Investigating locus-specific effects on transgene silencing and DNA methylation in *Arabidopsis thaliana*

Eleanor Joy Walton Bsc.

Doctor of Philosophy

University of York

Biology

January 2015

Abstract

The repression of transgene expression by RNA silencing can affect the development of genetically modified plants for both research and agronomical purposes. It remains unclear why some transgenes when introduced into plants undergo silencing whereas other identical transgenes do not. Additionally, some transgenes when silenced can support DNA methylation whereas others cannot. One possibility for this variation is that there are locus-specific effects influencing a transgene's potential to undergo RNA silencing and/or DNA methylation. These locus-specific effects may be a result of either the site of transgene integration or be due to the properties of the transgene locus itself.

This research has aimed to address these questions by generating and characterizing multiple independent single locus GFP transgenic Arabidopsis lines that have been triggered to undergo silencing by crossing with an amplicon trigger line (AMP243). The ability of these lines to undergo RDR6-dependent Post Transcriptional Gene Silencing (PTGS) and/or RNA directed DNA Methylation (RdDM) has been investigated in an attempt to correlate characteristics of the transgene in each independent line with RNA silencing outcomes. Additionally, epitope-tagged versions of RDR2 and RDR6 have been generated in order to investigate their potential recruitment to target loci.

The data presented in this body of work supports there being no locus-specific effect on RNA silencing. For example, amplicon-triggered RDR6-dependent PTGS occurred in all the independent lines suggesting that RDR6 activity is not influenced by either genome location or locus structure. Despite the production of high levels of small interfering RNAs, not all lines supported detectable levels of RdDM in the F1 of the crosses with the trigger line. However, the results presented here clearly demonstrate that the levels of RdDM increase over a generation.

Additionally, Chromatin Immunoprecipitation experiments have revealed that amplicon-triggered RdDM was not associated with changes in the histone marks H3Ac or H3K9me. This suggests that this non-canonical pathway of DNA methylation may not be linked with changes in histone modification, although testing will be required in order to strengthen this conclusion.

Table of Contents

Abstract....2
Table of Contents....3
List of Figures....9
List of Tables....11
Acknowledgments....12
Author's Declaration....13
Chapter 1 Introduction....14
1.1 The Variety of RNA silencing pathways....15
1.1.1 RNA silencing is involved in a variety of functions....15
1.1.2 The core elements of RNA silencing....15
1.1.3 The different classes of small RNA....16
1.2 Post-transcriptional Gene Silencing....17
1.2.1 Trans-acting siRNA biogenesis....18
1.2.2 Amplification of Post-transcriptional Gene Silencing by RDR6....21
1.2.3 RDR6 template recognition....21
1.3 RNA silencing and DNA methylation....22
1.3.1 DNA methylation and its maintenance....22
1.3.2 RNA directed DNA Methylation....24
1.3.3 Non-canonical RdDM pathways....28
1.3.4 Silencing of Transposable elements....29
1.4 Overlap between PTGS and DNA methylation....29
1.5 Factors influencing transgene silencing....31
1.5.1 Locus specific effects....32
1.5.2 Transcript threshold....33
1.5.3 Matrix scaffold attachment regions....33

1.6 Project aims....33
Chapter 2 Materials and Methods....35
2.1 Nucleic acid extraction and manipulation....35
2.1.1 Arabidopsis genomic DNA extraction....35
2.1.2 Plasmid DNA extraction....35
2.1.3 DNA extraction from Agarose gel....36
2.1.4 RNA extraction....36
2.1.5 Production of cDNA....36
2.1.6 Precipitation of DNA....36
2.2 PCR....37
2.2.1 Primers....37
2.2.2 Taq PCR....40
2.2.3 Variations of the Taq PCR....40
2.2.4 qPCR....40
2.2.5 Thermal Asymmetric Interlaced PCR (TAIL PCR)....41
2.3 Manipulation of Bacteria and plasmids....43
2.3.1 Bacterial growth....43
2.3.2 Cloning....43
2.3.3 Bacterial transformation....43
2.4 Propagation and manipulation of Plants....45
2.4.1 *Arabidopsis thaliana* lines....45
2.4.2 Growth media....45
2.4.3 Chlorine gas sterilization of seeds....45
2.4.4 Plant growth conditions....45
2.4.5 Transforming Arabidopsis by floral dip....46
2.4.6 Crossing Plants....46
2.4.7 Infiltration of *Nicotiana benthamiana*....46

2.5 Blotting and detection of nucleic acids 47
2.5.1 Southern blot 47
2.5.2 Northern blot of mRNA 47
2.5.3 Northern blot optimised for small RNA detection 48
2.5.4 Detection of nucleic acid by probing blots 48
2.6 Sequencing 50
2.6.1 Sanger sequencing 50
2.6.2 Bisulfite sequencing 50
2.7 Chromatin Immunoprecipitation (ChIP) 51
2.7.1 Solutions required 51
2.7.2 Chromatin crosslinking 52
2.7.3 Chromatin preparation 52
2.7.4 Pre-clearing and immune precipitation 53
2.7.5 Collection, washes and elution of immune complexes 53
2.7.6 Reverse crosslinking 53
2.7.7 DNA clean up 53
2.8 Protein detection by western blot 54
2.8.1 Solutions required for western blot 54
2.8.2 Protein extraction and separation 55
2.8.3 Transfer 55
2.8.4 Blocking 56
2.8.5 Primary antibody incubation 56
2.8.6 Secondary antibody incubation 56
2.8.7 Western blot detection using chemiluminescence 56
2.9 Imaging GFP 56
2.9.1 Hand-held UV lamp 56
2.9.2 Fluorescence microscopy 56

2.9.3 Confocal microscopy 57

Chapter 3 Genomic location does not affect amplicon-triggered RNA silencing or RdDM of a transgene 58

