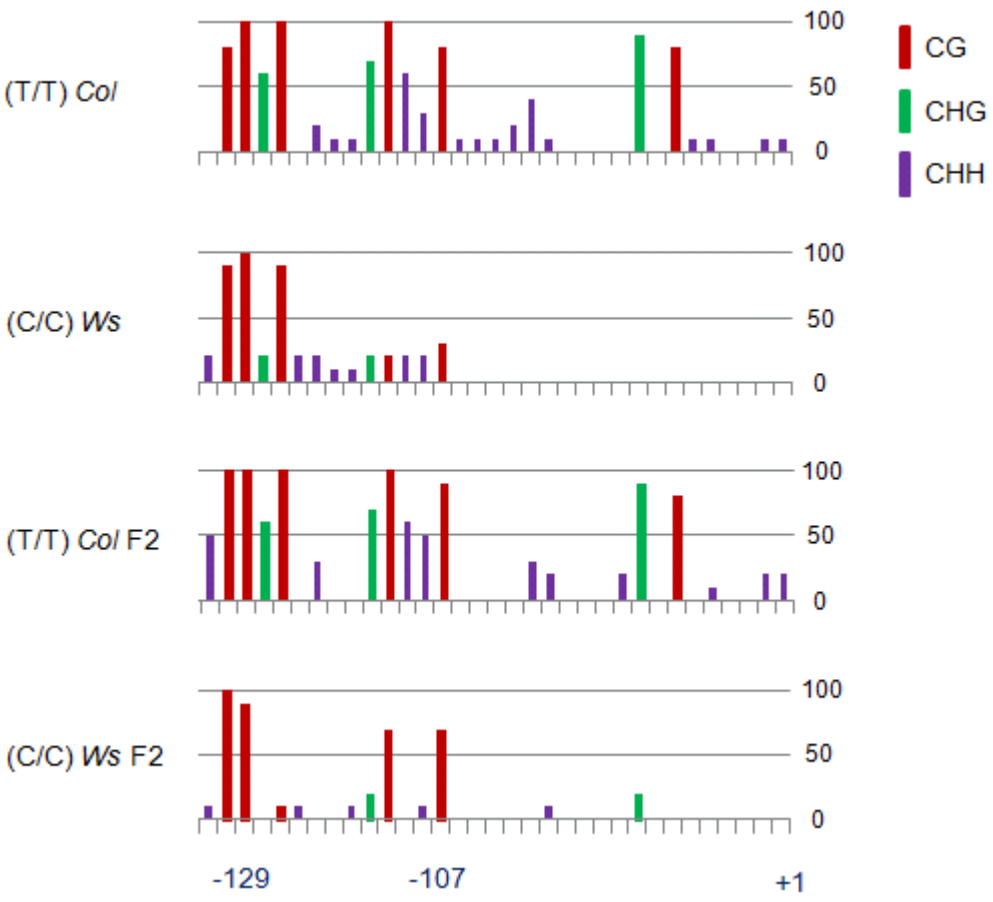
Associated Polymorphisms

Substitution

Species Variant (attribution)	Length	Polymorphic Sequence	Polymorphism Verified
Col-0	1 bp	T	yes
Bay-0; Goettingen-7; Sha; Br-0; Bur-0; C24; Fei-0; Ler-1; Lov-5; NFA-8; RRS-10; RRS-7; Tamm-2; Ts-1	1 bp	C	yes

Figure 3.8. Generating an *Arabidopsis Col/Ws* segregating population to analyse the stability of the *ncRNA* epialleles between the accessions. **A**. The figure shows a schematic diagram of crosses between *Arabidopsis Wassilewskija (Ws)* and *Columbia (Col)* and subsequent self pollinations of the resulting hybrids. Crosses were carried out according to Section 8.1.1 of the Materials and Methods. Generations are shown on the left and *Ws* and *Col* derived alleles are shown in red and green, respectively. Beginning from the top, F0 wild-types were crossed to generate a *Ws/Col* F1 hybrid. The hybrid was self fertilised to produce a segregating F2 population. **B**. The figure shows a screenshot highlighting a *Col* specific polymorphism located upstream of the *ncRNA locus* identified using a polymorphism search tool (http://www.arabidopsis.org/). The top red box highlights the name and type of the polymorphism and the lower red box highlights the base (T) specific to the *Col* accession.

A.



B.

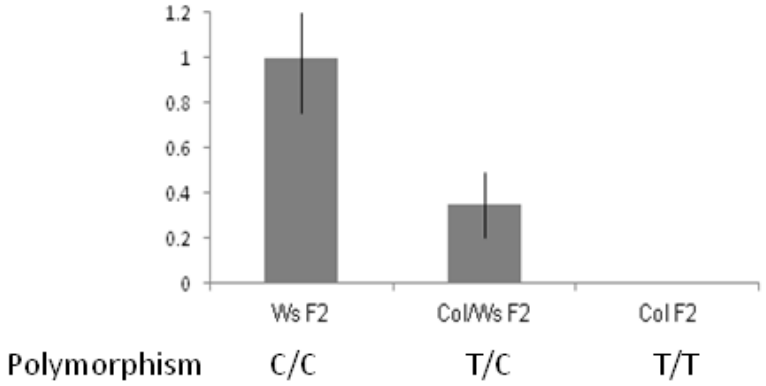


Figure 3.9. Analysing *Arabidopsis* accession specific epiallele stability. **A**. The figure shows DNA methylation levels at individual cytosines at a target region (-129 to +1) within the *ncRNA* promoter in *Arabidopsis Col, Ws* and *Col/Ws* F2 segregating lines. DNA methylation levels were analysed as described in Figure 3.2 using DNA from basal rosette leaves. The respective polymorphism for each accession and hybrid is provided on the left. A percentage scale is on the Y-axis of each graph and methylation types are colour coded according to the key. **B**. The figure shows a qRT-PCR analysis to determine expression levels of the *ncRNA locus* in *Col/Ws* F2 segregating lines, performed as described in Figure 3.7 using RNA from one basal rosette leaf. Expression levels on the Y-axis, which were normalised against *ELONGATION FACTOR* gene expression, are shown in comparison to expression in *Ws* F2. The respective polymorphism for each line is provided below the X-axis.

3.2.7. Analysing the activity of the *ncRNA* promoter using a reporter gene

The absence of DNA methylation within the transcribed region of the *ncRNA locus* in *Ws* highlights the possibility that methylation marks within the transcribed region could contribute to expression. To analyse if the transcribed region is required for expression of the *ncRNA locus*, the *ncRNA* promoter was fused with a reporter gene (*β–GLUCURONIDASE, GUS)* and transferred into *Arabidopsis Col* (Figure 3.10). Transferring the reporter gene construct into a *met1* homzygous mutant was attempted via floral dip but was not possible, most likey due to the reduced fertilitfy of the mutant. *GUS* expression could not be detected in *Col* reporter gene lines using the histochemical assay. One reason for the inability to detect *GUS* expression using the histochemical assay could be a weak activity of the *ncRNA* promoter, which made it necessary to use a more sensitive sqRT-PCR analysis to detect *GUS* expression. cDNA from a *Col* reporter gene line and *met1* were analysed for both *GUS* and *ncRNA locus* expression to compare the promoter activity of the *ncRNA locus* with and without its transcribed region. Expression of the *ncRNA locus* was detected at 30 cycles in *met1* and *GUS* expression was detected at 39 cycles in the reporter gene line (Figure 3.11). While differences in transcript stability, primer pair efficiencies and the mutant background could all contribute to the different expression levels between *GUS* in *Col* and the *ncRNA locus* in *met1*, a drastic difference of 9 cycles is an indication that the transgenic promoter in *Col* may be weaker than the endogenous promoter in *met1* (Figure 3.7). To determine if the weaker expression of the reporter gene was due to DNA methylation, the methylation profiles of the 107 bp promoter regions were analysed by bisulfite sequencing ten clones from *met1* and the reporter gene line. It was evident that methylation levels were lower at the transgene promoter than that of the endogenous promoter in *met1* (Figure 3.11). These data indicate that the transgenic *ncRNA* promoter itself may be weaker than the endogenous promoter in *met1* supporting a model that predicts a role of the transcribed region in *ncRNA locus* expression.

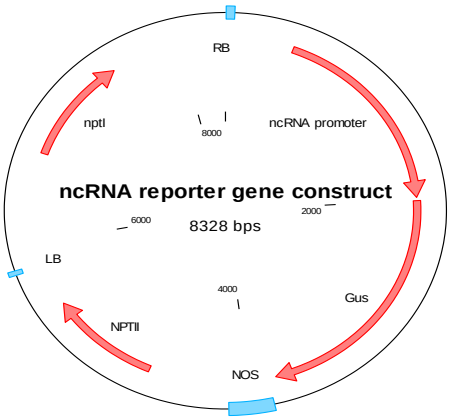


Figure 3.10. A map of the reporter gene construct transferred into *Arabidopsis* to analyse *ncRNA* promoter activity. The *ncRNA* promoter (-1069 to -1) was amplified and inserted into the polylinker region of the pGreen 0029 *GUS* vector as described in Sections 8.1.4.6 and 8.2.1.7.7 of Materials and Methods. *Right boarder* (*RB*) and *left boarder* (*LB*) mark the T-DNA boundaries. *NPTI* and *NPTII* mark the bacterial and plant *NEOMYCIN PHOSPHOTRANSFERASE* genes, respectively, which confer resistance to kanamycin. *NOS* is the terminator region from the *NOPALINE SYNTHASE* gene. Red arrows indicate gene orientation with the arrow head corresponding the to 3' end of the respective gene.

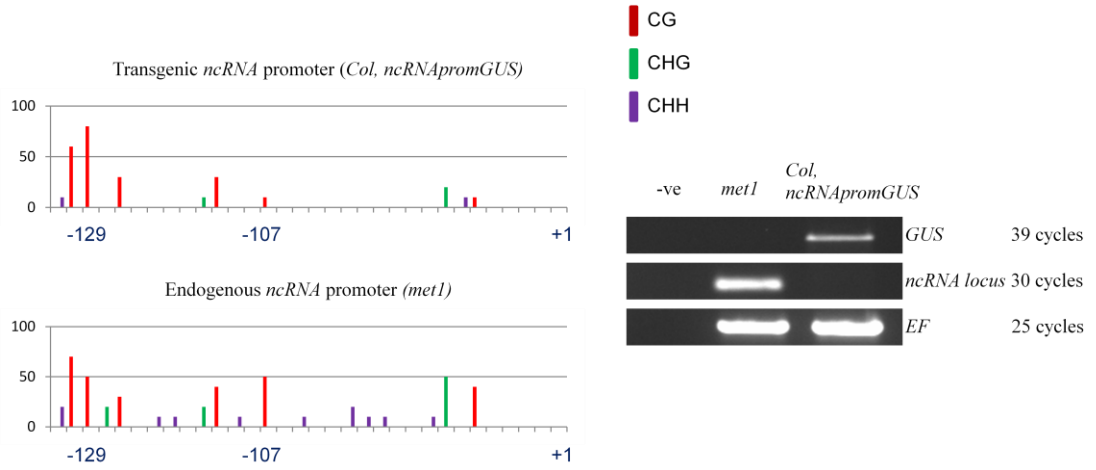


Figure 3.11. DNA methylation levels and activity of the transgenic *ncRNA* promoter in *Arabidopsis Col* and the endogenous *ncRNA* promoter in *met1*. The methylation profiles of the transgenic and endogenous *ncRNA* promoter between positions -129 and +1 are shown on the left for *Col* and *met1,* respectively. Methylation levels at individual cytosines were analysed as described in Figure 3.2. A sqRT-PCR analysis, preformed as described in Figure 3.1, is shown on the right. Samples are labelled on the top and, from left to right, include a water only control (-ve), *met1* mutant and *Arabidopsis Col* transformed with the *ncRNA* reporter gene construct. Targets genes and the PCR cycle numbers are provided on the right. *ELONGATION FACTOR* (*EF*) gene expression was used as an internal control to compare target gene expression levels.

3.3. Discussion

In this chapter, the *Arabidopsis met1* mutant has been exploited to introduce heritable epigenetic variation. The effect of the *met1* mutation was analysed by identifying epigenetic changes induced in *met1* and following their stability when MET1 function was restored. Heritable gene expression changes induced in *met1* were detected for *AT4G15242*, which codes for a *ncRNA* (Figure 3.4). In contrast, for *AT4G25530*, which codes for the FWA transcription factor (Koornneef et al, 1991), the *met1* induced gene expression changes were reset when one wild-type *MET1* allele was restored (Figure 3.4). Interestingly, previous studies using the same method but with a *ddm1* mutant to initiate DNA methylation changes, found that the *FWA* gene remained active upon restoration of the wild-type alleles (Kinoshita et al, 2007). The reason for these differences in *FWA* epiallele stability when demethylation is induced by the different epigenetic modifier mutants is unknown. The possibility that pre-existing DNA methylation enhances silencing has been shown for *FWA* (Chan et al, 2006). Therefore, these differences may arise if demethylation induced by inactivating DDM1 is more efficient at this locus than in *met1,* which could be possible, because DDM1 is involved in facilitating both MET1-regulated CG methylation (Soppe et al, 2002) and CMT2 regulated CHH methylation (Zemach et al, 2013) at some loci. The speculation that pre-existing DNA methylation facilitates the reseting of *FWA* expression observed in this study is also appealing considering *met1-1* is not a null (Kankel et al, 2003).

Heritable DNA demethylation at the *FWA* promoter induced in *met1* has previously been reported, using a *Cfo*I methylation-sensitive Southern blot analysis (Kankel et al, 2003). Heritable DNA methylation changes at the *FWA* promoter would not support the resetting of *FWA* expression when MET1 is restored. It is, however, important to consider that only four *Cfo*I sites exist within the *FWA* promoter and using *Cfo*I for methylation-sensitive Southern blot analysis would only define methylation at four cytosines in a CG sequence context (Kankel et al, 2003). This highlights the importance of analysing target gene expression in addition to DNA methylation when defining epialleles, and a thorough analysis of locus-specific methylation profiles by bisulfite sequencing.

The stable expression of the *ncRNA locus* and the resetting of *FWA* expression in *MET1* restored lines, indicates that *met1* induced changes are reset at some loci, while at others they are not. A possible explanation for this variation in stability could be that some loci are under strict epigenetic control, such as loci that are associated with 24-nt siRNAs that would target *de novo* DNA methylation. *De novo* methylation can then only be effectively maintained by the presence of a functional MET1 protein. Such epigenetic control has been observed for some transposable elements (TEs) (Teixeira et al, 2009). TEs are repetitive and therefore their transcripts are capable of forming double-stranded RNA. Double-stranded RNA is cleaved by a DICER enzyme complex into 24-nt siRNAs, which function in RdDM to establish DNA methylation at homologues regions of the genome by DRM2 (Teixeira et al, 2009). Interestingly, *FWA* contains two direct repeats within its 5' coding region, which have been shown to be a target for RdDM when the *FWA* locus is introduced into wild-type as a transgene (Chan et al, 2006) and the *FWA* locus remains unmethylated and expressed when introduced into a *drm2* mutant (Chan et al, 2006). Direct repeats and DRM2 silencing are features lacked by the *ncRNA locus*. These observations indicate that *FWA* has distinct features required for RdDM, which the *ncRNA locus* lacks. Therefore, the possibility that DNA methylation targeting systems are contributing factors in target gene resetting becomes an appealing speculation.

An interesting aspect with regard to the DNA methylation changes induced at the *ncRNA locus* by inactivating MET1, were the stable changes to DNA methylation at CHG and CHH sequence types. Changes in CHG and CHH methylation were interesting as MET1, CMT3 and DRM2 are the classical CG, CHG and CHH methyltransferases, respectively (Cokus et al, 2008). While some redundancy in function has been observed between DRM2 and CMT3 (Cao & Jacobsen, 2002a), the evidence in this chapter would indicate that MET1 also contributes to DNA methylation at some CHG and CHH sequences in wild-type. While it is still not known whether these effects of MET1 on CHG and CHH methylation are direct or indirect, the same has been observed for the *REPETITIVE PETUNIA SEQUENCE* (*RPS*) element. The *RPS* element attracts DNA methylation when introduced into wild-type *Arabidopsis* but loses DNA methylation at all three sequence types when transferred into *met1* by a genetic cross (Singh et al, 2008). Singh et al, (2008) proposes that these effects are likely indirect, as MET1 binding may be required to

recruit CMT3 or DRM2 guiding factors (Singh et al, 2008). It has also recently emerged that MET1 can *de novo* methylate the body of some genes (Zubko et al, 2012), which provides another discrepancy from the classical model for MET1 as a CG-specific maintenance methyltransferase. These results emphasise the complexity of epigenetic mechanisms, stress the importance of not over-simplifying scientific concepts and prompts a re-assessment of functions and target specificity of plant DNA methyltransferases.

In addition to direct DNA methylation changes and an increase in expression of target genes, some target genes altered their expression in the *met1* mutant with no direct changes in DNA methylation. These indirect changes included an increase in expression for *AT4G10850*, which encodes a Nodulin MtN3 family protein and a reduction in expression for *AT1G19070* and *AT2G41380*, which encode an F-box family protein and an S-adenosyl methionine dependent methyltransferase, respectively. It is unlikely that transcriptional competition from local over-lapping or antisense genes results in the indirect changes in gene expression of *AT4G10850*, *AT1G19070* and *AT2G41380* as antisense genes were inactive in *met1*. At this stage, changes in expression of *AT4G10850, AT1G19070* and *AT2G41380* in *met1* are caused by unknown indirect effects. The fact that *AT2G41380* encodes a protein methyltransferase, which requires the same substrate as MET1, S-adenosyl methionine, led to the speculation about a substrate-specific feedback loop between these methyltransferase proteins. *AT2G41380* would be an ideal target to analyse when *MET1* is over-expressed, as more MET1 could result in reduced SAM levels. AT2G41380 would have to compete more strongly for SAM and may be up-regulated to meet this competition.

The stability of the *ncRNA locus* and its epigenetic variation between *Arabidopsis Col* and *Ws* accessions was demonstrated when characterising its epigenetic regulation. While the *ncRNA locus* is silent in *Arabidopsis Ler* and *Col*, which requires DDM1 and MET1 but not DRM2, it is expressed in *Arabidopsis Ws* wild-type. The reasons for epigenetic variation between *Arabidopsis* accessions are unknown. Speculative models that could explain epigenetic variation include genetic variation, *trans*-acting factors and environmental signals (Fujimoto et al, 2012). The DNA sequence of the *ncRNA* upstream region, transcribed region and downstream region are identical

between *Col* and *Ws*. Additionally, if there were sequence-specific differences influencing expression of the *ncRNA locus* it would not be possible to activate a silent *Col* allele in *met1*. It is therefore unlikely that genetic variation is responsible for the difference in *ncRNA locus* expression between the accessions. It is also unlikely that *trans*-acting factors are responsible because of the surprising stability of DNA methylation and expression profiles observed in the F2 offspring of a *Col/Ws* hybrid parent. However, at this stage, the involvement of environmental factors in *ncRNA* accession specific epigenetic variation is unknown and cannot be eliminated. The default state of the *ncRNA locus* is also not clear but *Ws* was the only accession with an active *ncRNA* allele, which would argue that the default state is inactive.

A comparison of DNA methylation profiles between the active *Ws* allele and the silent *Col* allele of the *ncRNA locus* identified differences in the 3' end of the promoter and in the transcribed region. Both epialleles show a similar methylation profile in the upstream promoter region (-748 to -129), while in *Ws* methylation is almost completely eliminated -107nt upstream of the transcription start site and within the transcribed region. This small 107nt hypomethylated region within the *ncRNA* promoter of *Ws* raises questions with regards to the relevance of the transcribed region itself and methylation marks within the transcribed region for expression of the *ncRNA locus*. These questions are encouraged by weak expression of a reporter gene under the control of the *ncRNA* promoter. Evidence that DNA methylation inhibits transcriptional elongation but not initiation has been observed in *Neurospora crassa* (Rountree & Selker, 1997). This could be a model that applies to the *ncRNA locus* if the methylation within the transcribed region has a role in expression. Enhancer elements have been found within the transcribed region of some genes, including *ELONGATION FACTOR 1β* (Gidekel et al, 1996) and *AGAMOUS* (Sieburth & Meyerowitz, 1997) in *Arabidopsis*. The presence of an enhancer element in the transcribed region of the *ncRNA locus* could explain its silent state in *Col* if the methylation prevented access of regulatory proteins to the enhancer element.

To analyse the role of the *ncRNA* transcribed region in expression of the *ncRNA locus* would require the transfer of the *ncRNA locus* into *Arabidopsis*. The expression level of the transgene when free of DNA methylation could be compared to the endogenous

gene in *Ws* and *met1*. This would provide some insight into the role of the *ncRNA* transcribed region in its own expression or whether there could be regulatory DNA elements further upstream of the endogenous gene. To analyse the role of methylation marks within the transcribed region in *ncRNA locus* expression would require targeting of DNA methylation to the transcribed region of the transgenic *ncRNA locus* and subsequent expression analysis. Targeting of DNA methylation could be achieved using an inverted repeat to generate 24-nt siRNAs homolgous to the *ncRNA* transcribed region. Silent lines could be crossed with the *met1* mutant to analyse if methylation in the transcribed region is lost and the transgene expressed.

The objective of this chapter was to identify a strategy to induce heritable DNA methylation changes in the model organism *Arabidopsis*. This was accomplished for a target gene, which codes for a *ncRNA*, using the epigenetic modifier mutant *met1* to induce DNA methylation changes. This is a strategy that can now be exploited in other plants to induce variation without changing the DNA sequence. In addition, the project unexpectedly found that the epigenetic and transcriptional states of the *ncRNA locus* vary between *Arabidopsis Col* and *Ws* wild-type accession. The stability of the epigenetic and transcriptional states was demonstrated by almost identical methylation and expression profiles in hybrid offspring.

4.0. Investigating *MET1* over-expression and methylation-independent functions in *Arabidopsis*

4.1. Introduction

In *Arabidopsis*, maintenance of methylation at CG sequences is catalysed by DNA METHYLTRANSFERASE 1 (MET1) (Kankel et al, 2003). MET1 consists of 1534 amino acids and shares significant homology with the mammalian maintenance methyltransferase DNMT1. Amino acids 1 to 1093 of MET1 contribute the N-terminal domain that contains the Nuclear Localisation Sequence (NLS), Replication Foci Targeting Sequence (RFTS), and two Bromo-Adjacent Homology (BAH) domains, which may facilitate protein-protein interactions (Pavlopoulou & Kossida, 2007). Protein interaction assays and mutational analysis suggest that interactions of the RFTS domains result in DNMT1 dimerisation (Fellinger et al, 2009), although direct evidence of whether MET1 acts as a dimer via its RFTS domain is yet to be demonstrated. The C-terminal region of MET1, which is joined to the N-terminal domain via Glycine-Lysine (GK) repeats, contributes the catalytic domain. Functional motifs within the catalytic domain are identified by Roman numerals of which motifs I and X are required for SAM binding, and the active site is represented as motif IV (Figure 4.1) (Pavlopoulou & Kossida, 2007).

Knockdown (Finnegan et al, 1996) and knockout (Kankel et al, 2003) approaches, which result in DNA hypomethylation and developmental abnormalities, have been used to study MET1. The effects of increasing *MET1* levels have never been assessed in plants, as the current CG maintenance model would suggest over-expression of MET1 would result in maintenance of CG methylation. However, the recent evidence described in Chapter 3 suggests that MET1 functions deviate from the CG maintenance model. A reduction in DNA methylation at CHG and CHH sequences was detected in the *met1* mutant, which highlights a new role for MET1 in influencing non-CG methylation and raises questions with regards to the role of its catalytic domain in this function. Down-regulation of target genes without local DNA methylation changes in *met1* indicate MET1 levels may have a quantitative effect on gene expression. Another example in which MET1 functions deviate from the maintenance model is provided by Zubko et al, (2012) when they found that

methylation lost from the body of an endogenous target gene in a *met1 Arabidopsis* mutant, was partially restored at CG sites when MET1 was re-introduced. Re-methylation did not require passage through the germline, which suggests MET1 may have *de novo* activity at CG sequence contexts (Zubko et al, 2012). These examples of MET1 functions that deviate from the classical CG maintenance model prompt a re-assesment of the effects of MET1 over-expression, especially if increasing MET1 levels was to alter target gene expression or methylation levels.

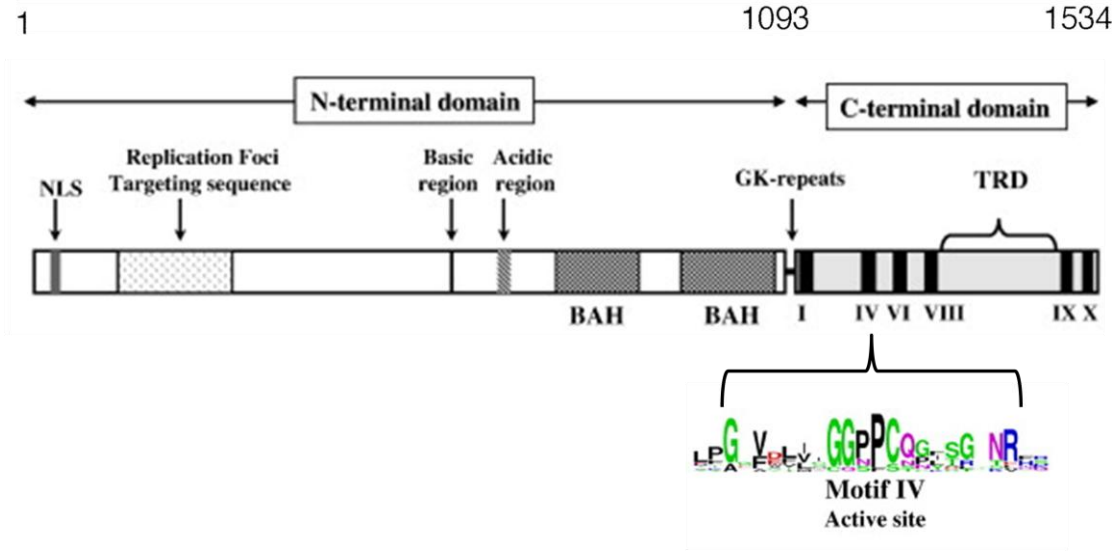


Figure 4.1. MET1 protein structure. The figure shows a schematic diagram of the MET1 protein. Amino acid numbers are shown above the schematic. The N-terminal domain of MET1 (left) contains the Nuclear Localisation Sequence (NLS), replication foci targeting sequence, basic and acid regions and two Bromo-Adjacent Homology (BAH) domains. The C-terminal domain of MET1 (right) contains the catalytic motifs indicated by Roman numerals. A sequence logo is provided below the active site of MET1 (Motif IV), which shows the most conserved amino acids in this motif between plant DNA methyltransferases. Image modified from Kankel et al, (2003) and Pavlopoulou & Kossida, (2007).

To analyse quantitative effects of MET1, the *MET1* gene was over-expressed in *Arabidopsis* using the 35S CaMV promoter. In parallel, a mutated *MET1* gene (*MET1mut*) was over-expressed to analyse methylation-independent effects of MET1 over-expression. Target gene expression changes and developmental phenotypes that are caused by inactivation of *met1* were analysed in the over-expression lines.

4.2. Results

4.2.1. The production of *MET1* and *MET1mut* over-expression constructs

To analyse the quantitative effects of MET1, it was first necessary to produce an over-expression construct. This was accomplished by removing the *MET1* cDNA sequence from the p-GEM T easy vector (Promega) and inserting it into the polylinker region of the plant transformation vector 35S pGreen 0179 (Figure 4.3).

