

**Understanding the interaction between
RNA-directed DNA methylation and
DNA demethylases and its role in
Fusarium oxysporum disease response
in *Arabidopsis thaliana***

Joanne R.M. Lee, B. Biotech (Hons.)

A thesis submitted for the degree of Doctor of Philosophy of the
Australian National University

January 2016

© Copyright by [Joanne R.M. Lee] [2016]

All Rights Reserved

Statement of Authorship

The research carried out in this thesis was conducted at the CSIRO between March 2012 – January 2016 for submission of a PhD from The Australian National University. I declare I am the sole author and have made specific references to work done by other sources. The content of this thesis has only been generated since the commencement of my higher degree candidature and does not include work submitted for the award of any other diploma or degree at The Australian National University or any other university.

Joanne Lee

Acknowledgements

I would like to begin with my heart felt gratitude to my supervisor Dr Ming-Bo Wang. He has supported my research endeavors since the very beginning and more importantly encouraged me to be an independent researcher and to develop and grow in all aspects of my PhD training. Dr Tony Millar has been my ANU supervisor and has given his time and support whenever I needed general or specific advice. I am eternally grateful to both supervisors.

Across the course of my PhD I have been fortunate to have my advisors, Dr Danny Llewellyn, Dr Iain Wilson and Dr Liz Dennis who I'd like to thank. The combination of different personalities and the broad experience they brought with them in my supervision helped develop my personal resilience, scientific creativity and goal targeted approach to the thesis. Dr Jim Peacock, Dr Kemal Kazan, Dr Ian Greaves and Dr Ben Trevaskis have also been instrumental to my success and I thank them for their encouragement and enthusiasm.

Many of the technical skills I have learnt would not have been possible without a great group of people. Firstly Neil Smith, our lab technician, has been instrumental in helping me with Northern Blots, DNA and RNA extractions, cloning, plant husbandry and many other experiments. Neil is also an all-round decent bloke and made working in the lab a pleasure and I'm immensely grateful. The career advice and epigenetic knowledge provided by Dr Ulrike Schumann helped me adjust to my PhD studies. I'd also like to thank her for teaching me fungal infection, chitin biomass experiments and RT-PCR. Dr Ming-Bo Wang, as well as supervising, helped me with viral infection assays and across the board in the other experiments I performed. It is rare to have a supervisor keen to help you out in the lab and I (along with my results) certainly benefited. Bioinformatics was a key part of this thesis and I'd like to thank Dr Phil Au, Dr Michael Groszmann and Dr Steve Eichten for their help with the small RNA, mRNA and bisulfite sequencing respectively.

I thank Jian-Kang Zhu for providing the *rdd* mutant, Robert Fischer for the *ros1*, *dml2* and *dml3* mutants and Kemal Kazan's group for supplying *Fusarium oxysporum* f.sp. *conglutinans* (strain 4287) and whole genome 454 sequencing contigs. I thank Ming-Bo Wang for the use of the TuMV and CMV viruses. I appreciate the photography taken by Carl Davies.

Lastly I'd like to thank my friends and family for helping me still have fun during the PhD. Thanks to my parents for making the trips to Canberra to visit and to my brother for calling me 'just because'. And finally to Andrew, my husband. Thanks for giving us the gift of our beautiful children Elaine and Graeme who gave me all the motivation I needed to finish this thesis...

Publications

Work related to this thesis has been published:

Lee J.R.M., Duxbury Z., Wang M.-B. (2013) Small RNAs and Transgenerational Epigenetic Variation. *American Journal of Plant Sciences*. 4(3); 641-646

Fang Y.-Y., Smith N.A., Zhao J.-H., **Lee J.R.M.**, Guo H.-S., Wang M.-B. (2014) Cloning and profiling of small RNAs from Cucumber mosaic virus satellite RNA. *Methods in Molecular Biology*. 1236:99-109. DOI 10.1007/978-1-4939-1743-3_9

Abstract

RNA-directed DNA methylation (RdDM), directed by 24 nt small interfering RNA (siRNA), is a plant-specific *de novo* cytosine methylation pathway responsible for natural suppression of repeat sequences such as transposons and centromeric repeats. RdDM can be counteracted by DNA demethylation, which in *Arabidopsis* is catalyzed by four DNA demethylases; Demeter (DME), Repressor of silencing (ROS1) and Demeter-like 2 and 3 (DML2, DML3). Early studies in our laboratory have suggested DNA demethylases and RdDM work together to regulate plant defense-related genes and mediate *Arabidopsis* resistance against the fungal pathogen *Fusarium oxysporum*. How these two pathways interact to regulate defense-related genes remained unclear. The aim of this study is to investigate how RdDM and DNA demethylases play a role in *Fusarium* resistance by analysing changes in small RNA (sRNA) accumulation, DNA methylation, and gene expression in wild-type *Arabidopsis* (Col-0) and the DNA demethylase mutant *rdd* (*ros1 dml2 dml3*) with and without *Fusarium* infection.

Whole-genome sRNA deep sequencing plus northern blot hybridisation detected a large number of differentially accumulated (DA) siRNAs Col-0 and *rdd* plants. A large majority of these DA 24 nt siRNAs occurred in gene bodies, the 3kb flanking regions or transposable elements (TEs). These results indicate that DNA demethylases are involved in the accumulation of RdDM associated siRNAs, and that RdDM and DNA demethylases function co-ordinately to regulate gene expression through targeting TE sequences.

Comparative analysis of gene expression and DNA methylation between Col-0 and *rdd*, using whole genome mRNA and bisulfite sequencing, supported by qRT-PCR analysis and existing DNA methylation data, provided evidence that DA siRNAs participate in DNA demethylase mediated regulation of nearby defense related gene targets. These analyses revealed;

- i) genes differentially expressed (DE) between Col-0 and *rdd* are enriched for stress response functions;
- ii) siRNAs and DNA methylation co-exist near transcription start sites (TSS) of the DE genes in the *rdd* background but are absent in Col-0; and
- iii) the new methylation in *rdd* spreads towards the TSS, and is immediately adjacent to the end of TEs that are methylated in Col-0 as well as in *rdd*.

These features suggest that TEs in the promoters are the target of RdDM and have the potential to induce spread of RdDM to adjacent sequences near TSS resulting in gene repression. DNA demethylases are required to counter this effect ensuring active gene expression.

In contrast to the Col-0 vs *rdd* comparison, of which a large proportion of DE genes have differential siRNA and DNA methylation, a very small proportion of the Fusarium responsive genes are associated with differential siRNA and DNA methylation between mock treated and Fusarium infected plants. This suggests the Fusarium responsive expression pattern is not directly controlled by RdDM changes. I propose the Fusarium response is determined by the intrinsic property of cis-elements present in stress response gene promoters, but their accessibility by transcription factors can be inhibited by DNA methylation. DNA demethylases therefore are required to maintain a low-level methylation state to regulate gene transcription activity.

Previous studies state the DME demethylase is expressed and functional only in developing seed, and therefore Ros1, DML2 and DML3 should account for the bulk of demethylase activity in vegetative tissues. However, my mRNA sequencing data and our previous microarray expression data both indicated a high level of *DME* expression in vegetative tissues. This suggested the *rdd* mutant may retain some DME DNA demethylase activity in vegetative tissues, which would affect the interpretation of our data. To address this issue, I generated *DME* knockdown transgenic lines in the *rdd* background using hairpin RNA technology, and assayed the transgenic plants for Fusarium resistance. These *ros1-dml2-dml3-hpDME* plants exhibited enhanced Fusarium susceptibility than *rdd*, suggesting DME contributes to DNA demethylase activity in vegetative tissues. These transgenic materials will be valuable materials for future functional studies of DNA demethylases in Arabidopsis.

Important Abbreviations

AGO: argonaute

CMT3: chromomethylase 3

DA: differential abundance

DCL: dicer like

DE: differential expression

DME: Demeter DNA demethylase mutant

DML: Demeter like DNA demethylase mutant

DMR: differentially methylated region

DRM2: domains re-arranged methyltransferase 2

FOX: *Fusarium oxysporum*

MET1: methyltransferase 1

miRNA: micro RNA

nat-siRNA: natural antisense siRNA

phasRNA: phased, secondary, small interfering RNAs

PTGS: post transcriptional gene silencing

rdd: triple DNA demethylase mutant

RdDM: RNA directed DNA methylation

RDR: RNA-dependent RNA polymerase

RISC: RNA induced silencing complex

ROS1: repressor of silencing 1 DNA demethylase mutant

sRNA: small RNA

siRNA: small interfering RNA

tasiRNA: trans-acting siRNA

TE: transposable element

Table of Contents

Statement of Authorship	ii
Acknowledgements	iii
Publications.....	iv
Abstract	v
Important Abbreviations	vii
Chapter 1: Introduction	12
1.0 General Introduction	13
1.1 RNA Silencing Pathways	14
1.2 Post-Transcriptional RNA Silencing Pathways.....	15
1.2.1 miRNA Pathway.....	15
1.2.2 Post-Transcriptional sRNA Pathway	15
1.3 RdDM and DNA Demethylation.....	17
1.3.1 Overview of RdDM	17
1.3.2 Maintenance of DNA Methylation	18
1.3.3 DNA Demethylation.....	21
1.3.3.1 DME DNA Demethylase.....	21
1.3.3.2 ROS1 DNA Demethylase	22
1.3.3.3 DML2 and DML3 DNA Demethylase.....	22
1.3.3.4 Known Functions of DNA Demethylases	23
1.3.3.5 <i>rdd</i> , a Triple DNA Demethylase Mutant	24
1.4 Alternative RdDM Pathways	25
1.5 Dynamics of DNA Methylation and its Influence on Gene Expression	26
1.5.1 Promoter Methylation and Gene Expression.....	26
1.5.2 Gene Body Methylation and Gene Expression.....	27
1.5.3 Spread of DNA Methylation	28
1.6 The Relationship between Changes in 24 nt siRNA, DNA Methylation and Gene Transcription	28
1.7 Influence of RdDM and DNA Demethylation on TE Expression	29
1.8 The Role of RdDM in the Stress Response of Plants	32
1.8.1 RdDM in Abiotic Stress Response.....	32
1.8.2 RdDM and DNA demethylation in Biotic Stress Response	33
1.8.3 miRNA and phasiRNA in Stress Response	34
1.9 The <i>Fusarium oxysporum</i> – <i>Arabidopsis thaliana</i> Phyto-Pathogen system	35

1.20 Thesis Aims	36
Chapter 2: Materials and Methods	37
2.0 Plant Materials, Virus and Fungal Strains.....	38
2.01 Plant Growth Conditions and Fungal and Virus Infections.....	38
2.02 Phenotypic Assays	39
2.1 Fungal Biomass Measurement	39
2.2 General Molecular Methods	40
2.21 Rapid CTAB DNA Extraction.....	40
2.22 Total RNA Extraction using Trizol	40
2.23 DNA and RNA Quantification and Visualisation	42
2.24 Phenol-Chloroform Extraction	42
2.25 Polymerase Chain Reaction (PCR)	42
2.3 Northern Blot Hybridization	42
2.31 Electrophoresis of Small/Large RNA using a Urea-Polyacrylamide (PAGE)/Formaldehyde-Agarose Gel	43
2.32 Hybridisation with Radioactive Probes	43
2.33 Membrane Wash, Exposure and Strip.....	44
2.34 Preparation of Radioactive Probes.....	44
2.4 General Molecular Cloning.....	45
2.41 PCR purification and Ligation	45
2.42 Transformation of DH5a <i>E. coli</i> cells	47
2.43 Colony Screening and Sequencing	47
2.5 Whole Genome Sequencing Library Construction and Sequencing.....	48
2.51 Small RNA and mRNA Library Construction and Sequencing.....	48
2.52 Bisulfite Sequencing Library Construction and Sequencing.....	49
2.6 Bioinformatics for the Whole Genome Sequencing Data	49
2.61 Bioinformatic Analysis of Small RNA Sequencing Data	49
2.62 Bioinformatic Analysis of mRNA Sequencing Data.....	51
2.63 Bioinformatic Analysis of Bisulfite Sequencing Data.....	51
2.64 General Bioinformatics.....	52
2.7 Quantitative Real-Time PCR	53
2.71 qRT-PCR Primer Design	53
2.72 DNase Treatment of RNA	53
2.73 cDNA Synthesis.....	53

2.74 Quantitative PCR	55
Chapter 3: Col-0 and DNA Demethylase Mutants Show Differential Phenotypic Responses and Small RNA Accumulation upon <i>Fusarium oxysporum</i> Infection	56
3.0 Introduction.....	57
3.1 Results	58
3.1.1 DNA Demethylase Mutants Show no Phenotypic Differences to Col-0 Under Normal Growth Conditions	58
3.1.2 DNA Demethylase Mutants Show Differential Phenotypic Differences to Col-0 in Response to <i>Fusarium</i> Infection	58
3.1.3 DNA Demethylase Mutants show Differential Phenotypic Responses to Viral Infection	62
3.1.4 Assessing Spraying as an Alternative way of <i>Fusarium</i> Inoculation.....	62
3.1.5 Quantitative Measurement of <i>Fusarium</i> Biomass in Infected <i>rdd</i> and Col-0 Plants.....	65
3.1.6 Col-0 and <i>rdd</i> Show Differential Small RNA Accumulation in Response to <i>Fusarium</i> Infection	67
3.2 Discussion	70
3.3 Conclusions	73
Chapter 4: Genome-wide Survey of Small RNA Accumulation in Col-0 and <i>rdd</i> upon <i>Fusarium oxysporum</i> Infection	74
4.0 Introduction.....	75
4.1 Results	77
4.1.1 Quality Assessment of sRNA Sequencing Data	77
4.1.2 Identification of Differentially Abundant sRNA using DESeq	79
4.1.3 The Majority of DA sRNAs Are Located In Or Within 3 kb of Protein Coding Genes	83
4.1.3.1 Distribution of DA sRNA across Arabidopsis Chromosomes	83
4.1.3.2 Distribution of DA sRNA across Arabidopsis Genomic Features	83
4.1.3.3 Distribution of Protein Coding Gene Associated DA sRNA along Gene Features ..	88
4.1.4 The Majority of TEs and Transposon Genes are not Targeted by DA sRNA.....	89
4.1.5 Analysis of Differential miRNA Accumulation	91
4.1.6 Verification of sRNA Expression Pattern by Northern Blot Hybridisation of DA miRNA	91
4.2 Discussion	94
4.3 Conclusions.....	97
Chapter 5: Identification of Differentially Expressed sRNA Target Genes Across the Genome	98
5.0 Introduction.....	99

5.1 Results	100
5.1.1 A large number of genes are potentially targeted by DA sRNAs	100
5.1.2 Identification of DE genes using mRNA sequencing.....	105
5.1.3 Identification of DE genes associated with DA sRNA	107
5.2 Discussion	115
5.3 Conclusions.....	118
Chapter 6: Whole Genome Identification of Differentially Expressed sRNA Target Genes with Differentially Methylated Regions	119
6.0 Introduction.....	120
6.1 Results	122
6.1.1 Identification of genes with DMRs	122
6.1.2 Identification of DMRs associated with DA sRNAs at 3 DPI	125
6.1.3 Identification of DE genes associated with DMRs.....	125
6.1.4 Identification of DE genes associated with both DA sRNA and DMRs	126
6.1.5 DE genes downregulated in <i>rdd</i> are associated with increased sRNA and DNA methylation in the promoter region	128
6.1.6 Fusarium induced gene expression changes show weak correlation with sRNA and DNA methylation changes.....	132
6.1.7 Expression pattern and DMRs of DE genes determined by whole genome sequencing can be validated	134
6.2 Discussion	135
6.3 Conclusions.....	138
Chapter 7: Conclusions and future prospects	139
Chapter 8: Bibliography	151
Chapter 9: Supplementary Figures and Tables	164

Chapter 1: Introduction

1.0 General Introduction

The classical definition of a complex heritable trait is a phenotype influenced by alleles of multiple genes and the environment (Hartl and Jones, 2005). Studies since the 1990s on RNA directed gene silencing have shown that in addition to the conventional genetic factors, stable, long-term alterations in gene transcription are also possible through epigenetic modifications. Epigenetics can be defined as control of gene expression through changes in chromatin and DNA organization rather than changes in DNA sequences (Bender, 2004). Recent studies have suggested that rapid adaption to environment via epigenetic mutations of protein-coding gene targets can occur at high frequencies not possible with classical genetics (Johannes et al., 2009). As environmental changes, such as those associated with global warming, occur at an increasing pace it is imperative that scientists understand the potential epigenetic responses of flora and fauna.

One of the key epigenetic silencing pathways in plants is RNA directed DNA methylation (RdDM). Initiated by 24 nucleotide (nt) small interfering RNA (siRNA), the RdDM pathway plays an essential role in maintaining genome stability by repression of transposable elements (TEs) and repetitive sequences and in regulating gene expression. De novo DNA methylation induced by RdDM can be partly maintained during DNA replication by maintenance DNA methyltransferases in plants. Active expression can be reintroduced through a dynamic process involving DNA demethylases that remove excess DNA methylation marks. For example stress response genes may be silenced until a defence response is required when DNA demethylases can help to initiate gene transcription.

Numerous studies have paid attention to siRNA and DNA methylation levels in various RdDM mutants or under various stress conditions (Stroud et al., 2013, Lister et al., 2008, Downen et al., 2012). However, the relationship between siRNA, DNA methylation and gene expression remains poorly understood. The plant epigenetics field is also just beginning to explore the relationship between epigenetics and disease resistance in plants (Downen et al., 2012, Yu et al., 2013, Kim et al., 2010a). An understanding of this relationship may lead to development of novel technology for disease control. In this thesis I aim to explore the role of siRNA and DNA demethylation in regulating the complex molecular responses of plants to infection by the fungal pathogen *Fusarium oxysporum*. The goal of this study is to enhance our understanding of the interaction between RdDM and DNA demethylases and their role in *Fusarium oxysporum* disease response in *Arabidopsis thaliana*.

Besides siRNA and DNA methylation, other epigenetic modifications such as histone 3 lysine 27 methylation (H3K27me) and histone lysine 9 methylation (H3K9me) influence chromatin structure and affect gene expression. Studies now show that RdDM, DNA demethylases and histone methylation may work together to regulate gene expression (Fan et al., 2012, Qian et al., 2012, Bernatavichute et al., 2008). However, the focus of this thesis is on the RdDM and DNA demethylation pathway and therefore histone modifications will not be discussed in detail. It is however an avenue which must be explored to gain a complete understanding of the impact the RdDM and demethylation pathway have on gene expression in response to stress.

This literature review will; i) outline the RNA silencing pathways, ii) provide an overview on RdDM, maintenance DNA methylation and DNA demethylases, iii) discuss the role of small RNA in defense; and iv) introduce the *Fusarium oxysporum* – *Arabidopsis thaliana* pathogen-plant system used in this thesis.

1.1 RNA Silencing Pathways

Various small RNA silencing pathways exist in eukaryotes that may influence gene expression at the transcriptional and post-transcriptional level. There are three major classes of small RNA (sRNA); microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs) and small interfering RNAs (siRNAs). MiRNAs are found in both animal and plant species, piRNAs are specific to animals although functionally analogous to the RdDM associated siRNAs in plants (to be discussed later).

piRNAs are not present in plants or fungi and are best described in *Drosophila melanogaster*, *Caenorhabditis elegans* and mice (Matzke and Mosher, 2014). piRNA are derived from germ line single-stranded RNA and function to inactivate transposons by guiding repressive chromatin modifications. siRNAs in contrast are present only in plants, fungi and the fission yeast *Schizosaccharomyces pombe*. They are processed from long double-stranded RNA (dsRNA) and can direct both post-transcriptional and transcriptional gene silencing. Both piRNA and siRNA rely on the core RNAi proteins: DICER (DCR) processes long double-stranded RNAs (dsRNA) into siRNA/piRNA and ARGONAUTE (AGO) proteins guide sRNA, sequence-specific chromatin modifications (described in more detail below).

In plants, dsRNA is synthesized from single-stranded RNA by a RNA-DIRECTED RNA POLYMERASE (RDR). Arabidopsis encodes six RDRs, of which RDR1, RDR2 and RDR6 have been shown to function in generating dsRNA precursor for the various small RNA pathways. RDR1 is involved in the antiviral siRNA pathway, RDR6 in the trans-acting siRNA (tasiRNA) as well as antiviral siRNA pathways, whereas RDR2 is required for the RdDM pathway. dsRNA is processed

to sRNA by a ribonuclease (RNase) III-like endonuclease termed DICER. Arabidopsis encodes four DICER-LIKE (DCL) proteins responsible for processing 21 nt (DCL4) and 22 nt (DCL2) posttranscriptional silencing siRNAs, 24 nt (DCL3) RdDM associated siRNAs, and 20-22 nt microRNA (DCL1) (Lee et al., 2013). Mature siRNAs and miRNAs interact with ARGONAUTE (AGO) proteins to form an RNA-induced silencing complex (RISC), and guide the complex to homologous target mRNA or DNA inducing posttranscriptional or transcriptional gene silencing (Wang et al., 2011). Ten AGO family proteins have been described in Arabidopsis. AGO1 has been shown through the use of *ago1* null mutants to be the main effector for the miRNA pathway (Vaucheret et al., 2004). Viral siRNA, transgene siRNA and tasiRNA have also been shown to associate with AGO1 complexes (Wang et al., 2011). AGO4 associates with the 24 nt siRNA to induce RdDM (Vaucheret, 2008).

1.2 Post-Transcriptional RNA Silencing Pathways

1.2.1 miRNA Pathway

In the miRNA pathway, primary miRNA transcript is generated by RNA polymerase II from miRNA genes, which folds into a stem-loop structure and is processed to 21 nt miRNA by DCL1. Mature miRNA is loaded to AGO1 to form RISC, which directs RNA cleavage or translational repression of mRNA targets (Figure 1.1). miRNA normally silences genes in trans, namely genes encoded in genomic locations different to the miRNA gene (Carthew and Sontheimer, 2009). Many miRNA targets in plants are regulatory genes such as transcription factor genes. Some miRNAs of 22-nt size trigger the biogenesis of trans-acting siRNA (tasiRNA) or phased secondary siRNA (phasiRNA) (see below).

1.2.2 Post-Transcriptional sRNA Pathway

TasiRNA is a commonly recognized minor class of plant endogenous sRNA with 21-nt size and acts post-transcriptionally in trans to regulate mRNA involved in plant development, metabolism and biotic/abiotic stress responses (Felippes and Weigel, 2009, Hu et al., 2011, Chen et al., 2007). tasiRNA is derived from *TAS* gene transcript, which is first cleaved by a 22-nt miRNA into single stranded RNA (ssRNA) fragments, which are then converted to dsRNA by RDR6 following by DCL4 processing, giving 21-nt phased tasiRNAs (reviewed in Fei et al. 2013). *Arabidopsis thaliana* has four major families of *TAS* genes targeted by three miRNA (Table 1.1).

Another group of sRNAs, called phasiRNAs, resemble tasiRNAs as their biogenesis follows the same pathway as tasiRNA. Protein coding genes, including genes encoding Ca^{2+} -ATPase (Wang et al., 2011), transcription factors (Si-Ammour et al., 2011) and pentatricopeptide repeat proteins (Howell et al., 2007) are targets of phasiRNA.

Figure 1.1: miRNA pathway

A primary miRNA transcript that forms a hpRNA molecule is cleaved by DCL1 in the nucleus to produce a shorter precursor miRNA dsRNA molecule. The miRNA duplex has 2 nucleotide 3' overhangs which are methylated by HUA ENHANCER1 (HEN1). HASTY (HST) exports the miRNA duplex to the cytoplasm where mature single-stranded miRNA is loaded onto ARGONAUTE, AGO1 or AGO10, (the catalytic centre of plant RNA-induced silencing complexes, RISC) to mediate transcript repression.

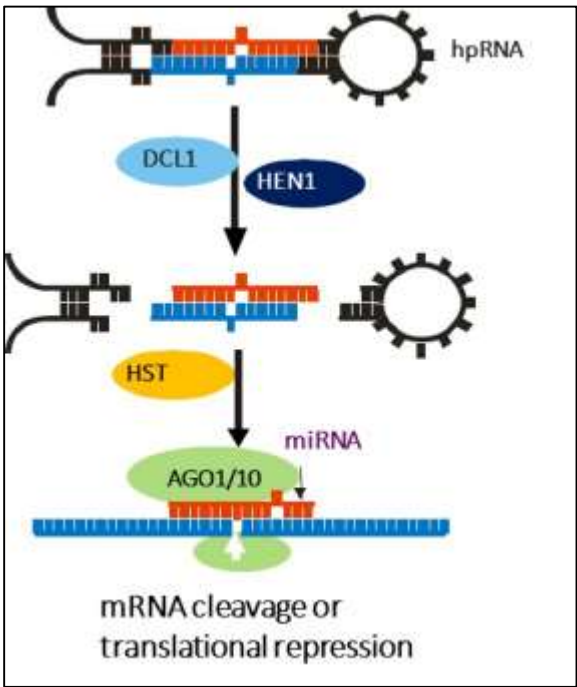


Table 1.1: Known tasiRNA genes and target in Arabidopsis

Updated data from Felippes and Weigel (2009).

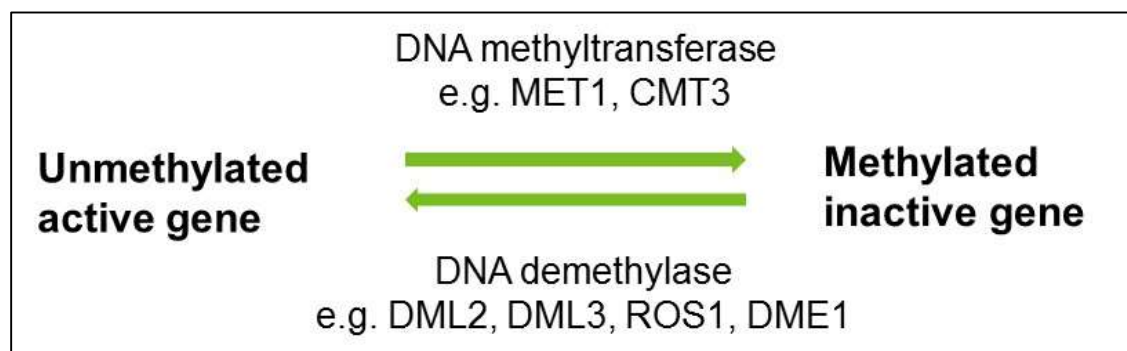
<i>TAS</i> gene	Targeted by	Production of tasiRNAs from the <i>TAS</i> transcript is triggered by miRNAs__of the cleavage site
<i>TAS1</i>	miR173	Downstream
<i>TAS2</i>	miR173	Downstream
<i>TAS3</i>	miR390	Upstream
<i>TAS4</i>	miR828	Downstream

1.3 RdDM and DNA Demethylation

1.3.1 Overview of RdDM

DNA methylation level and transcriptional activity is under dynamic control by DNA methyltransferase and demethylase enzymes (Figure 1.2). The RdDM pathway is not yet completely understood, but mutagenesis studies in *Arabidopsis* have established the key molecular machinery involved in RdDM (Xie and Yu, 2015, Eamens et al., 2008) (Figure 1.3). Briefly, RdDM begins with the biogenesis of 24-nt siRNAs: methylated or repetitive DNA acts as a template for DNA-dependent RNA POLYMERASE IV (Pol IV) to transcribe ssRNA. RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) then converts the ssRNA to double stranded (dsRNA), which is subsequently processed by DCL3 into 24 nt siRNA duplex. ARGONAUTE 4 (AGO4) then binds to a single strand of the siRNA duplex to form RISC, which then interacts with nascent RNA transcript synthesized by DNA-dependent RNA polymerase V (Pol V). This AGO4-Pol V complex then recruits other downstream RdDM factors including DOMAINS RE-ARRANGED METHYLASE 2 (DRM2), inducing *de novo* cytosine DNA methylation to the target DNA locus. RdDM-induced cytosine DNA methylation occurs in all sequence contexts, including CG, CHG and CHH where H = A, T or C.

Figure 1.2. Overview of DNA methyltransferase and demethylase enzymes



1.3.2 Maintenance of DNA Methylation

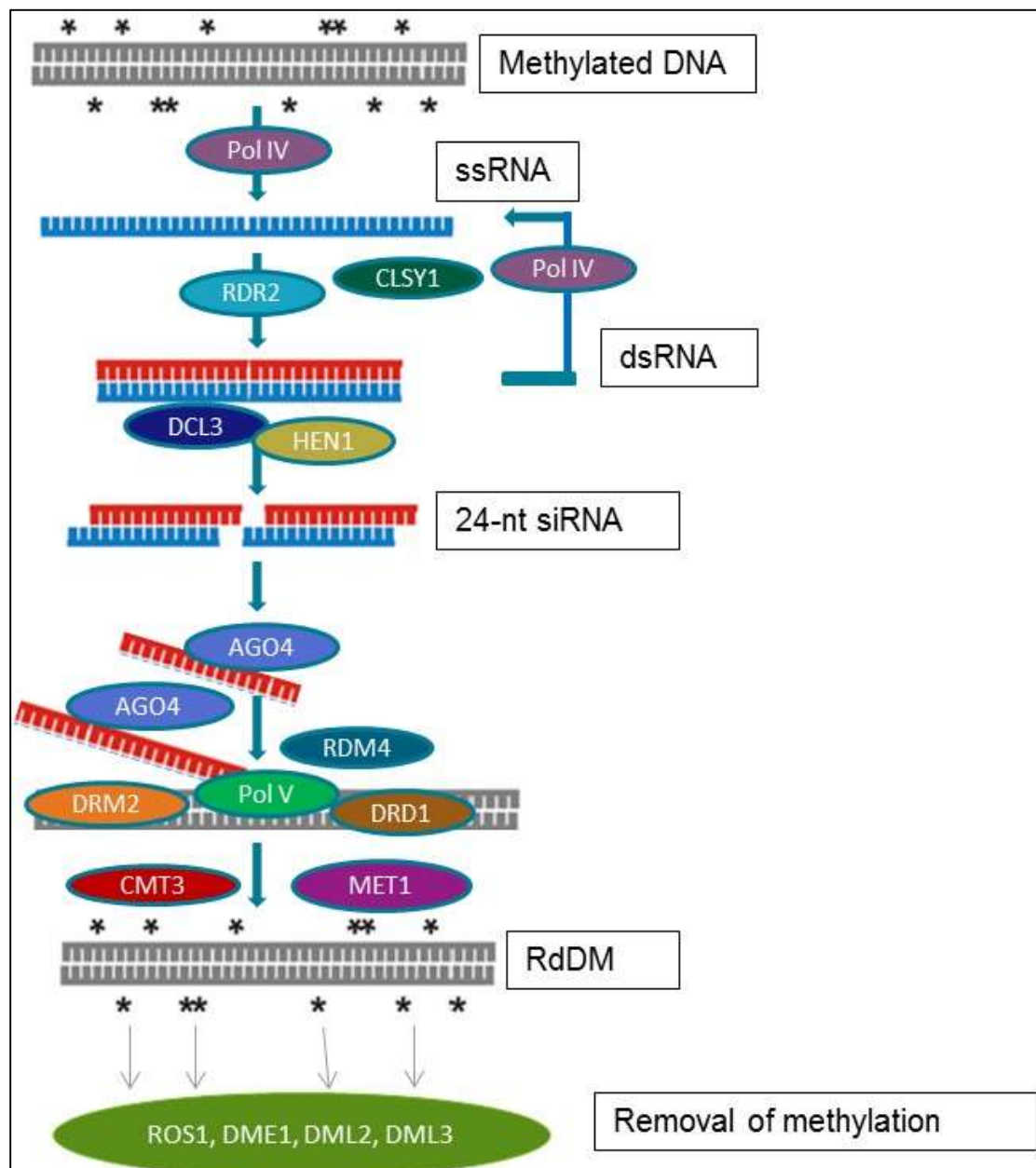
As mentioned earlier, RdDM in plants can induce cytosine methylation in all sequence contexts, including CG, CHG and CHH (Schmitz et al., 2011, He et al., 2009). Once induced, methylation in the symmetric contexts CG and CHG can be maintained during DNA replication by Methyltransferase 1 (MET1) and Chromomethylase 3 (CMT3) respectively. Methylation at CHH sites cannot be maintained during DNA replication and needs to be continuously generated *de novo* by RdDM through the function of DRM2 (Figure 1.3).

MET1 can act independently of CMT3 and DRM2 in maintaining CG methylation (Huang et al., 2014, Singh et al., 2008). However, MET1 seems to cooperate with CMT3 and DRM2 which contributes to the maintenance of non-CG methylation. For instance, loss of MET1 function in the *met1* mutant results in not only a genome-wide loss of CG methylation, but also some loss of CHG and CHH methylation at certain sites. In addition, MET1 also function in conjunction with the DNA demethylase REPRESSOR OF SILENCING 1 (ROS1) to maintain DNA methylation. *ROS1* expression is positively regulated by DNA methylation and repressed in *met1* (Huettel et al., 2006). The *Ros1* locus can therefore ensure epigenetic stability by acting as an epigenetic rheostat (Williams et al., 2015). The *met1* mutants have been shown to have phenotypic reversion through successive generations (Mathieu et al., 2007). Low levels of DNA methylation are seen in newly generated homozygous *met1* mutants but by the third or fourth generation the *met1* progeny have no DNA methylation defect (Mathieu et al., 2007). While residual DNA methyltransferase activity from DRM2 and CMT3 may contribute to the methylation recovery, a lack of ROS1 expression in the mutant are believed to be the principal cause of this reversion phenotype. Future experiments which overexpress ROS1 in *met1* mutants could be performed to see if DNA methylation recovery can be abolished or reduced in *met1* progenies.

MET1 has also been shown to work in collaboration with Variant in Methylation 1 (VIM1) to coordinate DNA methylation (Kim et al., 2014, Shook and Richards, 2014). VIM1 target genes had severe DNA hypomethylation (decrease in methylation) across all sequence contexts, and showed transcriptional activation in the *vim1/2/3* triple mutant. In the *met1* background VIM1-binding capacity to target genes was reduced indicating CG methylation by MET1 is recognized by VIM1.

Figure 1.3: Major enzymes of the RNA directed DNA methylation (RdDM) pathway

Methylated or repetitive DNA acts as a template for single stranded RNA (ssRNA) to be transcribed by RNA POLYMERASE IV (Pol IV). Aberrant RNA is then converted to double stranded RNA (dsRNA) by RNA-DIRECTED RNA POLYMERASE 2 (RDR2). Additional aberrant RNA molecules can be formed by Pol IV in a self-perpetuating loop. The CLASSY1 (CLSY1) protein is a putative SNR1-like chromatin remodelling factor suggested to function together with POL IV or RDR2 in siRNA biogenesis. The dsRNA is diced by DICER LIKE 3 (DCL3) into 24-nucleotide siRNA duplexes and methylated by HUA ENHANCER1 (HEN1) to stabilize the siRNAs. ARGONAUTE 4 (AGO4) then bind to one strand of the siRNA duplexes and interact with nascent RNA transcript synthesized by RNA POLYMERASE V (Pol V) to direct cytosine methylation in the DNA by DOMAINS RE-ARRANGED METHYLASE2 (DRM2) and the chromatin remodelling protein, DEFECTIVE IN RNA-DIRECTED DNA METHYLATION1 (DRD1). RNA DIRECTED DNA METHYLATION 4 (RDM4) acts as a regulatory factor for RNA polymerase V. The *de novo* methylation at the CG and CHG sites can be maintained by METHYLTRANSFERASE (MET1) and CHROMOMETHYLASE3 (CMT3) respectively. Methylation levels can also be affected by the DNA demethylases REPRESSOR OF SILENCING (ROS1), DEMETER (DME) and DEMETER-LIKE (DML2 and DML3) removing the methyl group.



CMT3 functionally primarily to maintain CHG methylation with *cmt3* mutants showing genome-wide decreases in CHG methylation (Tompas et al., 2002, Lindroth et al., 2001, Bartee et al., 2001). CMT3 has shown preferential methylation of transposons, even with single copies within the genome (Tompas et al., 2002). CHG methylation appears to be maintained through a reinforcing loop involving histone (H3K9) and DNA methylation (Bernatavichute et al., 2008).

DRM2, the final major DNA methylase, generates CHH methylation in *de novo* RdDM. No maintenance of CHH methylation occurs during DNA replication, however CMT3 and DRM2 have been shown to jointly maintain CHH methylation at some loci (Cao et al., 2003). DRM1/2 may also regulate CHG methylation in a non-redundant fashion as found in two independent studies (Stroud et al., 2013, Cao and Jacobsen, 2002). A forward genetic screen (Naumann et al., 2011) supported the previous study that DRM2 is the major *de novo* DNA methyltransferase. Further work suggests that an evolutionarily conserved yet catalytically inactive DNA methyltransferase, DRM3, acts to promote Pol V stabilization or Pol V transcriptional elongation through physical interaction with Pol V (Zhong et al., 2015). Additionally, the *drm3* mutant shows moderate DNA hypomethylation across all sequence contexts and a reduction in small RNA abundance (Zhong et al., 2015, Henderson et al., 2010). DRM3 presumably plays a supportive role to DRM2 under normal conditions although further studies should be carried out to investigate the possible role of DRM3 under stress conditions.

Recently, it has been proposed that INVOLVED IN DE NOVO 2 (IDN2), an RNA-binding protein, is required for DNA methylation establishment (Ausin et al., 2012). Known to bind to dsRNA, IDN2 is thought to act downstream of the upstream siRNA biogenesis steps in the RdDM pathway. It was found that either IDN2-LIKE1 (IDNL1) or IDN2-LIKE 2 (IDNL2) is required in cooperation with IDN2 to complete DRM2-mediated genome methylation. Only minimal activity of DRM2 can occur in IDN2 complex mutants. Owing to the complexity of the RdDM pathway it is likely that more proteins will be uncovered that help maintain DNA methylation.

Maintenance of DNA methylation for transposons and repetitive sequences does not always occur through DNA methyltransferases. Indirect maintenance can arise through Decreased DNA Methylation 1 (DDM1), a switch2/sucrose nonfermentable2 chromatin-remodeling protein (Gendrel et al., 2002). The *ddm1* mutant has DNA hypomethylation across all three cytosine contexts. The lost methylation marks are inherited and result in various developmental abnormalities (Saze et al., 2012). Mutant studies revealed the DNA methyltransferase CMT3 was necessary for DDM1 DNA methylation maintenance however RdDM was indispensable (Sasaki et al., 2012).

Another minor DNA methyltransferase, Chromomethylase 2 (CMT2), has been shown to methylate CHH context cytosines predominantly within the body of long transposable elements (Shen et al., 2014). Plants of *cmt2* mutant have also been shown to have improved heat-stress tolerance implying this alternative methyltransferase may be involved in DNA methylation abiotic stress defence.

A key whole genome bisulfite methylation study was undertaken on leaf tissue across 86 *Arabidopsis thaliana* methylation or methylation-related mutants (Stroud et al., 2013). The main findings showed that genic and heterochromatic methylated sites may be maintained by different pathways. For example, genic methylation is unable to be restored in the *met1* mutant by the RdDM pathway. In contrast, siRNA (indicating RdDM) exclusively associated with heterochromatin is required for DNA methylation restoration in *ddm1* mutants (Teixeira et al., 2009). DDM1 appears to maintain DNA methylation specifically at heterochromatin as genic methylation remains largely intact in a *ddm1* mutant (Stroud et al., 2013).

1.3.3 DNA Demethylation

Maintenance of gene silencing occurs when the stable epigenetic cytosine DNA methylation mark is recognized by the plant. As a reversible modification of DNA methylation, DNA demethylation may be required to initiate specific gene expression or to reset the epigenetic genome in response to environment or during development. In mammals, DNA methylation is removed immediately after fertilization through passive DNA demethylation (Zhu, 2009, Lee et al., 2013). Unlike mammals, DNA methylation can be transgenerationally inherited in plants with DNA demethylation occurring by passive and active demethylation mechanisms. Passive DNA demethylation is initiated when the maintenance methyltransferases are inactive in the cell cycle immediately following DNA replication. The newly synthesized strand retains the unmethylated state. Active demethylation occurs via DNA demethylases, a subfamily of DNA glycosylases, through the base excision repair mechanism where DNA glycosylase enzymes remove 5-methylcytosine and replace it with an unmethylated cytosine. In plants four demethylases are known; DEMETER (DME), REPRESSOR OF SILENCING (ROS1) and DEMETER-LIKE 2 and 3 (DML2, DML3).

1.3.3.1 DME DNA Demethylase

DME is shown to be expressed in maternal specific tissues during seed development and in the vegetative cell of male gametophytes during pollen development, and is involved in endosperm and embryo development (Choi et al., 2002). Genomic imprinting is an epigenetic phenomenon where one parental allele may be imprinted (silenced) and the other parental allele is

expressed. Mutation of *DME* is found to disrupt gene imprinting in developing Arabidopsis seed and results in seed abortion. For instance, *MEA*, an imprinted gene, is actively expressed in the endosperm (maternal tissues) but silenced in the embryo of developing seed (Penterman et al., 2007b). In somatic tissues *MEA* is methylated, but in the endosperm of developing seed, DME excises 5-methylcytosine establishing hypomethylation and activating *MEA* transcription. It has been shown that DME interacts with the DNA methyltransferase MET1 to regulate the imprinting of *MEA* (Xiao et al., 2003) suggesting that both DNA methylation maintenance and demethylation are important for imprinting.

1.3.3.2 ROS1 DNA Demethylase

ROS1 is a homolog of DME with widespread expression in all tissues (Gong et al., 2002). Transgenic studies of a reporter gene driven by the promoter of a cold regulated gene (RD29A:LUCIFERASE) found ROS1 to be required for active transcription. In the *ros1* mutant RD29A:LUC and the endogenous RD29A gene were transcriptionally silenced due to heavy methylation. It is therefore clear ROS1 is required to suppress methylation. In various DNA methylation mutants (*met1*, *pol IV*, *rdr2*, *dcl3* and *drm2*) the *ROS1* mRNA level is very low (Mathieu et al., 2007, Huettel et al., 2006) suggesting that *ROS1* expression responds to methylation levels. Presumably *DML2* and *DML3* expression is also sensitive to DNA methylation but due to their low expression in wild-type plants it is more difficult to determine (Mathieu et al., 2007). *ROS3*, an RNA-binding protein involved in DNA demethylation is also regulated by DNA methylation levels (Zheng et al., 2008). In a *ros1* mutant, *ROS3* expression is enhanced and vice versa indicating all the demethylation components, not just demethylases, are responsive to DNA methylation. *ROS1* can be suppressed by Zinc finger and OCRE domain-containing protein 1 (ZOP1), a nucleic acid-binding protein required for RdDM and pre-mRNA splicing (Zhang et al., 2013).

1.3.3.3 DML2 and DML3 DNA Demethylase

Until recently little was known about the function of these two Demeter-like (DML2 and DML3) demethylases. A genome-tiling microarray was used to identify DNA demethylated targets of DML2 and DML3 in Col-0 (Penterman et al., 2007b). It was found that greater than 80 % of the 180 DML demethylated loci were within the 5' and 3' gene ends. The DML genes therefore function in a locus specific manner to protect genes from DNA methylation that may mark protein coding genes for silencing. DNA methylation analysis of the long terminal repeat retroelement *AtGP1* and the flowering gene *FWA* using bisulfite sequencing found that *dml2* and *dml3* mutants had low or no methylation where wild type plants had strong methylation (Ortega-Galisteo et al., 2008). RT-PCR analysis showed that hypomethylation in the *dml* mutants

were not due to a decrease in *MET1* expression. These results suggest that DML2 and DML3 are important for methylation removal but also for maintenance of highly methylated sites.

1.3.3.4 Known Functions of DNA Demethylases

Despite the belief that DME functions specifically in developing seed, recent studies indicate that all four Arabidopsis DNA demethylase genes, including *DME*, are expressed in all plant organs. RT-PCR analysis of T-DNA loss of function alleles detected *ROS1*, *DML2* and *DML3* expression in roots, stems, leaves and inflorescences (Penterman et al., 2007b). An independent study (Ortega-Galisteo et al., 2008) measured expression of all four demethylase genes (including *DME*) in cauline leaves, flowers, stems, siliques and roots using RT-PCR. The study found *DML2* and *DML3* expression in all tissues except mature seeds therefore supporting the previous study by Penterman and colleagues (2007). The results for *ROS1* expression was in agreement with an earlier *ROS1:GUS* transgenic study by the same lab (Gong et al., 2002), showing expression in all tissues analyzed. Significantly, *DME* expression was detected in cauline leaves, flowers, stems and roots but at very low levels in siliques and seeds (Ortega-Galisteo et al., 2008). The *DME* expression pattern challenged an earlier report stating that *DME* was primarily expressed in the central cell of the female gametophyte (Choi et al., 2002). DME expression will be discussed further in Chapter 6.

DNA demethylase activity has been measured to determine if a preference for methylated cytosine context exists. An *in vitro* assay was used to detect oligonucleotides bearing 5-methylcytosine (Penterman et al., 2007b). DML3 and DML2 were found to actively interact with all contexts (CG, CHG and CHH) but DML2 had weaker activity at the CG site. No activity in either DML enzyme was found with non-methylated oligonucleotides. A similar experiment which expressed DML2 and DML3 cDNA fused to a maltose-binding protein in *E. coli* also found DML3 activity across the three cytosine contexts similar to DME and ROS1 activity (Ortega-Galisteo et al., 2008). This study also indicated DML3 has a preference for the CG and CHG sequence contexts, and DML2 has low, but measureable activity across the three cytosine contexts. A preference for demethylation of CHG over CG sites was found *in vitro* for ROS1 (Agius et al., 2006). An opposite result, showing a preference for CG over CHG sites by ROS1 and DME, was observed by Morales-Ruiz et al. (2006). Collectively these findings indicate that the demethylases do not have strong sequence preferences and their targets may differ depending on the tissue or gene. Indeed the ROS1 demethylase removes methylated cytosines *in vivo* from all sequence contexts in the *RD29A* promoter but only removes CHG and CHH contexts in most other target loci (Zhu et al., 2007).

1.3.3.5 *rdd*, a Triple DNA Demethylase Mutant

Two *rdd* (*ros1 dml2 dml3*) *Arabidopsis* mutants were developed independently using different T-DNA insertion lines, by Robert Fischer (Penterman et al., 2007b) and Jian-Kang Zhu's (Le et al., 2014) laboratories. The *rdd* mutant from the Zhu lab was used in this thesis as it had a genetic background of Col-0. No mutant phenotypic characterization studies had been published on the *rdd* mutant despite the statement that no developmental defects exist under normal conditions in *rdd* (Le et al., 2014). This is addressed in Chapter 3 with comparisons made between *rdd*, the single demethylase mutant *ros1*, *dml2* and *dml3* and wild-type Col-0.

Beside DNA demethylases, DNA demethylation can be regulated by DNA methylation factors, as described above in the *met1* and *ros1* comparison (Huettel et al., 2006, Mathieu et al., 2007). Recent studies also demonstrate a role of other enzymes such as INCREASED DNA METHYLATION 1 (IDM1) in DNA demethylation (Qian et al., 2012). Higher levels of DNA methylation in all cytosine contexts at the 3' region of the gene AT1G26400 (a FAD-binding Berberine family protein) was found in the *ros1-4* and *idm1* mutants compared to Col-0. The non-additive nature of *ros1* and *idm1*, found through double-mutant analysis, means that these two genes function in the same genetic pathway but do not colocalize. IDM1 controls methylated DNA levels to create an environment permissible for DNA demethylase function. In the DNA demethylation triple mutant, *rdd*, an increase in DNA methylation at tasiRNA loci (TAS 1, 2 and 3) is seen (Lister et al., 2008). This indicates that tasiRNA genes are targeted for *de novo* methylation when DNA demethylases are not present and therefore may have silenced expression. The tasiRNA will be briefly examined in Chapter 4.

DNA methylation studies have compared *rdd* and Col-0 and found only a small number of loci showing DNA methylation differences (Lister et al., 2008, Penterman et al., 2007b). These loci generally have hypermethylation in *rdd* compared to Col-0 across all three cytosine contexts near genes or at the 5' and 3' untranslated regions (UTRs) (Penterman et al. 2007, Lister et al. 2008, Qian et al. 2012, Stroud et al. 2013). This coincides with 24 nt siRNA commonly concentrated very close to gene ends (Gent et al., 2013). Limited overlap in the demethylation target genes was found among the independent studies possibly due to differences in growth conditions or plant tissues used.

mRNA sequencing analysis showed that 0.53 % of annotated *Arabidopsis thaliana* genes had altered transcript abundance in *rdd* compared to Col-0 (Lister et al. 2008). Of these genes, there were 20 % fewer transposons in *rdd* than in Col-0 that had mRNA reads. Hypermethylated transposons were also observed in *rdd* which were close to genes and often accompanied by

increased sRNA abundance. This indicates that in the wild-type Col-0 a subset of transposons is actively maintained in a demethylated state by DNA demethylases to protect neighbouring genes from a local sRNA increase which could potentially silence the genes.

1.4 Alternative RdDM Pathways

De novo cytosine methylation can also be induced in plants by non-canonical RdDM mechanisms induced by 21 nt siRNA associated with posttranscriptional gene silencing (PTGS) pathways. It has been recently shown that *de novo* DNA methylation, but not maintenance of DNA methylation, can be initiated at some TE loci through RDR6-dependent 21-nt siRNA, and involving AGO2, Pol V and DRM2 (Nuthikattu et al., 2013). This RDR6-dependent DNA methylation triggers 24 nt siRNA production for function in the canonical RdDM pathway to allow maintenance of DNA methylation at the target TE loci. RDR6-dependent siRNA was only produced when TEs were transcriptionally active. Another study showed that when DCL2 and DCL4 were saturated the RDR6-dependent dsRNA were processed into 24 nt siRNA for canonical RdDM to silence a newly integrated Evade retrotransposon (Mari-Ordonez et al., 2013).

The involvement of 21-nt siRNAs in *de novo* RdDM was revealed when studying NERD (NEEDED FOR RDR2-INDEPENDENT DNA METHYLATION), a GW repeat- and PHD finger-containing protein that binds AGO2 and Histone H3 (Pontier et al., 2012). Termed the NERD-dependent RdDM pathway it requires RDR1/6, 21 nt siRNA, AGO2, SUPPRESSOR OF GENE SILENCING 3 (SGS3), SILENCING DEFECTIVE 3 (SDE3), SDE5 and canonical RdDM Pol IV and Pol V components (Pontier et al., 2012, Garcia et al., 2012). In addition, genome-wide methylation analysis of 86 Arabidopsis methylation-associated mutants (Stroud et al., 2013) revealed that mutants of PTGS factors, such as *RDR1* and *DCL2/DCL4*, showed significant hypomethylation. In rice an *rdr1* mutant had some sRNA clusters associated with DNA methylation changes along with altered expression of miRNA target genes (Wang et al., 2014b). Despite these changes, the rice *rdr1* mutant did not display phenotypic differences to wild type plants. Changes were however seen under abiotic stress conditions suggesting that the 21-nt siRNA-dependent RdDM pathway is utilized during abiotic stress response.

Gene expression can also be repressed by miRNA or tasiRNA triggered RdDM. In miRNA-triggered DNA methylation DCL3 is able to produce 24 nt miRNA which can interact with AGO4 to direct DNA methylation at loci from which the miRNA transcript is derived as well as at miRNA target loci (Wu et al., 2010, Chellappan et al., 2010). A 200 bp bidirectional extension of DNA methylation at the miRNA binding site is possible through miRNA-triggered DNA methylation

(Hu et al., 2014). In a tasiRNA-triggered RdDM pathway RDR6-dependent dsRNA are cleaved by DCL1 to produce siRNA, which is loaded to AGO4 or AGO6 to direct DNA methylation via canonical RdDM mechanisms for gene and chromatin regulation (Wu et al., 2012, Wu, 2013).

1.5 Dynamics of DNA Methylation and its Influence on Gene Expression

Limited studies have looked at changes in sRNA abundance and associated gene transcription (Lister et al., 2008, Groszmann et al., 2011) but many studies have examined DNA methylation and corresponding gene transcription. In Arabidopsis it was found that 91 % of transposons, 58 % of pseudogenes and 20 % of expressed, non-overlapping genes free of known TEs were methylated (Zilberman et al., 2007). Similar numbers were acquired in an earlier study (Zhang et al., 2006). Methylated transposons are silenced while the coding sequence of genes across many tissues seems to still be moderately expressed despite CG methylation (Zhang et al., 2006, Zilberman et al., 2007). The relationship between DNA methylation and gene expression is different with regard to gene body methylation and promoter methylation. While promoter methylation is known to repress gene expression, the role of gene body methylation remains unclear.

1.5.1 Promoter Methylation and Gene Expression

Transcriptional repression is widely recognized to be associated with strong DNA methylation within promoters however the quantitative relationship between DNA methylation and gene expression levels, and the role of promoter methylation is not fully understood. When comparing promoter and gene body methylation regardless of gene expression, promoter methylation is more frequent than gene body methylation (Vining et al., 2012). Only approximately 5 % of expressed genes have DNA methylation within the promoter (Zhang et al., 2006) but the percentage of DNA methylation is much larger for genes with low or no expression. The importance of epigenetic marks on transcription upstream of the gene is quite clear. Promoter-methylated genes may be expressed in specific tissues while body methylated genes are constitutively expressed at a greater level (Zhang et al., 2006).

More refined mapping demonstrated that methylation at the 5' as well as 3' gene ends may be evolutionarily selected against to allow for transcription initiation and termination. An inverse correlation of DNA methylation with gene expression was seen where the CG and CHG methylation decreased at 5' and 3' gene ends (Gent et al., 2013, Zilberman et al., 2007). CHH methylation was generally low in abundance in the 1 kb upstream or downstream flanking regions of the gene (Gent et al., 2013). Unlike CG and CHG methylation, the level of CHH

methylation near the 5' end of genes does not appear to show an inverse correlation with gene expression. Rather, it appears to show a positive correlation with gene expression (Gent et al., 2013) suggesting a distinct function of CHH methylation.

1.5.2 Gene Body Methylation and Gene Expression

DNA methylation within transcribed regions or gene bodies is found in over one-third of expressed genes in *Arabidopsis* (Zhang et al., 2006), almost exclusively in the CG context (Ahmed et al., 2011). Genic CG methylation, but not CHG or CHH methylation, is similar across species as distantly related as angiosperms and ferns (Takuno et al., 2016). Methylated gene body corresponds to less siRNAs than methylated flanking regions (Cokus et al., 2008). An interplay between transcription and gene body methylation was suggested as highly expressed genes were generally unmethylated (Zilberman et al., 2007). In addition, short genes with DNA methylation were poorly expressed and gene bodies with low DNA methylation were often transcribed linking transcriptional changes to the influence of DNA methylation. In the DNA methyltransferase *met1* mutant some body-methylated genes were found transcriptionally up regulated (Zhang et al., 2006). Similarly, in the model tree species, *Populus trichocarpa*, unmethylated genes generally had higher expression than methylated genes (Vining et al., 2012). Compared to *Arabidopsis*, *Populus trichocarpa* had transcription repressed at a greater level with gene body methylation than promoter methylation (Vining et al., 2012). A modeling study on humans found that, like in *P. trichocarpa* gene body methylation is more indicative of gene expression level than promoter methylation (Lou et al., 2014). Furthermore, it was found that either promoter or gene body methylation is sufficient to signify low gene expression.

However, a recent study in rice showed that while heavy gene body methylation is associated with lower levels of gene expression, intermediate body methylation tends to be associated with high levels of gene expression (Wang et al., 2013). In addition, RdDM induced against an intron sequence can enhance gene expression (Deng and Chua, 2015). Thus, gene body methylation could also play a positive role in gene expression. It has been suggested in animal studies that gene body methylation may serve to suppress cryptic promoter of antisense transcript (Druker et al., 2004). However, in *met1* mutants antisense transcripts were rarely reported and did not correlate with gene body methylation (Zhang et al., 2006), which argues against this function. However, this result does not rule out CHG or CHH methylation functioning to inhibit cryptic transcription. RdDM-directed methylation, could occur to some sequence elements such as TEs in the gene body, and may function in this manner. Recent studies showed that a high incidence of TE sequences may overlay with exons and that 53 % of exonic TE sequences were unmethylated and 24 % were highly methylated (Ahmed et al., 2011).

It remains to be understood if DNA methylation of these TE sequences in the gene body has a function in gene expression.

1.5.3 Spread of DNA Methylation

DNA methylation “spreading” is an interesting phenomenon which might play a role in transcriptional gene regulation. It was shown that tandem repeats had a sharp drop in DNA methylation and siRNA abundance at flanking DNA (Cokus et al., 2008). A more gradual reduction in DNA methylation was seen moving from an inverted repeat into flanking DNA but siRNA abundance still sharply decreased at the flanks of the inverted repeat. siRNA therefore may correlate with repeat initiation and termination but DNA methylation abundance flanking repeats may be determined by the type of repetitive sequence.

TEs are repetitive sequences within the genome and are a known target of RdDM with 74% of annotated TEs methylated across the three cytosine contexts in Arabidopsis (Ahmed et al., 2011). A significant fraction of methylated TEs have no or few siRNA at the TE site and exhibit higher DNA methylation and siRNA levels at the TE extremities. Evidence showed DNA methylation may spread over short distances (~500 bp) from adjacent siRNA-targeted regions to target nearby TE sequences. CHG methylation was shown to spread the shortest distance. In maize, transposons with 24 nt siRNA within 1kb of the start or end of a gene were strongly associated with de novo RdDM (Gent et al., 2013). DNA methylation gradients are abolished in DNA methyltransferase mutants *met1* or become less steep in *drm1/2/cmt3 (ddc)* while *rdd* displays similar DNA methylation gradients as wild type (Ahmed et al., 2011). It was therefore proposed that DNA methyltransferases act in promoting or limiting spread of DNA methylation while DNA demethylases do not significantly contribute. This is opposite to the suggestion that active DNA demethylation contributes to the majority of DNA methylation maintenance (Zhu, 2009). Importantly, a main function of DNA demethylation is thought to prevent methylation spreading from repetitive sequences to neighbouring genes (Eichten et al., 2012). Through DNA methylation spreading in promoter regions they propose that neighbouring genes can be repressed by siRNA-targeted TE sequences (Zhu, 2009).

1.6 The Relationship between Changes in 24 nt siRNA, DNA Methylation and Gene Transcription

As mentioned earlier, the relationship between 24 nt small RNA accumulation and gene expression has not been extensively investigated. However, it can be anticipated that as the inducer of RdDM, a change in 24-nt siRNAs can alter gene expression levels. Indeed this was shown to be the case with several genes that had altered 24 nt siRNA and DNA methylation

levels in their flanking regions, and showed changes in expression in Arabidopsis hybrids compared to the parents (Groszmann et al., 2011). This was not an exhaustive search but demonstrated that altered hybrid sRNA levels can be associated with gene activity changes. An earlier study found in the gene bodies, 9% of DNA methylation was associated with siRNAs compared to the genome average of 37% (Zhang et al., 2006) suggesting that genic methylation is independent of siRNA targeting will be examined in Chapter 6.

Lister and colleagues (2008) performed genome-wide analysis of small RNA, methylation and transcriptome in the floral buds of various Arabidopsis mutants. Of particular interest were findings from the *rdd* triple demethylase mutant. An increase in sRNA levels was highly associated with an increase in DNA methylation density compared to wild-type Col-0, illustrating a dual directional relationship between siRNA and DNA methylation. This relationship was also seen in the DNA methyltransferase mutants *met1* and *ddc* where both DNA methylation and sRNA decreased. Small RNA that map to multiple locations do not act at every locus as determined by 85.4 % of unique sRNA-mapping loci containing methylcytosines compared to 47.6 % of multiple sRNA mapping loci.

1.7 Influence of RdDM and DNA Demethylation on TE Expression

Transposable elements (TEs) are mobile genetic elements that drive genome evolution through genomic DNA rearrangements and mutations. They fall under two major classes; 1) retrotransposons, that move via a copy and paste mechanism and 2) DNA transposons, that move via a cut and paste mechanism (Wheeler, 2013, Roy et al., 2015). This thesis refers to TEs as all sequences with homology to defined TEs, including the full length transposon genes that can encode proteins. A transposon gene refers to only the later full length, protein encoding TEs. Retrotransposons (consisting of Copia, Gypsy, LINE and SINE) represent the largest fraction of TE sequences in both plants and animal genomes, which makes up a large proportion of the genome. For instance, in maize around 49-78% of the genome is made up of retrotransposons (Sanmiguel and Bennetzen, 1998). DNA transposons consist of helitron, Tc1/mariner, en-Spm, hAT, harbinger, MuDR and pogo superfamilies (Ahmed et al., 2011). The highest proportion of methylated TE sequences (~90%) is from the Gypsy and En-Spm superfamilies while the lowest fraction (~40-50%) is the RC/Helitrons and Tc1/mariner superfamilies.

Transposons and other repetitive DNA sequences are generally silenced by heavy DNA methylation, particularly at CG sites, to maintain genome integrity. Hypermethylation of transposons has been shown in the *rdd* and *ros1* mutants leading to more significant reduction

in transposons expression (Zhu, 2009, Lister et al., 2008). These results suggest a basal level of transposon expression is maintained by DNA demethylation. Another study focused on 348 genes differentially expressed in *rdd* compared to wild type and found that a significant proportion of these genes were TE genes, or stress-responsive genes containing TE sequences in their promoters (Le et al., 2014). The role of TEs and DNA demethylation may be to maintain genome plasticity in genome structure and genetic variability for efficient responses following environmental challenges.

Recent studies in both *Arabidopsis* and maize showed that a strong correlation exists between stress-inducible gene expression and presence of TE insertions in these genes (Makarevitch et al., 2015). Non-stressed plants typically have low defence gene expression while stressed plants express their defence genes to enable pathogen associated molecular patterns (PAMP)-mediated resistance. Multiple up-regulated TE families (between four and nine) are associated upstream of stress-inducible abiotic defence related genes in maize (Makarevitch et al., 2015).