3.1 Introduction 58

3.2 Results 60

3.2.1 Creating homozygous single locus GFP transgenic lines 60

3.2.2 Silencing was triggered by AMP243 in all 35S:GFP lines tested 62

3.2.3 Deveoplment of lines homozygous for both the AMP243 trigger and the 35S:GFP transgene 72

3.2.4 Attempting to trigger silencing using the SUC-GF:IR trigger line 73

3.2.5 Ability to undergo RdDM did not seem to be influenced by genome location 74

3.3 Discussion 81

3.3.1 Summary 81

3.3.2 Production of homozygous single locus GFP transgenic lines 82

3.3.3 Locus-specific effects 85

3.3.4 Overlap between PTGS and RdDM 87

3.3.5 Other areas of inquiry 88

3.3.6 Conclusions 89

3.4 Acknowledgments 89

Chapter 4 Development of epitope-tagged RNA silencing components 90

4.1 Introduction 90

4.2 Results 92

4.2.1 AMP243 silencing is RDR6-dependent and is delayed in *rdr2* mutants 92

4.2.2 Using the Gateway cloning system 94

4.2.4 Attempts to generate NRPD1a tagged lines were unsuccessful 97

4.2.5 The production and characterization of GFP-tagged lines 97

4.2.6 Validating a mutant *rdr2 Arabidopsis* line 100

4.2.7 Generating and characterising RDR2:FLAG and RDR6:FLAG transgenic Arabidopsis lines. 101

4.2.8 Characterising NRPE1: FLAG lines 109

4.3 Discussion.....110

4.3.1 AMP243 triggered silencing is affected by both *rdr6* and *rdr2* mutants 110

4.3.2 The generation of epitope tagged constructs and potential future uses.....111

4.4 Acknowledgments.....113

Chapter 5 RdDM may be influenced by transgene structure and copy number.....114

5.1 Introduction 114

5.2 Results.....116

5.2.1 Analysis of transgene copy number in the multiple independent 35S:GFP lines 116

5.2.2 Analysis of integration complexity in the multiple independent 35S:GFP lines 119

5.2.3 Analysis of DNA methylation in the F2 generation of the AMP243 X 35S:GFP Crosses 123

5.2.4 Spontaneous silencing in the 35S:GFP parent lines.....129

5.2.5 Two Histone modifications are not altered by AMP243 triggered silencing.....130

5.2.6 Attempts to use the FLAG fusion proteins to investigate recruitment of RNA silencing machinery to the GFP loci were unsuccessful 135

5.3 Discussion.....136

5.3.1 Summary of results 136

5.3.2 Copy number.....136

5.3.3 Transgene integration structure 139

5.3.4 Zygosity 141

5.3.5 Spontaneous silencing in the parent lines 142

5.3.6 The use of FLAG tagged proteins 143

5.3.7 Histone modifications 145

5.3.8 Conclusion.....146

5.4 Acknowledgments.....146

Chapter 6 Discussion.....147
6.1 AMP243-triggered PTGS was not affected by the genomic location.....147
6.2 AMP243-triggered RdDM is influenced by multiple -factors.....149
6.3 Histone modifications.....150
6.4 Attempting to investigate recruitment of RNA silencing proteins to targets using FLAG tagged lines.....151
6.5 Limitations of the work and alternative techniques.....151
6.6 Other areas of research into increasing expression of transgenes and reducing variability153
6.7 Concluding remarks.....154
Appendix 1. Chi square analysis of T1 35S:GFP lines.....155
List of Abbreviations.....160
References.....162

List of Figures

Figure 1.1. Overview of the main RNA silencing pathways in plants adapted from Molnar *et al*., (2011) 16

Figure 1.2. Overview of the biogenesis of *trans*-acting siRNA. 20

Figure 1.3. The production of secondary siRNA during PTGS by RDR6 21

Figure 1.4. Key steps of RNA-directed DNA methylation 25

Figure 3.1. Schematic of the work discussed in this chapter 59

Figure 3.2. Diagram of the viral section of the 35S-*PVX:GFP* transgene in trigger line AMP243. 60

Figure 3.3. Overview of homozygous single locus GFP transgenic line production 62

Figure 3.4. Typical result of visual analysis observed for all F1 AMP243 X 35S:GFP Crosses 63

Figure 3.5. AMP243-triggered loss of GFP fluorescence is associated with loss of *GFP* mRNA 64

Figure 3.6. siRNAs accumulate in the AMP243 X 35S:GFP cross but not in the parent lines. 66

Figure 3.7. Restriction map of the 35S:GFP transgene for AluI and HaeIII. 67

Figure 3.8. Analysis of methylation status of AMP243 X 35S:GFP crosses by Southern blot. 69

Figure 3.9. Bisulfite sequencing data confirm results from the Southern blots. 71

Figure 3.10. Typical result of visual analysis observed for all SUC-GF:IR x 35S:GFP crosses 74

Figure 3.11. Primer plan for thermal asymmetric interlaced (TAIL) PCR 74

Figure 3.12. 1.2% Agarose gels for TAIL PCR products for lines 18, 19, 22, 23, 28 and 38 76

Figure 3.13. Output from the Epigenome browser ANNO-J for lines 67 and 103. 77

Figure 3.14. Overview of work performed in this chapter 82

Figure 4.1. AMP243 triggered silencing of the GFP transgene is RDR6 dependent 93

Figure 4.2. AMP243-triggered silencing is delayed in *rdr2* mutants 94

Figure 4.3. Schematic of work involved in creating epitope tagged constructs and desired outcomes 95

Figure 4.4. RDR6:GFP showed GFP fluorescence when transiently expressed in *Nicotiana benthamina* as determined by Confocal microscopy. 99

Figure 4.5. Validation of the SALK_059661 *rdr2* mutant by Southern Blot. 101

Figure 4.6. The introduction of RDR2:FLAG to *rdr2* plants results in partial restoration of wild type 5S rDNA DNA methylation levels as detected by Southern blot. 102

Figure 4.7. Phenotype rescue of the RDR6 wild type phenotype was seen in T2 progeny from transformations of the *rdr6-11* mutant with the RDR6:FLAG construct. 103