To analyse methylation-independent effects of *MET1* over-expression in *Arabidopsis*, a construct was produced to over-express *MET1mut*, which encodes the MET1 protein with its catalytic activity removed. To remove the catalytic function from MET1 the strategy employed by Hsieh, (1999b) was exploited (Hsieh, 1999b). In this study the active site cysteine of DNMT3B is replaced with a serine. The analogous region of DNMT3Bs active site loop region in MET1 is GGPP**C**QGFSGMNRFN (Figure 4.1). In *MET1*, the region which produces these amino acids is located between the unique restriction enzyme sites, *Bsu*36I and *Ppu*MI. Two pairs of primers were used to amplify the 5' and 3' regions, which encode the MET1 catalytic domain (Figure 4.2). One primer from each pair overlaps at the TGT sequence encoding the cysteine, and primers were therefore designed so that the TGT sequence was replaced by TCT to encode a serine when the two amplicons were assembled (Figure 4.2). After assembly the amplicon was cut with *Bsu*36I and *Ppu*MI and inserted into the 35S *MET1* construct cut with the same restriction enzymes, to produce the *MET1mut* over-expression construct (Figure 4.3).

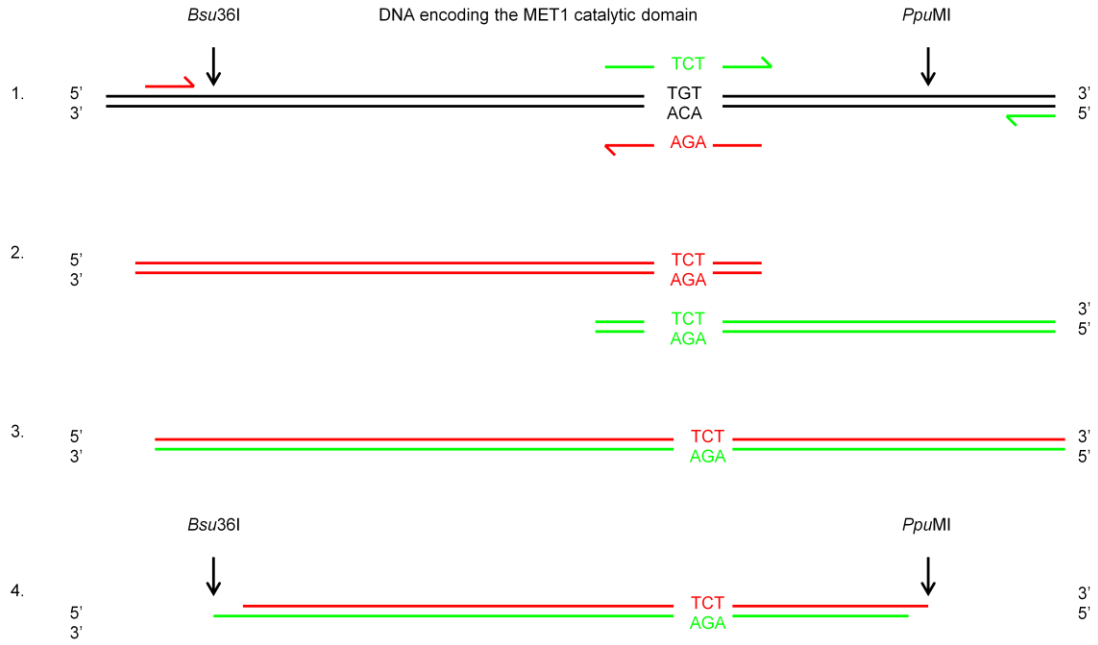


Figure 4.2. Assembly PCR for site directed mutagenesis of the MET1 catalytic domain. The figure shows a schematic diagram of an assembly PCR reaction to introduce a mutation within the sequence encoding the MET1 catalytic motif. **1**. Two pairs of primers were designed to amplify two overlapping regions that encode parts of the MET1 catalytic domain. These primer pairs are colour coded red and green. The reverse primer (red) used to amplify the 5' region and the forward primer (green) used to amplify the 3' region were designed to introduce a G to C mutation in each amplicon, **2**. **3**. The two amplicons were mixed in a PCR reaction where they prime each other producing 1 mutated amplicon. **4**. The amplicon contains unique restriction enzymes sites *Bsu*36I and *Ppu*MI allowing the exchange of the wild-type sequence with the mutated sequence after restriction enzyme digestion.

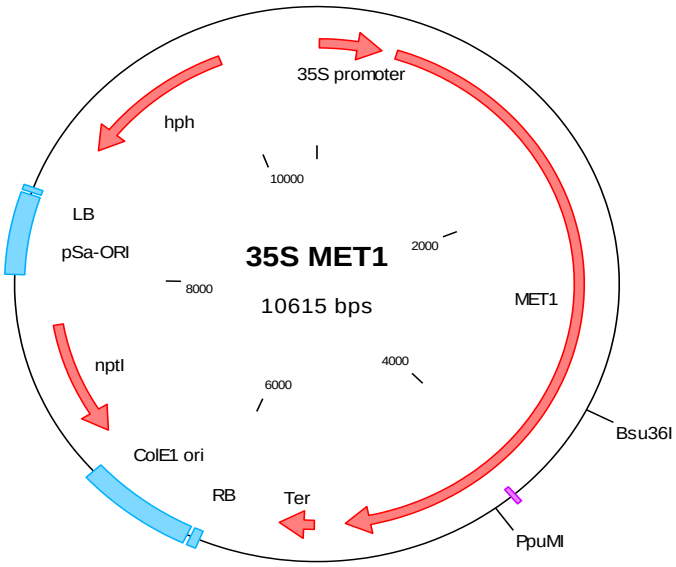


Figure 4.3. Maps of the 35S *MET1* and 35S *MET1mut* constructs. The 35S *MET1* and 35S *MET1mut* constructs were produced according to Sections 8.2.1.7.1 and 8.2.1.7.2 of Materials and Methods, respectively. The sequences of the primers used to produce the constructs are provided in Section 8.1.4.6 of Materials and Methods. *LB* (left boarder) and *RB* (right boarder) mark the T-DNA region. The T-DNA region contains the plant selectable marker hygromycin (*HPH*) and the 35S *MET1* cassette. ColE1-ori and pSa-Ori mark the replication of origin in *E. coli* and *Agrobacterium,* respectively. Red arrows indicate gene orientation with the arrow head corresponding to the 3' end of the respective gene. A mutation was introduced into the 35S *MET1* construct to replace a cysteine with a serine codon within the active site (Figure 4.2) and is highlighted by the purple marker.

4.2.2. The production of 35S *MET1* and 35S *MET1mut* *Arabidopsis* transformants

The two transgenic constructs were transferred into *Arabidopsis*, and for each, four T1 transgenic lines were selected that over-express *MET1* or *MET1mut*, respectively (Figure 4.4). At this stage of the analysis the copy number status and the zygosity of the transformants is unknown.

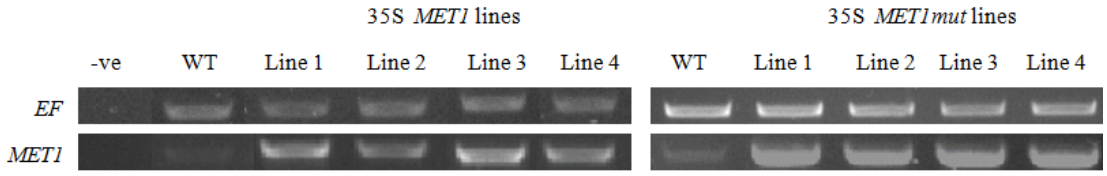


Figure 4.4. *MET1* and *MET1mut* expression in *Arabidopsis* transformants. The figure shows a sqRT-PCR analysis of *MET1* over-expression in *Arabidopsis* 35S *MET1* and 35S *MET1mut* transformants (T1). The analysis was performed as described in Figure 3.1 using RNA from one basal rosette leaf from plants grown as described in Materials and Methods Section 8.2.5. *ELONGATION FACTOR* (*EF*) was used as an internal control to compare gene expression levels to wild-type (WT) and "-ve" refers to a water only reaction. 4 individual over-expressing transformants for each construct were identified and are indicated above each gel lane on the left and right for *MET1* and *MET1mut,* respectively.

4.2.3. Identification and analysis of target genes in *Arabidopsis MET1* and *MET1mut* over-expression transformants

To analyse the effects of *MET1* over-expression it was necessary to identify target genes. Target genes were identified using two strategies. The first strategy selected genes that were targets for both methylation and demethylation functions by screening the epigenome of the DNA glycosylase mutants using the methylome mapping tool (http://signal.salk.edu/cgi-bin/methylome) (Mathieu et al, 2007) and the epigenome browser (http://neomorph.salk.edu/epigenome/epigenome.html). The second strategy selected genes identified in Chapter 3 that altered their expression in the *met1* mutant, to investigate if they responded to *MET1* levels quantitatively. Two target genes altered their expression in some of the transformed lines (Figure 4.5). This included *AT5G52310*, which encodes a cold regulated gene *RD29A* (Sako et al, 2012) and *AT2G41380,* which encodes a SAM dependent methyltransferase. Interestingly, *RD29A* was down-regulated in all four lines over-expressing *MET1* but was unchanged when over-expressing *MET1mut*. Contrarily, the expression of *AT2G41380,* which is down-regulated in the *Arabidopsis met1* mutant, was up-regulated when over-expressing both *MET1* and *MET1mut* in three out of four lines. These data suggest that over-expression of *MET1* and *MET1mut* in *Arabidopsis* results in changes in the expression of genes, and over-expressing *MET1* and *MET1mut* seems to be required but not sufficient for expression changes.

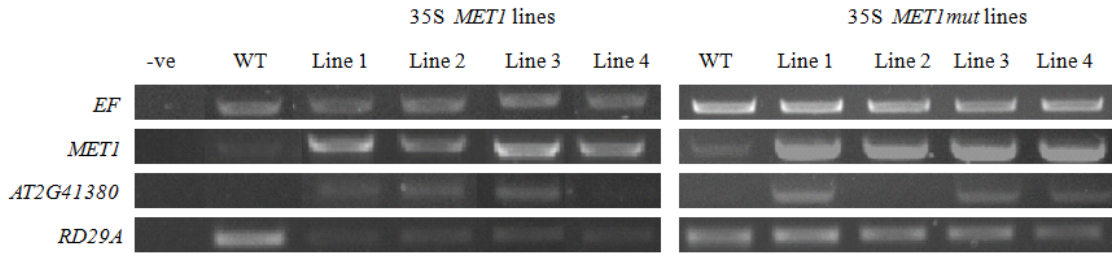


Figure 4.5. Analysing target gene expression in *Arabidopsis* lines over-expressing *MET1* and *MET1mut*. The figure shows a sqRT-PCR to analyse target gene expression in T1 35S *MET1* and 35S *MET1mut* transformants. The analysis was performed according to Figure 4.4. The lines analysed are indicated above the gel image and target genes are provided on the left. "–ve" refers to water only control and *ELONGATION FACTOR* (*EF*) was used as an internal control to compare gene expression levels to wild-type (WT).

4.2.4. Analysing *RD29A* promoter methylation in *Arabidopsis MET1* and *MET1mut* over-expression lines

A reduction in *RD29A* gene expression when over-expressing the *MET1* gene highlighted the possibility of direct repressive DNA methylation changes at this locus in these lines. To investigate direct DNA methylation changes, a target region for the demethylase ROS1 was analysed within the *RD29A* promoter (Sako et al, 2012) by sequencing eight bisulfite converted clones produced from 35S *MET1* line 1 and 35S *MET1mut* line 1 (Figure 4.6). No consistent DNA methylation changes were detected.

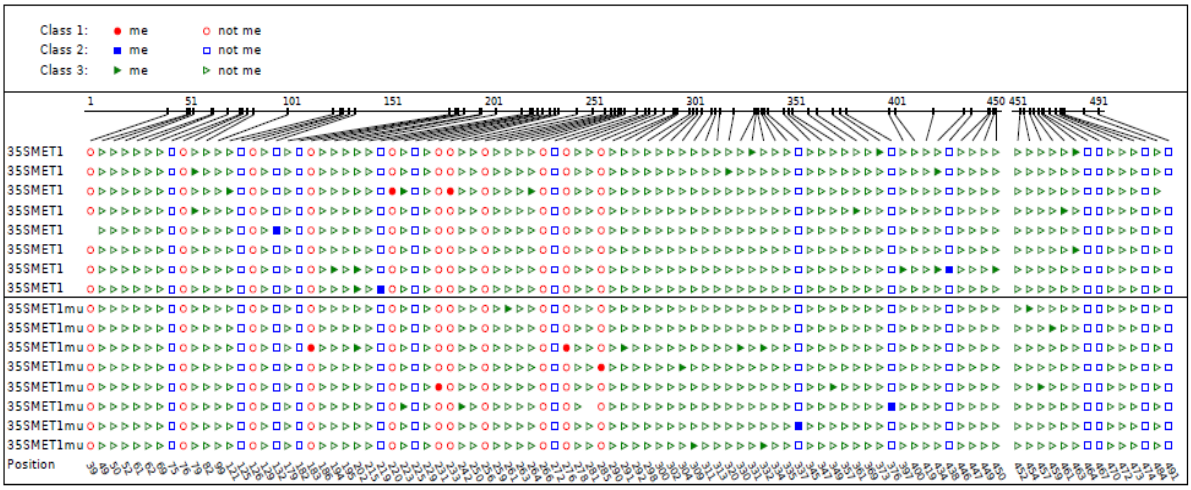


Figure 4.6. Bisulfite sequencing analysis of the *RD29A* promoter in *Arabidopsis* lines over-expressing *MET1* and *MET1mut*. The figure shows a region of the *RD29A* promoter in T2 *Arabidopsis* lines over-expressing *MET1* and *MET1mut* analysed by bisulfite sequencing as described in Section 8.2.1.11 of Materials and Methods using DNA from 2 week old seedlings. The top boxed region contains sequencing data calculated by Cymate (Hetzl et al, 2007) from line 1 over-expressing the *MET1* gene. The bottom boxed region contains sequencing data calculated by Cymate (Hetzl et al, 2007) from line 1 over-expressing the *MET1mut* gene. 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.

4.2.5. Analysing flowering time in *Arabidopsis* over-expressing *MET1* and *MET1mut* genes

A developmental consequence of inactivating *MET1* using an antisense RNA in *Arabidopsis* is a delay in flowering time (Ronemus et al, 1996). To analyse if flowering time is affected by increasing *MET1* levels, flowering time was analysed in the offspring from three primary *MET1* and *MET1mut* over-expression transformants. Flowering time was analysed by counting basal rosette leaf numbers upon bolting in long day conditions (Soppe et al, 2000), and was carried out with the help of Marianne Shewell (P. Meyer lab, the University of Leeds). Data for individual plants (Figure 4.7 A and B) and averages for each line (Figure 4.7 C and D) are shown. Data for *MET1* and *MET1mut* over-expression lines can only be compared to their respective wild-type because the two experiments were not performed in parallel. The T2 plants used in this analysis were selected on kanamycin and contain at least one copy of the transgene. At this stage the zygosity of the lines is unknown. *MET1* over-expression significantly delayed flowering time in line 3 by an average of 25 days compared to wild-type (Figure 4.7C) and *MET1mut* over-expression significantly delayed flowering by an average of seven days in line 2 compared to wild-type

(Figure d 4.5D). Therefore, *MET1* and *MET1mut* over-expression delays flowering in some plants.

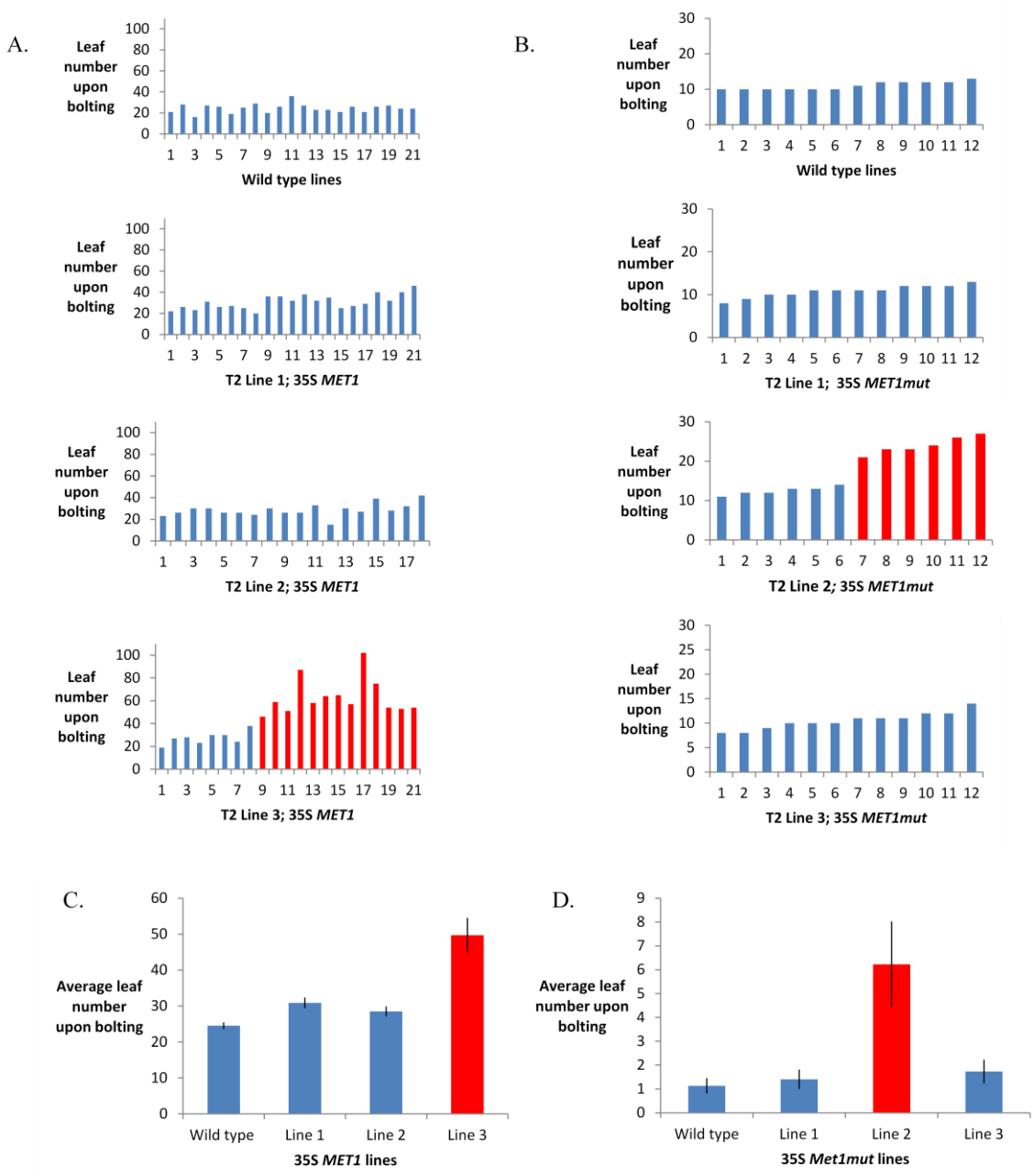


Figure 4.7. Flowering analysis of *Arabidopsis* lines over-expressing the *MET1* and *MET1mut* genes. The figure shows a flowering analysis of *Arabidopsis* lines over-expressing the *MET1* and *MET1mut* genes. T2 plants were grown on selection and analysed as described in Section 8.2.5 of Materials and Methods. **A**. The figure shows data for individual wild-type control plants and offspring from *MET1* over-expression lines 1, 2 and 3. **B**. The figure shows data for individual wild-type control plants and offspring from *MET1mut* over-expression lines 1, 2 and 3. Individual plants are presented on the X-axis and leaf numbers of the Y-axis. **C** and **D** show average values for each line from **A** and **B**, respectively. Significant delays in flowering are highlighted in red.

4.3. Discussion

In this chapter target gene expression changes and a delay in flowering time were found in some lines when over-expressing the *MET1* and *MET1mut* genes in *Arabidopsis* (Figure 4.5 and 4.7).

The target gene *RD29A* was down-regulated when over-expressing *MET1* and was unchanged when over-expressing *MET1mut*. This suggests that *RD29A* expression changes require the catalytic activity of MET1 when *MET1* is over-expressed. While this was an encouraging indicator for DNA methylation changes, no direct DNA methylation changes were found when analysing a demethylase target region within the *RD29A* promoter by bisulfite sequencing. However, not all regions relevant for expression have been analysed for DNA methylation. Recently, it was reported that ten WRKY8 binding sites are present within the *RD29A* promoter (Hu et al, 2013), of which three are within the range of the bisulfite sequencing analysis. Therefore, at this stage, it is not possible to eliminate that direct DNA methylation changes are responsible for *RD29A* down-regulation when over-expressing *MET1*. Additionally, *RD29A* expression and promoter methylation could be analysed in MET1 over-expression lines after stress treatment. In this study plants were grown at 25 ^{o}C under 16/8 h day night conditions. Sako et al, (2012) show that DNA methylation levels increase within the *RD29A* promoter in an *rpt2a* mutant, which has a T-DNA insertion within the *RPT2A* gene encoding a 19S proteasome subunit (Sako et al, 2012). The increase in DNA methylation within the *RD29A* promoter in *rpt2a* was assigned to reduced degradation and accumulation of DNA methyltransferases. Interestingly, Sako et al, (2012) were able to increase DNA methylation levels at the *RD29A* promoter further by exposing the *rpt2a* mutant to 4 ^{o}C for 12 h (Sako et al, 2012), a stress treatment that could be tested on MET1 over-expression lines.

The target gene *AT2G41380* was up-regulated when over-expressing *MET1*. This gene produces a SAM dependent methyltransferase. I hypothesized, because of its down-regulation observed in the *Arabidopsis met1* mutant, described in Chapter 3, that AT2G41380 may be involved in a substrate specific feedback loop with MET1. *AT2G41380* up-regulation when *MET1* levels increase would support this hypothesis. This is not, however, the first time that feedback regulation involving epigenetic

modifiers has been suggested in plants. *ROS1* and *DME*, which encode demethylation functions, are down-regulated in the *Arabidopsis met1* mutant (Mathieu et al, 2007). No direct DNA methylation changes were detected in the *ROS1* promoter and the down-regulation was assigned to indirect DNA methylation changes (Mathieu et al, 2007). This feedback loop would be a useful system to prevent further loss of DNA methylation in these genomes.

To analyse if SAM levels are involved in the potential feedback loop, it would be required to analyse mutants in feedback loop components (Figure 4.8). For example, *MET1* and *AT2G41380* levels could be analysed in *S-adenosyl methionine synthetase (sam1/2)* or *S-adenosyl homocysteine hydrolase (hog1)* mutants (Rocha et al, 2005). A feedback system will likely involve multiple regulators. To characterize the feedback system would require a reporter gene construct using the *AT2G41380* promoter to drive expression. When transferred into plants the expression of the reporter gene could be analysed in a T2 population following EMS mutagenesis. Plants with changes in reporter gene expression would have mutations in the *MET1* gene but may also have mutations within other genes required for a feedback system to function.

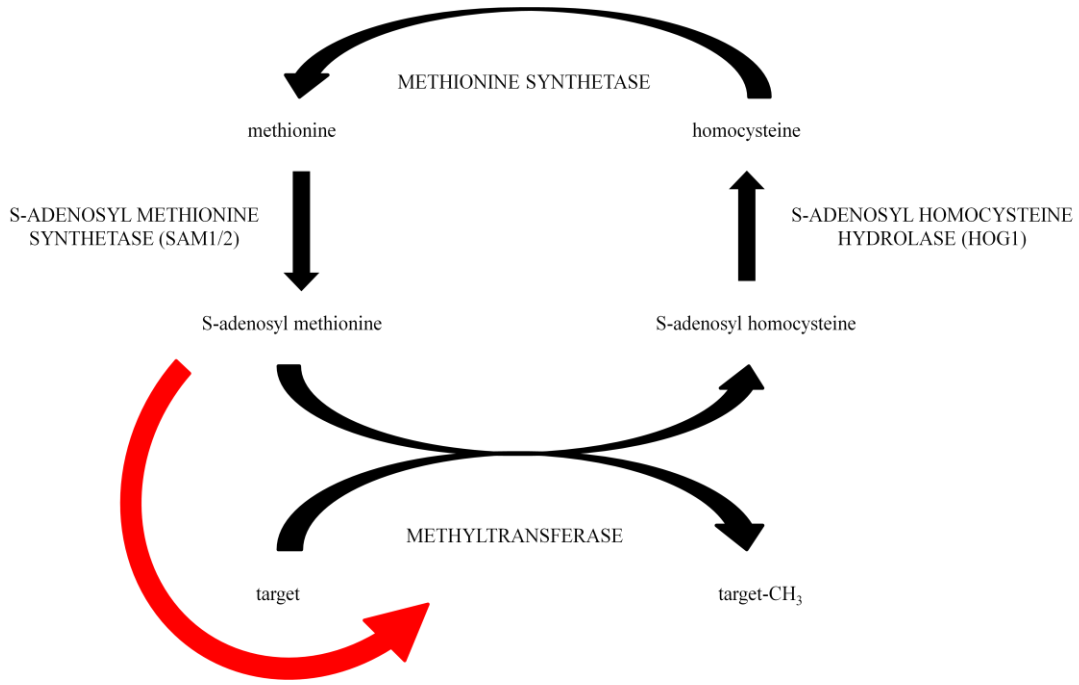


Figure 4.8. A simplified diagram of S-adenosyl methionine and S-adenosyl homocysteine metabolism. Beginning from the bottom left, METHYLTRANSFERASE proteins transfer a methyl-group (-CH_3) from S-adenosyl methionine to target proteins and DNA. The by-product of methylation, S-adenosyl homocysteine is metabolized back to S-adenosyl methionine via an S-ADENOSYL HOMOCYSTEINE hydrolase and METHIONINE and S-ADENOSYL METHIONINE synthetases. It has been speculated that a feedback system exists between S-adenosyl methionine levels and METHYLTRANSFERASE proteins, which is indicated by the red arrow.

AT2G41380 was also up-regulated when over-expressing the *MET1mut* gene. This does not eliminate the feedback hypothesis, as MET1mut may retain SAM binding capacity. *MET1mut* target gene expression changes strengthen the possibility that MET1 may have methylation-independent roles in plants. The same suggestion has been made for the mammalian homologue DNMT1 (Espada et al, 2011). In this case, increases and decreases in target gene expression without DNA methylation changes were observed when expressing DNMT1 with mutations in its N-terminal BAH domains (Espada et al, 2011). Elucidating the methylation-independent functions of DNA methyltransferases will add a new perspective to these epigenetic modifiers.

It is important to notice that *AT2G41380* is not up-regulated in all over-expression lines. It is, in fact, up-regulated in six out of the eight lines. Variation in gene expression caused by epigenetic processes is classically exemplified by position effect variegation (PEV), which is a variegation caused by a change in expression of a gene

in some cells. PEV could explain why *AT2G41380* is up-regulated in some over-expression lines if an epigenetic change, induced by increased MET1 levels, occurred early in development for it to be mitotically inherited and represented either an 'on' or 'off' state throughout the organism.

Over-expressing *MET1* results in delayed flowering of some plants indicating that *MET1* over-expression seems to be required but not sufficient to delay flowering. A delay in flowering time only in some over-expression lines could arise if loci controlled by methylation are regulated by concentration-dependent stochastic effects of *MET1*. For example, lines with higher *MET1* expression levels would have increased probability of a MET1 induced change. MET1 proteins may stochastically associate with a particular locus at a certain developmental stage to induce an epigenetic change in some lines, which is mitotically inherited and which causes a delay in flowering time. Alternatively, MET1 proteins could stochastically bind flowering regulators to delay flowering in some lines. Interestingly, a delay in flowering of some plants also occurs when the *MET1mut* gene is expressed. While this may eliminate a role of CG maintenance, it does not eliminate a role for DNA methylation changes in delayed flowering when over-expressing *MET1mut,* as it is unknown if the catalytic function of MET1 is required to influence non-CG methylation.