TEs seem to interact with siRNAs to influence gene expression. In fact it has been shown that the loss of 24 nt TE-associated siRNA in maize is strongest on TEs located near genes (Gent et al., 2013). In *Arabidopsis* 95 % of unmethylated TE sequences and 48-58 % of poorly methylated TEs have no 24 nt siRNA loci targeting the TE region (Ahmed et al., 2011). In contrast, 82 % (in repetitive rich heterochromatin) to 89 % (in gene rich euchromatin) of densely methylated TEs have matching 24 nt siRNA, indicating that densely methylated TEs are associated with siRNA in heterochromatin. *Arabidopsis thaliana* has a comparatively low TE copy number (~17%) compared to crops such as maize (85%) (Bucher et al., 2012). To examine the genome-wide epigenetic TE silencing patterns in other plant species *Arabidopsis thaliana* was compared to the closely related *Arabidopsis lyrata* which has a two to three fold higher TE copy number (Hollister et al., 2011). A greater fraction of siRNAs (86 %) targeted *A. lyrata* TEs compared to 68 % in *A. thaliana* and was associated with more effective TE silencing. siRNA density therefore closely mirrors TE distribution. A study in wheat grain found that TEs were in general similarly targeted by siRNAs regardless of tissue or TE family (Sun et al., 2013). The study in *Arabidopsis* showed that siRNA-targeting TEs are more likely to be methylated and have an effect on nearby gene expression (Hollister et al., 2011). A rise in gene expression was seen as the distance from the nearest TE increased in both *Arabidopsis* species but the two species differed in rate of increase. For example, maximal gene expression levels were reached when the nearest TE was greater than 2.5 kilo-base (kb) away in *A. thaliana* but only approximately 1 kb away in *A. lyrata*. Therefore, it appears that in a species with high TE copy number (like *A. lyrata*) siRNA levels are

also high and gene expression level is more tightly regulated by TEs than plant species (like *A. thaliana*) which has a relatively low TE copy number.

Analysis of differentially methylated regions (DMRs) in maize found that these regions were also located near TEs (Eichten et al., 2013a). Over 300 genes were found with differential expression patterns associated with DNA methylation variation. Cumulatively these results indicate changes in siRNA may target RdDM to promote DNA methylation variation at or near TE sites which leads to dynamic gene expression changes.

The mechanism by which TEs mediate gene expression remains to be understood. It is clear however that epigenetic misregulation often associates with endogenous TEs in plants (Bucher et al., 2012). Recent studies have focused on TEs and epigenetic regulators to address the question of how nearby gene expression is altered by TEs in plants. The current paradigm suggests that the chromatin state of genes neighbouring TEs can be influenced by heterochromatic silencing that spreads past the borders of TEs. Indeed this was found to be true in maize for lowly expressed genes 800 – 1,200 bp from a subset of retrotransposon TEs present in moderate to high copy numbers (Eichten et al., 2012). Not all retrotransposon TE families studied exhibited heterochromatic spreading which lead to the question: does the TE family determine if spreading occurs or does the insertion-site preference determine the chromatin state of the gene. While Eichten and colleagues (2012) concluded that the retrotransposon family determines spreading, other studies hint at a more complex regulation between TEs and gene expression. For instance in Arabidopsis, TEs near stress-responsive genes did not show a TE family preference (Le et al., 2014). Furthermore, tissue-specific transcriptional reactivation of repetitive DNA was found to be influenced by DNA methylation rather than TE or DNA transposon type (Baubec et al., 2014). A maize study however found that inbred lines had more differentially methylated regions than different tissues within the same plant line, but the authors did not link their findings to repetitive sequences (Eichten et al., 2013b). The evidence in Arabidopsis does suggest that TEs of mostly less than 1 kb in size are capable of spreading their DNA methylation marks into nearby promoters or transcriptional start sites of neighbouring genes to influence their expression (Le et al., 2014). No single factor (TE family, TE size, tissue type, or other factors) has been attributed to gene expression changes in this study. It is likely that multiple factors are involved in altering DNA methylation of TEs to influence gene expression.

It has been suggested sRNA effectors evolved from TEs and that TEs have shaped plant resistance gene evolution (Weiberg et al., 2014). It is proposed that by being present nearby,

TEs carry epigenetic information for resistance gene clusters which in turn are DNA methylated via sRNA generated RdDM from TEs (Wheeler, 2013). TEs and sRNA are therefore regulators in the plant-pathogen evolutionary arms race.

1.8 The Role of RdDM in the Stress Response of Plants

The growth and development of plants can be affected by their exposure to abiotic stresses such as heat, drought, salinity and nutrient deprivation or by biotic stress such as viral, nematode or pathogen attack. These primary stresses can lead to oxidative and thermal secondary stresses which cause genomic changes that disturb normal growth and may lead to death. Two broadly categorized responses occur in plants to general stress (Guleria et al., 2011). Firstly, a physiological response may occur which involve stress-associated proteins and metabolites. Secondly, a genomic response may be initiated through epigenetic regulations. Epigenetic responses may include histone and DNA modifications mediated by mechanisms such as RdDM. Several excellent reviews have recently covered RdDM and defence response, an area of high research interest due to the major economic effects that abiotic and biotic stress has on crop production (Gutzat and Scheid, 2012, Khraiweh et al., 2012, Ruiz-Ferrer and Voinnet, 2009, Seo et al., 2013, Wheeler, 2013).

1.8.1 RdDM in Abiotic Stress Response

A role for miRNA in a plants response to various abiotic stresses such as cold, hypoxia, mechanical and nutrient stress has been established (Khraiweh et al., 2012, Jeong and Green, 2013). Only recently has a role for siRNA-directed RdDM been found against abiotic stress. The effect of cold in maize was studied due to sensitivity of maize to frost. DNA methylation was found to not directly associate with cold acclimation however a specific and rapid demethylation of cold response genes in seedlings and roots was reported suggesting that epigenetics might contribute to maize adaptation to cold stress (Shan et al., 2013, Steward et al., 2002). It is hoped that artificially altering DNA methylation state may induce heritable changes in a plant phenotype as a way to develop stress resistance. In another example for the involvement of RdDM in abiotic stress response, Pol IV and Pol V, components of the RdDM pathway, are shown to be required for the transcriptional response to heat stress involved in conferring basal heat tolerance in *Arabidopsis* (Popova et al., 2013). Beside RdDM factors, *DDM1*, involved in maintenance of DNA methylation, and *ROS1* involved in DNA demethylation, have both been implicated in direct, or indirect, repair of oxidative DNA damage (Questa et al., 2013).

1.8.2 RdDM and DNA demethylation in Biotic Stress Response

The role of RdDM and DNA demethylation in biotic stress response is of major importance to this study. Many studies have examined biotic stress due to the economic benefit of developing disease free crops. Changes in DNA methylation in response to infection with the bacterial pathogen *Pseudomonas syringae* has been examined in many studies (Pavet et al., 2006, Agorio and Vera, 2007, Downen et al., 2012). Global DNA methylation is substantially reduced at a 180 bp centromeric repeat and other loci following bacterial attack (Pavet et al., 2006). This DNA hypomethylation occurs in the absence of DNA replication, indicating that it is due to active demethylation which may play a role in bacterial resistance. Consistently, bacterial growth was enhanced in infected *ros1* compared to wild-type Arabidopsis, although not in *dml2* or *dml3* plants, suggesting a ROS1-dependant role in *Pseudomonas* resistance (Yu et al., 2013).

Widespread dynamic DNA methylation changes have been shown to associate with differentially expressed genes in response to biotic stress. As previously shown (Pavet et al., 2006), hypomethylation occurred in Arabidopsis in response to infection with *P. syringae*, which suggests that DNA demethylation may play a positive role in bacterial resistance. This was also suggested by the enhanced *Pseudomonas* resistance of DNA methylation and RdDM-deficient mutants *met1*, *ddc*, *dcl2/3/4*, *rdr2*, and *drd1* (Downen et al., 2012). The study by Downen and colleagues further investigated the transcriptional consequence of the widespread hypomethylation in the DNA methylation mutants, and found that disruption of DNA methylation induces genes with functions in defense response. In another study, AGO4, Pol IV, Pol V, IDN2 and DRD1, components of the RdDM pathway, are shown to be transcriptionally down-regulated in Arabidopsis during the early stages of *Pseudomonas* infection (Yu et al., 2013). This result suggests that *Pseudomonas* infection may repress RdDM activity causing hypomethylation and inhibiting transcriptional gene silencing. By examining the responses to a virulent and an avirulent *P. syringae* strain and salicylic acid hormone, it was shown that differentially methylated cytosine patterns are coupled to differential gene expression (Downen et al., 2012). Some of the immune response genes were shown to contain repeats and TEs in their promoter regions which were targeted for DNA hypomethylation in response to bacterial infection (Yu et al., 2013). Therefore, DNA demethylation is involved in plant-induced immune responses, potentially activating defence genes through release of the TE and/or repeat silencing within their promoters to enable transcription. Promoters of *Fusarium oxysporum*-inducible stress response genes in Arabidopsis have also been suggested to be regulated by DNA demethylases targeting TEs in the promoters (Le et al., 2014).

The Vera lab (Lopez et al., 2011) found that Pol V, another key RdDM component, is required for plant immunity but only in response to specific pathogens. The *Pol V* mutant along with *ago4*, *drd1*, *rdr2* and *drm1drm2* all showed an increase in disease susceptibility to the necrotrophic fungal pathogens *Botrytis cinerea* and *Plectosphaerella cucumerina*. The *Pol V* mutant, but not the *Pol IV* mutant, showed enhanced resistance towards *P. syringae*. The genetic evidence clearly shows that the RdDM pathway is involved in plant immunity, but how the RdDM pathway regulates disease resistance gene expression remains unclear.

Another layer in plant defense mechanisms involves chromatin modification. Enhanced resistance to *Pseudomonas syringae* in a loss-of-function mutant of the histone deacetylase HDA19, was associated with increased salicylic acid-mediated defense responses (Choi et al., 2012). An earlier study showed that HDA19 over expression increased the ethylene and jasmonic acid pathways and enhanced resistance to the fungal pathogen *Alternaria* (Zhou et al., 2005). Taken together these results illustrate the importance of RdDM and histone modifications and their interaction with the hormonal pathways in plant defence against pathogen infections.

1.8.3 miRNA and phasiRNA in Stress Response

Other types of small RNA, miRNA and phasiRNA, have also been shown to play a role in abiotic and biotic stress response. miRNA402 over-expression in *Arabidopsis* promoted seed germination and seedling growth under salt stress conditions (Kim et al., 2010b). Seedling growth, but not seed germination, was also induced by miRNA402 over-expression during dehydration and cold stress conditions. RT-PCR analysis showed that miR402 down regulated the DNA demethylase gene *DML3*. These results highlight the role that miRNA402 plays in DNA demethylation and as a positive regulator of seed germination and seedling growth under abiotic stress conditions. Another positive miRNA regulator of plant stress response is miRNA160a, which targets pathogen associated molecular pattern (PAMP) induced callose deposition genes (Li et al., 2010). Some miRNAs appear to be negative regulators of plant stress response. PAMP induced callose deposition and disease resistance to bacteria are negatively regulated by miRNA398b and miRNA773 (Li et al., 2010). miRNA472 has been shown to post transcriptionally control a CC-NB-LRR gene, such as the RPS5 gene, to increase susceptibility to *Pseudomonas syringae* (Boccardo et al., 2014).

A role of miRNA-mediated production of phasiRNAs in defence against biotic stress has been suggested in various independent studies (Xiao et al., 2014, Wei et al., 2014, Wong et al., 2014, Zhai et al., 2011). In *Nicotiana benthamiana* miRNAs with differential expression profiles were

found upon infection with Tomato yellow leaf curl China virus (Xiao et al., 2014). Some of these differentially expressed miRNAs induced phasiRNA production, suggesting a function of the phasiRNAs in the interaction between the virus and *N. benthamiana*. The I2 resistance gene in tomato (*Solanum lycopersicum*) encodes resistance against *Fusarium oxysporum* and the Tomato yellow leaf curl China virus (Wei et al., 2014). miRNA6024 was shown to trigger phasiRNAs from I2 homologues which may positively influence tomato resistance during pathogen attack. Legume studies found phasiRNA generally target defense related nucleotide-binding site leucine-rich repeat (NB-LRR) proteins but phasiRNA could also be produced from NB-LRRs (Zhai et al., 2011, Wong et al., 2014)

1.9 The *Fusarium oxysporum* – *Arabidopsis thaliana* Phyto-Pathogen system

Generally, the epigenetics-related studies to date have focused on leaf-infecting pathogens. In fact, most of our knowledge regarding the role of RdDM came from the *Arabidopsis*-*Pseudomonas syringae* system. Mechanisms of plant resistance to root-infecting fungus are poorly understood which poses a significant barrier for the development of resistant crops. *Fusarium oxysporum* is a soil-borne, widespread hemibiotrophic fungal pathogen that is the causal agent of root rot or wilt diseases across many commercially important crops such as sugarcane, tomato and cotton (Gordon and Martyn, 1997). *Fusarium* has an extensive list of non-pathogenic and pathogenic isolates which are grouped into more than 120 different forma specialis (f. sp.) based on host specificity (Gordon and Martyn, 1997, Michielse and Rep, 2009).

Extensive study of gene transfer, pathogenicity, signal transduction and infection process has occurred in this model fungal organism (Czymmek et al., 2007, Chen et al., 2014, Peschen et al., 2004, Di Pietro et al., 2003). The infection process of the hemibiotrophic pathogen is best examined in tomato (Di Pietro et al., 2003). In the biotrophic phase spores germinate in the soil to allow the fungal hyphae to penetrate the root tips and intercellularly entering the root cortex and xylem vessels. Initially yellowing and wilting of the leaves occurs as a result of the vascular tissue becoming clogged with fungal spores. After the establishment of infection, the fungus turns necrotrophic and produces toxins or effectors that enter the vascular tissue and contribute to the vein obstruction that eventually can lead to the death of the plant. Vascular tissue and roots may turn brown or rot depending on the forma specialis that infects the plant.

The model plant *Arabidopsis thaliana* is an ideal species to use for investigating the role of RdDM in plant-fungal interaction because of the availability of genetic mutants, its well

characterized genome, and the susceptibility to infection by *Fusarium oxysporum* (Edgar et al., 2006). The *Arabidopsis thaliana*-*Fusarium oxysporum* system was therefore chosen for this study.

1.20 Thesis Aims

An early study in our laboratory using the *Arabidopsis*-*Fusarium* system has implicated RdDM and DNA demethylases in *Fusarium* resistance and interaction between the two mechanisms. These studies have shown that:

- 1) the RdDM mutants *Pol V* and *AGO4* and the DNA demethylation mutant *rdd* were more susceptible to *Fusarium* infection (Le et al., 2014);
- 2) many stress-response genes were downregulated in *rdd* and these genes contain promoter TEs that showed increased DNA methylation in *rdd* in regions corresponding to published RdDM associated siRNAs (Le et al., 2014).

However, while this study suggested that RdDM and DNA demethylases interact to coordinately regulate *Fusarium* resistance in *Arabidopsis*, how these two pathways interact remained unclear. Also, how these two pathways function in response to *Fusarium* infection remained unknown.

The aim of this thesis was therefore to examine the interaction of RdDM and DNA demethylases, and the role of this interaction in plant disease response using the *Arabidopsis*-*Fusarium oxysporum* system, with a focus on 24-nt RdDM associated siRNAs. In particular, using whole genome sRNA, mRNA and bisulfite sequencing, coupled with bioinformatics and wet lab gene expression analyses, I aimed to address the following questions

1. Is 24-nt siRNA accumulation responsive to *Fusarium* infection and do DNA demethylases participate in this response?
2. Do *Fusarium*-induced or *rdd*-affected changes in 24-nt siRNAs correspond to changes in gene expression, particularly stress response genes?
3. Are siRNA-related gene expression changes associated with changes in DNA methylation?
4. Which DNA methylase genes participate in these plant responses to *Fusarium*?

Chapter 2: Materials and Methods

2.0 Plant Materials, Virus and Fungal Strains

In all experiments performed in this study, *Arabidopsis thaliana* ecotype Columbia-0 (Col-0), was used as the wild type plant. Single demethylation mutants used included *ros1-3* and *dml2-1* both of WS background introgressed six times into the Col-0 background, and *dml3-1* of Col-0 background (described in Penterman *et al.*, 2007 and provided by Dr Robert Fischer) and the triple mutant *rdd* of Col-0 background (*dml2*: Salk_056440, *dml3*: Salk_131712, *ros1*: Salk_045303 T-DNA insertion lines) (described in Le *et al.*, 2014 and provided by Dr Jian-Kang Zhu).

Virus strains used were the Fny strain of Cucumber mosaic virus (CMV), and Turnip mosaic virus (TuMV) (available in CSIRO, Australia). The fungal strain used was *Fusarium oxysporum* f.sp. *conglutinans* (strain 5176) (obtained from Dr Roger Shivas, Queensland Department of Primary Industries and Fisheries, Australia).

2.01 Plant Growth Conditions and Fungal and Virus Infections

All seeds were sterilised for three hours in a desiccator using chlorine gas generated by mixing 100 mL of household bleach (White King) with 3 mL of concentrated hydrochloric acid (Clough and Bent, 1998). Sterilized seed were sown on Murashige and Skoog (MS) agar (see Supplementary Table 2.1) with 3 % sucrose and incubated for two-three days at 4 °C for stratification before being placed in a 22 °C growth room with a light regime of 16-hours light and 8-hours dark.

Fusarium oxysporum from a glycerol stock was inoculated into half strength (12 g per litre) potato dextrose broth (Sigma Aldrich), and grown at 28 °C for three-four days with constant shaking. Spores were harvested by filtering through Miracloth and centrifugation and resuspended in water to a concentration of 1×10^6 spores/mL. Two and half week-old *Arabidopsis* seedlings were infected via root dip or spray inoculation as specified. Dipping roots in, or spraying plants with sterile water, was used for mock infection. Inoculated plants were placed on sucrose-free MS medium (Supplementary Table 2.2) and incubated at 22°C in the 16-hour light/8-hour dark growth room. For RNA analysis, 10 or more whole plants were collected at 1, 3 and 6 days post infection and frozen in liquid nitrogen. Four biological replicates were harvested for each time point.

For soil infection, plants were grown on MS media for two weeks, root-inoculated with the 1×10^6 spores/mL *Fusarium* culture, and allowed to grow on sucrose-free MS for two-three days,

before being planted in 10 cm pots with Debco soil. The plants were grown in a growth cabinet under 16-hour light at 25 °C simulating day conditions and 8-hour dark, 18 °C night conditions. A humid environment was created by covering pots with glad wrap. The glad wrap was removed after three days, and plants were watered every second day.

For virus inoculation, TuMV-infected *Nicotiana benthamiana* leaves, and CMV-infected *N. benthamiana* leaves were ground in a mortar and pestle in the presence of 0.1 M phosphate buffer at pH 7.0. Approximately 3 week-old plants grown at 25°C in the soil were used for virus inoculation. Two or three leaves were dusted with the abrasive Carborundum, and then hand-rubbed with the extract of TuMV or CMV-infected *Nicotiana* leaves.

2.02 Phenotypic Assays

For measuring disease phenotype after *Fusarium* infection, plants were photographed and disease severity was measured by visually assessing symptom development on the leaves using a scale of 0 – 4. Similar to previous scales used to measure disease symptoms (Le et al., 2014, Lyons et al., 2015), 0 = no chlorotic leaves, 1 = 1-3 leaves showing chlorosis, 2 = 4-6 leaves showing chlorosis, 3 = all leaves showing chlorosis, 4 = dead plant. At least 30 plants were assessed per line.

Rosette diameter and root length were measured at 14 DPI with a ruler. Roots were then removed from the plant with scissors and the remainder of the plants was weighed on a measuring balance (Sartorius). To determine germination rate, sterilized and stratified seed on MS agar was placed in a 22°C growth room, monitored daily and photographed upon seedling emergence.

For measuring flowering time plants were grown in soil at 25 °C under a 16 hour light period watering every second day. Flowering time was measured for the 3 single demethylase mutants, the triple *rdd* demethylase mutant and Col-0.

2.1 Fungal Biomass Measurement

A chitin fluorescence assay developed by Ayliffe *et al.*, (2013) was used to measure fungal biomass in *Fusarium*-infected *Arabidopsis* plants. In brief, whole plant, root, and aerial tissues were harvested from *Fusarium*-infected plants at 1, 3 and 6 days post infection, weighed and covered with 1 M KOH. Tissue was autoclaved and then washed with 50 mL of 50 mM Tris, pH 7.0 to neutralize. After neutralization, the tissue was sonicated at 30-40 amplitudes using a UP 4000S ultraschall-prozessor (a brand by Dr. Hielscher) to generate a uniform, macerated

suspension. Approximately 100 μL of tissue suspension was added to 10 μL of 1 mg/mL solution of WGA-FITC (Sigma Aldrich) dissolved in 200 μL H_2O and left to stand for at least 1 hour. Following staining, samples were centrifuged at 600 $\times g$ for 3 minutes. Unbound stain was removed from the supernatant by pipetting and the pellet resuspended in 200 μL of 50 mM Tris, pH 7.0. Washing was done three times and resuspended in a final 100 μL of 50 mM Tris, pH 7.0. Fluorometric measurements were made with 485-nm adsorption, 535-nm emission wavelengths and a 1.0-s measurement time on a Wallac Victor 1420 multilabel counter (Perkin-Elmer Life Science, Waltham, MA, U.S.A.).

2.2 General Molecular Methods

2.21 Rapid CTAB DNA Extraction

For rapid DNA extraction a metal ball bearing and 75 μL of 0.5x cetyltrimethylammonium bromide (CTAB, pH 8.0) (Table 2.1) was added to a small amount of tissue (e.g. a cotyledon) in a 1.5 ml Eppendorf tube before shaking the tube in the Tissue Lyser (Qiagen) for 2 minutes at a frequency of 30. Chloroform (50 μL) was added before the tube was shaken vigorously by hand to dislodge the ball bearing from the base of the tube. Sample was centrifuged at 3,800 rpm for 5 minutes and the supernatant was used for PCR reactions.

2.22 Total RNA Extraction using Trizol

Approximately 50 – 100 mg of tissue was collected for RNA extraction. The tissue was ground with a mortar and pestle in liquid nitrogen before homogenization with 2 mL Trizol reagent (Ambion, Life Technologies). Samples were incubated for 5 minutes at room temperature, 300 μL chloroform was added, and the samples were shaken well for 15 seconds before incubating again for 5 minutes at room temperature. Following centrifugation at 13,200 rpm for 15 minutes at 4 °C the upper aqueous phase was transferred to a fresh 1.5 mL tube. To the supernatant 600 μL of isopropanol was added and the mixture was incubated overnight at -20 °C. RNA was then precipitated by centrifugation at 13,200 rpm for 10 minutes at 4 °C. The RNA pellet was washed with 1 mL of 75 % ethanol in DEPC-treated H_2O . Short vortexing dislodged the RNA pellet from the side of the tube before centrifugation at 9,500 rpm for 5 minutes at 4 °C. Supernatant was discarded and the pellet was air dried before being resuspended in 20 μL DEPC-treated H_2O .

Table 2.1: Buffers

Buffer	Final Reagent Concentration	How to Prepare
0.5X CTAB, pH 8.0	140 mM Sorbitol	Add Together
	220 mM Tris-HCL, pH 8.0	
	22 mM EDTA	
	800 mM NaCl	
	1% Sarkosyl	
	1% Cetyltrimethyl Ammonium Bromide (CTAB)	
	50% Distilled water	
17% Acrylamide Gel – 25 mL sized gel	10.6 mL 40% Acrylamide/bis-acrylamide (19:1)	Room Temperature then 50°C until dissolved
	2.5 mL 10 X Tris/Borate/EDTA (TBE)	
	4.3 mL Distilled water	
	10.5 g Urea	
	150 µL 10% Ammonium Persulfate (APS)	Add on ice
	12.5 µL TEMED	Add on ice last and pour gel immediately
Small RNA Northern Blot Pre- hybridisation Buffer	50% Formamide	Add together
	1% sodium dodecyl sulfate (SDS)	
	5x SSPE (pH 7.4) – see buffer description below	
	5x Denhardt's solution – see buffer description below	
	Distilled water to final volume	
100 x Denhardt's solution	2% (w/v) Ficoll 400	Add together
	2% (w/v) polyvinylpyrrolidone	
	2% (w/v) bovine serum albumen	
	Distilled water to final volume	
20x SSPE solution	3.0 M Sodium chloride (NaCl)	Mix together and adjust to pH 7.4 with 10M sodium hydroxide (NaOH)
	0.2M Monosodium phosphate (NaH ₂ PO ₄)	
	0.02 M Ethylenediaminetetraacetic acid (EDTA)	
Small RNA Northern Blot wash buffer	2x SSC	Add together
	0.2% SDS	
	Distilled water to final volume	

2.23 DNA and RNA Quantification and Visualisation

RNA and DNA concentration was determined using the NanoDrop® ND-1000 Spectrophotometer (Thermo Scientific, City CA, USA).

Quality of RNA or DNA was determined by running a 1 % agarose gel in 1x Tris/Borate/EDTA (TBE). The gel tank was soaked in concentrated sodium hydroxide (Sigma) prior to running RNA to prevent RNA degradation.

2.24 Phenol-Chloroform Extraction

Phenol-chloroform extraction was performed on DNA that needs to be purified such as following digestion reactions (see general molecular cloning experiments). To 50 µL of DNA, 50 µL H₂O was added to bring the total volume up to 100 µL. One hundred microlitres of Phenol-Chloroform-isoamyl alcohol (Ambion) was added then the samples were vortexed and spun at top speed for 5 minutes before the upper-phase was transferred to a new 1.5 mL microcentrifuge tube. The aqueous phase was extracted again, but using chloroform, to remove the residual amount of phenol. To the final aqueous phase 10 µL 3M sodium acetate and 250 µL 100 % ethanol was added and the tube vortexed before incubation on ice for 10-15 minutes. Samples were spun for 10 minutes at top speed, supernatant discarded and the pellet washed with 500 µL 70 % ethanol by spinning at top speed for 5 minutes. DNA pellet was dried and dissolved in H₂O.

2.25 Polymerase Chain Reaction (PCR)

A 20 µL PCR reaction generally consisted of 1 µL template DNA (~200 ng), 2 µL each of 10 µM primer pairs, 2 µL 10X reaction buffer (Fischer), 2 µL 25 mM MgCl₂ (Fischer), 0.8 µL 5 mM dNTP (Fischer), 0.2 µL Taq Polymerase (Fischer) and 10 µL dH₂O. PCR was run in a thermo cycler (My Cycler) with the following cycles:

Initial denaturing: 95 °C for 3 minutes for 1 cycle

Amplification: 95 °C for 45 seconds, 55 °C for 45 seconds, and 72 °C for 1 minute by 30 cycles

Extension: 72 °C for 5 minutes then hold at 10 °C until being collected.

2.3 Northern Blot Hybridization

Northern Blots were run using total RNA, method described above in section 2.2, to determine the size class of small RNA in 2.5 week old Arabidopsis either infected with Fusarium, CMV, TUMV and/or in the *rdd* mutant compared to Col-0.

2.31 Electrophoresis of Small/Large RNA using a Urea-Polyacrylamide (PAGE)/Formaldehyde-Agarose Gel

Formaldehyde agarose gel electrophoresis of RNA was performed for CMV and TUMV Northern Blot hybridisation. To prepare formaldehyde agarose gels (100 mL), 1.3 grams agarose was melted in 85 ml of water in a microwave, cooled to ~50 °C, and then 10 mL of 10x MOPS buffer and 5 mL of 40 % formaldehyde was added. The gel mix was poured into a gel rig and allowed to set before being used for electrophoresis. For sample loading, a 5 µL aliquot of each RNA sample was mixed with 1 µL of 10x MOPS buffer, 10 µL formamide, 3.5 µL formaldehyde and 0.1 µL RNA stain (10 mg/mL Ethidium bromide), and denatured at 65 °C for 10 – 15 minutes, then 15 µL of RNA loading dye (10 mL formamide + 10 mg bromophenol blue [BPB] + 10 mg xylene cyanol FF [XCFF]) was added to the sample immediately before loading the gel. The gel was run in 1x MOPS buffer at 50-80 V for approximately two hours or until the lower blue dye had migrated through three quarters of the gel. RNA gel image was captured using a UV transilluminator (Gel Doc TM XR+, Biorad). RNA was transferred to HyBond-N membrane (GE healthcare, Piscataway, NJ, USA) using capillary blotting overnight in 20x SSC. The RNA membrane was UV crosslinked using the UV stratalinker 1800 (Stratagene, Sydney, Australia).

For small RNA analysis, 20 µg total RNA was added to 15 µL of deionised formamide/BPB/XCFF loading dye, heated for 2 minutes at 95 °C and separated on a 17 % urea-PAGE gel in 1 x TBE running buffer. The 17 % urea-PAGE (Table 2.1) was poured between glass plates in an upright Bio-Rad mini Protean II rig, and allowed to set for approximately one hour before use. The gel was run at 250-300 volts until the lower bromophenol blue dye had migrated to the bottom of the gel.

Following electrophoresis the separated RNA was electroblotted onto Hybond-N⁺ membrane (Ge Healthcare, Piscataway, NJ, USA) in 0.5 x TBE at 30 volts for 1-1.5 hours. The RNA membrane was UV crosslinked using the UV stratalinker 1800 (Stratagene, Sydney, Australia). The membrane was left overnight to dry as evidence suggests this may improve sensitivity (Pall and Hamilton, 2008). Once dry, the membrane was cut in half at approximately 1 cm above the upper xylene cyanol FF dye marker. The upper part was used to probe for the U6 small nuclear RNA, an indicator of relative sample loading, and the lower part for small RNA hybridization.

2.32 Hybridisation with Radioactive Probes

The membrane was placed, RNA side up, in a hybridisation bottle and was pre-hybridised with 50 mL of 50 % formamide buffer (Table 2.1) in a rotator oven at 40 °C for 2-3 hours. The buffer

volume was reduced to 20 mL, the radiolabelled probe added and hybridisation allowed to continue at 40 °C overnight.

2.33 Membrane Wash, Exposure and Strip

Following hybridisation, the hybridization solution was poured off (stored for further use) and the membrane quickly rinsed with approximately 50 mL of small RNA Northern Blot wash buffer (Table 2.1) to remove the majority of the unbound radioactive probe. The membrane was washed twice with approximately 100 mL of wash buffer for 15 minutes in the rotator oven at 42 °C.

To visualise RNA abundance, the membrane was sealed in a plastic bag and exposed against a phosphor screen overnight. The membrane containing the U6 loading control was normally exposed for 1-2 hours. The phosphor screen was scanned in a phosphorimager (FLA-5000, Fujifilm) with 100 pixel resolution.

To strip the membranes for re-hybridization with other small RNA probes, a solution of 0.1% SDS was brought to the boil using a microwave and subsequently poured directly onto the membrane in a container, which was incubated on a shaker for 10 minutes. This was repeated again. Membranes were visualized by phosphorimaging, as previously described, to ensure complete removal of the radioactive label. Membranes were either re-probed directly or dried and stored for future use.

2.34 Preparation of Radioactive Probes

To detect siRNA abundance, radioactive RNA probe was prepared as follows: plasmid DNA containing the target sequence was linearized by restriction digestion, and RNA was transcribed using either SP6 or T7 RNA polymerase in a 20 µL reaction containing [α -³²P]-UTP (Perkin Elmer). The radioactive RNA transcript was purified by precipitation with 10 µL of 7.5M NH₄Ac and 75 µL of ethanol on ice for 10 minutes followed by centrifugation at 13000 rpm for 15 minutes. The pellet was resuspended in 100 µL TE, and precipitated again with 50 µL 7.5M NH₄Ac and 375 µL of ethanol on ice for another 10 minutes. The probe was pelleted as above, air-dried for 5 minutes and resuspended in 20 µL H₂O or TE.

For detecting miRNA and tasiRNA, oligonucleotides complimentary to the specific sRNA molecules were end labelled using ³²P-ATP and T4 DNA Kinase (Invitrogen, Life technologies). Briefly, 2 µL oligonucleotide (10 µM) were end labelled with 4 µL 5 x Forward Buffer, 10 µL sterilized water, 3 µL [γ -³²P]-ATP, 1 µL T4 Kinase and left to incubate at 37 °C for 1 hour. To remove unincorporated nucleotides, the probes were purified using microspin G-25 columns

(GE healthcare) according to the manufacturer's instructions. Briefly, the column was vortexed to resuspend the resin and spun at 3,000 rpm for 1 minute to remove the suspension buffer. 30 µl of sterilized water was added to the labelling reaction to give a final volume of 50 µl, before being loaded onto a spin column and spun at 3,000 rpm for 2 minutes. The flow-through liquid was collected and used as probe. A list of oligonucleotide probes used to detect small RNA is seen in Table 2.2.

2.4 General Molecular Cloning

Using the cloning procedures detailed below an RNAi hairpin construct was made to silence DME (AT5G04560), a DNA demethylase gene in Arabidopsis. In summary the DME fragment in pGEM T-easy was excised using the restriction enzymes XbaI and ClaI and ligated in an antisense orientation into XbaI and ClaI-digested pKannibal vector. The DME fragment in pGem T-easy was then excised using the restriction enzymes XhoI and KpnI and ligated in a sense orientation into XhoI and KpnI-digested pKannibal vector harbouring the antisense fragment (pKannibal: DME hp). The *Nicotiana tabacum* rubisco small subunit (SSU) promoter in pBC (stratagene) was excised using the restriction enzymes SacI and SalI and ligated into the pKannibal: DME hp vector restricted with SacI and XhoI (replacing the 35S promoter sequence), see Figure 2.1. The completed SSU promoter /DME hairpin construct was excised using NotI digestion and ligated into the NotI site of the binary vector pBART.

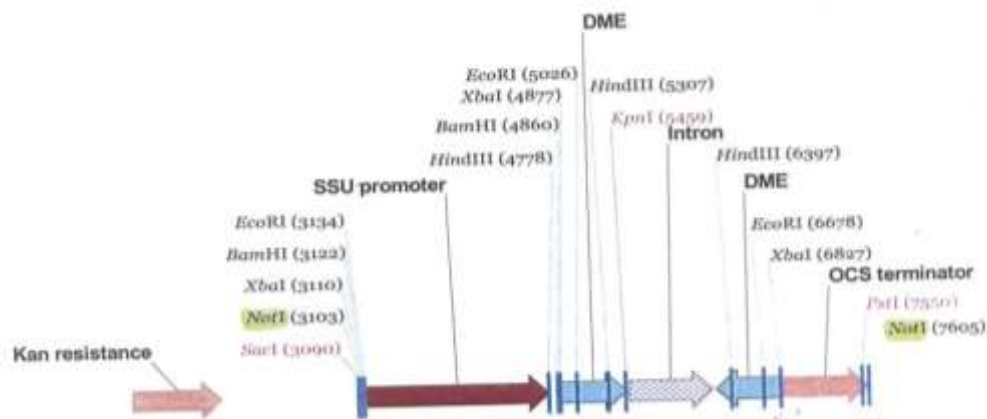
2.41 PCR purification and Ligation

A PCR fragment corresponding to 556 base pairs of the fourth DME exon was amplified using the forward primer ACTCGAGTCTAGACACATGTACTAGACATAGAGG, (containing 5' end XhoI and XbaI sites) and the reverse primer AGGTACCATCGATATCCCACTGGCTGTAGGC (containing 5' end KpnI and ClaI sites). PCR conditions were followed as outlined in section 2.2. The DME PCR product was purified using a Qiagen PCR purification column. Briefly, 5 x volume of PB buffer was added to the PCR reaction before being applied to a spin column. After centrifugation for 1 minute at 13,000 rpm, the column was washed with 750 µL of wash buffer PE, spun again for 1 minute, and the product eluted with 40-50 µL of elution buffer.

Table 2.2: small RNA Northern Blot Probes

Probe	5'-3'
AtSN1 siRNA probe	ACCAACGTGTTGTTGGCCAGTGGTAAATCTCTCAGATAGAGG
siRNA 1003 probe	ATGCCAAGTTTGGCCTCACGGTCT
siRNA 02 probe	GTTGACCAGTCCGCCAGCCGAT
Cluster 4 siRNA probe	AAGATCAAACATCAGCAGCGTCAGAGGCTT
Ta-siRNA 255 probe	TACGCTATGTGGACTTAGAA
microRNA 171 probe	GATATTGGCGCGGCTCAATCA
Locked nucleic acid siRNA 02 probe ("+" stands for locked nucleic acid residue)	G+TTG+ACC+AGT+CCG+CCA+GCC+GAT
LNA cluster 4 siRNA probe	AA+GATC+AAAC+ATCA+GCA+GCGTC+AG+AGG+CTT
AGO4 promoter siRNA probe	ATATATGATTTGTATGTAAAAGATGAACAACAACAAGCGCTAG AAAAGGATGAACGTTGTAGAGGTAATAATATGATAACATTAGA TCCAAATAGCAAAAGCTCATTAGAATAGTTATGTTACTTTATAA TTGGAGACAAAAATTTCACTTTGCTTTCAAAGCAGATTTAATAT ATAATTTTGTCTCATATTGAAGGTACCATTGCTTTATAGGAACA ATGTTGACGATTCCAGTTTCGGTTGCATCTGCTCAAAGAAGTTT TTCAAAGCTGAAGTTGATCAAATCGTATCTTCGGTCGACTATGT CACAAGAAAGATTGAGTGAGTTGGCTATTTTATCAATCGAAAG AGAATTGGTTAGAGACATTGATTTTGAAAGCTTGGTTAACGATT TAGTGGAGAGGAAGGGTCAACAACATTTAAGAACTAATATTT AGTTTGTTAATTTTATTTTCTTACATGACAATGATGTTTATTTT TGTAGACGATTTTTTTTTCGTAAGGATATTTATTTTATTCGCTT
miRNA 164 probe	TACGCGGAATTTCTGATATA
miRNA 172 probe	ATGCAGCATCATCAAGATTCT
miRNA 159 probe	TTTGGATTGAAGGGAGCACTT
LNA miRNA 165 probe	GGGTGGA+TGA+AGC+CTGTG

Figure 2.1: DME HP in pKannibal driven by the SSU promoter (7610 bp)



Ligation of the PCR products into the pGEM®-T Easy vector (Promega) was performed according to the manufacturer's instructions. Briefly, 5 µL of 2x ligation buffer, 1 µL of the pGEM®-T Easy vector, 3 µL of PCR product and 1 µL of T4 DNA ligase were mixed in an eppendorf tube and allowed to incubate for two hours or overnight. Similarly, to engineer the DME hairpin construct the ligation reactions involved mixing the restricted insert and vector fragments in a ratio of 10:1 in the presence of 10x ligation buffer (Promega), water to 20 µL, and 1 µL of T4 DNA ligase. The ligation was incubated at room temperature for at least two hours or overnight.

2.42 Transformation of DH5α *E. coli* cells

Ligated plasmids were transformed into *E. coli* by mixing 1 µL of ligation product with 50 µL of electro-competent *E. coli* DH5α cells via electroporation at 2.0 kV using an electroporator (Electro Cell Manipulator, ECM-395, BTX). Immediately after electroporation 500 µL of Luria Broth (LB, Thermo Fischer) was added, and then incubated at 37 °C for 40 minutes. The bacterial suspension was plated onto LB plates containing appropriate antibiotics to select for colonies containing the cloned plasmid vector. pGEM®-T Easy-based clones were also screened using blue/white selection with isopropyl β-D-a-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). The bacterial plates were incubated overnight at 37 °C.

2.43 Colony Screening and Sequencing

Putative clones were screened using PCR analysis. A PCR master mix was prepared with oligonucleotide primers designed to amplify clones containing the desired insert and orientation. Twenty microlitres of the PCR reaction mix was aliquotted into PCR tubes. Then, using a pipette tip, a sample of each colony was transferred to the reaction mix for PCR amplification. Positive colonies were grown overnight in LB liquid culture (4 mL LB with 100 mg/L ampicillin) at 37 °C with constant shaking at 200 rpm. The QIAprep® Spin Miniprep Kit (Qiagen) was used to purify the plasmid DNA.

To confirm the presence and orientation of plasmid in a colony, insert-specific restriction enzyme digestion was designed using restriction enzymes and buffers according to the manufacturer's instructions. Generally, a total digestion volume of 20 µL was used containing, 2 µL 10X Buffer, 2 µL plasmid DNA, 1 µL restriction enzyme and dH₂O. Digests were incubated overnight at 37 °C and then separated in a 1% agarose gel in 1X TBE buffer, and visualized with a UV transilluminator (Gel Doc™ XR+, Biorad).

For sequencing, between 100 – 500 ng of plasmid DNA (2 µL) was added to 1 µL Big Dye (Applied Biosystems, Life Technologies), 3.5 µL 5X BD Buffer, 1.5 µL of 5 µM Primer and 12 µL dH₂O. Following the sequencing reaction (96 °C for 1 minute, 30 cycles of 96 °C for 1 minute, 50 °C for

5 seconds, 60 °C for 4 minutes) the mix was transferred to a 1.5 mL microcentrifuge tube before 3 µL of 3 M Sodium acetate and 50 µL of 100% ethanol was added to precipitate the DNA. The tube was vortexed, incubated at room temperature for 10 minutes and spun at 13,000 rpm for 20 minutes at room temperature. The supernatant was removed, 500 µL of 75 % ethanol added to wash the DNA pellet before the tube was again spun at top speed for 20 minutes at room temperature and the supernatant discarded. The pellet was air dried and sequenced by the ANU Biomolecular Resource Facility (John Curtin School of Medical Research) using the Sanger sequencing platform.

2.5 Whole Genome Sequencing Library Construction and Sequencing

2.51 Small RNA and mRNA Library Construction and Sequencing

Whole plant samples of both mock and Fusarium-infected of Col-0 and *rrd* lines harvested at 1, 3 and 6 days post infection (DPI) were used for RNA extraction. Ten or more plants were pooled for each biological replicate. Of four biological replicates, two were used for small RNA sequencing library construction. The same two 3 DPI replicates, plus an additional 3 DPI replicate from an independent infection experiment, were chosen for mRNA-sequencing library construction.

Total RNA was isolated using Trizol reagent described previously. RNA quality was assessed using a RNA 6000 Bioanalyser (Agilent) to ensure samples recorded a high quality RNA integrity number (>7.9). If needed the samples were further purified using phenol-chloroform extraction and ethanol precipitation. The RNA pellet was covered with 100% ethanol and sent for sequencing.

Small RNA cDNA library construction and high-throughput sequencing were performed by Beijing Genomics Institute (BGI; <http://www.genomics.cn/>). Briefly total RNA was size-fractionated to 18-30 nucleotides using denaturing PAGE gel and 5' and 3' adapters were ligated to the RNA prior to cDNA synthesis and RT-PCR. Resulting cDNA libraries were then sequenced on an Illumina HiSeq 2000. Twelve samples were run per lane. Complete small mRNA sequencing data sets generated by BGI will be deposited in NCBI's Gene Expression Omnibus upon publication.

mRNA sequencing library construction and high-throughput sequencing was performed by the Australian Genome Research Facility Ltd (AGRF). Briefly, the first step was to enrich poly-A mRNA from 10 µg of total RNA using oligo (dT) magnetic beads. Next the mRNA was broken into

short fragments which were then used as templates for first strand cDNA synthesis using reverse transcriptase and random hexamer primers. RNase H, DNA polymerase I, buffer and dNTPs were added for second strand cDNA synthesis. These cDNA fragments were then end repaired and adenine nucleotides were added to create a Poly-A tail. Sequencing adaptors were ligated to the fragments which were then purified and enriched by PCR to produce the final cDNA library for Illumina HiSeq 2000 RNA sequencing.

2.52 Bisulfite Sequencing Library Construction and Sequencing

Bisulfite-sequencing libraries were constructed similar to that outlined in Lister *et al.*, 2008. To summarise, genomic DNA was extracted using a CTAB DNA extraction method (Small and Leaver, 1988) and 5 µg was fragmented by sonication to ~100-500 bp using Convaris 200 bp protocol (Convaris, USA). End repair and a 3' A-tail over-hang was added to gDNA followed by Agencourt AMPure XP purification (Beckman Coulter, USA) before ligation of NEXTflex methylated adapters provided by Bio Scientific (Bio Scientific, USA) as per manufacturers guidelines. gDNA was again purified using Agencourt AMPure XP purification (Beckman Coulter, USA). The concentration of the adapter-ligated gDNA was checked on a Qubit Fluorometer (Life Technologies, Australia) to ensure 450 ng or less was used for bisulphite conversion performed with the MethylCode Bisulfite Conversion kit (Life Technologies, Australia) as per manufactures' instructions. Ten–fifteen nmol of bisulphite-converted, adapter-ligated DNA molecules (~ 20 µL) were enriched by PCR using 4-6 cycles with the following: 25 µL of KAPA SYBR FAST qPCR Master Mix (2x) (Kapa, USA), which contains DNA polymerase, dNTPs, SYBR Green I dye and MgCl₂ at a final concentration of 2.5 mM, 2 µL NEXTflex Primer mix (Bio Scientific, USA) and 3 µL of dH₂O. The thermocycling of the PCR was as follows: 95 °C for 2 minutes, then 5 cycles of 98 °C for 20 seconds, 64 °C for 15 seconds, 72 °C for 1 minute, completed with one 72 °C, 10 minute step. The enriched library was purified with Agencourt AMPure XP purification (Beckman Coulter, USA) before quantity and quality was examined by a LabChip GXII bioanalyzer (PerkinElmer, USA) and a Qubit Fluorometer (Life Technologies, Australia). qPCR was performed according to the manufacturers' instructions (Kapa, USA) to determine the concentration of the libraries.

2.6 Bioinformatics for the Whole Genome Sequencing Data

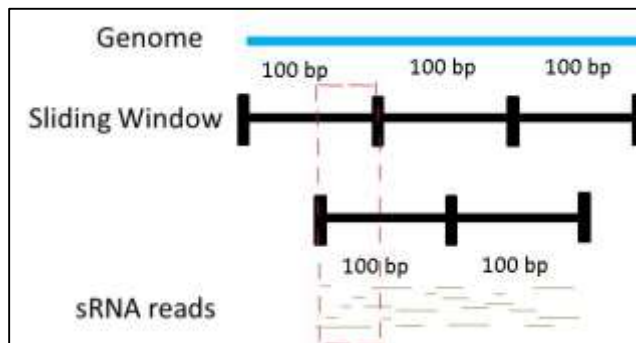
2.61 Bioinformatic Analysis of Small RNA Sequencing Data

The small RNA-seq reads were received with adaptors and low quality reads removed. I then removed reads less than 18 nt and greater than 30 nt using the CLC genomics workbench (version 4.9). Trimmed and filtered reads were mapped to the TAIR 10 *Arabidopsis thaliana* genome (from the CSIRO Plant Industry Bioinformatics Group Genome Browser) and *Fusarium*

oxysporum f.sp. *conglutinans* (strain 4287) contigs previously sequenced by 454 technology by Donald Gardiner (CSIRO, Agriculture Flagship, Brisbane). Reads which mapped to *A. thaliana* were retained and the reads which mapped to *F. oxysporum* were removed. The CLC genomics workbench RNA-seq analysis application was used to allow no mismatches. Mapping up to 10 or 30 locations was allowed to avoid removing reads that map to repeat sequences. Mapping up to 1000 locations with one mismatch per 100bp was also done. Using a similar approach to previous sRNA sequencing studies (Lister et al., 2008, Ha et al., 2009) a 100 bp sliding window with a 50 bp offset was generated against the trimmed and filtered *A. thaliana* reads which mapped up to 10, 30 or 1000 locations to ensure overlapping reads were not counted twice using Galaxy (Goecks J, 2010). Small RNA reads had to locate entirely within the 100 bp window with no overlap in the following window. Reads with overlaps between windows was calculated in the subsequent window that had the 50bp overlap as shown in Figure 2.2.

Figure 2.2: Whole-genome sliding window

The blue line represents the genome while the green lines represent the sRNA reads. The genome is broken up into 100bp segments with a sliding window 50 bp overlap as shown by the red dotted lines. sRNA reads within the red box are counted for the 100 bp window.



A negative binomial distribution comparative analysis for 1, 3 and 6 DPI were performed using EdgeR version 3.0 (Robinson et al., 2010, Robinson and Smyth, 2007, Robinson and Smyth, 2008, McCarthy et al., 2012) and DESeq version 1.13.0 (Anders and Huber, 2010). The reads were normalised to the size of the library for each replicate. sRNA density between wild type and the mutant or mock and *Fusarium* infection were regarded as having significant, altered sRNA abundance if the read count had a $\log_2$ fold difference greater than 1 and an adjusted P-value of less than 0.05.

2.62 Bioinformatic Analysis of mRNA Sequencing Data

The mRNA sequencing reads were received from the Australian genome research facility (AGRF). Reads longer than 100 nt were filtered out using the CLC genomics workbench (version 4.9) to ensure that only high-quality mRNA reads without adaptors remained. Trimmed and filtered reads were mapped to the TAIR 10 *Arabidopsis thaliana* genome (from the CSIRO Plant Industry Bioinformatics Group Genome Browser) using the CLC genomics workbench RNA sequencing analysis application. No mismatches were allowed in the alignment of reads against the genome.

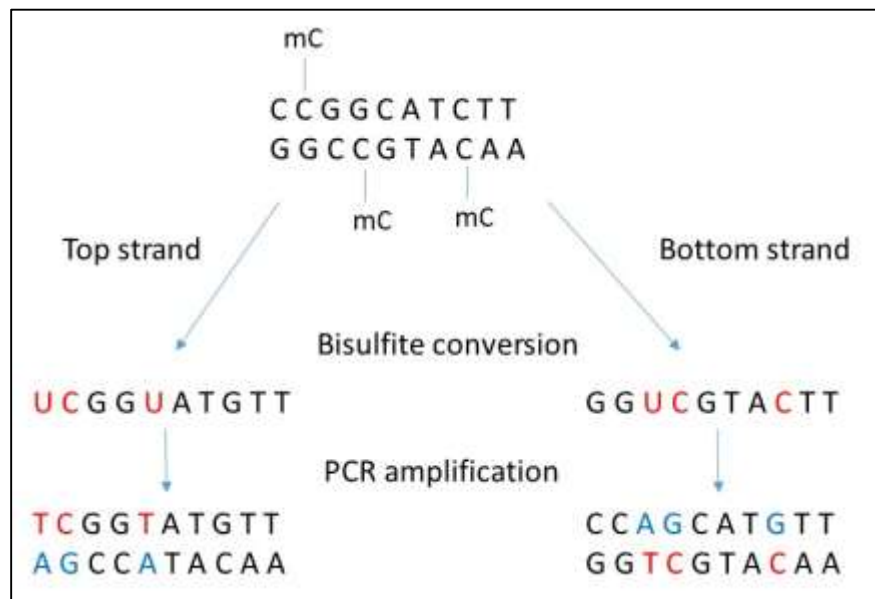
A negative binomial distribution comparative analysis was performed using DESeq (Anders and Huber 2010). The reads were normalised to the size of the library for each replicate. To determine the threshold of *p* value in multiple tests the FDR (false discovery rate) was applied. A FDR of ≤ 0.05 and $\log_2\text{fold} \geq 1$ or ≤ -1 was used to judge significant differential expression.

2.63 Bioinformatic Analysis of Bisulfite Sequencing Data

Bisulfite sequencing libraries were sequenced using the HiSeq 2500 as per manufacturer's instructions. The bisulphite converted genomic DNA reads were mapped to the *Arabidopsis thaliana* genome using the Bismark aligner version 0.10.0 (Krueger and Andrews, 2011). Sequence reads are first altered to represent fully bisulfite-converted Watson (C→T) and Crick read (G→A) versions before alignment to a similarly converted *A. thaliana* genome (Figure 2.3). The four alignment processes are run in parallel with two mismatches in the 'seed' length, the first 20 bp of the read, allowed. The read which has fewer mismatches to the genome than any other alignment, or has only one alignment is considered to align uniquely and kept for further analysis. The other reads were removed. Short read sequences that mapped to the same starting position were collapsed into a single consensus read to reduce clonal bias. Individual cytosine methylation calls were assigned to one of the three cytosine contexts (CG, CHG or CHH where H=T, C or A) using the Bismark methylation extractor.

Differentially methylated regions (DMRs) were called from 100 bp non-overlapping windows for the Col-0 vs *ddd* comparisons and the mock vs *Fusarium* infected comparisons. To be considered a DMR, both samples of the comparison had to show at least 10x coverage with 80 % difference in CG, 50 % in CHG, and/or 20 % in CHH within the 100 bp window. The percentage of methylation was based on general methylation levels for the *Arabidopsis* genome (Hsieh et al., 2009). Adjacent windows meeting the criteria were combined and average methylation levels across all samples were gathered for the defined DMR regions.

Figure 2.4. Explanation of bisulfite conversion-PCR



Annotated TAIR10 sequences of protein coding gene were retrieved from the CSIRO Plant Industry Bioinformatics Group Genome Browser. Regions 3 kb upstream and 3 kb downstream of annotated protein coding genes were also obtained. The protein coding genes and 3kb flanking regions were aligned against the DMRs using Bedtools intersect run in Cygwin.

The whole genome bisulfite methylation data was validated using the same genes selected for RT-PCR gene expression validation. The DNA methylation pattern of reads sequenced for this study were shown in IGV (Thorvaldsdottir et al., 2013) and compared against the DNA methylation pattern of reads from the browser of an independent whole genome bisulfite sequencing experiment; available at <http://genomes.mcdb.ucla.edu/AthBSseq/> (Stroud et al., 2013). Only the DNA methylation patterns of Col-0 and *ddd* under mock conditions were compared against this DNA methylation browser because of the comparable plant conditions.

2.64 General Bioinformatics

Annotated sequences of protein coding genes, transposons, pseudogenes, ncRNAs and tRNAs were retrieved from the CSIRO Plant Industry Bioinformatics Group Genome Browser. Regions 3 kb upstream and 3 kb downstream of annotated protein coding genes were also utilised. This distance was chosen as promoter sequences can vary in size from 500 – 3000 nt (Baev et al., 2010). These sequences were aligned against the differentially abundant (DA) mRNA using Bedtools intersect version 2.17.0 (Quinlan and Hall, 2010) run in Cygwin (Racine, 2000). Reads were visualised along the chromosome using the Integrated Genome Viewer (IGV) platform.

Functional pathway characterisation of gene annotations was performed by Go annotations with MapMan (Thimm et al., 2004) and statistical $\log_2$ fold change determined using Excel (Microsoft, Washington, DC, USA).

2.7 Quantitative Real-Time PCR

Quantitative real-time PCR (qRT-PCR) was performed for gene expression verification of 2.5 week old plants either mock (water) or *Fusarium* treated, and placed on MS agar without sucrose for 1, 3 or 6 DPI, as used for sRNA and mRNA sequencing analyses.

2.71 qRT-PCR Primer Design

qRT-PCR primers were designed using Primer 3 (Rozen and Skaletsky, 2000). Default settings were used except for product size range which was chosen to be 150-200 base pairs and the last 2 base pairs of the primer were selected as a G or a C (GC clamp: 2) to increase primer binding strength. The gene specificity of primers was checked by blasting the full length CDS in The Arabidopsis Information Resource (TAIR, <https://www.arabidopsis.org/>). Primers were selected if they matched only the intended target gene sequence in particular at the 3'-region of the primer.

Primer specificity during qRT-PCR amplification was confirmed by 1) a single peak in the melting temperature curve analysis; and 2) a single band of the correct size as shown in products run on a 2 % agarose gel electrophoresis. A list of RT-PCR primers used in this study are shown in Table 2.3.

2.72 DNase Treatment of RNA

Ten microlitres of RNA was DNase-treated with 5 μ L RQ1 10x DNase buffer (Promega), 3 μ L RQ1 DNase (Promega) and 32 μ L DEPC water at 37 °C for 30 minutes. RNA was then purified by phenol/chloroform extraction (Ambion) and chloroform (POCD healthcare) extraction followed by ethanol precipitation as for purification of restriction digested DNA. RNA pellet was air-dried before being resuspended in 15 μ L DEPC water.

2.73 cDNA Synthesis

Into a 1.5 mL Eppendorf tube 10 μ L of DNase-treated RNA (2-5 μ g), 0.5 μ L of 50 pmol oligo dT (Sigma), 4 μ L 5x 1st strand buffer (Invitrogen) and 5.5 μ L water was added. The tube was placed in boiling water for 3 minutes then cooled and reaction mixture transferred to a PCR tube on ice. To the tube 5 μ L of 5 mM dNTPs (Fischer), 1 μ L RNase Inhibitor (RNase Out, Invitrogen), 4 μ L of 5x 1st Strand Buffer (Invitrogen), 2 μ L 0.1 M DTT (Invitrogen), 2 μ L SuperScript III (Invitrogen) and 6 μ L water was added. The tube was incubated at 55 °C for 1 hour.

Table 2.3: RT-PCR Primers

Probe	Description	5'-3'
AT3G18780	Actin2	AGTGGTCGTACAACCGGTATTGT (F)
		GATGGCATGAGGAAGAGAGAAAC (R)
AT5G43940	FDH	TTGGACTTGCTGTTGCCGA (F)
		TGGCTTGTCGTGATCCTTTGG (R)
AT1G58602	CC-NBS-LRR	TGGAAGAAGAGATTGTCGGAAGGAGG (F)
		TCCTGGCTCTGTCTGTGATTGCTG (R)
AT3G10010	DML2	CCTTTGAATGGCACATACTTCC (F)
		GCTTACGATTCTCCCTGTCG (R)
AT2G04795	Unknown Protein	GACGGTTAATGGAGATCACG (F)
		TTAACCTCCGAGACCCTTCC (R)
AT2G33830	Dormancy/Auxin associated family protein	GACCTAAACCGGAGCATGG (F)
		CTTGTCGAAGAGGTTTGTGC (R)
AT4G01350	Cysteine/histidine-rich C1 domain family protein	TTCTCACGTCTCGTGCTAGG (F)
		CCGATTGTATTCATTACGTTCC (R)
AT1G34245	Epidermal patterning factor 2	TTTGGTCGTAACTCCATTCG (F)
		CACACGCGTATGAACAATCC (R)
AT5G06730	Peroxidase superfamily protein	CTGTCGGGCTAAAGACAACC (F)
		CATCAGGTGTGCTCAGATCG (R)

Four microlitres of cDNA from each of the samples was combined and a 1:10, 1:100 and 1:1000 serial dilution made. These dilutions were used as the standard curve for the qRT-PCR. The remaining cDNA was diluted and transferred to the source plate for qRT-PCR.

2.74 Quantitative PCR

Quantitative PCR was run on a Rotor-Gene 6000 thermo cycler (Corbett Life Science, San Francisco, CA, USA). PCR reaction consisted of 5 μ L cDNA, 0.8 μ L of 5 μ M primer pairs, 0.75 μ L dH₂O, 1 μ L buffer (Invitrogen), 0.7 μ L of 50 mM MgCl₂ (Invitrogen), 0.4 μ L 5 mM dNTP (Fischer) and 0.5 μ L Syber Green (1:1000, Invitrogen). Each sample was analysed in triplicates. The PCR condition was shown below:

Stage 1: 95 °C for 10 minutes

Stage 2: 95 °C for 10 seconds, 60 °C for 20 seconds, 72 °C for 20 seconds for 40 cycles

Stage 3: 95 °C for 15 seconds, 60 °C for 15 seconds, 95 °C for 15 seconds

The $\Delta\Delta$ Ct method (Livak and Schmittgen, 2001) was used to determine relative expression of target genes against mock treated Col-0 plants. The house-keeping genes AT3G18780 (Actin 2) and AT2G26250 (FDH) were used as an internal reference for calculating the relative gene transcript levels under investigation. The cDNA sample without reverse transcriptase was used as a negative control to prove no genomic DNA contamination in the RNA samples.

**Chapter 3: Col-0 and DNA Demethylase Mutants Show Differential
Phenotypic Responses and Small RNA Accumulation upon
Fusarium oxysporum Infection**

3.0 Introduction

As reviewed in Chapter 1, both RdDM and DNA demethylases have been implicated in plant disease resistance. For instance, the Pol V mutant, *nrpe1*, shows increased disease susceptibility to the necrotrophic fungal pathogens *Botrytis cinerea* and *Plectosphaerella cucumerina* and enhanced resistance to the bacterial pathogen *Pseudomonas syringae* (Lopez et al., 2011). The *rdr2*, *drd1* and the triple mutants *dcl2/3/4* and *drm1/drm2/cmt3* (*ddc*) showed enhanced resistance to *P. syringae* (Yu et al., 2013). In the DNA demethylation mutant, *ros1*, an increase in *P. syringae* growth was reported suggesting a role of the ROS1 demethylase in bacterial resistance. Our own laboratory also demonstrated that the triple demethylation mutant (*rdd*) was more susceptible to *Fusarium oxysporum* infection (Le et al., 2014) compared to Columbia. These studies have implicated an interaction between RdDM and DNA demethylases in plant defence. Additional studies have shown the RdDM mutants *ago4* and *nrpe1* were also more susceptible than wild type Col-0 to *Fusarium* infection, and there was a strong overlap in downregulated stress response genes between *rdd* and the RdDM mutants *nrpe1* and *nRPD1* (Pol IV) (Le et al., 2014). However, it remains unclear how these two pathways interact. Furthermore, how these pathways respond to *Fusarium* infection needs to be investigated. In addition, while the results using the *rdd* triple mutant indicated a role of the DNA demethylases in *Fusarium* resistance, how the individual demethylases ROS1, DML2 and DML3 contribute to defence has not been well studied.

In this chapter I will focus on phenotypic assessment of the *rdd* mutant, as well as the single DNA demethylase gene mutants *dml2*, *dml3* and *ros1* prior to and following *Fusarium* infection. In an attempt to understand if DNA demethylases have a broad function in biotic stress response, I will compare the phenotypes of *rdd* and Col-0 in response to infection by two viruses Cucumber mosaic virus (CMV) and Turnip mosaic virus (TuMV). As a preparation for genome-wide small RNA (sRNA) deep sequencing analysis, I will examine if sRNAs are responsive to *Fusarium* infection and determine suitable time points for RNA sample preparation using sRNA northern blot hybridization. Specifically, in this chapter I will address the following questions:

1. Does the *rdd* mutant and the single demethylase mutants *dml2*, *dml3*, and *ros1* differ from Col-0 under standard conditions or under *Fusarium* infection?
2. How does *rdd* respond to CMV or TuMV infection?
3. Do Col-0 and *rdd* show differential sRNA accumulation in response to *Fusarium* infection and what is the time course of sRNA response following *Fusarium* infection?

3.1 Results

3.1.1 DNA Demethylase Mutants Show no Phenotypic Differences to Col-0 Under Normal Growth Conditions

The phenotypes of *rdd* and the single DNA demethylase mutants *ros1*, *dml2* and *dml3* were compared with Col-0 both in soil and in MS-medium without sucrose (sucrose-free MS; Le et al., 2014). Based on visual observations, all mutant lines were phenotypically identical to Col-0 plants both in soil (Figure 3.1) and on sucrose-free MS plates (Supplementary Figure 3.1).

Seed germination and flowering rates were visually monitored on a daily basis to further assess if any phenotypic differences were shown in the mutants compared to Col-0. Germination of mutant seed was observed three days after sowing which was consistent with Col-0 (Supplementary Figure 3.2). No obvious difference in the time or frequency of flowering was seen among the five Arabidopsis lines (Supplementary Figure 3.3). Root length was also examined 16 days post sowing for the DNA demethylase mutants and compared to Col-0 (Supplementary Figure 3.4). No significant differences were seen although *ros1* and *dml2* may have slightly shorter roots than Columbia, *rdd* and *dml3*.