Figure 1.8 Complementation of the *rdr6* phenotype. 104

Figure 4.9. The RDR6:FLAG transgene is present in T2 plants. 105

Figure 4.10. Genotyping of T2 plants reveals that the epitope tagged construct is present in plants with the *rdr6-11* mutant background 106

Figure 4.11. The RDR6:FLAG fusion protein is detectable by western blot in some RDR6:FLAG lines using an anti-FLAG primary antibody. 108

Figure 4.12. The FLAG tag is detectable by Western Blot in NRPE1:FLAG, RDR2:FLAG and RDR6:FLAG plants. 109

Figure 5.1. Copy number may affect the ability of F1 crosses to undergo methylation 118

Figure 5.2. Schematic of the GFP T-DNA insert and possible outcomes of the BamHI Southern Blot 120

Figure 5.3. Analysis of transgene integration structure in a selection of 35S:GFP transgenic lines 121

Figure 5.4. GFP DNA methylation is prevalent in the F2 AMP243 X 35S:GFP crosses. 124

Figure 5.5 Analysis of DNA methylation status in parent, homozygous and hemizgous samples for four lines 129

Figure 5.6. Amplicon-mediated silencing does not lead to a reducation in H3 acetylation levels at target loci 133

Figure 5.7. Amplicon-mediated silencing does not lead to a increase in H3K9me2 levels at target loci 134

List of Tables

Table 2.1 Primers used for PCR reactions in this project. 38

Table 2.2 *Arabidopsis thaliana* lines used in this project 45

Table 2.3 Primers Used to Generate Probes for Nucleotide Blots 49

Table 3.1 Summary of Bisulfite sequencing results which confirm Southern Blot data 72

Table 3.2. Analysis of TAIL PCR results using ANNOJ 79

Table 3.3. Information from TAIR database on genes in Table 3.2 that have been the site of T-DNA insertion in this study 81

Table 3.4. Centromeres in *Arabidopsis thaliana* 84

Table 5.1. Summary of the copy number and integration data 122

Table 5.2 Percentage of GFP transgene DNA methylation in parent, homozygous and hemizygous samples for 4 lines. 126

Table 5.3. Spontaneous silencing in next generation of 35S:GFP plants 130

Acknowledgments

I would first like to thank my supervisor Louise Jones for all of the help and support she has provided me throughout my PhD. I would also like to thank my TAP panel, Harv Issacs and Michael Schultze, for all of their advice and encouragement during this project. I would also like to thank Fabian Vaistij for teaching me the finer points of ChIP and qPCR. I would like to thank the horticultural staff in the Department of Biology for not only looking after my plants but for also always being there when I needed to see a friendly face.

I have been very lucky when it comes to people I share lab space with. There are so many amazing people from the L2 floor over the years that have made my time here so enjoyable that I cannot name them all here. Special mention does need to go to former members of the Jones group, Tom Brabbs and Tom Smith. Their advice and support at the start of my time in York was invaluable to me. I am also indebted to the Pownall and Issac groups for their advice on everything from lab work, to computers, to life in York and beyond. Mostly I would like to thank them for always reminding me that however tough things got, at least I didn't work on frogs!

Outside of the lab I would like to thank all of the friends I have made at the university who have been a tremendous source of support. Special mention goes to my surrogate families – my house mates from Lilac Ave and YUCC.

I could not have done any of this without the help and support of my family who are always at the end of the phone when I need them. I would like to thank my Mum for teaching me not to give up and for reading all of the SMP books to me. I would like to thank my Dad for his endless encouragement and support of me pursuing science. Both of my sisters deserve a special mention here. Hannah for keeping me sane and always giving me the best advice and Jess for always pushing me when I need it and for teaching me a new meaning of the phrase "work ethic".

Lastly, but by no means least, I would like to thank Ben for putting up with me over the last few years of my PhD. His seemingly endless support and encouragement have been immensely valuable to me and he has always believed I would succeed, even when I couldn't see it myself. Thank you.

This PhD was funded by the BBSRC

Author's Declaration

None of the work presented in this thesis has previously published or submitted for a qualification either at the University of York or at any other institution. Some parts of the work however have been presented in a poster at ICAR 2014 (Vancouver). All of the work presented here is that of the author except those started in the acknowledgements section of each results chapter.

Chapter 1 Introduction

Transgenes allow the production of new plant varieties that may have important economical and social value whist also providing a useful research tool to further our basic knowledge of plants. One of the areas in which transgenes have been particular useful is the field of RNA silencing (Meyer, 2013). Transgenes act as both an important tool to aid research into this area and as a conundrum to solve due to the fact that transgenes very often undergo RNA silencing when introduced into higher plants. RNA silencing was partially responsible for the variability in transgene expression originally observed in the infancy of the field of genetically modifying plants (Meyer, 2013). Much has since been discovered about transgenes, RNA silencing and the connection between the two. However, there are still many unanswered questions surrounding both fields.

One such question is why some transgenes when introduced into plants undergo silencing whereas other identical transgenes do not. Additionally, it is unclear why some transgenes can support DNA methylation when silenced whereas others cannot. This project aims to investigate these questions by exploring the possibility that locus-specific effects influence a transgene's potential to undergo RNA silencing and/or DNA methylation. These locus-specific effects may be a result of either the site of transgene integration or be due to the properties of the transgene locus itself. These properties could include complex integration of repeats of the transgene at a single integration site for example.

This research has aimed to address these questions by generating and characterizing multiple independent single locus GFP transgenic Arabidopsis lines that have been triggered to undergo silencing by crossing with an amplicon trigger line (AMP243). The ability of these lines to undergo RNA-dependent RNA polymerase (RDR)6-dependent post transcriptional gene silencing (PTGS) and/or RNA directed DNA methylation (RdDM) has been investigated in an attempt to correlate characteristics of the transgene to silencing outcomes.

Another question related to transgene silencing is how proteins of the PTGS/RdDM pathway become recruited to target loci and not others. To investigate this epitope-tagged versions of RDR2 and RDR6 have been generated in order to investigate their potential recruitment to target loci.