FWA expression is induced when methylation is lost from two direct repeats within the 5' coding region of the *FWA* gene (Chan et al, 2006). This activation partially explains the delay in flowering time in hypomethylated mutants (Kankel et al, 2003), as FWA acts as a floral repressor (Ikeda et al, 2007). While direct DNA methylation loss is not a predicted outcome of MET1 over-expression, it has not been eliminated that MET1 over-expression may have a dominant negative effect. If MET1 over-expression has dominant negative effects in *Arabidopsis* it would become interesting to analyse *FWA* expression in MET1 over-expression lines that show wild-type flowering time or a delay in flowering time.

Detecting over-expression of the *MET1* gene and phenotypes are good indicators that MET1 protein levels are elevated in MET1 over-expression lines. Despite these indications MET1 protein levels have never been directly analysed in MET1 over-

expression lines. Over-expression of the MET1 protein could be determined by Western blot analysis after creating FLAG-tagged MET1 over-expression transformants. This is an important consideration as post-transcriptional regulatory mechanisms can interfere with mRNA over-expression being translated into protein over-expression. A final consideration with regards to MET1 over-expression is the possibility that other factors required for MET1 activity, which if not increased with MET1 levels, may be limiting. For example, the chromatin remodelling factor DDM1 is required for MET1 activity (Chan et al, 2005). To determine if and what factors were limiting when MET1 is over-expressed was not within the project objectives.

This chapter presents data which highlight plants as ideal candidates to study methylation-independent functions of DNA methyltransferases. Viable offspring produced from *MET1* over-expression lines suggests that plants can tolerant increases as well as deficiencies (Kankel et al, 2003) in *MET1* levels, while, both increases and deficiencies of *DNMT1* are lethal to the embryo in mammals (Biniszkiewicz et al, 2002). The high tolerance of plants to *MET1* changes will allow catalytically inactive DNA methyltransferases but also methyltransferases with domain deletions and replacements to be fully characterized in wild-type and the respective DNA methyltransferase mutant backgrounds. Analysing methylation-independent functions of mammalian epigenetic modifiers could be extended into plants, which would avoid the lethal complications that arise in mammals from epigenetic modifier inactivation and over-expression. It should, however, be acknowledged that using a heterologous system could complicate interpretations.

5.0. Activity of mammalian DNA demethylation functions in plants

5.1. Introduction

The stability of DNA methylation is exemplified in Chapter 3 when analysing the *ncRNA* promoter methylation across generations. However, DNA methylation marks are epigenetic modifications that can be removed. In plants, this is regulated by two DNA demethylation systems.

Passive DNA demethylation refers to the loss of methylation coupled with replication (Hsieh, 1999a). It occurs when daughter cells inherit hemi-methylated DNA and maintenance methyltransferases fail to propagate DNA methylation onto the newly synthesised strand (Wolffe et al, 1999). Passive demethylation is promoted by nucleoprotein blockage, which describes the occlusion of methyltransferase targets by regulatory nucleoproteins (Hsieh, 1999a) and acetylated histone aversion, when acetylated histones repel methyltransferases (Wolffe et al, 1999).

Active demethylation in plants is initiated when DNA glycosylases remove a methylated cytosine leaving an abasic site, which is replaced by an unmethylated cytosine in the base excision repair (BER) pathway (Gong & Zhu, 2011). Plant DNA glycosylases include REPRESSOR OF SILENCING 1 (ROS1) (Gong et al, 2002), DEMETER (DME), DEMETER-LIKE 1 (DML1), DEMETER-LIKE 2 (DML2) and DEMETER-LIKE 3 (DML3) (Choi et al, 2002). Plant DNA glycosylases function throughout the genome targeting the 5' and 3' ends of genes but not gene body methylation (Penterman et al, 2007). Demethylation targeting mechanisms are still poorly defined but recent studies have identified REPRESSOR OF SILENCING 3 (ROS3). ROS3 is found in discrete foci localised with ROS1 and has the ability to bind single stranded RNAs, which could function in ROS1-targeted demethylation (Zheng et al, 2008).

No orthologs of plant DNA glycosylases are found in mammals but both active global demethylation and sequence specific demethylation have been documented. Global paternal demethylation occurs in the zygote and in primordial germ cells (PGCs) and

sequence specific demethylation occurs in somatic cells in response to various signals (Bruniquel & Schwartz, 2003).

The underlying mechanisms in mammalian DNA demethylation appear to be divided into two systems. ACTIVATION-INDUCED DEAMINASE (AID) and APOLIPOPROTEIN B RNA EDITING CATALYTIC COMPONENT 1 (APOBEC1) can deaminate methylated cytosines to thymine, which creates a thymine/guanine mismatch. The thymine is removed by the thymine DNA glycosylases THYMINE DNA GLYCOSYLASE (TDG) and METHYLCYTOSINE-BINDING PROTEIN 4 (MBD4) to create an abasic site, which is repaired in the BER pathway (Rai et al, 2008). Alternatively, active demethylation in mammals can occur via the oxidation of methylcytosine to either 5-hydroxy-methylcytosine (hmC), 5-formyl-cytosine (fC) or 5-carboxyl-cytosine (caC) (Ito et al, 2011). This oxidation is catalysed by three TEN-ELEVEN TRANSLOCASE (TET1-3) proteins (Figure 5.1) (Ito et al, 2011; Tan & Shi, 2012). Oxidised bases are removed by glycosylase activity and replaced in the BER pathway (Guo et al, 2011).

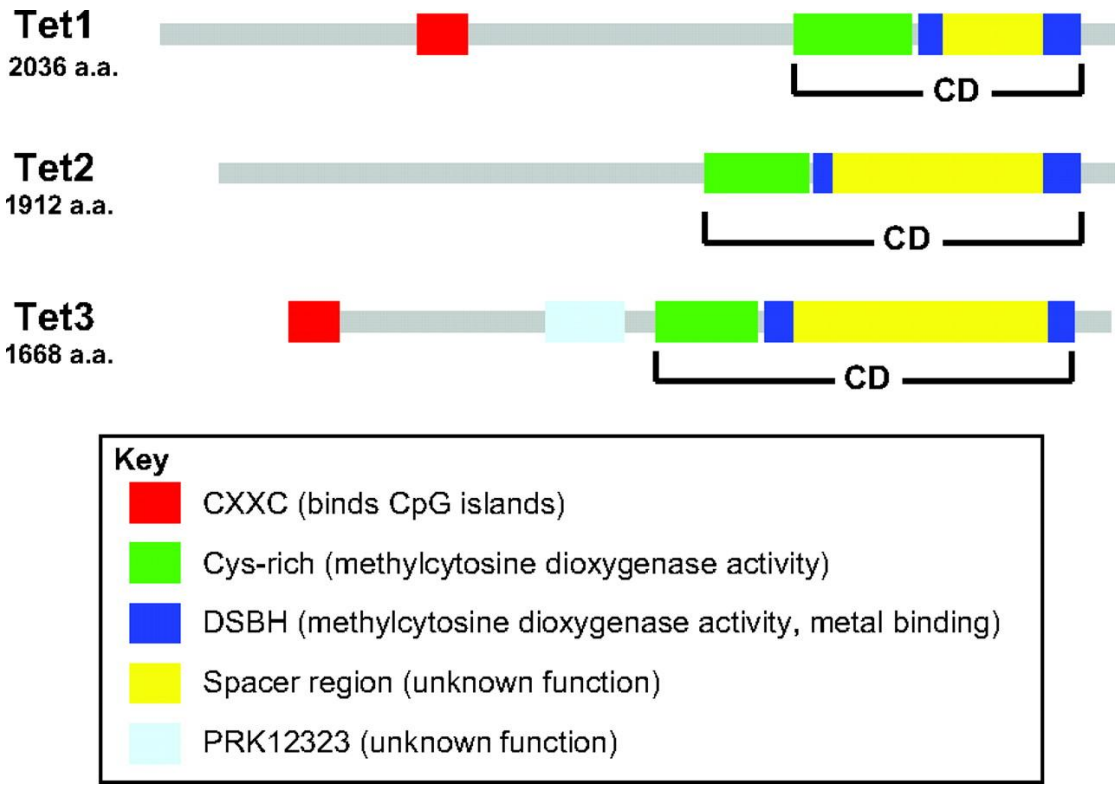


Figure 5.1. Mammalian TET protein structures. The TET protein and size is provided on the left. Diagrams of the TET proteins are provided on the right. The CXXC domain and Cys-rich, double-stranded β-helix (DSBH) and Spacer regions are colour coded and identified using the key. Possible functions of the TET domains and regions are in brackets. The C-terminal domain (CD) is provided below each protein. Taken from Tan & Shi, (2012).

These differences between plant and mammalian DNA demethylation systems provide an opportunity to exploit plants to study mammalian epigenetic modifiers and to test if mammalian demethylase functions can be exploited in plants. The requirement for this is emphasised by the high tolerance of plants to DNA methylation changes, which are lethal to the embryo in mammals (Gruntman et al, 2008). In this chapter, the effects of expressing a region of the *TET3* gene, which encodes the catalytic domain and nuclear localisation sequence (NLS) were analysed in plants.

5.2. Results

5.2.1. Expressing a region of the human *TET3* gene in *Arabidopsis*

To analyse TET3 effects in plants it was necessary to produce *Arabidopsis* transformants that express *TET3*. The human *TET3* cDNA clone was provided by Professor R.Meehan (The University of Edinburgh). The 3' region, which encodes the TET3 catalytic domain and NLS (Appendix 10.4) was inserted into the polylinker of the plant expression vector, 35S pGreen 0179 (Figure 5.2A). The 5' region of the gene was not used, as this encodes the N-terminal region of TET3, which among other functions is likely required for the recruitment of regulatory complexes (Koh et al, 2011) that may not exist in *Arabidopsis*. Four T2 *Arabidopsis* lines that express the 3' region of the *TET3* gene were selected for subsequent analysis (Figure 5.2B). The zygosity of the T2 TET3 *Arabidopsis* lines at this stage of the analysis is unknown.

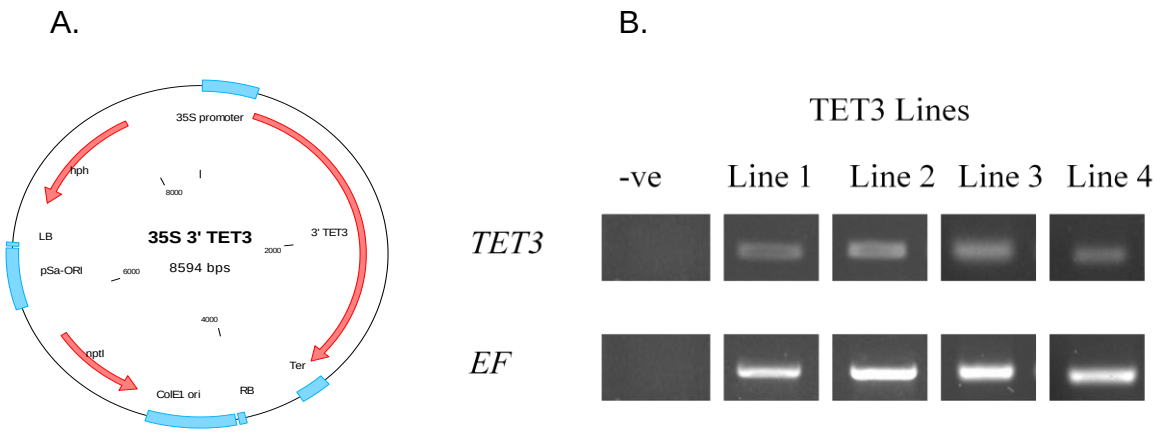


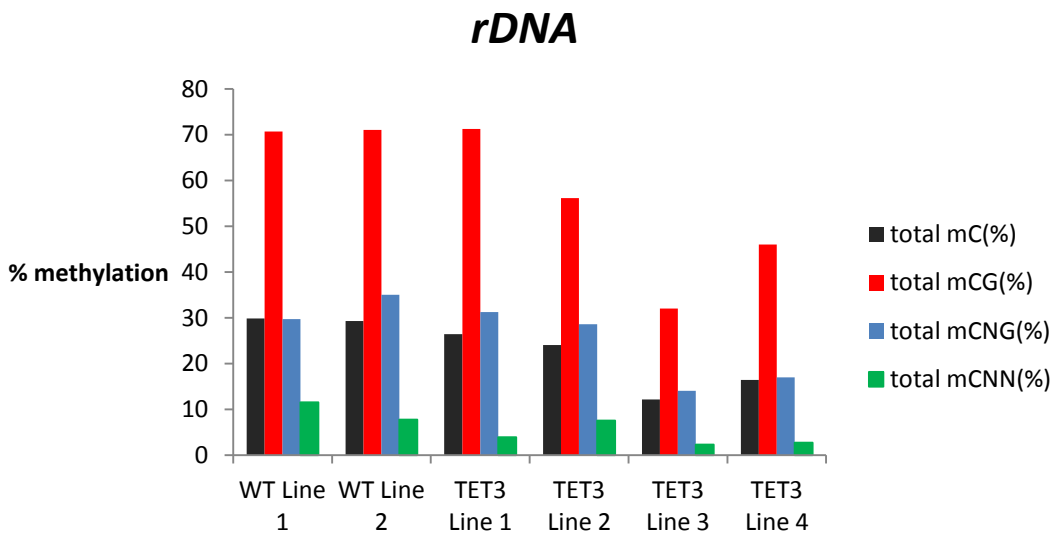
Figure 5.2. *Arabidopsis* lines that express the 3' region of the human *TET3* gene. **A**. The figure shows the 35S *TET3* construct, which was produced as described in Sections 8.1.4.6 and 8.2.1.7.6 of Materials and Methods. Plasmid labels are as described in Figure 4.3. **B**. The figure shows a sqRT-PCR analysis of four *Arabidopsis* TET3 lines (T2) performed as described in Figure 4.4. *TET3* expression is shown in the top boxes. *ELONGATION FACTOR* (*EF*) gene expression was used as an internal control and "-ve" refers to a water only sample.

5.2.2. Analysing DNA demethylation in *Arabidopsis* plants that express the 3' region of the *TET3* gene

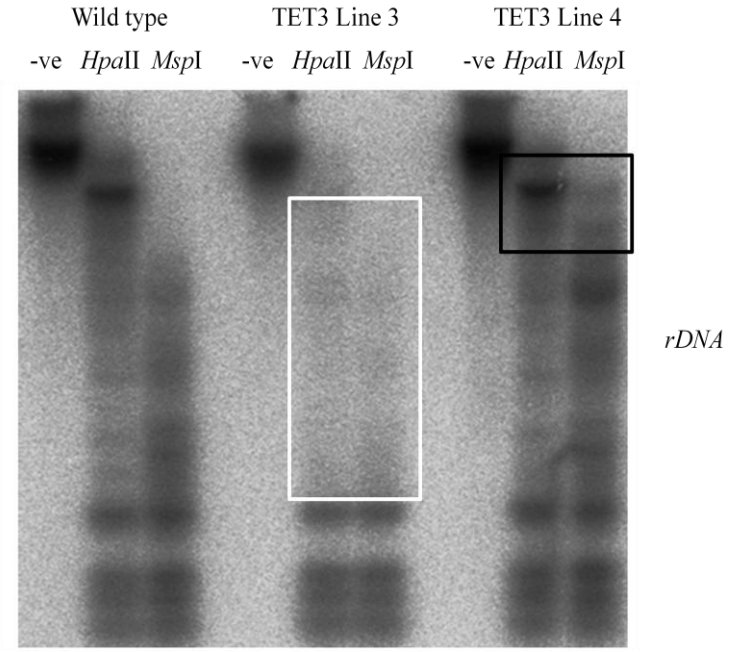
After identifying *Arabidopsis* lines that expressed the 3' region of the *TET3* gene, potential target regions were analysed for DNA methylation changes by bisulfite sequencing 10 clones amplified for different putative target regions from the four *Arabidopsis* lines. DNA methylation levels were calculated as the percentage of methylated cytosine at individual sequence types. To follow up stability of changes any targets that showed a significant change in DNA methylation when analysed by bisulfite sequencing were subsequently analysed by methylation-sensitive Southern blot. *rDNA* is a known target of TET3 in mammalian cells (Guo et al, 2011) and was therefore the first target analysed in the TET3 *Arabidopsis* lines (Figure 5.3).

DNA methylation changes within the gene body of *rDNA* coding for the 18S subunit were found in three lines (T2) when analysed by bisulfite sequencing (Figure 5.3A). The largest changes occurred in TET3 lines 3 and 4. For this reason they were selected for methylation-sensitive Southern blot analysis (Figure 5.3B and Figure 5.3C). Methylation sensitivities for *Hpa*II and *Msp*II are as described in Sections 2.2.1 and 2.2.2. *BstU*I, *Eco*RII and *Mnl*I are CG, CNG and CNN methylation-sensitive, respectively, and will therefore only cleave their recognition sequence when the respective sequence type of DNA methylation is absent or lost. Methylation-sensitive Southern blot analysis confirmed demethylation in TET3 line 3 when compared to wild-type (white boxes, Figure 5.2B and white asterisks, Figure 5.3C). DNA demethylation is indicated by a weaker signal from higher-molecular-weight DNA. However, based on this logic, it appears that TET3 line 4 gains methylation when analysed by methylation-sensitive Southern blot (black box, Figure 5.3B). This contrast in data for TET3 line 4 raised questions with regards to the TET3 intermediates including hmC, fC and caC and their sensitivities to the methylation detection techniques.

A.


rDNA
% methylation
80
70
60
50
40
30
20
10
0
WT Line 1
WT Line 2
TET3 Line 1
TET3 Line 2
TET3 Line 3
TET3 Line 4
total mC(%)
total mCG(%)
total mCNG(%)
total mCNN(%)

B.


Wild type
TET3 Line 3
TET3 Line 4
-ve HpaII MspI
-ve HpaII MspI
-ve HpaII MspI
rDNA

C.

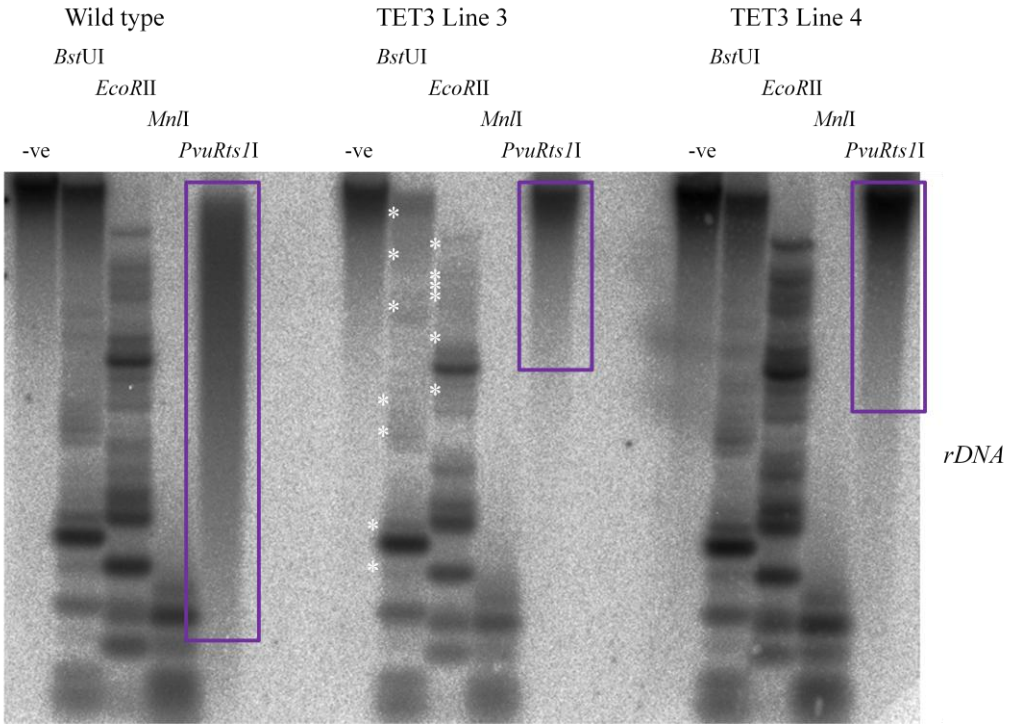
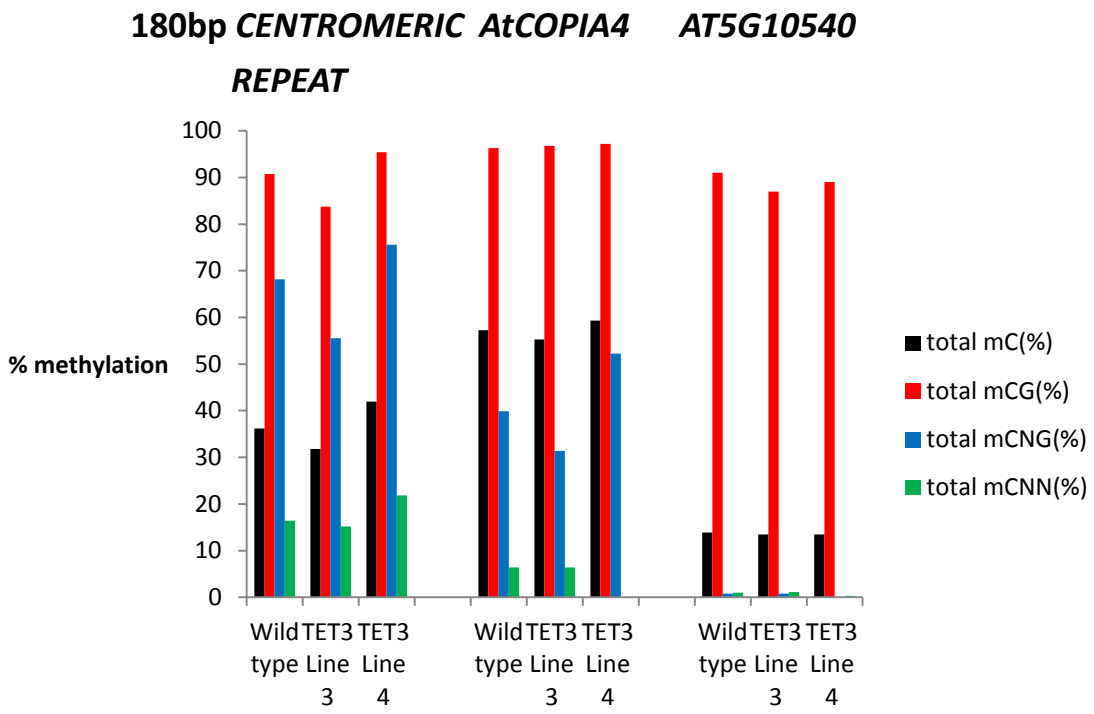


Figure 5.3. DNA methylation analysis at *rDNA* in TET3 *Arabidopsis* lines. **A**. The figure shows DNA methylation levels within a 310 bp region of *rDNA* among wild-types and *Arabidopsis* TET3 lines (T2) determined by bisulfite sequencing analysis as described in Sections 8.2.1.11 and 8.1.4.3 of Materials and Methods using DNA from basal rosette leaves. DNA methylation levels are represented by the Y-axis and each sequence type is identified with a unique colour indicated by the key on the right. **B**. The figure shows a *Hpa*II and *Msp*I methylation-sensitive Southern blot analysis of wild-type and *Arabidopsis* TET3 lines 3 and 4 (T2) using DNA from basal rosette leaves and an *rDNA* specific probe performed as described in Sections 8.2.1.13 and 8.1.4.4 of Materials and Methods. *Hpa*II and *Msp*I methylation sensitivities are described in Section 2.2.1 and Figure 2.4. DNA methylation changes, which are evident when the fingerprint varies from the wild type pattern, are highlighted by a white and black box for TET3 lines 3 and 4, respectively. The lane labelled "-ve" refers to undigested DNA and is used as a control for non-specific digestion or DNA degradation. **C**. The figure shows a *BstU*I, *EcoR*II *Mnl*I and *PvuRts1*I methylation-sensitive Southern blot analysis of wild-type and *Arabidopsis* TET3 lines 3 and 4 (T2) performed as described in Figure 5.3B. Some DNA methylation changes in TET3 line 3, which are evident when the fingerprint varies from the wild type pattern, are highlighted by white asterisks. Variation in *PvuRts1*I digestion among wild-type and TET3 lines is highlighted by purple boxes.

The identification of TET3-induced DNA methylation changes at *rDNA* in *Arabidopsis* prompted the analysis at other target loci known to contain DNA methylation. Bisulfite sequencing was used to analyse TET3-induced methylation changes at *180 bp CENTROMERIC REPEATS*, the retrotransposon *AtCOPIA4* and the gene body of *AT5G10540* (Figure 5.4). No DNA methylation changes could be detected at the regions analysed for *AtCOPIA4* or *AT5G10540* in lines 3 or 4 (T2) (Figure 5.4A). Very subtle changes were found at *180 bp CENTROMERIC REPEATS*,

which prompted a methylation-sensitive Southern blot analysis (Figure 5.4B) but no DNA methylation changes were detected.

A.



B.

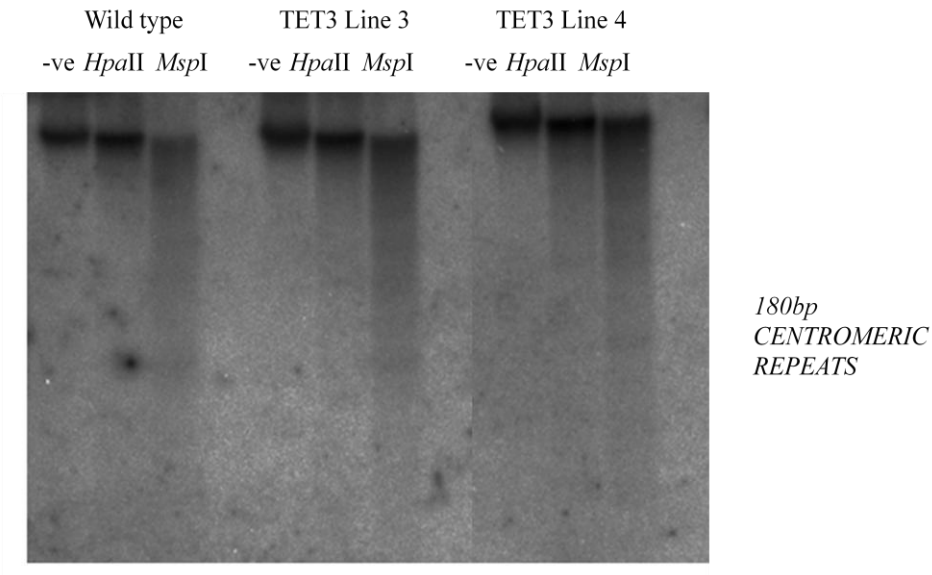


Figure 5.4. Analysing DNA methylation at target loci in *Arabidopsis* TET3 expressing lines. **A**. The figure shows DNA methylation levels within a region of *180 bp CENTROMERIC REPEATS, AtCOPIA4* and *AT5G10540* in TET3 expressing *Arabidopsis* lines (T2). DNA methylation levels were determined by bisulfite sequencing analysis as described in Sections 8.2.1.11 and 8.1.4.3 of Materials and Methods using DNA from basal rosette leaves. Methylation levels are represented by the Y-axis and each sequence type is identified with a unique colour indicated by the key on the right. **B**. The figure shows a methylation-sensitive Southern blot analysis of *180 bp CENTROMERIC REPEATS* in wild-type and TET3 expressing *Arabidopsis* lines performed as described in Figure 5.3B.

5.2.3. Analysing the stability of TET3-induced changes in *Arabidopsis*

To test if the TET3-induced DNA methylation changes at the *rDNA locus* were stable the methylation-sensitive Southern blot analysis was extended into the next generation (T3 lines, Figure 5.5). T3 TET3 lines were selected and at this stage of the analysis the zygosity of the T3 TET3 lines is unknown. A weaker signal is detected in T3 plants of TET3 line 3 compared to wild-type (white box, Figure 5.5). Signals from higher-molecular-weight DNA are observed again in TET3 line 4 T3 plants (Black box, Figure 5.5). These observations indicate that the TET3-induced changes in *Arabidopsis* are stable across one generation in the presence of the transgene.