3.1.2 DNA Demethylase Mutants Show Differential Phenotypic Differences to Col-0 in Response to Fusarium Infection

To assess phenotypic differences among the DNA demethylase mutants and Col-0 plants upon Fusarium infection, I first examined disease development of soil-infected plants. As shown in Figure 3.2 A-E, all plant lines showed disease symptoms with various degrees of leaf yellowing or chlorosis. To measure subtle differences in disease response among the five lines, a quantitative disease score was given to each plant at 10 days post inoculation (DPI) resulting in an average disease score for each line (Figure 3.2F). The result shows *rdd* was more susceptible to Fusarium infection than Col-0 as previously reported (Le et al., 2014). The single mutants were also more susceptible than Col-0 to Fusarium infection, with *ros1* showing the highest disease score among the three. However, the single mutants were less susceptible than the triple mutant *rdd*. This result suggested that the three DNA demethylase genes have some functional redundancy in plant response to Fusarium infection.

Figure 3.1: DNA demethylase mutants show no visible phenotypic differences to Col-0



Figure 3.2: Phenotypes of DNA demethylase mutants and Col-0 at 10 days post *Fusarium oxysporum* infection in soil (A-E), (F) Mean disease score of chlorotic leaves

Disease symptom scoring (all plants had black roots): 0 = no chlorotic leaves, 1 = 1-3 leaves showing chlorosis, 2 = 4-6 leaves showing chlorosis, 3 = all leaves showing chlorosis, 4 = dead plant. Standard error bars are shown on the histogram. At least 30 plants were assessed per line.

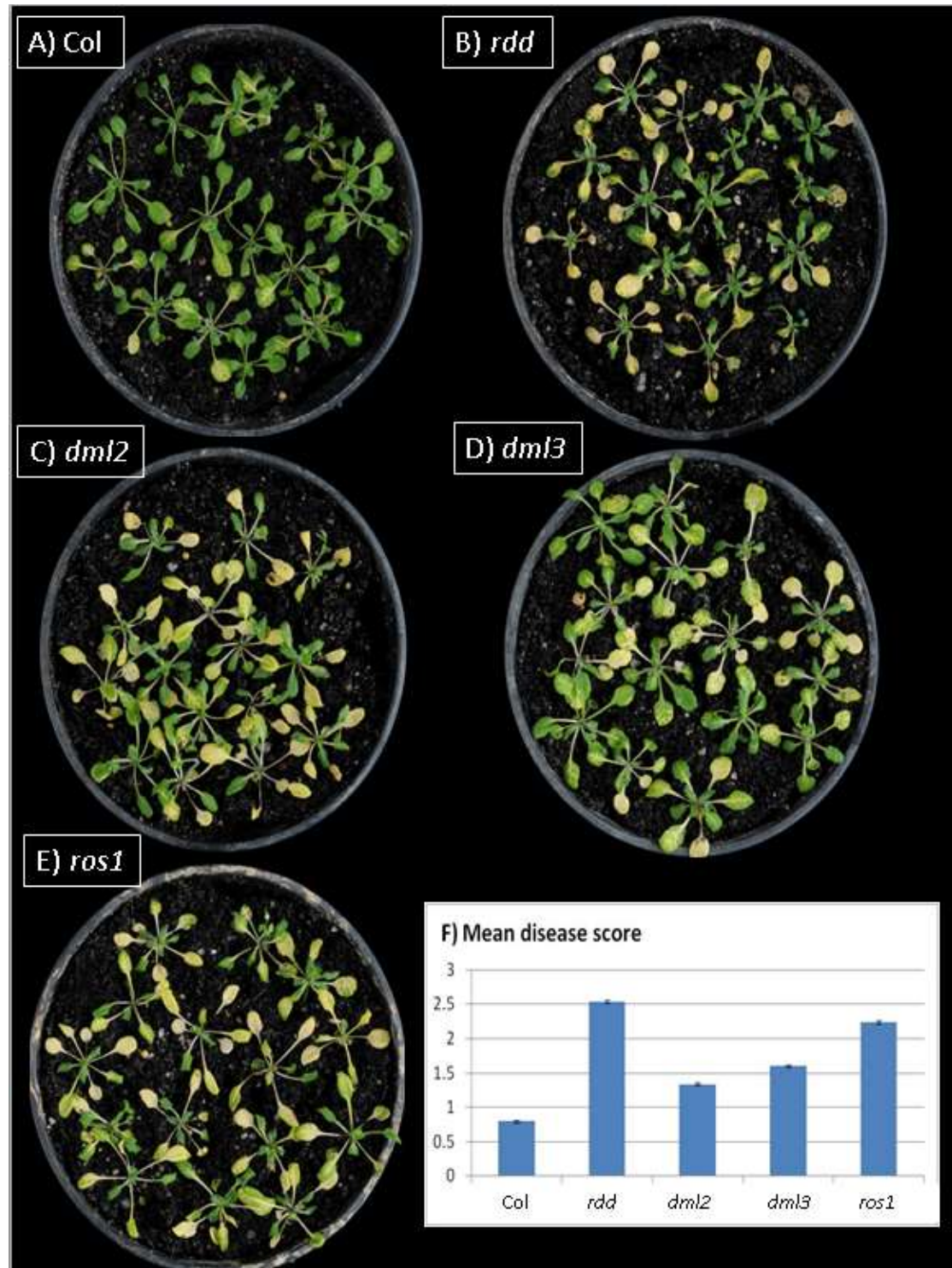
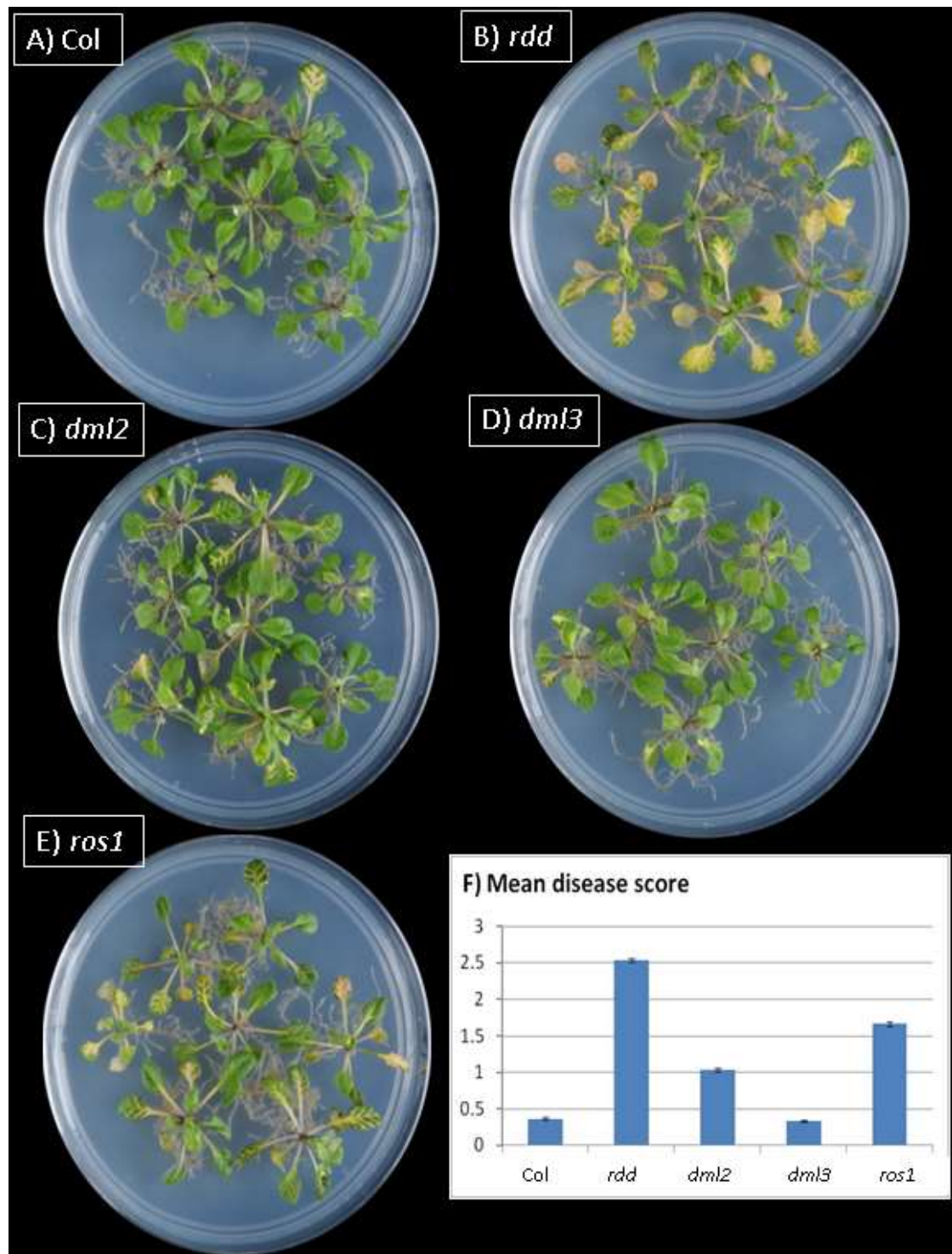


Figure 3.3: Phenotypes of DNA demethylase mutants and Col-0 8 days post *Fusarium oxysporum* infection on sucrose-free MS (A-E), (F) Mean disease score of chlorotic leaves

Disease symptom scoring (all plants had black roots): 0 = no chlorotic leaves, 1 = 1-3 leaves showing chlorosis, 2 = 4-6 leaves showing chlorosis, 3 = all leaves showing chlorosis, 4 = dead plant. Standard error bars are shown on the histogram. At least 30 plants were assessed per line.



On sucrose-free MS plates *Fusarium* infection of *Arabidopsis* is faster and more uniform than in soil. I therefore assessed the *Fusarium*-response phenotypes using plate infection. As shown in Figure 3.3, *rdd* plants displayed the most severe response to *Fusarium oxysporum* infection at 8 DPI as demonstrated by the significant number of chlorotic leaves. The single DNA demethylase mutants *ros1* and *dml2* showed intermediate phenotypes between *rdd* and Col-0. Different to the soil infection experiment, the *dml3* mutant showed almost identical phenotype to the wild type Col-0. These visual observations were supported by the mean disease score based on the number of chlorotic leaves (Figure 3.3F).

3.1.3 DNA Demethylase Mutants show Differential Phenotypic Responses to Viral Infection

One question asked was if DNA demethylases play a broad role in plant biotic stress response. While focusing on *Fusarium oxysporum* as the pathogen system in this thesis, I was interested in assessing DNA demethylase function using non-fungal pathogens such as the bacteria *Pseudomonas syringae* and viruses. However due to time limitation I only analysed the response of *rdd* and Col-0 to two viruses, Cucumber mosaic virus (CMV) and Turnip mosaic virus (TuMV). Interestingly, *rdd* appears to be more resistant to TuMV (Figure 3.4) and to a lesser extent to CMV (Supplementary Figure 3.5) when compared to the wild type Col-0, as indicated by stronger growth and less necrosis of leaves. This was confirmed by RNA northern blot hybridization showing lower levels of CMV virus accumulation in *rdd* than in Col-0 (Figure 3.5). While more experiments are needed to confirm and understand this differential viral response by *rdd*, this result did raise the possibility that DNA demethylases may play a broad role in plant disease resistance.

3.1.4 Assessing Spraying as an Alternative way of *Fusarium* Inoculation

The experiments described so far involved *Fusarium* inoculation via root dipping. With the root dipping method, plants needed to be temporarily uprooted which may cause physical injury to roots that could result in non-*Fusarium*-specific responses. To minimize such non-specific responses, I tested if spraying could be used to inoculate *rdd* and Col-0 plants. I then compared spraying with the root dipping method, details in the materials and methods Chapter 2.

Figure 3.4: Phenotypes of *rdd* and Col-0 at 10 days following Turnip mosaic virus (TuMV) infection

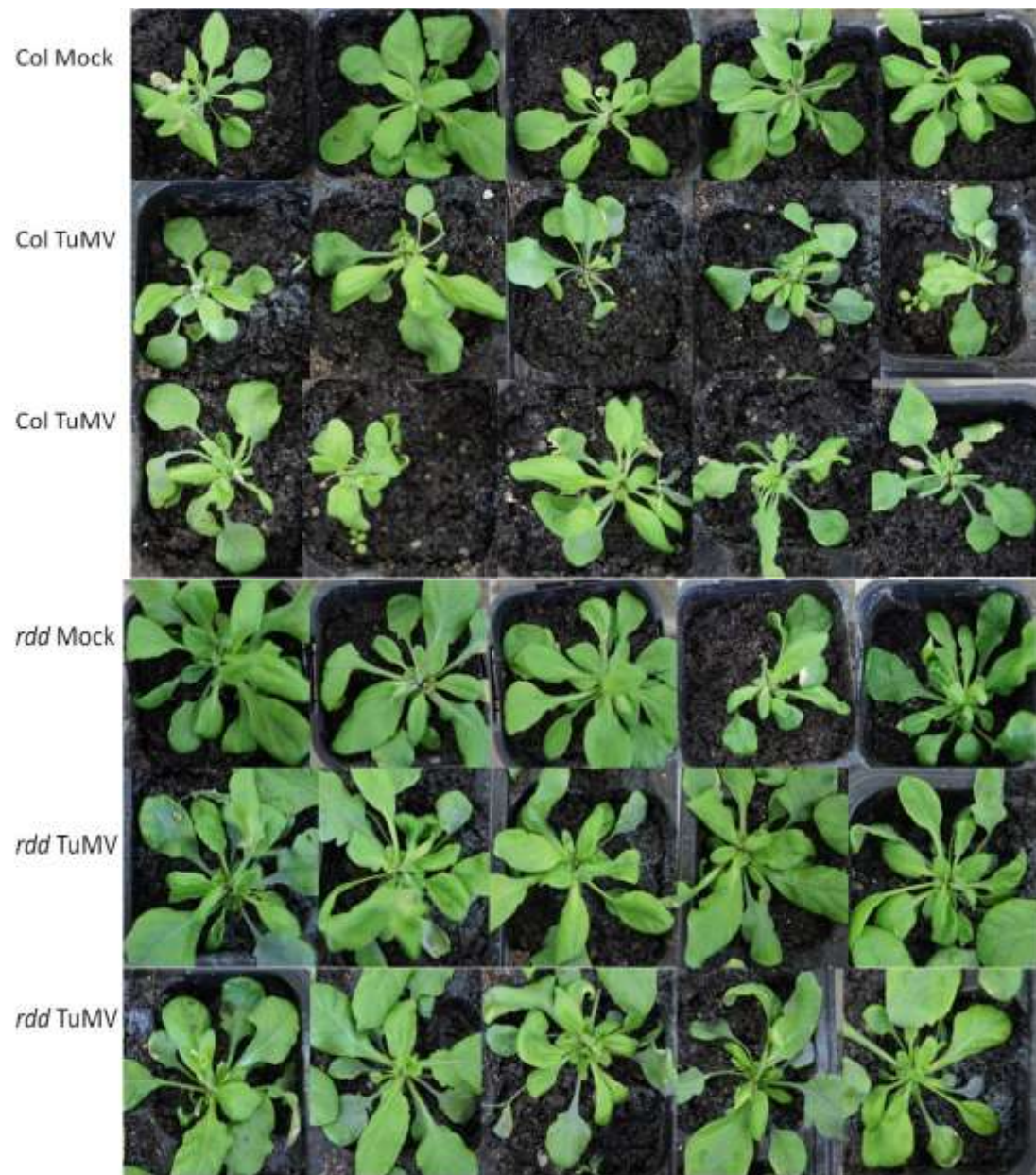
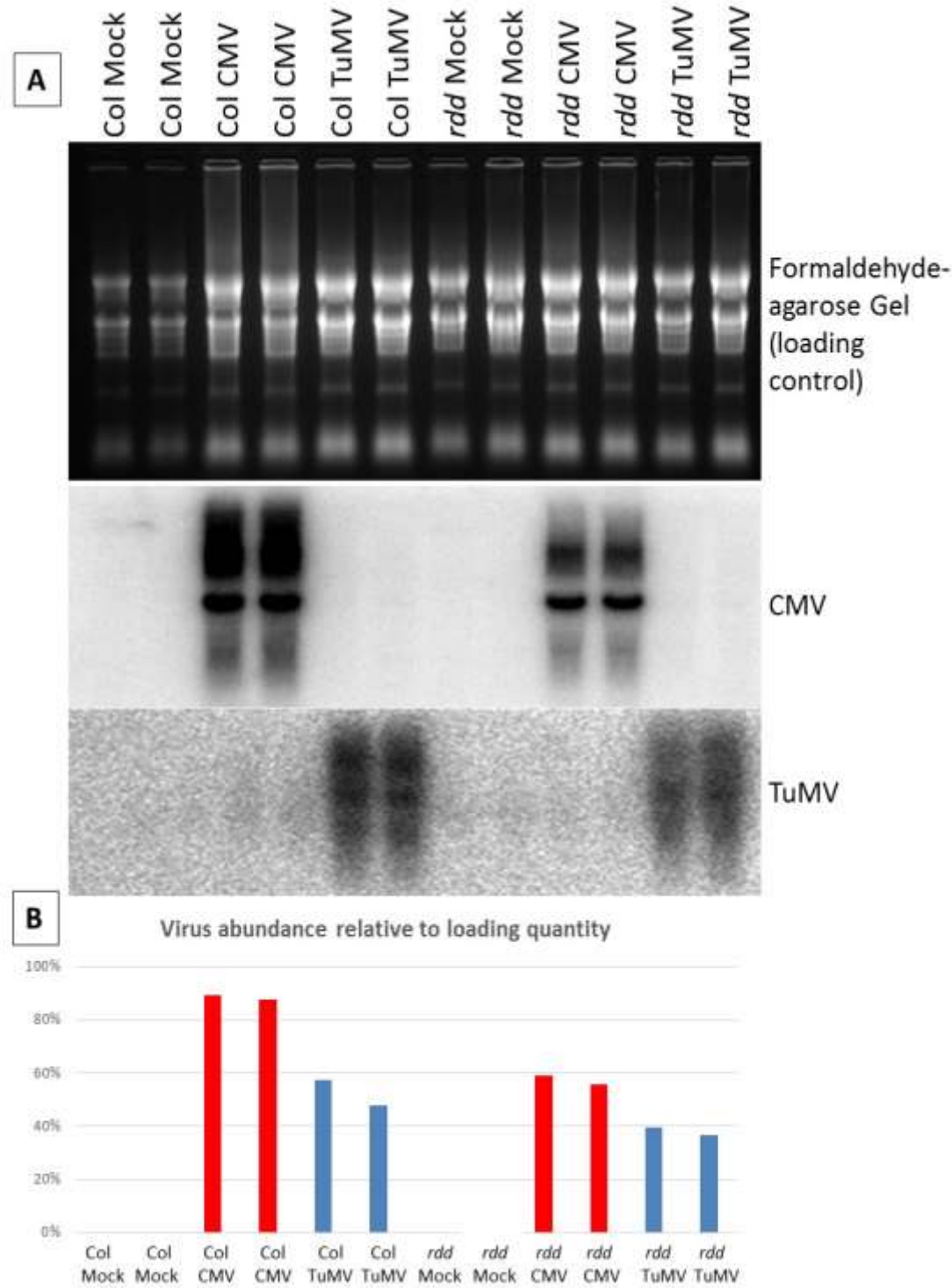


Figure 3.5: TuMV and CMV RNA accumulation in infected Col-0 or *rdd* plants at 10 DPI A) Northern blot analysis B) Quantification of bands relative to loading quantity



Basically, Arabidopsis plants to be spray-inoculated were grown on MS plates for 1 week and then transferred to sucrose-free MS plates where they continued to grow for an additional 2 weeks. Suspension of 1×10^6 Fusarium spores/mL was then sprayed on the 3 week old plants. Photos were taken to monitor the symptom progression over time. With Fusarium spray inoculation, Figure 3.6A, the *rdd* plants clearly showed stronger disease susceptibility to Fusarium infection than Col-0. Many *rdd* plants died at 10 DPI whereas the Col-0 plants only displayed leaf chlorosis. This indicated that spray inoculation could be used to infect *rdd* and Col-0. However, as shown in Figure 3.6A, inoculation with root dipping gave a clearer phenotypic difference between *rdd* and Col-0. More importantly mock inoculated plants, corresponding to the spray inoculated plants, showed stunted growth. This was presumably due to an extended period of growth on sucrose-free MS medium. In contrast, the mock-treated plants corresponding to plants with the roots dipped in inoculum, which had been grown on normal MS plates until infection, did not show such the stunting phenotype. Consequently, all future Fusarium inoculations were performed with the root dip method.

3.1.5 Quantitative Measurement of Fusarium Biomass in Infected *rdd* and Col-0 Plants

To further assess the difference in Fusarium response between *rdd* and Col-0, I performed a quantitative assay of fungal biomass using the chitin method described in Chapter 2. Basically, the assay is based upon chitin detection using wheat germ agglutinin (WGA) conjugated to a fluorophore, which allows chitin to be measured based on the level of fluorescence. Three biological replicates were used for each sample, with four technical duplicates for each biological replicate. As fluorescence can vary depending on the staining time results from separate assays are not comparable. The mock treated plants generally were similar in fluorescence between samples. This is the background fluorescence noise in plant tissue as non-fungal eukaryotes do not have chitin to which the WGA can bind (Ayliffe et al., 2013).

Figure 3.6: Phenotype of Col-0 and *rdd* at 10 DPI after *Fusarium oxysporum* infection

For spray inoculation, plants were grown on normal MS plates until 1 week of age then were transferred to sucrose-free MS plates. At 3 weeks plants were inoculated with the fungus.

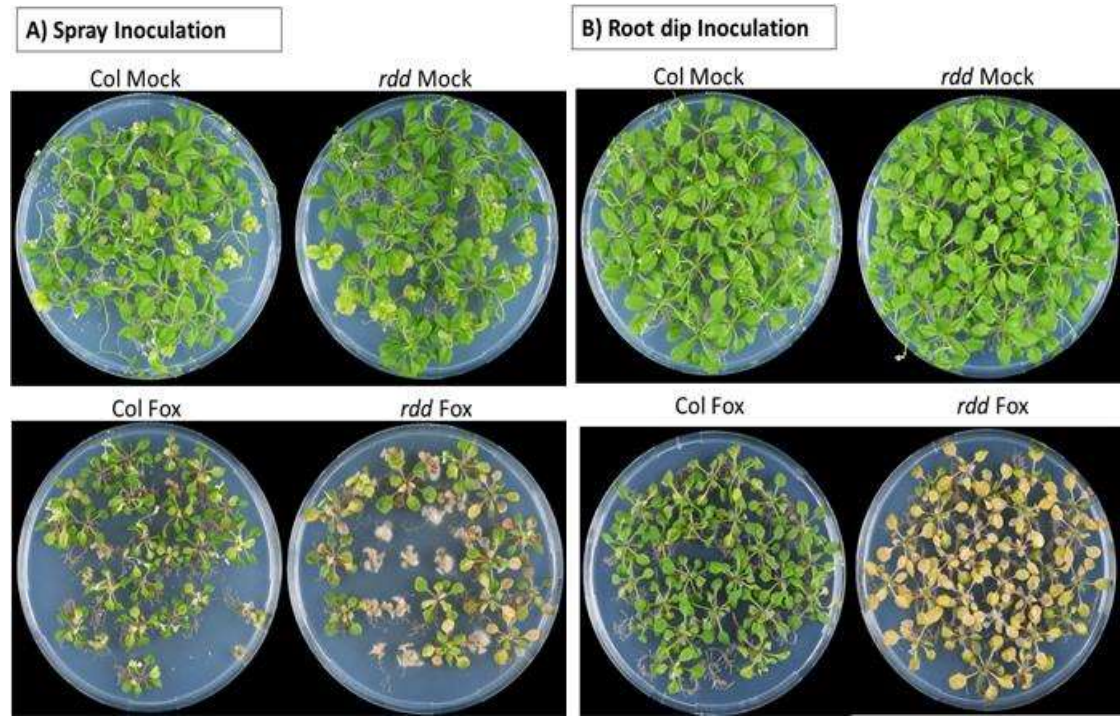
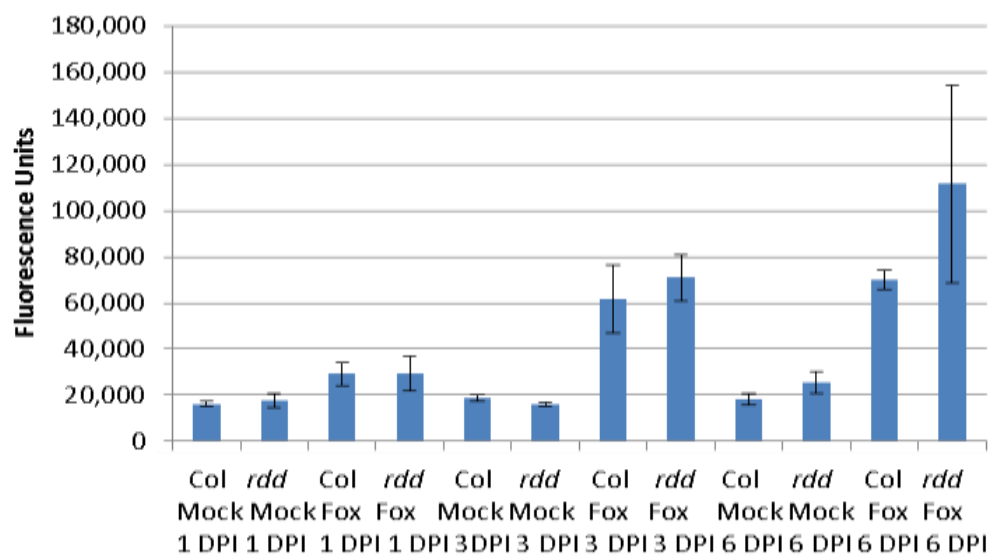


Figure 3.7: Measurement of *Fusarium* biomass in whole plant tissue using florescent labelled WGA binding of chitin



Fungal biomass in whole tissue of plate-inoculated plants was measured at 1, 3 and 6 DPI (Figure 3.7). Across all time points the mock samples showed low background fluorescence levels as expected. At 1 DPI in *Fusarium* infected samples, Col-0 and *rdd* had slightly higher fluorescence than the mock samples indicating low levels of *Fusarium* in plant tissues. Fungal biomass showed a strong increase at 3 DPI, as indicated by the high levels of fluorescence. However, from 3 to 6 DPI, fungal biomass in Col-0 did not show further increase, although there was some increase in *rdd*. This result indicated that the first three days were important for colonization of *Fusarium* in Col-0 and *rdd* plants, providing a basis for time point selection in subsequent sRNA and mRNA sequencing analysis. The result in Figure 3.7 was consistent with that from a separate experiment that measured root and shoot tissue separately (Supplementary Figure 3.6), therefore validating the use of whole plant tissue in the chitin quantitative assay.

Interestingly, the *rdd* plants showed no difference in fungal biomass at 1 DPI to Col-0, only a slight increase at 3 DPI and some, but not dramatic, increase at 6 DPI (Figure 3.7). This was consistent with the previous RT-PCR and chitin assay showing only a subtle increase in *Fusarium* accumulation in roots of soil infected *rdd* at 10 DPI, and in roots and shoots of plate inoculated *rdd* at 3 DPI, compared to Col-0 (Le et al., 2014). This suggests that the increased disease susceptibility observed in *rdd* is not due to enhanced presence of *Fusarium* in root tissues but is likely caused by increased sensitivity of the plant to disease symptom development.

3.1.6 Col-0 and *rdd* Show Differential Small RNA Accumulation in Response to *Fusarium* Infection

To examine if the accumulation of RdDM associated 24 nt siRNA is responsive to *Fusarium* infection, and to determine the time points suitable for subsequent high-throughput sRNA sequencing analysis, I performed sRNA northern blot hybridization analysis of selected sRNA species in Col-0 and *rdd* at different time points following *Fusarium* infection.

The siRNAs chosen for analysis were from a set of sRNAs known to be associated with RdDM, including the TE-derived siRNA1003, siRNA02 and AtSN1 siRNAs, and the cluster 4 siRNA of an unknown genomic source (He et al., 2009, Gao et al., 2010). All these siRNAs have been used as hallmarks to determine RdDM function in *Arabidopsis*. 24 nt siRNAs corresponding to a repetitive element upstream of the AGO4 gene, were also used as a target in my sRNA northern blot analysis. A miRNA, miRNA173, and a tasiRNA, ta-siRNA255, which function in PTGS, were used as non-RdDM pathway controls. U6 RNA was hybridized as a sample loading control. The sequences of all the probes used for the sRNA northern hybridization are listed in Chapter 2, Table 2.2.

An analysis of AGO4 and siRNA02 accumulation in *Fusarium*-infected Col-0 at 0, 8 and 24 hours post inoculation did not detect clear differences among the time points (Supplementary Figure 3.7). The *rdd* plants were not analysed at these specific time points, but an analysis of AGO4-derived siRNAs at 6 hours post inoculation (HPI) and 3 DPI for both Col-0 and *rdd* revealed an effect of *Fusarium* infection on siRNA, particularly in *rdd* at 3 DPI, showing a clear increase in siRNA signals in *Fusarium*-infected compared to uninfected plants (Figure 3.8A). I also analysed the accumulation of AGO siRNA and siRNA02 at 1 and 6 DPI, which showed again that AGO4 siRNA accumulation was increased in *rdd* upon *Fusarium* infection, more so at 6 DPI than 1 DPI (Figure 3.8B). Interestingly, an increase in AGO4 siRNA accumulation could also be observed for Col-0 at 6 DPI (Figure 3.8B). A similar trend was observed for siRNA02, although the band pattern looked more complex than AGO4 siRNA. In addition to AGO4 siRNA and siRNA02, a clear increase in sRNA abundance was also observed for siRNA1003 and cluster 4 siRNA at 6 DPI compared to 1 and 3 DPI (Figure 3.9). The different temperatures (26 °C as opposed to 22 °C) used for *Fusarium* infection did not affect small RNA abundance (Figure 3.8).

In contrast to the RdDM associated siRNAs, the accumulation of the PTGS-associated sRNAs, miR173 and ta-siRNA255, did not show significant changes between Col-0 and *rdd*, or between mock-treated and *Fusarium*-infected plants (Figures 3.8 and 3.9, and Supplementary Figure 3.7). Taken together, the sRNA northern blot hybridization results suggest the accumulation of RdDM associated siRNAs is responsive to *Fusarium* infection, and that DNA demethylases play a role in this response. Furthermore, these results indicated that this *Fusarium* response is increasing with time, becoming particularly prominent at 6 DPI. We therefore decided to perform sRNA deep sequencing on 1, 3 and 6 DPI plant samples.

Figure 3.8: sRNA northern blot analysis of mock treated and *Fusarium* infected Col-0 and *rdd* at different time points as specified

The top panel (A), plants were incubated at 22°C following *Fusarium* infection, while for the lower panel (B), plants were incubated at 26°C. M, mock-treated; F, *Fusarium*-infected. Uniform loading was verified by the U6 probe.

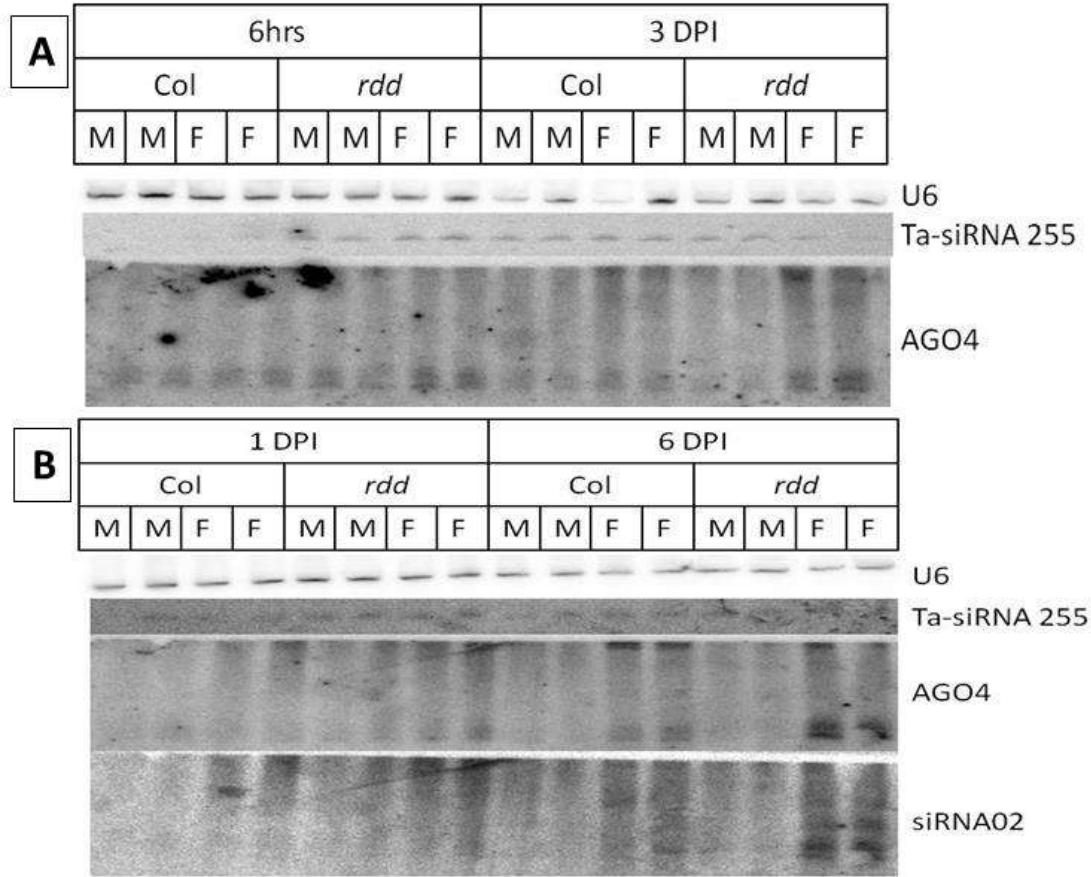
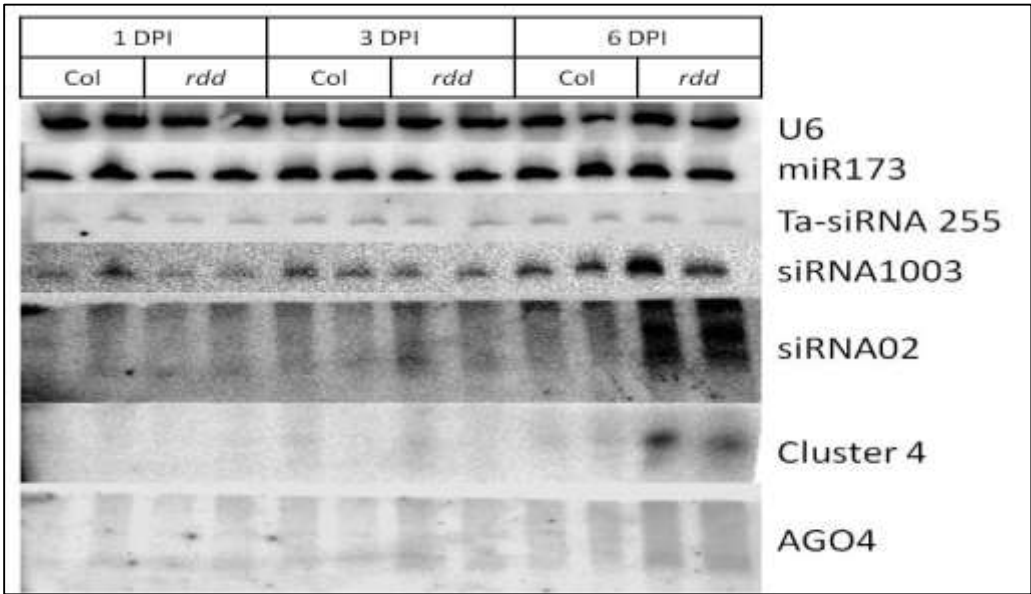


Figure 3.9: sRNA northern blot analysis of *Fusarium*-infected Col-0 and *rdd* plants at 1, 3 and 6 DPI

Uniform loading was verified by the U6 probe.



3.2 Discussion

In this chapter I examined i) the phenotypes of triple and single DNA demethylase mutants prior to and following *Fusarium* infection, ii) response of *rdd* plants to infection by viruses, iii) spray inoculation of Col-0, and *rdd* with *Fusarium* and its comparison to root dipping inoculation, iv) biomass difference of *Fusarium* accumulation in infected *rdd* and Col-0 plants, and v) time course of *Fusarium*-induced sRNA changes. The objectives of these experiments were to verify stress response differences between the DNA demethylases and wild type Col-0, and to establish the time course suitable for sRNA, mRNA and DNA methylation analyses in subsequent chapters.

Consistent with previous reports (Le et al., 2014, Penterman et al., 2007b) the DNA demethylase mutants showed no visible phenotypic differences to Col-0 under normal growth conditions, both in soil and on MS plates. This was true for both the appearance of plants, and germination and flowering time. Our laboratory has previously shown that the triple DNA demethylation mutant, *rdd*, was more susceptible to *Fusarium* infection compared to Col-0 (Le et al., 2014, Zhu et al., 2013). To further verify this and test the redundancy of the three individual DNA demethylase genes, I assessed Col-0, *rdd* as well as the single DNA demethylase mutants *ros1*, *dml2* and *dml3* for *Fusarium* response. While all plant lines developed disease symptoms with time, the *rdd* mutant showed the quickest onset and most severe symptoms and the Col-0 plants the slowest and weakest. For the three single DNA demethylase mutants, *ros1* showed the strongest disease response in both soil and plate infection; *dml2* also showed enhanced disease susceptibility compared to Col-0 in both soil and plate infection, but was weaker than *ros1*. The *dml3* mutant, although weaker than *ros1*, showed significant enhancement of *Fusarium* susceptibility compared to Col-0 in soil infection but did not display phenotypic difference to Col-0 on sucrose free MS plates. These results confirmed the *Fusarium* susceptibility of the DNA demethylase mutants compared to Col-0, and suggested functional redundancy among the three DNA demethylase genes. As *rdd* consistently showed the strongest disease response to *Fusarium*, all further experiments were carried out with *rdd* and Col-0.

The visualization of *Fusarium* colonization is important for understanding the dynamics of specific cell colonization. An earlier study using fluorescent proteins and microscopic imagery showed that *Fusarium* entered the vascular system via the roots as early as 2-3 days after inoculation, depending on the *Fusarium* species and plant host, which subsequently resulted in interveinal chlorosis and eventual necrosis of leaves (Bishop and Cooper, 1983). More recent work (Czymmek et al., 2007, Diener, 2012) has been done on the invasion of *Arabidopsis*

thaliana by *Fusarium oxysporum* in soil. These soil studies revealed that *Fusarium* will have colonised the root at least 4 days post infection. These are excellent studies that provide significant insight into the process of *Fusarium* colonization during soil infection. However, in my thesis the sRNA, mRNA and DNA demethylation analyses would be based on using plants inoculated on sucrose free MS plates as this plate inoculation gave a more consistent and uniform plant infection than soil inoculation. It was therefore useful to examine the time course of colonization by *Fusarium oxysporum* f. sp. *conglutinans* strain Fo5176 in Col-0 and *rdd* plants on sucrose-free MS plate. This was done through fungal biomass quantification using the chitin method developed by Ayliffe and colleagues (2013).

Compared to the background florescent level (0 DPI), *Fusarium* appeared to have already colonised *Arabidopsis* cells at 1 DPI, but at 3 and 6 DPI the colonisation became very clear in both *rdd* and Columbia. However, no clear difference in *Fusarium* biomass was observed at the 1 and 3 DPI between *rdd* and Columbia. At 6 DPI, the *rdd* mutant plants seemed to have increased *Fusarium* biomass but a statistical significance could not be confirmed due to moderate errors between the replicates. This biomass result provided a quantitative clue to the time course of *Fusarium* infection on sucrose free MS plates, and suggested that the increased *Fusarium* susceptibility in *rdd* compared to Col-0 is less likely due to increased fungal accumulation but more likely due to a difference in physiological responses of the two plant lines. The result also showed that fungal biomass in Col-0 no longer increases after 3 DPI, suggesting that *Fusarium* established its infection by 3 DPI.

As a preliminary investigation of whether DNA demethylases may have a broad function in biotic stress response, I analysed the *rdd* mutant with virus infection. Interestingly, the *rdd* plants appeared to be more resistant than Col-0 to infections by both CMV and TuMV, belonging to two different viral families. Anti-viral defence in plants is known to occur through the post-transcriptional siRNA silencing pathway directed by 21/22 nt viral siRNA (Hu et al., 2011). In addition to 24 nt siRNAs, TEs in plants can also give rise to PTGS-associated 21 and 22 nt siRNAs, particularly in DNA methylation deficient mutant such as *ddm1* where TE transcription becomes more active (Creasey et al., 2014). It is possible that in the *rdd* mutant TEs are more repressed generating less RNA transcript for PTGS siRNA production, therefore allowing for PTGS factors such as DCL4 and RDR6 to become more available for processing viral siRNAs, resulting in enhanced virus resistance. However, this requires further investigation and it is also possible that other mechanisms are responsible for the enhanced virus resistance in *rdd*. Nevertheless, the viral infection experiment did suggest that DNA demethylases play a role in plant response

to other biotic stresses. Further studies using bacterial and other fungal pathogens will help establish a broad role of DNA demethylases in biotic stress response.

Previously published studies involving *Fusarium* infection are in general based on soil inoculation using root dipping (Diener, 2012, Hall et al., 2013, Czymmek et al., 2007). Most of my experiments involved root dipping inoculation on sucrose-free MS plate because of the uniform and consistent symptom development. To avoid a non-specific effect of root injury on subsequent sRNA, mRNA and DNA methylation analyses, I attempted to develop a spray-based inoculation method. Unfortunately, this method required the plants to be placed on sucrose free MS plates for an extended period to allow root establishment prior to *Fusarium* inoculation. Consequently, the mock-treated plants, to be used as controls for the molecular analyses, also developed abnormal phenotypes presumably due to lack of sucrose. Therefore, although spray inoculation did induce differential *Fusarium* symptoms on Col-0 and *rdd*, the method was abandoned and all future experiments were based on root-dipping inoculation.

In this chapter I also carried out preliminary assessment of sRNA responses to *Fusarium* infection in Col-0 and *rdd*. Importantly, these small scale sRNA northern blot analyses indicated that i) the accumulation of RdDM associated 24 nt siRNAs, but not the PTGS-associated miRNA and tasiRNA, are responsive to *Fusarium* infection, and ii) the sRNA response is stronger in *rdd* than Col-0. This result therefore supported an interaction between RdDM and DNA demethylases, and a role of this interaction in *Fusarium* response. Furthermore, through this sRNA northern blot analysis I established a time course of 1, 3 and 6 DPI for further molecular analysis. Interestingly, a clear sRNA response was usually observed from 3 DPI following *Fusarium* infection, which coincides with the result of fungal biomass analysis showing a high fungal level at 3 DPI.

3.3 Conclusions

Under normal growth conditions, no phenotypic differences were observed among the triple DNA demethylase mutant, *rdd*, the single DNA demethylase mutants (*ros1*, *dml2* and *dml3*), and wild type Col-0. However, all DNA demethylases showed increased susceptibility to *Fusarium oxysporum* infection compared to Col-0 in soil, although the *rdd* triple mutant had the strongest response, suggesting functional redundancy of the individual DNA demethylase genes. The *rdd* plants appeared to be more resistant to CMV and TuMV viral infection suggesting a broad function of DNA demethylases in biotic stress response.

Increased *Fusarium* susceptibility could be accounted for by two possibilities: increased fungal colonisation or enhanced disease sensitivity. Quantification of fungal biomass using the chitin method detected only subtle increases in *Fusarium* titre in *rdd* compared to Col-0, suggesting that the latter mechanism, namely increased disease sensitivity, accounts for the observed disease susceptibility in *rdd*.

RdDM associated siRNAs, but not PTGS-associated sRNAs, were found to be responsive to *Fusarium* infection. Furthermore, the 24 nt siRNAs are differentially accumulated between Col-0 and *rdd* in response to *Fusarium* infection, suggesting RdDM and DNA demethylase interact in *Fusarium* disease resistance. Phenotypic, fungal biomass and sRNA analyses point to the suitability of 1, 3 and 6 DPI time points for subsequent molecular analyses to be discussed in following three chapters.

Chapter 4: Genome-wide Survey of Small RNA Accumulation in Col-0 and *rdd* upon *Fusarium oxysporum* Infection

4.0 Introduction

A major focus of this thesis is to understand the role of RdDM and its interaction with DNA demethylases in *Fusarium* resistance through the analysis of RdDM associated siRNAs. As described in Chapter 3, preliminary sRNA analysis using northern blot hybridization indicated that the accumulation of RdDM associated siRNAs is affected by *Fusarium* infection, particularly in the *rdd* mutant. This implies that RdDM and DNA demethylases interact with each other in response to fungal infection, presumably by regulating stress response genes. To further understand the role of RdDM and DNA demethylation in plant defence, a genome-wide sRNA analysis in Col-0 and *rdd* with or without *Fusarium* infection was carried out using Illumina deep sequencing technology, which will be the focus of this chapter

To identify differentially abundant (DA) sRNAs from the sRNA sequencing data, I employed two well tested methods, DESeq and EdgeR (Anders et al., 2013, Reeb and Steibel, 2013). Both rely on a negative binomial distribution to detect DA sRNA but the program differences lie in their normalization features and the methods to estimate dispersion. I found DESeq to be the preferred program for measuring differential abundance of the sRNA sequencing reads due to the large number of DA sRNA that were obtained for this study. The DA sRNA determined bioinformatically usually requires experimental validation. This was performed by northern blot hybridization of DA miRNAs.

Recent interest has surrounded repeats and TEs in promoter regions as targets of RdDM which upon pathogen infection may experience transcriptional changes. DNA hypomethylation, a decrease in DNA methylation, was seen in response to bacterial infection at these sites (Yu et al., 2013). Stress responsive gene promoters have also been suggested to be regulated by DNA demethylases targeting TEs in *Arabidopsis* infected with *Fusarium oxysporum* (Le et al., 2014). Therefore, DNA demethylation is involved in plant-induced immune responses potentially activating defence genes through release of the TE and/or repeat silencing within their promoters to enable transcription. This chapter will provide a comprehensive analysis on the genomic location of DA sRNA to ascertain if the DNA demethylases or *Fusarium* infection affects siRNA accumulation in genes, 3 kb flanking regions and/or TE rich loci.

Chapter 4 will address on a genome-wide scale the sRNA changes in Columbia and *rdd* upon Fusarium infection. The main aims are to answer the following questions:

1. Is sRNA accumulation affected in the *rdd* triple DNA demethylation mutant under normal growth conditions?
2. Is sRNA accumulation affected by Fusarium infection?
3. What are the genomic targets of differentially abundant sRNA?
4. Are non-RdDM associated sRNAs, namely miRNAs, differentially accumulated in *rdd* or upon Fusarium infection?

4.1 Results

4.1.1 Quality Assessment of sRNA Sequencing Data

High-throughput sRNA sequencing was performed by BGI on mock-treated and *Fusarium*-infected Col-0 and *rdd* at 1, 3 and 6 days' post inoculation (DPI), with each treatment having two biological replicates. The 1 DPI time point represents *Fusarium* in a biotrophic infection phase, 3 DPI in a transient phase and 6 DPI taking on a necrotrophic lifestyle (Lyons et al., 2015). The average number of sRNA reads for each treatment is summarised in Table 4.1, while the number of sRNA reads for each of the 24 sequenced samples are shown in Supplementary Table 4.1. On average, approximately 14 million reads were obtained for each sample after the removal of adaptors and low quality reads.

Reads were filtered to keep only the 18-30 nucleotide (nt) sequences. Size distributions of the sRNA reads are shown in Figure 4.1. Approximately 50 % of all small RNA are 23-24 nt in size followed by approximately 40 % of sRNA 20-21 nt in size. For all samples, the 23-24 nt size fraction is the most dominant, which is also consistent with previous reports, but there appears to be a reduction in the representation of this size fraction in *rdd*, particularly the *Fusarium*-infected *rdd*, in comparison to Col-0.

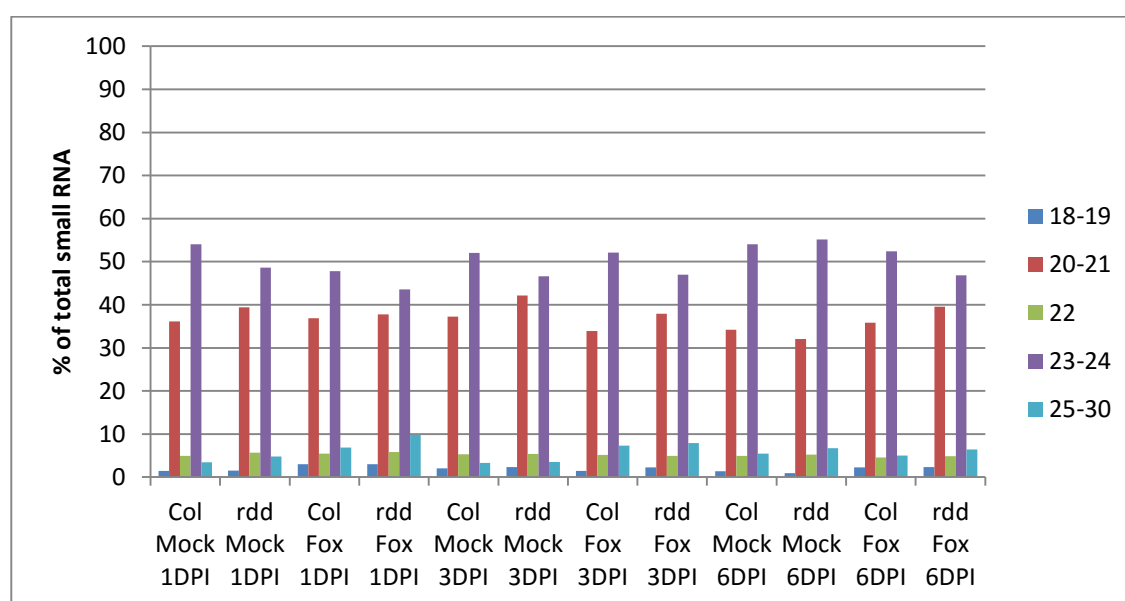
A unique sRNA refers to a specific sRNA sequence in the sequencing data. Approximately 75% of the unique sRNA is of the 23-24 nt size fraction (Supplementary Figure 4.1). Among the 4,000,000 to 6,000,000 total 23-24 nt sRNA reads in each sample (Supplementary Figure 4.2A), there were approximately 1,500,000 unique sRNA (Supplementary Figure 4.2B). The rest of the 18-30 nt sRNA reads contain roughly 200,000 unique sRNA in each sample. The small number of 20-21 nt unique sRNA is due to the relatively high number of reads for some of the miRNA species. The overall frequencies of sRNA size classes are consistent with published data (Ha et al., 2009, Kasschau et al., 2007).

sRNA reads of 18-30 nt in size were then mapped to the *Arabidopsis thaliana* genome. Stringent mapping criteria were employed to remove sRNA reads that aligned to more than ten loci within the genome (Table 4.1). This resulted in 82-83 % of the trimmed and filtered reads mapping to the genome. These sRNA reads were used in subsequent analyses of differential sRNA accumulation. However, when reads that aligned to up to 1000 genomic loci were included, the proportion of mapped reads could reach over 97% (Supplementary Table 4.2), indicating the *Arabidopsis* origin of these sRNA reads. These results confirmed the high quality of the small

Table 4.1: Average number of small RNA sequencing reads from plants of each treatment

	Col Mock	<i>rdd</i> Mock	Col Fusarium	<i>rdd</i> Fusarium
	1DPI	1DPI	1DPI	1DPI
Total reads	13,509,603	13,581,654	13,295,958	13,338,938
Reads after filtering (18-30nt)	13,342,379	13,418,973	13,088,135	13,111,732
Reads aligning in up to 10 different loci against <i>A. thaliana</i>	11,084,821	11,066,140	10,806,006	10,620,523
	3DPI	3DPI	3DPI	3DPI
Total reads	13,095,726	14,771,043	14,987,986	14,443,558
Reads after filtering (18-30nt)	12,893,320	14,514,839	14,725,236	14,197,310
Reads aligning in up to 10 different loci against <i>A. thaliana</i>	10,678,354	11,912,016	11,928,071	11,275,359
	6DPI	6DPI	6DPI	6DPI
Total reads	15,505,589	13,892,546	13,991,886	13,954,852
Reads after filtering (18-30nt)	15,220,243	13,732,543	13,730,433	13,614,541
Reads aligning in up to 10 different loci against <i>A. thaliana</i>	12,523,104	11,119,952	10,973,786	10,366,960

Figure 4.1: Small RNA size distribution as a percentage of total small RNA



RNA sequencing data. A small percentage of reads mapped to the *Fusarium oxysporum* genome in *Fusarium* inoculated samples, which increased at the later time points of 3 and 6 DPI, but only represented less than 0.5 % of all trimmed and filtered reads. These reads were excluded from further analysis in *Arabidopsis thaliana* but confirm the presence of *Fusarium* in infected plants.

sRNAs loaded to different AGO proteins in *Arabidopsis* have specific 5' terminal nucleotide preferences (Mi et al., 2008). For instance, AGO4-bound 24 nt sRNAs, involved in RdDM, show a strong enrichment for 5' terminal adenosine, whereas AGO1-bound sRNAs, involved in PTGS, have a preference for 5' terminal uridine. As shown in Supplementary Figure 4.3A, adenine and thymine (indicating uridine) are the most prevalent 5' terminal nucleotides in the total population of the trimmed and filtered sRNA reads. The 5' adenosine become more dominant when unique sRNA reads were analysed (Supplementary Figure 4.3B), which complements Supplementary Figure 4.2B highlighting 24 nt sRNAs making up the majority of unique sRNA. Together these results confirmed the high-quality of the sRNA sequencing data.

4.1.2 Identification of Differentially Abundant sRNA using DESeq

The main aim of this chapter is to investigate if sRNA have differential accumulation in *rdd* and upon *Fusarium* infection. This involved the identification of differentially abundant (DA) sRNA between Co-0 and *rdd* and between mock-treated and *Fusarium*-infected samples. Two well tested methods (Anders et al., 2013, Reeb and Steibel, 2013) of negative binomial distribution exist to detect differential abundance between sRNA count data. These programs are called DESeq and edge R with differences in their normalization features and in the methods used to estimate dispersion. Edge R uses the trimmed mean values while DESeq creates a virtual library that every sample is compared against using a relative log expression approach. Edge R uses a dispersion-mean relationship whereby a moderate feature-level dispersion estimates toward a mean trend. DESeq in contrast uses the maximum of individual dispersion estimates and the dispersion-mean trend.

In order to find an optimal method for identifying DA sRNA, I compared edge R and DESeq. Similar trends were observed in the total DA sRNA identified using edge R (Figure 4.2A-C) and DESeq (Figure 4.2D-F). For example, relatively small numbers of DA sRNA are found in the mock vs *Fusarium* comparisons as opposed to the relatively large numbers seen between Col-0 and *rdd*. There are both up and downregulated sRNAs, but the amount of DA sRNAs that are upregulated in *rdd* or upon *Fusarium* infection (indicated by the negative bars) is larger than that of downregulated DA sRNAs (Figure 4.2). The number of DA sRNA across the three time

points is relatively consistent except for the *rdd* mock vs *rdd* Fusarium comparison, where the 6 DPI comparison has a large increase in DA sRNA reads compared to the other two DPI time points. Furthermore, there are more upregulated than downregulated DA sRNAs in Fusarium-infected *rdd* plants, particularly at 6 DPI, suggesting that Fusarium infection increases sRNA accumulation in *rdd*. This appeared to be consistent with the northern blot result shown in Chapter 3 showing increased accumulation of the RdDM associated sRNAs in *rdd* by Fusarium infection. As a whole the DESeq method resulted in larger numbers of DA sRNAs than the Edge R method, particularly in the Col-0 vs *rdd* comparison. Consequently, DA sRNA called by the DESeq program were used for subsequent small RNA analysis.

A 100 bp sliding window with a 50 bp overlap was used to calculate the number of small RNA. The *A. thaliana* genome was broken down to encompass 2,393,345 windows, 100 bp in size. In total 135,199 windows were found to contain small RNAs which were differentially abundant (significant P value ≤ 0.05) and had a $\log_2$ fold change greater than 1 or lower than -1 between Col-0 and *rdd* under either Fusarium or mock conditions.

A large number of significant, differentially abundant small RNA was found at each time point across the three comparisons (Figure 4.3). In each comparison the 6 DPI time point had the largest number of DA sRNA. Overall the Col-0 vs *rdd* comparison had the largest number of DA sRNA, approximately 2-9 times more DA sRNA than the mock vs Fusarium comparisons of both the Col-0 and *rdd* backgrounds. This indicated that the mutation in the DNA demethylases causes more changes in sRNA accumulation than Fusarium infection.

DA sRNAs from the three Col-0 mock vs *rdd* mock comparisons showed a large overlap, with greater than 50 % of the DA sRNA at each individual time point shared between two or all three time points (Supplementary Table 4.3). This result indicated that most of the DA sRNAs are not affected by developmental timing of Arabidopsis and are specific to the genetic difference between Col-0 and *rdd*. It also indicated that the sRNA sequencing data from the different time points and samples have good reproducibility. The mock vs Fusarium comparisons showed fewer DA sRNAs than the Col-0 vs *rdd* mock comparisons, indicating that Fusarium infection caused less sRNA changes than the mutation of DNA demethylases. However, the *rdd* mock vs *rdd* Fusarium comparison generally gave more DA sRNAs than the Col-0 mock vs Col-0 Fusarium comparisons, particularly at 6 DPI. This suggested that the DNA demethylases play a role in sRNA accumulation both under normal growth conditions and in response to Fusarium infection. The three time points following Fusarium infection showed relatively small overlap in DA sRNAs, although the 3 DPI and 6 DPI shared more DA sRNAs than the 1 DPI and 3 DPI. It was presumed the same would be true for the 1 DPI and 6 DPI overlap however this comparison was not done due to time restrictions.

Figure 4.2: DA small RNA across 1, 3 and 6 DPI using either the DESeq or EdgeR programs

Positive bars show an increase in reads in mock treatment or Col-0, and negative bars show an increase in reads in Fusarium treatment or *rdd*. DA sRNA were determined based on $\log_2$ fold change ≥ 1 or ≤ -1 and P value ≤ 0.05 .

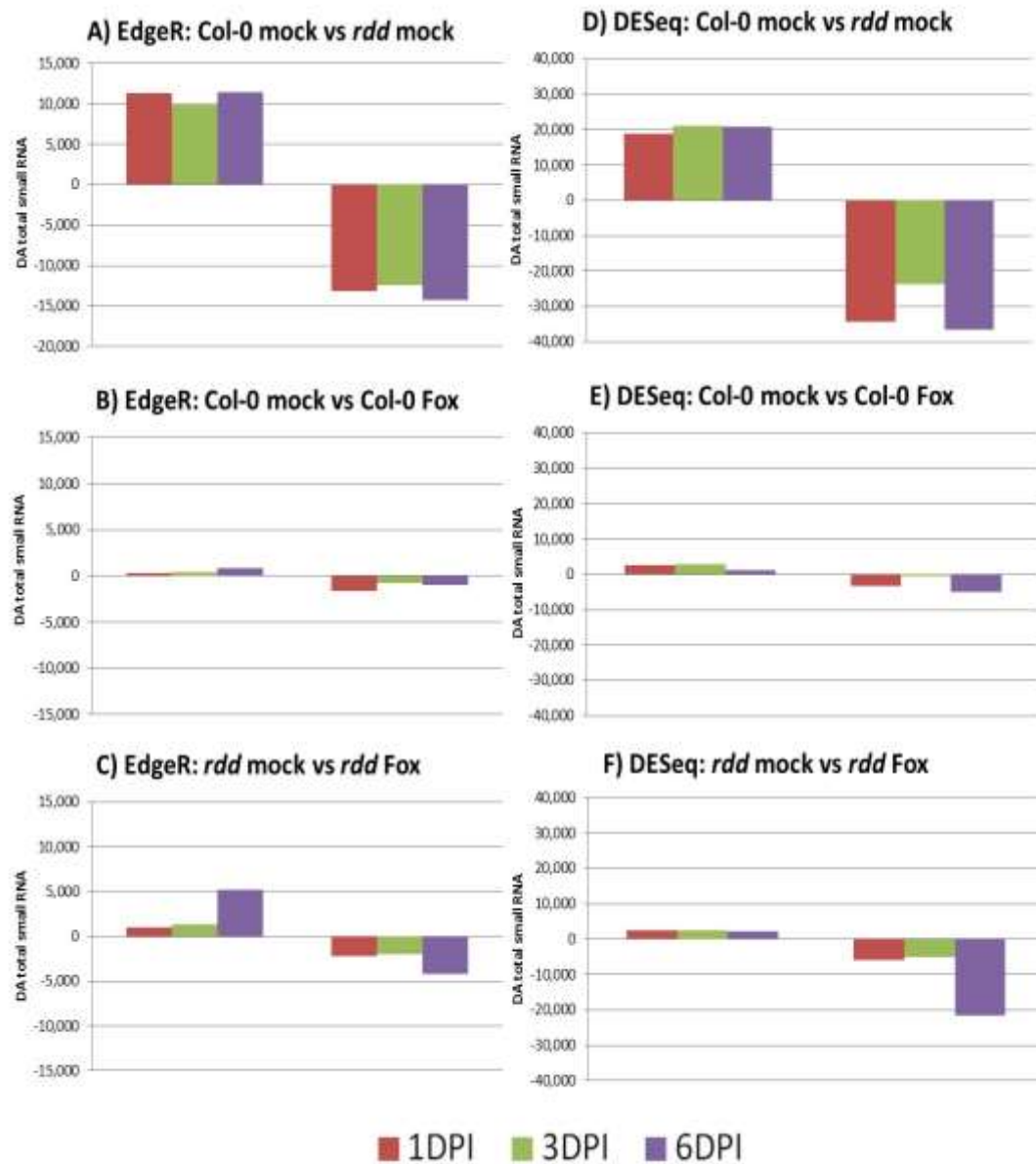
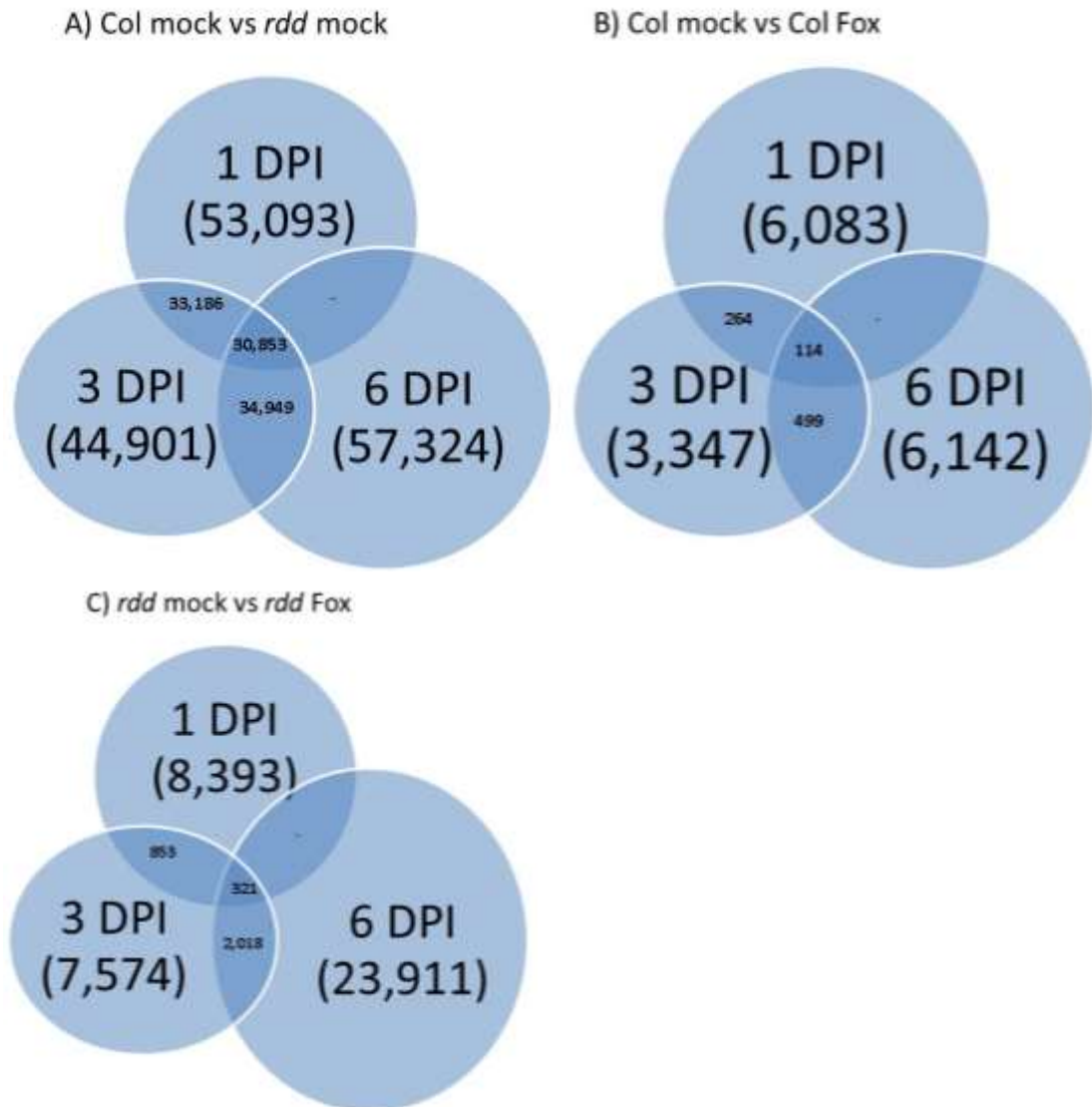


Figure 4.3: Overlap of DA sRNAs among different time points in different comparisons

DA sRNA were determined based on $\log_2$ fold change ≥ 1 or ≤ -1 and P value ≤ 0.05 . Numbers in brackets represent the total DA sRNA found in the individual time point.