1.1 The Variety of RNA silencing pathways

1.1.1 RNA silencing is involved in a variety of functions

RNA silencing is a conserved molecular process in eukaryotes that allows sequence-specific regulation of gene expression. In plants, RNA silencing has many functions including protection of the genome against invasive nucleic acids such as transposable elements and viruses, as well as acting as a key regulator of gene expression during development and response to stress (Jones-Rhoades & Bartel, 2004). Each of these functions is performed by different classes of the small RNAs (sRNAs) which mediate RNA silencing. These sRNAs range in size from 20 nucleotides (nt) to 30nt in length (Chen, 2009) and allow silencing pathways to act in a sequence-specific manner. Classes of sRNA found in Arabidopsis include but are not limited to: micro RNAs (miRNAs), small-interfering RNAs (siRNAs), trans-acting siRNAs (*ta*-siRNAs), natural-antisense siRNAs (nat-siRNAs), endo-inverted repeats siRNAs and long miRNAs (Reviewed by Ghildiyal & Zamore, 2009).

1.1.2 The core elements of RNA silencing

The core elements of sRNA biogenesis are similar for all classes (Figure 1.1). The initial stage involves the processing of long double stranded RNA (dsRNA) into short duplexes by the endonuclease activity of enzymes in the DICER-like family (DCL) (Park et al., 2002, Law & Jacobsen, 2010). Hua Enhancer 1 (HEN1) is involved in stabilizing the sRNA by causing the methylation of the 3'OH group (Matzke & Mosher, 2014, Yang et al., 2006). One strand of the resulting sRNA duplex is incorporated into an Argonaute (AGO)-associated complex (Khvorova et al., 2003). There is a bias as to which strand gets incorporated into the AGO complex depending on the thermodynamic stability of the duplex: the strand with the least stable 5' end is incorporated most often (Khvorova et al., 2003). The AGO complex becomes incorporated into the RNA-induced silencing complex (RISC) (Reviewed by Molnar et al., 2011). The sRNA then directs the RISC to the sequence to be silenced by Watson-Crick base pairing.

The different RNA silencing pathways can act at either the transcriptional or post-transcriptional level. At the post-transcriptional level the sRNA directs the AGO protein to messenger RNA (mRNA) and causes protein production to be inhibited by mRNA cleavage or translational inhibition (Chen, 2004). Alternatively, silencing at the transcriptional level involves RNA-directed DNA methylation (RdDM) and/or histone modifications which prevent the production of mRNA (Zhang & Zhu, 2011).

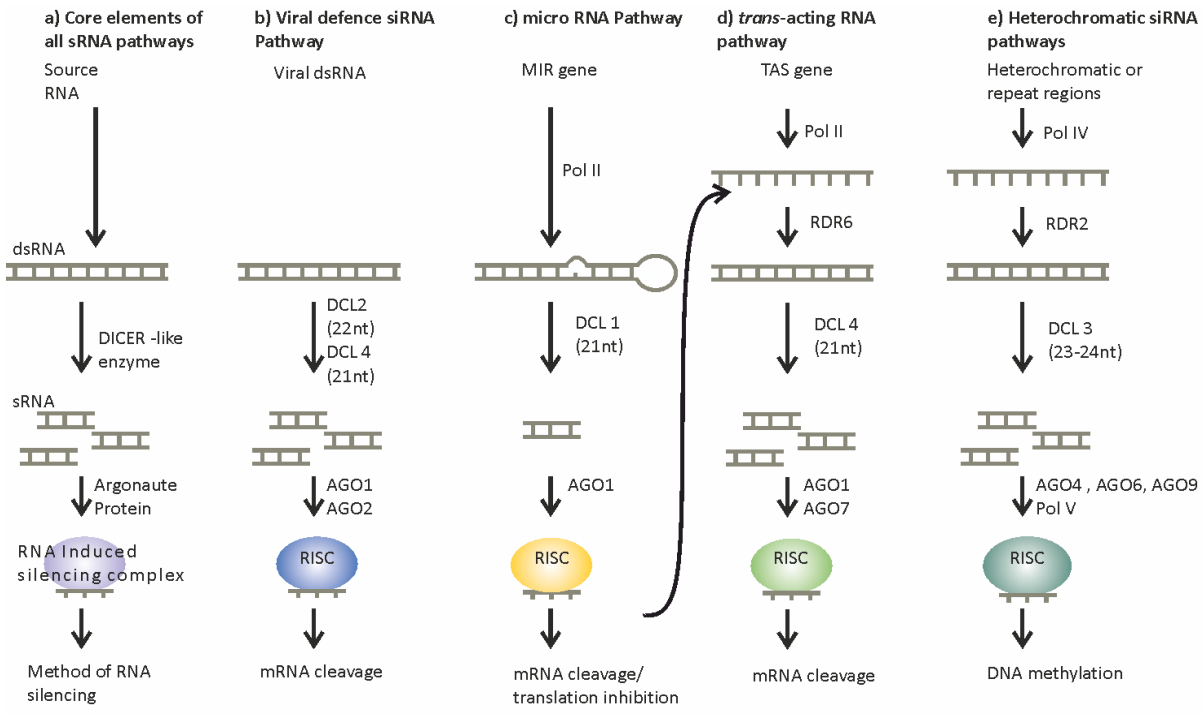


Figure 1.1. Overview of the main RNA silencing pathways in plants adapted from Molnar *et al*., (2011)

A) Schematic of the basic silencing pathway with the core elements. Double stranded RNA is processed by a member of the DICER-like (DCL) family of enzymes into sRNA duplexes. One strand of the sRNA duplex is incorporated into a protein belonging to the Argonaute (AGO) family. Other proteins associate with the sRNA containing AGO to form the RNA Induced Silencing Complex (RISC). RISC then goes on to mediate RNA silencing. Each RISC complex is a different colour to indicate the different composition of proteins present. The main pathways include the **B)** Viral siRNA, **C)** microRNA, **D)** *trans*-acting siRNA and **E** the heterochromatic pathways. In addition to the core elements some pathways require RNA-dependent RNA polymerases (RDRs) and/or DNA-dependent RNA polymerases (Pol II, Pol IV, Pol V) to generate double-stranded RNA.