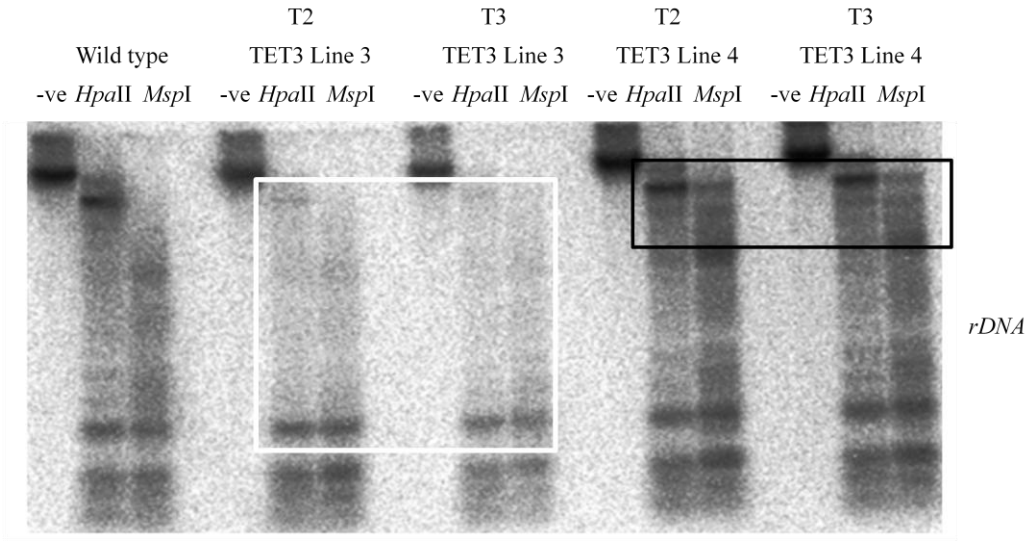


Figure 5.5. Methylation-sensitive Southern blot to analyse the stability of TET3-induced DNA methylation changes in *Arabidopsis*. The figure shows a *Hpa*II and *Msp*I methylation-sensitive Southern blot analysis of wild type and T2 and T3 TET3 lines 3 and 4 using an *rDNA* probe. The analysis was performed as described in Figure 5.3B. DNA methylation changes, which are evident when the fingerprint varies from the wild type pattern, are highlighted by a white and black box for TET3 lines 3 and 4, respectively

5.2.4. *rDNA* copy number and mutation analysis in *Arabidopsis* TET3 lines

Changes in DNA methylation are not the only potential cause of the observed differences in fragment sizes, as *rDNA* copy number changes in TET3 lines would also result in a different Southern blot fingerprint. To analyse if *rDNA* expansion was responsible for the differences in signal detection in TET3 lines, Southern blot analysis was repeated with the methylation insensitive restriction enzyme *Ase*I (Figure 5.6). There is little change in signal strength between TET3 lines and wild-type when

analysed by methylation-insensitive Southern blot using *180 bp CENTROMERIC REPEAT* and *rDNA* specific probes. Therefore, *rDNA* copy number does not vary at a level which is detectable by Southern blot analysis between TET3 lines and wild-type. It is therefore unlikely that the differences seen in the methylation-sensitive Southern blots are due to changes in *rDNA* copy number.

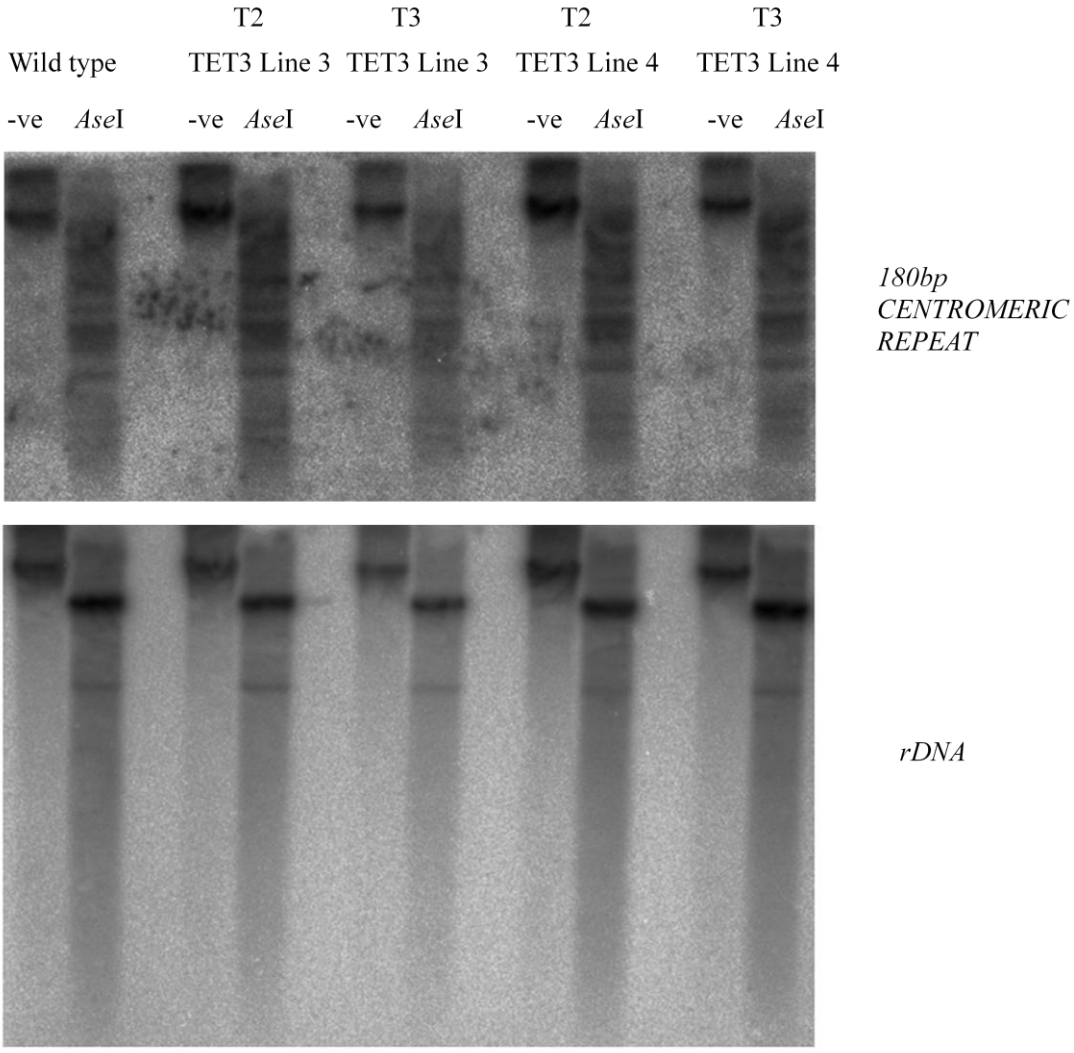


Figure 5.6. A Southern blot to analyse *rDNA* expansion in *Arabidopsis* TET3 lines. The figure shows DNA from wild-type and TET3 lines 3 and 4 analysed by Southern blot performed as described in Figure 5.3B using the methylation-insensitive restriction enzyme *Ase*I. Both T2 and T3 generations were included for each TET line. The blot was analysed using *180bp CENTREMERIC REPEAT* (top panel) and *rDNA* (bottom panel) probes amplified with primers provided in Section 8.1.4.4 of Materials and Methods.

5.2.5. Analysing hmC, fC and caC intermediates among *Arabidopsis* TET3 lines

After analysing DNA demethylation at *rDNA* in *Arabidopsis* TET3 lines it was logical to question if any of the TET3 intermediates could be detected. *rDNA* analysis by Southern blot was repeated with the restriction enzyme *PvuRts1*I, which specifically cleaves DNA that contains hmC (purple boxed regions, Figure 5.3C). *PvuRts1*I appears to cleave *rDNA* in wild-type but not *rDNA* in TET3 lines 3 or 4 (T2). This indicates that wild-type *Arabidopsis rDNA* contains hmC, whereas TET3 lines 3 and 4 either lack hmC or have low levels undetectable by Southern blot. hmC in TET3 lines could be processed to other intermediates preventing digestion by *PvuRts1*I. *PvuRts1*I, however, is a novel restriction enzyme and caution should be taken when interpreting this data.

5.3. Discussion

In this chapter, the effects of expressing the 3' region of the human *TET3* gene coding for the catalytic domain and NLS were analysed in *Arabidopsis*. The most obvious effects were found at *rDNA* in TET3 line 3. Both bisulfite sequencing and methylation-sensitive Southern blot indicate a reduction in DNA methylation levels at CG and CNG sequence contexts. Interestingly, DNA demethylation is induced at *rDNA* when over-expressing TET1 in mammalian human embryonic kidney (HEK) 293 cells (Guo et al, 2011). However, no DNA methylation changes were found in *Arabidopsis* at the retroelement *AtCOPIA4* or the gene body of *AT5G10540*. Similarly, LINE-1 retroelements are unaffected when TET1 is over-expressed in HEK293 cells (Guo et al, 2011). These data are not sufficient to suggest that *rDNA* is the only region that responds in this manner to TET3 activity in *Arabidopsis*. To identify all regions would require large scale bisulfite sequencing. However, it is unlikely that even large scale bisulfite sequencing would reflect the full extent of TET3 activity because TET3 proteins have to compete with endogenous epigenetic modifiers. Therefore, TET3 activity could be compensated for by DNA methyltransferase activity. Regions under strict epigenetic control such as those which produce siRNAs may be extremely difficult to detect TET3 activity because remethylation would follow demethylation. Another model which could explain different target sensitivities is that genomic regions differ in their accessibility to the

TET3 enzyme. Heterochromatin regions may be less accessible to TET proteins than highly transcribed regions, such as *rDNA*.

Demethylation identified only at the *rDNA locus* suggests that *rDNA* may be particularly sensitive to stable DNA methylation changes or TET3 activity. The stability of DNA methylation changes induced by TET3 at the *rDNA locus* is reinforced by the conservation of the detected DNA methylation pattern in T2 and T3 lines. Stable changes may be due to the continuous presence of TET3 or the lack of a remethylation mechanism that restores TET3-induced demethylation effects. The lack of a remethylation mechanism at the *rDNA locus* was also observed when *rDNA* demethylation was not restored in a *ddm1* hypomethylated background segregated back into wild-type (Kakutani et al, 1999). To confirm this, however, would require the removal of the transgene to test if remethylation occurs. At this stage it is only possible to speculate about the reason why *rDNA* would be particularly sensitive to TET3 activity. For example, epigenetic modifiers have been found localised to discrete foci within the nucleus (Zheng et al, 2008). If TET3 was localised at *rDNA* then this could make *rDNA* a hotspot for DNA methylation changes. A similar outcome would occur if *rDNA* was more accessible to TET3 than other genomic loci.

Identifying TET3 changes only at the *rDNA locus* suggests that demethylation occurred or was stably maintained at specific loci only. This led to the question if TET3 could be more efficiently targeted to other genomic loci if it was linked to specific target domains. TET3 fusion constructs have now been produced with help from visiting internship student Lennard Ganß, Ruprecht-Karls-University of Heidelberg. *TET3* has been fused with the region that encodes the CHD domain of CHROMOMETHYLTRANSFERASE 3 (CMT3) and the MBD domain of the human METHYL-CpG BINDING DOMAIN PROTEIN 1 (MBD1). While, analysis of TET3-CHD and TET3-MBD lines was not within the project timeframe, 24 TET3-CHD and 14 TET3-MBD transformants have been produced and analysing these lines would be ideal for a new research candidate.

The reason for differences in demethylation levels detected among the TET3 lines by bisulfite sequencing is unknown. Only a 310 bp region of the 18S *rDNA* gene body was analysed by bisulfite sequencing and this may not provide a full representation of

the level of TET3 activity at *rDNA*. Within the 5' end of the TET3 gene is the sequence which encodes the CXXC domain and it is critical for TET3 targeting (Xu et al, 2012). The region which encodes the CXXC domain was excluded from the 35S TET3 construct design. Removing this specificity could lead to random activity, which may explain the difference in methylation levels between the TET3 lines.

A surprising result was the detection of demethylation in TET3 line 4 by bisulfite sequencing analysis but an increased sensitivity to methylation-sensitive restriction enzymes when analysed by Southern blot. It would be logical if a level of demethylation had to be reached before changes were detectable by Southern blot analysis. However, the fact that this increased sensitivity is stronger than wild-type raised questions with regards to TET3 intermediates and their detection with current techniques. No studies have analysed *Hpa*II methylation sensitivities to TET3 intermediates but studies have investigated the isoschizomer *Msp*I. *Msp*I will cleave DNA when the sequence CCGG is both unmethylated and methylated at the internal C, $C^{m}CGG$. It will not cleave the sequence when fC or caC are present (Ito et al, 2011). Bisulfite sequencing analysis does not convert mC or hmC (Yu et al, 2012) but does convert caC (He et al, 2011) and fC (Yu et al, 2012) to uracil. These limitations prevent distinguishing between mC or hmC and C, caC or fC. These different sensitivities offer an explanation to the conflicting TET3 line 4 bisulfite sequencing and Southern blot data. If the *rDNA* region analysed in TET3 line 4 contained fC and caC intermediates then they would be converted by sodium bisulfite. Conventionally this would then be interpreted as unmethylated DNA. However, fC and caC intermediates would prevent cleavage by methylation-sensitive restriction enzymes, which would then conventionally be interpreted as methylated DNA.

To follow up the analysis of TET3 intermediates, the other TET proteins could be expressed in *Arabidopsis*. TET1 and TET2 have stronger oxidative capacities compared to TET3 *in vitro* (Ito et al, 2011). If these capacities were maintained in *Arabidopsis*, then higher intermediate or demethylation levels may be the result, which would be easier to detect with current methods. However, strong hypomethylation in *Arabidopsis* causes phenotypic abnormalities, including reduced fertility (Bartee & Bender, 2001). This would make lines more difficult to work with and was the reason why TET3 was initially chosen for expression in *Arabidopsis*. To

examine the demethylation mechanisms the TET proteins could be expressed in an *Arabidopsis* glycosylase mutant. This would preferably be the *ros1;dml2;dml3* triple mutant because, while demethylation at some loci is controlled by single DNA glycosylases, some loci are controlled redundantly by multiple DNA glycosylases (Penterman et al, 2007). Expressing TET3 proteins in an *Arabidopsis* glycosylase mutant would provide a larger window of opportunity to detect TET3 intermediates, if the DNA glycosylases are responsible for excision of the oxidised bases. In the mutant oxidised bases would not be excised and therefore accumulate making detection with current methods more likely.

It would also be useful to express the TET3-MBD and TET3-CHD constructs in the glycosylase mutant and analyse the transformed wild-type lines. There is currently no known method to target DNA demethylation to specific loci, despite the benefits of such a system. In plants, targeting DNA methylation changes could create epigenetic variation at particular loci, which may avoid the phenotypic abnormalities produced by current global demethylation approaches. In mammals, it would be useful if hypermethylated tumor suppressor genes could be targeted for demethylation in cancer cells.

In future experiments it may be useful to include *Arabidopsis* lines transformed with mutated TET constructs that remove the catalytic function of the TET protein. This would help to determine the effects of TET proteins in plants that were directly caused by catalytic activities and would be particularly useful for TET-fusion constructs, when dominant negative effects could result from expressing MBD or CHD domains.

The data presented in this chapter indicate that the human TET3 protein is catalytically active in *Arabidopsis* and can be used to induce DNA methylation change at the *rDNA locus*. It also highlights the complexities and limitations of current detection techniques for the sixth, seventh and eighth bases. New techniques to reliably investigate these marks are emerging. For example, oxidative bisulfite sequencing (oxBS-Seq) uses potassium perruthenate which oxidises hmC to fC (Booth et al, 2012). fC can then be converted by sodium bisulfite to distinguish between mC and hmC. Despite these advances there is still no technique to

differentiate between fC and caC modifications *in vivo*. Future studies should develop reliable detection methods for these novel bases and create conditions which favour their detection, for example, via expressing TET proteins in glycosylase mutant backgrounds.

6.0. Generating epigenetic diversity in tomato by inactivating the tomato *DNA METHYLTRANSFERASE 1*

6.1. Introduction

Traditional crop breeding strategies can require up to a decade to produce a new variety. Genetic-hybridisation methods require the generation of a pure line by multiple self pollinations. Pure lines with desirable phenotypes can then be crossed (Rommens et al, 2007). Crop breeding has been assisted by advancements in technology, such as marker assisted selection (MAS). MAS allows specific traits to be followed in breeding programmes, using DNA amplification and sequencing techniques to identify single nucleotide polymorphisms (SNPs) associated with a trait (Tester & Langridge, 2010). This is particularly useful for traits that are recessive, do not display a visible phenotype or when multiple genes are required for a desired characteristic (Tester & Langridge, 2010). Traditional crop breeding strategies have limitations. Breeding can only be carried out between plants that can mate sexually with each other. When plants are crossed unwanted traits may be introduced and pure lines have to be constantly maintained to allow F1 lines to be generated (Rommens et al, 2007).

Agronomics and breeding techniques have linearly increased global crop yield by an average of 32 million metric tons per year since 1961 (Tester & Langridge, 2010). The Declaration of the World Summit on Food Security target requires 70% more food by 2050 (Tester & Langridge, 2010). To meet this target the average annual increase in yield has to rise from 32 to 44 million metric tons. This target becomes even more challenging with unpredictable global environment change and reducing arable land (Tester & Langridge, 2010).

Tradition breeding strategies rely on genetic changes but there are examples where stable changes to the expression of genes without altering the DNA sequence can contribute to variation. These examples are called epi-mutants because there has been a change in their epigenome. DNA methylation is the most frequent modification of plant DNA and changes to the DNA methylation patterns are often found among epi-mutants. A core feature of DNA methylation is transcriptional repression, either by

direct obstruction of transcriptional proteins or by serving as a target for specific proteins, which signal chromatin condensation (Klose & Bird, 2006). Examples of epi-mutants include, the *Arabidopsis superman* (*sup*) epi-mutant, which has *sup*-like flowers. These are flowers characterised by an increase in the number of stamens due to hypermethylation within the promoter and 5' coding region of the *sup* gene (Jacobsen & Meyerowitz, 1997; Kishimoto et al, 2001). A late flowering phenotype is displayed in the *Arabidopsis fwa* epi-mutant due to hypomethylation at direct repeats within the 5' region of the *fwa* gene (Soppe et al, 2000). Silencing of the *bonsai* gene in a *ddm1 Arabidopsis* mutant produces compact stems (Saze & Kakutani, 2007). A rice epi-mutant, *Epi-d1* has dwarfed tillers (vegetative branch shoots), assigned to hypermethylation at the *dwarf1* (*d1*) gene promoter (Miura et al, 2009). Hypermethylation at the *colourless non ripening* (*cnr*) locus in tomato inhibits fruit ripening (Zhong et al, 2013).

Inducing stable expression changes by interfering with epigenetic systems in commercially viable crops could be an additional strategy to create variation and reduce the burden on traditional breeding to meet future demands. Therefore, the aim of work described in this chapter was to inactivate the tomato maintenance methyltransferase, *DNA METHYLTRANSFERASE 1* (*tomMET1*) to examine if this would lead to enhanced epigenetic diversity.

6.2. Results

6.2.1. Knockdown of *tomMET1* in tomato using an inverted repeat cassette driven by the 35S CaMV promoter

A working method was described in Chapter 3 for the generation of stable epi-alleles in the model organism *Arabidopsis*. To apply the same strategy to tomato would require the production of a *tomMET1* knockout or knockdown line. A knockdown approach using an inverted repeat construct was chosen because a knockout line was not available. The first objective was to obtain the *tomMET1* gene sequence. This was acquired using the recently published tomato genome sequence browser (Sato et al, 2012). Then, a 609 bp region that encodes part of the *tomMET1* catalytic domain was amplified from tomato cDNA and inserted into the pHannibal inverted repeat vector

in opposite orientations. This produced a sense and antisense sequence separated by an intron, driven by the 35S promoter. The inverted repeat cassette was then sub-cloned into the plant transformation vector pGreen 0029 (Figure 6.1A). Tomato transformations were carried out in parallel using the 35S *tomMET1* inverted repeat vector and a pGreen 0029 empty vector control (Figure 6.1B). 1625 explants were cultured with the inverted repeat and no positive transformants could be regenerated in an antibiotic selection process. 535 explants were cultured with the empty vector and five transformants were regenerated in an antibiotic selection process (Figure 6.2). The lack of 35S *tomMET1* inverted repeat tomato transformants highlighted the possibility that tomato may be sensitive to the down-regulation of *tomMET1* by expressing the inverted repeat construct, at least during explant regeneration.

A. B.

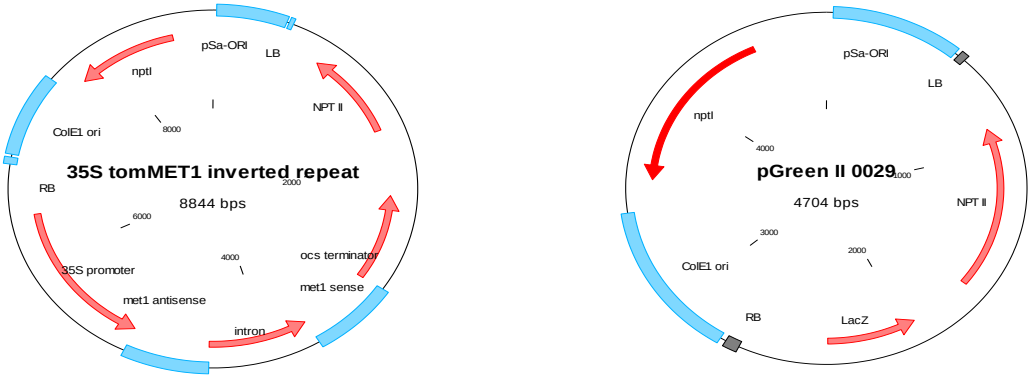


Figure 6.1. A map of the 35S inverted repeat constructs used for the knockdown of *tomMET1* in tomato. Labels and arrows are as described in Figure 4.3. **A**. The figure shows the 35S *tomMET1* inverted repeat construct produced as described in Section 8.2.1.7.3 of Materials and Methods. **B**. The figure shows a map of pGreen 0029 used as an empty-vector control.

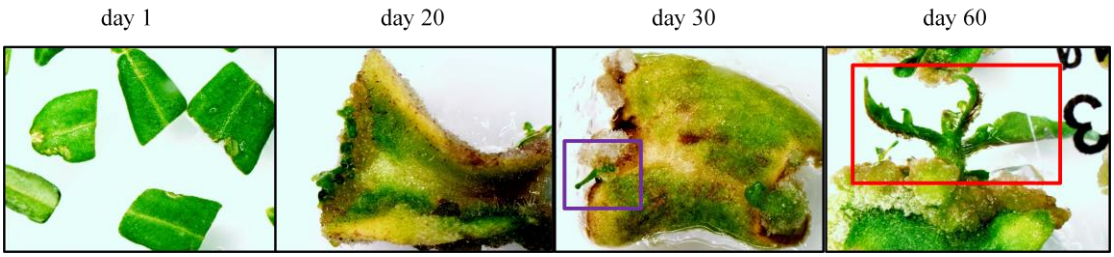


Figure 6.2. A photographic timeline showing the generation of tomato transformed with the pGreen 0029 empty vector. Tomato transformations were performed as described in Section 8.2.3.2 of Materials and Methods. On day one the explants are inoculated with *Agrobacterium.* Between days 20 and 30 leaf primordia begin to emerge highlighted by the purple box. Transgenic tomato leaf and stem structures appear after 60 days shown within the red box and require transfer onto rooting medium. Regeneration of transgenic lines was only possible for explants transformed with the empty vector control, pGreen 0029.

6.2.2. Knockdown of *tomMET1* in tomato using an inverted repeat cassette driven by a heat inducible promoter

The lack of transformants from the 35S *tomMET1* inverted repeat transformations prompted a change in strategy. A heat inducible (HI) promoter, which had previously been characterised in *Arabidopsis* (Gallois et al, 2002) was chosen to drive the *tomMET1* inverted repeat. This involved amplifying the HI promoter by PCR using primers to incorporate restriction enzyme sites. This enabled the replacement of the 35S promoter in the 35S *tomMET1* inverted repeat construct (Figure 6.3A). The activity and tissue specificity of the HI promoter was unknown in tomato. Therefore, a HI *β–GLUCURONIDASE (GUS)* reporter gene construct (provided by Dr Elena Zubko, the University of Leeds) was transferred in parallel into tomato and *Arabidopsis* (Figure 6.3B). Tomato transformants could be generated for both constructs. 630 explants were cultured with the HI *GUS* construct and 11 transformants were regenerated. 1100 explants were cultured with the HI *tomMET1* inverted repeat construct and 5 transformants were regenerated.

A. B.

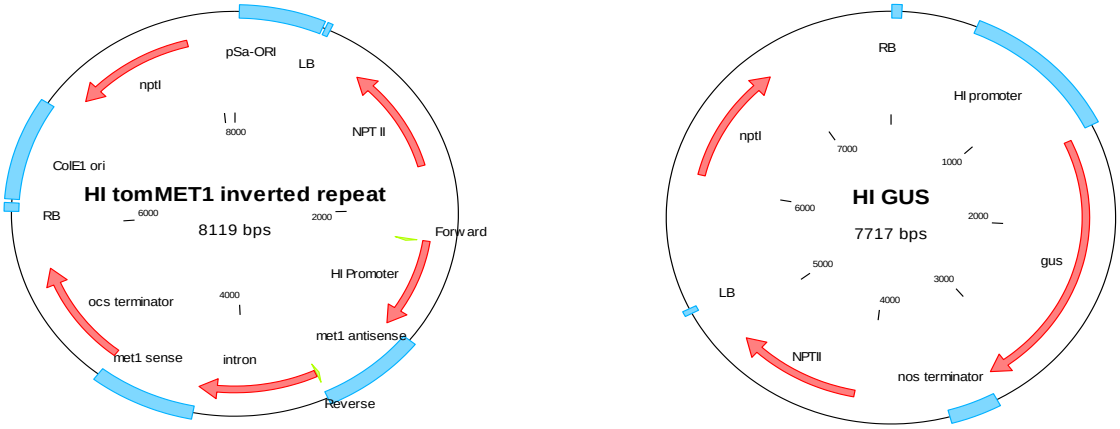


Figure 6.3. Maps of the HI *tomMET1* inverted repeat and HI *GUS* constructs. Labels and arrows are as described in Figure 4.3. **A**. The figure shows a map of the HI *tomMET1* inverted repeat construct, which was produced as described in Section 8.2.1.7.4 of Materials and Methods. **B**. The figure shows a map of the HI *GUS* construct provided by Dr Elena Zubko, the University of Leeds.

Prior to analysing the HI inverted repeat transformants the activity of the HI promoter was characterised using the HI *GUS* reporter transformants. *Arabidopsis* and tomato HI *GUS* transformants were grown on MS30 medium with selection. Positive transformants were exposed to 30, 35 and 40^{o}C overnight and immediately analysed for GUS activity. This temperature range was selected because it includes 38^{o}C, a temperature at which promoter induction was previously found in *Arabidopsis* (Gallois et al, 2002). At 40^{o}C, the blue GUS product was detected in the root tips, some lateral roots and cotyledon leaf tips of tomato and in roots, root hairs and cotyledons of *Arabidopsis* (Figure 6.4). This data demonstrates that the HI promoter is active in both tomato and *Arabidopsis*.