4.1.3 The Majority of DA sRNAs Are Located In Or Within 3 kb of Protein Coding Genes

4.1.3.1 Distribution of DA sRNA across Arabidopsis Chromosomes

The DA sRNA distribution across the five Arabidopsis chromosomes, plus the chloroplast and mitochondrial genomes, was examined to determine if DA sRNA clustered preferentially in one area (Supplementary Figure 4.4). The amount of DA sRNA is similar across the five different chromosomes or among the different time points. The slight increase of DA sRNA from chromosome 1 in the Col mock vs *rdd* mock comparisons is likely due to this chromosome being the largest in Arabidopsis. There was almost no DA sRNA reads from the chloroplast or mitochondrial genome in the Col-0 vs *rdd* mock comparison, consistent with sRNAs derived from the nuclear genome. However, in the mock vs Fusarium comparisons there were some DA sRNAs matching the chloroplast and mitochondrial genome, suggesting that these sRNA reads might be due to RNA degradation caused by Fusarium infection.

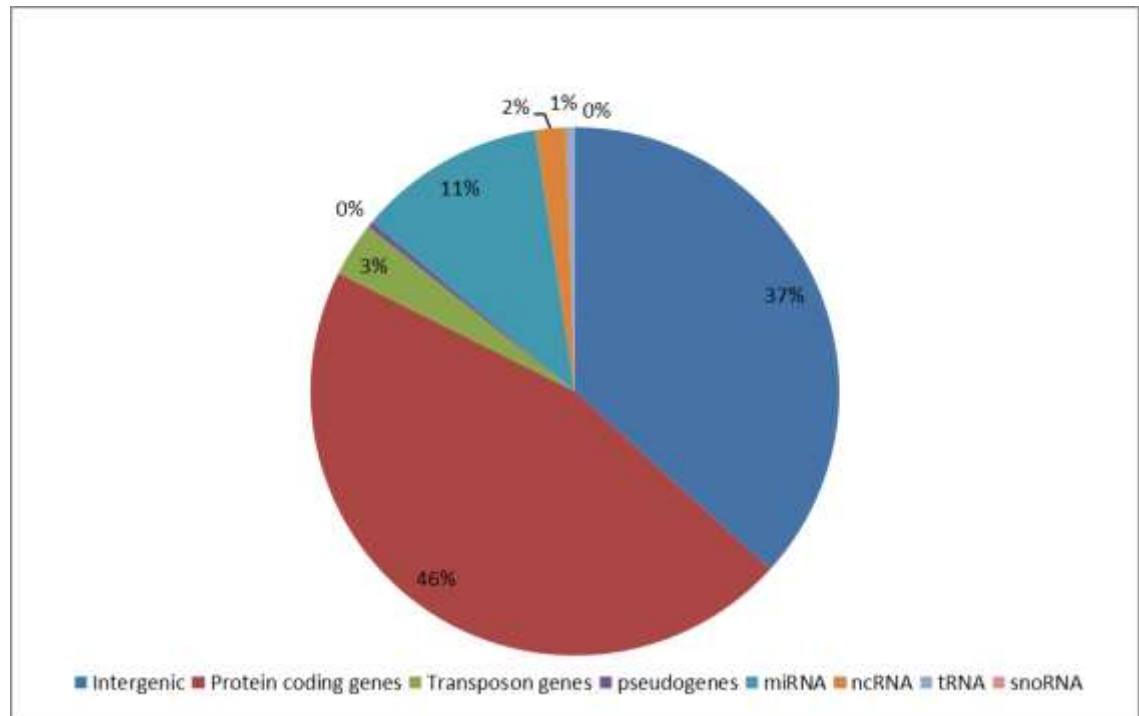
4.1.3.2 Distribution of DA sRNA across Arabidopsis Genomic Features

Total small RNAs from all samples, regardless of differential or non-differential abundance, were aligned against different regions in the Arabidopsis genome (Figure 4.4). “Intergenic” areas within the genome is the region upstream of a gene’s 5’ UTR and downstream of the 3’ UTR. “Protein coding genes” include the 5’ and 3’ UTR as well as 3kb upstream and downstream of annotated genes that may encompass intergenic areas. “Transposon genes” are grouped separately to protein-coding genes, which included the annotated transposon gene sequences but not the up or downstream regions. Transposable element and promoter annotations were excluded from Figure 4.4 as these regions may overlap with protein coding genes, transposon genes and ncRNA.

Transposable elements (TEs) and transposon genes are mobile genetic elements that drive genome evolution through genomic DNA rearrangements and mutations. They fall under two major classes; 1) Retrotransposons, that move via a copy and paste mechanism and 2) DNA transposons, that move via a cut and paste mechanism (Wheeler, 2013, Roy et al., 2015). This thesis refers to TEs as all sequences with homology to defined TEs, including the full length transposon genes that can encode proteins. Transposon gene refers to only the full length, protein coding TEs. The majority of sRNA from this dataset is located within protein coding and intergenic regions. Not surprisingly, 11 % of the total sRNA reads are miRNAs, reflecting the relatively high abundance of individual miRNA species.

Figure 4.4: Average distribution of total sRNA in the *A. thaliana* genome

The protein coding gene region includes 3 kb upstream and 3 kb downstream of annotated genes. Intergenic areas within the genome is the region upstream of a gene's 5' UTR and downstream of the 3' UTR. Transposon genes are grouped separately to protein-coding genes, which included the annotated transposon gene sequences but not the up or downstream regions.



Compared to the total sRNA, DA sRNAs are more enriched for protein coding genes (Figure 4.5). In addition, the proportion of DA sRNAs associated with protein coding genes is in general higher for the DA sRNAs down-regulated in *rdd* or Fusarium-infected plants than those up-regulated in *rdd* or upon Fusarium infection, although the total number of upregulated DA sRNA is in general larger. Another interesting aspect is the DA sRNAs associated with transposon genes, which are strongly enriched in the up-regulated DA sRNAs in *rdd* or upon Fusarium infection, particularly at the 1 DPI time point. This result suggests that transposons are a major target of DNA demethylases and stress-induced epigenetic changes. It also suggests that stress-induced changes may occur very early following stress treatment, in this case Fusarium infection. The distribution of DA sRNA showed differences across the three time points following Fusarium infection (Figure 4.5). For instance, the 6 DPI *rdd* mock and *rdd* Fusarium comparison gave a much higher number of DA sRNAs than the 1 or 3 DPI comparisons (Figure 4.5C), with the majority of the DA sRNAs being upregulated in Fusarium-infected plants, which again was consistent with the northern blot results in Chapter 3.

Figure 4.5: Total DA sRNA distribution across genomic features

The protein coding gene region includes 3 kb upstream and 3 kb downstream of annotated genes.

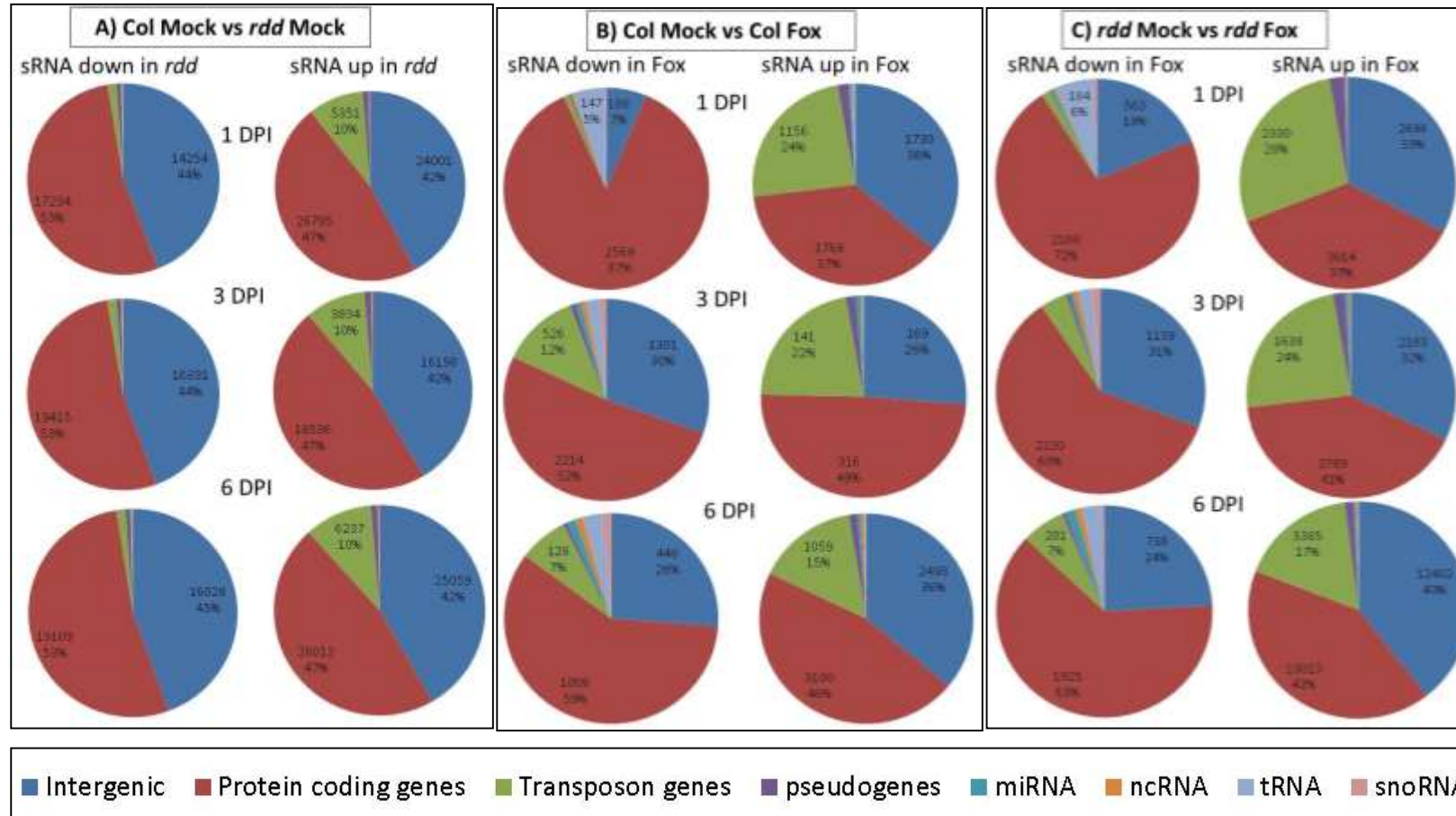


Figure 4.6: Distribution of protein coding gene associated DA sRNA reads across gene features

5' upstream and 3' downstream extend 3kb from the 5'UTR and 3' UTR respectively. 2 fold up or 2 fold down means that DA sRNA reads are ≥ 2 fold up or down-regulation in the respective comparisons. For instance, 2 fold up in Col-0 mock vs *rdd* mock comparison means ≥ 2 fold downregulation in *rdd* compared to Col-0.

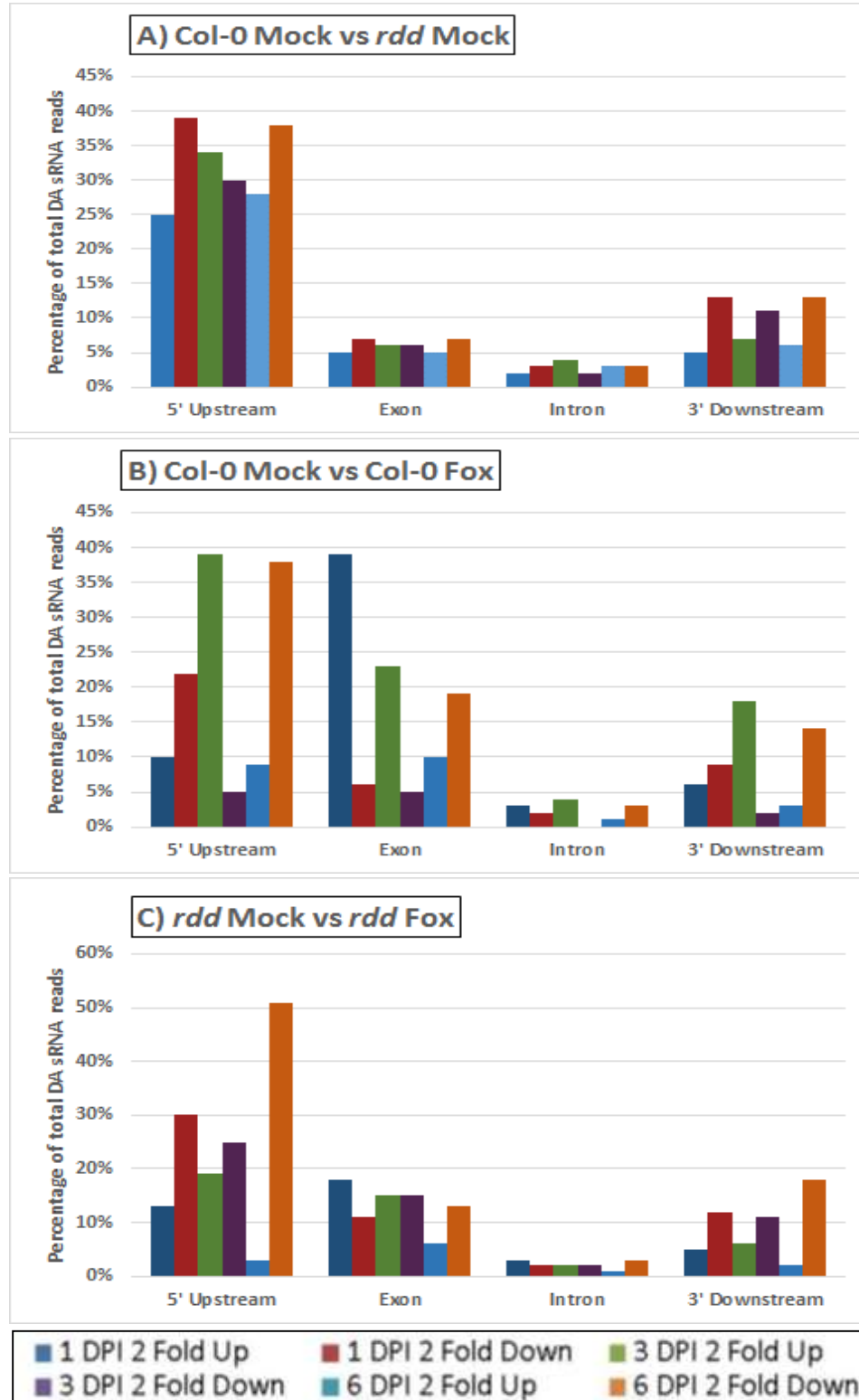
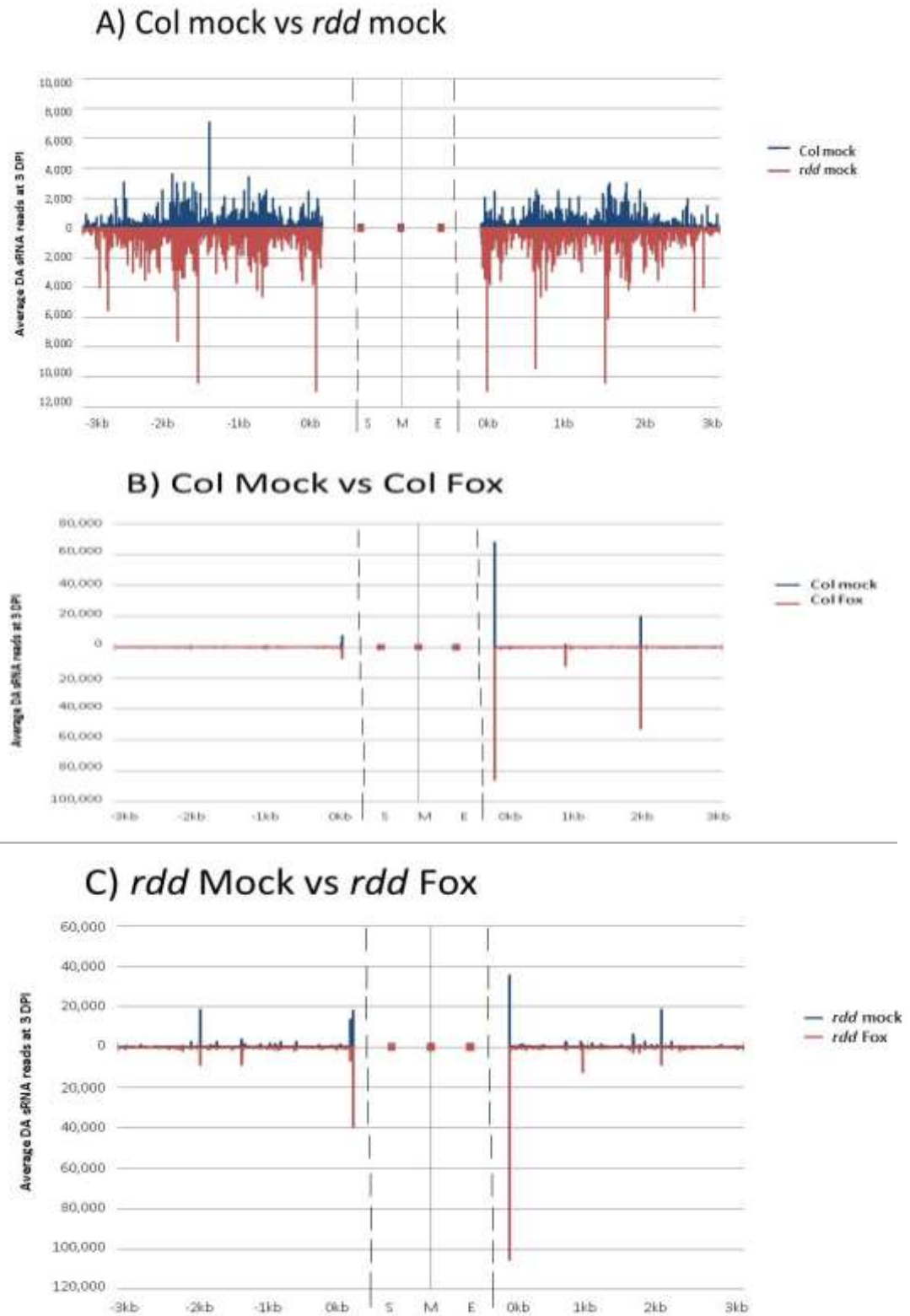


Figure 4.7: The physical distribution of DA sRNA at 3 DPI upstream, downstream and within genes

Negative kb represents distance upstream from the gene while positive kb represents distance downstream from the gene. S stands for start of the gene; M stands for middle of the gene, while E stands for end of the gene. DA sRNA were determined based on $\log_2$ fold change ≥ 1 or ≤ -1 and P value ≤ 0.05 . This is a true representation of the physical distribution of DA sRNA (ie no data was removed).



4.1.3.3 Distribution of Protein Coding Gene Associated DA sRNA along Gene Features

Protein coding gene associated DA sRNAs in the Col-0 mock vs *rdd* mock comparison are located primarily in the 5' upstream followed by the 3' downstream region (Figure 4.6A). There were also DA sRNAs from the gene body, with exon DA sRNAs more representative than intron DA sRNAs. This result is consistent with gene flanking regions the main target of epigenetic gene regulation.

In the mock vs Fusarium comparisons, DA sRNAs within exon regions was more prominent than the Col-0 vs *rdd* mock comparisons (Figure 4.6C and 4.6D). The 5' upstream and 3' downstream regions are still enriched for DA sRNAs across the three comparisons. In the mock vs Fusarium comparisons, an increase in the exons was observed that was not seen in the Col-0 vs *rdd* comparison. This suggests that the increase in exon DA RNA is not due to increased RNA degradation. Consistent with the distribution of total DA sRNA, the majority of DA sRNAs in the 6 DPI mock vs Fusarium comparisons showed increased expression upon Fusarium infection, particularly in the *rdd* background (Figure 4.6B and 4.6C).

To better examine the distribution of DA sRNAs along protein coding genes, I divided the gene features into 3 kb upstream, 3 kb downstream, and three equal segments (start, middle and end) of the gene body, and presented the DA sRNA reads along these features (Figure 4.7). The DA sRNAs at 3 DPI were chosen as this was the time point that would be further analysed against the transcriptome (Chapter 5) and bisulfite (Chapter 6) sequencing data, but an analysis of DA sRNA at 1, 3 and 6 DPI showed a similar trend (data not shown). sRNA read count between two replicates was averaged and presented among the upstream, downstream and three gene body segments. Consistent with Figure 4.6, DA sRNA reads preferentially locate upstream or downstream of the gene in all three comparisons. This result is also consistent with previous findings in maize (Gent et al., 2013) and Arabidopsis (Lister et al., 2008). The Col-0 vs *rdd* comparison showed a relatively even distribution of DA sRNAs along the up and downstream regions, but there is a dip in DA sRNA abundance around 500 bp in both regions (Figure 4.7A). There also appeared to be a dip at around 1.5 and 2.5 kb from the gene ends. There are several sharp DA sRNA peaks in the Col-0 vs *rdd* mock comparison, particularly in *rdd*, including ones located near the 5' and 3' gene ends. There are fewer DA sRNA reads overall in the mock vs Fusarium comparisons (Figure 4.7B and 4.7C). Interestingly, there are a number of sharp DA sRNA peaks, the strongest of which overlap with the gene ends particularly the 3' gene end. This result raises the possibility that these gene end DA sRNAs are important for the expression of genes following Fusarium infection.

4.1.4 The Majority of TEs and Transposon Genes are not Targeted by DA sRNA

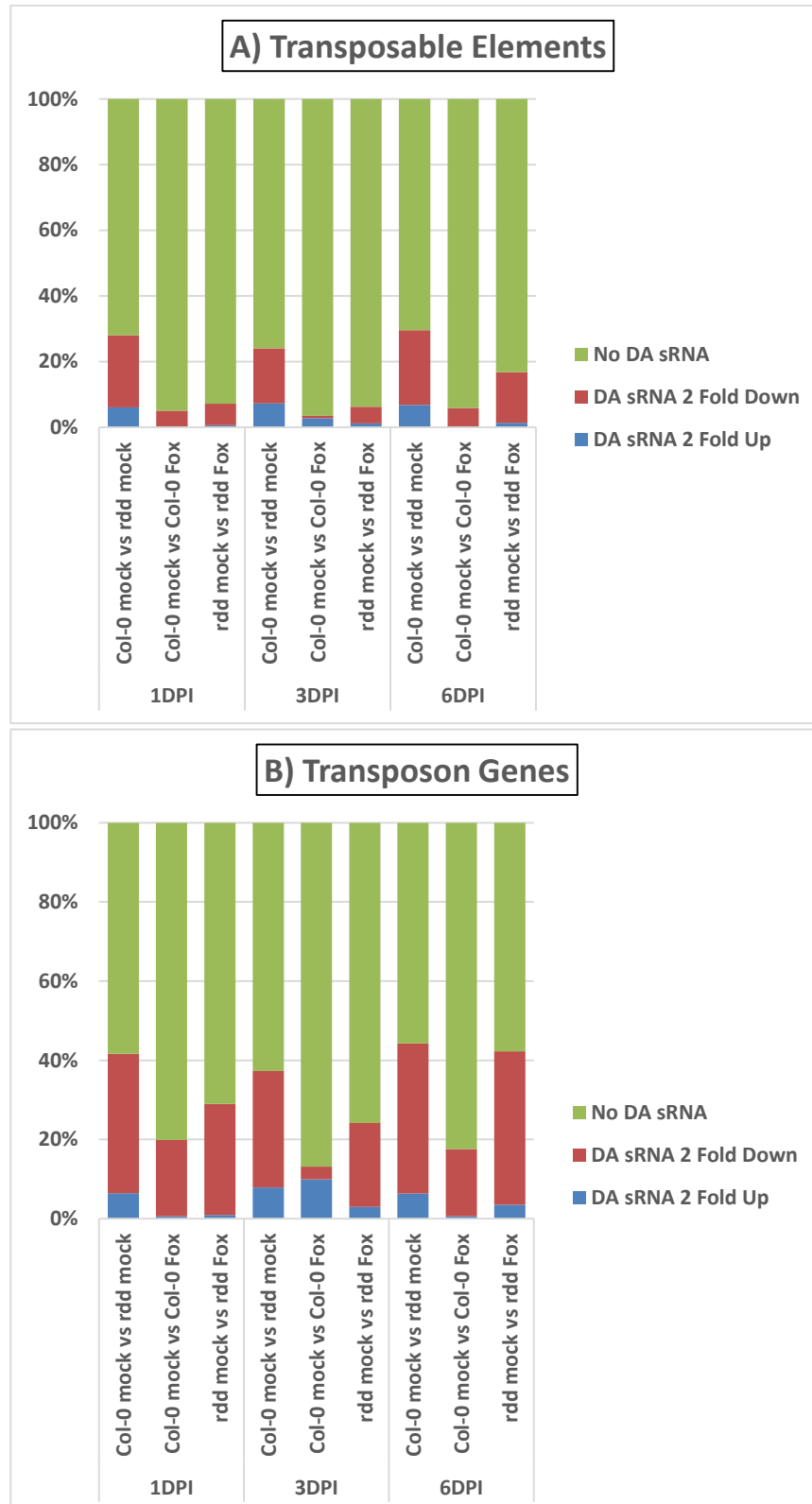
TEs are the main target of RdDM and the main source of 24-nt siRNAs in plants. As explained previously this thesis refers to TEs as all sequences with homology to defined TEs, including the annotated full length transposon genes that can encode proteins. A previous report from our laboratory suggested that TE sequences near genes are also a major target for DNA demethylases. Figure 4.5 shows that transposon genes are an important target of DA sRNAs, particularly upon *Fusarium* infection. Figure 4.8 identifies the percentage of unique TEs or only transposon genes for which DA sRNA was present. The majority of TEs and transposon genes are not targeted by DA sRNA. However, up to 20 % of the total TEs, and almost 40 % of transposon genes, show upregulated DA sRNA in *rdd* compared to Col-0. This is in contrast to the downregulated DA sRNA which occurs in even fewer TEs or transposon genes (Figure 4.8A and 4.8B). This indicates that DA sRNAs associated with TEs are mostly upregulated in *rdd* compared to Col-0. The DA sRNA distribution in transposon genes (Figure 4.8B) complements the total DA sRNA distribution across the genomic features shown in Figure 4.5.

TEs fall under two major classes; 1) retrotransposons, that move via a copy and paste mechanism and 2) DNA transposons, that move via a cut and paste mechanism (Wheeler, 2013). Supplementary Figure 4.5A shows in the Col-0 vs *rdd* comparison, DA sRNA occur most frequently near the RC/Helitron DNA transposon followed by the MuDR DNA transposon. The HAT DNA transposon, two LTR (Copia and Gypsy) and the LINE one retrotransposon superfamilies are also relatively enriched for DA sRNA in the Col-0 vs *rdd* comparison. For the Helitron TE, DA sRNA tends to be upregulated in *rdd* compared to Col-0. The DA sRNA associated with other TEs either show no difference between *rdd* and Col-0 or are slightly downregulated in *rdd* (Supplementary Figure 4.5A).

A different pattern of TE superfamily distribution for DA sRNA is seen in the mock vs *Fusarium* comparisons (Supplementary Figure 4.5B and Figure 4.5C). While the RC/Helitron TE superfamily is still among the highly represented ones, the Gypsy LTR retrotransposon superfamily becomes highly enriched, accounting for around 30 % of DA sRNAs as opposed to around 10 % in the Col-0 vs *rdd* mock comparison. The MuDR DNA transposons are also more enriched in the mock vs *Fusarium* comparisons than in the Col-0 vs *rdd* mock comparisons.

Figure 4.8: Distribution of DA sRNA targeting unique TE (A) and transposon genes (B)

TEs are all sequences with homology to defined TEs, including the annotated full length transposon genes that can encode proteins. Transposon genes only include the later. Unique refers to TEs only counted once regardless of the number of times it is found throughout the genome. The percentage refers to the number of unique transposable elements, out of 31189, (or transposons genes – out of 3903) in the Arabidopsis genome for which differentially abundant (DA) sRNA was present or absent.



4.1.5 Analysis of Differential miRNA Accumulation

My thesis focuses on RdDM associated sRNAs. The northern blot results in Chapter 3 suggested that 24 nt RdDM associated sRNAs, but not miRNAs or tasiRNAs, are affected in *rdd* or by Fusarium infection. However, miRNAs play important roles in plant development, and some miRNAs are also found to play a role in plant defence (Kim et al., 2010b, Li et al., 2010, Boccara et al., 2014). I therefore briefly analysed the sRNA sequencing data to examine if miRNA accumulation might be affected in the *rdd* mutant or by Fusarium infection. The result from the 3 DPI data is shown in Table 4.2.

A total of 61 annotated miRNAs, representing 35 miRNA families, were found DA at 3 DPI from the Col-0 vs *rdd* (28), Col-0 mock vs Col-0 Fusarium (22) and *rdd* mock vs *rdd* Fusarium (34) comparisons. Eleven miRNAs were found to be differentially abundant in both of the mock vs Fusarium comparisons. Three members of the miR395 family were represented in the DA sRNAs of all three comparisons. Interestingly, the expression of miR395 is known to be responsive to stress treatments such as sulfur deficiency (Gao et al., 2013).

4.1.6 Verification of sRNA Expression Pattern by Northern Blot Hybridisation of DA miRNA

To verify the DA sRNA results from the bioinformatics analysis, I performed northern blot hybridization against the DA miR164, chosen because of its relatively high abundance for easy detection and specific nucleotide sequence for easy probe design. miR164 is a senescence regulator and declines in abundance with age leading to an increase in the gene targets which promote leaf longevity (Rubio-Somoza and Weigel, 2011). The additional letter is given for miRNA with nearly identical sequences except for one or two nucleotides. In the sRNA sequencing data the three miRNA (miR164A, miR164B and miR164C) were treated as the one miRNA. miR164 was found to be differentially abundant at 3 DPI in Col-0 vs *rdd* and in *rdd* mock vs *rdd* Fusarium (Table 4.2; Figure 4.9B). This pattern was largely replicated in the sRNA northern blot (Figure 4.9A).

Table 4.2: Differentially abundant miRNA at 3 DPI

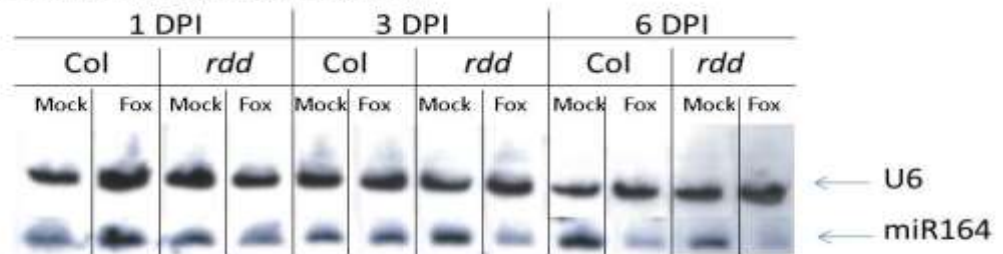
miRNA highlighted in bold are found also differentially abundant at the 1 and 6 DPI. Increasing positive Log₂ fold values are highlighted using orange (relatively low value) to red (high value) colours and represents a downregulation in *rdd* or Fusarium infected samples. Increasing negative Log₂ fold changes (LFC) are highlighted using yellow (relatively low value) to green (high value) and represents an upregulation in *rdd* of Fusarium infected samples. A white/blank cell means the miRNA was not differentially abundant.

	Col mock vs <i>rdd</i> mock (LFC)	Col mock vs Col Fox (LFC)	<i>rdd</i> mock vs <i>rdd</i> Fox (LFC)
MIR156G		-1.658	-1.318
MIR157D	-5.672		
MIR159C		-2.807	
MIR160B		2.205	1.014
MIR164A			1.097
MIR164B			1.015
MIR164C	-1.138		
MIR165A	-1.658		1.146
MIR165B			1.868
MIR167D			1.762
MIR169A			1.012
MIR169B	-1.212		1.142
MIR169D			1.247
MIR169E	-1.585		2.755
MIR169F		-1.489	
MIR169G	-1.465		2.021
MIR169H			1.919
MIR169I	-1.015		
MIR169L			1.773
MIR169M	-2.046		
MIR169N	-2.644		
MIR171B			1.675
MIR172C		-2.000	
MIR172D	-2.747		
MIR319A			1.277
MIR319B			-1.283
MIR390B			1.747
MIR394A		3.087	
MIR394B		2.939	
MIR395A			-2.202
MIR395B	-1.760	-3.165	-1.228
MIR395C	-1.455	-2.585	-1.585
MIR395D		-2.807	-2.874
MIR395E		-2.807	-2.248
MIR395F	-1.350	-2.415	-1.529
MIR396A	-2.280		
MIR397A		-1.962	-1.171
MIR397B	-3.322	-2.022	-1.598
MIR398A	-2.087		
MIR398B		-1.331	
MIR398C		-1.065	
MIR399A	-4.781		
MIR399B		-1.609	-1.807
MIR399F	-2.585		

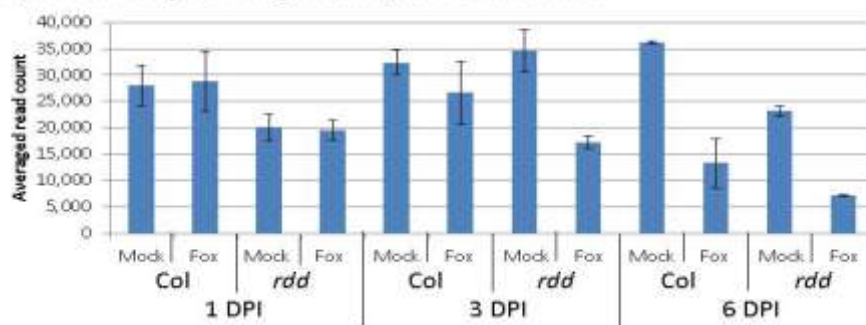
	Col mock vs <i>rdd</i> mock	Col mock vs Col Fox	<i>rdd</i> mock vs <i>rdd</i> Fox
MIR404	1.027		-1.489
MIR405D		1.379	
MIR406	1.188		
MIR408		-1.389	
MIR447A	-1.429	-1.387	
MIR822			1.181
MIR831			-1.773
MIR833A			2.000
MIR835	-1.127		
MIR839	-1.447	-1.246	
MIR841A	-1.235		
MIR843	7.325		
MIR850			1.348
MIR857		-1.524	-2.000
MIR858A	-9.691		
MIR860	-1.087		
MIR869	-1.264		1.699

Figure 4.9: Northern blot hybridisation of miR164 (A) and comparison with the sRNA sequencing read count (B)

A) sRNA Northern Blot hybridisation



B) sRNA sequencing averaged read count



4.2 Discussion

In Chapter 3 I used northern blot hybridization to show that RdDM associated sRNAs might be affected by mutation of DNA demethylases and responsive to *Fusarium* infection. I also determined in Chapter 3 that 1, 3 and 6 DPI time points were suitable for genome-wide analysis. Consequently, 24 samples, including two replicates each of mock-treated and *Fusarium*-infected Col-0 and *rrd* mutant plants were sent for sRNA deep sequencing by BGI, resulting in an average of 14 million sRNA reads for each sample. Preliminary analysis of the sRNA sequencing data confirmed the high quality of the data. Almost all of the sRNA reads can be mapped to the Arabidopsis genome with a small proportion in infected samples also mapping to the *Fusarium* genome. The reads showed the expected size distribution of plant sRNAs with 24 nt species being the most dominant followed by 21 nt ones. Differentially abundant sRNAs show high levels of overlap among three Col-0 vs *rrd* mock comparisons. Lastly very little chloroplast or mitochondrial genome sequences were detected.

In this chapter I focused on genome-wide identification of differentially abundant (DA) Arabidopsis sRNA in *rrd* and upon *Fusarium* infection using bioinformatics. DESeq and Edge R are the top performers for analysis of differential expression of counts. The recent papers by Anders et al. (2013) and Reeb and Steibel (2013) detail the differences in normalisation, sequencing quality analysis and dispersion estimation between the two programs. Neither method is singled out as being the best for analysing quantitative readouts (in the form of counts) from sRNA sequencing data. My preliminary analysis of DA sRNA showed that the DESeq method gave more DA sRNAs than the Edge R method for Col-0 vs *rrd* comparison, but similar numbers of DA sRNA between the two methods when comparing mock vs *Fusarium* treatment. The DESeq dataset was used for further analysis due to the larger number of DA sRNA which could be mapped against mRNA and DNA methylation features of interest in Chapter 5 and 6.

Through this bioinformatics analysis I identified a large number of sRNAs that were differentially abundant between Col-0 and *rrd*, most of which are 24 nt in size. This is a significant finding as it suggests that DNA demethylases play a role in sRNA accumulation and supports our hypothesis that DNA demethylases interact with RdDM in regulating gene expression (Le et al., 2014). Interestingly, sRNA levels tend to increase in the *rrd* mutant compared to the wild type Col-0. This suggests that increased DNA methylation may result in increased sRNA production, consistent with previous findings such as found in the maintenance MET1 methyltransferase mutant that results in reduction in siRNA accumulation (Lister et al., 2008). A significant number

of DA sRNAs, mostly 24 nt in size, were also identified between *Fusarium*-infected and mock-treated samples. This suggested that sRNAs, or RdDM, plays a role in plant response to *Fusarium* infection, by regulating the expression of some stress response genes, as suggested by previous studies discussed in Chapter 1. The *Fusarium*-responsive DA sRNAs showed different patterns between the Col-0 and the *rdd* backgrounds, suggesting that DNA demethylases also play a role in sRNA accumulation during stress response. Interestingly, there was a strong increase in downregulated *rdd* DA sRNA at 6 DPI following *Fusarium* infection, which is consistent with the northern blot result (Chapter 3).

In support of their role in gene regulation, a large proportion of the DA sRNAs in all three comparisons are located in the gene body and 3 kb up and downstream of genes. For the Col-0 vs *rdd* comparison, the vast majority of the gene associated DA sRNAs reside in the up and downstream flanking regions, particularly the upstream regions that form promoters of genes. This is consistent with RdDM targeting promoter sequences to regulate transcriptional activity of genes. However, exon sequences were also found to be frequently targeted by DA sRNAs, particularly in the mock vs *Fusarium* comparisons. How sRNA targeting the gene body might regulate gene expression, particularly under stress conditions, remains unknown and will be an interesting question to investigate in future studies. The DA sRNAs matching both the upstream and downstream sequences showed uneven distributions, particularly for the mock vs *Fusarium* comparisons. A clear peak of DA sRNAs occur near the transcription start site (5' end) and the 3' ends of genes, suggesting that sRNAs matching these gene ends are important for gene regulation. Consistent with this possibility studies in maize and *Arabidopsis* have shown a close association of 24 nt siRNAs and gene ends (Gent et al., 2013, Zhang et al., 2006).

TEs are the major target of RdDM in plants. A study of stress response genes downregulated in *rdd* showed that these genes are enriched for TE presence in the promoter sequences, suggesting that DNA demethylases target the promoter TEs to regulate the expression of these genes (Le et al., 2014). I therefore examined the association of the DA sRNAs with TEs and transposon genes, and found that a significant fraction of TEs are targeted by these DA sRNAs. A previous mRNA sequencing study in Col-0 and *rdd* found transcriptome reads associated with approximately 20 % of transposons and pseudogenes in *Arabidopsis* (Lister et al., 2008). A similar percentage of TEs was found to be associated with DA sRNA in this study, raising the possibility that DA sRNAs target primarily transcriptionally active TEs, but this has yet to be examined. These TE associated DA sRNAs tend to be upregulated in *rdd* or upon *Fusarium* infection following the same trend as the DA sRNA associated with protein coding genes. In fact,

some of the protein coding gene associated DA sRNAs could also be associated with TE sequences residing in the gene body and the 3 kb flanking regions.

A recent study on easiRNAs (epigenetically activated small interfering RNAs), 21 nt siRNAs derived from TEs in the *ddm1* mutant, showed that these siRNAs originated mostly from the *GYPSY* superfamily of retrotransposons (Creasey et al., 2014). Interestingly, the DA sRNAs detected in *Fusarium* infected plants were highly enriched for *GYPSY* TEs, suggesting that these TEs are the main target of dynamic sRNA mediated epigenetic regulation in *Arabidopsis*. It will be interesting to examine the function of these DA sRNAs in regulating the expression of stress response genes.

The northern blot hybridization in Chapter 3 suggested that 24 nt sRNAs, but not the PTGS-associated 21 nt sRNAs, are primarily affected in *rdd* or by *Fusarium* infection. Indeed, the DA sRNAs found in the sRNA sequencing data are primarily of the 24 nt size class. While focusing on these RdDM associated DA sRNAs in my analysis, I was interested in knowing if miRNAs, important for plant development or stress response, also undergo changes in *rdd* or upon *Fusarium* infection. 35 annotated miRNA families were found differentially abundant at 3 DPI in three comparisons. The most prominent are miR395B/C/F and miR397B, which were differentially abundant in *rdd* compared to Col-0 as well as in mock samples compared to *Fusarium* treated samples. miR395 targets the sulphate uptake and assimilation pathway and has been shown to be up-regulated by sulphate deficiency (Matthewman et al., 2012). Up-regulation of miR395 has also been shown to facilitate sulphate enhanced defence after infection with *Hibiscus* chlorotic ringspot virus (Gao et al., 2013). The results from this study suggest DNA demethylases are involved in sulphate uptake and that miR395 provides targeted up regulation of the sulphate pathway as a defence response against *Fusarium oxysporum*. Meanwhile miR397B has been shown to regulate lignin content and seed number in *Arabidopsis* (Wang et al., 2014a). This study (Wang et al., 2014a) found that overexpression of miR397B resulted in reduced cell wall thickness due to diminished lignin deposition. Therefore, *rdd* and plants infected with *Fusarium* may also have a thin cell wall, due to the upregulation of miR397B, which may lead to the increased susceptibility to fungal hyphae that penetrate the cell walls during infection. To illustrate the potential importance of the DA miRNA found, the mRNA expression of the associated gene targets is elaborated in Chapter 5. However, this thesis focuses on RdDM associated sRNAs and therefore miRNAs were not studied in details.

4.3 Conclusions

High quality whole genome small RNA sequencing data was obtained for mock treated and Fusarium infected Col-0 and *rdd* plants at 1, 3 and 6 DPI, which provided a foundation for identification of sRNAs affected by mutation of DNA demethylases and by Fusarium infection. Indeed, from the sRNA sequencing data I identified a large number of sRNAs that were differentially abundant between Col-0 and *rdd* and between mock treated and Fusarium infected plants. These DA sRNAs are located predominantly within protein coding genes and their flanking sequences. Transposon genes are also enriched for DA sRNAs, particularly in Fusarium-infected plants. These results suggest that i) DNA demethylases are involved in sRNA accumulation, ii) sRNA accumulation is responsive to Fusarium infection, and iii) protein coding genes and transposon genes are the major targets of DNA demethylases and sRNAs. Together these results support the view that DNA demethylases interact with RdDM to regulate stress response gene expression (Le et al., 2014). The questions now are if the DA sRNAs are associated with differential expression of the overlapping or nearby genes, and if DA sRNAs affect gene expression through changes in DNA methylation. These questions will be investigated in Chapters 5 and 6. The genetic targets of the DA sRNA and their expression are further discussed following the whole genome mRNA sequencing examined in Chapter 5.

Chapter 5: Identification of Differentially Expressed sRNA Target Genes Across the Genome

5.0 Introduction

In Chapter 4 I investigated the genome-wide changes of sRNA accumulation in *rdd* and upon Fusarium infection using sRNA deep sequencing. A large number of differentially accumulated (DA) sRNAs were identified from the following comparisons:

- Col-0 mock vs *rdd* mock
- Col-0 mock vs Col-0 Fusarium
- *rdd* mock vs *rdd* Fusarium

The next important question is if these DA sRNAs are associated with differential expression of neighbouring genes that might contribute to the disease phenotypes of *rdd* and Col-0. While some previous studies (Lister et al., 2008, Groszmann et al., 2011) have examined genome-wide sRNA accumulation and gene expression in plants, the relationship between the two remains poorly understood. The 24 nt size class of sRNAs, representing the dominant fraction of the DA sRNAs I identified, has the potential to regulate gene expression by directing RdDM. A recent study on the role of sRNAs in hybrid vigour in Arabidopsis showed that F1 hybrid plants have altered 24 nt siRNA accumulation and DNA methylation in the regions flanking genes compared to parental plants, which correlates with gene expression changes (Groszmann et al., 2011). But do changes in RdDM associated sRNAs in *rdd* or upon Fusarium infection alter gene expression levels?

The main aim of Chapter 5 is to investigate if DA sRNAs are associated with differential gene expression between Col-0 and *rdd*, and between mock-treated and Fusarium-infected plants. Furthermore, Chapter 5 endeavours to examine the potential functions of these differentially expressed (DE) genes. Both mRNA sequencing and bioinformatics were employed to identify DA sRNA target genes. In particular, Chapter 5 will focus on the following questions:

1. What are the potential genetic targets of DA sRNAs and
2. What are the biological functions of the DA sRNA affected genes?

5.1 Results

5.1.1 A large number of genes are potentially targeted by DA sRNAs

To identify genes associated with DA sRNAs, DA sRNAs identified in Chapter 4 were aligned against *Arabidopsis thaliana* TAIR9 gene annotations from the CSIRO Plant Industry Bioinformatics Genome Browser. The number of genes with DA sRNA (including the 3 kb upstream, or 3 kb downstream regions were listed separately in Table 5.1). The numbers of overlapping gene associated DA sRNA among the 1, 3 and 6 DPI time points are shown in Table 5.1B. The size of gene promoters can be anywhere from 500 to 3000 bp long (Baev et al., 2010). I included 3 kb flanking regions in this analysis to ensure that regulatory regions of promoters are well covered.

Consistent with the relatively high numbers of DA sRNAs, a large number of genes were found to be associated with DA sRNAs in the Col-0 vs *rdd* comparison (Table 5.1), suggesting that they might be regulated by sRNAs through interaction with DNA demethylases. Many DA sRNA associated genes were also found in the mock vs *Fusarium* comparisons, particularly in the 6 DPI *rdd* comparison. This result suggests a role of DA sRNAs in the plants response to *Fusarium* infection, and an involvement of DNA demethylases in this response process. Also consistent with the DA sRNA distribution, is the number of genes associated with flanking region DA sRNAs is much bigger than that with gene body DA sRNAs. This result was in line with flanking sequences being more important in controlling gene transcription than gene body regions. It thereby suggests that DA sRNAs regulate the expression of genes by targeting primarily the flanking sequences. However, a significant number of genes are associated with gene body DA sRNAs, suggesting that gene body DA sRNAs also play a regulatory role in gene expression.

Table 5.1: Number of genes associated with differentially abundant small RNA

All DA sRNA had a log₂ fold change (LFC) value of ≥1 or ≤-1 and P value of ≤0.05. The same DA sRNA and/or gene may be counted more than once across the gene body and 3 kb flanking regions.

A) Singular time points

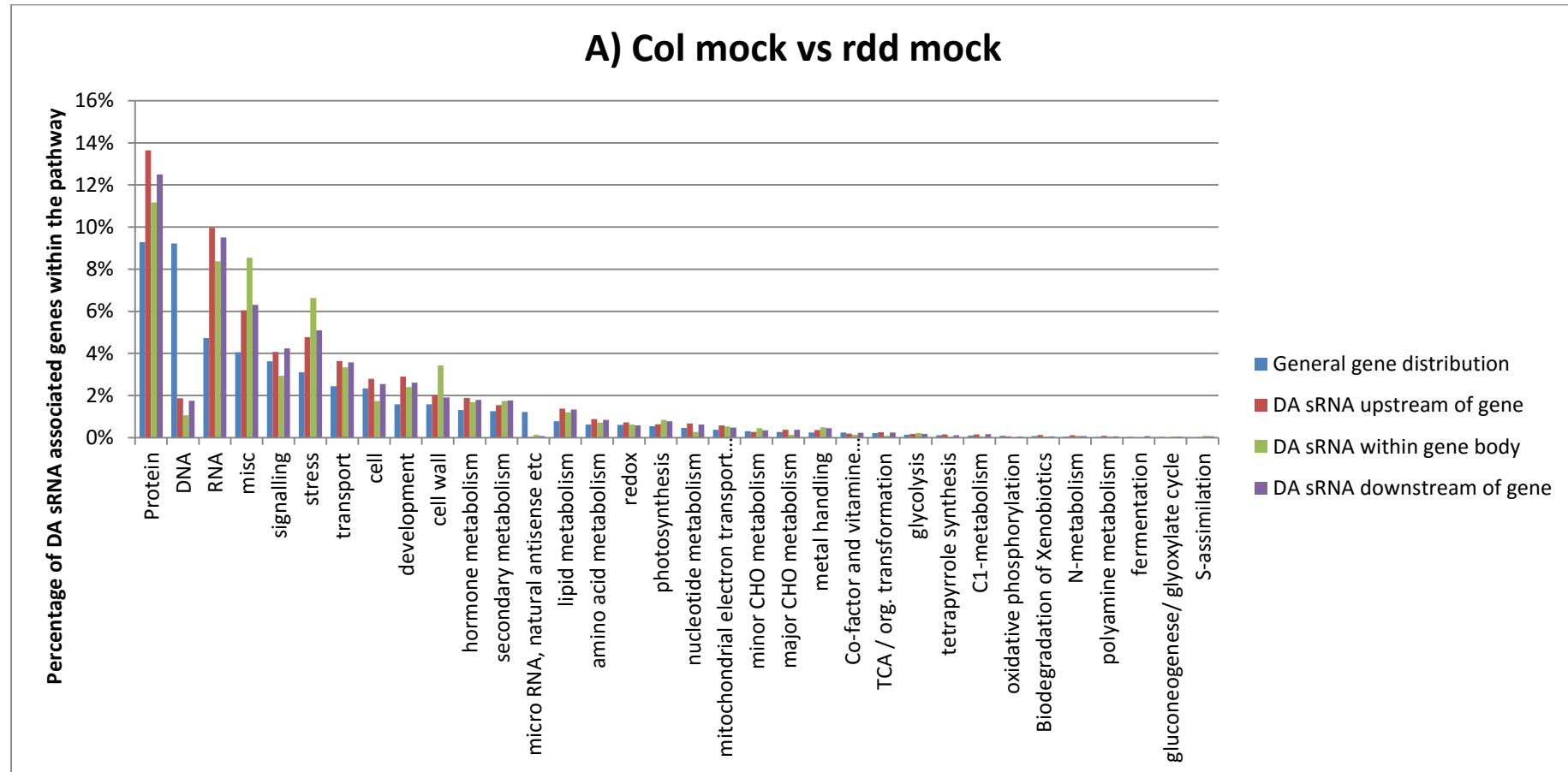
Comparison	Time point: DPI	Total DA sRNA	DA sRNA within the			DA sRNA gene targets within the		
			3kb upstream	Gene body	3kb downstream	3kb upstream	Gene body	3kb downstream
Col mock vs <i>rdd</i> mock	1	53,093	43,926	7,876	34,025	8,738	2,349	7,329
	3	44,901	38,283	7,002	29,380	8,323	2,240	6,880
	6	57,324	46,781	8,306	36,172	8,931	2,395	7,372
Col mock vs Col Fox	1	6,083	3,303	2,272	3,903	1,694	783	1,656
	3	3,347	2,099	933	2,231	1,394	491	1,384
	6	6,142	3,542	1,495	3,408	1,983	645	1,765
<i>rdd</i> mock vs <i>rdd</i> Fox	1	8,393	4,315	1,900	4,126	2,397	857	2,146
	3	7,574	4,478	1,937	4,264	2,413	869	2,181
	6	23,911	13,486	3,932	11,911	5,173	1,469	4,472

B) 1,3, and 6 DPI Overlapping time points

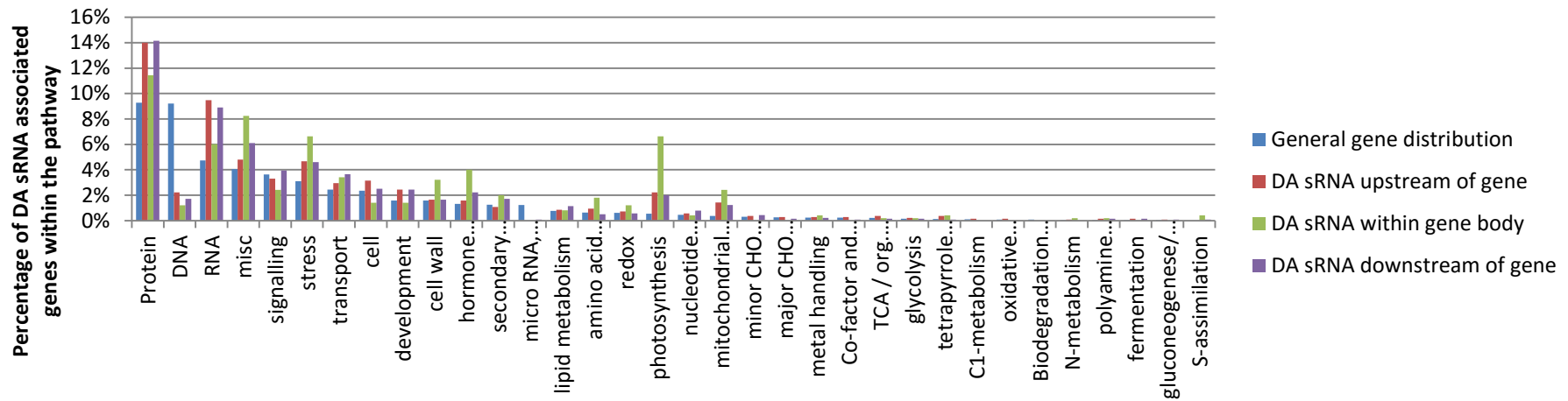
Comparison	Total DA sRNA	DA sRNA within the			DA sRNA gene targets within the		
		3kb upstream	Gene body	3kb downstream	3kb upstream	Gene body	3kb downstream
Col mock vs <i>rdd</i> mock	30,853	28,272	4,969	21,352	6,980	1,682	5,674
Col mock vs Col Fox	114	61	85	103	43	36	60
<i>rdd</i> mock vs <i>rdd</i> Fox	321	214	177	209	89	72	107

Figure 5.1: MapMan Gene Ontology analysis of DA sRNA associated genes at 3 DPI

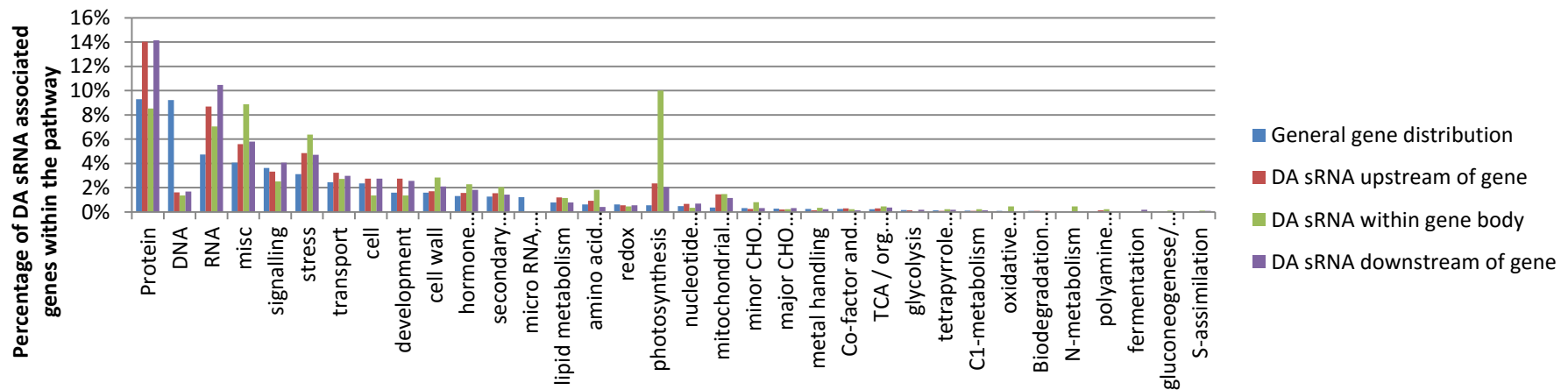
The histograms show the percentage of DA sRNA associated genes of a particular pathway in total DA sRNA associated genes. The percentage of all genes of a particular pathway in total Arabidopsis genes is indicated by the blue bar on the left for comparison. The list of genes that are part of each pathway is available from Thimm et al., 2004.



B) Col mock vs Col Fox



C) rdd mock vs rdd Fox



Further in depth analysis of the DA sRNA gene targets, including functional categorization was performed on the 3 DPI samples. I focused on this time point because i) Fusarium accumulation and sRNA accumulation differences become noticeable at this time point between Col-0 and *rdd*, and ii) subsequent mRNA sequencing and DNA methylation analysis were performed on only the 3 DPI samples due to limited funds for sequencing. Functional categorization of the DA sRNA associated genes at 3 DPI (Figure 5.1) was done using gene ontology (GO) analysis combined with MapMan (Thimm et al., 2004) to determine the biological pathways of genes. The distribution of the DA sRNA associated genes was compared to the overall pathway distribution of Arabidopsis genes to determine pathway enrichment of these genes. The list of genes that are part of each pathway is able to be downloaded on the MapMan website <http://mapman.gabipd.org/web/guest>. This list was used to generate the general gene distribution. Each category has subcategories which provide further information on the genes included in that pathway. For example, the protein category includes genes involved in protein synthesis and degradation.

For both Col-0 vs *rdd* and mock vs Fusarium comparisons, the DNA and miRNA pathways genes are depleted, whereas the protein, RNA, miscellaneous, and stress pathway genes are enriched, in all the three sets of genes associated with upstream, gene body and downstream DA sRNAs, respectively. The depletion of miRNA genes is consistent with very few miRNAs being differentially accumulated. For the two mock vs Fusarium comparisons, the photosynthesis and mitochondrial pathway genes become highly enriched, suggesting that these genes might be under sRNA regulation during stress response. Interestingly, genes associated with gene body DA sRNAs are unique in that they showed stronger enrichments for the relevant pathways than those associated with upstream and downstream DA sRNAs. This is particularly prominent for the photosynthesis pathway, where the genes with gene body DA sRNAs are 3-5 times more enriched than genes with flanking region DA sRNAs. In addition, for the cell wall pathway, only genes associated with gene body DA sRNAs showed clear enrichment. These results raised the possibility that gene body DA sRNAs may play a special role in stress responsive gene expression.

5.1.2 Identification of DE genes using mRNA sequencing

To examine the expression pattern of DA sRNA associated genes, whole genome mRNA sequencing was performed on the same 3 DPI total RNA samples used for sRNA sequencing plus one additional replicate to allow for improved statistical analysis of the sequencing data. mRNA sequencing was conducted by the Australian Genome Research Facility Ltd.

mRNA sequencing data was of high quality. Approximately 91 % of unique RNA reads could be mapped to the *A. thaliana* genome (supplementary Table 5.1). The percentage of mapped reads for the Fusarium infected *rdd* samples in comparison was low but still reached an average of 86%.