1.1.3 The different classes of small RNA

Variation of these basic core elements allows for the biogenesis of the different classes of sRNAs through various silencing pathways (Figure 1.1). The initial dsRNA can come from a variety of sources including inverted repeats, endogenous RNA with secondary structure and by the action of RNA dependent RNA polymerases (RDRs) which produce dsRNA from single stranded RNA (ssRNA) templates (Figure 1.1) (Law & Jacobsen, 2010). Plant genomes encode for multiple DCL and AGO proteins. For example, Arabidopsis has four DICER-like enzymes (DCL 1- DCL4) (Henderson et al., 2006) and 10 AGOs (AGO1- AGO10) (Reviewed Mallory & Vaucheret, 2010). These proteins work together in different combinations in the RNA silencing pathways to produce the different classes of sRNAs and to effect changes in gene expression (Figure 1.1). For example miRNAs are commonly 21nt in size and originate from endogenously encoded long strands of

ssRNA, which contain imperfect hairpin loop structures (Figure 1.1b). The miRNAs are processed from the double-stranded regions by DCL1 and are subsequently incorporated into complexes with AGO1 (Qi et al., 2006, Vaucheret et al., 2004). Many miRNAs are involved in regulating developmental processes (Chen, 2009).

It should be noted that this is a simplified view of RNA silencing. There are many cases of redundancy between the protein families in the pathway. For example DCL proteins produce a specific size of sRNAs each but there is an overlap between different DCL proteins and the sources of dsRNA processed (Zhang et al., 2007). Additionally, there is overlap between the AGO proteins with miRNAs reportedly making up 91% of the sRNAs that associate with AGO1 and 3% of the sRNAs that associate with the RdDM Argonaute AGO4 (Qi et al., 2006).

1.2 Post-transcriptional Gene Silencing

RNA silencing occurs at either the transcriptional or post-transcriptional level. At the post-transcriptional level silencing causes protein production to be inhibited by mRNA cleavage or translational inhibition, whereas silencing at the transcriptional level involves preventing the production of mRNA by RdDM and/or histone modifications (Chen, 2004, Zhang & Zhu, 2011). Both work in a sequence-specific manner due to the actions of sRNA. Both transgenes and endogenous genes can undergo Post-Transcriptional Gene Silencing (PTGS) and Transcriptional gene silencing (TGS).

The process of PTGS can be thought of in two stages. There is the initiation stage where targets of PTGS are identified and silenced which can be followed by a self-maintenance stage which allows the persistence of silencing in the absence of the initial trigger (Eamens et al., 2008, Ruiz et al., 1998, Vaistij et al., 2002). The self-maintenance stage can occur through the action of RDR6. During this maintenance stage RDR6 can amplify the RNA silencing signal through the production of secondary siRNA. However, not all variants of the PTGS pathway undergo self-maintenance; the reasons for this are unclear.

PTGS can involve the actions of AGO1, DCL2, DCL4, RDR6 and SUPPRESSOR OF GENE SILENCING 3 (SGS3) (Reviewed by Mourrain et al., 2007). However, unlike the miRNA pathway, the proteins involved in PTGS show greater variation due to the broader nature of the pathways involved. For example, PTGS of transgenes can be either RDR6-dependent or independent. Additionally, PTGS of transgenes has been demonstrated to involve DNA methylation that is dependent on Polymerase IV (Pol IV) and Polymerase V (Pol V) from the RdDM pathways (Eamens et al., 2008, Herr et al., 2005). This will be discussed in more detail later.

The role of RDR6 in RNA silencing in Arabidopsis was reported by Dalmay et al. (2000a) using a reverse-genetics screen based upon PTGS of a transgene (Mourrain et al., 2000). RDR6 has also been shown to play a role in the biogenesis of *ta*-siRNA; the discovery of which provided the first role for RDR6 in endogenous gene regulation (Curaba & Chen, 2008, Vazquez et al., 2004). The biogenesis of *ta*-siRNA provides a well-studied example of RDR6 amplifying the effect of an RNA silencing pathway.

1.2.1 Trans-acting siRNA biogenesis

Biogenesis of ta-siRNAs, also referred to as phased, secondary, small interfering RNAs (phasiRNA), starts with the transcription of a non-coding TAS gene by RNA polymerase II (Figure 1.2) (Fei et al., 2013). This single-stranded RNA transcript is then cleaved by the action of AGO1 or AGO7 guided by a miRNA. The cleaved fragments are stabilized by SGS3 before being converted to dsRNA by RDR6 (reviewed by Vazquez et al., 2010). This may occur in SGS3/RDR6 granules that have been found in the cytoplasm (Kumakura et al., 2009). The resultant dsRNA is then processed by DCL4 to produce a new 21nt sRNA. The biogenesis of the 21nt siRNAs begins adjacent to the initial cleavage site, as created by the original miRNA, and moves along the dsRNA in 21nt portions. The resulting *ta*-siRNAs are therefore phased from the original miRNA, covering a different sequence of the original TAS mRNA (Yoshikawa et al., 2005). The phased siRNAs can either cover the downstream fragment of the cleaved mRNA or the upstream fragment depending on if a “one hit” or “two hit” pathway of biogenesis occurs (Axtell et al., 2006). In the one hit model the mRNA is targeted at one site by one TAS-derived miRNA which leads to the production of ta-siRNAs for the 3’ fragment downstream of the miRNA cleavage site (Fei et al., 2013). In the “two hit” system the mRNA is targeted by two miRNAs with two sites but cleavage only occurs at the 3’ target site and results in the production of phased siRNAs to the 5’ fragment upstream of the cleavage site and between the two miRNA target sites (Axtell et al., 2006). The *ta*-siRNAs produced then target other mRNAs in *trans*, causing the silencing of genes from different gene families. The cleaved fragments from the mRNA targeted by the *ta*-siRNAs can then become templates for RDR6 themselves creating a reinforcing mechanism of silencing (Figure 1.2 and Figure 1.3). Although the *ta*-siRNA class of sRNA works at the post transcriptional level, it differs from other forms of PTGS due to the involvement of TAS genes.