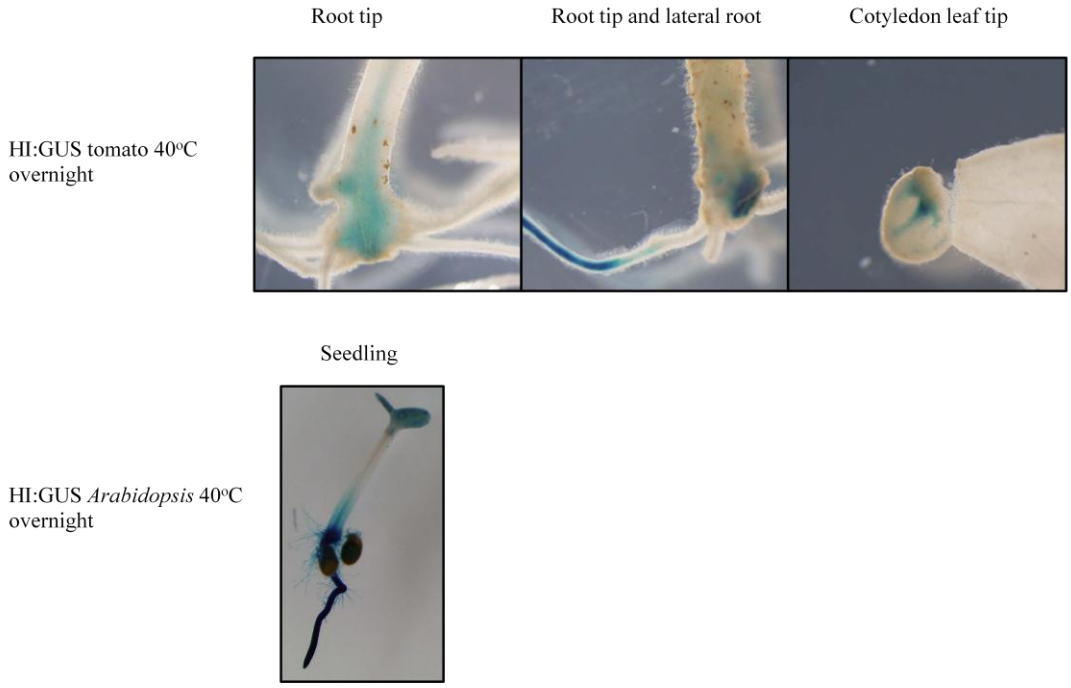


Figure 6.4. GUS histochemical analysis of the HI *GUS* tomato and *Arabidopsis* transformants. From left to right, the top panels show root tips, lateral roots and cotyledons of 2 week old tomato HI *GUS* transformants after an overnight incubation at 40°C and subsequent GUS staining performed as described in Section 8.2.4 of Materials and Methods. The bottom panel shows a 2 week old *Arabidopsis* HI *GUS* transformant after an overnight incubation at 40°C and subsequent GUS staining performed as described in Section 8.2.4 of Materials and Methods.

Once it had been confirmed that the HI promoter could drive expression of the *GUS* gene in tomato, the HI *tomMET1* inverted repeat transformants were analysed using the same strategy. First, 5 transformants were confirmed using PCR to amplify a region of the transgene. A forward primer that binds to the HI promoter and a reverse primer that binds to the *tomMET1* antisense insert were used (Figure 6.5).

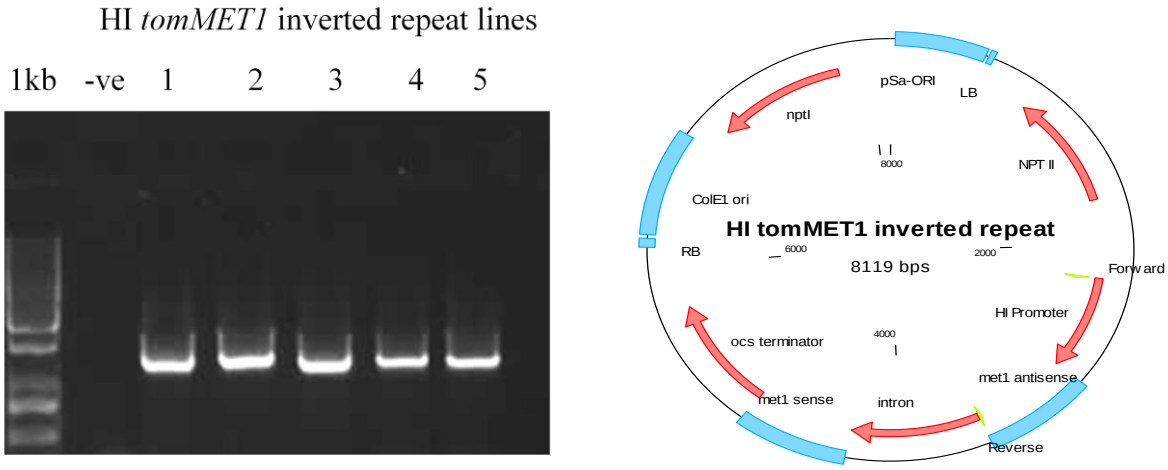


Figure 6.5. Analysing transformants using PCR to amplify a region of the HI *tomMET1* inverted repeat transgene. On the left is a PCR analysed using an ethidium bromide stained agaorse gel performed as described in Section 8.2.1.3 using primers from Section 8.1.4.2 of Materials and Methods. HI *tomMET1* inverted repeat transformant numbers are provided above each lane alongside the 1kb DNA marker (Bioline) and a water only control (-ve) on the left. On the right is a map of the HI *tomMET1* inverted repeat construct. Labels are as described in Figure 4.3. The forward primer binds to the 5' end of the HI promoter and the reverse primer binds to the 3' end of the antisense insert. Primer positions are shown on the diagram as green arrows labelled forward and reverse.

After analysing a region of the transgene by PCR, the transformants could be analysed for heat induction of the inverted repeat and down-regulation of *tomMET1*. The transformants were exposed to 40^{o}C overnight. The expression of the *tomMET1* inverted repeat could be detected at 24 hours after the heat shock (Figure 6.6) using the same primers for detecting the presence of the transgene (Figure 6.5) but *tomMET1* levels were not down-regulated (Figure 6.6). Subsequent samplings after 44 and 54 hours were scheduled but after 44 hours tissue damage prevented sampling (Figure 6.7).

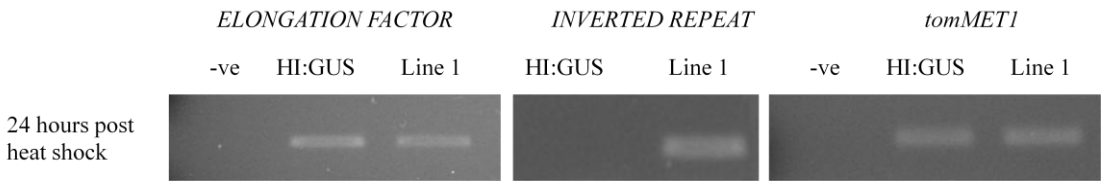


Figure 6.6. Analysing the expression of the inverted repeat and endogenous *tomMET1* in the HI *tomMET1* inverted repeat tomato transformant line 1. The figure shows a sqRT-PCR analysis using RNA from 2 week old tomato seedlings and an ethidium bromide stained agarose gel. The time of RNA sampling after heat shock is provided on the left of the image. Expression of *ELONGATION FACTOR, INVERTED REPEAT* and *tomMET1* in a HI GUS line and HI *tomMET1* inverted repeat line1 is indicated above the gel lanes. A water only sample was included as a negative control (-ve).

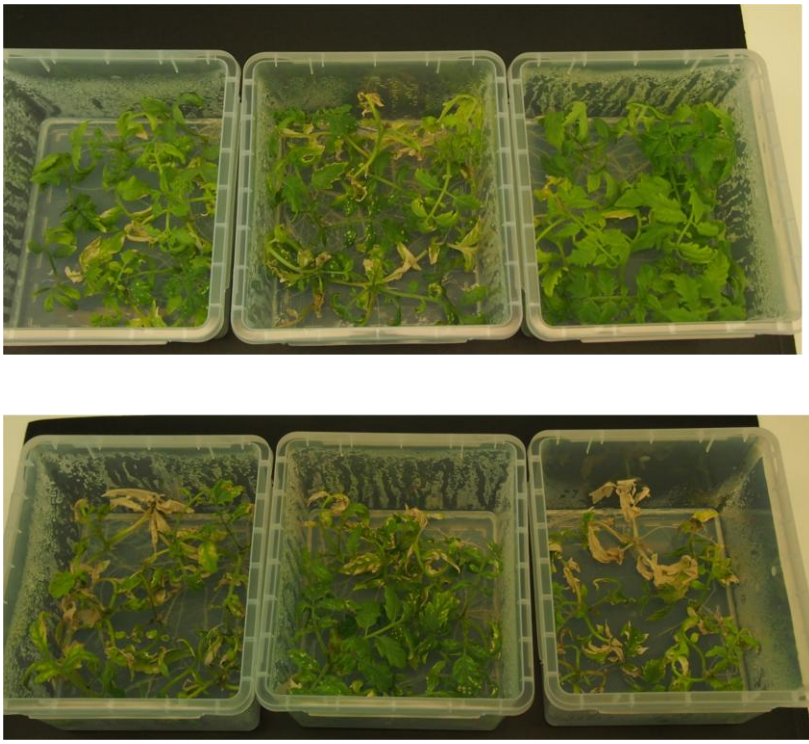


Figure 6.7. Tomato transformants 48 hours after heat treatment. The top and bottom panels respectively show HI GUS and HI *tomMET1* inverted repeat tomato transformants 44 hours after an overnight 40^{o}C heat exposure.

The approach needed to be optimised to avoid heat induced tissue damage. This would allow *tomMET1* expression analysis at increasing time intervals after heat exposure. In an attempt to avoid tissue damage the 40^{o}C exposure period was reduced to five and ten hours. These conditions allowed sampling at 15, 44 and 54 hours but they were not sufficient to induce the expression of the transgene in three of the *tomMET1* inverted repeat lines analysed (Figure 6.8).

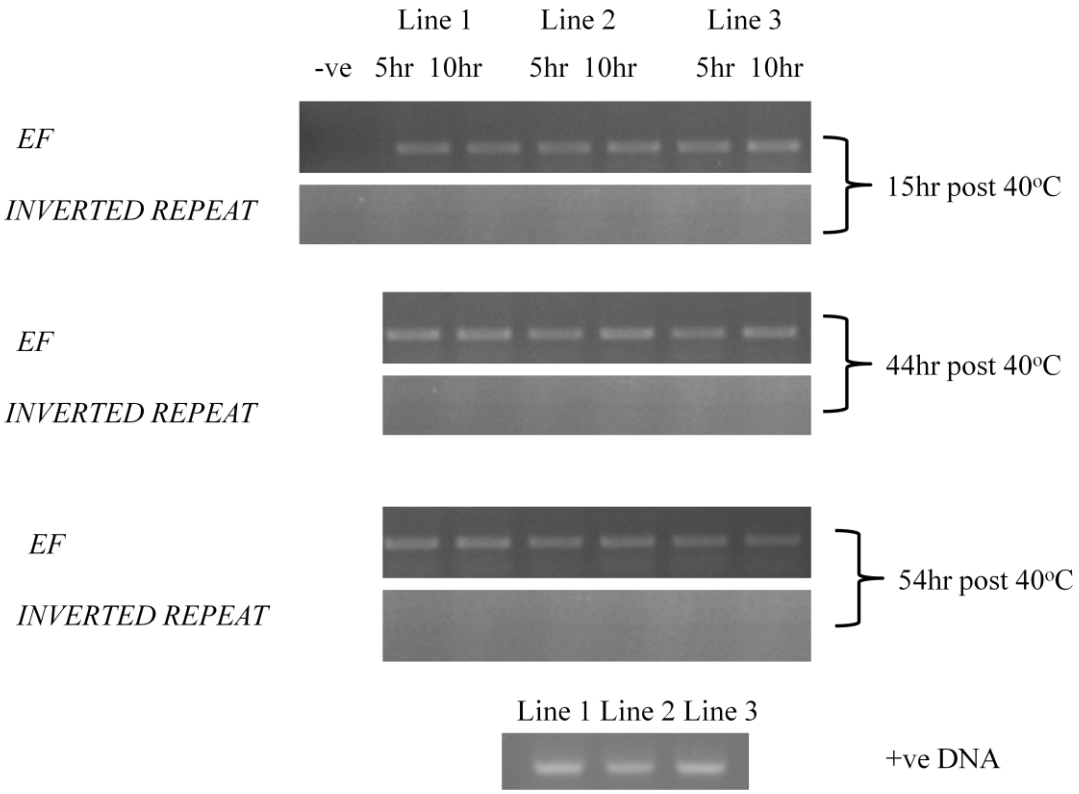


Figure 6.8. Analysing the expression of the inverted repeat in the HI *tomMET1* inverted repeat tomato transformants. The figure shows a sqRT-PCR analysis using an ethidium bromide stained agarose gel, and RNA harvested at 15, 44 and 54 hours from HI *tomMET1* inverted repeat lines 1, 2 and 3 after 40^{o}C heat exposure periods of five and ten hour. Sampling times are indicated on the right and the HI *tomMET1* inverted repeat line and respective heat exposure periods are indicated above each gel lane. *ELONGATION FACTOR* (*EF*) was used as a reference gene and DNA isolated from HI *tomMET1* inverted repeat lines 1, 2 and 3 was used as a positive control (+ve DNA) for *INVERTED REPEAT* amplification. A water only sample was used as a negative control (-ve).

6.3. Discussion

In this chapter, the inactivation of the tomato maintenance methyltransferase tomMET1 was attempted using an inverted repeat cassette. The initial strategy used the 35S CaMV promoter to drive inverted repeat expression. Using this strategy it was not possible to regenerate transformed tomato, despite repeated attempts and successful regeneration of empty vector control lines in parallel. The lack of 35S *tomMET1* inverted repeat tomato transformants prompted the speculation that inactivating *tomMET1* in tomato may have extreme consequences for viability, at least during explant regeneration. Such an extreme outcome was not anticipated because knockout of *met1* by mutagenesis has been carried out successfully in *Arabidopsis* (Kankel et al, 2003) and knockdown of *met1* by RNA interference (RNAi) (Chen et al, 2008; Kim et al, 2008; Oh et al, 2009) has been carried out successfully in both *Arabidopsis* (Chen et al, 2008; Kim et al, 2008) and tobacco (Oh et al, 2009). Tobacco was also transformed by explant inoculation and regeneration (Oh et al, 2009) similar to the procedure used in this thesis for tomato (Rai et al, 2012), which would suggest tobacco can tolerate *met1* inactivation during explant regeneration, unlike tomato. However, *Arabidopsis* and tobacco do display stunted growth and abnormal flower morphology due to DNA hypomethylation (Kim et al, 2008; Oh et al, 2009), highlighting the fact that DNA methylation changes in plants do have developmental consequences.

Lethal effects of interfering with DNA methyltransferases have been documented in mammals. For example, mouse *dnmt1* (Li et al, 1992), *dmnt3a* and *dnmt3b* (Okano et al, 1999) mutants are not viable and even quantitative changes can have detrimental effects (Loughery et al, 2011). One potential cause of lethality was identified when an inverted repeat targeting *DNA METHYLTRANSFERASE 1* (*DNMT1*) activated apoptosis in human fibroblast cells (Loughery et al, 2011). This was assigned to a drop in steady state protein levels for key mismatch repair (MMR) components, resulting in elevated mutations at microsatellite repeats, including those which are unmethylated (Loughery et al, 2011). An effective inducible strategy will allow any phenotypes of the *tomMET1* inverted repeat in tomato to be characterised at specific developmental stages.

Inverted repeats can target non-specific RNA (RNAi off-targeting) if their sequence is also homologous to other sequences in addition to the target. RNAi off-targeting can result in translation blocking or cleavage of non-specific targets. It can complicate the interpretation of phenotypic effects and result in toxicity (Jackson & Linsley). When all full length cDNAs in the *Arabidopsis* genome are used as initiators of RNAi, it was found that an average of 68.7% of transcripts could potentially silence 3.9 off-target genes (Xu et al, 2006). The *tomMET1* inverted repeat sequence was screened against the tomato genome (Sato et al, 2012; Xu et al, 2006) and no homology with any non-specific targets was found. It is therefore unlikely that the *tomMET1* inverted repeat has off-target effects, which are responsible for the phenotype. Using a different region of the *tomMET1* cDNA in the inverted repeat construct and observing the same phenotypes would, however, increase confidence in siRNA specificity.

A heat inducible promoter was chosen for the controlled induction of the *tomMET1* inverted repeat. Using this promoter allowed activation of the transgene with an overnight 40^{o}C heat treatment. However, 24 hours after induction *tomMET1* levels were unchanged. At this stage the competence of the inverted repeat could be questioned but the inability to amplify the inverted repeat cassette in one amplicon using PCR from a plasmid template is a good indicator of hairpin formation. It is more likely that a 24 hour period is not sufficient to induce silencing as it can take up to 48 hours for siRNAs to degrade target mRNA, especially for transcripts with high, stable expression levels (Hahn et al, 2004). Tissue damage appeared 48 hours after heat exposure preventing sampling and subsequent *tomMET1* expression analysis. Shorter heat shock periods of five and ten hours were insufficient to induce detectable inverted repeat expression. It could be argued that instability of the primary transcript would prevent detection by RT-PCR, but the ability to detect inverted repeat expression using an overnight 40^{o}C heat treatment would agrue against this. A Northern blot to analyse siRNAs produced by the inverted repeat would be an alternative strategy to analyse inverted repeat expression, and could be performed when an effective induction strategy, which does not cause tissue damage after induction, is developed to allow *tomMET1* expression levels to be analysed over an extended time course.

The activity of the HI promoter was analysed using the *GUS* reporter gene in *Arabidopsis* and tomato. GUS activity was detected after a 40^{o}C overnight incubation in *Arabidopsis* roots, root hairs and cotyledons and in tomato root tips, some lateral roots and cotyledon leaf tips. This demonstrates that the HI promoter is active in both *Arabidopsis* (Gallois et al, 2002) and tomato in specific locations after an overnight 40^{o}C heat exposure. HI promoter activity in other locations in which GUS product was not detected, such as the stem, cannot be used as an indicator for null activity. HI promoter activity in these areas may be exceptionally weak preventing GUS product detection or it may be more difficult to infiltrate with substrate. RNA isolated from null expressing tissue could be analysed for *GUS* expression by PCR based methods after heat treatment.

On reflection, the 35S *tomMET1* inverted repeat strategy created an unpredictable outcome and the heat inducible system caused tissue damage at a temperature which activated the promoter. The outlook for this project is now to develop and test a system for controlled induction of the *tomMET1* inverted repeat that avoids tissue damage complications in the heat inducible approach. To test if the use of an inducible promoter which does not involve increased temperatures is less damaging, knockdown of *tomMET1* will be attempted using the same inverted repeat cassette under the control of an alcohol inducible (AI) promoter (Caddick et al, 1998; Roslan et al, 2001). While an alcohol inducible system should be less damaging than the heat inducible system, it is not without its side effects. For example, ethanol treatment of tomato inhibits fruit ripening (Kelly & Saltveit, 1988), however, this should not prevent molecular analysis. Targeting *tomMET1* will remain the focus, as this aims to create genome wide epigenetic changes and such large scale changes provide capacity for scientific discovery. If the strategy is successful, *tomMET1* knockdown lines will be phenotypically characterised, which may require large scale expression profiling and bisulfite sequencing.

7.0. General Discussion

Inducing stable DNA methylation changes that transcend into expression changes could be a source of phenotypic variation in plants. In this study, four strategies have been developed and tested for their capacity to induce stable DNA methylation and expression changes.

7.1. Chemical induction of DNA methylation changes

Zebularine has a half life of approximately 508 hours and effectively demethylates the *Arabidopsis* genome (Baubec et al, 2009). These two properties highlighted the chemical as an ideal candidate to induce DNA methylation changes in commercial crops. Tomato was appealing due to the recent completion of the genome sequencing project (Sato et al, 2012) and the stable epigenetic change demonstrated by the tomato *cnr* mutant (Manning et al, 2006).

Treatment of tomato with zebularine induced DNA methylation changes within *rDNA* in two independent repetitions. However, while these changes were somatically heritable they reverted in subsequent generations. These data support an epigenetic reprogramming mechanism that functions in the germline or during early embryogenesis of tomato. Unlike mammals, there is currently no clear evidence in plants for an extensive global epigenetic reprogramming in the developing embryo. There is, however, an example for demethylation of an imprinted gene in maize. The *MATERNALLY EXPRESSED IN EMBRYO 1 (MEE1)* maternal allele is demethylated during fertilisation and methylated later in embryogenesis (Jahnke & Scholten, 2009), which suggests that some plants have evolved a mechanism for resetting of epigenetic marks at imprinted loci. Considering that *rDNA* methylation is not dictated by a maternal or paternal imprint, it will be interesting to determine whether epigenetic reprogramming in tomato is sequence specific or a global event, as in mammals.

7.2. MET1 inactivation

If DNA methylation changes are to be a source of stable variation in plants, then they must be heritable across generations. Therefore, while chemical treatment was

suitable for the induction of somatically heritable changes, a strategy was required to induce trans-generational changes. The detection of methylated loci unaffected by zebularine treatment, highlighted the possibility that trans-generational changes required minimal threshold levels. To maximise DNA methylation loss genetic strategies were developed to inactivate MET1. The strategy was first tested in *Arabidopsis* and then applied to tomato.

One surprising observation was the different sensitivity of *Arabidopsis* and tomato to MET1 inactivation. While the *Arabidopsis met1* mutant was viable and produced viable offspring, it was not possible to produce tomato lines transformed with a constitutively expressed *MET1* inverted repeat. The sensitivity of tomato to the MET1 inverted repeat was unanticipated since both tobacco (Oh et al, 2009) and rice (Teerawanichpan et al, 2004) have been successfully transformed with inverted repeats to target their MET1 homologues. However, the requirement of certain epigenetic modifiers is critical at early developmental stages in some plants. In rice the requirement of the epigenetic modifier ROS1a appears indispensible for viable offspring, as the progeny of a *ros1a* heterozygous mutant contains no homozygous mutant lines, due to embryonic lethalility (Ono et al, 2012). It is therefore tempting to speculate that MET1 may have an essential role in the regeneration of tomato explants.

Alternatively inactivating MET1 in tomato could be lethal. Lethal effects of inactivating DNA methyltransferases have been documented in mammals. For example mouse *dnmt1* (Li et al, 1992), *dmnt3a* and *dnmt3b* (Okano et al, 1999) mutants are not viable. One potential cause of lethality was identified when an inverted repeat targeting *DNA METHYLTRANSFERASE 1* (*DNMT1*) activated apoptosis in human fibroblast cells (Loughery et al, 2011). This was assigned to a drop in steady state protein levels for key mismatch repair (MMR) components, resulting in elevated mutations at microsatellite repeats, including those which are unmethylated (Loughery et al, 2011). Therefore, if MET1 inactivation is lethal in tomato, secondary functions of MET1 as well as DNA methylation changes should be considered.

7.3. Novel MET1 targets

A search for novel epigenetically regulated target genes in the *met1 Arabidopsis* mutant identified a *ncRNA locus*, which has several surprising features with regards to its epigenetic regulation. While loss of CG methylation at this locus in *met1* was expected as MET1 is the CG maintenance methyltransferase, a reduction in non-CG methylation was unanticipated. At the *ncRNA locus* CG methylation is reduced by 71% in *met1* but in addition CHG and CHH methylation levels are reduced by 60% and 78%, respectively. This suggests that the separation between CG and non-CG methylation pathways is lost at this locus and that MET1 is required for cytosine methylation in all sequence contexts. There are other reports of non-CG methylation loss in *met1*. For example, CHG methylation is modestly reduced at the *SADHU1-3* retroelement (Rangwala & Richards, 2007) and non-CG methylation is lost at a *REPETITIVE PETUNIA SEQUENCE (RPS)* element when introduced into *met1* (Singh et al, 2008). The *ncRNA locus*, however, is the first example of an endogenous gene to show drastic losses of both CHG and CHH methylation in *met1*, and could therefore become a useful model gene to investigate the mechanisms responsible for this unusual DNA methylation pattern.

The epigenetic changes at the *ncRNA locus* were stably inherited over three generations, proving that epigenetic modifier mutants can be exploited to induce heritable epigenetic variation in plants. This was not consistent among all the target genes analysed. The *FWA* gene, which is activated in *met1*, reverts when a wild-type *MET1* allele is restored by a genetic cross. The mechanisms which distinguish target gene stability are unknown but it is tempting to speculate that target genes under strict epigenetic control are less likely to stably change their DNA methylation profile. This may especially apply to targets that are associated with siRNAs, which function as part of the RdDM pathway. As soon as a functional *MET1* allele is restored methylation woud be restored and gene expression switched ‘off’ (Teixeira et al, 2009). This speculation is encouraged as two direct repeats within the 5’ coding region of the *FWA* gene are targets for RdDM when the locus is introduced into wild-type as a transgene (Chan et al, 2006).

The stable nature of the *ncRNA* epi-alleles was again demonstrated by the silent methylated *Columbia* allele and active *Wassilewskija* allele, which were stably transmitted through a *Columbia/Wassilewskija* hybrid. The inactive *Columbia* allele contains methylation throughout the promoter and transcribed regions, while methylation in the active *Wassilewskija* allele is restricted to the upstream promoter region. Only a small 107 bp promoter region of the *Wassilewskija ncRNA* allele remains unmethylated, which raises two possibilities. Either the small unmethylated promoter region is responsible for expression of the *ncRNA locus* or expression is influenced by methylation within the transcribed region. The observation that the *ncRNA* promoter-reporter transgene has a much lower expression level than the *ncRNA locus* in *met1* and *Wassilewskija*, argues in favour of a contribution from the transcribed region to the expression control of the locus, for example via hosting an enhancer that could be affected by methylation. Analysing the influence of methylation within the transcribed region of the *ncRNA locus*, would be a particularly interesting outlook considering that the role of gene body methylation is still speculative, and there are very few examples in which body methylation influences transcription. One of the few examples in which body methylation influences transcription are *CYSTEINE RICH PEPTIDASE (CRP)* genes. In comparison to the *ncRNA locus* in *Columbia*, *CRP* genes are methylated in leaves at CG and non-CG sequences within their transcribed regions (You et al, 2012). The methylation within the coding region of *CRP* genes likely contributes to gene silencing because its loss in synergid cells correlates with *CRP* gene expression (You et al, 2012).

7.4. *MET1* over-expression effects

If MET1 was restricted to a CG maintenance function then the predicted outcome of increasing MET1 levels would be maintenance of CG methylation. The loss of non-CG methylation at the *ncRNA locus* in *met1* suggests MET1 functions deviate from the classical CG maintenance model. To investigate novel MET1 functions both a catalytically active and inactive MET1 were over-expressed in *Arabidopsis*. In this study no direct DNA methylation changes were detected when over-expressing *MET1*. *RD29A* down-regulation when over-expressing the catalytically active MET1 is encouraging, as this could be an outcome of increased DNA methylation within the *RD29A* promoter. However, a complete analysis of the *RD29A* promoter is still

required to determine if DNA methylation is involved in *RD29A* down-regulation in *MET1* over-expression lines. In contrast to *RD29A*, some target genes were up-regulated. This was the case for *AT2G41380*, which was up-regulated in both MET1 catalytically inactive and active lines. *AT2G41380* codes for a protein methyltransferase that requires the same substrate as MET1. These observations formed the hypothesis that *AT2G41380* could be involved in a competitive feedback loop with MET1. Feedback mechanisms involved in the regulation of epigenetic modifiers have previously been demonstrated. For example, *ROS1* levels are down-regulated in a hypomethylated background presumably to prevent further loss of methylation (Mathieu et al, 2007). These observations highlight the potential complexity of feedback networks, which makes them interesting subjects for further analysis.