Table 5.2: DE *Arabidopsis thaliana* genes identified from the 3 DPI mRNA sequencing data

The number in parentheses indicate the number of DE genes common between the Col-0 mock vs Col-0 Fusarium and the *rdd* mock vs *rdd* Fusarium comparisons. All DE genes had a log₂ fold change (LFC) value of ≥1 or ≤-1 and P value of ≤0.05.

	Col mock vs <i>rdd</i> mock	Col mock vs Col Fox	<i>rdd</i> mock vs <i>rdd</i> Fox
Protein coding gene	33	1308	3175
Pseudogene	1	17	47
Transposon genes	11	39	114
Transfer RNA (tRNA)	0	1	8
Non-coding RNA (ncRNA)	4	16	48
Total	49	1381 (1206)	3392 (1206)

mRNA sequencing reads that mapped against Arabidopsis were run through the DESeq program to determine differential gene expression across the three comparisons. A heatmap was drawn to show the expression levels of all DE genes (up or downregulated) across the three comparisons (Supplementary Table 5.2). A small percentage of the DE genes had a highly significant $\log_2$ fold change (LFC) value of ≥ 4 or ≤ -4 . All the DE genes were annotated against protein coding genes, pseudogenes, transposons, transfer RNA (tRNA) and non-coding RNA (ncRNA) to determine their gene features (Table 5.2). A large number of DE genes were found in the mock vs Fusarium comparisons, with 1381 genes for the Col-0 background and 3392 for the *rdd* background. There was a significant overlap between the two sets of DE genes, with 1206 DE genes in common between the two genetic backgrounds, representing 87 % and 36 % of the total DE genes in Col-0 and *rdd* backgrounds, respectively. This result indicated that Fusarium infection had a major impact on genome-wide gene expression patterns. The much larger number of Fusarium induced DE genes in the *rdd* background compared to the Col-0 background suggested that DNA demethylases are involved in the regulation of these genes during stress response. Surprisingly, despite the high abundance of DA sRNA between *rdd* and Col-0 mock samples, the numbers of DE genes were much smaller for the Col-0 vs *rdd* comparison (49) than for the mock vs Fusarium comparisons. This result indicated that most of the DA sRNAs in the Col-0 vs *rdd* comparison are not associated with DE genes. It has previously been shown that the *rdd* mutations only affected a small number of genes or genomic loci in Arabidopsis (Penterman et al., 2007a, Lister et al., 2008, Le et al., 2014). The number of DE genes identified here, 49, is even smaller than that of the previously reported DE genes detected by microarray (348 genes) (Le et al., 2014) and mRNA sequencing (168 genes) (Lister et al., 2008), which was likely due to the stringent criterion used in the DESeq analysis.

To categorize the function of these DE genes, I performed gene ontology analysis using MapMan, and the result is shown in Supplementary Figure 5.1. The enriched pathways for DE genes in the Fusarium infected samples included RNA, miscellaneous, signalling, stress, transport, development, cell wall, hormone metabolism, and secondary metabolism. The DE genes from the Col-0 vs *rdd* comparison were also enriched for the miscellaneous, signalling, stress, and hormone metabolism pathways, but not for the RNA and transport pathways. Depletion in genes targeting the development, cell wall and secondary metabolism pathways is seen in Col-0 vs *rdd*. In addition, the DE genes from the Col-0 vs *rdd* comparison were also depleted in lipid and amino acid metabolism and redox pathways, but strongly enriched in the

DNA, miRNA, CHO metabolism and N-metabolism pathways. It is interesting that the stress, signalling and hormone metabolism pathways were enriched for the DE genes in all comparisons, consistent with a potential role of these genes in stress response. This result supports a role of DNA demethylases in regulating the expression of stress response genes.

5.1.3 Identification of DE genes associated with DA sRNA

To investigate if DA sRNAs play a role in the differential expression of the identified DE genes, the DE genes identified from mRNA sequencing were overlapped with the genes associated with DA sRNAs in the gene body and 3 kb flanking sequences. The result of the 3 DPI samples is presented in Table 5.3A, and that of all three time points listed in Table 5.3B. The DE genes from the Col-0 vs *rrd* mock comparison showed a high level of association with DA sRNAs particularly DA sRNAs in the upstream region (59 %). This result is consistent with the view that DNA demethylases interact with RdDM to regulate gene expression. The number of DE genes from the mock vs *Fusarium* comparisons was relatively small, which suggested that only a small fraction of the DE genes in *Fusarium* infected samples are regulated by sRNAs and the *Fusarium* responsive expression patterns of many other genes are independent of sRNAs. The total number of DE genes associated with DA sRNAs is larger for the *rrd* mock vs *rrd* *Fusarium* comparison than for the Col-0 mock vs Col-0 *Fusarium* comparison, suggesting the involvement of DNA demethylases in the sRNA mediated regulation of stress response genes.

Table 5.3: Number of DE genes of 3 DPI associated with DA sRNA

The value in the brackets indicates the percentage of total DE genes identified by mRNA-seq (as shown in Table 5.1) that overlaps with DA sRNA associated genes. The same gene may be counted more than once across the gene body and 3 kb flanking regions. All DA sRNA targeting DE genes had a log₂ fold change (LFC) value of ≥1 or ≤-1 and P value of ≤0.05.

A) Singular 3 DPI time point

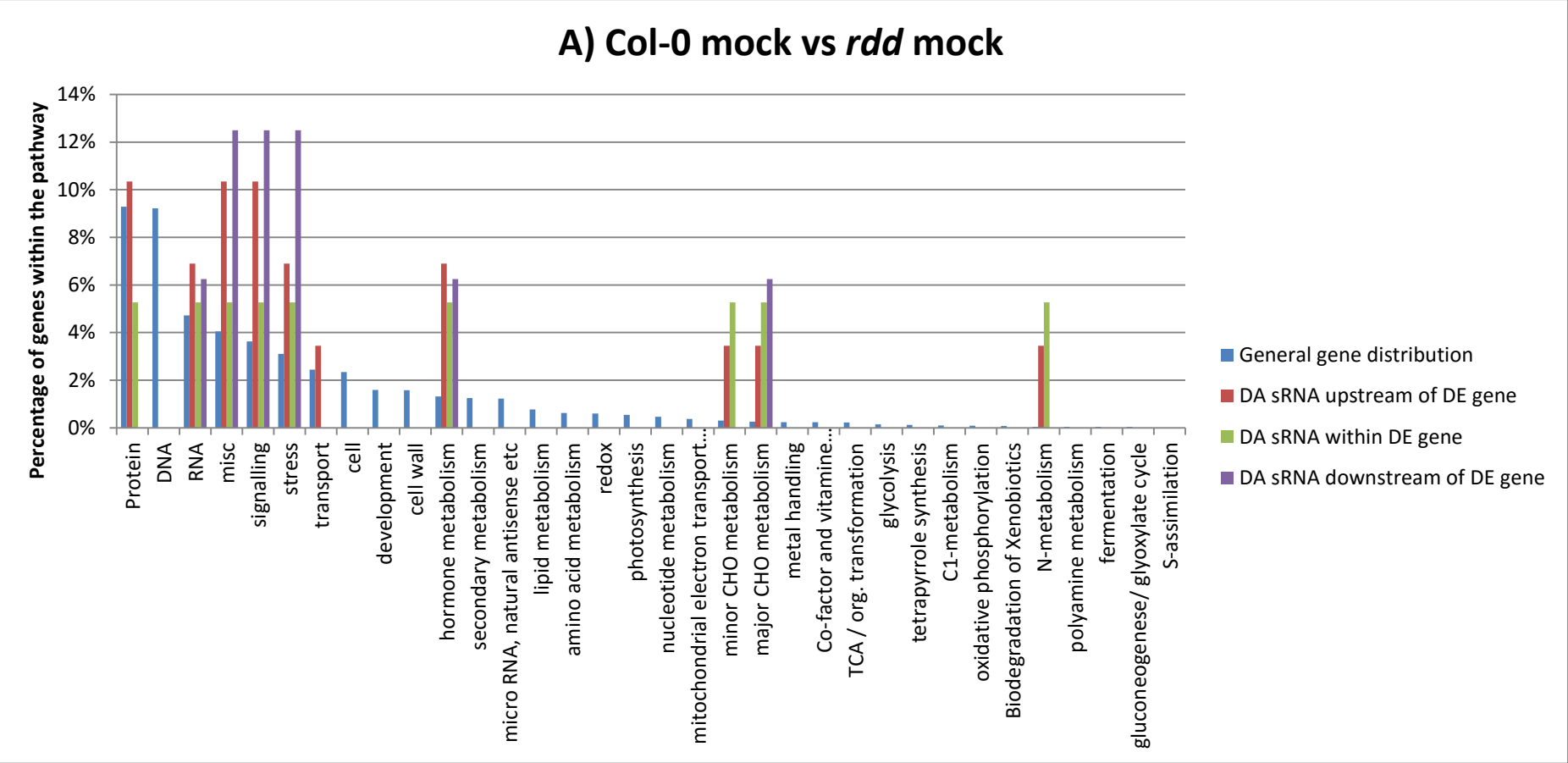
Comparison	Number of DA sRNA gene targets associated with DE genes			Unique (no duplicates) DA sRNA gene targets associated with DE genes		
	3kb upstream	Gene body	3kb downstream	3kb upstream	Gene body	3kb downstream
Col-0 mock vs <i>rdd</i> mock	270	85	91	29 (59%)	19 (39%)	16 (32%)
Col-0 mock vs Col-0 Fox	123	326	186	71 (5%)	104 (8%)	95 (7%)
<i>rdd</i> mock vs <i>rdd</i> Fox	610	872	644	301 (9%)	256 (8%)	300 (9%)

A) 1,3, and 6 DPI Overlapping time points

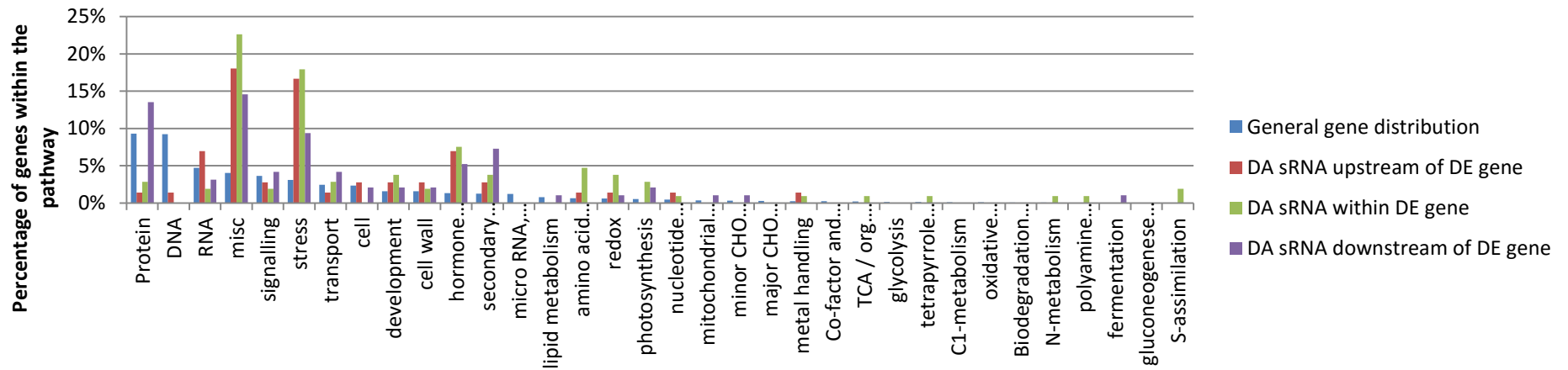
Comparison	Number of DE genes associated with DA sRNA			Unique (no duplicates) DE genes associated with DA sRNA		
	3kb upstream	Gene body	3kb downstream	3kb upstream	Gene body	3kb downstream
Col-0 mock vs <i>rdd</i> mock	222	64	64	29 (59%)	16 (32%)	13 (26%)
Col-0 mock vs Col-0 Fox	13	41	12	8 (0.6%)	14 (1%)	4 (0.3%)
<i>rdd</i> mock vs <i>rdd</i> Fox	40	111	54	18 (0.5%)	41 (1%)	21 (0.6%)

Figure 5.2: Function characterisation of DA sRNA associated DE genes at 3DPI using MapMan gene ontology analysis

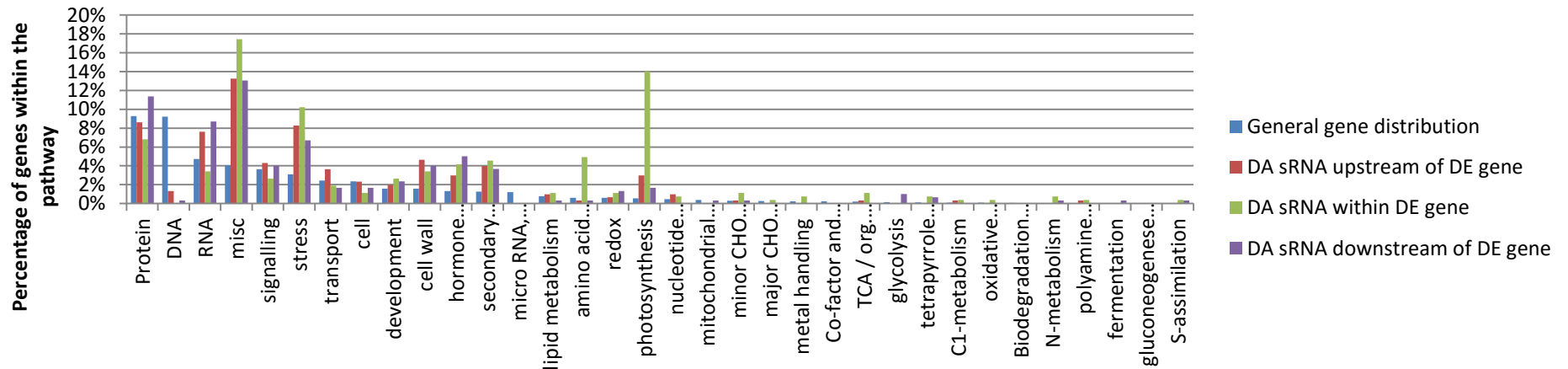
The histograms show the percentage of DA sRNA associated DE genes of a particular pathway in total DA sRNA associated DE genes. The percentage of all genes of a particular pathway in total Arabidopsis genes is indicated by the blue bar on the left for comparison. The list of genes that are part of each pathway is available from Thimm et al., 2004.



B) Col-0 mock vs Col-0 Fox



C) *rdd* mock vs *rdd* Fox



DE genes which had DA sRNA either within the gene body or in the surrounding 3 kb flanking regions were run through MapMan to categorize the function of these genes. Interestingly, the DA sRNA associated DE genes from all comparisons are enriched for the stress and hormone metabolism pathways plus the miscellaneous pathway (Figure 5.2). The miscellaneous pathway means genes targeted a variety of pathways. This result is consistent with these genes being potentially involved in stress response and with the DNA demethylases and with sRNAs playing a potential role in regulating stress response genes. DE genes associated with gene body and flanking sequence targeting DA sRNAs showed similar trends, but with slight variations. For DE genes from the Col-0 vs *rdd* comparison, those associated with flanking sequence DA sRNAs are more enriched than those associated with gene body DA sRNAs. On the other hand, DE genes associated with gene body DA sRNA in the above mentioned pathways become more prominent in the mock vs *Fusarium* comparisons. These results suggest that DA sRNAs located in all regions of genes play a role in gene regulation, but gene body DA sRNAs might be particularly important in stress responsive expression of genes. There were some clear differences in the distribution of DA sRNA associated DE genes between the Col-0 vs *rdd* and the mock vs *Fusarium* comparisons (Figure 5.2). For instance, the CHO and N-metabolism pathways were strongly enriched for the Col-0 vs *rdd* comparison but not for the mock vs *Fusarium* comparison.

The analysis so far in this chapter was focused on DA sRNAs located within the gene body or flanking region of DE genes, which are primarily of the 24 nt size class. In Chapter 4 I also identified a number of differentially accumulated miRNAs in the three comparisons. miRNAs target genes in trans and it would be interesting to know if these DA miRNAs are associated with differential expression of their target genes. The 35 miRNA families found to have differential accumulation in Chapter 4 have 1009 potential target genes. None of these genes were differentially expressed between Col-0 and *rdd* (Table 5.4). In contrast, 14 of the 35 DA miRNA families were associated with DE genes at 3 DPI in *Fusarium* infected plants (Table 5.4). Negative correlation between DA miRNA expression pattern and DE gene expression pattern was observed for miR156G, miR395B,C,E, miR397A,B in the Col-0 background, and for miR156G, miR164A,B, miR397A,B, miR831, miR850 and multiple members of miR169 and miR395 families in the *rdd* backgrounds. miR156G, miR395 and miR397 were consistently associated with DE target genes in both genetic backgrounds, suggesting that they play a role in regulating the expression of these genes during *Fusarium* infection.

Table 5.4: Expression pattern of DE genes potentially targeted by DA miRNAs

All DA miRNA and DE gene targets had a log₂ fold change (LFC) value of ≥1 or ≤-1 and P value of ≤0.05.

	DA miRNA	LFC	DE miRNA Target Gene	LFC mRNA-seq
Col mock vs <i>rdd</i> mock	-	-	-	-
Col mock vs Col Fox	miR156G	1.658	AT2G42200	-1.184
	miR169F	1.489	AT2G39210	1.297
	miR395B	3.165	AT5G13630	-1.236
	miR395C	2.585		
	miR395D	2.807	AT4G14680	2.330
			AT3G22890	1.220
	miR395E	2.807	AT5G13630	-1.236
			AT4G14680	2.330
			AT3G22890	1.220
	miR397A	1.962	AT2G29130	-1.540
	miR397B	2.022		
	miR398B	1.331	AT2G04040	1.166
	miR398C	1.065		
	miR857	1.524	AT5G58390	1.036
<i>rdd</i> mock vs <i>rdd</i> Fox	miR156G	1.318	AT2G42200	-1.996
	miR164A	-1.097	AT3G12977	-1.697
			AT5G39610	2.510
			AT1G10530	1.446
	miR164B	-1.015	AT3G12977	-1.697
			AT5G39610	2.510
			AT1G10530	1.446
	miR165A	-1.146	AT5G01090	-1.076
	miR169A	-1.012	AT1G54160	1.433
			AT1G72830	1.227

	DA miRNA	LFC	DE miRNA Target Gene	LFC mRNA-seq
<i>rdd</i> mock vs <i>rdd</i> Fox	miR169B	-1.142	AT1G54160	1.433
			AT1G72830	1.226
	miR169H	-1.919	AT1G54160	1.433
			AT1G72830	1.226
	miR169L	-1.773	AT1G54160	1.433
			AT1G72830	1.226
	miR169D	-1.247	AT1G68560	-1.045
			AT2G39210	1.058
	miR169E	-2.755	AT1G68560	-1.045
			AT2G39210	1.058
	miR169G	-2.021	AT1G68560	-1.045
			AT2G39210	1.058
	miR171B	-1.675	AT4G21990	1.477
			AT1G73300	-1.407
			AT1G73270	-1.047
			AT5G61480	-1.190
			AT4G18340	1.658
	miR390B	-1.747	AT3G17185	-1.393
			AT5G03640	1.428
			AT2G28780	-1.891
	miR395A	2.202	AT2G28780	-1.891
			AT5G13630	-1.531
			AT4G14680	2.617
			AT3G22890	1.811
	miR395B	1.228	AT2G28780	-1.891
	miR395B	1.228	AT5G13630	-1.531

	DA miRNA	LFC	DE miRNA Target Gene	LFC mRNA-seq
<i>rdd</i> mock vs <i>rdd</i> Fox	miR395C	1.585	AT2G28780	-1.891
			AT5G13630	-1.531
	miR395D	2.874	AT2G28780	-1.891
			AT5G13630	-1.531
			AT4G14680	2.617
			AT3G22890	1.811
	miR395E	2.248	AT2G28780	-1.891
			AT5G13630	-1.531
			AT4G14680	2.617
			AT3G22890	1.811
	miR395F	1.529	AT2G28780	-1.891
			AT5G13630	-1.531
	miR397A	1.171	AT2G29130	-2.361
			AT4G33230	1.110
	miR397B	1.598	AT2G29130	-2.361
	miR822	-1.181	AT4G08160	-1.284
	miR831	1.773	AT1G63100	-1.516
			AT1G68765	1.366
	miR850	-1.348	AT3G28007	2.193
	miR857	2.000	AT4G27850	1.469
	miR869	-1.699	AT1G54020	6.549
			AT5G52280	-1.103
			AT1G67120	-1.321

5.2 Discussion

The relationship between genome-wide changes in sRNA accumulation and gene expression in the *rdd* mutant or following *Fusarium* infection has not been investigated before. In this chapter I investigated this first by finding the list of genes that had overlap with DA sRNAs in the transcribed, 5' and 3' flanking regions, and then examining the expression pattern of these DA sRNA associated genes using mRNA sequencing. Only the 3DPI samples were included in the mRNA sequencing analysis mainly due to limited funds available for sequencing.

Thousands of genes were found to be associated with DA sRNAs in the Col-0 vs *rdd* comparison. This large number is in sharp contrast to the relatively small number of genes that are differentially expressed between Col-0 and *rdd* found in the mRNA sequencing analysis here and in previous studies (Lister et al., 2008, Le et al., 2014). This indicated that the majority of the DA sRNAs is not associated with differential gene expression in *rdd*. This lack of association between DA sRNAs and DE genes could suggest that only specific regions of genes, like the exons or introns are targeted by sRNAs. Pre-existing DNA methylation may enhance 24 nt siRNA production by helping proper recruitment of Pol IV which will be further analysed in Chapter 6. This study only included whole regions in the analysis, namely the gene body and the 3 kb flanking sequences. DA sRNA within the 5' upstream region for all three comparisons did have the largest number of genes targeted suggesting DA sRNA may influence the transcription start site. What specific regions are the main targets of sRNAs require further investigation. Regardless of the DA sRNA location a large proportion of the 49 DE genes between Col-0 and *rdd* were associated with DA sRNAs. Furthermore, these DE genes are particularly enriched for pathways such as stress, signalling and hormone metabolism, which are related to stress response functions. These results suggest that these DA sRNA associated DE genes are regulated by the coordinated action of sRNAs and DNA demethylases, supporting the view that RdDM and DNA demethylases interact in regulating stress gene expression (Le et al., 2014).

Among the DE genes in the Col-0 vs *rdd* comparison were transposon genes, representing 22 % of the total 49 DE genes. TEs are known to be regulated by RdDM (Sun et al., 2013, Eichten et al., 2013a). This analysis suggests that the expression of TEs is maintained by DNA demethylases in conjunction with RdDM.

A much larger number of DE genes were found in the mRNA sequencing analysis from the Fusarium infected samples of both Col-0 and *rdd* than from the Col-0 vs *rdd* comparison. The number in the Col-0 background, 1381, was significantly larger than those identified previously using mRNA sequencing of mock treated and Fusarium infected Col-0, namely 207 at 1 DPI and 696 at 6 DPI DE genes (Zhu et al., 2013). This increased number was likely due to the greater depth of Illumina HiSeq 2000 mRNA sequencing data in this thesis (over 14 million mapped reads for each sample) than the SOLiD strand-specific mRNA sequencing data (around 7 million mapped reads per sample) (Zhu et al., 2013). The relatively large number of DE genes in Fusarium infected samples indicated that Fusarium infection has a much stronger impact on gene expression than the mutation of DNA demethylases. However, the majority of these DE genes are not associated with DA sRNAs, indicating that the Fusarium responsive expression of most genes is not regulated by sRNAs. Over 100 DE genes in the Col-0 mock vs Col-0 Fusarium comparison and over 300 DE genes in the *rdd* mock vs *rdd* Fusarium comparison were associated with DA sRNAs. Like those in the Col-0 vs *rdd* comparison, these DE genes were also enriched for stress response related pathways including the stress and hormone metabolism pathways, suggesting the involvement of sRNAs and RdDM in stress gene regulation. The significantly larger number of DE genes in the *rdd* background than the Col-0 background following Fusarium infection further suggest an involvement of DNA demethylases in this sRNA mediated regulation of stress response gene expression.

To establish a correlation between DA sRNA and DE gene expression the list of genes targeting the photosynthetic pathway were compared between the list of DA sRNA with DE genes (Figure 5.2) against the DA sRNA associated genes (Figure 5.1). The number of DE genes with DA sRNA was less than 10 across all regions (upstream, downstream and within the gene body) of the mock vs Fusarium comparisons. The only exception was the *rdd* mock vs *rdd* Fox comparison within the gene body that had 37 photosynthetic genes targeted. These genes were also found to be the targets of DA sRNA although the DA sRNA targeted between 31-88 photosynthetic genes depending on the comparison.

An interesting difference between DA sRNA associated DE genes in *rdd* and upon Fusarium infection is the differential representation of gene body DA sRNAs and flanking region DA sRNAs. The Fusarium induced DA genes are in general more associated with gene body DA sRNAs, whereas the *rdd* affected DE genes are more associated with flanking region DA sRNAs. This raises the possibility that gene body DA sRNAs are particularly important in Fusarium

inducible gene expression. It would be interesting to examine the features of the gene body target sequences for these DA sRNAs.

In this chapter I also looked at the potential effect of miRNAs on differential gene expression associated with the *rdd* mutation or Fusarium infection. I found a number of miRNA target genes are potentially regulated by some of the DA miRNAs because their expression showed an inverse correlation with the abundance of miRNAs. These miRNA target genes are worth further investigation. However, as this thesis focuses on RdDM and DNA demethylases, the biological function of these DA miRNAs and associated DE genes was not examined further in this thesis.

5.3 Conclusions

Some previous studies have analysed the differential gene expression between *rdd* and wild-type *Arabidopsis* plants under normal growth conditions (Le et al., 2014, Lister et al., 2008) and in Col-0 following *Fusarium* infection (Zhu et al., 2013) implicating both DNA demethylases and RdDM in disease resistance. No studies to date have examined the relationship between sRNA accumulation and gene expression changes upon *Fusarium* infection in *rdd*. In this chapter I i) performed a genome-wide survey of genes associated with DA sRNAs identified in Chapter 4; ii) obtained and analysed mRNA sequencing data of 3 DPI mock treated and *Fusarium* infected Col-0 and *rdd* samples; iii) identified DE genes associated with DA sRNAs; and iv) performed gene ontology analysis to categorize the function of these genes. My results show the DA sRNA associated DE genes were enriched for biological pathways related to stress responses, providing support for the view that DNA demethylases interact with RdDM to regulate stress response genes. In addition, I also identified a number of miRNA target genes that appeared to be regulated by the DA miRNAs during *Fusarium* infection. In conclusion, in this chapter I successfully generated a list of genes that were associated with DA sRNAs and differentially expressed in *rdd* or upon *Fusarium* infection. I have therefore provided a foundation for further investigation of the role of sRNAs and DNA demethylases in plant response to *Fusarium* or other fungal pathogens.

Chapter 6: Whole Genome Identification of Differentially Expressed sRNA Target Genes with Differentially Methylated Regions

6.0 Introduction

In Chapters 4 and 5 I investigated the relationship between sRNAs and gene expression in the DNA demethylase mutant *rdd* and upon *Fusarium* infection. I showed that many differentially expressed (DE) genes were associated with differentially abundant (DA) sRNA, suggesting a role of these sRNAs in the regulation of these genes. These sRNAs were predominantly 24 nt in size capable of directing RdDM, which implied that they regulate the DE genes by inducing changes in DNA methylation.

A genome-wide study (Lister et al., 2008) found that in *rdd* an increase in DNA methylation occurs most frequently in promoters and 3' UTRs, suggesting that DNA demethylation is most active in flanking gene regions. However, while the same study also examined transcriptome and sRNA changes between *rdd* and Col-0, no detailed discussion was made linking gene expression changes with changes in sRNA accumulation or DNA methylation.

Transposable elements (TEs) are the major target of sRNA directed DNA methylation. It was previously shown that TE sequences in the promoters of some *rdd* downregulated stress response genes showed localized hypermethylation (an increase in DNA methylation) in *rdd* compared to Col-0 (Le et al., 2014), suggesting that these near-gene TE sequences are targeted by sRNAs. It would be interesting to examine if sRNA changes occur in these hypermethylated TE sequences in *rdd*. The *rdd* downregulated stress response genes in Arabidopsis (Le et al., 2014) and stress response genes in maize (Makarevitch et al., 2015), are enriched for near-gene TE sequences, implicating TEs as a regulator for stress responsive expression patterns. However, if the stress responsive expression of these genes is regulated through sRNA and DNA methylation changes in the TE sequences remains unclear.

Genome-wide DNA hypomethylation (decrease in DNA methylation) has been shown to be associated with infection by the bacterial pathogen *Pseudomonas syringae* (Pavet et al., 2006). This hypomethylation is likely to be important for plant defence against the bacterial pathogen, as DNA methylation deficient mutants, with genome-wide hypomethylation, are generally more resistant to *P. syringae* than wild type plants, presumably due to increased expression of defence genes (Downen et al., 2012). On the other hand the DNA demethylase mutant, *ros1*, showed enhanced susceptibility to *P. syringae*, presumably due to decreased defence gene expression (Yu et al., 2013). It would be interesting to examine if *Fusarium* infection also causes genome-wide changes in DNA methylation.

In this chapter I aim to investigate the possible link of sRNA and gene expression changes with DNA methylation changes in *rdd* and upon Fusarium infection by performing genome-wide DNA methylation analysis in mock treated and Fusarium infected Col-0 and *rdd*. Whole genome bisulfite sequencing was performed on the 3 DPI samples of mock treated and Fusarium infected Col-0 and *rdd* that were used for sRNA and mRNA sequencing analyses. Specifically, in this chapter I will address the following questions:

1. What are the genes that show differential DNA methylation in *rdd* and upon Fusarium infection and
2. What are the biological functions of the genes targeted by differential methylation?
3. What is the relationship between DA sRNAs, differential DNA methylation and differential gene expression in *rdd* and upon Fusarium infection?

6.1 Results

6.1.1 Identification of genes with DMRs

Genome-wide bisulfite sequencing was performed to identify differentially methylated regions (DMRs) between Col-0 and *ddd* and between mock treated and Fusarium infected samples at 3 DPI. I collaborated with the Pogson Lab at the ANU to construct the bisulfite sequencing libraries, which were sequenced by the Lister Lab at the UWA, resulting in high quality sequencing data. As summarised in Supplementary Table 6.1, approximately 29-37 million reads were obtained for each sample, which were mapped using the Bismark aligner as outlined in Chapter 2. This resulted in between 20-30 million uniquely aligned reads for each sample. Individual cytosine methylation calls were assigned to one of the three cytosine contexts (CG, CHG or CHH where H stands for A, T, or C). The number of cytosines covered as a proportion of the total cytosines present in *Arabidopsis* is high (95 or 96 %) across all samples, confirming the high quality of the sequencing data.

The distribution of individual cytosine context as a percentage of total cytosine sites was the same in all samples, which were 73%, 14% and 13% for the CHH, CHG and CG contexts, respectively. When comparing methylated cytosines of the individual context against the total methylated cytosines (supplementary Figure 6.1), the CHH context still represents the highest proportion of the methylated cytosines (49-53 %) followed by the CG (29-32 %) and the CHG contexts (18-19 %). Interestingly, the proportion of methylated cytosines in the CG context increased slightly (by 2-3 %) in both Col-0 and *ddd* backgrounds in Fusarium-infected samples, which was accompanied by a slight decrease in the CHH context (Supplementary Figure 6.1). The proportion of methylated cytosines in the CHG context remained stable compared to the other two contexts. In general, there was no difference in the distribution of methylated cytosine contexts between Col-0 and *ddd*.

DMRs were called as described in Chapter 2. Briefly, to be considered a DMR, both samples of the comparison had to show at least 10x read coverage with 80 % difference in CG, 50 % in CHG, and/or 20 % in CHH methylation within a 100 bp window. These values were based on the DNA methylation patterns found by Cokus and colleagues (2008).

The overall distribution of differentially methylated cytosine contexts as a percentage of total DMRs within the 100 bp windows is shown in Figure 6.1. The proportions of differentially methylated cytosines across the three contexts in the Col-0 mock vs *ddd* (a different *ddd* mutant)

mock comparison are similar to published data (Lister et al., 2008), with the CG context representing over 50 % of the DMRs while the CHH and CHG contexts represent approximately 20 % of the DMRs.

Upon *Fusarium* infection the largest proportion of DNA methylation changes occur in the CHH context (Figure 6.1). This was also observed for DMRs found in gene rich regions in response to *P. syringae* (Downen et al., 2012), suggesting that CHH methylation is the major target of stress induced methylation changes. The proportion of differentially methylated CHH sites in *rdd* (69%) is smaller than in Col-0 (82%) following *Fusarium* infection, suggesting that DNA demethylases play a role in the *Fusarium* induced CHH methylation changes. The distribution of differentially methylated CG and CHG sites upon *Fusarium* infection is also different between Col-0 and *rdd* backgrounds. The DMRs in *Fusarium* infected Col-0 have 12 % of cytosines in the CG context and 6 % in the CHG context, whereas for the *rdd* mock vs *rdd* *Fusarium* comparison only 3 % of DMRs was in the CG context and 28 % in the CHG context.

A large number of genes were found to be associated with DMRs, particularly in the Col-0 vs *rdd* comparison (Table 6.1). Future analysis should be done to identify the function of the genes associated with DMR across the three cytosine contexts (ie CG, CHG and CHH). The DMRs are located primarily in the 5' and 3' flanking regions, but some DMRs also occur in the gene body (Supplementary Figure 6.2). In *rdd* the CG DMRs have the largest number of associated genes. Consistent with the CHH DMRs being the most frequent in the mock vs *Fusarium* comparison, they are also associated with the largest number of genes in this comparison. As CHH methylation depends entirely on RdDM, this result suggests that RdDM plays a role in the regulation of DMR associated genes during *Fusarium* infection.

Figure 6.1: Overall distribution of differentially methylated cytosine as a percentage of total DMRs

The total DMR loci may be counted twice if that loci have more than one cytosine context differentially methylated.

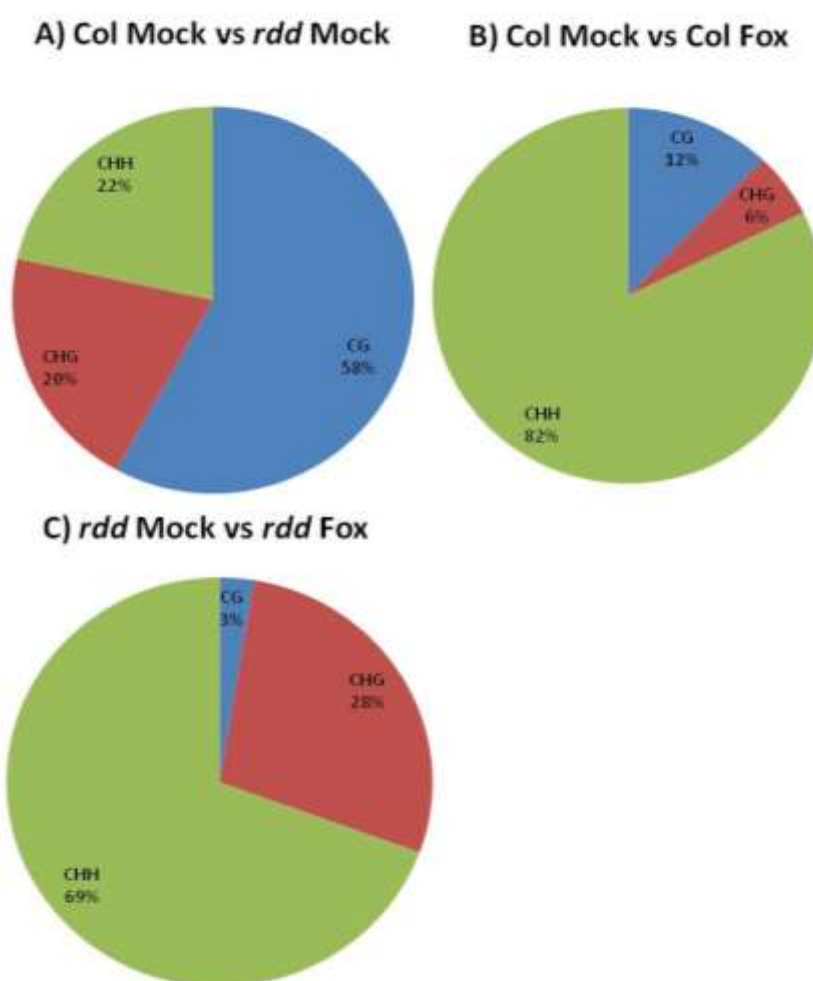


Table 6.1: DMRs associated with genes at 3DPI

The number of DMR includes the differentially methylated cytosines found within the gene body and the 3 kb flanking regions. The total number of DMR does not count a locus twice if that loci has more than one cytosine context differentially methylated.

Comparison	Total DMR	Number of DMR			Total number of gene targets	Number of gene targets		
		CG	CHG	CHH		CG	CHG	CHH
Col mock vs <i>rdd</i> mock	14,132	8,903	3,136	3,308	13,848	11,833	4,356	4,115
Col mock vs Col Fox	1,451	258	119	1,750	1,528	562	98	898
<i>rdd</i> mock vs <i>rdd</i> Fox	2,138	34	368	913	2,049	72	417	1,647

6.1.2 Identification of DMRs associated with DA sRNAs at 3 DPI

The genes with DA sRNAs identified in the last chapter were overlapped with DMR associated genes, resulting in a list of genes with both DA sRNA and DMRs (Table 6.2). The total numbers of genes in Table 6.2 having both DA sRNA and DMR is much smaller than those in Table 6.1 showing the number of genes associated with DMRs or those in Table 5.1 associated with DA sRNA. This indicates the majority of the genes are associated with either DA sRNA or DMRs but not with both. For the genes with both DA sRNA and DMR (Table 6.2), the DA sRNA sites may overlap with the DMRs and/or may fall within more than one of the stated region (ie upstream, gene body or downstream) of the genes. Consistent with the relatively large number of DA sRNA or DMR associated genes, the Col-0 vs *rrd* comparison had 5,585 genes with both DA sRNA and DMRs, whereas the mock vs *Fusarium* comparisons gave a much smaller number (98 genes in Col-0 and 352 genes in *rrd*). The majority of these DA sRNAs and DMRs are located in the flanking regions of genes, especially for the Col-0 vs *rrd* comparison. In addition, the CG DMRs are the most abundant for DA sRNA associated genes in the Col-0 vs *rrd* comparison, whereas the CHH DMRs were the most frequent for genes in the mock vs *Fusarium* comparison.

6.1.3 Identification of DE genes associated with DMRs

To examine if the DMRs are associated with differential gene expression, genes with DMRs (Table 6.1) were overlapped with the list of differentially expressed (DE) genes identified using mRNA sequencing for the 3 DPI samples described in Chapter 5 (Table 5.2). As shown in Table 6.3, 29 of the 49 genes differentially expressed between Col-0 and *rrd* were associated with DMRs, which were located in the 3 kb upstream region (for 21 genes), the gene body (for 10 genes) and/or the 3 kb downstream region (for 11 genes). Thus, over half (59 %) of the DE genes in *rrd* are associated with DMRs, suggesting that these DMRs are involved in the differential gene expression in *rrd*. The total number of DE genes with DMRs for the mock vs *Fusarium* comparison is 65 and 273 for the Col-0 and *rrd* backgrounds, respectively. The majority of the associated DMRs are in the CHH context, indicating that the differential methylation is due to RdDM. This suggests that RdDM plays a role in the regulation of these genes during *Fusarium* infection. The much larger number of DMR associated genes in the *rrd* background (273) than the Col-0 background (65) suggests that RdDM interacts with DNA demethylases in the regulation of these *Fusarium* responsive genes. However, the numbers of DMR associated DE genes are relatively small compared to the total numbers of DE genes at 3 DPI (1381 for Col-0 and 3392 for *rrd*, Table 5.2), suggesting that most of the DE genes are not regulated through differential DNA methylation at this time point.

Gene ontology analysis was performed using MapMan on the DMR associated DE genes. Because of the relatively small gene number, a comparison with the overall Arabidopsis gene distribution was not done. Nevertheless, the stress pathway was well represented in all three comparisons: 3, 9 and 21 of the DMR associated DE genes from the Col-0 vs *rdd*, Col-0 mock vs Col-0 Fusarium, and *rdd* mock vs *rdd* Fusarium comparisons, respectively, respond to stress.

6.1.4 Identification of DE genes associated with both DA sRNA and DMRs

The purpose of the whole genome mRNA transcriptome and bisulfite sequencing was to examine if the DA sRNAs identified in Chapter 4 play a role in the regulation of *rdd* affected and Fusarium responsive genes, and if this regulation involves changes in DNA methylation that may lead to gene expression differences. The three gene lists, namely the DE genes, genes associated with DA sRNAs and genes associated with DMRs, were therefore overlapped to identify DE genes with both DA sRNA and DMRs. The 3 DPI data were used as the mRNA sequencing and bisulfite sequencing was performed for only this time point. The same 3 DPI plant samples were used for the sRNA, mRNA and bisulfite sequencing analyses, allowing for cross comparison void of any sample variations. Table 6.4 displays the total numbers of DE genes, genes with DA sRNA, and genes with DMRs at 3 DPI, and Table 6.5 shows the number of genes showing differential expression and having DA sRNA and DMRs. Over 50 % of the 49 DE genes in the Col-0 vs *rdd* comparison have both DA sRNA and DMRs, which is a remarkably high proportion suggesting that the *rdd* affected genes are regulated by the coordinated action of DNA demethylases and sRNA directed DNA methylation (or RdDM). In contrast to the *rdd* affected genes that were compared to Col-0, the mock vs Fusarium comparisons had a very small proportion of the DE genes (0.6 % in Col-0 and 1.5 % in *rdd* plants) that overlapped with DA sRNA and DMR regions. This indicates that the majority of Fusarium responsive genes is not regulated through sRNA directed DNA methylation at the 3 DPI time point. Overall each comparison showed the upstream and downstream flanking 3 kb regions to be the most significant regions for DA sRNA to potentially influence DE gene expression and cytosine methylation.

Gene ontology analysis was performed using MapMan on the DE genes associated with DA sRNA and DMRs (Supplementary Table 6.2). Because of the relatively small gene number, a comparison with the overall Arabidopsis gene distribution was not done. Nevertheless, the stress and signalling pathway was well represented in all three comparisons: 5, 3 and 5 of the DE genes belonged to these pathways and were associated with DA sRNA and DMR from the Col-0 vs *rdd*, Col-0 mock vs Col-0 Fusarium, and *rdd* mock vs *rdd* Fusarium comparisons, respectively. The RNA pathway was also enriched in *rdd* in response to Fusarium.

Table 6.2: Number of genes with both DA sRNA and DMR that overlap with or are close to each other within the gene body or at 3 kb flanking regions

An overlap may exist between the DA sRNA and the different cytosine contexts gene targets.

Comparison	Total number of genes targeted by DA sRNA with DMR	Upstream of the gene target				Within the gene body				Downstream of the gene target			
		Total number of genes targeted by DA sRNA with DMR	DA sRNA with CG DMR	DA sRNA with CHG DMR	DA sRNA with CHH DMR	Total number of genes targeted by DA sRNA with DMR	DA sRNA with CG DMR	DA sRNA with CHG DMR	DA sRNA with CHH DMR	Total number of genes targeted by DA sRNA with DMR	DA sRNA with CG DMR	DA sRNA with CHG DMR	DA sRNA with CHH DMR
Col mock vs <i>rdd</i> mock	5,585	3,333	2,291	1,688	1,513	555	384	243	173	2,695	1,893	1,218	1,158
Col mock vs Col Fox	98	45	0	2	43	6	1	1	4	47	0	3	44
<i>rdd</i> mock vs <i>rdd</i> Fox	352	201	5	36	169	27	0	7	22	129	2	27	104

Table 6.3: Number of DE genes associated with DMRs at 3 DPI

An overlap in the total DE genes and the different cytosine contexts may exist.

Comparison	Total number of DMR targeting DE genes	Upstream of the gene target				Within the gene body				Downstream of the gene target			
		Total number of DMR targeting DE genes	DE genes with CG DMR	DE genes with CHG DMR	DE genes with CHH DMR	Total number of DMR targeting DE genes	DE genes with CG DMR	DE genes with CHG DMR	DE genes with CHH DMR	Total number of DMR targeting DE genes	DE genes with CG DMR	DE genes with CHG DMR	DE genes with CHH DMR
Col mock vs <i>rdd</i> mock	29	21	12	9	17	10	8	2	3	11	9	3	4
Col mock vs Col Fox	65	31	5	3	23	4	2	0	2	33	10	1	22
<i>rdd</i> mock vs <i>rdd</i> Fox	273	133	4	21	113	23	0	6	15	133	1	21	113

Table 6.4: Total numbers of DE genes, genes with DA sRNA, and genes with DMRs at 3 DPI

	DA sRNA	DA sRNA gene targets	DE mRNA	DMR	DMR gene targets
Col mock vs <i>rdd</i> mock	44,901	12,548	49	14,132	13,848
Col mock vs Col Fox	3,347	2,975	1,381	1,451	1,528
<i>rdd</i> mock vs <i>rdd</i> Fox	7,574	4,720	3,392	2,138	2,049

Table 6.5: 3 DPI DA small RNA overlapping with DE genes and DMRs

	Upstream	Gene body	Downstream	Total gene targets
Col mock vs <i>rdd</i> mock	20	6	9	25
Col mock vs Col Fox	3	1	4	8
<i>rdd</i> mock vs <i>rdd</i> Fox	23	6	26	53

6.1.5 DE genes downregulated in *rdd* are associated with increased sRNA and DNA methylation in the promoter region

To examine the DA sRNA and DMR associated DE genes in more detail, 15 genes were chosen from the gene list of the Col-0 mock vs *rdd* mock comparison, which also had differential sRNA abundance, gene expression and DNA methylation in the Col-0 Fusarium vs *rdd* Fusarium comparison (results not presented in this thesis). These chosen genes were therefore most likely to be regulated by DNA demethylases. The 15 genes all had DA sRNA upstream of the gene body. Based on MapMan analysis these genes covered a range of biological functions including hormone metabolism, signalling, RNA and the stress pathway (Supplementary Table 6.2).

A heatmap was generated to establish if the 15 genes had a similar pattern of sRNA, DMR and gene expression changes between Col-0 and *rdd* (Figure 6.2). The heatmap was split into sRNA, mRNA and the three DMR contexts. Most of the DA sRNA cells were divided in two halves because two sRNA clusters were found with varying log fold changes. For one gene, AT2G33830, the CG and CHH DMR cells are also split into two halves to indicate two different DMRs with different fold changes. The two or three rows for some of the genes represent multiple sRNA clusters targeting different DMR regions. White space indicates no changes were found for this region. Interestingly, these 15 genes shared two common features, firstly their expression

levels, based on mRNA sequencing reads, were all downregulated in the DNA demethylation mutant *rdd* compared to Col-0. This downregulation was also observed in the previous microarray study (Le et al., 2014). Secondly, the DMR regions generally occurred in all three cytosine contexts with an increase in DNA methylation in *rdd* compared to Col-0.

The sRNA, mRNA and bisulfite sequencing reads were loaded onto IGV, a visualisation browser, to visually examine the relationship between sRNA, DNA methylation and gene expression in the 15 genes. In the example for AT2G33830, a dormancy associated gene, the decreased gene expression in *rdd* is associated with an increase in sRNA abundance and in DNA methylation of all 3 cytosine contexts (Figure 6.3). It is important to note that while DNA methylation exist across a wide range of the upstream sequence in both the wild-type Col-0 and *rdd*, the DMR, characterized by the existence of CG, CHG and CHH methylation in *rdd* and the absence of detectable methylation in Col-0, is located at the same region as the DA sRNA. This result suggests that sRNAs are directly responsible for the localized DNA methylation in *rdd*, and the DNA methylation is responsible for the downregulation of the gene in *rdd*. Furthermore, the DMR occurred immediately adjacent to the existing methylation region in Col-0, which overlaps with a TE sequence (Figure 6.3). This suggested that the newly occurred methylation in *rdd* results from a sRNA directed spread initiated from a methylated TE region. A similar distribution pattern of DMR and DA sRNA was observed for 13 out of the 15 genes. That is, DNA methylation occurred in *rdd* but was absent in Col-0, and the DMRs overlapped with the sRNA sites in *rdd* and was immediately adjacent to an existing methylated region in Col-0 which overlaps or is close to a TE sequence (see Supplementary Figures 6.3-6.5 for three more examples). The DNA methylation coverage in *rdd* tends to be upstream spreading towards the transcription start site of the DE genes (Figure 6.3 and Supplementary Figure 6.3-6.5). The region near the transcription start site is likely to be an essential part of the gene promoter important for transcription initiation. Therefore, this DNA methylation spread, triggered by DA sRNA from methylated TE or near-TE regions, is likely to account for the repressed gene expression in *rdd*.

The presence of DNA demethylation in *rdd* but absence of methylation in Col-0 were also observed for several regions by Lister and colleagues (2008). In addition to such dramatic changes in sRNA and DNA methylation, some other minor sRNA and DNA methylation changes were also detected between Col-0 and *rdd* in my analysis but these did not appear to correlate with one another.

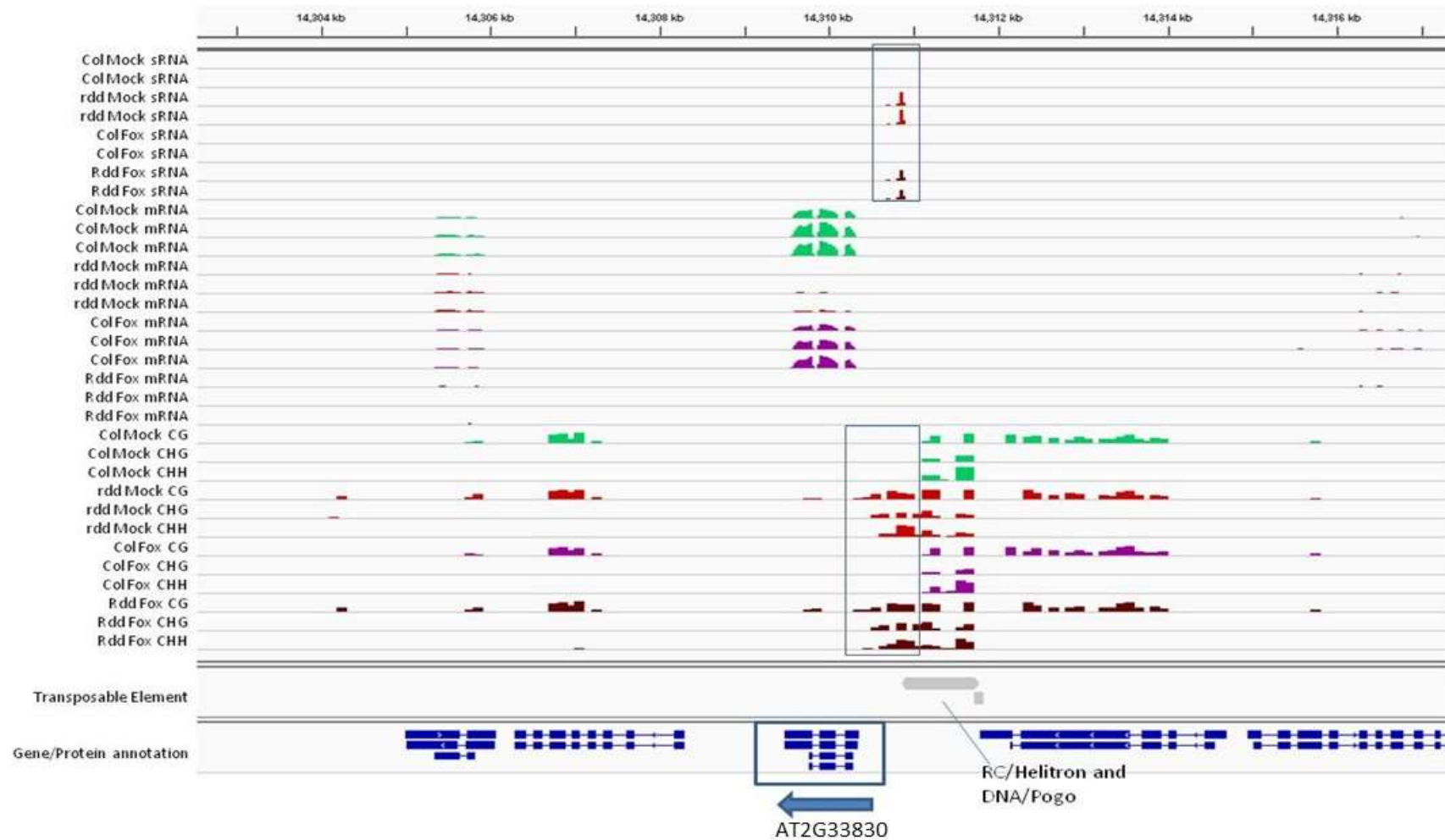
Figure 6.2: Heatmap analysis of representative DE genes with DA sRNA and DMR in the Col-0 mock vs *rdd* mock comparison

All genes had DA sRNA in the 3kb upstream region. Increasing colour from orange (log2 fold change value of 2) to red (log2 fold change value of 6) indicate downregulation in *rdd* compared to Col-0. Increasing colour from yellow to green indicates upregulation in *rdd* with the log2 fold change value from 1 to 6. The two columns in the sRNA cells indicate two different sRNA clusters with different log 2 fold changes within the 3kb flanking upstream of the gene. Similarly, two columns in the DMR cells indicate two different DMRs with different log 2 fold changes. The multiple rows per gene indicate different sRNA clusters targeting different DMR regions. White space indicates that no changes were found for this region.

Gene	DA sRNA		DE mRNA	DMR CG		DMR CHG	DMR CHH
AT1G34245	Orange	Green	Orange	Green	Green	Green	Green
AT2G33830	Green	Orange	Orange	Green	Red	Yellow	Orange
AT2G43290	Green	Orange	Orange	Green	Green	Green	Green
AT3G27940	Green	Orange	Orange	Green	Yellow		
	Orange	Orange	Orange		Yellow		
AT4G01350	Green	Orange	Orange	Green		Green	Green
AT4G04570	Orange	Green	Red	Green	Green	Orange	Orange
AT5G35732	Orange	Green	Orange	Green		Green	Green
	Green	Orange	Orange	Green	Green		
AT5G38280	Green	Orange	Red	Green	Yellow		
	Yellow	Orange	Red	Green			
	Orange	Orange	Red	Green	Orange		
AT1G58602	Green	Orange	Orange			Orange	Orange
	Green	Orange	Orange		Orange		
	Green	Orange	Orange	Green	Green		
AT1G51470	Orange	Yellow	Orange			Orange	Orange
AT2G04795	Orange	Green	Orange			Green	Green
AT2G14900	Orange		Orange			Orange	Orange
	Green	Orange	Orange	Green	Orange		
AT2G34655	Green		Orange			Green	Green
AT3G60520	Green	Orange	Orange			Green	Green
AT5G25120	Green	Green	Orange	Green	Yellow		
	Orange	Yellow	Orange	Green			

Figure 6.3: IGV visualization of DA sRNA and DMR in AT2G33830 (dormancy associated gene) downregulated in *rdd*

Green = Col-0 mock, red = *rdd* Mock, purple = Col-0 Fox, brown = *rdd* Fox.



6.1.6 Fusarium induced gene expression changes show weak correlation with sRNA and DNA methylation changes

As shown in Table 6.5, only eight DE genes in the Col-0 mock vs Col-0 Fusarium comparison were found to contain DA sRNA and DMR. A much larger number (53) of DE genes in the *rdd* mock vs *rdd* Fusarium comparison were associated with DA sRNA and DMR. A heatmap was generated to examine the correlation among gene expression, sRNA abundance and DNA methylation for these Fusarium responsive DE genes (Supplementary Table 6.3). The heatmap shows that these Fusarium responsive DE genes are different to the *rdd* affected DE genes in several aspects. Firstly, the overall fold change difference in the mock vs Fusarium comparison, as illustrated by the intensity of colour shading in Supplementary Table 6.3, was much smaller than in the Col-0 vs *rdd* comparison. This made it harder to find a relationship between sRNA, mRNA and DNA methylation changes. Secondly, changes in DNA methylation occur in just one cytosine context, most in CHH with a few in the CG or CHG context. Lastly, unlike in the Col-0 vs *rdd* comparison where the DE genes were all downregulated in *rdd*, the Fusarium responsive DE genes were either up or downregulated following Fusarium infection.

Unlike the Col-0 vs *rdd* comparison, no obvious correlation between sRNA and DNA methylation changes and gene expression changes could be seen using the IGV visualization browser. Nevertheless, some minor correlations could be observed in a number of Fusarium responsive genes, as exemplified by the two genes shown in Figure 6.4 and Figure 6.5. AT1G52000 is a mannose-binding lectin protein involved in salt stress, which showed a higher level of mRNA reads in Fusarium infected than uninfected Col-0 plants (Figure 6.4). There was a corresponding small increase in sRNA and CHH methylation levels in the Fusarium infected plants in a small localized region upstream of the gene. Despite the small number, sRNA reads were present in both the mock and Fusarium infected samples (Figure 6.4). AT2G01340 is a nematode responsive gene which showed an increased gene expression in *rdd* upon Fusarium infection (Figure 6.5). There was a small increase in sRNA abundance in the upstream region of AT2G01340 in the Fusarium infected *rdd* plant (Figure 6.5). However, there was no corresponding change in DNA methylation at the same locus. Instead, a small decrease in DNA methylation could be seen in a different location upstream of the gene in Fusarium infected *rdd* (Figure 6.5). Taken together, these results suggested that Fusarium responsive gene expression is in general not associated with strong changes in DNA methylation, which is consistent with the previous observation in our laboratory (Le et al., 2014).

Figure 6.4: IGV visualization of DA sRNA and DMR in AT1G52000 (mannose-binding lectin protein involved in salt stress gene) upregulated in Col-0 upon Fusarium infection

Green = Col-0 mock, red = Col-0 Fox

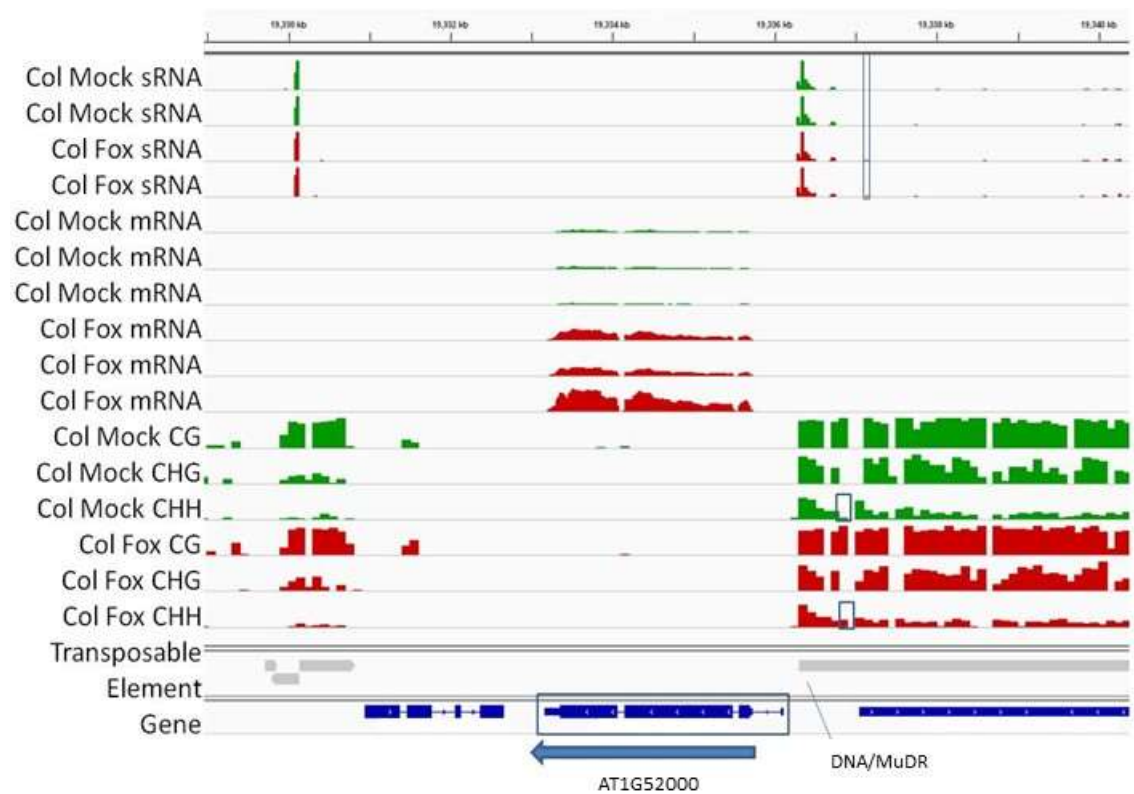
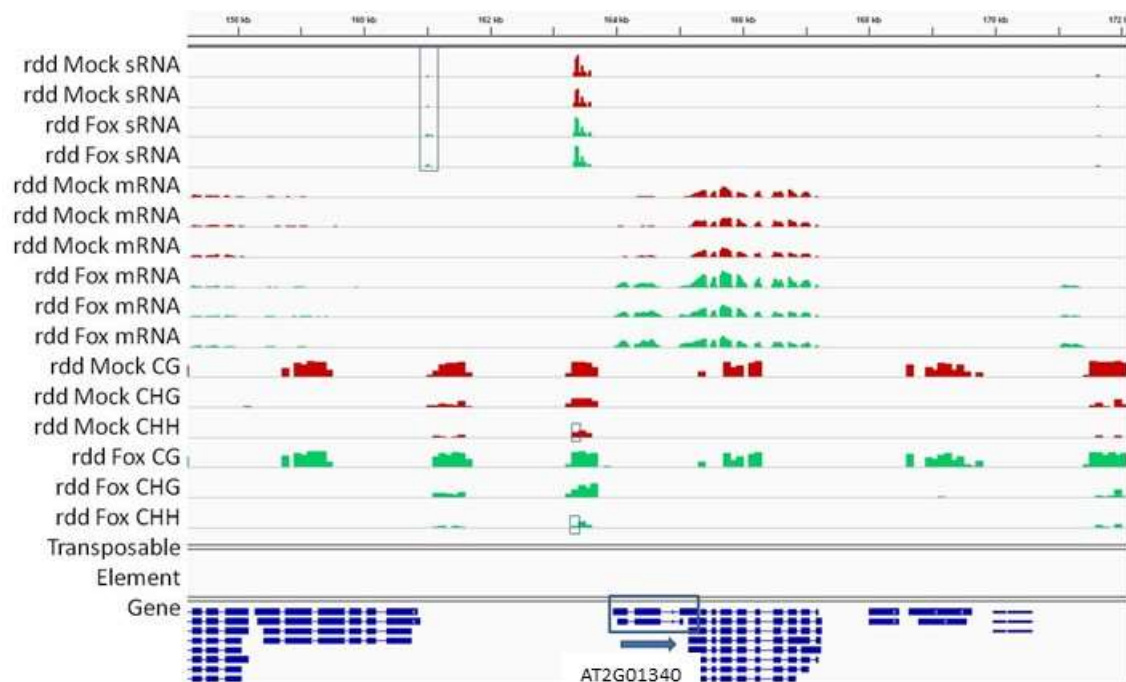


Figure 6.5: IGV visualization of DA sRNA and DMR in AT2G01340 (nematode responsive gene) upregulated in *rdd* upon Fusarium infection

Red = *rdd* Mock, green = *rdd* Fox.