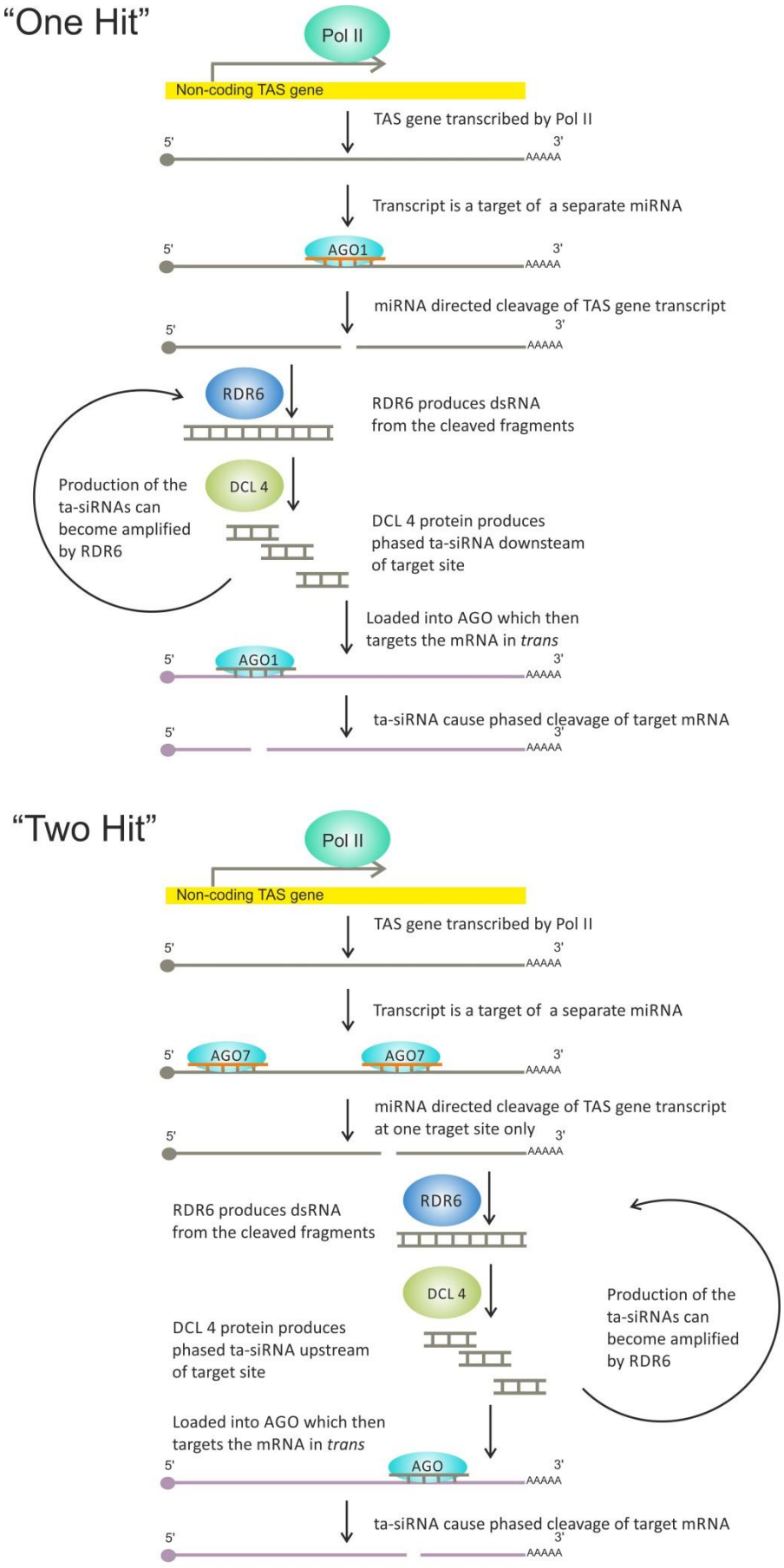

"One Hit"
Pol II
Non-coding TAS gene
TAS gene transcribed by Pol II
5'
3'
AAAAA
Transcript is a target of a separate miRNA
AGO1
miRNA directed cleavage of TAS gene transcript
RDR6
RDR6 produces dsRNA from the cleaved fragments
DCL 4
Production of the ta-siRNAs can become amplified by RDR6
DCL 4 protein produces phased ta-siRNA downsteam of target site
Loaded into AGO which then targets the mRNA in trans
ta-siRNA cause phased cleavage of target mRNA
"Two Hit"
Pol II
Non-coding TAS gene
TAS gene transcribed by Pol II
Transcript is a target of a separate miRNA
AGO7
AGO7
miRNA directed cleavage of TAS gene transcript at one traget site only
RDR6
RDR6 produces dsRNA from the cleaved fragments
DCL 4
Production of the ta-siRNAs can become amplified by RDR6
DCL 4 protein produces phased ta-siRNA upstream of target site
Loaded into AGO which then targets the mRNA in trans
AGO
ta-siRNA cause phased cleavage of target mRNA

Figure 1.2. Overview of the biogenesis of *trans*-acting siRNA.

One Hit model The transcript of a non-coding TAS gene is targeted by a miRNA. The cleaved fragments become templates for RDR6 which use them to produce long dsRNA. The dsRNA is processed by DCL4 to produce 21nt phased ta-siRNA downsteam of the initial target site. These ta-siRNAs guide the AGO-mediated cleavage of the mRNA of several unrelated genes, hence they work in *trans*. The cleaved fragments caused by the ta-siRNAs targeting can then become templates for RDR6 themselves creating a reinforcing mechanism of silencing. **Two Hit** The transcript of a non-coding TAS gene is targeted by two miRNAs and cleavage occurs at the 3' target site. The rest of the process is as for the one hit model but the phased ta-siRNAs occur upstream of the cleaved target site instead of downstream as for the one hit model (Adapted from Fei et al., 2013).