7.5. Active demethylation

The study of mammalian epigenetic modifiers is complicated by embryonic lethality caused by epigenetic modifier inactivation. The high tolerance of *Arabidopsis* to DNA methylation changes makes it an ideal candidate to investigate mammalian epigenetic modifiers. The fourth strategy therefore exploited *Arabidopsis* to study the mammalian TET3 methylcytosine dioxygenase. Intriguingly, methylation levels at a region within *AtCOPIA4*, *AT5G10540* and *180 bp CENTROMERIC REPEATS* remained stable, while a region within the *rDNA locus* was demethylated in *TET3*-expressing *Arabidopsis* lines. This implies that demethylation, similar to the conservation of hypomethylation described earlier, is locus-specific. The mechanisms which dictate selective demethylation within the *rDNA locus* and the stable conservation of methylation at other loci are unknown. It is important to acknowledge that the *TET3* gene is expressed in a wild-type background and therefore TET3 has to compete with endogenous methylation systems. The efficiency of endogenous methylation systems in reverting epigenetic changes is exemplified by *FWA* expression in the *met1* mutant, which is rapidly reset upon restoration of a functional *MET1* wild-type allele. A possible model could be that demethylation only affects loci that conserve epi-alleles such as the *rDNA locus.* At other regions, TET3 demethylation effects would be quickly erased by efficient re-methylation.

7.6. Outlook and open questions

7.6.1. Testing and exploiting the *ncRNA* promoter

Several aspects of this work require further study. It is still unclear, whether methylation within the promoter or the transcribed region influences expression of the *ncRNA locus*. To analyse whether methylation within the transcribed region has a role in expression of the *ncRNA locus*, an inverted repeat of the transcribed region could be introduced into *Arabidopsis Wassilewskija*. The inverted repeat would target DNA methylation via the RdDM pathway to the transcribed region. The inverted repeat, however, could also silence the *ncRNA locus* via post transcriptional silencing mechanisms. Therefore, to prevent transcript degradation the inverted repeat would either have to be introduced into a *rdr6* mutant that lacks posttranscriptional silencing or the transgene has to be segregated after DNA methylation establishment. *Arabidopsis Wassilewskija* lines containing methylation within the transcribed region would be analysed for *ncRNA locus* expression. The same strategy could be carried out in parallel to test promoter methylation.

If methylation within the 107 bp region of the *ncRNA* promoter was responsible for expression, then the *ncRNA* promoter could be exploited to analyse MET1 regulators. The *ncRNA* promoter would be fused with a reporter gene and transferred into *Arabidopsis Columbia*. To analyse MET1 regulators would require silent *Arabidopsis Columbia* reporter gene lines, which could be produced either by screening a large population or by exploiting the inverted repeat strategy. The progeny would be subjected to EMS mutagenesis and subsequently screened for the release of silencing by analysing reporter gene expression. Lines that express the reporter gene will contain mutations within the *MET1* and *DDM1* genes but also within genes coding MET1 regulators (Figure 7.1).

In addition silent reporter gene lines could be exploited to identify MET1 inhibitors. MET1 inhibitors may become useful as anti-cancer compounds if they were to also inhibit DNMT dysfunction that occurs in cancer cells. To analyse MET1 inhibitors, protoplast from silent reporter gene lines could be exposed to MET1 inhibitor

candidates in a chemical screen. Positive MET1 inhibitors would release the silencing of the reporter gene (Figure 7.2).

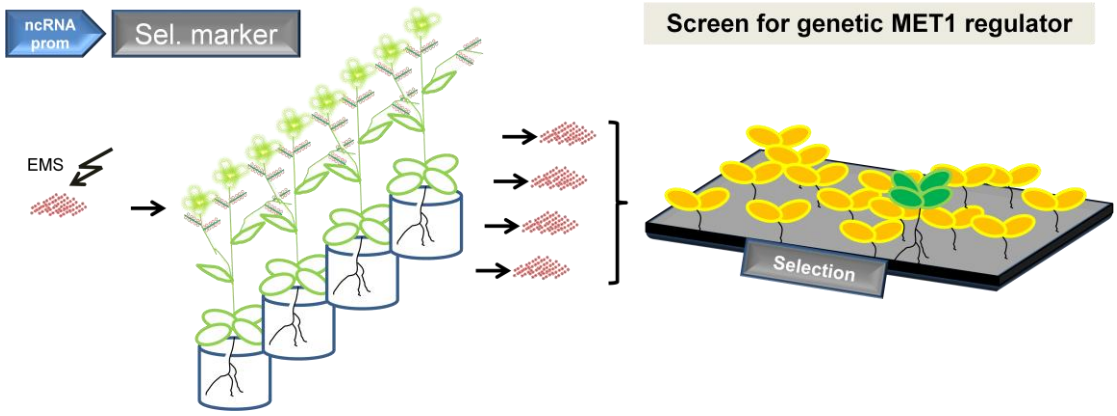


Figure 7.1. A screen for genetic MET1 regulators. An illustration of some key steps to identify genetic MET1 regulators using the *ncRNA* promoter. *ncRNA* promoter reporter gene transformants, which have a silent copy of the transgene due to DNA methylation, would be subjected to EMS mutagenesis indicated on the left by the staggered arrow. Offspring from the EMS mutagenised parent would be analysed for expression of the reporter gene to identify genetic MET1 regulators. In the example above one line shows resistance to selection (green) and therefore expression of the reporter gene, which could be due to loss of silencing from the *ncRNA* promoter via a mutation in MET1 or a MET1 regulator.

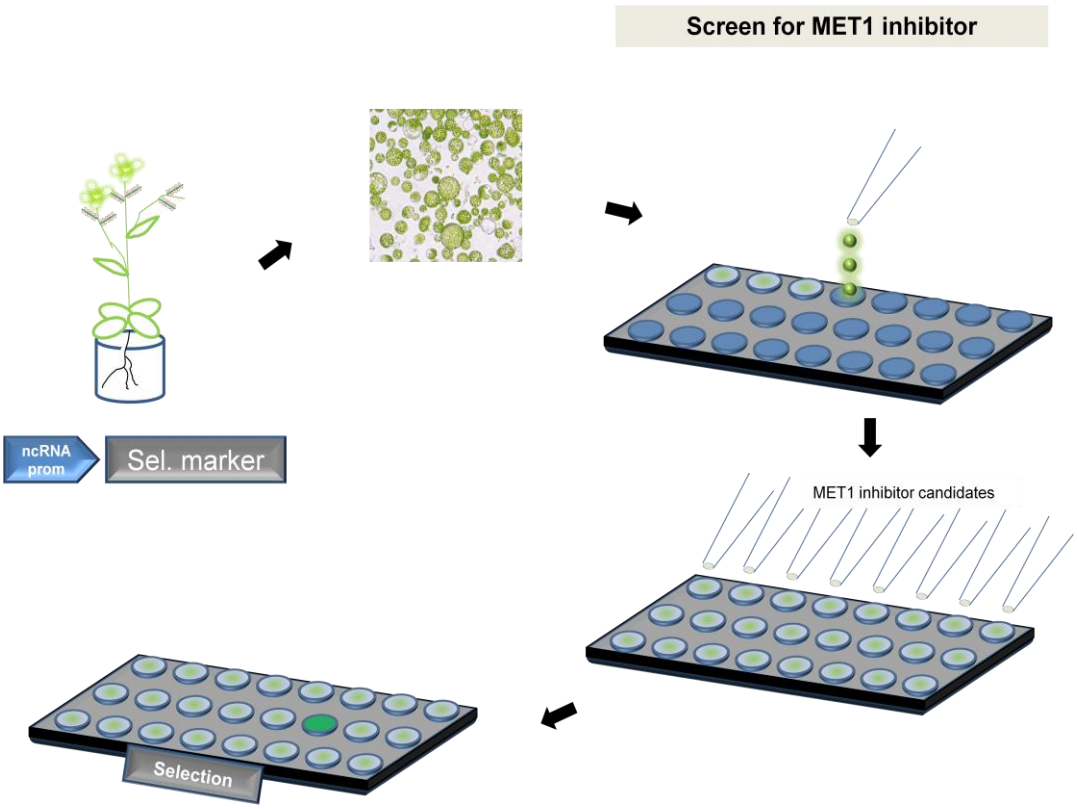


Figure 7.2. A screen for MET1 inhibitors. An illustration of some key steps to identify MET1 inhibitors using the *ncRNA* promoter. Protoplasts from a *ncRNA* promoter reporter gene transformant that has a silent copy of the transgene could be used in a chemical screen with MET1 inhibitor candidates. MET1 inhibitors would be identified by populations of protoplasts that express the reporter gene. The middle well third from the right in the final plate would highlight a potential MET1 inhibitor as silencing of the *ncRNA* promoter has been lost allowing expression of the reporter gene.

7.6.2. MET1 depletion in tomato

The inability to transform tomato with a MET1 inverted repeat driven by a constitutive promoter led to the speculation that MET1 interference is lethal to tomato. To test this theory MET1 will be inactivated using an inducible inverted repeat system. If MET1 depletion is lethal in tomato it would highlight its similarity to mammalian systems (Loughery et al, 2011). Future work should then investigate MMR components to analyse if, as in mammals (Loughery et al, 2011), MET1 depletion interferes with this system resulting in lethality.

Mammalian embryonic lethality is also caused by increasing DNA methyltransferase levels (Biniszkiewicz et al, 2002). Considering the similarity of tomato and mammals in response to a reduction of CG methyltransferases, it would be interesting to analyse

the sensitivity of tomato to *MET1* over-expression by transforming tomato with the *Arabidopsis MET1* construct used in this work.

7.6.3. SAM levels

It is not clear if expression levels of *AT2G41380* are altered by changing SAM resources. To investigate if SAM levels are involved in the potential feedback loop, the expression of *AT2G41380* could be analysed in mutants in feedback loop components, such as an *S-adenosyl methionine synthetase* mutant. *Arabidopsis* contains two SAM synthetase encoding genes, *SAM-1* and *SAM-2*, for which mutants are available. It may be necessary to cross the two mutant lines to generate a *sam-1, sam-2* double mutant to detect changes in *AT2G41380* expression.

A feedback system will likely involve multiple regulators. To characterize the feedback system would require a reporter gene construct using the *AT2G41380* promoter to drive expression. After transfer into plants the expression of the reporter gene could be analysed in a T2 population following EMS mutagenesis. Plants with changes in expression would have mutations in the *MET1* gene but also within other genes required for a feedback system to function.

7.6.4. Target DNA demethylation

The strategies in this work were developed to induce genome-wide DNA methylation changes. While a broad approach provides capacity for scientific discovery, it is also important to consider a target-specific system, which would be useful to induce DNA methylation changes at specific loci. In this work TET3 has been linked to several targeting domains, which could be tested for their target specificity in plants. An optimised target-specific system could be exploited to activate silent disease- or stress-resistant genes. The applications could be extended into mammals where TET3 could be targeted to tumour suppressor genes that become silenced in some cancers.

8.0. Materials and Methods

8.1. Materials

8.1.1. Plant material

All the *Arabidopsis* mutants and respective wild-types used in this study contain a *Columbia* background, unless stated otherwise. The *Arabidopsis met1-1* mutant was provided by Dr Ortrun Mittelsten Scheid (GMI, Vienna, Austria) and genotyped according to Singh et al, (2008). All other T-DNA lines were obtained from NASC and genotyped using primers in Section 2.1.4.1. Hybrid lines were generated using a standard crossing procedure (Scholl et al, 2000).

8.1.2. Bacterial strains

Plasmid cloning was carried out using *Escherichia coli* DH5α (New England Biolabs). Plant transformations were carried out using *Agrobacterium tumefaciens* GV3101::pMP90 (Hellens et al, 2000a).

8.1.3. Donated plasmids and DNA sequences

The pGreen plasmid collection (Hellens et al, 2000a; Hellens et al, 2000b) was provided by Mark Smedley (John Innes centre, Norwich research park, Colney, Norwich, NR4 7UH). pGreen 0179, containing the 35S-*NOS* cassette, was provided by Dr Andrea Kunova (P. Meyer lab, Centre for Plant Sciences, University of Leeds, Leeds, LS2 9JT). The heat inducible promoter (Hsp18.2) (Gallois et al, 2002) was provided by Robert Stablowski (John Innes centre, Norwich research park, Colney, Norwich, NR4 7UH). The *Arabidopsis MET1* cDNA cloned in p-GEM T easy (Promega) and the Hsp18.2:*GUS* construct were provided by Dr Elena Zubko (P. Meyer lab, Centre for Plant Sciences, University of Leeds, Leeds, LS2 9JT). The human *TET3* and MBD cDNA clones were provided by Dr Colm Nestor (R. Meehan lab, MRC Human Genetics Unit MRC IGMM, University of Edinburgh Western General Hospital, Crewe Road, Edinburgh EH4 2XU). The pHannibal inverted repeat

cloning vector (Wesley et al, 2001), was provided by Dr Wayne Charlton (Centre for Plant Sciences, University of Leeds, Leeds, LS2 9JT). The pACN and pSRNACNBin alcohol inducible system was obtained from Zeneca (Caddick et al, 1998; Roslan et al, 2001).

8.1.4. Primer sequences

All of the primer sequences are for *Arabidopsis* unless stated otherwise.

8.1.4.1. Primer sequences used for genotyping

Mutant	Forward primer	Reverse primer
met1-1 (+HaeIII)	CTCTTTAGTAGAAGTTGGCATG	GTTAAGCTCATTCATAGCCTTGC
drm2-2 (SALK_150863)	ATTCGTGTAGCCCTTGAGCC	TTACCGCCTGCCAGATTGTT
ncRNA SNP	CGCACGTCGACTTCTTCTTTT	GCAGACAACAAACAACCACAGA
ddm1-10	TATGCTGCTTTTGCGTTGCT	AGCAACGCCAGTATGTCCTC

8.1.4.2. Primer sequences used for expression analysis

Target gene	Forward primer	Reverse primer
ncRNA (AT4G15242) (qRT-PCR and sqRT-PCR)	CGATCTGTGCGCTTTACTCCC	GGCTTGGGAAATGGAAAGAGG
SAM dependent Methyltransferase (AT2G41380)	AGCAGAGCAATATGCAGCAGCC	TAAATCAACTGATTCAGG
EF1a (AT1G07940) (qRT-PCR)	CTCTCCTTGAGGCTCTTGACCAG	CCAATACCACCAATCTTGTAGACATCC
EF1a (AT1G07940) (sqRT-PCR)	GCGTGTCATTGAGAGGTTCG	GTCAAGAGCCTCAAGGAGAG
FWA (AT4G25530)	GAGAGAGTTGATTACATTGGG	CACTTTATGGGTTGATGCCAC
CACTA (AT2G12210)	CATGTGTAAACGACTGCTGTC	ATTTCCAGGTTGTTGTGGTCC
TET3	CTGCCCGAGCCTGCCAAGTC	GGGCTTCAGGCCTTGCTGGG
MET1 (AT5G49160)	GGGCTCGAGCTTCCATTATCATCAGTCAC	GGGGGTACCGCTGGTTTGGATGAGACAGC
Solanum lycopersicum MET1	CGGCTTGCGTTGAGGTTTAT	GATGACAAAGTCCCTGATGG
Solanum lycopersicum EF1a	GAGCGATGGATGGTGAATCT	TTGTACGTGCGTCCAGAAAG
Solanum lycopersicum MET1 inverted repeat	GCCAACAACAGGAAGATCTCCAA	TCTTCGTCTTACACATCACTTGTCA
UNKNOWN (AT5G15360)	GGCTTGTTGTTCGAGCCGGC	ACGACAGCCCGAGCCGAGA
NODULIN MtN3 (AT4G10850)	ATTAACAATAAACGGCACAGGG	GAAATCCAGCTACCGATAACC
F-BOX gene (AT1G19070)	TCAGATGAGATAATCTGCCACA	CCTAATGGGCAAATTGCTTG
RD29A (AT5G52310)	TGATCGATGCACCAGGCGTAAC	TCGGGGTCTCGACGTTGACCT

8.1.4.3. Primer sequences used for bisulfite sequencing

Target locus	Forward primer	Reverse primer
rDNA	GGGGAGGTAGTGAYAATAAATAA	CACTCTAATTTCTTCAAARTAACA
Upstream of the *AT4G15242* promoter	TGATTAYAATTATTAAAGATTATGTGA	ATTTATAAATARTAAATAAAAATTCA
AT4G15242 promoter	TATTTATAAATTGTGTATTGTAAG	CATCATTAAATATATCATTTAAAC
AT4G15242 promoter/body (fragment 1)	AAATTTATGATATAYTGATAAAATTA	TTTACCATTCATAARCTATAATCC
AT4G15242 promoter/body (fragment 2)	AGTTTATGTTTTAGGTTTTGAATGAATG	CCACRCACRTCRACTTCTTCTTTT
AT4G15242 body (fragment 3)	AAAAGAAGAAGTYGAYGTGYGTGG	CACCAAAAARARACCAACTTCCCC
AT4G15242 body (fragment 4)	TGTTTGTGTGTTGYTGTTTTAGGTGTAG	CCAAAARTRTTARRCAATRCTTACTCACTCTAAC
AT4G15242 body (fragment 5)	GATTGGTGGTGTTAGTATGGCTCTTTGTG	TAAATATTCATTATCACAATRAAAATTTC
AT4G15242 terminator (fragment 6)	GAGGATTAGGTTYAAGAATGTTGTATG	CACAACRAARCARTCACTTTC
AT4G15242 hypo/hyper mC junction	GGATTATAGYTTATGAATGGTAAA	CCACRCACRTCRACTTCTTCTTTT
180 bp CENTROMERIC REPEAT	CATAACGRCCCACTTCCTATATCCACACA	TTGAGTATTAGGGTTTCYGGAATATTTGG
AT5G10540	AGTTTGTGATAYAATTGTATATTGTTTTA	ACATATACTRAAACACRACATATTAA
EVD COPIA	AGATCCCAAAATTCTCTCCCCACC	AGATGAAGAAGAAGATAATAA
RD29A promoter	GTAAATGTAAAATGATTATATGATGGG	CTAAAATTAAAATCTACCTAAATACTAC
ncRNA GUS fusion	GGATTATAGYTTATGAATGGTAAA	CCCACCAACRCTRATCAATTCCAC

8.1.4.4. Primer sequencing used for southern blot probe amplification

Name	Forward primer	Reverse primer
rDNA	TACCTGGTTGATCCTGCCAGT	CAATGATCCTTCCGCAGGTTCAC
180 bp CENTROMERIC REPEAT	CGCCGCCCATGCATAACG	GTTTTCACCACCATTACCTGATC

8.1.4.5. Primer sequences used for methylation-sensitive amplified fragment length polymorphism (MS-AFLP) analysis

Adaptor and respective primer names		
*Eco*RI adapter oligos	CTCGTAGACTGCGTACC	AATTGGTACGCAGTCTAC
*Eco*RI pre-selective primer	GACTGCGTACCAATTCA	
*Eco*RI selective primer	GACTGCGTACCAATTCACTG	
*Hpa*II/*Msp*I adaptor oligos	GATCATGAGTCCTGCT	CGAGCAGGACTCATGA
*Hpa*II/*Msp*I pre-selective primer	ATCATGAGTCCTGCTCGGA	
*Hpa*II/*Msp*I selective primer	ATCATGAGTCCTGCTCGGACT	

8.1.4.6. Primer sequences used for plasmid construction

Construct and respective primer names	Forward primer	Reverse primer
8.2.1.7.1		
5' *Bsu*36I catalytic domain	GTTAAAGTGAGAAGGTTTTATAG	CAGAAAATCCCTGAGATGGAGGTCCACC
3' *PpU*MI catalytic domain	GGTGGACCTCCATCTCAGGGATTTTCTG	TATGCACCGGCCTCCAGGATTCC
8.2.1.7.3		
*Cla*I/*Xba*I Sense	GGGATCGATCCGTTCACTTACTGTCAGAG	GGGTCTAGACTAAGTGAGCCTATTTTTGC
*Xho*I/*Kpn*I Antisense	GGGGGTACCCCGTTCACTTACTGTCAGAG	GGGCTCGAGCTAAGTGAGCCTATTTTTGC
8.2.1.7.4		
*Xho*I/*Fsp*I Heat inducible promoter	GGGTGCGCAATGGTCATTTCTTCTGGTT	GGGCTCGAGCCTGCTGATTTGATCTGA
8.2.1.7.5		
5' *Nsi*I/*Pst*I met1 inverted repeat	GGGATGCATAATATGTCCTTTGCTAATCC	GGGCTGCAGATTGGGGTACCCCGTTCACTTACT
3' *Kpn*I/*Pst*I met1 inverted repeat	ACGGGGTACCCCAATTGGTAAGGAA	GGGCTGCAGACTAAGTGAGCCTATTTTTG
8.2.1.7.6		
*Hind*III/*EcoR*I *TET3*	CCAACCAAGCTTGGCCCCACGGTCGCCTCTAT	CCAGAATTCTGAGGTACGCTGGCTCCCT
8.2.1.7.7		
*Bst*XI/*BamH*I *ncRNA* promoter	GGGCCACCGCCGTGGGCAAGTTTTTGGTTTGTTTT	GGGGGATCCGCTAAGTATATTGGAAGTAT

8.2. Methods

8.2.1. DNA analysis and cloning

8.2.1.1. Isolation of genomic DNA from plants

Isolation of plant genomic DNA was carried out using the modified Vejlupkova and Fowler method (Vejlupkova & Fowler, 2003). 560 μl of extraction buffer (200 mM NaCl; 18 mM $NaHSO_3$; 200 mM Tris-HCl, pH 8.0; 0.07 mM EDTA) and 180 μl of 5% sarkosyl was added to 0.5 g of plant tissue ground in liquid nitrogen. 2 phenol:chlorophorm:isoamyl-alcohol (12:12:1) and 1 chlorophorm extractions were performed after a 2 h 65^{o}C incubation. The DNA was precipitated and washed with 100% isopropanol and 70% ethanol, respectively. Re-suspension was carried out using TE buffer (pH 8.0) supplemented with RNase A (20 mg/l).

8.2.1.2. Mini-prep isolation of plasmid DNA from *E. coli*

Mini-prep isolation of plasmid DNA from *E. coli* was carried out using a modified alkaline lysis method (Sambrook et al, 1989). Individual colonies were grown in a shaking incubator in 2 ml of liquid lysogeny broth (LB) media (10 g/l bacto-tryptone; 5 g/l bacto-yeast extract; 10 g/l NaCl) supplemented with the required antibiotics for 17 h at 37 ^{o}C. 100 μl, 200 μl and 150 μl of solution 1 (50 mM glucose; 25 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0), solution 2 (0.2 M NaOH; 1% SDS) and solution 3 (3 M KAc, pH 5.5) was added to the pelleted culture. Cell debris were removed by centrifugation at 17000 g and plasmid DNA was precipitated and washed with 100% isopropanol and 70% ethanol, respectively.

8.2.1.3. Isolation of plasmid DNA from *Agrobacterium tumefaciens* GV3101::pMP90

Mini-prep isolation of plasmid DNA from *Agrobacterium* was carried out using a modified alkaline lysis method (Wang, 2006). Individual colonies were grown in a shaking incubator in 10 ml of liquid lysogeny broth (LB) media (10 g/l bacto-

tryptone; 5 g/l bacto-yeast extract; 10 g/l NaCl) supplemented with the required antibiotics for 48 h at 28 ^{o}C. 100 µl, 200 µl and 150 µl of solution 1 (50 mM glucose; 25 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0; 4 mg/ml lysozyme), solution 2 (0.2 M NaOH; 1% SDS) and solution 3 (3 M KAc, pH 5.5) was added to the pelleted culture. A phenol:chlorophorm:isoamyl-alcohol extraction was carried out and the DNA was precipitated and washed with 99% ethanol and 70% ethanol, respectively.

8.2.1.4. *E. coli* chemically competent cells and plasmid transformation

E,coli competent cells were made according to (Sambrook et al, 1989). *E. coli* was grown in a shaking incubator in 500 ml of liquid lysogeny broth (LB) media (10 g/l bacto-tryptone; 5 g/l bacto-yeast extract; 10 g/l NaCl) at 37^{o}C. When an OD_{600} 0.4 was reached cells were pelleted (6000 g) and re-suspended three times using 100 mM $MgCl_2$, 100 mM $CaCl_2$ and 85 mM $CaCl_2$ 15% glycerol. The final re-suspension was aliquot, frozen using liquid nitrogen and stored at -80^{o}C. DNA for transformation was added to ice-thawed aliquots and transformed by heat shock.

8.2.1.5. *Agrobacterium tumefaciens* GV3101::pMP90 electro-competent cells and plasmid transformation

Agrobacterium competent cells were made and transformed according to (Mersereau et al, 1990; Shen & Forde, 1989). *Agrobacterium* was grown in a shaking incubator in 500 ml of liquid lysogeny broth (LB) media (10 g/l bacto-tryptone; 5 g/l bacto-yeast extract; 10 g/l NaCl) supplemented with the required antibiotics at 28 ^{o}C. When an OD_{600} 0.8 was reached the cells were pelleted (6000 g) and re-suspended in ice cold sterile water. This procedure was repeated 3 times with a final re-suspension in 10% glycerol. The final re-suspension was aliquot, frozen using liquid nitrogen and stored at -80^{o}C. DNA for transformation was added to ice-thawed aliquots and transformed by electroporation.

8.2.1.6. Polymerase chain reaction (PCR)

PCR for genotyping and semi-quantitative gene expression 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 Phusion (Finnzymes) according to the manufacturer's instructions.

8.2.1.7. Construction of plasmids

Restriction enzymes were used according to the manufacturer's instructions (New England Biolabs). DNA fragments with compatible ends were ligated in a reaction incubated for 17 h at 4 ^{o}C using 1 U of T4 DNA ligase (Promega). De-phosphorylation of recipient plasmids was carried out using calf intestinal alkaline phosphatase (Promega) according to the manufacturer's instructions. 5' overhangs produced after amplicon assembly were filled by PCR. Expression cassettes were transferred into both pGreen 0179 and pGreen 0029, which have in plant hygromycin and kanamycin selection, respectively.

8.2.1.7.1. Catalytically active *Arabidopsis MET1* over-expression

The *MET1* gene was cut from p-GEM T easy (Promega) using *Eco*RI and subsequently ligated into pGreen 0179 35S-*NOS*, which contains a single *Eco*RI site in the polylinker region between the promoter and terminator. *Apa*I and *Not*I allowed the cassette to be transferred into pGreen 0029.

8.2.1.7.2. Catalytically inactive (mutated) *Arabidopsis met1* over-expression

The 5' and 3' ends of the *MET1* region encoding the catalytic domain were amplified using the 5' *Bsu*361 catalytic domain and 3' *PpU*MI catalytic domain primers from met1 in p-GEM T easy (Promega). The 5' *Bsu*361 catalytic domain reverse primer and the 3' *PpU*MI catalytic domain forward primer overlap and contain a missense mutation with respect to the original *MET1* cDNA sequence. This missense mutation changed a conserved cysteine to a serine in the produced protein, when the two amplicons were assembled and replaced the *Bsu*361 and *PpU*MI flanked region in the

catalytically active *Arabidopsis MET1* over-expression plasmid, 8.2.1.7.1. *Apa*I and *Not*I allow the cassette to be transferred into pGreen 0029.

8.2.1.7.3. 35S *Solanum Lycopersicum MET1* inverted repeat

A 609 bp region of the *Solanum lycopersicum MET1* gene was amplified by PCR using two primer sets, *Cla*I/*Xba*I sense and *Xho*I/*Kpn*I antisense. This produced two identical amplicons with different primer incorporated restriction enzyme sites, allowing cloning into the inverted repeat vector pHannibal. The inverted repeat cassette, which consists of the 35S promoter, met1 sense and antisense regions separated by an intron and a *NOS* terminator, was transferred into pGreen 0179 and pGreen 0029 using the restriction enzyme *Not*I.

8.2.1.7.4. Heat inducible *Solanum Lycopersicum MET1* inverted repeat

The heat inducible promoter was amplified by PCR from Hsp18.2:*GUS*, using the *Xho*I/*Fsp*I heat inducible promoter primer pair. The heat inducible promoter amplicon with the incorporated *Xho*I and *Fsp*I sites replaced the 35S promoter in the 35S *Solanum Lycopersicum* met1 inverted repeat plasmid 8.2.1.7.3.