6.1.7 Expression pattern and DMRs of DE genes determined by whole genome sequencing can be validated

A number of the DE genes identified in this thesis using mRNA-sequencing were analysed by RT-PCR both by colleagues in the lab (Le et al., 2014) and myself (Supplementary Figures 6.6-6.9), which verified their expression pattern. Differential methylation for the DMR regions for some of the DE genes in the Col-0 and *rdd* comparison could be positively verified by comparing with a published online whole genome DNA methylation dataset (Stroud et al., 2013). It would be ideal to verify the DMRs using targeted manual bisulfite sequencing of the same plant materials used in the mRNA- and sRNA-sequencing experiments, as those used by Stroud and colleagues (2013) were from a different *rdd* mutant. Furthermore, there was no published data on Fusarium induced DNA methylation changes, making experimental validation particularly important for the DMRs in Fusarium infected samples. Unfortunately owing to time limitation I was not able to perform manual bisulfite sequencing.

6.2 Discussion

In this chapter I analysed the genome-wide bisulfite sequencing data of 3 DPI Col-0 and *rdd* samples in order to examine the relationship between gene expression, sRNA abundance and DNA methylation in the DNA demethylase mutant and upon *Fusarium* infection. A large number of differentially methylated regions (DMRs) were identified within the gene body or 3 kb flanking sequences of thousands of associated genes. These results provide the first insight into the genome-wide DNA methylation landscape following *Fusarium oxysporum* infection in both wild type Col-0 and *rdd*. The number of DMRs for the Col-0 vs *rdd* comparison is significantly larger than those by Lister et al (2008) and Penterman et al (2007). This is probably due to the different tissue (whole plant tissues here vs flower tissues by Lister et al) or technology (whole genome bisulfite sequencing here vs genome-tilling microarrays by Penterman et al) used in the different studies.

The sequence context distribution of DMRs is different between the Col-0 vs *rdd* comparison and the mock vs *Fusarium* comparisons. The CG context is most represented in the DMRs of Col-0 vs *rdd* comparison, with CG DMRs being 3 times more frequent than the CHG and CHH DMRs. For the mock vs *Fusarium* comparisons, which have much smaller numbers of gene associated DMRs, the CHH DMRs are the most abundant. CHH methylation is generated and maintained primarily by siRNA-induced RdDM in plants. The high level representation of CHH DMRs in *Fusarium* infected plants therefore suggests that dynamic DNA methylation changes in *Arabidopsis* during pathogen infection is mediated mainly by the RdDM pathway.

To examine if the DMRs are associated with gene expression changes, I overlapped the DMR associated genes with the DE genes identified by mRNA-sequencing in Chapter 5. The DE genes from the Col-0 vs *rdd* comparison showed a high frequency of association with DMRs, with 29 of the 49 DE genes containing DMRs, indicating that the differential expression of genes in *rdd* could largely be accounted for by changes in DNA methylation. The proportion of *Fusarium* induced DE genes that have DMRs is very small, indicating that most of the *Fusarium* responsive genes are not regulated by DNA methylation changes, and that the majority of the *Fusarium* induced DMRs do not cause changes in gene expression, at least not at the 3 DPI time point. It is worth noting that a previous study (Yu et al., 2013) on DNA methylation changes induced by the bacterial flagellin-derived peptide flg22 showed that DNA hypomethylation occurred in response to flg22 treatment, but only transiently at the early time point of 6 to 9 hours. It is therefore possible that our bisulfite sequencing analysis did not detect the DMRs of early time

points that might be important for *Fusarium* responsive gene expression. Nevertheless, 65 DE genes in the Col-0 mock vs Col-0 *Fusarium* comparison, and 273 DE genes in the *rdd* mock vs *rdd* *Fusarium* comparison were associated with DMRs, indicating that a subset of the DE genes is regulated by differential DNA methylation at 3 DPI.

A main objective of this thesis was to understand if sRNA-directed RdDM and DNA demethylases acts collectively to regulate gene expression. To examine this, I overlapped the list of DE genes derived from the mRNA-sequencing with those of the DA sRNA and DMR associated genes, creating the list of genes with differential expression as well as DA sRNA and DMRs. The DE genes from the Col-0 and *rdd* comparison showed a strong association with DA sRNA and DMRs, with 25 of the 49 DE genes having both DA sRNA and DMRs that usually occurred in all three cytosine contexts. These genes were mostly downregulated in *rdd* compared to Col-0, and tended to show increased sRNA abundance and DNA methylation level in a localized area of the upstream region, often near the transcription start site of the genes. This indicates that the repressed expression of these genes in *rdd* was caused by increased RdDM against the gene promoters.

For 15 of the *rdd* affected and DA sRNA and DMR associated genes examined in detail using the IGV visualization browser, 13 showed the existence of overlapping sRNA and DNA methylation in *rdd* but absence of both in Col-0 at the same locus. These new sRNA and DNA methylation loci in *rdd* are usually 100 bp to 1 kb in size and often immediately adjacent to an existing methylated region in Col-0 that overlap with a TE sequence. These results suggest that TE sequences in gene promoters not only attract methylation due to them being the target of RdDM, but also have the potential to spread DNA methylation via RdDM to adjacent regions, affecting gene expression. These results also suggest that DNA demethylases target the promoter TE sequences to prevent the spread of sRNA-directed DNA methylation ensuring active gene expression. Thus, my results not only support the view that DNA demethylases target promoter TEs to regulate gene expression (Le et al., 2014), but also provide a plausible model for DNA demethylase mediated gene regulation: DNA demethylases interact with siRNA-directed RdDM to maintain active gene expression of genes with promoter TEs (discussed further in Chapter 7, Figure 7.1).

TEs are targeted by epigenetic silencing mechanisms including RdDM, and recent studies have suggested that TEs near genes play a role in regulating the expression of neighboring genes. Currently it is thought that the chromatin state of the neighboring genes can be influenced by heterochromatic silencing of TEs that can spread past TE borders. Indeed this was found to be

true in maize for lowly expressed genes located 800 – 1,200 bp from a subset of moderately to highly abundant retrotransposon TEs (Eichten et al., 2012). Interestingly, the locations of the DA sRNA and DMR adjacent TEs in some of the downregulated genes in *rrd* from my analysis are consistent with this distance in maize, suggesting that these TEs in Arabidopsis, like those in maize, can repress the expression of neighboring genes by initiating spread of heterochromatic silencing. My results indicated that DNA demethylases can target these TE regions to prevent sRNA and methylation spread, suggesting that TE-mediated gene expression is a dynamic process involving both RdDM or heterochromatic silencing and DNA demethylation.

In contrast to the Col-0 vs *rrd* comparison, only eight DE genes in Fusarium infected Col-0, representing a very small proportion of the total 1,381 DE genes, showed an association with both DA sRNA and DMRs. Furthermore, the degree of differential gene expression, sRNA abundance and DNA methylation was relatively low compared to the DA sRNA and DMR associated genes in the Col-0 vs *rrd* comparison. The lack of strong DNA methylation changes with Fusarium induced differential gene expression in Col-0 was also observed previously (Le et al., 2014). Thus, it remains unclear if sRNAs and DNA methylation play an important role in Fusarium responsive gene expression. Interestingly, recent studies on pathogen induced defence priming is associated with changes in histone modification but not DNA methylation in defence gene promoters, although RdDM factors are involved in this priming (Lopez et al., 2011, Luna et al., 2012). If siRNAs are involved in the histone modification associated with defence priming remained unknown.

Unlike the Col-0 background, 53 DE genes in the *rrd* mock vs *rrd* Fusarium comparisons were found to have DA sRNA and DMR. None of these 53 genes were shared with the Fusarium responsive DE genes in Col-0. This result implied that DNA demethylases are involved in stabilizing the expression of these genes during Fusarium infection. These Fusarium responsive, DA sRNA and DMR associated genes are enriched for plant defence functions, particularly in the stress, signalling and RNA pathways. The RNA pathway incorporates TEs supporting the view that TE-mediated gene expression may be targeted by DNA demethylases under Fusarium infection to regulate stress responsive genes (Le et al., 2014).

6.3 Conclusions

In this chapter I examined the genome-wide methylation status of Col-0 and *rdd* with or without *Fusarium* infection at 3 DPI, and identified a large number of genes with differentially methylated regions (DMRs) in the gene body or 3 kb flanking sequences. Comparison with genes containing differentially abundant (DA) sRNAs and those with differential expression (DE) identified from sRNA sequencing and mRNA sequencing in the last two chapters resulted in a list of differentially expressed genes with both DA sRNA and DMRs. These list of genes allowed me to examine the relationship between DA sRNA, DMRs and DE gene expression.

DE genes in the Col-0 vs *rdd* comparison showed a strong relationship with DA sRNA and DMRs. A significant proportion of these genes showed downregulation in *rdd* and localized increase in sRNA abundance and DNA methylation levels, indicating that these genes are regulated by siRNA-directed DNA methylation in conjunction with DNA demethylases. I found that the increased sRNA and DNA methylation tended to occur at proximal promoter regions immediately adjacent to methylated TE sequences, suggesting that TEs can initiate spread of DNA methylation via siRNA-directed RdDM leading to repressed gene expression and that DNA demethylases target the TE region to prevent methylation spread in wild type plants. These results support the view that DNA demethylases target near gene TEs to positively regulate gene expression, and provide evidence indicating the involvement of siRNA-directed RdDM in this DNA demethylase mediated gene regulation.

DE genes in the Col-0 mock vs Col-0 *Fusarium* comparison showed a weak association with DA sRNAs and DMRs, suggesting that *Fusarium* responsive gene expression is usually independent of siRNAs and DNA methylation changes. *Fusarium* infection of *rdd* resulted in many more DE genes than that of Col-0, suggesting that DNA demethylases have a role in plant response to *Fusarium*. How DNA demethylases are involved in this process requires further investigation.

Chapter 7: Conclusions and future prospects

This thesis was designed to investigate if RNA directed DNA methylation (RdDM) interacts with DNA demethylases to coordinately regulate plant defence against the fungal pathogen *Fusarium oxysporum* by focusing on RdDM associated 24 nt siRNAs. Three major questions were addressed, including i) does 24 nt siRNA accumulation change upon *Fusarium* infection or in the triple DNA demethylase mutant *rdd*; ii) do *Fusarium* induced or *rdd* affected changes in 24 nt siRNA accumulation result in changes in gene expression, particularly the expression of stress response genes; and iii) are siRNA associated gene expression changes associated with changes in DNA methylation? In addition to these major questions, I also attempted to address two additional questions: iv) which DNA demethylase genes are required for *Fusarium* resistance; v) do DNA demethylases play a broader role in biotic stress response?

Genome-wide distribution of sRNA and DNA methylation has been examined in various DNA methylation related mutants including *rdd* (although the *rdd* mutant has a slightly different genetic background to the one used in our study) (Stroud et al., 2013, Lister et al., 2008), however little attention has been paid to the relationship between sRNA, DNA methylation and gene expression. While both RdDM and DNA demethylation have been implicated in plant defence against bacterial and fungal pathogens (Agorio and Vera, 2007, Baldrich et al., 2014, Yu et al., 2013, Downen et al., 2012), including *Fusarium oxysporum*, how these mechanisms regulate plant defence remains poorly understood. In particular, genome-wide sRNA distribution and DNA methylation in response to *Fusarium* infection were not investigated previously. Furthermore, our previous study (Le et al., 2014) suggested an interaction between RdDM and DNA demethylases in the regulation of stress response genes, how they interact is also unclear.

I started this project by re-evaluating the phenotypes of DNA demethylase mutants under normal conditions and upon *Fusarium* infection. This showed that the *rdd* mutant, and the single DNA demethylase mutants *ros1*, *dml2*, and *dml3*, had no clear phenotypic difference to the wild type Col-0 plants under normal growth conditions. However, upon *Fusarium* infection, these DNA demethylase mutants all displayed enhanced disease phenotypes compared to Col-0. The *rdd* mutant showed the strongest disease phenotypes, followed by *ros1*, with *dml2* and *dml3* showing relatively weak phenotypes, indicating a functional redundancy of these DNA demethylase genes. My results therefore confirmed our previous observation on *rdd* (Le et al., 2014) and extended this observation by providing data on the functional redundancy of DNA demethylase genes. These results also suggested that DNA demethylases are not important for plant development but play a role mainly in plant stress response.

In addition to *Fusarium oxysporum*, I assayed the phenotypes of *rdd* in response to virus infection, and showed that *rdd* appeared to be more resistant than Col-0 to Cucumber mosaic virus (CMV) and Turnip mosaic virus (TuMV) viruses. How the DNA demethylase mutant became more resistant to the viruses remains an interesting question and warrants further investigation in the future. One possibility is that increased transcriptional silencing of TEs and repetitive sequences, which is expected to occur in the genome of *rdd*, may reduce the accumulation of RNA transcript from these elements that would be targeted by posttranscriptional gene silencing (PTGS). Consequently, PTGS machinery would become more available for silencing viral RNAs, resulting in enhanced virus resistance. However, further studies, such as comparing viral siRNA accumulation and PTGS efficiency of transgenes between *rdd* and Col-0, are needed to test this hypothesis. Nevertheless, my results with virus infection suggested that DNA demethylases play a broader role in biotic stress response in Arabidopsis.

A major experimental approach in my study was high throughput sRNA and mRNA sequencing. In order to determine suitable time points for genome-wide sRNA sequencing analysis, I first examined sRNA accumulation in mock treated and *Fusarium* infected Col-0 and *rdd* plants using small scale northern blot hybridization. This preliminary experiment indicated that RdDM associated siRNAs, but not miRNAs and tasiRNAs, were differentially accumulated between Col-0 and *rdd* in response to *Fusarium* infection, particularly after 3 DPI when *Fusarium* colonization become clearly established based on my quantitative assay. The difference in siRNA accumulation became stronger at 6 DPI, however after 6 DPI the chlorosis symptoms of *rdd* usually become very severe preventing preparation of good quality RNA. Therefore, plant tissues of 1, 3 and 6 DPI were selected for sRNA sequencing analysis, and high quality data was obtained.

A large number of differentially abundant (DA) sRNA were identified between Col-0 and *rdd* and between mock treated and *Fusarium* infected Col-0 or *rdd* plants that are located predominantly in protein coding genes and their flanking sequences. The large number of DA sRNA between Col-0 and *rdd* indicates that DNA demethylases are involved in 24 nt siRNA accumulation implicating a direct functional interaction between RdDM and DNA demethylases. Furthermore, *Fusarium* infection affected sRNA distribution in both Col-0 and *rdd*, particularly in *rdd*, where RdDM associated siRNA tend to increase upon *Fusarium* infection, suggesting that 24 nt siRNA accumulation is responsive to *Fusarium* infection, and DNA demethylases are involved in this *Fusarium* induced response. A number of miRNAs also showed differential accumulation between Col-0 and *rdd* or between mock treated and

Fusarium infected plants. The potential function of these DA miRNAs in Fusarium response was not investigated due to time limitation. While this thesis focused on DA sRNA as a whole, further analysis could be performed specifically on the up or downregulated sRNA levels.

The DA sRNAs were found to be associated with thousands of protein coding genes, suggesting that they may play a role in the regulation of these genes. To examine this, mRNA sequencing was performed on the 3 DPI sample. Interestingly, in contrast to the large number of DA sRNA associated genes (~10,000) between Col-0 and *rdd*, only a small number of differentially expressed (DE) genes (49) was identified by mRNA sequencing, indicating that the majority of the *rdd* affected DA sRNAs do not play a role in the regulation of overlapping or adjacent genes. Nevertheless, a large proportion of the 49 *rdd* affected DE genes was associated with DA sRNAs, suggesting that they are regulated by both DNA demethylases and RdDM. Consistent with the result from a previous microarray analysis (Le et al., 2014), these genes are enriched for stress response function.

Unlike the Col-0 and *rdd* comparison, over a thousand genes were found to be differentially expressed in Col-0 upon Fusarium infection, and this number was even bigger in the *rdd* backgrounds (over three thousands), indicating that Fusarium infection has a major impact on genome-wide gene expression. However, less than 10 % of these DE genes were associated with DA sRNAs, indicating that most of these Fusarium responsive genes are not regulated by siRNAs or RdDM. Interestingly, those Fusarium responsive genes in both the Col-0 and *rdd* backgrounds, that are associated with DA sRNAs, are enriched for stress response functions, suggesting that for these genes sRNA or RdDM play a role in Fusarium responsive gene expression. The significantly larger number of Fusarium responsive DE genes in the *rdd* background (3,392 genes) than in the Col-0 background (1,381 genes) implied that DNA demethylases are involved in the siRNA-mediated regulation of Fusarium-responsive expression.

To further understand the relationship between RdDM and DNA demethylases and their function in Fusarium response, genome-wide bisulfite sequencing analysis was performed on the 3 DPI samples to examine genome-wide DNA methylation variations. A large number of differentially methylated regions (DMRs) was identified, with DMRs in the CG context being the dominant ones for Col-0 vs *rdd* comparison, and CHH DMRs being the dominant for the mock vs Fusarium comparisons. This indicated that Fusarium infection induces methylation changes mainly in the CHH context, indicative of RdDM. These DMRs are associated with over ten thousand genes for the Col-0 vs *rdd* comparison, and over a thousand genes for the mock vs

Fusarium comparisons. However, only a small fraction of these DMRs was associated with DE genes or DA sRNAs, indicating that the majority of the DMRs does not result in differential gene expression or was not caused by differential sRNA accumulation.

Nevertheless, a large proportion of the DE genes in the Col-0 vs *rdd* comparison were associated with both DA sRNAs and DMRs, which are often located at the same promoter region of the DE genes. This suggested that these *rdd* affected genes, enriched for stress response functions, are regulated by siRNA-directed DNA methylation or RdDM. A careful examination of a subset of these DA sRNA and DMR associated DE genes revealed two major features: i) siRNAs and DNA methylation exist near transcription start sites (TSS) in the *rdd* background but are absent in Col-0; and ii) the new methylation in *rdd* occurred in the direction towards the TSS, and was immediately adjacent to the end of TEs that are methylated in Col-0 as well as in *rdd*. These features suggest that TEs in the promoters are the target of RdDM and have the potential to induce spread of RdDM to adjacent sequences near TSS resulting in the repression of gene expression. DNA demethylases are required to counter this effect ensuring active gene expression. This model, illustrated in Figure 7.1, supports the view that DNA demethylases interact with RdDM and target promoter TEs to positively regulate the expression of a subset of stress response genes (Le et al., 2014). My data confirmed that siRNAs are involved in this regulation. One scenario for the interaction of RdDM with DNA demethylases is that siRNAs, or the resulting DNA methylation, are involved in the recruitment of DNA demethylases to promoter TE regions, and DNA demethylases in turn prevent the spread of DNA methylation (and the spread of siRNAs as increased DNA methylation can lead to increased siRNAs).

The genes downregulated in *rdd* tend to be induced by Fusarium infection (Le et al., 2014), suggesting that DNA demethylases are involved in Fusarium responsive gene expression. However, the previous study in our laboratory, using small scale manual bisulfite sequencing, detected only subtle changes in DNA methylation upon Fusarium infection in the promoter TE region of the *rdd* affected, Fusarium inducible genes (Le et al., 2014). Thus, the differential gene expression between mock treated and Fusarium infected plants could not be accounted for by differential DNA methylation in DNA demethylase targeted TE regions. In my thesis I showed a similar result: very few of the Fusarium responsive DE genes are associated with DMRs and DA sRNAs, suggesting that the Fusarium responsive expression pattern is controlled by many factors rather than just DNA methylation. Interestingly, a study in maize showed that no consistent DNA methylation changes occurs in maize genome in response to various stresses (Eichten et al., 2014). So does DNA demethylases play a role in stress responsive gene

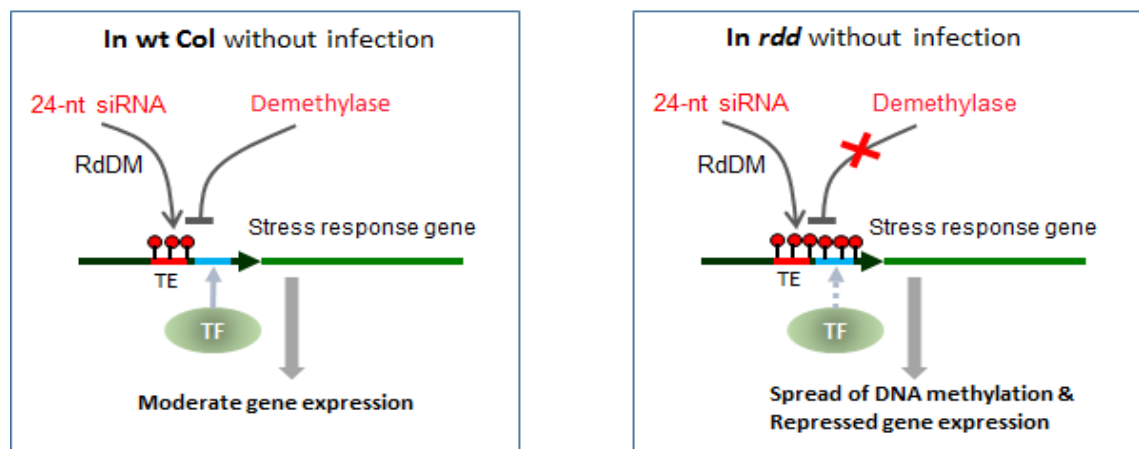
expression and if so how? The previous study in our laboratory showed that although the expression of several stress response genes with promoter TEs is downregulated in *rdd* compared to Col-0 under normal growth conditions due to increased DNA methylation in promoter TE sequences, these genes still showed Fusarium inducible expression patterns in *rdd* (although the final level of expression is lower than in Col-0) (Le et al., 2014). These observations raise the possibility that the stress inducibility of these genes is determined by the inherent property of transcription factor (TF)-binding sequence elements that may exist inside or near the promoter TE sequences, and the role of DNA demethylases is to keep a minimum methylation level allowing for TF binding. Thus, DNA demethylases may function to maintain a below threshold level of DNA methylation at stress responsive cis-elements near promoter TEs before and during stresses allowing for inducible expression to occur. However, there may be no direct role in the stress inducible expression pattern of the genes (Figure 7.1).

In this study, all genome-wide analysis was based on the use of the triple DNA demethylase mutant *rdd* in comparison to Col-0. It was previously reported that *ROS1*, *DML2* and *DML3* are expressed in vegetative tissues while the remaining DNA demethylase gene, *DME*, was expressed only in floral tissue (Choi et al., 2002), and therefore *ROS1*, *DML2* and *DML3* account for all DNA demethylase activities in somatic tissues of Arabidopsis. However, our previous microarray data (Le et al., 2014) showed that *DME* is expressed in vegetative tissues of Arabidopsis at a level lower than *ROS1* but higher than *DML2* and *DML3* (Figure 7.2a). The mRNA-sequencing data in this study also indicated a significant level of *DME* expression in vegetative tissues (Figure 7.2b). Significant expression of *DME* in vegetative tissues was also detected by RT-PCR analysis (Ortega-Galisteo et al., 2008). These observations raised the possibility that the *rdd* mutant still possesses some DNA demethylase activity in vegetative tissues due to the presence of *DME*, which could affect the interpretation of our data, and that of the data in previous reports involving the use of *rdd*. One way to address this potential issue is to obtain a quadruple DNA demethylase mutant, namely a *ros1 dml2 dml3 dme* mutant in which all four DNA demethylases are mutated. Unfortunately, genetic knock-out of *dme* would affect normal seed development therefore a quadruple mutant would be unviable.

Figure 7.1: Infection model explaining how RdDM interacts with DNA demethylases to regulate the expression of stress response genes.

Under normal growth conditions (upper panels), the promoter TE sequence is targeted by 24 nt siRNA-directed RdDM, resulting in its methylation. In wild type plants DNA demethylases are recruited by 24 nt siRNAs or DNA methylation to the TE sequence, which in turn function to prevent spread of DNA methylation to the adjacent transcription factor (TF)-binding site, allowing good TF binding and a moderate level of gene expression. In *rdd* mutant plants, DNA demethylase activity is lost, resulting in the spread of DNA methylation to TF-binding sites. This leads to weak TF binding (indicated by the dashed line) and hence repressed gene expression. Upon *Fusarium* infection, a stress factor (SF) interacts with TF, causing strong induction of gene expression in wild type plants due to strong TF-binding but weak induction of gene expression due to weak TF binding. Note that the interaction between SF and TF is independent of DNA methylation at the TF-binding site, which could explain why *Fusarium* responsive DE genes are usually not associated with DMRs.

Under normal growth condition



Upon *Fusarium* infection

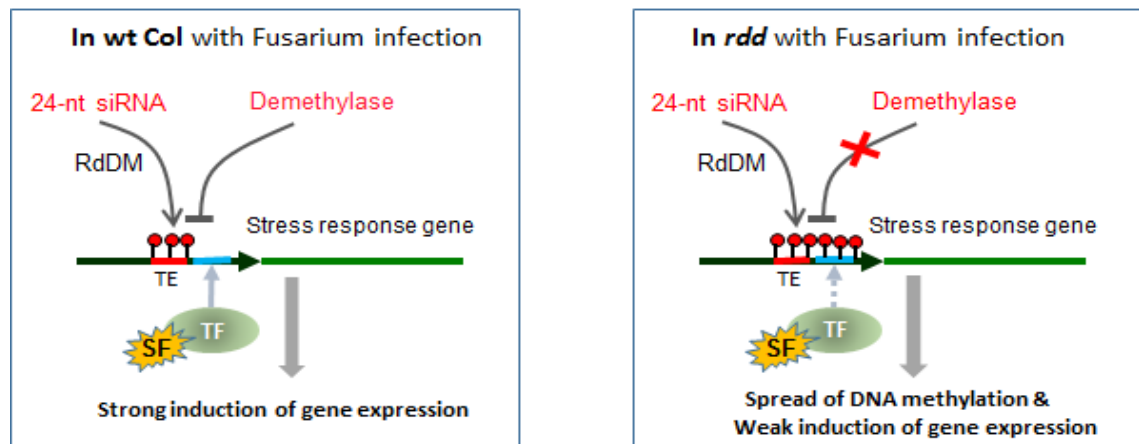


Figure 7.2: DME is expressed in vegetative tissues A) Expression levels of the four demethylases genes in Col-0 based on previous microarray data (Le et al., 2014). B) Expression levels of the four DNA demethylase gene in Col-0 based on mRNA sequencing data of this study.

Measured across three biological replicates and standard deviation recorded.

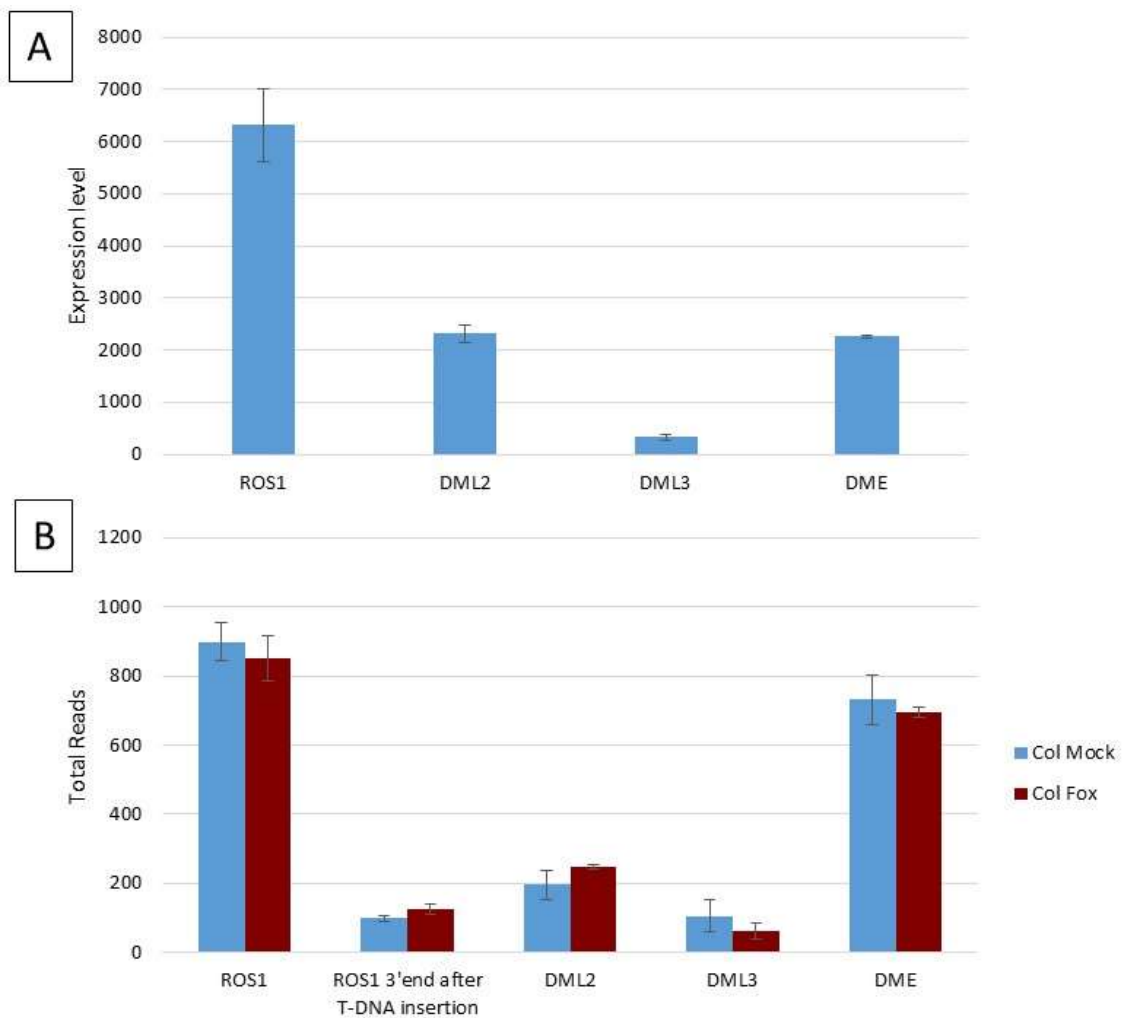


Figure 7.3: Silencing of *DME* in *rdd* using an hairpin RNA transgene driven by the *Arabidopsis* rubisco small subunit promoter (*rbcS-P*) A) Schematic of the hairpin RNA gene silencing construct. B) Transgenic hairpin RNA lines *hpDME-6* and *hpDME-7* show strong downregulation of *DME* mRNA compared to *Col-0* and *rdd*, which correlates with the expression of the *hpDME* RNA. Line *hpDME-5* show weak downregulation of *DME* mRNA, which correlates with a lack of clear expression of the *hpDME* RNA.

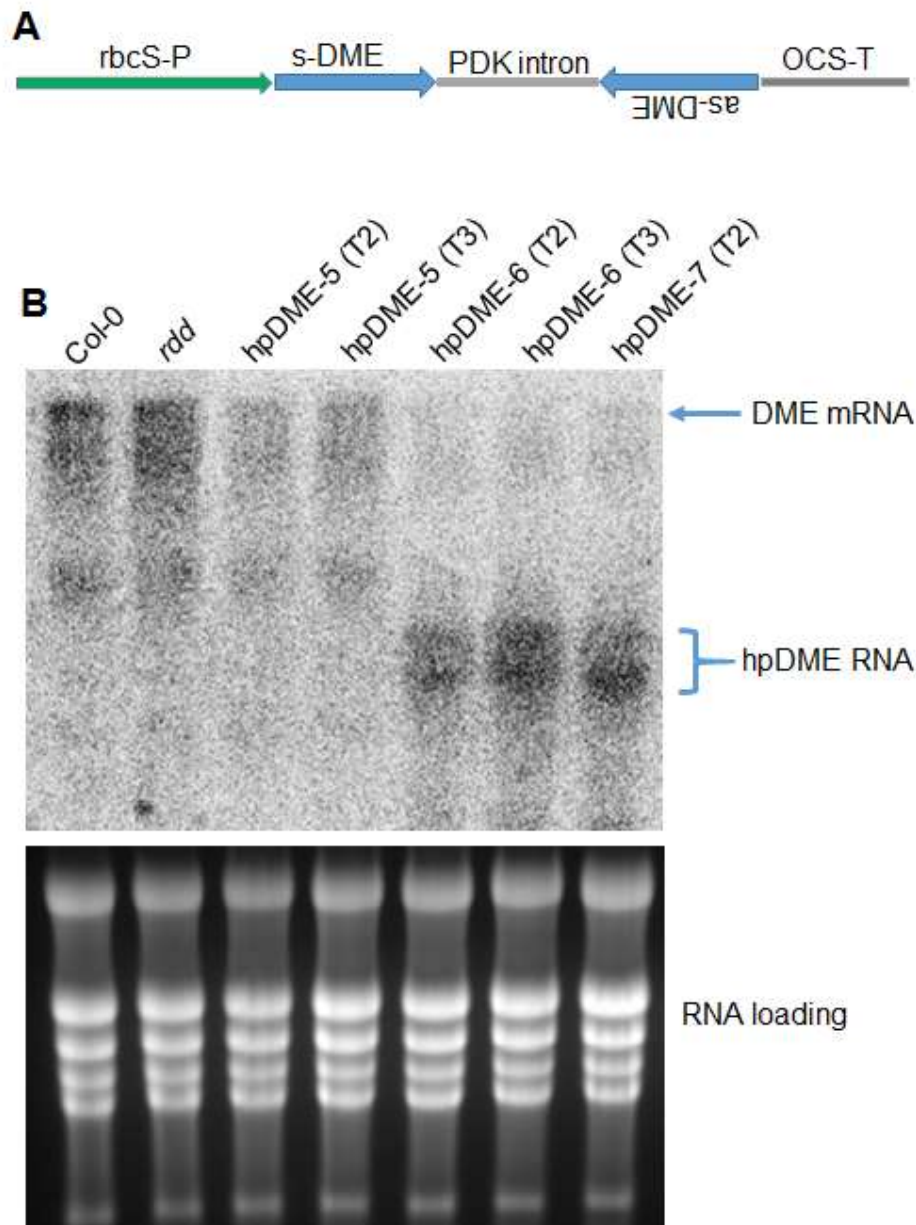
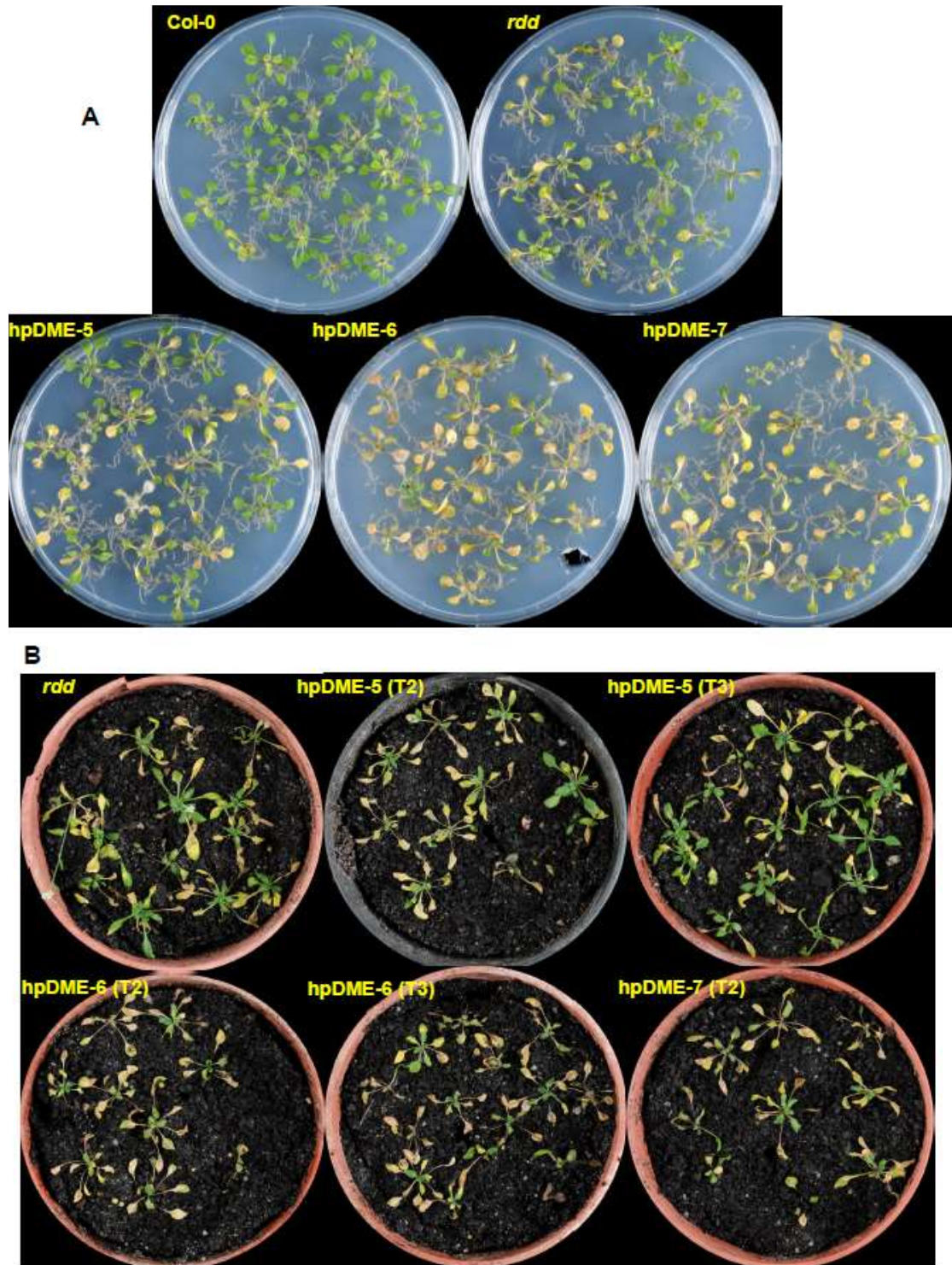
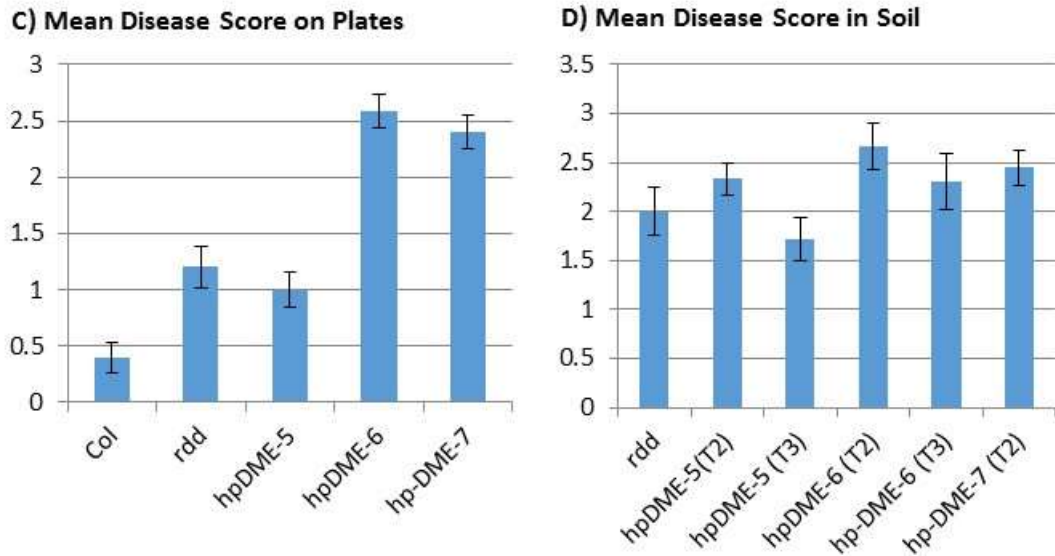


Figure 7.4: Silencing of *DME* in *rdd* using hairpin RNA transgene appears to increase susceptibility to *Fusarium* infection. A) Infected on sucrose-free MS plate. Picture taken at 11 dpi. B) Infected on sucrose-free MS plate for 3 days, then transferred to soil. Picture taken at 12 dpi. Mean disease score of chlorotic leaves on C) plates and D) soil.

Disease symptom scoring (all plants had black roots): 0 = no chlorotic leaves, 1 = 1-3 leaves showing chlorosis, 2 = 4-6 leaves showing chlorosis, 3 = all leaves showing chlorosis, 4 = dead plant. Standard error bars are shown.





In an attempt to examine if DME might contribute to DNA demethylase activity hence Fusarium resistance in vegetative tissues in Arabidopsis, a hairpin RNA transgene construct (hpDME) was prepared using the Arabidopsis rubisco small subunit promoter (*rbcS-P*) (Figure 7.3a). With this promoter the DME hpRNA is expected to be expressed only in the green tissues, preventing silencing of *DME* in developing seed. I then transformed this construct to the *rdd* mutant, resulting in transgenic lines in which DME expression was strongly downregulated (e.g., Figure 7.3b, hpDME lines 6 and 7). Consistent with the *rbcS* promoter functional only in the green tissues, the transgenic plants showed no abnormal seed development giving normal seed production.

To examine if the silencing of *DME* in *rdd* might increase susceptibility to *Fusarium oxysporum*, I infected the T2 and/or T3 progeny of the three hpDME lines shown in Figure 7.3 with Fusarium. In both plate and soil infection, the hpDME-6 and hpDME-7 plants appeared to be more susceptible to Fusarium infection than the untransformed *rdd*, developing more severe chlorotic symptoms (Figure 7.4). The hpDME-5 plants, which showed no strong DME silencing, displayed a similar disease phenotype to *rdd* (Figure 7.4). While further detailed analysis is needed to confirm this observation, particularly in soil, the results so far suggest that DME does contribute to DNA demethylase activity in vegetative tissues and plays a redundant role in Fusarium resistance. This finding, if confirmed, has important implications in understanding the role of DNA demethylases in regulating genome-wide methylation and gene expression. For instance, previous studies detected only limited genome-wide changes in DNA methylation and gene expression in *rdd*, which has led to the conclusion that DNA demethylases have a limited impact on genome-wide DNA demethylation (Lister et al., 2008). However, this limited changes

in *rdd* could be due to the residual DNA demethylase activity from DME. The hpDME lines generated in this thesis, which carries *ros1 dm12 dm13* triple knock-out mutations plus strong downregulation of *DME*, would be useful to investigate if DNA demethylases have a more profound effect on genome-wide DNA demethylation and gene expression than previously thought.

Chapter 8: Bibliography

- AGIUS, F., KAPOOR, A. & ZHU, J.-K. 2006. Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 11796-11801.
- AGORIO, A. & VERA, P. 2007. ARGONAUTE4 is required for resistance to *Pseudomonas syringae* in Arabidopsis. *Plant Cell*, 19, 3778-3790.
- AHMED, I., SARAZIN, A., BOWLER, C., COLOT, V. & QUESNEVILLE, H. 2011. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. *Nucleic Acids Research*, 39, 6919-6931.
- ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biology*, 11.
- ANDERS, S., MCCARTHY, D. J., CHEN, Y. S., OKONIEWSKI, M., SMYTH, G. K., HUBER, W. & ROBINSON, M. D. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, 8, 1765-1786.
- AUSIN, I., GREENBERG, M. V. C., SIMANSHU, D. K., HALE, C. J., VASHISHT, A. A., SIMON, S. A., LEE, T.-F., FENG, S., ESPANOLA, S. D., MEYERS, B. C., WOHLSCHLEGEL, J. A., PATEL, D. J. & JACOBSEN, S. E. 2012. INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 8374-8381.
- AYLIFFE, M., PERIYANNAN, S. K., FEECHAN, A., DRY, I., SCHUMANN, U., WANG, M. B., PRYOR, A. & LAGUDAH, E. 2013. A Simple Method for Comparing Fungal Biomass in Infected Plant Tissues. *Molecular Plant-Microbe Interactions*, 26, 658-667.
- BAEV, V., NAYDENOV, M., APOSTOLOVA, E., IVANOVA, D., DONCHEVA, S., MINKOV, I. & YAHUBYAN, G. 2010. Identification of RNA-dependent DNA-methylation regulated promoters in Arabidopsis. *Plant Physiology and Biochemistry*, 48, 393-400.
- BALDRICH, P., KAKAR, K., SIRE, C., MORENO, A. B., BERGER, A., GARCIA-CHAPA, M., LOPEZ-MOYA, J. J., RIECHMANN, J. L. & SEGUNDO, B. S. 2014. Small RNA profiling reveals regulation of Arabidopsis miR168 and heterochromatic siRNA415 in response to fungal elicitors. *Bmc Genomics*, 15.
- BARTEE, L., MALAGNAC, F. & BENDER, J. 2001. Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes & Development*, 15, 1753-1758.
- BAUBEC, T., FINKE, A., SCHEID, O. M. & PECINKA, A. 2014. Meristem-specific expression of epigenetic regulators safeguards transposon silencing in Arabidopsis. *Embo Reports*, 15, 446-452.
- BENDER, J. 2004. DNA methylation and epigenetics. *Annual Review of Plant Biology*, 55, 41-68.
- BERNATAVICHUTE, Y. V., ZHANG, X., COKUS, S., PELLEGRINI, M. & JACOBSEN, S. E. 2008. Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in Arabidopsis thaliana. *Plos One*, 3.
- BISHOP, C. D. & COOPER, R. M. 1983. AN ULTRASTRUCTURAL-STUDY OF VASCULAR COLONIZATION IN 3 VASCULAR WILT DISEASES .1. COLONIZATION OF SUSCEPTIBLE CULTIVARS. *Physiological Plant Pathology*, 23, 323-343.

- BOCCARA, M., SARAZIN, A., THIEBEAULD, O., JAY, F., VOINNET, O., NAVARRO, L. & COLOT, V. 2014. The Arabidopsis miR472-RDR6 silencing pathway modulates PAMP- and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *Plos Pathogens*, 10, e1003883-e1003883.
- BUCHER, E., REINDERS, J. & MIROUZE, M. 2012. Epigenetic control of transposon transcription and mobility in Arabidopsis. *Current Opinion in Plant Biology*, 15, 503-510.
- CAO, X. F., AUFSATZ, W., ZILBERMAN, D., METTE, M. F., HUANG, M. S., MATZKE, M. & JACOBSEN, S. E. 2003. Role of the DRM and CMT3 Methyltransferases in RNA-directed DNA methylation. *Current Biology*, 13, 2212-2217.
- CAO, X. F. & JACOBSEN, S. E. 2002. Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Current Biology*, 12, 1138-1144.
- CARTHEW, R. W. & SONTHEIMER, E. J. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136, 642-655.
- CHELLAPPAN, P., XIA, J., ZHOU, X., GAO, S., ZHANG, X., COUTINO, G., VAZQUEZ, F., ZHANG, W. & JIN, H. 2010. siRNAs from miRNA sites mediate DNA methylation of target genes. *Nucleic Acids Research*, 38, 6883-6894.
- CHEN, H.-M., LI, Y.-H. & WU, S.-H. 2007. Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 3318-3323.
- CHEN, R., JIANG, N., JIANG, Q., SUN, X., WANG, Y., ZHANG, H. & HU, Z. 2014. Exploring MicroRNA-Like Small RNAs in the Filamentous Fungus *Fusarium oxysporum*. *Plos One*, 9.
- CHOI, S.-M., SONG, H.-R., HAN, S.-K., HAN, M., KIM, C.-Y., PARK, J., LEE, Y.-H., JEON, J.-S., NOH, Y.-S. & NOH, B. 2012. HDA19 is required for the repression of salicylic acid biosynthesis and salicylic acid-mediated defense responses in Arabidopsis. *Plant Journal*, 71, 135-146.
- CHOI, Y. H., GEHRING, M., JOHNSON, L., HANNON, M., HARADA, J. J., GOLDBERG, R. B., JACOBSEN, S. E. & FISCHER, R. L. 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell*, 110, 33-42.
- CLOUGH, S. J. & BENT, A. F. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of Arabidopsis thaliana. *Plant Journal*, 16, 735-743.
- COKUS, S. J., FENG, S., ZHANG, X., CHEN, Z., MERRIMAN, B., HAUDENSCHILD, C. D., PRADHAN, S., NELSON, S. F., PELLEGRINI, M. & JACOBSEN, S. E. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 452, 215-219.
- CREASEY, K. M., ZHAI, J. X., BORGES, F., VAN EX, F., REGULSKI, M., MEYERS, B. C. & MARTIENSSSEN, R. A. 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature*, 508, 411-+.
- CZYMMEK, K. J., FOGG, M., POWELL, D. H., SWEIGARD, J., PARK, S.-Y. & KANG, S. 2007. In vivo time-lapse documentation using confocal and multi-photon microscopy reveals the mechanisms of invasion into the Arabidopsis root vascular system by *Fusarium oxysporum*. *Fungal Genetics and Biology*, 44, 1011-1023.

- DENG, S. & CHUA, N. H. 2015. Inverted-Repeat RNAs Targeting FT Intronic Regions Promote FT Expression in Arabidopsis. *Plant Cell Physiol*, 56, 1667-78.
- DI PIETRO, A., MADRID, M. P., CARACUEL, Z., DELGADO-JARANA, J. & RONCERO, M. I. G. 2003. Fusarium oxysporum: exploring the molecular arsenal of a vascular wilt fungus. *Molecular Plant Pathology*, 4, 315-325.
- DIENER, A. 2012. Visualizing and Quantifying Fusarium oxysporum in the Plant Host. *Molecular Plant-Microbe Interactions*, 25, 1531-1541.
- DOWEN, R. H., PELIZZOLA, M., SCHMITZ, R. J., LISTER, R., DOWEN, J. M., NERY, J. R., DIXON, J. E. & ECKER, J. R. 2012. Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences of the United States of America*, 109, E2183-E2191.
- DRUKER, R., BRUXNER, T. J., LEHRBACH, N. J. & WHITELAW, E. 2004. Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. *Nucleic Acids Res*, 32, 5800-8.
- EAMENS, A., WANG, M.-B., SMITH, N. A. & WATERHOUSE, P. M. 2008. RNA silencing in plants: Yesterday, today, and tomorrow. *Plant Physiology*, 147, 456-468.
- EDGAR, C. I., MCGRATH, K. C., DOMBRECHT, B., MANNERS, J. M., MACLEAN, D. C., SCHENK, P. M. & KAZAN, K. 2006. Salicylic acid mediates resistance to the vascular wilt pathogen Fusarium oxysporum in the model host Arabidopsis thaliana. *Australasian Plant Pathology*, 35, 581-591.
- EICHTEN, S. R., BRISKINE, R., SONG, J., LI, Q., SWANSON-WAGNER, R., HERMANSON, P. J., WATERS, A. J., STARR, E., WEST, P. T., TIFFIN, P., MYERS, C. L., VAUGHN, M. W. & SPRINGER, N. M. 2013a. Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations. *Plant Cell*, 25, 2783-2797.
- EICHTEN, S. R., ELLIS, N. A., MAKAREVITCH, I., YEH, C. T., GENT, J. I., GUO, L., MCGINNIS, K. M., ZHANG, X. Y., SCHNABLE, P. S., VAUGHN, M. W., DAWES, R. K. & SPRINGER, N. M. 2012. Spreading of Heterochromatin Is Limited to Specific Families of Maize Retrotransposons. *Plos Genetics*, 8.
- EICHTEN, S. R., SCHMITZ, R. J. & SPRINGER, N. M. 2014. Epigenetics: Beyond Chromatin Modifications and Complex Genetic Regulation. *Plant Physiology*, 165, 933-947.
- EICHTEN, S. R., VAUGHN, M. W., HERMANSON, P. J. & SPRINGER, N. M. 2013b. Variation in DNA Methylation Patterns is More Common among Maize Inbreds than among Tissues. *Plant Genome*, 6.
- FAN, D., DAI, Y., WANG, X., WANG, Z., HE, H., YANG, H., CAO, Y., DENG, X. W. & MA, L. 2012. IBM1, a JmjC domain-containing histone demethylase, is involved in the regulation of RNA-directed DNA methylation through the epigenetic control of RDR2 and DCL3 expression in Arabidopsis. *Nucleic Acids Research*, 40, 8905-8916.
- FELIPPE, F. F. & WEIGEL, D. 2009. Triggering the formation of tasiRNAs in Arabidopsis thaliana: the role of microRNA miR173. *Embo Reports*, 10, 264-270.
- GAO, R., TAN, D. J. X. & WONG, S. M. 2013. Upregulation of miR395 which targets ATP sulfurylase and sulfate transporter facilitates sulfur enhanced defence after Hibiscus chlorotic ringspot virus infection. *Plant Pathology Bulletin*, 22, 133-143.

- GAO, Z., LIU, H.-L., DAXINGER, L., PONTES, O., HE, X., QIAN, W., LIN, H., XIE, M., LORKOVIC, Z. J., ZHANG, S., MIKI, D., ZHAN, X., PONTIER, D., LAGRANGE, T., JIN, H., MATZKE, A. J. M., MATZKE, M., PIKAARD, C. S. & ZHU, J.-K. 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature*, 465, 106-U118.
- GARCIA, D., GARCIA, S., PONTIER, D., MARCHAIS, A., RENOU, J. P., LAGRANGE, T. & VOINNET, O. 2012. Ago Hook and RNA Helicase Motifs Underpin Dual Roles for SDE3 in Antiviral Defense and Silencing of Nonconserved Intergenic Regions. *Molecular Cell*, 48, 109-120.
- GENDREL, A. V., LIPPMAN, Z., YORDAN, C., COLOT, V. & MARTIENSSEN, R. A. 2002. Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. *Science*, 297, 1871-1873.
- GENT, J. I., ELLIS, N. A., GUO, L., HARKESS, A. E., YAO, Y. Y., ZHANG, X. Y. & DAWE, R. K. 2013. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Research*, 23, 628-637.
- GOECKS J, N. A., TAYLOR J 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11, R86.
- GONG, Z. H., MORALES-RUIZ, T., ARIZA, R. R., ROLDAN-ARJONA, T., DAVID, L. & ZHU, J. K. 2002. ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell*, 111, 803-814.
- GORDON, T. R. & MARTYN, R. D. 1997. The evolutionary biology of *Fusarium oxysporum*. *Annual Review of Phytopathology*, 35, 111-128.
- GROSZMANN, M., GREAVES, I. K., ALBERTYN, Z. I., SCOFIELD, G. N., PEACOCK, W. J. & DENNIS, E. S. 2011. Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2617-2622.
- GULERIA, P., MAHAJAN, M., BHARDWAJ, J. & YADAV, S. K. 2011. Plant Small RNAs: Biogenesis, Mode of Action and Their Roles in Abiotic Stresses. *Genomics Proteomics & Bioinformatics*, 9, 183-199.
- GUTZAT, R. & SCHEID, O. M. 2012. Epigenetic responses to stress: triple defense? *Current Opinion in Plant Biology*, 15, 568-573.
- HA, M., LU, J., TIAN, L., RAMACHANDRAN, V., KASSCHAU, K. D., CHAPMAN, E. J., CARRINGTON, J. C., CHEN, X., WANG, X.-J. & CHEN, Z. J. 2009. Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 17835-17840.
- HALL, C., HEATH, R. & GUEST, D. 2013. The infection process of *Fusarium oxysporum* f.sp. *vasinfectum* in Australian cotton. *Australasian Plant Pathology*, 42, 1-8.
- HARTL, D. L. & JONES, E. W. (eds.) 2005. *GENETICS Analysis of Genes and Genomes*, Sudbury, MA: Jones and Bartlett Publishers.
- HE, X.-J., HSU, Y.-F., ZHU, S., WIERZBICKI, A. T., PONTES, O., PIKAARD, C. S., 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, 498-508.
- HENDERSON, I. R., DELERIS, A., WONG, W., ZHONG, X., CHIN, H. G., HORWITZ, G. A., KELLY, K. A., PRADHAN, S. & JACOBSEN, S. E. 2010. The De Novo Cytosine

- Methyltransferase DRM2 Requires Intact UBA Domains and a Catalytically Mutated Paralog DRM3 during RNA-Directed DNA Methylation in *Arabidopsis thaliana*. *Plos Genetics*, 6.
- HOLLISTER, J. D., SMITH, L. M., GUO, Y.-L., OTT, F., WEIGEL, D. & GAUT, B. S. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2322-2327.
- HOU, P. Q., LEE, Y. I., HSU, K. T., LIN, Y. T., WU, W. Z., LIN, J. Y., NAM, T. N. & FU, S. F. 2014. Functional characterization of *Nicotiana benthamiana* chromomethylase 3 in developmental programs by virus-induced gene silencing. *Physiologia Plantarum*, 150, 119-132.
- HOWELL, M. D., FAHLGREN, N., CHAPMAN, E. J., CUMBIE, J. S., SULLIVAN, C. M., GIVAN, S. A., KASSCHAU, K. D. & CARRINGTON, J. C. 2007. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell*, 19, 926-942.
- HSIEH, T.-F., IBARRA, C. A., SILVA, P., ZEMACH, A., ESHED-WILLIAMS, L., FISCHER, R. L. & ZILBERMAN, D. 2009. Genome-Wide Demethylation of *Arabidopsis* Endosperm. *Science*, 324, 1451-1454.
- HU, Q., HOLLUNDER, J., NIEHL, A., KORNER, C. J., GEREIGE, D., WINDELS, D., ARNOLD, A., KUIPER, M., VAZQUEZ, F., POOGGIN, M. & HEINLEIN, M. 2011. Specific Impact of Tobamovirus Infection on the *Arabidopsis* Small RNA Profile. *Plos One*, 6.
- HU, W. X., WANG, T. Z., XU, J. H. & LI, H. Z. 2014. MicroRNA mediates DNA methylation of target genes. *Biochemical and biophysical research communications*, 444, 676-681.
- HUANG, J., WANG, H., LIANG, W., XIE, X. & GUO, G. 2014. Developmental expression of *Arabidopsis* methyltransferase genes MET1, DRM2, and CMT3. *Molecular Biology*, 48, 681-687.
- HUETTEL, B., KANNO, T., DAXINGER, L., AUFSATZ, W., MATZKE, A. J. M. & MATZKE, M. 2006. Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *Embo Journal*, 25, 2828-2836.
- JEONG, D. H. & GREEN, P. J. 2013. The role of rice microRNAs in abiotic stress responses. *Journal of Plant Biology*, 56, 187-197.
- JOHANNES, F., PORCHER, E., TEIXEIRA, F. K., SALIBA-COLOMBANI, V., SIMON, M., AGIER, N., BULSKI, A., ALBUISSON, J., HEREDIA, F., AUDIGIER, P., BOUCHEZ, D., DILLMANN, C., GUERCHE, P., HOSPITAL, F. & COLOT, V. 2009. Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. *Plos Genetics*, 5.
- KASSCHAU, K. D., FAHLGREN, N., CHAPMAN, E. J., SULLIVAN, C. M., CUMBIE, J. S., GIVAN, S. A. & CARRINGTON, J. C. 2007. Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *Plos Biology*, 5, 479-493.
- KHRAIWESH, B., ZHU, J.-K. & ZHU, J. 2012. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms*, 1819, 137-148.
- KIM, J.-M., TO, T. K., NISHIOKA, T. & SEKI, M. 2010a. Chromatin regulation functions in plant abiotic stress responses. *Plant Cell and Environment*, 33, 604-611.

- KIM, J., KIM, J. H., RICHARDS, E. J., CHUNG, K. M. & WOO, H. R. 2014. Arabidopsis VIM Proteins Regulate Epigenetic Silencing by Modulating DNA Methylation and Histone Modification in Cooperation with MET1. *Molecular Plant*, 7, 1470-1485.
- KIM, J. Y., KWAK, K. J., JUNG, H. J., LEE, H. J. & KANG, H. 2010b. MicroRNA402 Affects Seed Germination of Arabidopsis thaliana Under Stress Conditions via Targeting DEMETER-LIKE Protein3 mRNA. *Plant and Cell Physiology*, 51, 1079-1083.
- KRUEGER, F. & ANDREWS, S. R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27, 1571-1572.
- LE, T. N., SCHUMANN, U., SMITH, N. A., TIWARI, S., AU, P., ZHU, Q. H., TAYLOR, J., KAZAN, K., LLEWELLYN, D. J., ZHANG, R., DENNIS, E. S. & WANG, M. B. 2014. DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis. *Genome Biol*, 15, 458.
- LEE, J. R. M., DUXBURY, Z. & WANG, M. 2013. Small RNAs and transgenerational epigenetic variation. *American Journal of Plant Sciences*, 4, 641-646.
- LI, Y., ZHANG, Q., ZHANG, J., WU, L., QI, Y. & ZHOU, J.-M. 2010. Identification of MicroRNAs Involved in Pathogen-Associated Molecular Pattern-Triggered Plant Innate Immunity. *Plant Physiology*, 152, 2222-2231.
- LINDROTH, A. M., CAO, X. F., JACKSON, J. P., ZILBERMAN, D., MCCALLUM, C. M., HENIKOFF, S. & JACOBSEN, S. E. 2001. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science*, 292, 2077-2080.
- LISTER, R., O'MALLEY, R. C., TONTI-FILIPPINI, J., GREGORY, B. D., BERRY, C. C., MILLAR, A. H. & ECKER, J. R. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, 133, 523-536.
- LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods*, 25, 402-408.
- LOPEZ, A., RAMIREZ, V., GARCIA-ANDRADE, J., FLORS, V. & VERA, P. 2011. The RNA Silencing Enzyme RNA Polymerase V Is Required for Plant Immunity. *Plos Genetics*, 7.
- LOU, S., LEE, H.-M., QIN, H., LI, J.-W., GAO, Z., LIU, X., CHAN, L. L., LAM, V. K. L., SO, W.-Y., WANG, Y., LOK, S., WANG, J., MA, R. C. W., TSUI, S. K.-W., CHAN, J. C. N., CHAN, T.-F. & YIP, K. Y. 2014. Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. *Genome Biology*, 15.
- LUNA, E., BRUCE, T. J. A., ROBERTS, M. R., FLORS, V. & TON, J. 2012. Next-Generation Systemic Acquired Resistance. *Plant Physiology*, 158, 844-853.
- LYONS, R., STILLER, J., POWELL, J., RUSU, A., MANNERS, J. M. & KAZAN, K. 2015. Fusarium oxysporum Triggers Tissue-Specific Transcriptional Reprogramming in Arabidopsis thaliana. *Plos One*, 10, e0121902-e0121902.
- MAKAREVITCH, I., WATERS, A. J., WEST, P. T., STITZER, M., HIRSCH, C. N., ROSS-IBARRA, J. & SPRINGER, N. M. 2015. Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress. *Plos Genetics*, 11.
- MARI-ORDONEZ, A., MARCHAIS, A., ETCHEVERRY, M., MARTIN, A., COLOT, V. & VOINNET, O. 2013. Reconstructing de novo silencing of an active plant retrotransposon. *Nature Genetics*, 45, 1029-+.