1.2.2 Amplification of Post-transcriptional Gene Silencing by RDR6

As previously mentioned PTGS can become self-maintaining by the action of RDR6 (Figure 1.3). It is proposed that, as for the biogenesis of *ta*-siRNAs, cleaved ssRNAs generated from an initial round of sRNA-directed mRNA cleavage are used as a template by RDR6 to make dsRNA (Curaba & Chen, 2008). This dsRNA is then processed into so-called secondary sRNAs which then direct further mRNA cleavage events. In addition to the maintenance of the silencing signal, production of the secondary sRNAs leads to the spreading of silenced sites across the silenced gene (Vaistij et al., 2002). However, it still remains unclear why some transcripts are able to support this amplification cycle and act as RDR6 templates whereas others do not. Understanding what RDR6 uses as an initial template may hold the key to solving this conundrum.

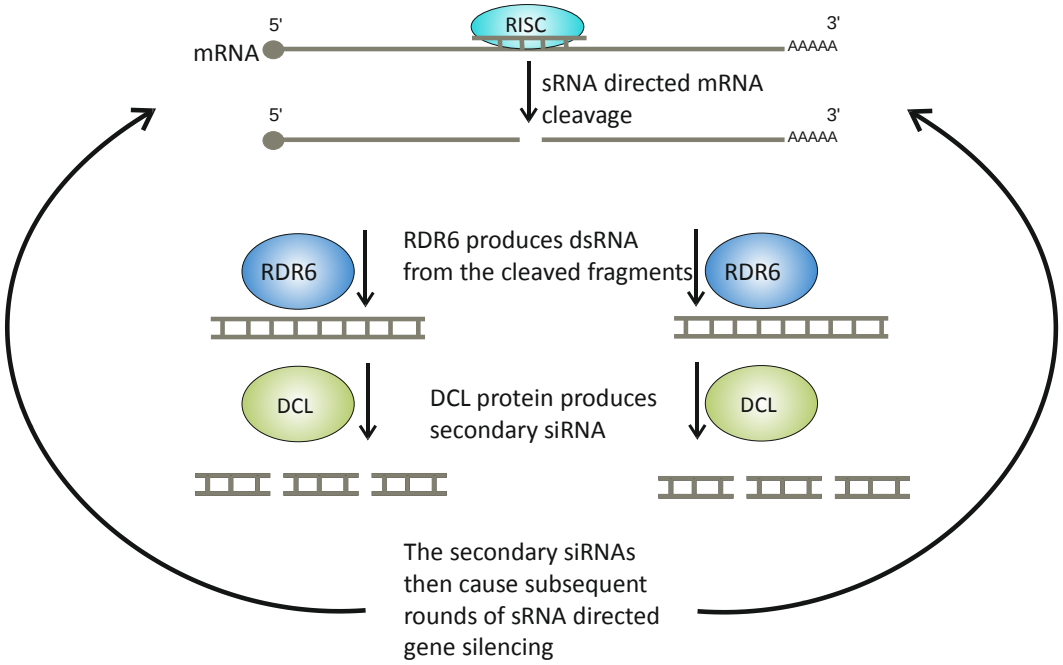


Figure 1.3. The production of secondary siRNA during PTGS by RDR6

RDR6 can use single stranded mRNA that has been cleaved through the action of sRNA mediated silencing as a template. After cleavage by RISC, RDR6 converses the ssRNA template into dsRNA. This can then be processed by a DCL protein to produce more sRNAs to guide silencing of mRNA with the same sequence by Watson-Crick base pairing. The sRNAs generated by RDR6 will be phased from the original cleavage site along the mRNA.

1.2.3 RDR6 template recognition

One proposed answer to the question of RDR6 template specificity suggests that RDR6 recognises transcripts that are "aberrant" in some way, for example missing the poly-A tail or lacking the 5' cap (Gazzani et al., 2004), and it is these differences that mark them for RDR6 activity. However, RDR6 does not appear to be able to distinguish between RNA with or without a poly-A tail (Curaba

& Chen, 2008). Additionally, it is not common for endogenous genes to act as RDR6 templates even when cleaved by miRNAs or siRNAs, suggesting that RDR6 is able to distinguish between cleaved transcripts from different sources (Vaistij et al., 2002). It has been suggested that targets cleaved by a 22 nt miRNA instead of a 21 nt miRNA tend to act as templates for RDR6, possibly due to the interaction of the 22 nt miRNA with AGO1 (Cuperus et al., 2010, Chen et al., 2010). Additionally, although many transgenes can readily undergo RDR6 amplification, this does not always occur. Locus-specific effects have been observed with identical transgenes (Jones et al., 1998). This raises the possibility that RDR6 may be interacting at the DNA level with the template producing-loci. Consistent with this idea it has been shown that RDR6-GFP fusion proteins localize in the nucleus which would means RDR6 has access to DNA and DNA-associated proteins (Hoffer et al., 2011, Kumakura et al., 2009, Luo & Chen, 2007). Furthermore, the DNA of genes that have undergone this amplification often become methylated via RdDM although the functional significance of this is not understood (Jones et al., 1998, Jones et al., 1999, Vaistij et al., 2002). It is possible that being transcribed from methylated DNA marks transcripts as being templates for RDR6 but such marked transcripts have yet to be identified (Eamens et al., 2008).

One of the main tasks of the RNA silencing machinery is to protect the plant genome against invading viruses and in some cases RDR6 plays an important role in this defence (reviewed by Ding, 2010). In these instances viral RNA acts as an RDR6 template (Garcia-Ruiz et al., 2010, Wang et al., 2010). However, siRNA-directed viral defence does not always depend on RDR6 and it is unclear what causes RDR6 to recognise some viruses as templates and not others.