8.2.1.7.5. Alcohol inducible *Solanum Lycopersicum MET1* inverted repeat

The *Solanum Lycopersicum MET1* inverted repeat was amplified in two fragments from the 35S *Solanum Lycopersicum MET1* inverted repeat plasmid, using the 5' *Nsi*I/*Pst*I met1 inverted repeat and 3' *Kpn*I/*Pst*I met1 inverted repeat primer pairs. Both fragments were transferred sequentially into pACN to produce an alcohol promoter; *Solanum Lycopersicum* met1 inverted repeat; *NOS* cassette. This cassette was transferred into pSRNACNBin using *Hind*III, which encodes a transcription factor for full functionality of the alcohol inducible system.

8.2.1.7.6. 35S *TET3*

The 35S *TET3* construct was cloned in collaboration with Dr Elena Zubko (P. Meyer lab, Centre for Plant Sciences, University of Leeds, Leeds, LS2 9JT).The *TET3* gene

was amplified from *TET3* cDNA cloned in pCMV-SPORT6, using the *Hind*III/*Eco*RI *TET3* primer pair. The amplicon was transferred into pGreen 0179 35S-*NOS*, which contains both *Hind*III and *Eco*RI sites in the polylinker region between the promoter and terminator.

8.2.1.7.7. *ncRNA* promoter *GUS*

The *ncRNA* promoter was amplified using the *Bst*XI/*BamH*I *ncRNA* promoter primer pair. The Hsp18.2 region was removed from the Hps18.2:*GUS* construct and replaced with the *ncRNA* promoter using the corresponding restriction enzymes.

8.2.1.8. DNA sequencing

All amplicons were A-tailed using My Taq DNA polymerase (Bioline). A-tailed amplicons were cloned into the T-A cloning vector system p-GEM T-easy (Promega), according to the manufacturer's instructions. Clones containing the target insert were sequenced by Beckman Genomics using the universal SP6 primer. Sequencing reads were aligned using the clustal function in Bioedit 7.0.9.0 (Higo et al, 1999) and methylated and un-methylated counts were illustrated using CyMate (Hetzl et al, 2007).

8.2.1.9. Agarose Gel Electrophoresis

Electrophoretic separation of DNA and RNA molecules was achieved using horizontal agarose gels (0.6-2.0%) containing 0.1 μg/ml ethidium bromide in TAE (Sambrook et al, 1989). DNA and RNA was visualised on a UV trans-illuminator linked to a digital imaging system (Syngene Bio-imager and Genesnap).

8.2.1.10. Polyacrylamide Gel Electrophoresis

Electrophoretic separation of DNA molecules was achieved using vertical polyacrylamide gels (10%) in TBE (Sambrook et al, 1989). DNA was visualised on a UV trans-illuminator linked to a digital imaging system (Syngene Bio-imager and

Genesnap), after staining with 0.1 μg/ml ethidium bromide solution. DNA isolation was achieved using the Q-spin gel extraction kit (Geneflow). DNA size was determined using either 1kb+ (Invitrogen) or 1kb (Bioline).

8.2.1.11. Bisulfite sequencing

The EZ DNA methylation-lightning kit (ZYMO Research) was used according to the manufacturer's instructions for the bisulfite conversion of plant genomic DNA. Target regions were subsequently amplified by PCR. All amplicons were A-tailed using My Taq DNA polymerase (Bioline). A-tailed amplicons were cloned into the T-A cloning vector system p-GEM T-easy (Promega), according to the manufacturer's instructions. Clones containing the target insert were sequenced by Beckman Genomics using the universal SP6 primer. Sequencing reads were aligned using the clustal function in Bioedit 7.0.9.0 (Higo et al, 1999) and methylated and un-methylated counts were illustrated using CyMate (Hetzl et al, 2007). Deamination of cytosines in a CHH sequence context was used as an indicator for conversion.

8.2.1.12. Methylation-sensitive amplified fragment length polymorphism (MS-AFLP) analysis

The MS-AFLP analysis is similar to the standard AFLP analysis described by (Vos et al, 1995) but has been modified to detect DNA methylation (Paun et al, 2010). 0.5 μg of genomic DNA was digested and ligated to double-stranded adapters in one step at 37 °C for 17 h using a MS-AFLP reaction mix (1.1 μl T4 DNA ligase buffer (Promega); 0.55 μl bovine serum albumin (1 mg/ml; New England Biolabs); 50 μM NaCl; 10 U *Msp*I or *Hpa*II (Fermentas); 10 U *Eco*RI (Promega); 1 U T4 DNA ligase (Promega); 4.5 μM *Msp*I/*Hpa*II adapters; 0.45 μM *Eco*RI adapters). The reaction mix was used as a template for PCR with pre-selective primers. The pre-selective PCR reaction was used as a template for PCR with selective primers. The selective PCR reaction was analysed using a 10% polyacrylamide gel. DNA isolation from polyacrylamide gels was achieved using dialyzer bags (Novagen). Sequenced MS-AFLP products were identified using NCBI Blast (*www.ncbi.nlm.nih.gov/*).

8.2.1.13. Southern blot analysis

Southern blot analysis was carried out according to a standard procedure (Sambrook et al, 1989). 3 μg of genomic DNA was digested with selected restriction enzymes according to the manufacturer's instructions (New England Biolabs). The digest reaction was run on an agarose gel (0.7%) for 16 h, 35 V at 4 ^{o}C. The gel was subsequently washed with depurination (0.125 M HCl), denaturation (0.5 M NaOH; 1.5 M NaCl) and neutralisation (1.5 M NaCl; 0.5 M Tris, pH 7.4) buffers for 10, 30 and 30 min, respectively. Membrane blotting was carried out according to (Sambrook et al, 1989) by positioning a positively charged membrane (Amersham GE Healthcare) between the wicked treated gel and weighted blotting paper. DNA was cross linked using UV and incubated in 100 μl/cm^2 hybridisation buffer (7% SDS; 0.5 M Sodium phosphate buffer, pH 7.2; 10 mM EDTA) at 60 ^{o}C for 1 h with rotation. A specific PCR amplified probe was radiolabeled, purified and added to the hybridisation buffer, using the Prime-It random labelling kit (Agilent technologies) and Sephadex G-50 spin columns (Roche), according to the manufacturer's instructions. After 4 h, the membrane was washed twice at 65 ^{o}C for 15 min with 2 x SSC (0.3 M NaCl; 0.03 M NaCi; 0.1% SDS) and once with 0.1 x SSC (0.15 M NaCl; 0.015 M NaCi; 0.1% SDS). Blots were sealed in plastic and visualised using the FLA5100 phospho-imager (Fuji).

8.2.2. RNA analysis

8.2.2.1. Isolation of RNA from plants

Isolation of total plant RNA was performed using a standard procedure (Stam et al, 2000). 750 μl of RNA extraction buffer (100 mM Tris-HCl, pH 8.5; 100 mM NaCl; 20 mM EDTA; 1% sarcosyl) was added to 0.5 g of plant tissue ground in liquid nitrogen. Two phenol:chlorophorm:isoamyl-alcohol (24:24:1) extractions were performed followed by 100% isopropanol, 4 M LiCl and 3 M, pH 7.0 NaAc precipitations. Extractions were quantified using the nano-drop spectrophotometer ND-1000 (Thermo Scientific). DNA was removed using the TURBO DNase kit (Ambion applied biosystems) according to the manufacturer's instructions.

8.2.2.2. cDNA synthesis

cDNA was synthesised using the Superscript II Reverse transcriptase kit (Invitrogen) according to the manufacturer's instructions.

8.2.2.3. Real time-PCR

Real time-PCR was carried out with the BioRad CFX96 real time C1000 thermal cycler, using SsoFast EvaGreen supermix (BioRad) as per manufacturer's instructions and cDNA prepared as described in sections 8.2.2.1 and 8.2.2.2 under the following conditions;

$95^{o}C$ 30 seconds
$95^{o}C$ 5 seconds
x 35
$60^{o}C$ 5 seconds

Target gene expression values were calculated as an average of 3 technical replicates normalised to the ELONGATION FACTOR 1α using a standard curve plotted from ½ serial dilutions of cDNA.

8.2.3. Plant transformation and tissue culture

8.2.3.1. *Arabidopsis* transformation by floral dip

Arabidopsis transformation was carried out by floral dip (Clough & Bent, 1998). A 500 ml transformed *Agrobacterium* culture was grown at 28 ^{o}C with the required antibiotics, until OD_{600} 1.0 was reached. Cells were pelleted (6000 g) and re-suspended to an OD_{600} 0.8 in 5% sucrose; 0.05% Silwett-L77. Wild-type plants grown at 25 ^{o}C, 16/8 h day/light conditions for 4 weeks were inverted into the re-suspended culture for 1 min. Seeds were harvested by bagging 6 weeks old plants. Positive transformants were identified by growing seeds on MS20 medium (4.405 g/l

Murashige and Skoog plus vitamins; 20 g/l Sucrose; 0.55% agar; pH 5.8) containing the required antibiotics, at 25 ^{o}C, 16/8 h day/light conditions for 2 weeks. Seeds were sterilised by washing in 70% ethanol for 2 min, soaking in 30% bleach (4.8% active hypochloride) for 10 min and washing 3 times with sterilised water.

8.2.3.2. Leaf disc transformation of *Solanum lycopersicum*

Leaf disc transformation of *Solanum lycopersicum* was carried out at the premises of ENZA ZADEN, Enkhuizen, The Netherlands (supervised by Iris Heidmann). Seeds of Moneyberg were washed in 99% ethanol, soaked in 25% bleach (4% active hypochloride) for 20 min and rinsed three times with sterilised water. Seeds were sown onto MSB530 medium (Murashige and Skoog salts, B5 vitamins, Duchefa M0231; 30 g/l Sucrose; 0.8% agar; pH 5.8) and germinated at 25 ^{o}C, 16/8 h day/light conditions for 10 days (until cotyledons expanded). Cotyledons were cut into 0.5cm pieces, placed onto solid co-cultivation medium (4.405 g/l MSB5; 3% glucose; 0.8% agar; 200 mg/l KH_2PO_4; 0.2 mg/l 2; 4D, 0.1 mg/l Kinetin; 0.1 mg/l indole-3-acetic acid; 46.8 μM Acetosyrringone; pH 5.8) and pre-cultured overnight. During this time *Agrobacterium* containing the required clone was grown in YEB media (5 g/l Yeast extract; 5 g/l Beef extract; 20 g/l Sucrose; pH 7.2; 2.5 mM MgSO4), with the appropriate antibiotics. The *Agrobacterium* was washed with liquid co-cultivation medium (4.405 g/l MSB5; 3% glucose; 200 mg/l KH_2PO_4; 0.2 mg/l 2; 4D, 0.1 mg/l Kinetin; 0.1 mg/l indole-3-acetic acid; 46.8 μM Acetosyrringone; pH 5.8), diluted to a density of OD_{600} 0.4 and poured over the explants. After one hour the explants were briefly dried, transferred onto fresh co-cultivation medium and incubated at 25 ^{0}C for 76 h under dim light conditions. The explants were transferred to selective medium (4.405 g/l MSB5; 3% glucose; 0.8% agar; 2 mg/l Zeatin; 500 mg/l Cefotaxime; selective antibiotic; pH 5.8) for regeneration.

8.2.3.3. Zebularine treatment

Solanum lycopersicum seedlings were sterilised and sown on MS30 medium (4.405 g/l Murashige and Skoog plus vitamins; 30 g/l Sucrose; pH 5.8) containing 80 μM zebularine (Sigma-Aldrich). Seedlings were incubated in dark conditions for 24 h at 25 ^{o}C, and then transferred to 25 ^{o}C, 16/8 h day/night conditions for 17 days. Seedlings were returned to MS30 medium and DNA was isolated.

8.2.4. GUS analysis

Plant tissue was harvested and vacuum infiltrated in GUS staining solution (100 mM $NaPO_4$, pH 7.0; 10 mM EDTA; 0.5 mM $K_4[Fe(CN)_6]$; 0.5 mM $K_3[Fe(CN)_6]$; 1% Triton X-100; 0.5 mg/ml X-Gluc) for 15 min. Samples were left at 37 ^{o}C for 17 h in the staining solution and subsequently washed with 70-100% ethanol and visualised under a light microscope (Jefferson et al, 1987).

8.2.5. Flowering analysis

Seeds were sterilised by washing in 70% ethanol for 2 min, soaking in 30% bleach (4.8% active hypochloride) for 10 min and washing 3 times with sterilised water. Sterilised seeds were sown on MS30 medium (4.405 g/l Murashige and Skoog plus vitamins; 30 g/l Sucrose; 0.8% agar; pH 5.8) and germinated under long day conditions (25 ^{o}C, 16/8 h day/light). After two weeks seedlings were transferred to soil and grown in long day conditions. Once the primary bolt reached 1cm in height from the base of the plant, leaves above 1cm in length were counted. Therefore, flowering time is measured as the total number of leaves before flowering, described by (Soppe et al, 2000).

9.0. References

Antequera F, Bird A (1999) CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr Biol* **9**(17)**:** R661-R667

Bartee L, Bender J (2001) Two Arabidopsis methylation-deficiency mutations confer only partial effects on a methylated endogenous gene family. *Nucleic Acids Res* **29**(10)**:** 2127-2134

Baubec T, Pecinka A, Rozhon W, Mittelsten Scheid O (2009) Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. *Plant J* **57**(3)**:** 542-554

Baylin SB (2005) DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol* **2 Suppl 1:** S4-11

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

Biniszkiewicz D, Gribnau J, Ramsahoye B, Gaudet F, Eggan K, Humpherys D, Mastrangelo MA, Jun Z, Walter J, Jaenisch R (2002) Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. *Mol Cell Biol* **22**(7)**:** 2124-2135

Bird A (2007) Perceptions of epigenetics. *Nature* **447**(7143)**:** 396-398

Boerjan W, Cervera MT, Delarue M, Beeckman T, Dewitte W, Bellini C, Caboche M, Van Onckelen H, Van Montagu M, Inze D (1995) Superroot, a recessive mutation in Arabidopsis, confers auxin overproduction. *Plant Cell* **7**(9)**:** 1405-1419

Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, Balasubramanian S (2012) Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science* **336**(6083)**:** 934-937

Brink RA (1959) Paramutation at the R-locus in maize plants trisomic for chromosome-10. *Proc Natl Acad Sci U S A* **45**(6)**:** 819-827

Bruniquel D, Schwartz RH (2003) Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nat Immunol* **4**(3)**:** 235-240

Brzeski J, Jerzmanowski A (2003) Deficient in DNA Methylation 1 (DDM1) Defines a Novel Family of Chromatin-remodeling Factors. *J Biol Chem* **278**(2)**:** 823-828

Burn JE, Bagnall DJ, Metzger JD, Dennis ES, Peacock WJ (1993) DNA methylation, vernalization, and the initiation of flowering. *Proc Natl Acad Sci U S A* **90**(1)**:** 287-291

Caddick MX, Greenland AJ, Jepson I, Krause KP, Qu N, Riddell KV, Salter MG, Schuch W, Sonnewald U, Tomsett AB (1998) An ethanol inducible gene switch for plants used to manipulate carbon metabolism. *Nat Biotechnol* **16**(2)**:** 177-180

Cao X, Jacobsen SE (2002a) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Natl Acad Sci USA* **99:** 16491-16498

Cao X, Jacobsen SE (2002b) Role of the Arabidopsis DRM Methyltransferases in De Novo DNA Methylation and Gene Silencing. *Current Biology* **12**(13)**:** 1138-1144

Casadesus J, Low D (2006) Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev* **70**(3)**:** 830-856

Celenza JL, Jr., Grisafi PL, Fink GR (1995) A pathway for lateral root formation in Arabidopsis thaliana. *Genes Dev* **9**(17)**:** 2131-2142

Chan SW (2004) RNA silencing genes control de novo DNA methylation. *Science* **303:** 1336

Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nature Rev Genet* **6:** 351

Chan SWL, Zhang X, Bernatavichute YV, Jacobsen SE (2006) Two-step recruitment of RNA-directed DNA Methylation to tandem repeats. *PLoS Biology* **4**(11)**:** e363

Chen CC, Wang KY, Shen CKJ (2012) The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J Biol Chem* **287**(40)**:** 33116-33121

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 S, Ishii T, Matsumoto N, Tanaka H, Eltayeb AE, Tsujimoto H (2011) Effects of the cytidine analogue zebularine on wheat mitotic chromosomes. *Chromosome Sci* **14:** 23-28

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* **110**(1)**:** 33

Chuang LSH, Ian HI, Koh TW, Ng HH, Xu GL, Li BFL (1997) Human DNA (cytosine-5) methyltransferase PCNA complex as a target for p21(WAF1). *Science* **277**(5334)**:** 1996-2000

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

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. *Nature* **452**(7184)**:** 215-219

Cortellino S, Xu JF, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A (2011) Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell* **146**(1)**:** 67-79

Crowson MS, Shull GE (1992) Isolation and characterization of a cDNA-encoding the putative distal colon H^+,K^+-Atpase - similarity of deduced amino-acid-sequence to gastric H^+,K^+-Atpase and Na^+,K^+-Atpase and messenger-RNA expression in distal colon, kidney, and uterus. *J Bioll Chem* **267**(19)**:** 13740-13748

Das PM, Singal R (2004) DNA methylation and cancer. *J Clin Oncol* **22**(22)**:** 4632-4642

Dechiara TM, Robertson EJ, Efstratiadis A (1991) Parental imprinting of the mouse insulin-like growth factor-li gene. *Cell* **64**(4)**:** 849-859

Dennis K, Fan T, Geiman T, Yan QS, Muegge K (2001) Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes & Dev* **15**(22)**:** 2940-2944

Espada J, Peinado H, Lopez-Serra L, Setien F, Lopez-Serra P, Portela A, Renart J, Carrasco E, Calvo M, Juarranz A, Cano A, Esteller M (2011) Regulation of SNAIL1 and E-cadherin function by DNMT1 in a DNA methylation-independent context. *Nucleic Acids Res* **39**(21)**:** 9194-9205

Ewald B, Sampath D, Plunkett W (2008) Nucleoside analogs: molecular mechanisms signaling cell death. *Oncogene* **27**(50)**:** 6522-6537

Fellinger K, Rothbauer U, Felle M, Langst G, Leonhardt H (2009) Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. *J Cell Biochem* **106**(4)**:** 521-528

Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. *Proc Natl Acad Sci U S A* **93:** 8449-8454

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669)**:** 806-811

Fujimoto R, Sasaki T, Ishikawa R, Osabe K, Kawanabe T, Dennis ES (2012) Molecular mechanisms of epigenetic variation in plants. *Int J Mol Sci* **13**(8)**:** 9900-9922

Gallois JL, Woodward C, Reddy GV, Sablowski R (2002) Combined SHOOT MERISTEMLESS and WUSCHEL trigger ectopic organogenesis in Arabidopsis. *Development* **129**(13)**:** 3207-3217

Galmarini CM, Mackey JR, Dumontet C (2001) Nucleoside analogues: mechanisms of drug resistance and reversal strategies. *Leukemia* **15**(6)**:** 875-890

Gehring M, Henikoff S (2007) DNA methylation dynamics in plant genomes. *Biochim Biophys Acta* **1769:** 276-286

Gidekel M, Jimenez B, Herrera-Estrella L (1996) The first intron of the Arabidopsis thaliana gene coding for elongation factor 1 beta contains an enhancer-like element. *Gene* **170**(2)**:** 201-206

Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74:** 481-514

Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu J-K (2002) ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* **111**(6)**:** 803-814

Gong Z, Zhu J-K (2011) Active DNA demethylation by oxidation and repair. *Cell Res* **21**(12)**:** 1649-1651

Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R (2008) Kismeth: analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics* **9:** 371

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, Iqbal K, Shi YG, Deng Z, Szabo PE, Pfeifer GP, Li J, Xu GL (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**(7366)**:** 606-610

Guo Junjie U, Su Y, Zhong C, Ming G-l, Song H (2011) Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell* **145**(3)**:** 423-434

Hahn P, Schmidt C, Weber M, Kang J, Bielke W (2004) RNA interference: PCR strategies for the quantification of stable degradation-fragments derived from siRNA-targeted mRNAs. *Biomol Eng* **21**(3-5)**:** 113-117

Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**(6)**:** 803-809

He X-J, Hsu Y-F, Zhu S, Wierzbicki AT, Pontes O, Pikaard CS, Liu H-L, Wang C-S, Jin H, Zhu J-K (2009) An effector of RNA-directed DNA Methylation in Arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell* **137**(3)**:** 498-508

He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song C-X, Zhang K, He C, Xu G-L (2011) Tet-mediated formation of 5-

carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**(6047)**:** 1303-1307

Hellens R, Mullineaux P, Klee H (2000a) Technical Focus: A guide to Agrobacterium binary Ti vectors. *Trends Plant Sci* **5**(10)**:** 446-451

Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM (2000b) pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol Biol* **42**(6)**:** 819-832

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

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* **27**(1)**:** 297-300

Hirochika H, Okamoto H, Kakutani T (2000) Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation. *Plant Cell* **12**(3)**:** 357-369

Hobbie L, Estelle M (1995) The axr4 auxin-resistant mutants of Arabidopsis thaliana define a gene important for root gravitropism and lateral root initiation. *Plant J* **7**(2)**:** 211-220

Hsieh CL (1999a) Evidence that protein binding specifies sites of DNA demethylation. *Mol and Cell Biol* **19**(1)**:** 46-56

Hsieh CL (1999b) In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. *Mol Cell Biol* **19**(12)**:** 8211-8218

Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D (2009) Genome-wide demethylation of Arabidopsis endosperm. *Science* **324**(5933)**:** 1451-1454

Hu Y, Chen L, Wang H, Zhang L, Wang F, Yu D (2013) Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. *Plant J* **74**(5)**:** 730-745

Ikeda Y, Kobayashi Y, Yamaguchi A, Abe M, Araki T (2007) Molecular basis of late-flowering phenotype caused by dominant epi-alleles of the FWA locus in Arabidopsis. *Plant Cell Physiol* **48**(2)**:** 205-220

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333:** 1300-1303

Iudanova SS, Maletskaia EI, Maletskii SI (2012) Epigenetic variation of the choriflowered-symflowered character in the symflowered sugar beet line SOAN-14. *Genetika* **48**(7)**:** 827-834

Jackson AL, Linsley PS Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat Rev Drug Discov* **9**(1)**:** 57-67

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

Jahnke S, Scholten S (2009) Epigenetic resetting of a gene imprinted in plant embryos. *Current Biology* **19**(19)**:** 1677-1681

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* **6**(13)**:** 3901-3907

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. *Current Biology* **17**(4)**:** 379

Jullien PE, Mosquna A, Ingouff M, Sakata T, Ohad N, Berger F (2008) Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis. *PLoS Biol* **6**(8)**:** e194

Kakutani T, Jeddeloh JA, Richards EJ (1995) Characterization of an Arabidopsis thaliana DNA hypomethylation mutant. *Nucleic Acids Res* **23**(1)**:** 130-137

Kakutani T, Munakata K, Richards EJ, Hirochika H (1999) Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of Arabidopsis thaliana. *Genetics* **151**(2)**:** 831-838

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

Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T (2003) Role of CG and Non-CG Methylation in Immobilization of Transposons in Arabidopsis. *Curr Biol* **13**(5)**:** 421

Kelly MO, Saltveit ME (1988) Effect of Endogenously Synthesized and Exogenously Applied Ethanol on Tomato Fruit Ripening. *Plant Physiology* **88**(1)**:** 143-147

Kim M, Ohr H, Lee JW, Hyun Y, Fischer RL, Choi Y (2008) Temporal and spatial downregulation of Arabidopsis MET1 activity results in global DNA hypomethylation and developmental defects. *Mol Cells* **26**(6)**:** 611-615

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T (2004) One-Way Control of FWA Imprinting in Arabidopsis Endosperm by DNA Methylation. *Science* **303**(5657)**:** 521-523

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M, Kakutani T (2007) Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J* **49**(1)**:** 38-45

Kishimoto N, Sakai H, Jackson J, Jacobsen SE, Meyerowitz EM, Dennis ES, Finnegan EJ (2001) Site specificity of the Arabidopsis METI DNA methyltransferase demonstrated through hypermethylation of the superman locus. *Plant Mol Biol* **46**(2)**:** 171-183

Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**(2)**:** 89-97

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, Rodig SJ, Daley GQ, Rao A (2011) Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**(2)**:** 200-213

Koornneef M, Hanhart CJ, Vanderveen JH (1991) A Genetic and Physiological Analysis of Late Flowering Mutants in Arabidopsis-Thaliana. *Mol Gen Genet* **229**(1)**:** 57-66

Larsen F, Gundersen G, Lopez R, Prydz H (1992) CpG islands as gene markers in the human genome. *Genomics* **13**(4)**:** 1095-1107

Lauria M, Rossi V (2011) Epigenetic control of gene regulation in plants. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1809**(8)**:** 369-378

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

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**(3)**:** 204-220

Ledford H (2008) Language: Disputed definitions. *Nature* **455**(7216)**:** 1023-1028

Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* **3**(9)**:** 662-673

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

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**(5524)**:** 2077-2080

Lindroth AM, Shultis D, Jasencakova Z, Jörg Fuchs J, Lianna Johnson L, Daniel Schubert D, Debasis Patnaik D, Sriharsa Pradhan S, Justin Goodrich J, Ingo Schubert I, Thomas Jenuwein T, Sepideh Khorasanizadeh S, Steven E Jacobsen SE (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J* **23:** 4146–4155

Liu X, Yu CW, Duan J, Luo M, Wang K, Tian G, Cui Y, Wu K (2011) HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable element silencing in Arabidopsis. *Plant Physiol* **158**(1)**:** 119-129

Loughery JEP, Dunne PD, O'Neill KM, Meehan RR, McDaid JR, Walsh CP (2011) DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response. *Hum Mol Genet* **20**(16)**:** 3241-3255

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648)**:** 251-260

Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A (2000) Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proc Natl Acad Sci U S A* **97**(19)**:** 10637-10642

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* **38**(8)**:** 948-952

Martel C, Vrebalov J, Tafelmeyer P, Giovannoni JJ (2011) The tomato MADS-box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner. *Plant Physiol* **157**(3)**:** 1568-1579

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

Melamed-Bessudo C, Levy AA (2012) Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proc Natl Acad Sci U S A* **109**(16)**:** E981–E988

Mersereau M, Pazour GJ, Das A (1990) Efficient transformation of Agrobacterium tumefaciens by electroporation. *Gene* **90**(1)**:** 149-151

Meyer P (2001) Chromatin remodelling. *Curr Opin Plant Biol* **4**(5)**:** 457-462

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

Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**(7262)**:** 427

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. *Nature* **411**(6834)**:** 212-214

Miura K, Agetsuma M, Kitano H, Yoshimura A, Matsuoka M, Jacobsen SE, Ashikari M (2009) A metastable DWARF1 epigenetic mutant affecting plant stature in rice. *Proc Natl Acad Sci U S A* **106**(27)**:** 11218-11223

Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, Husmann D, Zhan Y, Lajoie BR, McCord RP, Hale CJ, Feng W, Michaels SD, Frand AR, Pellegrini M, Dekker J, Kim JK, Jacobsen SE (2012) MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing. *Science* **336**(6087)**:** 1448-1451

Monk M, Boubelik M, Lehnert S (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ-cell lineages during mouse embryo development. *Development* **99**(3)**:** 371-382

Muller HJ (1930) Types of visible variations induced by x-rays in Drosophila. *Journal of Genetics* **22**(3)**:** 299-U297

Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2**(4)**:** 279-289

Oh K, Ivanchenko MG, White TJ, Lomax TL (2006) The diageotropica gene of tomato encodes a cyclophilin: a novel player in auxin signaling. *Planta* **224**(1)**:** 133-144

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

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:** 247

Ono A, Yamaguchi K, Fukada-Tanaka S, Terada R, Mitsui T, Iida S (2012) A null mutation of ROS1a for DNA demethylation in rice is not transmittable to progeny. *The Plant Journal* **71**(4)**:** 564-574

Ooi SKT, Qiu C, Bernstein E, Li KQ, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng XD, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**(7154)**:** 714-U713