- MATHIEU, O., REINDERS, J., CAIKOVSKI, M., SMATHAJITT, C. & PASZKOWSKI, J. 2007. Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell*, 130, 851-862.
- MATTHEWMAN, C. A., KAWASHIMA, C. G., HUSKA, D., CSORBA, T., DALMAY, T. & KOPRIVA, S. 2012. miR395 is a general component of the sulfate assimilation regulatory network in Arabidopsis. *Febs Letters*, 586, 3242-3248.
- MCCARTHY, D. J., CHEN, Y. & SMYTH, G. K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40, 4288-4297.
- MI, S., CAI, T., HU, Y., CHEN, Y., HODGES, E., NI, F., WU, L., LI, S., ZHOU, H., LONG, C., CHEN, S., HANNON, G. J. & QI, Y. 2008. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, 133, 116-127.
- MICHIELSE, C. B. & REP, M. 2009. Pathogen profile update: Fusarium oxysporum. *Molecular Plant Pathology*, 10, 311-324.
- NAUMANN, U., DAXINGER, L., KANNO, T., EUN, C., LONG, Q. A., LORKOVIC, Z. J., MATZKE, M. & MATZKE, A. J. M. 2011. Genetic Evidence That DNA Methyltransferase DRM2 Has a Direct Catalytic Role in RNA-Directed DNA Methylation in Arabidopsis thaliana. *Genetics*, 187, 977-979.
- NUTHIKATTU, S., MCCUE, A. D., PANDA, K., FULTZ, D., DEFRAIA, C., THOMAS, E. N. & SLOTKIN, R. K. 2013. The Initiation of Epigenetic Silencing of Active Transposable Elements Is Triggered by RDR6 and 21-22 Nucleotide Small Interfering RNAs. *Plant Physiology*, 162, 116-131.
- ORTEGA-GALISTEO, A. P., MORALES-RUIZ, T., ARIZA, R. R. & ROLDAN-ARJONA, T. 2008. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Molecular Biology*, 67, 671-681.
- PALL, G. S. & HAMILTON, A. J. 2008. Improved northern blot method for enhanced detection of small RNA. *Nature Protocols*, 3, 1077-1084.
- PAVET, V., QUINTERO, C., CECCHINI, N. M., ROSA, A. L. & ALVAREZ, M. E. 2006. Arabidopsis displays centromeric DNA hypomethylation and cytological alterations of heterochromatin upon attack by Pseudomonas syringae. *Molecular Plant-Microbe Interactions*, 19, 577-587.
- PENTERMAN, J., UZAWA, R. & FISCHER, R. L. 2007a. Genetic interactions between DNA demethylation and methylation in Arabidopsis(OA). *Plant Physiology*, 145, 1549-1557.
- PENTERMAN, J., ZILBERMAN, D., HUH, J. H., BALLINGER, T., HENIKOFF, S. & FISCHER, R. L. 2007b. DNA demethylation in the Arabidopsis genome. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 6752-6757.
- PESCHEN, D., LI, H. P., FISCHER, R., KREUZALER, F. & LIAO, Y. C. 2004. Fusion proteins comprising a Fusarium-specific antibody linked to antifungal peptides protect plants against a fungal pathogen. *Nature Biotechnology*, 22, 732-738.
- PONTIER, D., PICART, C., ROUDIER, F., GARCIA, D., LAHMY, S., AZEVEDO, J., ALART, E., LAUDIE, M., KARLOWSKI, W. M., COOKE, R., COLOT, V., VOINNET, O. & LAGRANGE, T. 2012. NERD, a Plant-Specific GW Protein, Defines an

- Additional RNAi-Dependent Chromatin-Based Pathway in Arabidopsis. *Molecular Cell*, 48, 121-132.
- POPOVA, O. V., DINH, H. Q., AUFSATZ, W. & JONAK, C. 2013. The RdDM Pathway Is Required for Basal Heat Tolerance in Arabidopsis. *Molecular Plant*, 6, 396-410.
- QIAN, W. Q., MIKI, D., ZHANG, H., LIU, Y. H., ZHANG, X., TANG, K., KAN, Y. C., LA, H. G., LI, X. J., LI, S. F., ZHU, X. H., SHI, X. B., ZHANG, K. L., PONTES, O., CHEN, X. M., LIU, R. Y., GONG, Z. Z. & ZHU, J. K. 2012. A Histone Acetyltransferase Regulates Active DNA Demethylation in Arabidopsis. *Science*, 336, 1445-1448.
- QUESTA, J. I., FINA, J. P. & CASATI, P. 2013. DDM1 and ROS1 have a role in UV-B induced- and oxidative DNA damage in *A. thaliana*. *Frontiers in Plant Science*, 4.
- QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841-842.
- RACINE, J. 2000. The Cygwin tools: A GNU toolkit for windows. *Journal of Applied Econometrics*, 15, 331-341.
- REEB, P. D. & STEIBEL, J. P. 2013. Evaluating statistical analysis models for RNA sequencing experiments. *Frontiers in genetics*, 4, 178.
- ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.
- ROBINSON, M. D. & SMYTH, G. K. 2007. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, 23, 2881-2887.
- ROBINSON, M. D. & SMYTH, G. K. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, 9, 321-332.
- ROY, N. S., CHOI, J. Y., LEE, S. I. & KIM, N. S. 2015. Marker utility of transposable elements for plant genetics, breeding, and ecology: a review. *Genes & Genomics*, 37, 141-151.
- ROZEN, S. & SKALETSKY, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, N.J.)*, 132, 365-86.
- RUBIO-SOMOZA, I. & WEIGEL, D. 2011. MicroRNA networks and developmental plasticity in plants. *Trends in Plant Science*, 16, 258-264.
- RUIZ-FERRER, V. & VOINET, O. 2009. Roles of Plant Small RNAs in Biotic Stress Responses. *Annual Review of Plant Biology*.
- SANMIGUEL, P. & BENNETZEN, J. L. 1998. Evidence that a Recent Increase in Maize Genome Size was Caused by the Massive Amplification of Intergene Retrotransposons. *Annals of Botany*, 82, 37-44.
- SASAKI, T., KOBAYASHI, A., SAZE, H. & KAKUTANI, T. 2012. RNAi-independent de novo DNA methylation revealed in Arabidopsis mutants of chromatin remodeling gene DDM1. *Plant Journal*, 70, 750-758.
- SAZE, H., TSUGANE, K., KANNO, T. & NISHIMURA, T. 2012. DNA Methylation in Plants: Relationship to Small RNAs and Histone Modifications, and Functions in Transposon Inactivation. *Plant and Cell Physiology*, 53, 766-784.
- SCHMITZ, R. J., SCHULTZ, M. D., LEWSEY, M. G., O'MALLEY, R. C., URICH, M. A., LIBIGER, O., SCHORK, N. J. & ECKER, J. R. 2011. Transgenerational

- Epigenetic Instability Is a Source of Novel Methylation Variants. *Science*, 334, 369-373.
- SEO, J.-K., WU, J., LI, Y., LI, Y. & JIN, H. 2013. Contribution of Small RNA Pathway Components in Plant Immunity. *Molecular Plant-Microbe Interactions*, 26, 617-625.
- SHAN, X., WANG, X., YANG, G., WU, Y., SU, S., LI, S., LIU, H. & YUAN, Y. 2013. Analysis of the DNA methylation of maize (*Zea mays* L.) in response to cold stress based on methylation-sensitive amplified polymorphisms. *Journal of Plant Biology*, 56, 32-38.
- SHEN, X., DE JONGE, J., FORSBERG, S. K. G., PETTERSSON, M. E., SHENG, Z. Y., HENNIG, L. & CARLBORG, O. 2014. Natural CMT2 Variation Is Associated With Genome-Wide Methylation Changes and Temperature Seasonality. *Plos Genetics*, 10.
- SHOOK, M. S. & RICHARDS, E. J. 2014. VIM proteins regulate transcription exclusively through the MET1 cytosine methylation pathway. *Epigenetics*, 9, 980-986.
- SI-AMMOUR, A., WINDELS, D., ARN-BOULDOIRES, E., KUTTER, C., AILHAS, J., MEINS, F., JR. & VAZQUEZ, F. 2011. miR393 and Secondary siRNAs Regulate Expression of the TIR1/AFB2 Auxin Receptor Clade and Auxin-Related Development of Arabidopsis Leaves. *Plant Physiology*, 157, 683-691.
- SINGH, A., ZUBKO, E. & MEYER, P. 2008. Cooperative activity of DNA methyltransferases for maintenance of symmetrical and non-symmetrical cytosine methylation in Arabidopsis thaliana. *Plant Journal*, 56, 814-823.
- SMALL, I. D. & LEAVER, C. J. 1988. PLANT GENETIC-TRANSFORMATION AND GENE-EXPRESSION - A LABORATORY MANUAL - DRAPER, J., SCOTT, R., ARMITAGE, P., WALDEN, R. *Nature*, 335, 216-216.
- STEWART, N., ITO, M., YAMAGUCHI, Y., KOIZUMI, N. & SANO, H. 2002. Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *Journal of Biological Chemistry*, 277, 37741-37746.
- STROUD, H., GREENBERG, M. V. C., FENG, S. H., BERNATAVICHUTE, Y. V. & JACOBSEN, S. E. 2013. Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. *Cell*, 152, 352-364.
- SUN, F. L., GUO, W. W., DU, J. K., NI, Z. F., SUN, Q. X. & YAO, Y. Y. 2013. Widespread, abundant, and diverse TE-associated siRNAs in developing wheat grain. *Gene*, 522, 1-7.
- TAKUNO, S., RAN, J.-H. & GAUT, B. S. 2016. Evolutionary patterns of genic DNA methylation vary across land plants. *Nature Plants*, 2, 15222.
- TEIXEIRA, F. K., HEREDIA, F., SARAZIN, A., ROUDIER, F., BOCCARA, M., CIAUDO, C., CRUAUD, C., POULAIN, J., BERDASCO, M., FRAGA, M. F., VOINNET, O., WINCKER, P., ESTELLER, M. & COLOT, V. 2009. A Role for RNAi in the Selective Correction of DNA Methylation Defects. *Science*, 323, 1600-1604.
- THIMM, O., BLASING, O., GIBON, Y., NAGEL, A., MEYER, S., KRUGER, P., SELBIG, J., MULLER, L. A., RHEE, S. Y. & STITT, M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal*, 37, 914-939.
- THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14, 178-192.

- TOMPA, R., MCCALLUM, C. M., DELROW, J., HENIKOFF, J. G., VAN STEENSEL, B. & HENIKOFF, S. 2002. Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Current Biology*, 12, 65-68.
- VAUCHERET, H. 2008. Plant ARGONAUTES. *Trends in Plant Science*, 13, 350-358.
- VAUCHERET, H., VAZQUEZ, F., CRETE, P. & BARTEL, D. P. 2004. The action of ARGONAUT1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes & Development*, 18, 1187-1197.
- VINING, K. J., POMRANING, K. R., WILHELM, L. J., PRIEST, H. D., PELLEGRINI, M., MOCKLER, T. C., FREITAG, M. & STRAUSS, S. H. 2012. Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *Bmc Genomics*, 13.
- WANG, C.-Y., ZHANG, S., YU, Y., LUO, Y.-C., LIU, Q., JU, C., ZHANG, Y.-C., QU, L.-H., LUCAS, W. J., WANG, X. & CHEN, Y.-Q. 2014a. MiR397b regulates both lignin content and seed number in *Arabidopsis* via modulating a laccase involved in lignin biosynthesis. *Plant Biotechnology Journal*, 12, 1132-1142.
- WANG, N. N., ZHANG, D., WANG, Z. H., XUN, H. W., MA, J., WANG, H., HUANG, W., LIU, Y., LIN, X. Y., LI, N., OU, X. F., ZHANG, C. Y., WANG, M. B. & LIU, B. 2014b. Mutation of the RDR1 gene caused genome-wide changes in gene expression, regional variation in small RNA clusters and localized alteration in DNA methylation in rice. *Bmc Plant Biology*, 14.
- WANG, X. G., DUAN, C. G., TANG, K., WANG, B. S., ZHANG, H. M., LEI, M. G., LU, K., MANGRAUTHIA, S. K., WANG, P. C., ZHU, G. H., ZHAO, Y. & ZHU, J. K. 2013. RNA-binding protein regulates plant DNA methylation by controlling mRNA processing at the intronic heterochromatin-containing gene IBM1. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 15467-15472.
- WANG, Y., ITAYA, A., ZHONG, X., WU, Y., ZHANG, J., VAN DER KNAAP, E., OLMSTEAD, R., QI, Y. & DING, B. 2011. Function and Evolution of a MicroRNA That Regulates a Ca²⁺-ATPase and Triggers the Formation of Phased Small Interfering RNAs in Tomato Reproductive Growth. *Plant Cell*, 23, 3185-3203.
- WEI, C. H., KUANG, H. H., LI, F. & CHEN, J. J. 2014. The I2 resistance gene homologues in *Solanum* have complex evolutionary patterns and are targeted by miRNAs. *Bmc Genomics*, 15.
- WEIBERG, A., WANG, M., BELLINGER, M. & JIN, H. L. 2014. Small RNAs: A New Paradigm in Plant-Microbe Interactions. *Annual Review of Phytopathology*, Vol 52, 52, 495-516.
- WHEELER, B. 2013. Small RNAs, big impact: small RNA pathways in transposon control and their effect on the host stress response. *Chromosome Research*, 21, 587-600.
- WILLIAMS, B. P., PIGNATTA, D., HENIKOFF, S. & GEHRING, M. 2015. Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genet*, 11, e1005142.
- WONG, J., GAO, L., YANG, Y., ZHAI, J., ARIKIT, S., YU, Y., DUAN, S., CHAN, V., XIONG, Q., YAN, J., LI, S., LIU, R., WANG, Y., TANG, G., MEYERS, B. C., CHEN, X. & MA, W. 2014. Roles of small RNAs in soybean defense against *Phytophthora sojae* infection. *Plant Journal*, 79, 928-940.

- WU, L. 2013. DICER-LIKE1 processed trans-acting siRNAs mediate DNA methylation: case study of complex small RNA biogenesis and action pathways in plants. *Plant signaling & behavior*, 8, e22476.
- WU, L., MAO, L. & QI, Y. 2012. Roles of DICER-LIKE and ARGONAUTE Proteins in TAS-Derived Small Interfering RNA-Triggered DNA Methylation. *Plant Physiology*, 160, 990-999.
- WU, L., ZHOU, H., ZHANG, Q., ZHANG, J., NI, F., LIU, C. & QI, Y. 2010. DNA Methylation Mediated by a MicroRNA Pathway. *Molecular Cell*, 38, 465-475.
- XIAO, B. G., YANG, X. L., YE, C. Y., LIU, Y., YAN, C. H., WANG, Y., LU, X. P., LI, Y. P. & FAN, L. J. 2014. A diverse set of miRNAs responsive to begomovirus-associated betasatellite in *Nicotiana benthamiana*. *Bmc Plant Biology*, 14.
- XIAO, W. Y., GEHRING, M., CHOI, Y., MARGOSSIAN, L., PU, H., HARADA, J. J., GOLDBERG, R. B., PENNELL, R. I. & FISCHER, R. L. 2003. Imprinting of the MEA polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Developmental Cell*, 5, 891-901.
- XIE, M. & YU, B. 2015. siRNA-directed DNA Methylation in Plants. *Current Genomics*, 16, 23-31.
- YU, A., LEPERE, G., JAY, F., WANG, J., BAPAUME, L., WANG, Y., ABRAHAM, A.-L., PENTERMAN, J., FISCHER, R. L., VOINNET, O. & NAVARRO, L. 2013. Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 2389-94.
- ZHAI, J., JEONG, D. H., DE PAOLI, E., PARK, S., ROSEN, B. D., LI, Y., GONZALEZ, A. J., YAN, Z., KITTO, S. L., GRUSAK, M. A., JACKSON, S. A., STACEY, G., COOK, D. R., GREEN, P. J., SHERRIER, D. J. & MEYERS, B. C. 2011. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev*, 25, 2540-53.
- ZHANG, C.-J., ZHOU, J.-X., LIU, J., MA, Z.-Y., ZHANG, S.-W., DOU, K., HUANG, H.-W., CAI, T., LIU, R., ZHU, J.-K. & HE, X.-J. 2013. The splicing machinery promotes RNA-directed DNA methylation and transcriptional silencing in Arabidopsis. *Embo Journal*, 32, 1128-1140.
- ZHANG, X. Y., YAZAKI, J., SUNDARESAN, A., COKUS, S., CHAN, S. W. L., CHEN, H. M., HENDERSON, I. R., SHINN, P., PELLEGRINI, M., JACOBSEN, S. E. & ECKER, J. R. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell*, 126, 1189-1201.
- ZHENG, X., PONTES, O., ZHU, J., MIKI, D., ZHANG, F., LI, W.-X., IIDA, K., KAPOOR, A., PIKAARD, C. S. & ZHU, J.-K. 2008. ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. *Nature*, 455, 1259-U70.
- ZHONG, X. H., HALE, C. J., NGUYEN, M., AUSIN, I., GROTH, M., HETZEL, J., VASHISHT, A. A., HENDERSON, I. R., WOHLSCHEGEL, J. A. & JACOBSEN, S. E. 2015. DOMAINS REARRANGED METHYLTRANSFERASE3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 911-916.
- ZHOU, C. H., ZHANG, L., DUAN, J., MIKI, B. & WU, K. Q. 2005. HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signaling of pathogen response in Arabidopsis. *Plant Cell*, 17, 1196-1204.
- ZHU, J.-K. 2009. Active DNA Demethylation Mediated by DNA Glycosylases. *Annual Review of Genetics*.

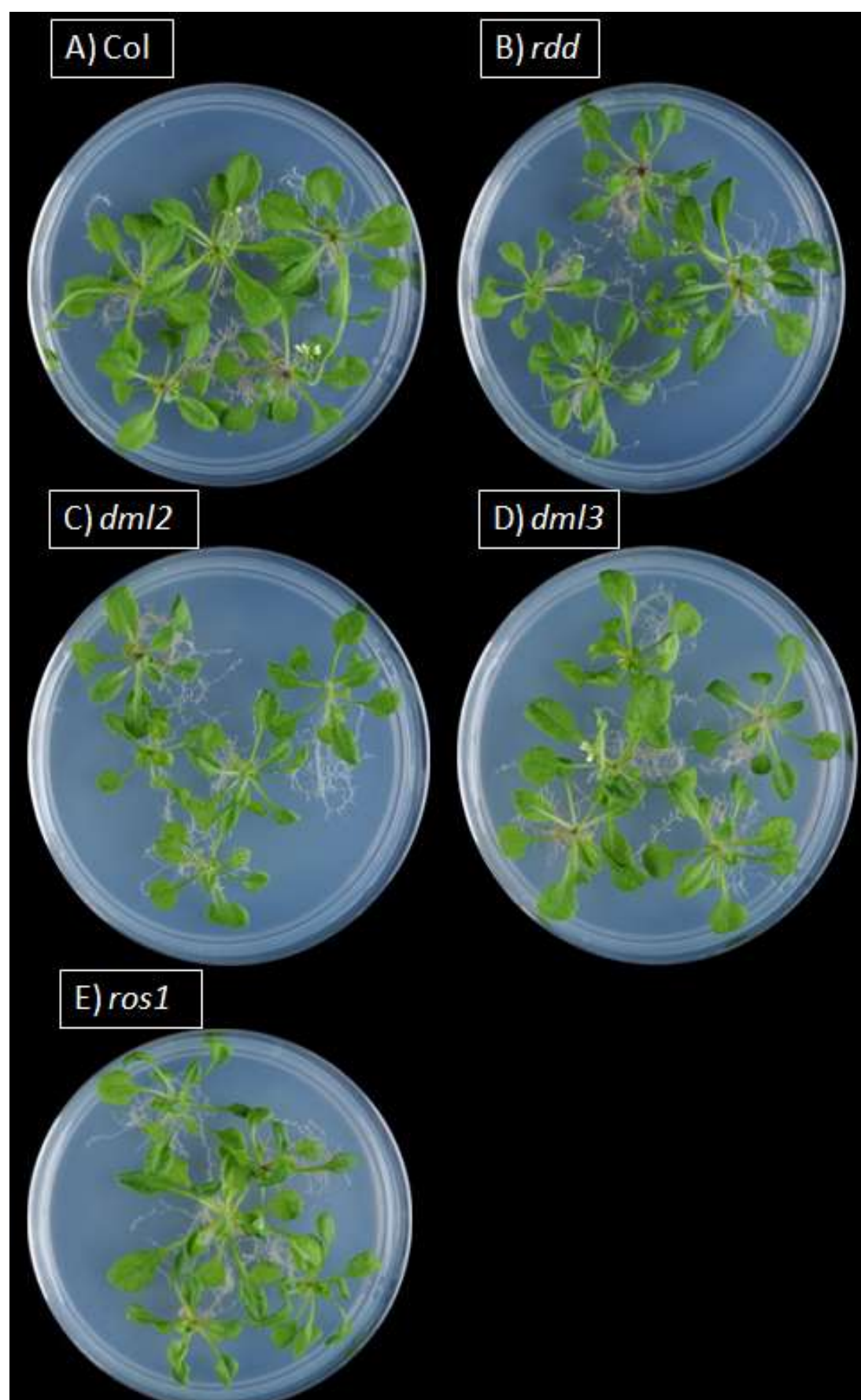
- ZHU, J., KAPOOR, A., SRIDHAR, V. V., AGIUS, F. & ZHU, J.-K. 2007. The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis. *Current Biology*, 17, 54-59.
- ZHU, Q.-H., STEPHEN, S., KAZAN, K., JIN, G., FAN, L., TAYLOR, J., DENNIS, E. S., HELLIWELL, C. A. & WANG, M.-B. 2013. Characterization of the defense transcriptome responsive to Fusarium oxysporum-infection in Arabidopsis using RNA-seq. *Gene*, 512, 259-266.
- ZILBERMAN, D., GEHRING, M., TRAN, R. K., BALLINGER, T. & HENIKOFF, S. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, 39, 61-69.

Chapter 9: Supplementary Figures and Tables

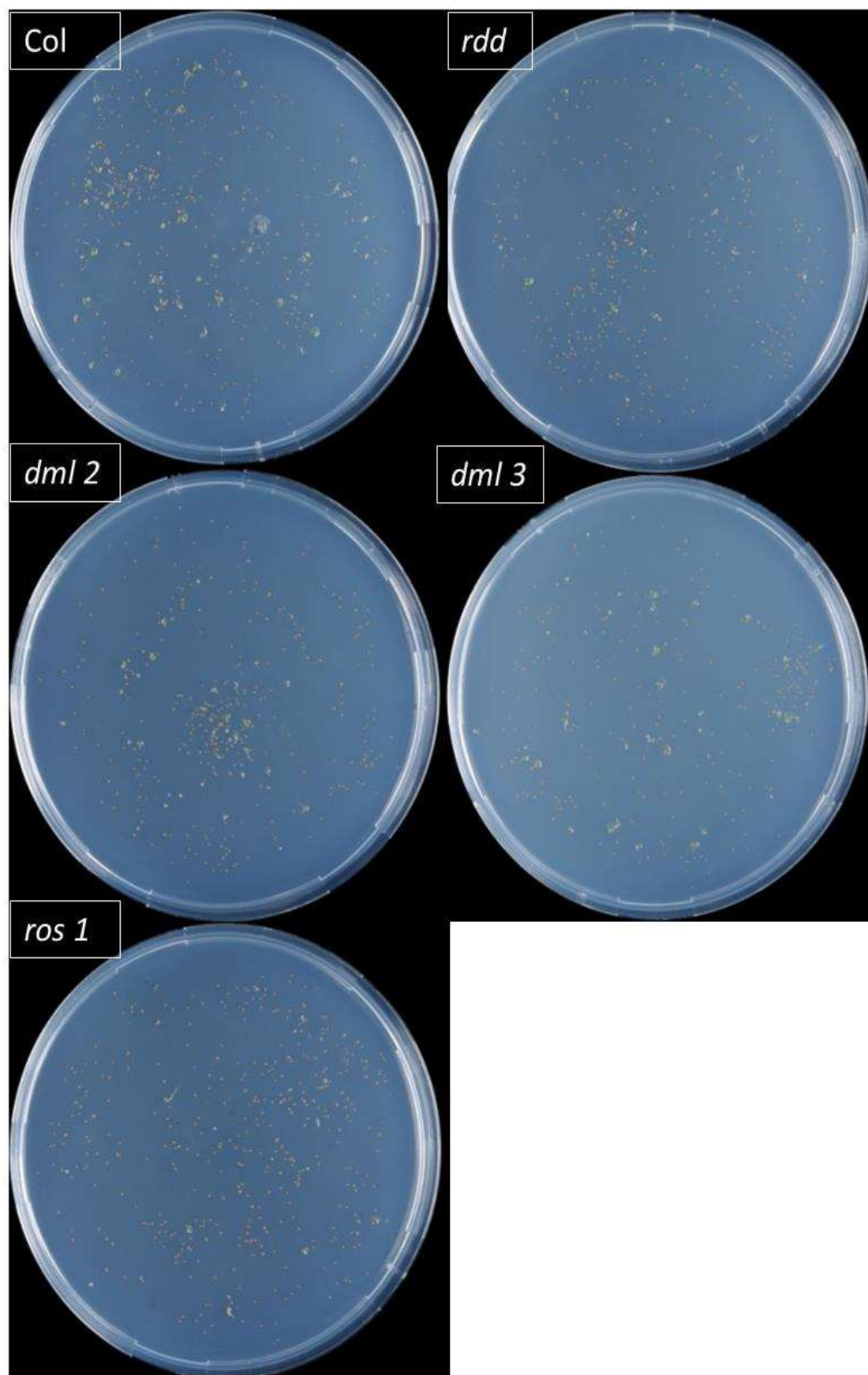
Supplementary Table 2.1: MS media

Buffer	Final Reagent Concentration	How to Prepare
20X Macro	33 g NH_4NO_3	Mix together
	8.8 g $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$	
	7.4 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	
	38.0 g KNO_3	
	3.4 g KH_2PO_4	
	dH ₂ O to 1 L	
1000X Micro	11.15 g $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$	Mix together
	0.125 g $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$	
	3.11 g H_3BO_3	
	4.3 g $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$	
	0.0125 g $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$	
	0.0125g $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$	
	0.115g KI	
	dH ₂ O to 500 mL	
200X Fe.EDTA	3.35g Na_2EDTA	Mix together
	$\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$	
	dH ₂ O to 500 mL	
100X Vitamins	5.0 mg Nicotinic acid	Mix together
	5.0 mg Pyridoxine HCl	
	1.0 mg Thiamine HCl	
	20 mg Glycine	
	dH ₂ O to 100 mL	
MS media	50 mL 20X Macro	Combine ingredients and adjust to pH 5.7, add 8.0 g Difco Agar (Bacto Laboratories) and autoclave. Plates containing MS media with no sucrose was used for <i>Fusarium oxysporum</i> inoculated samples.
	1.0 mL 1000X Micro	
	5.0 mL 200X Fe.EDTA	
	10.0 mL 100X Vitamins	
	30 g Sucrose	
	0.1 g Myoinositol	
	dH ₂ O to 1 L	

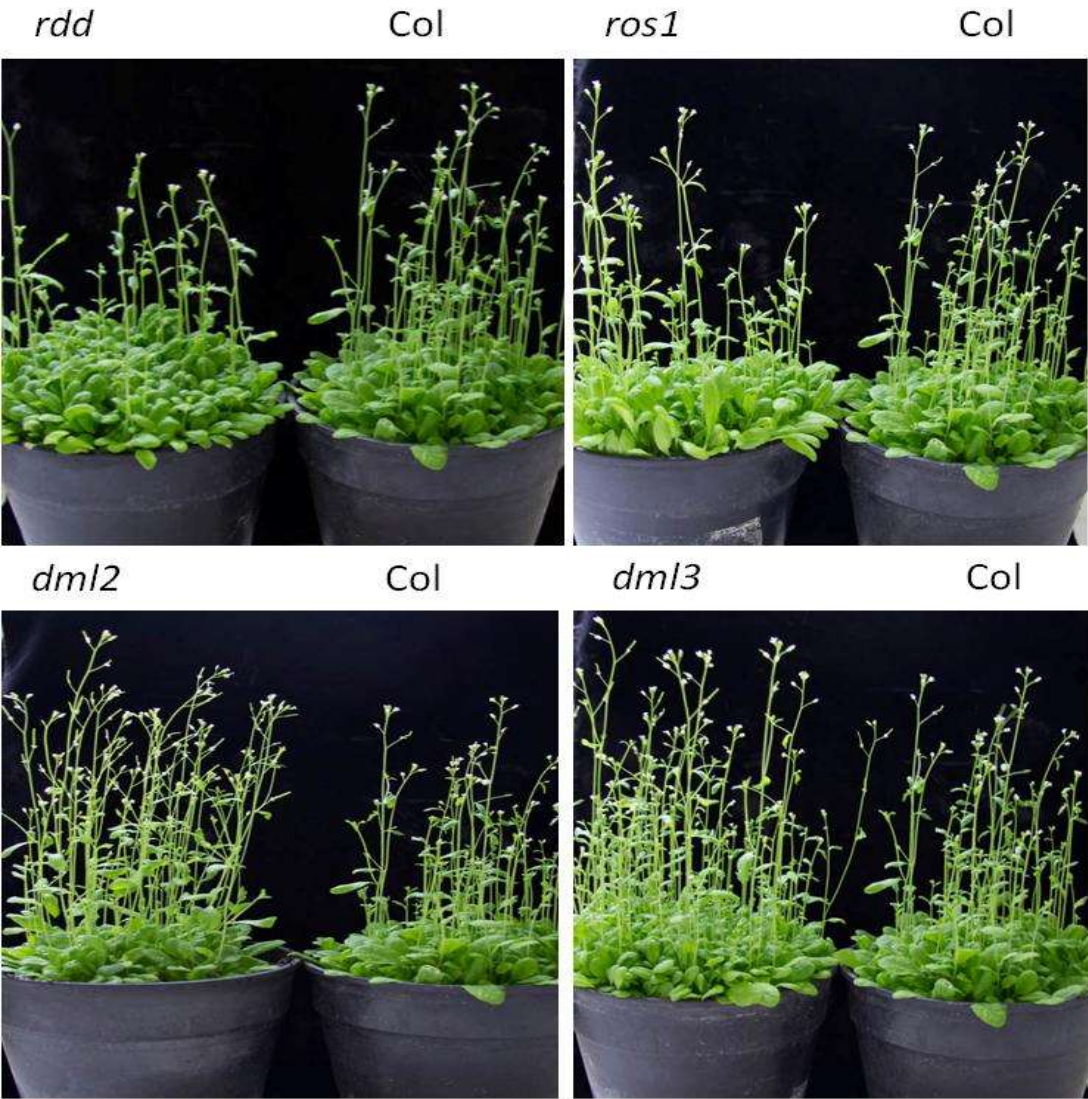
Supplementary Figure 3.1: Single demethylase comparison with *rdd* and Col-0 on plates



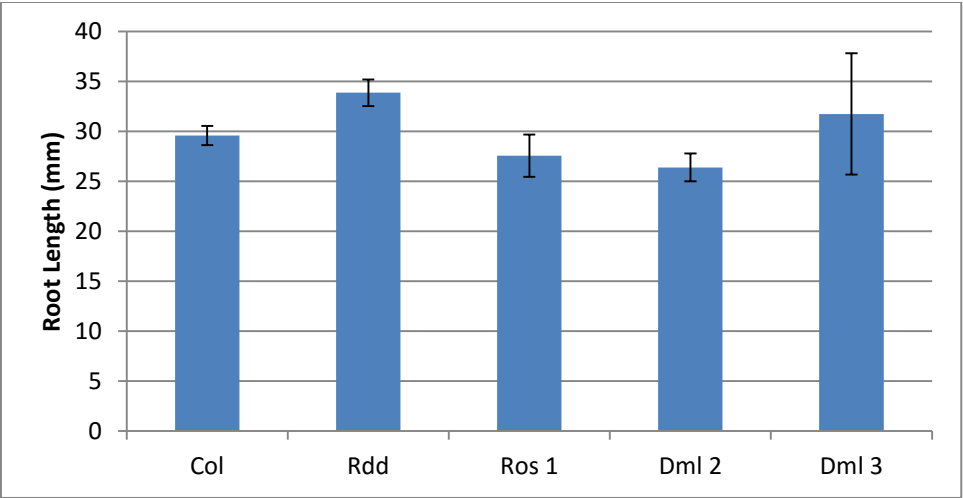
Supplementary Figure 3.2: Germination rate of demethylase mutants compared to Col-0 at 8 days after sowing



Supplementary Figure 3.3: Flowering time of demethylase mutants compared to Col-0 at approximately 6 weeks after sowing



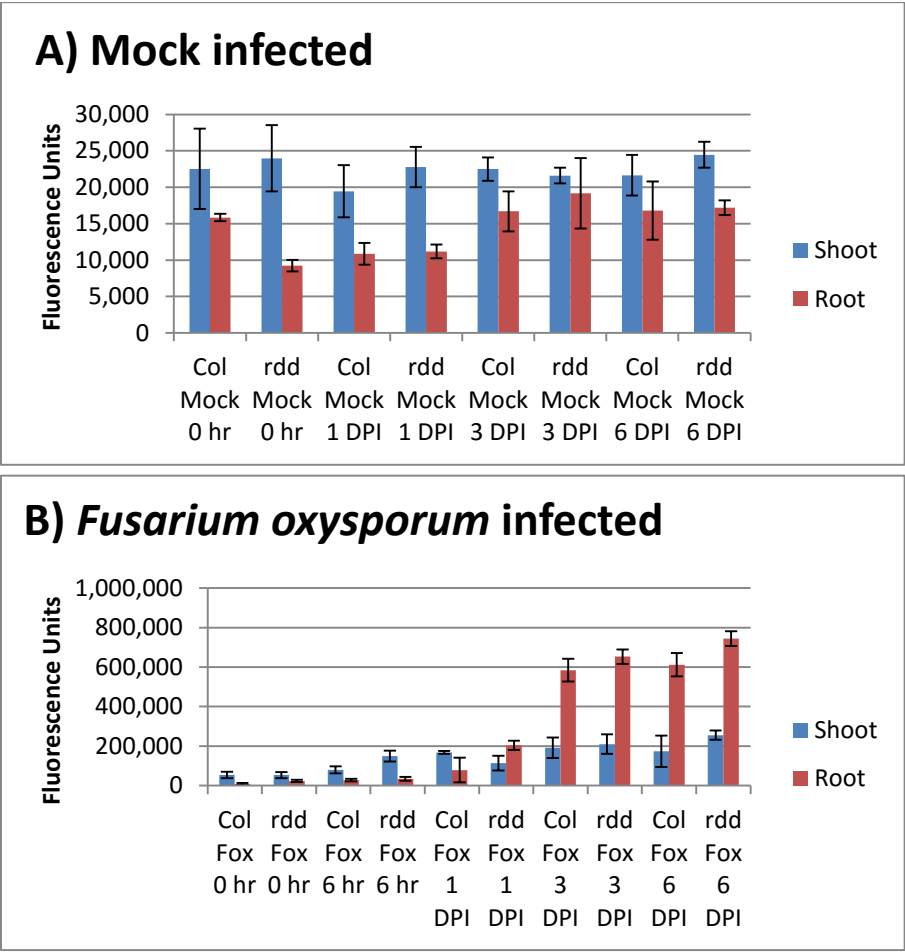
Supplementary Figure 3.4: Average root length under standard conditions 16 days post sowing in the demethylase mutants and Col-0 grown on MS agar with standard error



Supplementary Figure 3.5: Phenotypes of *rdd* and Col-0 10 days following cucumber mosaic virus (CMV) infection

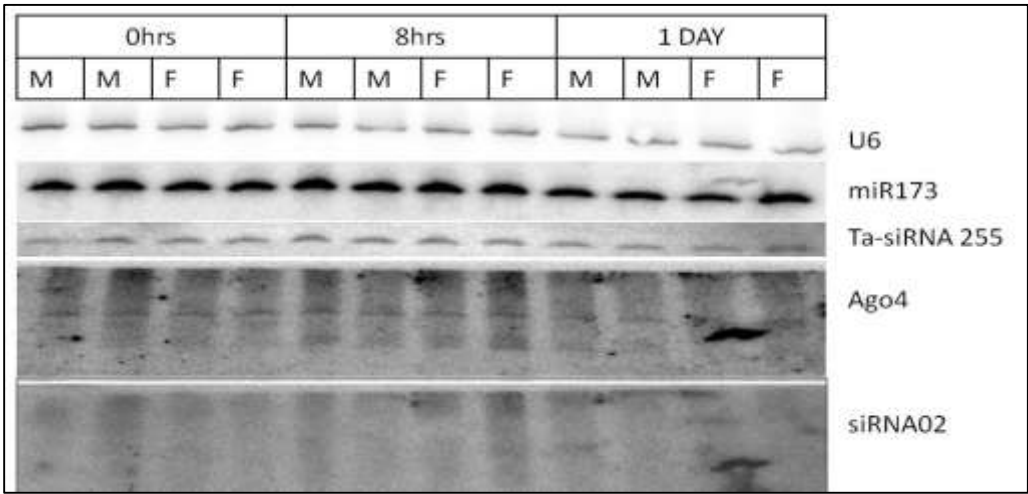


Supplementary Figure 3.5: Binding of WGA Fluorescence to Fungal Chitin in *Arabidopsis thaliana* separated roots and shoot tissue



Supplementary Figure 3.6: small RNA northern blot analysis for Col-0 mock and Col-0 *Fusarium* infected samples at 0hrs, 8hrs and 1 day post infection

U6 was used as a loading control and miRNA173 and ta-siRNA 255 were positive controls, while AGO4 and siRNA02 target 24nt small RNA involved in the RdDM pathway. The M stands for Mock treated and F stands for *Fusarium oxysporum* infected samples.

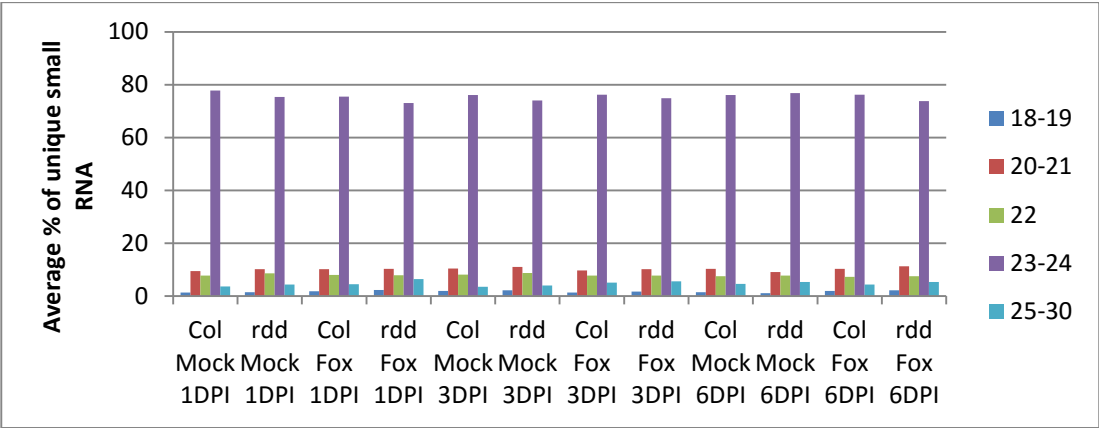


Supplementary Table 4.1: Summary of sRNA sequencing data for the 24 individual samples

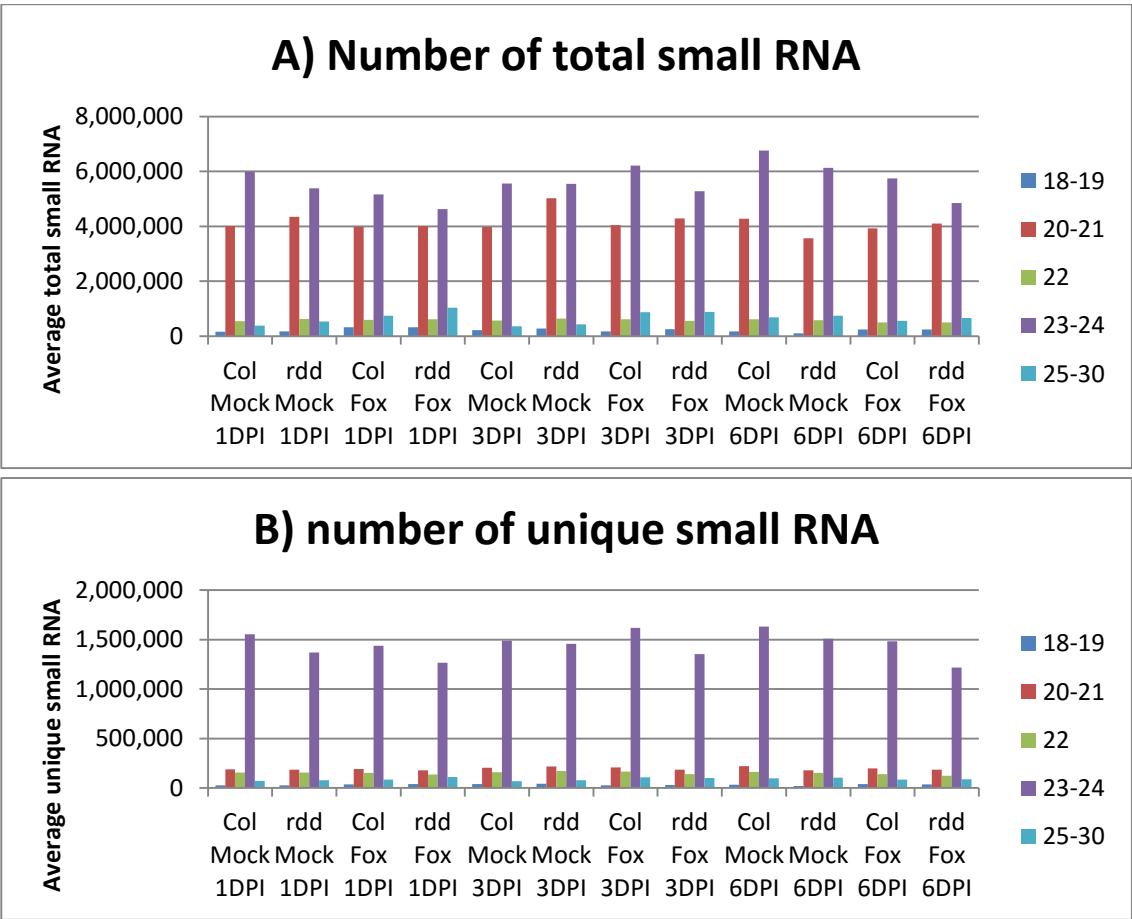
	Col Mock	Col Mock	<i>rdd</i> Mock	<i>rdd</i> Mock	Col Fox	Col Fox	<i>rdd</i> Fox	<i>rdd</i> Fox
	1DPI	1DPI	1DPI	1DPI	1DPI	1DPI	1DPI	1DPI
Total reads (adaptors and low quality reads already removed)	13,841,857 (100%)	13,177,349 (100%)	13,910,137 (100%)	13,253,170 (100%)	12,907,320 (100%)	13,684,596 (100%)	13,565,226 (100%)	13,112,649 (100%)
Reads after filtering (18-30nt) (percentage of total reads)	13,684,736 (98.865%)	13,000,022 (98.654%)	13,751,739 (98.861%)	13,086,207 (98.74%)	12,762,227 (98.876%)	13,414,043 (98.023%)	13,365,730 (98.529%)	12,857,733 (98.056%)
Reads mapped to <i>F. oxysporum</i> genome (percentage of trimmed and filtered reads)	1,312 (0.012%)	1,008 (0.009%)	1,545 (0.011%)	1,419 (0.011%)	6,538 (0.051%)	10,851 (0.081%)	14,004 (0.105%)	14,860 (0.116%)
Reads aligning in up to 10 different loci against <i>A. thaliana</i> a (percentage of trimmed and filtered reads)	11,371,238 (83.094%)	10,798,403 (83.064%)	11,267,338 (81.934%)	10,864,941 (83.026%)	10,555,988 (82.713%)	11,056,024 (82.421%)	10,844,792 (81.139%)	10,396,253 (80.856%)
	Col Mock	Col Mock	<i>rdd</i> Mock	<i>rdd</i> Mock	Col Fox	Col Fox	<i>rdd</i> Fox	<i>rdd</i> Fox
	3DPI	3DPI	3DPI	3DPI	3DPI	3DPI	3DPI	3DPI
Total reads (adaptors and low quality reads already removed)	13,229,443 (100%)	12,962,009 (100%)	15,548,701 (100%)	13,993,384 (100%)	15,028,978 (100%)	14,946,993 (100%)	13,494,438 (100%)	15,392,677 (100%)
Reads after filtering (18-30nt) (percentage of total reads)	13,034,273 (98.525%)	12,752,367 (98.383%)	15,321,520 (98.383%)	13,708,158 (97.962%)	14,805,886 (98.516%)	14,644,586 (97.977%)	13,283,733 (98.439%)	15,110,887 (98.169%)
Reads mapped to <i>F. oxysporum</i> genome (percentage of trimmed and filtered reads)	1,397 (0.011%)	1,600 (0.013%)	1,975 (0.013%)	1,408 (0.01%)	23,180 (0.157%)	18,191 (0.124%)	26,402 (0.199%)	29,398 (0.195%)
Reads aligning in up to 10 different loci against <i>A. thaliana</i> a (percentage of trimmed and filtered reads)	10,777,885 (82.689%)	10,578,823 (82.956%)	12,523,348 (81.737%)	11,300,685 (82.438%)	12,037,601 (81.303%)	11,818,540 (80.702%)	10,529,214 (79.264%)	12,021,504 (79.555%)

	Col Mock	Col Mock	<i>rdd</i> Mock	<i>rdd</i> Mock	Col Fox	Col Fox	<i>rdd</i> Fox	<i>rdd</i> Fox
	6DPI	6DPI	6DPI	6DPI	6DPI	6DPI	6DPI	6DPI
Total reads (adaptors and low quality reads already removed)	16,205,604 (100%)	14,805,573 (100%)	13,894,018 (100%)	13,891,074 (100%)	14,524,219 (100%)	13,459,553 (100%)	13,841,575 (100%)	14,068,129 (100%)
Reads after filtering (18-30nt) (percentage of total reads)	15,890,526 (98.056%)	14,549,961 (98.274%)	13,729,996 (100%)	13,735,090 (98.877%)	14,270,559 (98.254%)	13,190,307 (98%)	13,506,424 (97.579%)	13,722,658 (97.544%)
Reads mapped to <i>F. oxysporum</i> genome (percentage of trimmed and filtered reads)	2,303 (0.014%)	1,696 (0.012%)	1,408 (0.01%)	1,179 (0.009%)	34,114 (0.239%)	34,067 (0.258%)	66,758 (0.494%)	62,827 (0.458%)
Reads aligning in up to 10 different loci against <i>A. thaliana</i> (percentage of trimmed and filtered reads)	13,076,796 (82.293%)	11,969,412 (82.264%)	11,061,686 (80.566%)	11,178,217 (81.384%)	11,381,057 (79.752%)	10,566,514 (80.108%)	10,204,361 (75.552%)	10,529,559 (77.189%)

Supplementary Figure 4.1: Small RNA size distribution as a percentage of unique small RNA



Supplementary Figure 4.2: Average number of small RNA prior to mapping against *Arabidopsis thaliana* A) number of total small RNA B) number of unique small RNA



Supplementary Table 4.2: Summary of averaged small RNA sequencing read distribution using different parameters across the *Arabidopsis thaliana* genome

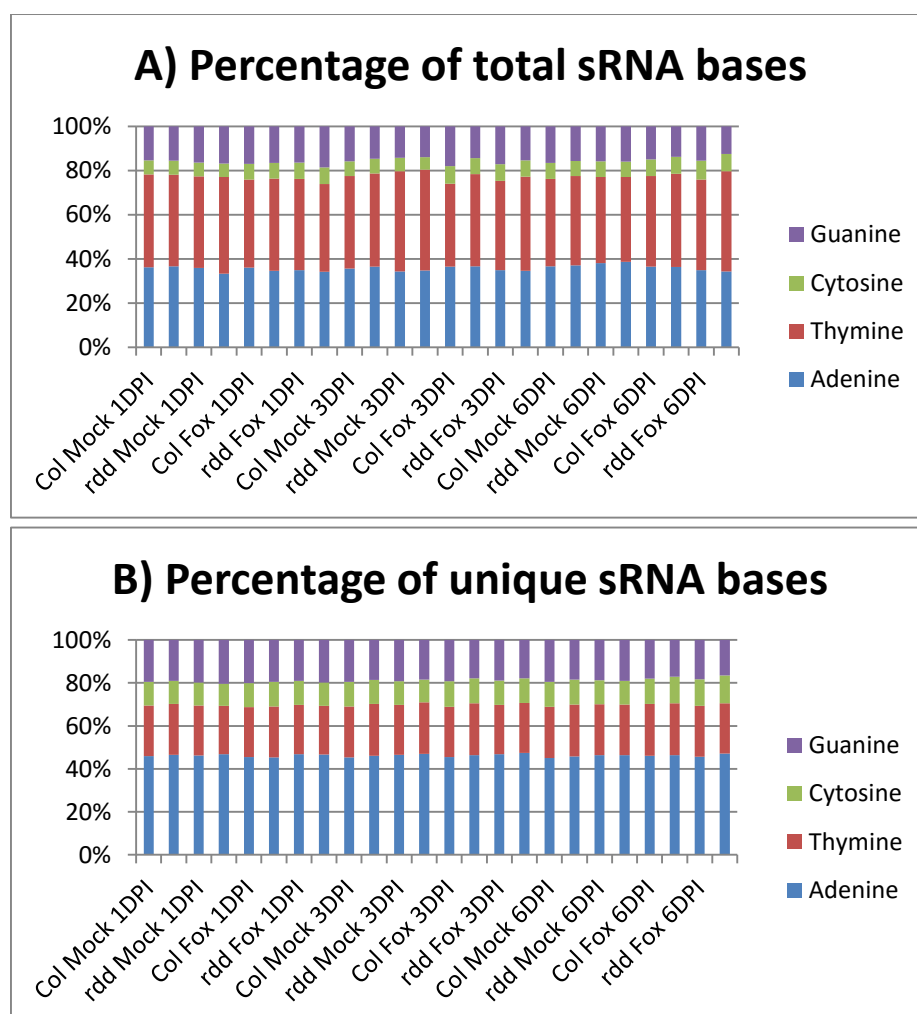
A) Reads aligning in up to 30 different loci against *A. thaliana*

B) Reads aligning in up to 1000 different loci against *A. thaliana* with 1 mismatch allowed per 100bp of read length

DPI; Days post inoculation

	Col Mock	<i>rdd</i> Mock	Col Fox	<i>rdd</i> Fox	Col Mock	<i>rdd</i> Mock	Col Fox	<i>rdd</i> Fox	Col Mock	<i>rdd</i> Mock	Col Fox	<i>rdd</i> Fox
	1DPI	1DPI	1DPI	1DPI	3DPI	3DPI	3DPI	3DPI	6DPI	6DPI	6DPI	6DPI
A	11,544,677 (85.455%)	11,469,645 (84.449%)	11,347,878 (85.348%)	11,116,847 (83.341%)	11,165,311 (85.259%)	12,386,397 (83.855%)	12,609,932 (84.133%)	11,890,101 (82.321%)	13,118,958 (84.607%)	11,550,462 (83.141%)	11,659,640 (83.331%)	10,861,901 (77.836%)
B	13,173,698 (97.513%)	12,874,742 (94.795%)	12,834,125 (96.526%)	12,468,028 (93.471%)	12,707,361 (97.034%)	13,894,541 (94.066%)	14,364,294 (95.838%)	13,368,708 (92.558%)	14,981,782 (96.621%)	13,118,656 (94.429%)	13,290,239 (94.985%)	12,366,786 (88.619%)

Supplementary Figure 4.3: 5' end nucleotide distribution of small RNA A) as a percentage of total small RNA bases B) as a percentage of unique small RNA bases



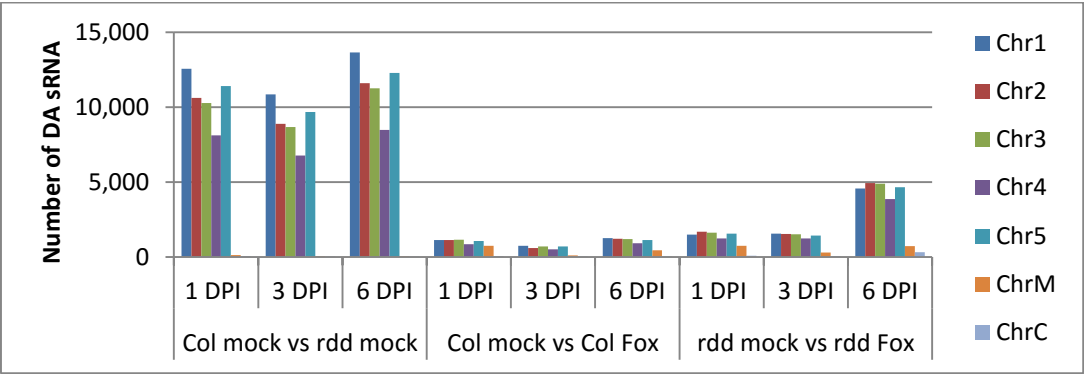
Supplementary Table 4.3: Overlap of differentially abundant (DA) sRNA between different time points, as a percentage of total DA sRNA at each time point

Days post inoculation (DPI)

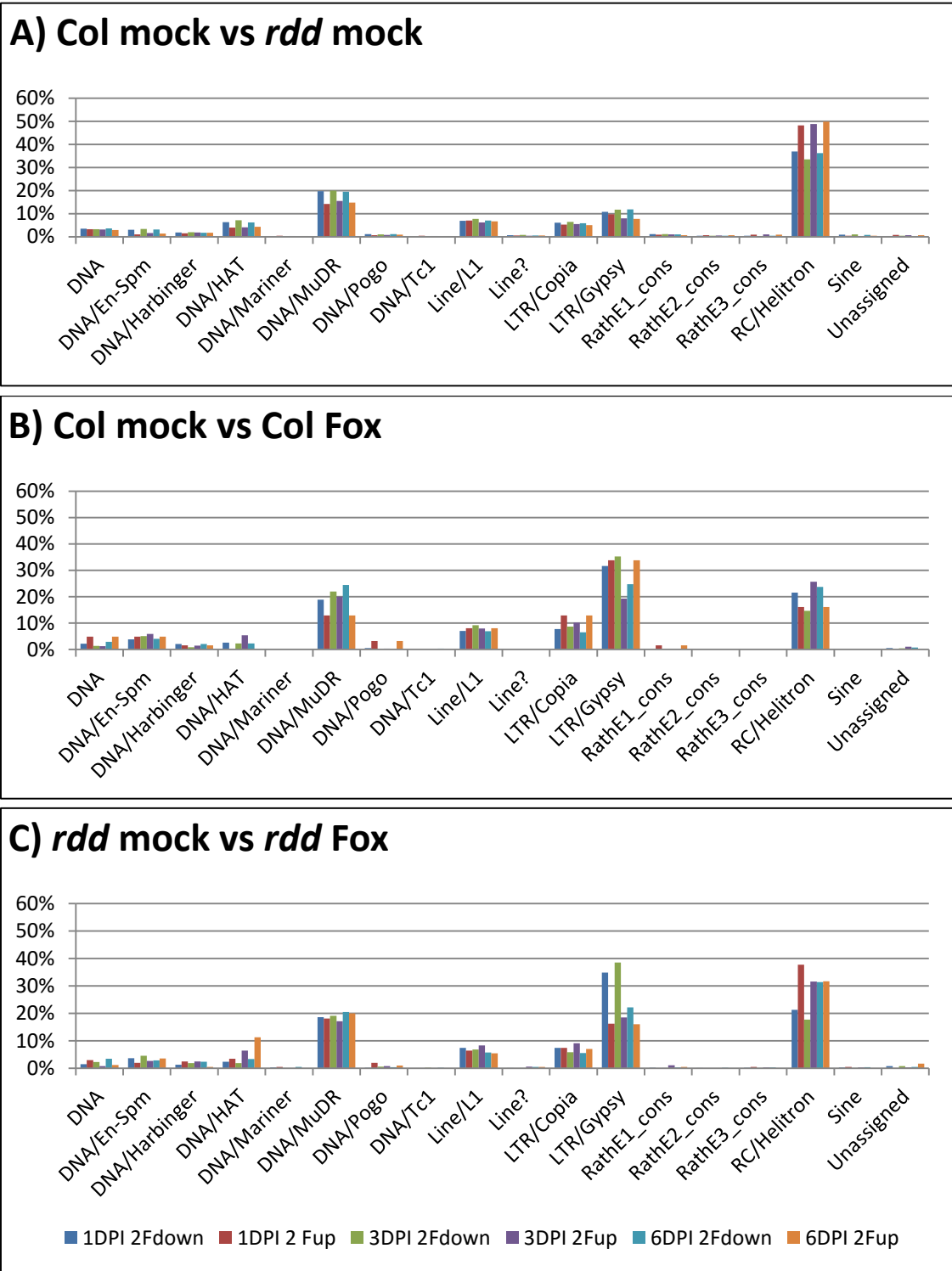
Time point of DA sRNA comparison	DA sRNA at	Genotype / treatment comparison		
		Col mock vs <i>rdd</i> mock	Col mock vs Col Fox	<i>rdd</i> mock vs <i>rdd</i> Fox
1 DPI vs 3 DPI	1 DPI	63%	4%	10%
	3 DPI	74%	8%	11%
3 DPI vs 6 DPI	3 DPI	78%	15%	27%
	6 DPI	61%	8%	8%
1, 3 and 6 DPI combined	1 DPI	58%	2%	4%
	3 DPI	69%	3%	4%
	6 DPI	54%	2%	1%

Supplementary Figure 4.4: DA sRNA distribution across the 5 *Arabidopsis thaliana* chromosomes plus the chloroplast and mitochondrial genomes

The number of DA sRNA with L2FC ≥ 1 or ≤ -1 and P value ≤ 0.05 .



Supplementary Figure 4.5: Distribution of DA sRNA-associated TE superfamilies as a percentage of total DA sRNA-associated TEs



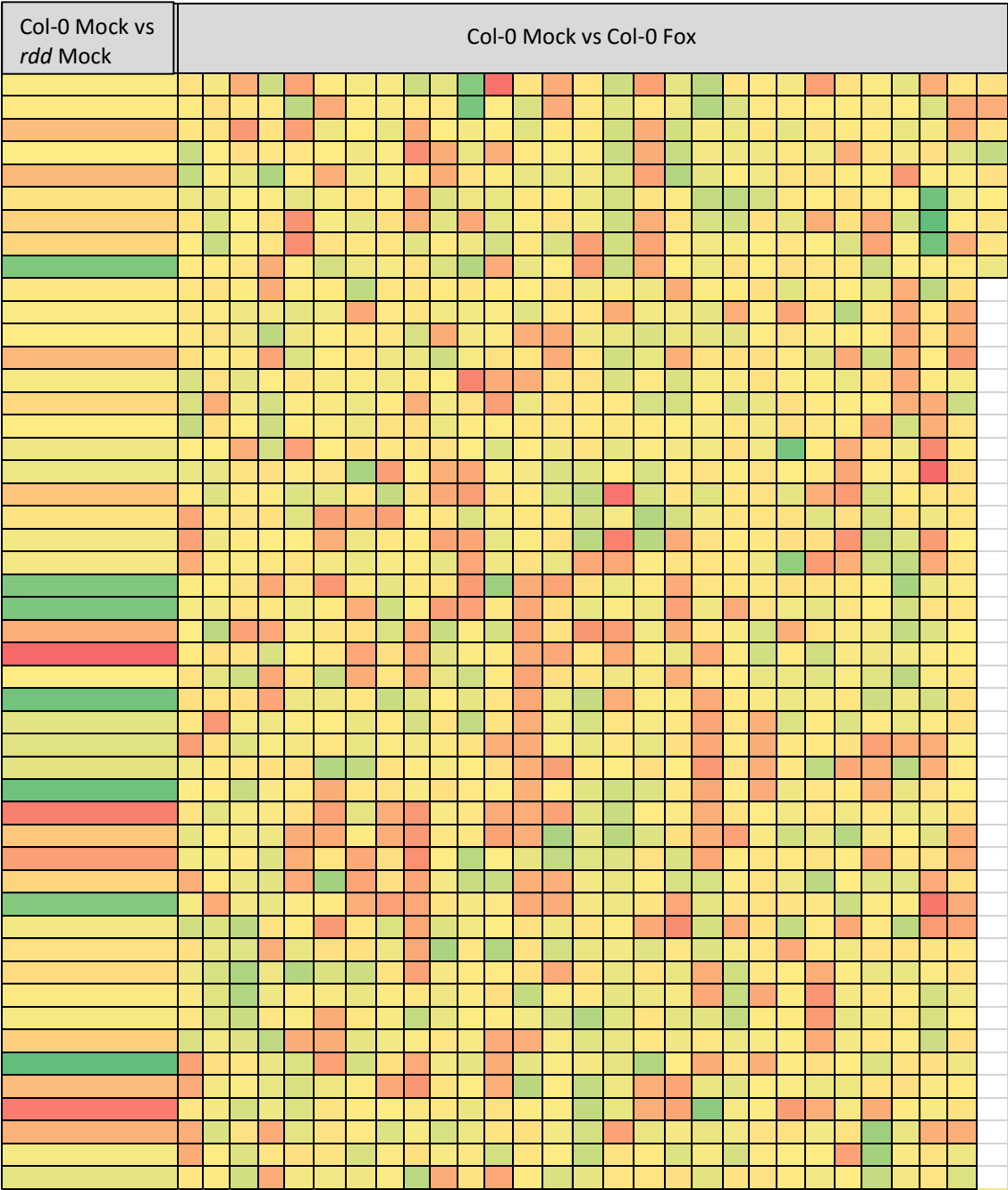
Supplementary Table 5.1: Summary of RNA-seq and mapping results

The numbers in parentheses are the percentage of reads against the number of raw reads.

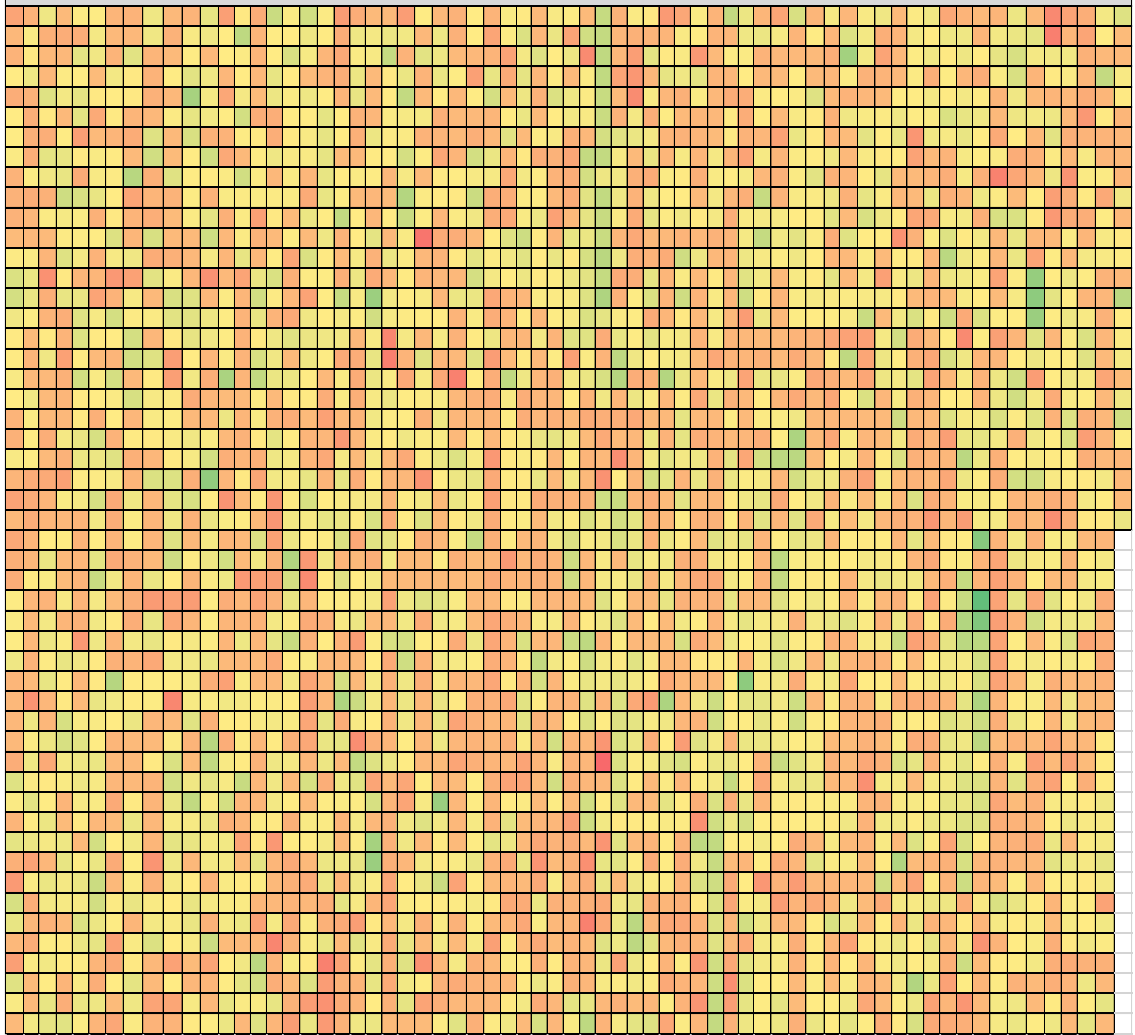
Treatment	Replicate	Number of single end raw reads	Number of unique reads	Number of uniquely reads mapped to <i>A. thaliana</i>
Col mock	1	14,820,249	13,823,822	13,823,182 (93.2%)
	2	15,374,175	14,394,795	14,394,157 (93.6%)
	3	15,454,526	14,468,711	14,468,122 (93.6%)
<i>rdd</i> mock	1	16,397,936	15,269,255	15,268,654 (93.1%)
	2	16,288,062	15,199,908	15,199,342 (93.3%)
	3	16,952,378	15,792,787	15,792,164 (93.2%)
Col Fusarium	1	15,568,786	14,416,253	14,210,531 (91.2%)
	2	16,406,918	15,210,623	14,999,838 (91.4%)
	3	16,401,606	15,158,292	14,842,567 (90.5%)
<i>rdd</i> Fusarium	1	16,758,461	15,589,844	15,354,833 (91.6%)
	2	15,791,083	12,891,287	12,694,328 (80.4%)
	3	17,072,141	15,564,566	14,913,828 (87.4%)

Supplementary Table 5.2: Heatmap analysis of DE genes

Increasing colour from orange (Log2 fold change value of 2) to red (Log2 fold change value of 6) indicate gene downregulation. Increasing colour from yellow (Log 2 fold change value of -2) to green (Log 2 fold change value of -6) indicates gene upregulation.

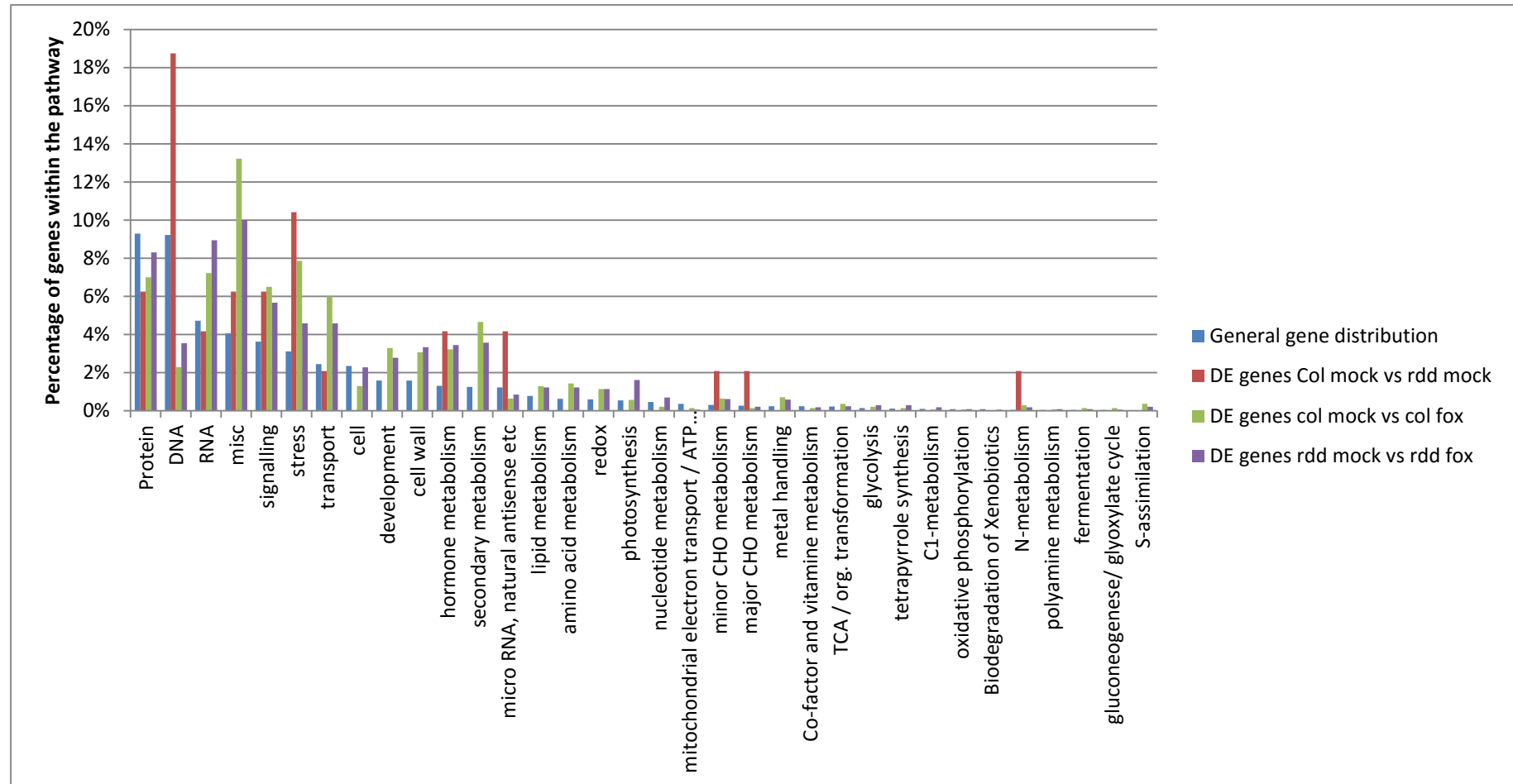


rdd Mock vs rdd Fox



Supplementary Figure 5.1: Mapman Gene Ontology for differentially expressed genes at 3 DPI

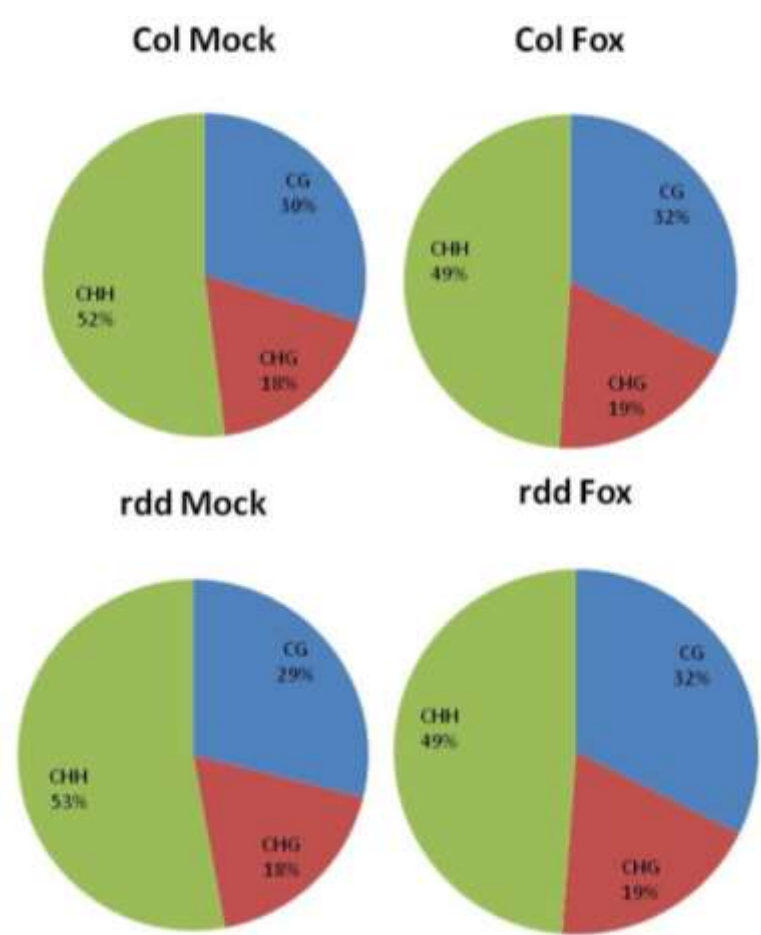
The histograms show the percentage of DA sRNA associated genes of a particular pathway in differentially expressed genes. The percentage of all genes of a particular pathway in total Arabidopsis genes is indicated by the blue bar on the left for comparison. The list of genes that are part of each pathway is available from Thimm et al., 2004.



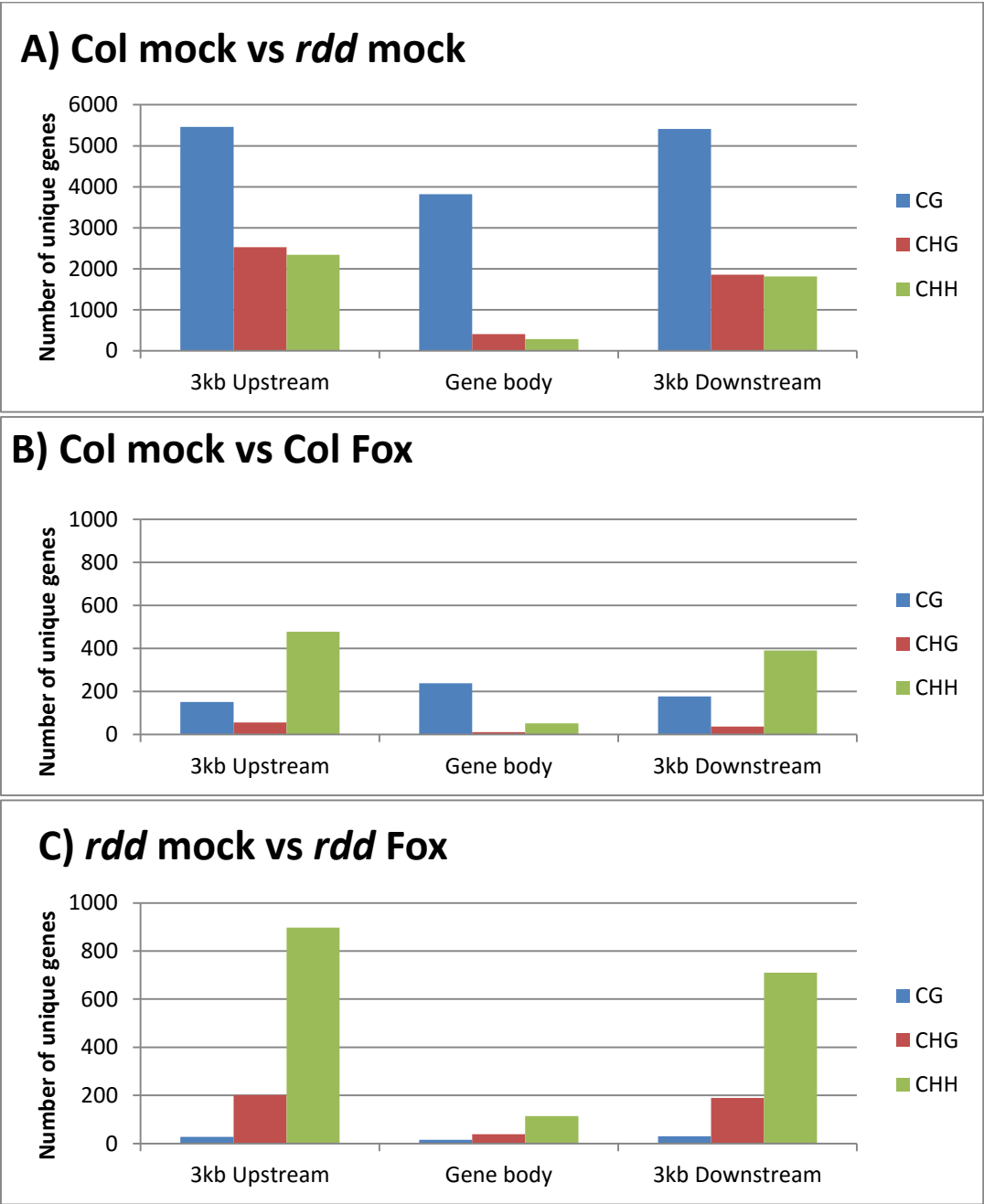
Supplementary Table 6.1: Summary of whole genome bisulfite sequencing reads of 3 DPI samples

Sample	Number of reads	Uniquely aligned reads to <i>Arabidopsis thaliana</i>	The number of cytosines covered as a proportion of the total cytosines present in <i>Arabidopsis thaliana</i>
Columbia Mock	32,705,946	25,507,248	96%
Columbia Fusarium	29,501,104	19,925,071	95%
<i>rdd</i> Mock	36,957,919	29,574,380	96%
<i>rdd</i> Fusarium	29,645,007	20,055,251	95%

Supplementary Figure 6.1: Individual methylated cytosine sites as a percentage of total methylated cytosines



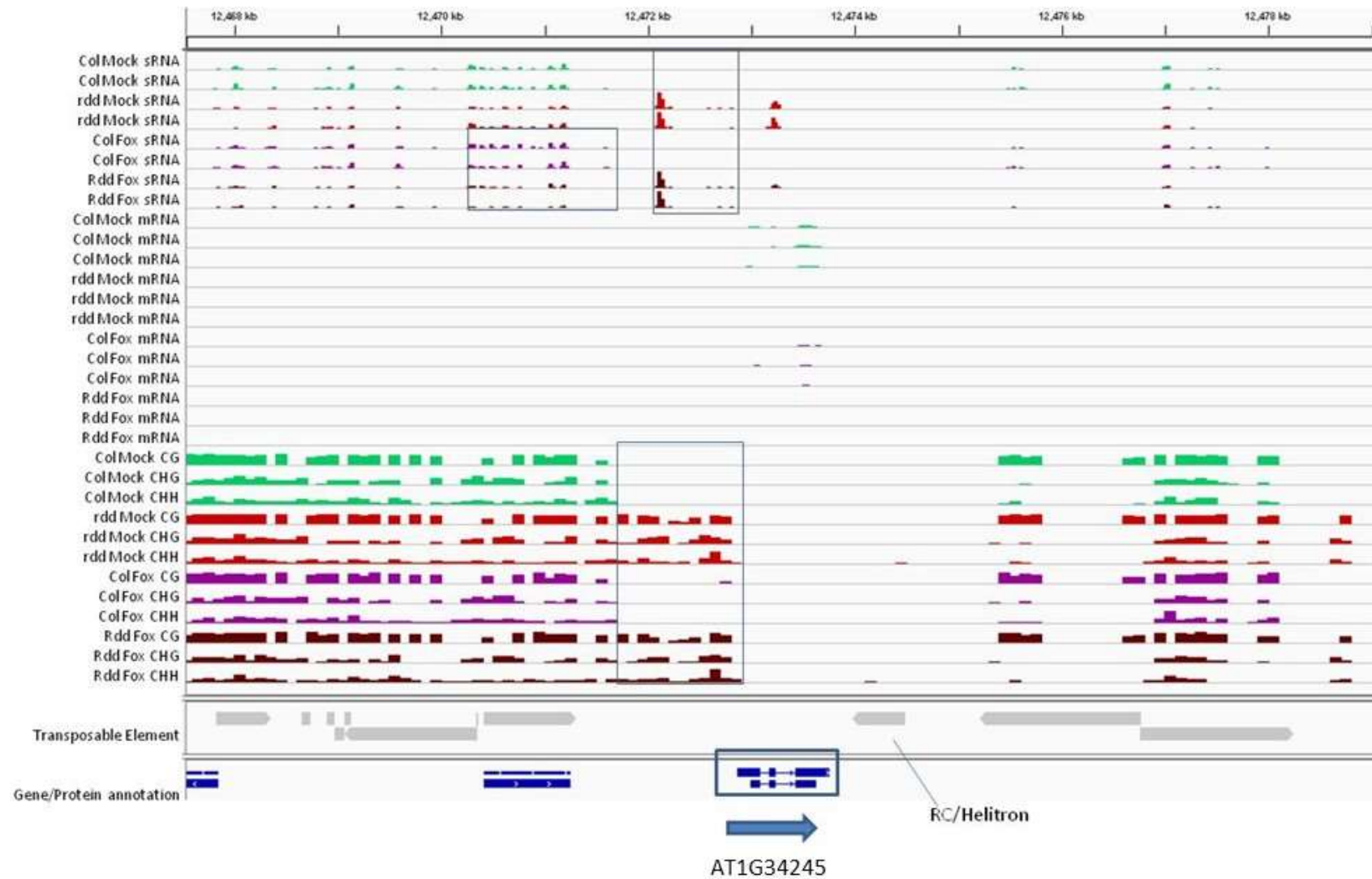
Supplementary Figure 6.2: Gene location of differentially methylated regions



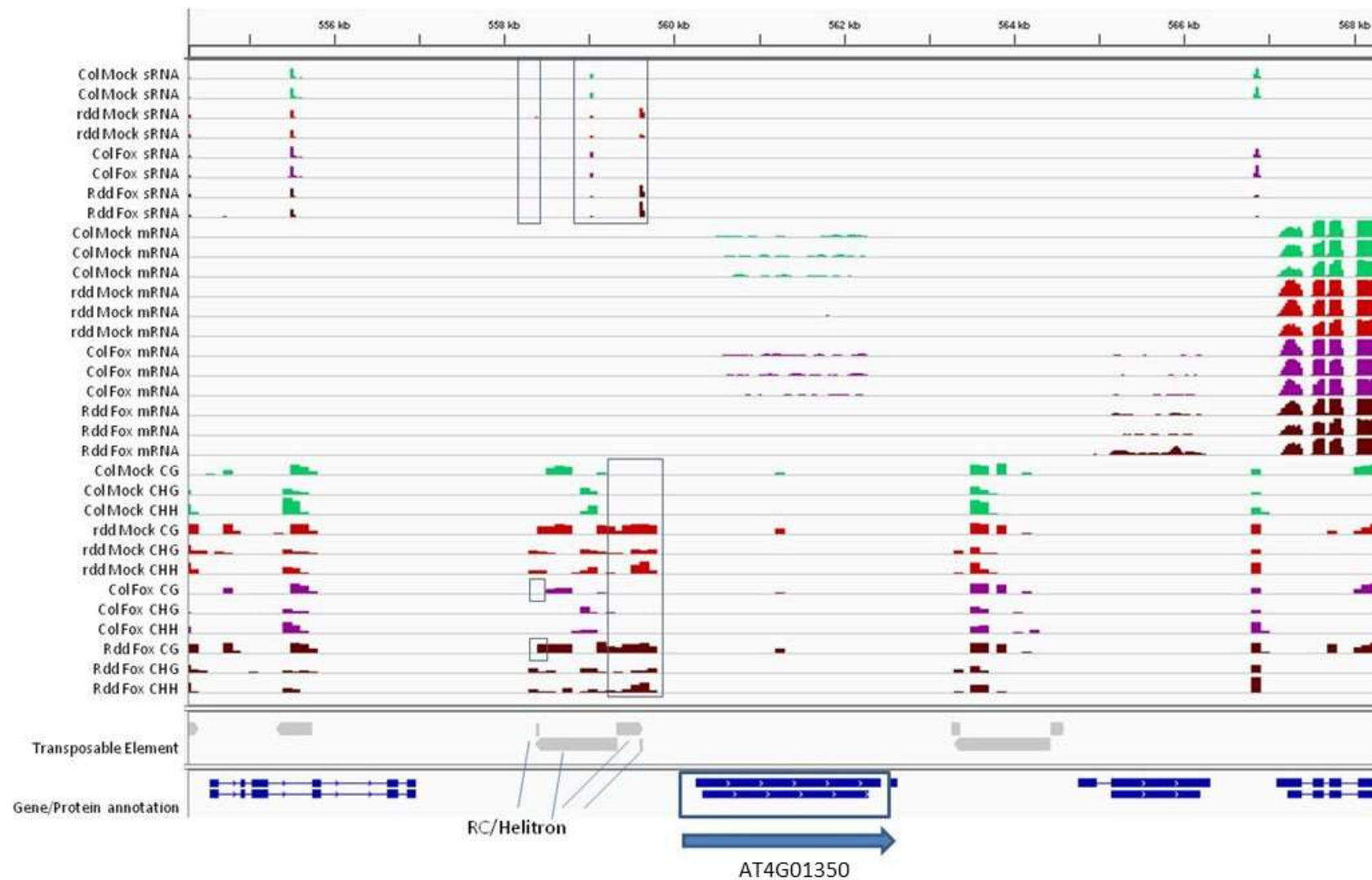
Supplementary Table 6.2: Gene ontology of the DE genes associated with DA sRNA and DMR

Col mock vs rdd mock		Col mock vs Col Fox		rdd mock vs rdd Fox	
Gene	Pathway	Gene	Pathway	Gene	Pathway
AT3G30720	major CHO metabolism	AT1G48510	mitochondrial	AT1G42970	photosynthesis
AT5G36870	minor CHO metabolism	AT5G36910	stress	AT1G60450	minor CHO metabolism
AT5G16570	N-metabolism	AT1G74430	RNA	AT5G10100	minor CHO metabolism
AT2G33830	hormone metabolism	AT1G52410	signalling	AT4G09030	cell wall
AT2G14900	hormone metabolism	AT5G03960	signalling	AT1G26240	cell wall
AT2G15042	stress	AT4G03450	cell	AT5G26670	cell wall
AT1G58602	stress	AT3G30730	not assigned	AT5G28237	amino acid metabolism
AT3G27940	RNA	AT1G52000	not assigned	AT5G18060	hormone metabolism
AT3G09160	RNA			AT2G38210	hormone metabolism
AT3G54720	protein			AT4G10457	stress
AT5G38280	signalling			AT3G13404	stress
AT4G04570	signalling			AT1G79690	nucleotide metabolism
AT2G43290	signalling			AT4G34730	RNA
AT4G01350	not assigned			AT2G45100	RNA
AT2G05915	not assigned			AT4G17810	RNA
AT3G60520	not assigned			AT4G01580	RNA
AT5G35732	not assigned			AT4G04830	RNA
AT2G34655	not assigned			AT4G04810	RNA
AT2G04795	not assigned			AT5G38435	DNA
AT5G25120	not assigned			AT5G35970	DNA
AT5G52320	not assigned			AT2G24200	protein
AT1G51470	not assigned			AT3G16550	protein
AT3G09162	not assigned			AT2G38970	protein
AT2G13550	not assigned			AT5G38590	protein
AT1G34245	not assigned			AT1G26230	protein
				AT2G27060	signalling
				AT1G79680	signalling
				AT5G66790	signalling
				AT2G28490	development
				AT2G16660	development
				AT1G03820	not assigned
				AT1G42980	not assigned
				AT1G54095	not assigned
				AT3G45200	not assigned
				AT5G46770	not assigned
				AT5G54400	not assigned
				AT2G14560	not assigned
				AT3G42725	not assigned
				AT3G49210	not assigned
				AT3G29575	not assigned
				AT2G05185	not assigned
				AT2G26530	not assigned
				AT2G37940	not assigned
				AT2G01340	not assigned
				AT5G35640	not assigned
				AT1G31710	not assigned
				AT5G06720	not assigned
				AT2G01890	not assigned
				AT2G01880	not assigned
				AT1G52030	not assigned
				AT5G38550	not assigned
				AT3G22142	not assigned

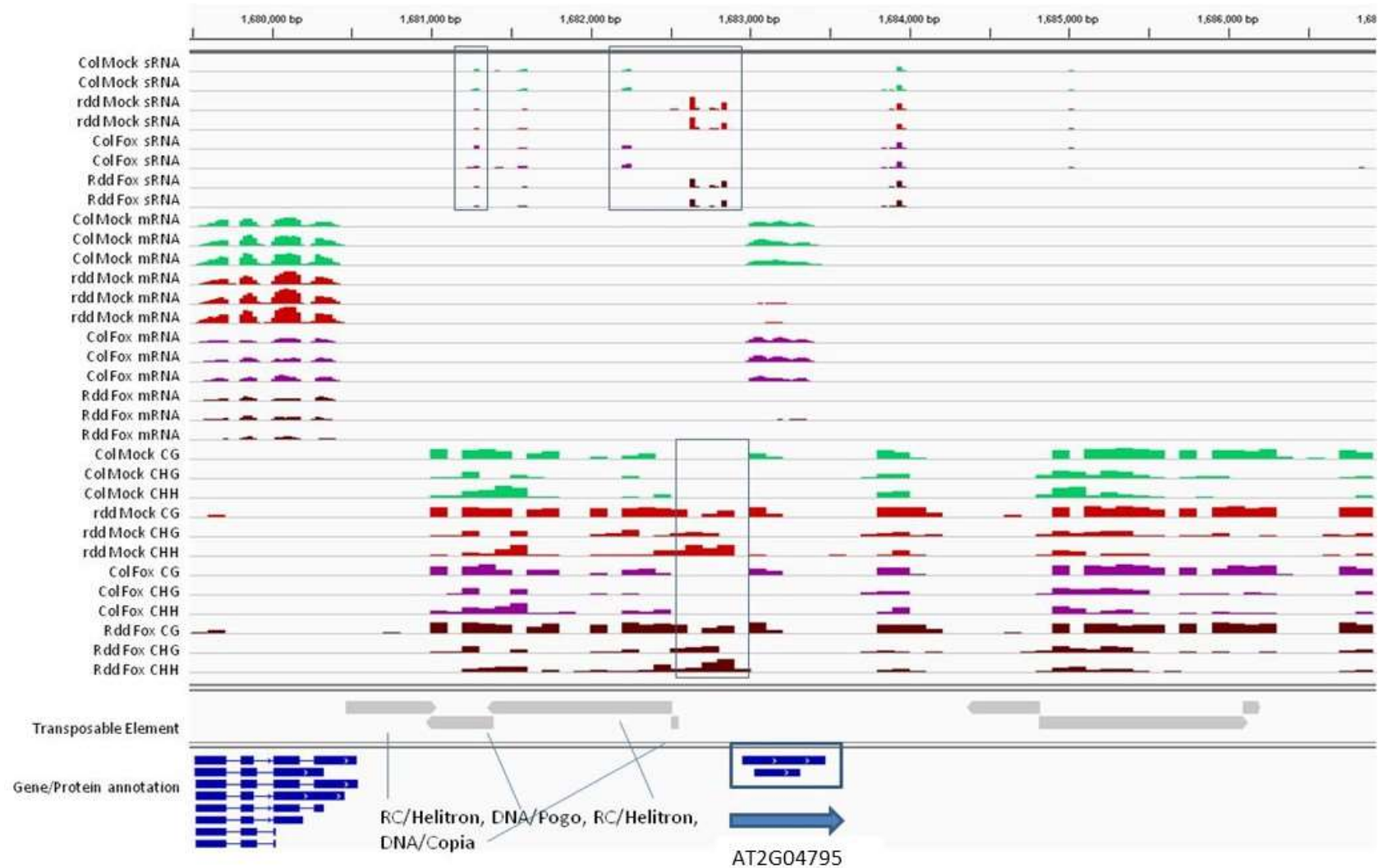
Supplementary Figure 6.3: IGV visualization of DA sRNA and DMR in AT1G34245 (epidermal patterning factor 2 of stomata/guard cell gene) downregulated in *rdd*



Supplementary Figure 6.4: IGV visualization of DA sRNA and DMR in AT4G01350 (cysteine/histidine-rich C1 domain family protein involved in chitin response) downregulated in *rdd*



Supplementary Figure 6.5: IGV visualization of DA sRNA and DMR in AT2G04795 (an unknown protein involved in oxidative stress) downregulated in *rdd*



Supplementary Table 6.3: Heatmap analysis of DE genes with DA sRNA and DMR in mock vs Fusarium comparisons

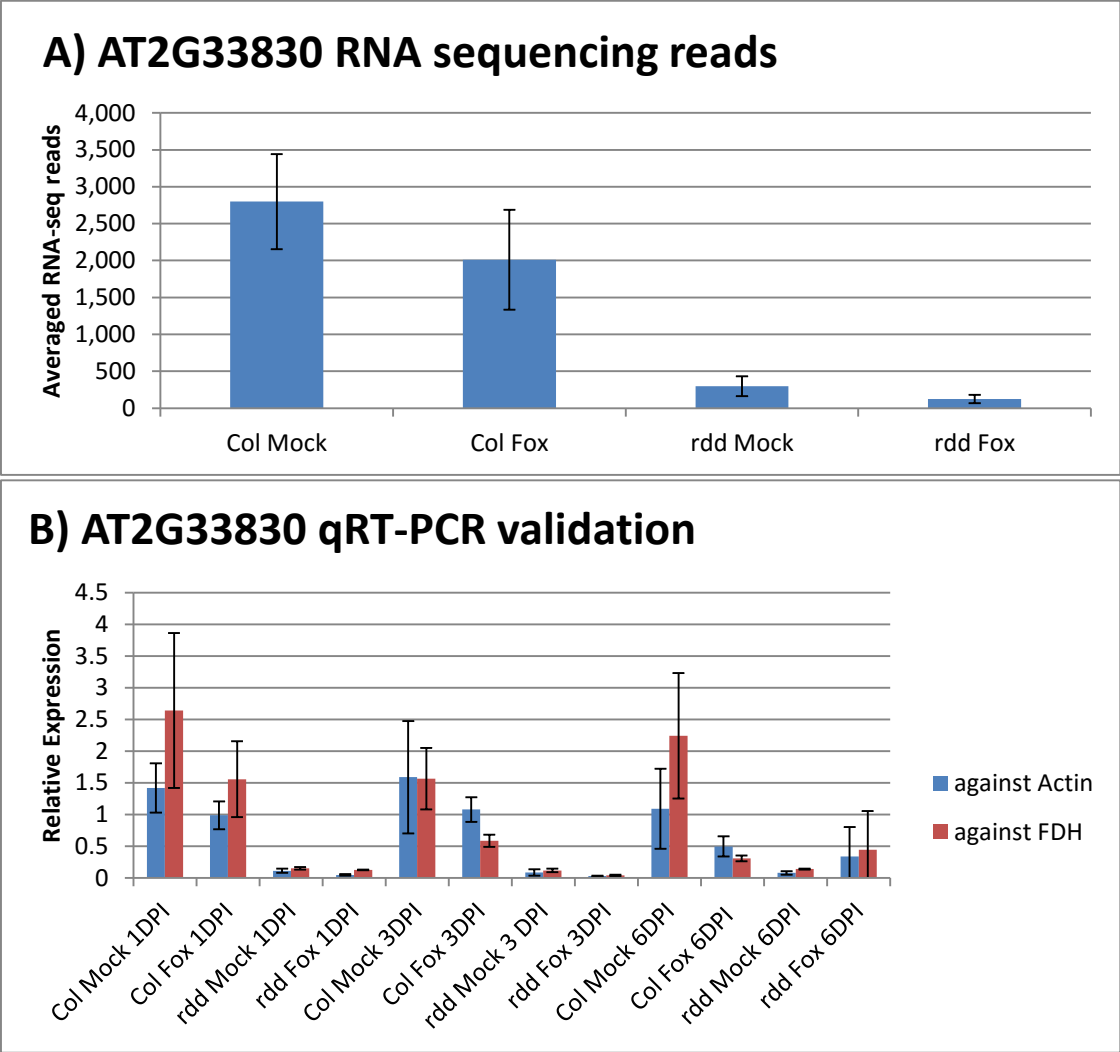
Increasing colour from orange (Log2 fold change value of 2) to red (Log2 fold change value of 6) indicate downregulation in *rdd* compared to Col-0. Increasing colour from yellow to green indicates upregulation in *rdd* with the log2 fold change value from 1 to 6. The two halves in the sRNA or DMR cells indicate two different sRNA clusters or DMRs with different log 2 fold changes within the 3kb flanking upstream of the gene. White space indicates that no changes were found for this region.

	Gene	DA sRNA	DE mRNA	DMR - CG	DMR - CHG	DMR - CHH
3kb Upstream Col mock vs Col Fox	AT1G52000					
	AT1G74430					
	AT4G03450					
Gene body Col mock vs Col Fox	AT1G52410					
3kb Downstream Col mock vs Col Fox	AT1G48510					
	AT3G30730					
	AT5G03960					
	AT5G36910					
3kb Upstream <i>rdd</i> mock vs <i>rdd</i> Fox	AT2G38970					
	AT5G35640					
	AT1G26230					
	AT1G79680					
	AT1G79690					
	AT5G35970					
	AT1G42980					
	AT2G01880					
	AT2G01340					
	AT2G16660					
	AT2G26530					
	AT2G28490					
	AT2G38210					
	AT3G13404					
	AT3G42725					
	AT3G45200					
	AT3G49210					
	AT4G01580					
	AT4G04830					
	AT4G09030					
	AT5G38435					
	AT5G38550					
	AT5G38590					
	AT2G24200					
	AT5G06720					
	AT1G31710					
	AT3G22142					

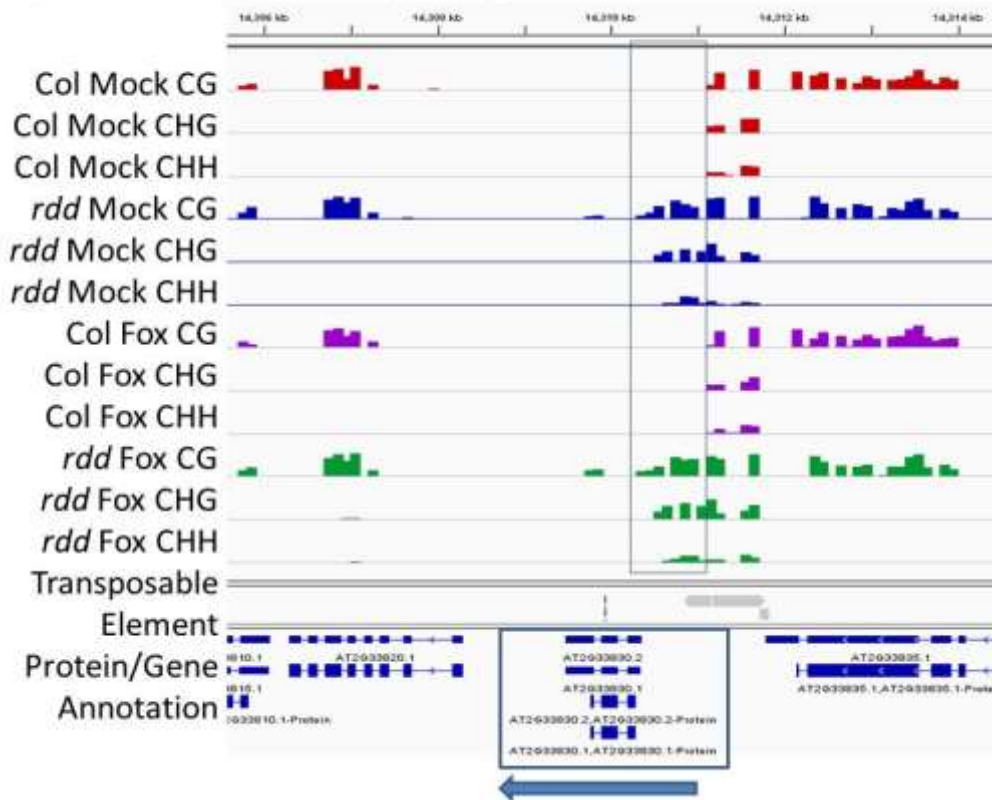
3kb Downstream <i>rdd</i> mock vs <i>rdd</i> Fox	AT4G34730					
	AT5G10100					
	AT1G03820					
	AT1G26240					
	AT1G54095					
	AT1G60450					
	AT4G10457					
	AT5G06730					
	AT1G42970					
	AT1G52030					
	AT2G01890					
	AT2G05185					
	AT2G14560					
	AT2G27060					
	AT2G37940					
	AT2G38210					
	AT2G45100					
	AT3G16550					
	AT3G29575					
	AT4G04810					
	AT4G04830					
	AT4G17810					
	AT5G18060					
	AT5G26670					
	AT5G28237					
	AT5G46770					
	AT5G54400					
	AT5G66790					

Supplementary Figure 6.6: Validation of gene expression using qRT-PCR and comparison of DNA methylation pattern with published data for AT2G33830, a DE gene in the Col-0 vs *rdd* comparison associated with DA sRNA and DMR in the upstream

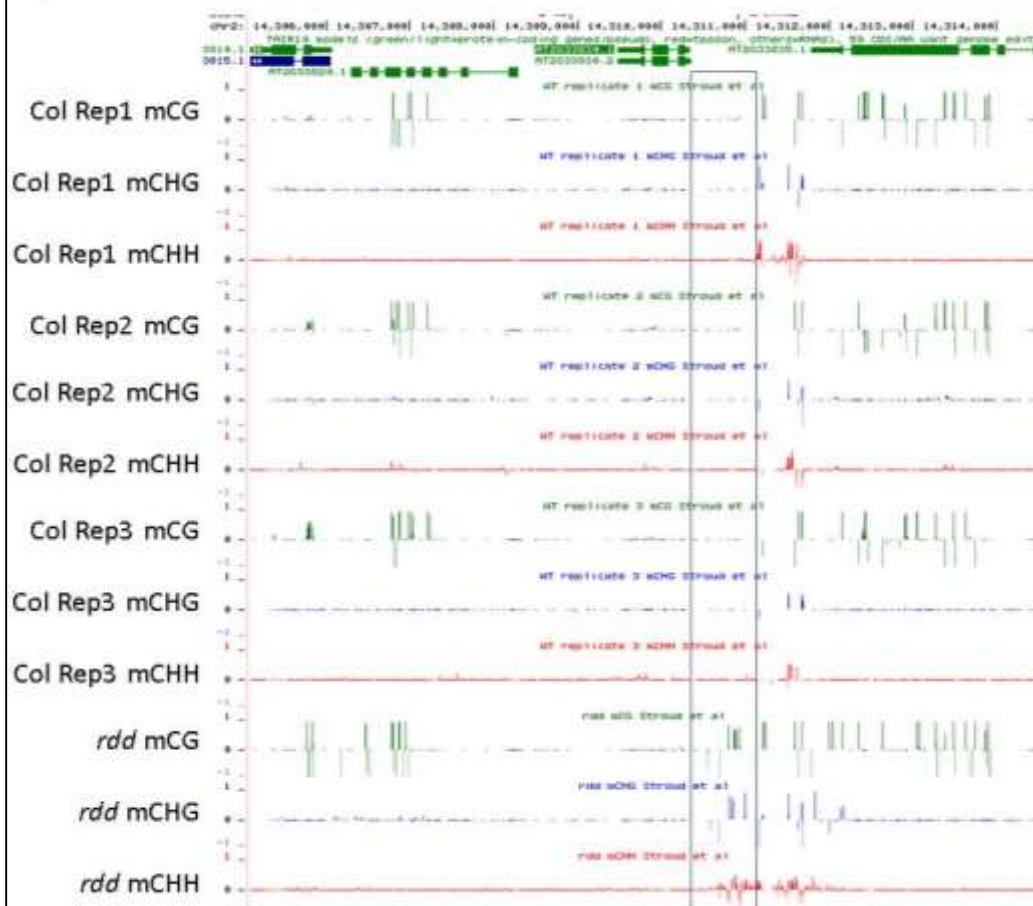
- A) RNA-seq reads averaged across three biological replicates with standard deviation.
- B) qRT-PCR validation averaged across two biological replicates with standard deviation.
- C) IGV visualisation of bisulfite sequencing reads with differentially methylated region boxed
- D) Bisulfite sequencing data from Stroud et al. 2013



C) AT2G33830 Bisulfite sequencing reads

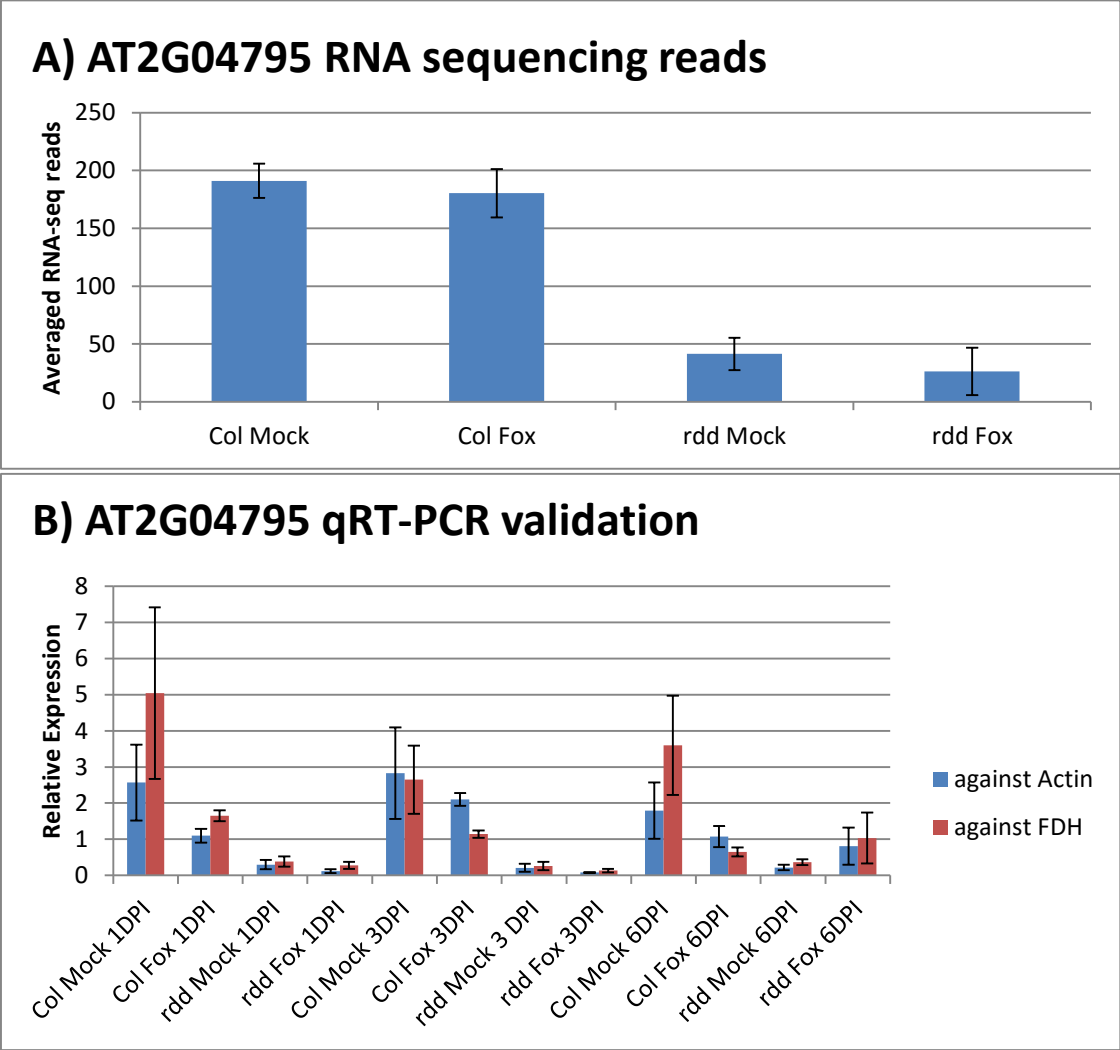


D) AT2G33830 Stroud dataset validation

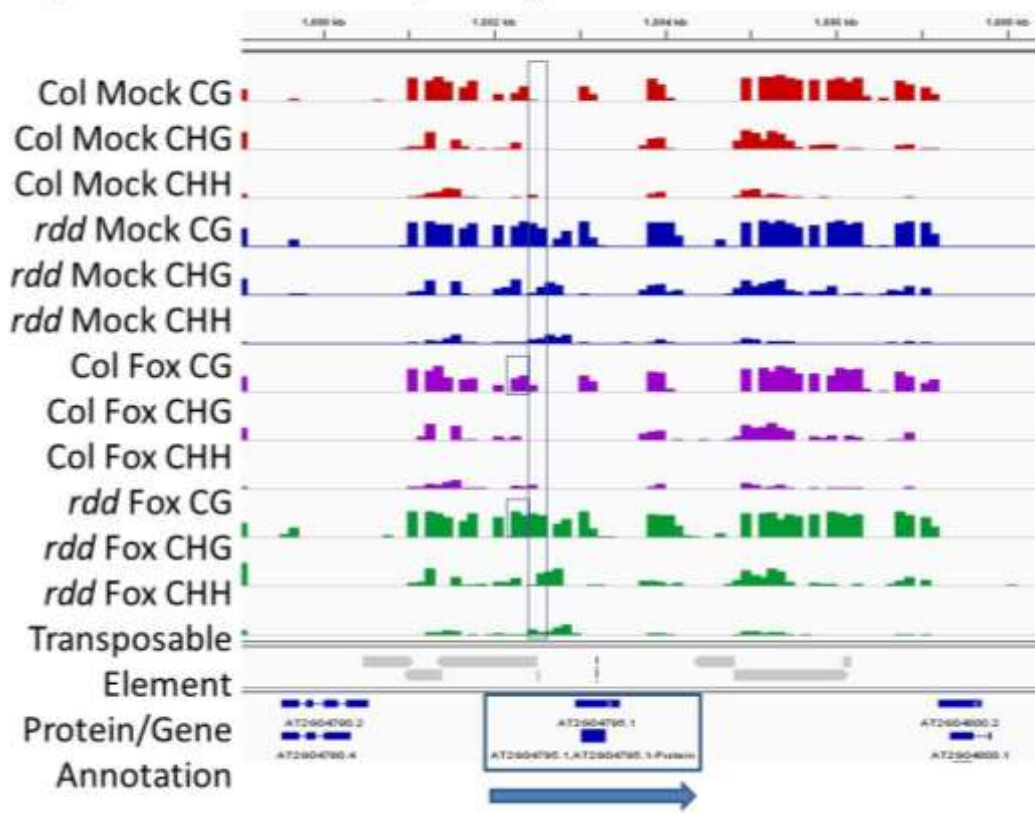


Supplementary Figure 6.7: Validation of gene expression using qRT-PCR and comparison of DNA methylation pattern with published data for AT2G04795, a DE gene in the Col-0 vs *rdd* comparison (under both mock and *Fusarium* infection conditions) associated with DA sRNA and DMR in the upstream

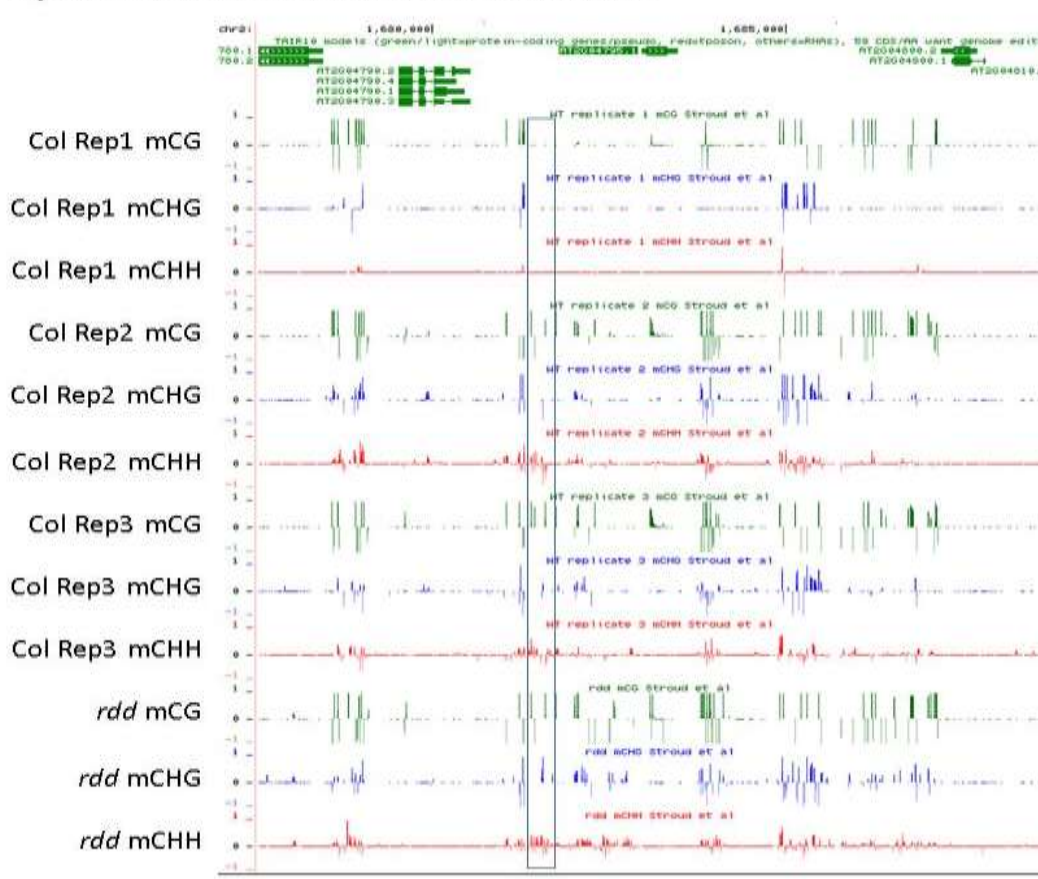
- A) RNA-seq reads averaged across three biological replicates with standard deviation.
- B) qRT-PCR validation averaged across two biological replicates with standard deviation.
- C) IGV visualisation of bisulfite sequencing reads with differentially methylated region boxed
- D) Bisulfite sequencing data from Stroud et al. 2013



C) AT2G04795 Bisulfite sequencing reads

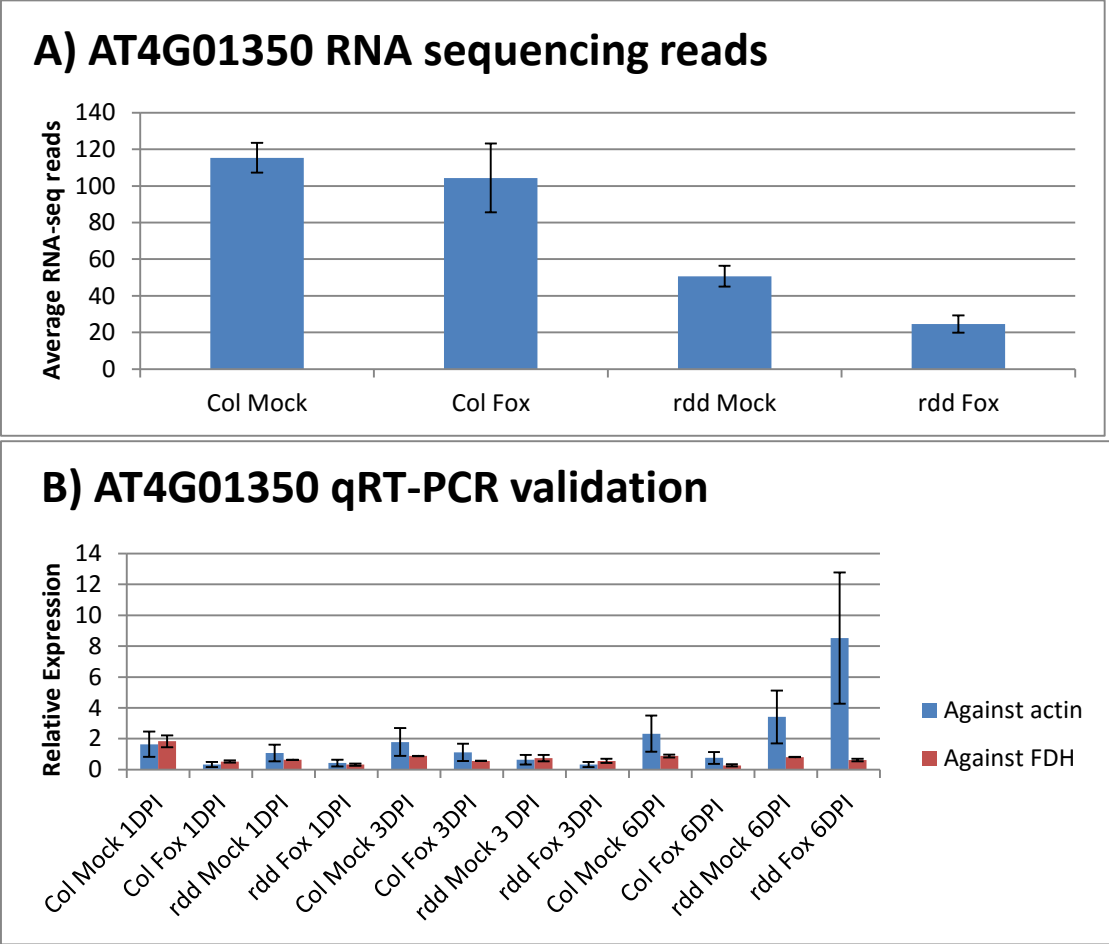


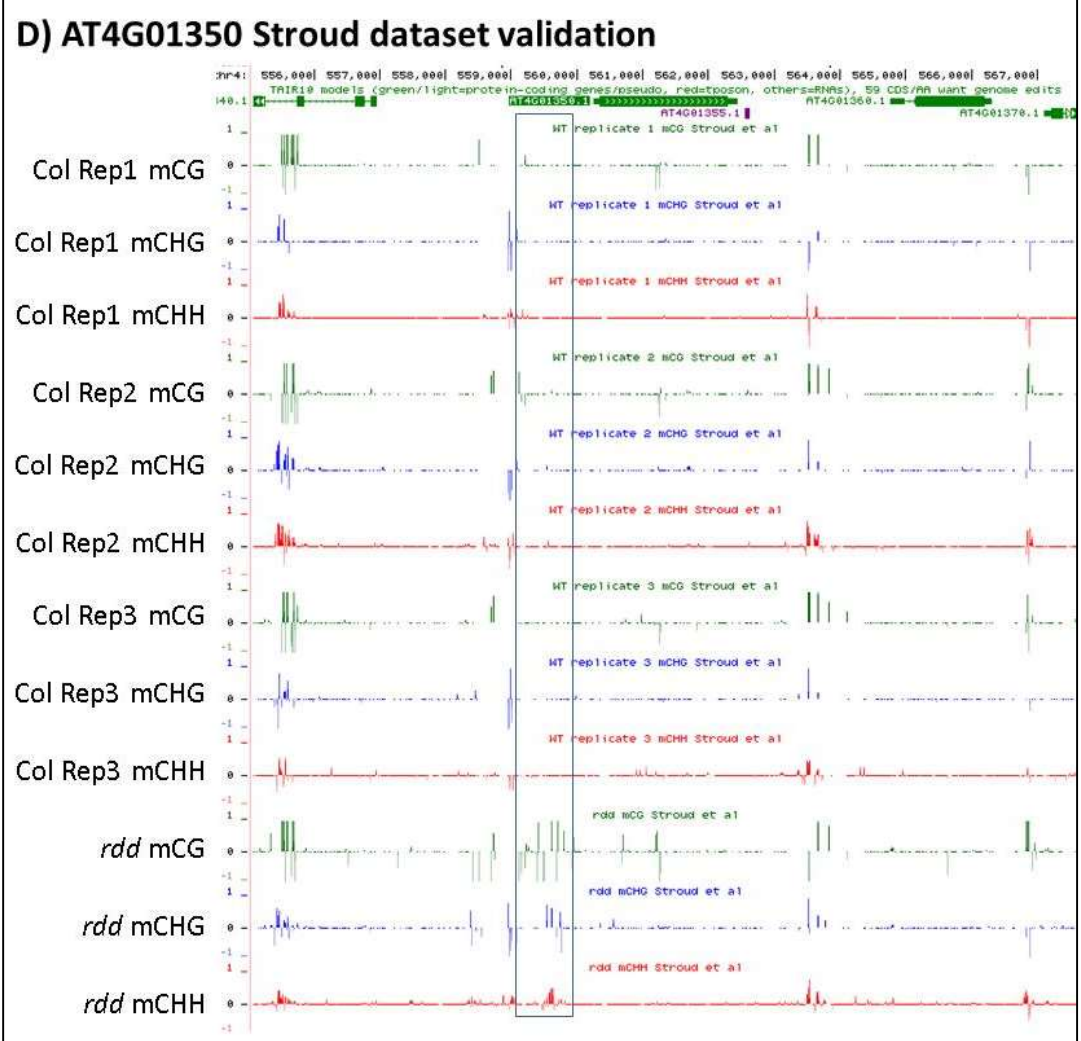
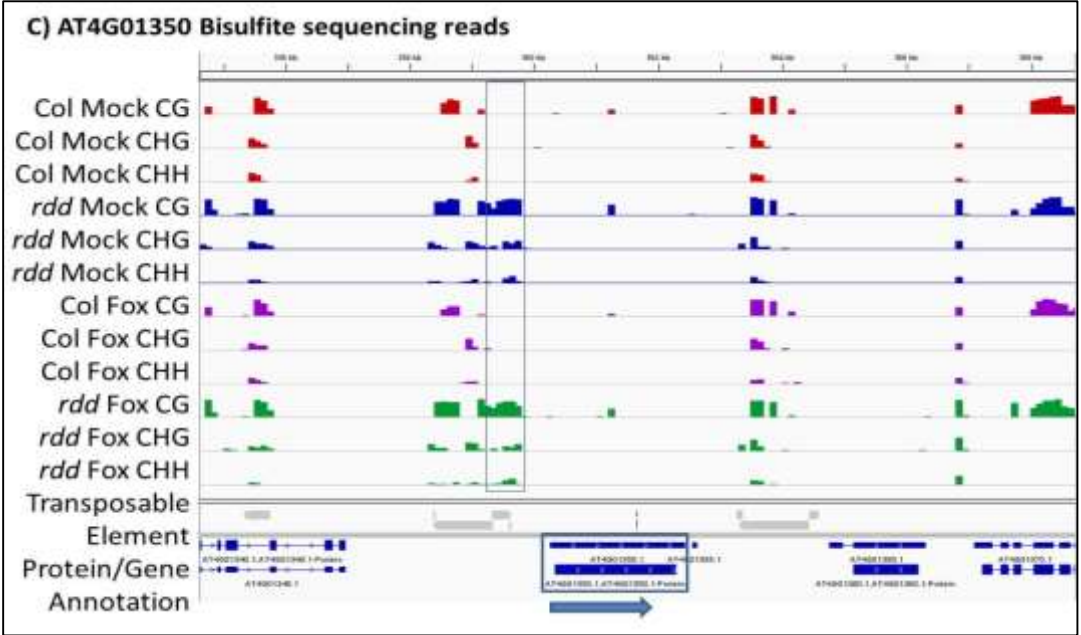
D) AT2G04795 Stroud dataset validation



Supplementary Figure 6.8: Validation of gene expression using qRT-PCR and comparison of DNA methylation pattern with published data for AT4G01350, a DE gene in the Col-0 vs *rdd* comparison (under both mock and *Fusarium* infection conditions) associated with DA sRNA and DMR in the upstream

- A) RNA-seq reads averaged across three biological replicates with standard deviation.
- B) qRT-PCR validation averaged across two biological replicates with standard deviation.
- C) IGV visualisation of bisulfite sequencing reads with differentially methylated region boxed
- D) Bisulfite sequencing data from Stroud et al. 2013





Supplementary Figure 6.9: Validation of gene expression using qRT-PCR and comparison of DNA methylation pattern with published data for AT1G34245, a DE gene in the Col-0 vs *rdd* comparison (under both mock and *Fusarium* infection conditions) associated with DA sRNA and DMR in the upstream

- A) RNA-seq reads averaged across three biological replicates with standard deviation.
- B) qRT-PCR validation averaged across two biological replicates with standard deviation.
- C) IGV visualisation of bisulfite sequencing reads with differentially methylated region boxed
- D) Bisulfite sequencing data from Stroud et al. 2013