Paun O, Bateman RM, Fay MF, Hedren M, Civeyrel L, Chase MW (2010) Stable epigenetic effects impact adaptation in allopolyploid orchids (Dactylorhiza: Orchidaceae). *Mol Biol Evol* **27**(11)**:** 2465-2473

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

Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL (2007) DNA demethylation in the Arabidopsis genome. *Proc Natl Acad Sci U S A* **104**(16)**:** 6752-6757

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

Pikaard CS (2013) Methylating the DNA of the Most Repressed: Special Access Required. *Mol Cell* **49**(6)**:** 1021-1022

Portis E, Acquadro A, Comino C, Lanteri S (2004) Analysis of DNA methylation during germination of pepper (Capsicum annuum L.) seeds using methylation-sensitive amplification polymorphism (MSAP). *Plant Sci* **166**(1)**:** 169-178

Prakash AP, Kush A, Lakshmanan P, Kumar PP (2003) Cytosine methylation occurs in a CDC48 homologue and a MADS-box gene during adventitious shoot induction in Petunia leaf explants. *J Exp Bot* **54**(386)**:** 1361-1371

Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**(7114)**:** 1008

Rai GK, Rai NP, Kumar S, Yadav A, Rathaur S, Singh M (2012) Effects of explant age, germination medium, pre-culture parameters, inoculation medium, pH, washing medium, and selection regime on Agrobacterium-mediated transformation of tomato. *In Vitro Cell Dev-Pl* **48**(5)**:** 565-578

Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**(7)**:** 1201-1212

Raj K, Mufti GJ (2006) Azacytidine (Vidaza(R)) in the treatment of myelodysplastic syndromes. *Ther Clin Risk Manag* **2**(4)**:** 377-388

Ramsahoye BH (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci USA* **97:** 5237-5242

Rangwala SH, Richards EJ (2007) Differential epigenetic regulation within an Arabidopsis retroposon family. *Genetics* **176**(1)**:** 151-160

Rigal M, Kevei Z, Pelissier T, Mathieu O (2012) DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. *EMBO J* **31**(13)**:** 2981-2993

Rocha PS, Sheikh M, Melchiorre R, Fagard M, Boutet S, Loach R, Moffatt B, Wagner C, Vaucheret H, Furner I (2005) The Arabidopsis HOMOLOGY-DEPENDENT GENE SILENCING1 gene codes for an S-adenosyl-L-homocysteine

hydrolase required for DNA methylation-dependent gene silencing. *Plant Cell* **17**(2)**:** 404-417

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

Ronemus MJ, Galbiati M, Ticknor C, Chen JC, Dellaporta SL (1996) Demethylation-induced developmental pleiotropy in Arabidopsis. *Science* **273**(5275)**:** 654-657

Roslan HA, Salter MG, Wood CD, White MR, Croft KP, Robson F, Coupland G, Doonan J, Laufs P, Tomsett AB, Caddick MX (2001) Characterization of the ethanol-inducible alc gene-expression system in Arabidopsis thaliana. *Plant J* **28**(2)**:** 225-235

Rountree MR, Selker EU (1997) DNA methylation inhibits elongation but not initiation of transcription in Neurospora crassa. *Genes Dev* **11:** 2383

Sako K, Maki Y, Kanai T, Kato E, Maekawa S, Yasuda S, Sato T, Watahiki MK, Yamaguchi J (2012) Arabidopsis RPT2a, 19S proteasome subunit, regulates gene silencing via DNA methylation. *PLoS ONE* **7**(5)**:** e37086

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

Sano H, Kamada I, Youssefian S, Katsumi M, Wabiko H (1990) A Single Treatment of Rice Seedlings with 5-Azacytidine Induces Heritable Dwarfism and Undermethylation of Genomic DNA. *Mol Gen Genet* **220**(3)**:** 441-447

Santi DV, Garrett CE, Barr PJ (1983) On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* **33**(1)**:** 9-10

Sato S, et al, (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**(7400)**:** 635-641

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)**:** 3641-3652

Saze H, Scheid OM, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* **34**(1)**:** 65-69

Scholl RL, May ST, Ware DH (2000) Seed and molecular resources for Arabidopsis. *Plant Physiol* **124**(4)**:** 1477-1480

Shearwin KE, Callen BP, Egan JB (2005) Transcriptional interference - a crash course. *Trends Genet* **21**(6)**:** 339-345

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

Sieburth LE, Meyerowitz EM (1997) Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. *Plant Cell* **9**(3)**:** 355-365

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

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijó JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**(3)**:** 461-472

Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJM (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

Soppe WJJ, Jasencakova Z, Houben A, Kakutani T, Meister A, Huang MS, Jacobsen SE, Schubert I, Fransz PF (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO Journal* **21**(23)**:** 6549-6559

Stam M, de Bruin R, van Blokland R, van der Hoorn RA, Mol JN, 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

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**(6)**:** 465-476

Takuno S, Gaut BS (2011) Body-methylated genes in Arabidopsis thaliana are functionally important and evolve slowly. *Mol Biol and Evol* **9:** 219-222

Tan L, Shi YG (2012) Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**(11)**:** 1895-1902

Tariq M, Saze H, Probst AV, Lichota J, Habu Y, Paszkowski J (2003) Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. *Proc Natl Acad Sci USA* **100**(15)**:** 8823-8827

Teerawanichpan P, Chandrasekharan MB, Jiang YM, 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

Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, Voinnet O, Wincker P, Esteller M, Colot V (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**(5921)**:** 1600-1604

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. *Science* **327**(5967)**:** 818-822

Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* **67**(3)**:** 946-950

Vanyushin BF, Sevast'yanova SS, Kirnos MD, Saidova NS (1990) Increasing the protein content of wheat grain under the influence of 5-azacytidine, a DNA methylation inhibitor *Biol Bull Acad sci USSR* **17**(1)**:** 60-66

Vaughn MW, Tanurdzi M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, Colot V, Doerge RW, Martienssen RA (2007) Epigenetic natural variation in Arabidopsis thaliana. *PLoS Biology* **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

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, et al. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**(21)**:** 4407-4414

Waalwijk C, Flavell RA (1978) Mspl, an Isoschizomer of Hpaii Which Cleaves Both Unmethylated and Methylated Hpaii Sites. *Nucleic Acids Res* **5**(9)**:** 3231-3236

Waddington CH (1942) The epigenotype. *Int J Epidemiol* **41**(1)**:** 10-13

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

Wesley SV, Helliwell CA, Smith NA, Wang MB, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J* **27**(6)**:** 581-590

Wierzbicki AT, Haag JR, Pikaard CS (2008) Non-coding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**(4)**:** 635-648

Wolffe AP, Jones PL, Wade PA (1999) DNA demethylation. *Proc Natl Acad Sci USA* **96**(11)**:** 5894-5896

Woo HR, Richards EJ (2008) Natural variation in DNA methylation in ribosomal RNA genes of Arabidopsis thaliana. *BMC Plant Biol* **8:** 92

Xu P, Zhang Y, Kang L, Roossinck MJ, Mysore KS (2006) Computational estimation and experimental verification of off-target silencing during posttranscriptional gene silencing in plants. *Plant Physiol* **142**(2)**:** 429-440

Xu Y, Xu C, Kato A, Tempel W, Abreu JG, Bian C, Hu Y, Hu D, Zhao B, Cerovina T, Diao J, Wu F, He HH, Cui Q, Clark E, Ma C, Barbara A, Veenstra GJ, Xu G, Kaiser UB, Liu XS, Sugrue SP, He X, Min J, Kato Y, Shi YG (2012) Tet3 CXXC

domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* **151**(6)**:** 1200-1213

You WH, Tyczewska A, Spencer M, Daxinger L, Schmid MW, Grossniklaus U, Simon SA, Meyers BC, Matzke AJM, Matzke M (2012) Atypical DNA methylation of genes encoding cysteine-rich peptides in Arabidopsis thaliana. *Bmc Plant Biology* **12**

Yu M, Hon GC, Szulwach KE, Song CX, Jin P, Ren B, He C (2012) Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc* **7**(12)**:** 2159-2170

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

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

Zhong S, Fei Z, Chen YR, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J, Shao Y, Giovannoni JJ (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat Biotechnol* **31**(2)**:** 154-159

Zhu B, Zheng Y, Angliker H, Schwarz S, Thiry S, Siegmann M, Jost JP (2000) 5-methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res* **28**(21)**:** 4157-4165

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**(1)**:** 61

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

10.0. Appendix

Figure 10.1. Sequencing results from MS-AFLP polymorphisms. Targets labelled somatic change and unaffected from Figure 2.3 were sequenced as described in Section 8.2.1.8 of Materials and Methods. For each target the sequencing result is shown first and the database alignment result with the highest homology follows. Alignments were performed using the NCBI basic nucleotide blast search tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Sequencing result for the amplicon labelled somatic change.

CGGATACGGTAGACGCAGTGGGCATGGGGCCTTCACCGGCTTCTATCTGC
CCAAAACGAATGCTCCTTGCGAATGACTGCCGCGCTCGCCTTGGACCCGA
CCGTGCCCGAAAGGGCGCGCCGGGCTCATGCGGCGCGCGGCGTCGTTGAG
GAATGCTACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATT
AAGCCATGCATGTGTAAGTATGAACAAATTCAGACTGTGAAACTGCGAAT
GGCTCATTAAATCAGTTATAGTTTGTTTGATGGTATCTACTACTCGGATAA
CCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCGACTTCTGGAAG
GGATGCATTTATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCTGCG
ATGATTCATGATAACTCGACGGATCGCACGGCCATCGTGCCGGCGACGCA
TCATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACC
ATGGTGGTGACGGGTGACGGAGAATTAGGGTTCGAT

NCBI database blast result for the above sequencing read.

It is homologous with an 18S ribosomal DNA gene.

CGGATACGGTAGACGCAGTGGGCATGGGGCCTTCACCGGCTTCTATCTGC
CCAAAACGAATGCTCCTTGCGAATGACTGCCGCGCTCGCCTTGGACCCGA
CCGTGCCCGAAAGGGCGCGCCGGGCTCATGCGGCGCGCGGCGTCGTTGAG
GAATGCTACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATT
AAGCCATGCATGTGTAAGTATGAACAAATTCAGACTGTGAAACTGCGAAT
GGCTCATTAAATCAGTTATAGTTTGTTTGATGGTATCTACTACTCGGATAA
CCGTAGTAATTCTAGAGCTAATACGTGCAACAAACCCCGACTTCTGGAAG
GGATGCATTTATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCTGCG
ATGATTCATGATAACTCGACGGATCGCACGGCCATCGTGCCGGCGACGCA
TCATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACC
ATGGTGGTGACGGGTGACGGAGAATTAGGGTTCGAT

Sequencing result for the amplicon labelled unaffected.

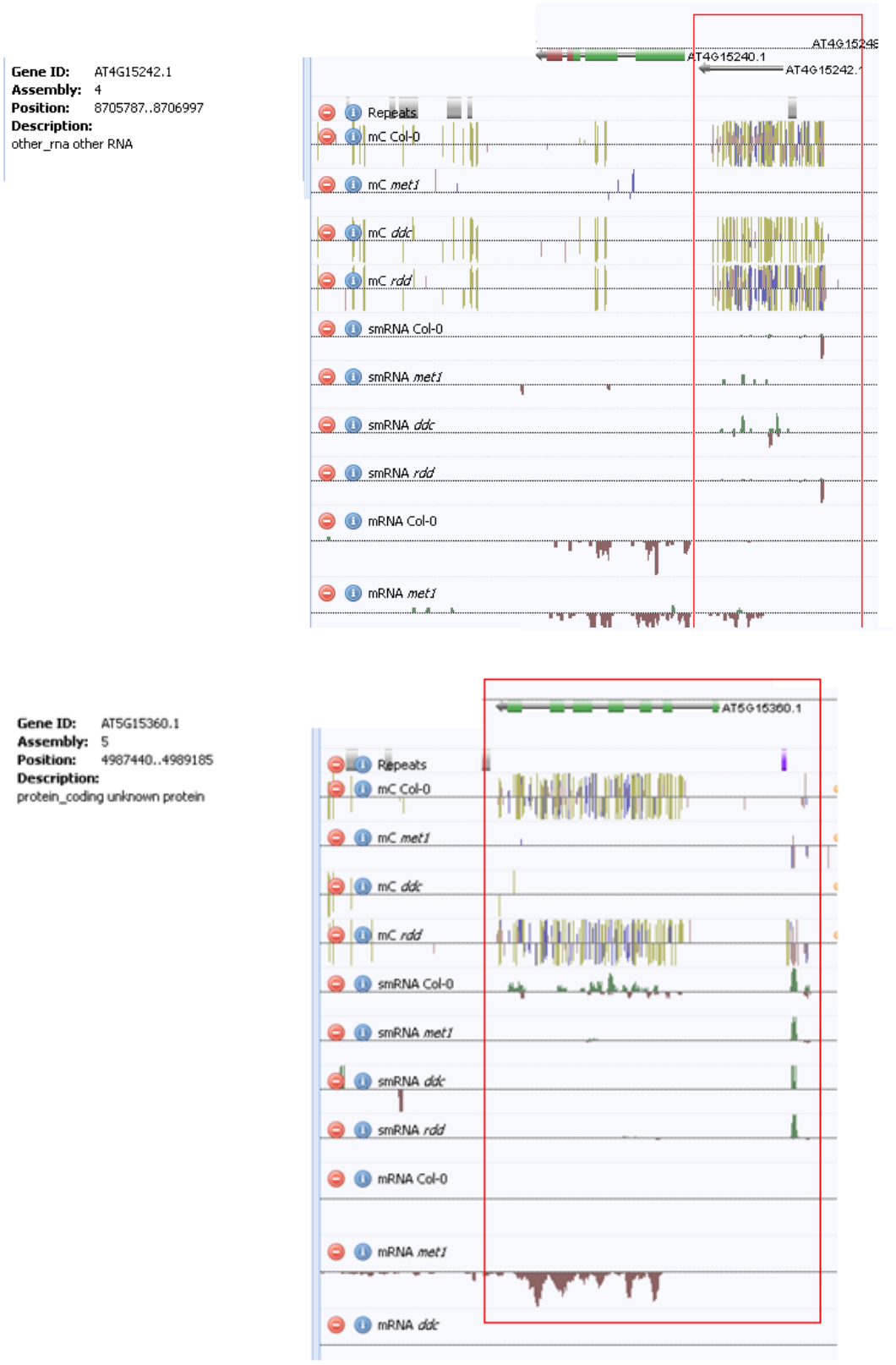
TGCACTCCGGACGGTCCTGGGGGCAAGGCTAATTCATAAGCTGCCTCCCC
CACGCGCTTCAGAACTTCAAATGGACGAATATACCTAGGACTAAGCATAC
CTCTCTTACCAAACCGCATCACTCCTTTCATGGGTGAAACCTTCAGCAAGA
CTTGCTCACCCTCCATAAACTCTAAATCTCTAACCTTTCGATCTGCATATTC
TTTTTGTCTACTTTGAGCCGCTAAAAGCTTTTCTTGAATGCATTTCACTTTA
TCTAATGATTCCCTCAGAAGGTCAGTACCCCAAGGTCTAACCTCGAATGC
ATCAAATCAAACCAACCAATGGGAGATCTACATCTCCTCCCATACAGTCC
CTCAAATGGGGCCATATCAATACTTGAGTGATAGCTATTGTTGTACG

NCBI database blast result for the above sequencing read.

It is homologous with a member of the *JINLING* retrotransposon family in Tomato.

GTGCACTCCTGACAGTCCTGGGGGCAAGGCTAATTCATAAGCTACCTCCC
CCACGCGCTTCAGAACTTCAAATGGACCAATATACCTCGGACTAAGCTTA
CCTCGCTTACCAAACCGCATCACCCCTTTCATGGGCGAAACCTTCAGCAA
GACTTGCTCACCCTCCATAAACTCCAAGTCTCTAACCTTTCGATCTGCATA
TTCTTTTTGTCTACTTTGAGCCGCTAAAAGCTTTTCTTGAATGGATTTCACT
TTATCTAACGAATCCCTCAGAAGGTCAGTACCCCAAGGTCTAACCTCAAA
TGCATCAAACCAACCAATGGGAGACCTACATCTCCTCCCATACAATGCTT
CAAATGGGGCCATATCAATACTTGAGTGATAGCTATTGTTGTATGA

Figure 10.2. Epigenome browser screen shots of MET1 targets. The screen shots were obtained for the respective gene (red boxed region) from (http://neomorph.salk.edu/epigenome/epigenome.html). Methylation types including CG, CHG and CHH are colour coded and are shown on the top four tracts for *Arabidopsis Columbia* (*Col*-0), the *Arabidopsis met1-3* mutant (*met1*), a *drm1, drm2* and *cmt3* triple mutant (*ddc*) and a *ros1, dml2* and *dml3* triple mutant (*rdd*). The next four tracts and the final tracts indicate small RNAs and expression levels, respectively, for *Col*-0 and the mutants.



Gene ID: AT4G10850.1
Assembly: 4
Position: 6674994..6676977
Description:
protein_coding nodulin MtN3 family protein similar to nodulin MtN3 family protein [Arabidopsis thaliana] (TAIR:AT1G66770.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO69485.1); contains InterPro domain MtN3 and saliva related transmembrane protein (InterPro:IPR004316)

Gene ID: AT1G19070.1
Assembly: 1
Position: 6584445..6584696
Description:
protein_coding F-box family protein similar to F-box family protein [Arabidopsis thaliana] (TAIR:AT4G22280.1); similar to F-box family protein [Arabidopsis thaliana] (TAIR:AT4G22280.2); similar to unnamed protein product [Vitis vinifera] (GB:CAO63569.1); contains InterPro domain Cyclin-like F-box (InterPro:IPR001810)

Gene ID: AT2G41380.1
Assembly: 2
Position: 17259019..17260135
Description:
protein_coding embryo-abundant protein-related similar to embryo-abundant protein-related [Arabidopsis thaliana] (TAIR:AT5G10830.1); similar to unknown [Populus trichocarpa x Populus deltoides] (GB:ABK96242.1); contains InterPro domain Methyltransferase type 11; (InterPro:IPR013216)

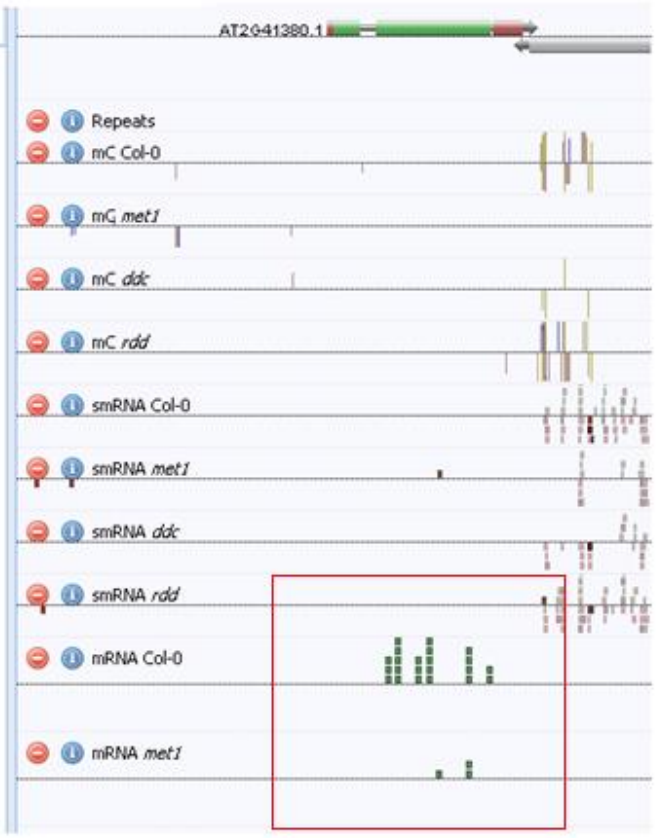
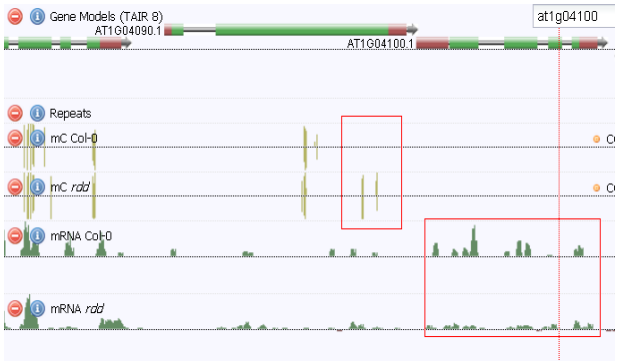
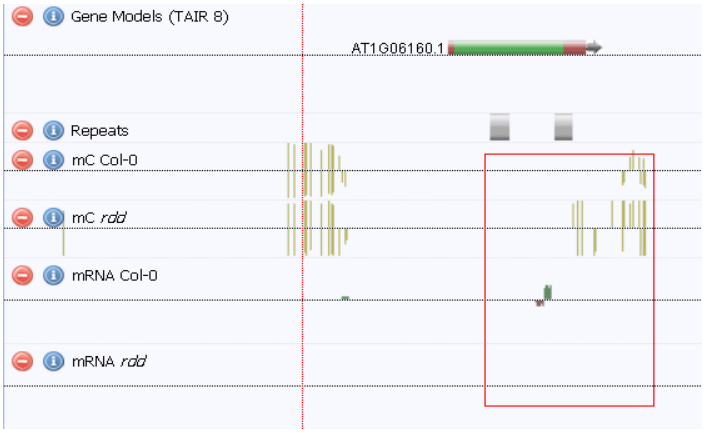


Figure 10.3. Epigenome browser screen shots of ROS1 targets. The screen shots were obtained for the respective gene (red boxed region) from (http://neomorph.salk.edu/epigenome/epigenome.html). Methylation types including CG, CHG and CHH are colour coded and are shown on the top two tracts for *Arabidopsis Columbia* (*Col*-0) and a *ros1, dml2* and *dml3* triple mutant (*rdd*). The next two tracts indicate expression levels for the respective gene in the two lines.

Gene ID: AT1G04100.1
Assembly: 1
Position: 1059505..1061190
Description:
protein_coding IAA10 (indoleacetic acid-induced protein 10); transcription factor Auxin induced gene, IAA10 (IAA10). Identical to Auxin-responsive protein IAA10 (IAA10) [Arabidopsis Thaliana] (GB:Q38828); similar to IAA11 (indoleacetic acid-induced protein 11) [Arabidopsis thaliana] (TAIR:AT4G28640.2); similar to IAA11 (indoleacetic acid-induced protein 11), transcription factor [Arabidopsis thaliana] (TAIR:AT4G28640.1); similar to aux/IAA protein [Populus tremula x Populus tremuloides] (GB:CAC84709.1); contains InterPro domain AUX/IAA protein; (InterPro:IPR003311); contains InterPro domain Aux/IAA-ARF-dimerisation; (InterPro:IPR011525)



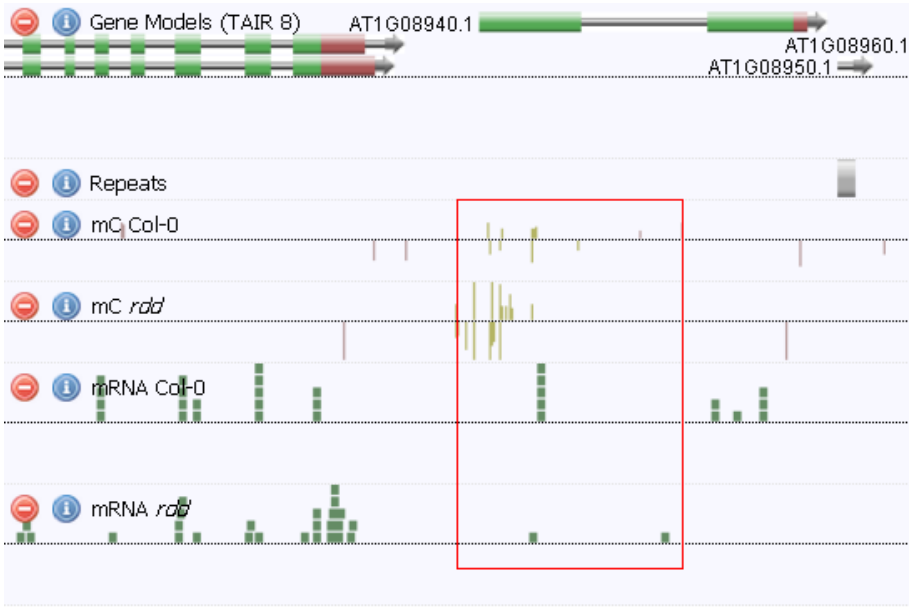
Gene ID: AT1G06160.1
Assembly: 1
Position: 1883003..1883933
Description:
protein_coding ethylene-responsive factor, putative encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5. Identical to Ethylene-responsive transcription factor ERF094 (ERF094) [Arabidopsis Thaliana] (GB:Q9LND1); similar to ATERF15 (ETHYLENE-RESPONSIVE ELEMENT BINDING FACTOR 15), DNA binding / transcription activator/ transcription factor [Arabidopsis thaliana] (TAIR:AT2G31230.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO65033.1); similar to hypothetical protein [Vitis vinifera] (GB:CAN78568.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO70707.1); contains InterPro domain DNA-binding, integrase-type; (InterPro:IPR016177); contains InterPro domain Pathogenesis-related transcriptional factor and ERF, DNA-binding; (InterPro:IPR001471)



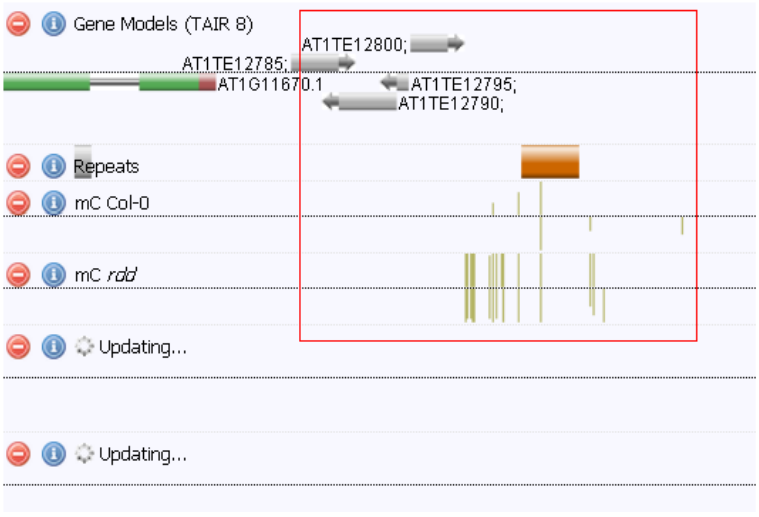
AT1G08940.1

protein_coding

Phosphoglycerate mutase family protein; FUNCTIONS IN: catalytic activity; INVOLVED IN: metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Histidine phosphatase superfamily, clade-1 (InterPro:IPR013078), Phosphoglycerate/bisphosphoglycerate mutase, active site (InterPro:IPR001345); BEST Arabidopsis thaliana protein match is: Phosphoglycerate mutase family protein (TAIR:AT3G05170.1); Has 1152 Blast hits to 1140 proteins in 473 species: Archae - 9; Bacteria - 736; Metazoa - 2; Fungi - 189; Plants - 85; Viruses - 0; Other Eukaryotes - 131 (source: NCBI BLink).



Gene ID: AT1TE12790;
Assembly: 1
Position: 3932218..3932527
Description:



Gene ID: AT2G46640.2
Assembly: 2
Position: 19155800..19158065
Description:
protein_coding unknown protein

Figure 10.4. TET3 cDNA and protein sequence. The TET3 protein and cDNA sequence obtained from NCBI (http://www.ncbi.nlm.nih.gov/) are shown on the left and right of the figure, respectively. The 3' region coding for the catalytic domain and nuclear localisation sequence are highlighted in blue.