1.3 RNA silencing and DNA methylation

1.3.1 DNA methylation and its maintenance

There are three sequence contexts in which DNA methylation can occur in plants CG, CHG or CHH, (where H can be either a C, T or A). Using bisulfite sequencing Cokus et al. (2008) have determined that the overall genome-wide levels of DNA methylation in Arabidopsis are 24% for CG, 6.7% for CHG, and 1.7% for CHH methylation. CG and CHG context methylation occurs in a symmetrical state with both strands of the DNA helix exhibiting the same cytosine methylation whereas CHH methylation can occur in an asymmetrical state (Cokus et al., 2008).

Arabidopsis displays a mosaic-like DNA methylation pattern which includes both heterochromatic and euchromatic regions. In Arabidopsis DNA methylation primarily occurs at repeat sequences, TEs, and heterochromatic regions such as pericentromeric regions etc. The silencing and methylation of TEs will be discussed later in this introduction. Arabidopsis also displays DNA methylation in euchromatin with targets including pseudogenes, promoter regions, and gene

bodies of actively transcribed genes. Over a third of expressed genes contain methylation within their transcribed region (Zhang et al., 2006). According to Zhang et al. (2006) these tend to be highly expressed genes such as housekeeping genes. This was a surprise upon its discovery as DNA methylation was seen only as a repressive mark. Gene body methylation has been shown to be evolutionarily ancient between plants and animals, whereas the methylation of TEs, as seen in Arabidopsis, is not (Zemach et al., 2010). In Arabidopsis gene body methylation tends to occur in the middle of the gene, not at the 5' and 3' terminals. DNA methylation of promoter regions in Arabidopsis occurs for only around 5% of genes. Unlike gene body methylation, genes with methylated promoters tend to have tissue specific expression patterns (Zhang et al., 2006). In Arabidopsis CG, CHG and CHH methylation predominately occurs at repeat regions and the pericentromeric regions, whereas CG methylation alone has a strong correlation to gene bodies (Cokus et al., 2008).

With regard to RNA silencing, targets of TGS tend to be the promoters of genes (English et al., 1996, Morel et al., 2000) whereas RdDM linked to PTGS tends to result in gene body methylation. This may be a result of PTGS requiring transcription of the target loci before it causes silencing (Hobbs et al 1993).

In Arabidopsis DNA methylation is maintained by three classes of DNA methyltransferases, each of which predominantly operates at a defined range of sequence contexts with some redundancy (Cokus et al., 2008, Watson et al., 2014). These include Methyltransferase 1 (MET1), Domains Rearranged Methyltransferase 1 (DRM2) and Chromomethylase 3 (CMT3) (reviewed by Law & Jacobsen, 2010). CMT3 is a plant specific methyltransferase while both MET1 and DRM2 have mammalian counterparts in DNMT1 and DNMT3, respectively (Reviewed by Meyer, 2011).

The majority of *de novo* methylation in plants is performed by DRM2 via the RdDM pathway for all sequence contexts (Cao et al., 2003, Greenberg et al., 2011, Naumann et al., 2011). RdDM along with DRM1 and DRM2 maintain the asymmetric CHH methylation (Cokus et al., 2008). A description of RdDM will be given in the next section.

MET1 is the main proponent of symmetric CG DNA methylation maintenance. The MET1 pathway often works in conjunction with the SWI/SNF chromatin remodelling factor Decrease in DNA methylation 1 (DDM1) (Hirochika et al., 2000, reviewed by Meyer, 2011). Additionally, the MET1 pathway involves three of the variant in methylation (VIM) proteins at some loci (Woo et al., 2008). These are SRA (SET- and RING-associated) domain proteins that can bind methylated cytosine. Single, double and triple mutants of *vim1*, *vim2* and *vim3* have revealed a role for these genes in CG methylation and that they operate together with partial redundancy (Woo et al., 2008).

The CMT3 pathway for DNA CHG methylation maintenance has a strong association with the repressive histone mark H3K9me2 (Bernatavichute et al., 2008). It is proposed that there is a feedback loop between histone and DNA methylation to maintain CHG methylation as loss of either CMT3 or the histone methyltransferase KRYPTONITE (KYP or SUVH4) results in the reduction of CHG methylation (reviewed by Law & Jacobsen, 2010). CMT3 has a chromodomain that interacts with methylated histones at the H3K9 and H3K27 positions (Du et al., 2012, Lindroth et al., 2004). KYP is one of several SRA domain proteins, including SUVH5 and SUVH6, that can bind CHG methylated DNA (reviewed by Law & Jacobsen, 2010). Together these proteins mediate the maintenance of CHG methylation.

There is increasing evidence that the roles of the three classes of methyltransferase are not independent and that there is overlap between MET1, CMT3 and DRM2 (Cokus et al., 2008, Watson et al., 2014). For example, Cokus et al. (2008) have reported that CMT3 and DRM1/DRM2 work redundantly, as single mutants of each of these genes do not display full loss of either CHG or CHH methylation. Mutants of *met1* have exceptionally low levels of CG methylation and rely heavily on stochastic RdDM to avoid detrimental phenotypes (Mathieu et al., 2007); this suggests that the function of MET1 to maintain DNA methylation cannot be replaced with either CMT3 or DRM2 (Cokus et al., 2008, Mathieu et al., 2007). Additionally, Watson et al. (2014) report that for a small number of loci MET1 may have a role in contributing to non-CG methylation.

1.3.2 RNA directed DNA Methylation

RdDM instigates *de novo* methylation in all three contexts and maintains the asymmetric CHH methylation (Reviewed by Zhang & Zhu, 2011, Meyer, 2011). Transcriptional silencing events involving RdDM can also become self-maintaining. The RdDM pathway can be described as a self-amplifying circle (Figure 1.4) (Law & Jacobsen, 2010). Although many of the components of this pathway have been well characterised, little is known about how the cycle starts, how it is amplified or how targets become assigned to this pathway.

In addition to the core RNA silencing elements described previously (Figure 1.1a) RdDM involves two plant-specific polymerases, Polymerase IV (Pol IV) and Polymerase V (Pol V) (Herr et al., 2005, Mosher et al., 2008) (Figure 1.4). Both polymerases share some subunit composition with Pol II, which they are thought to have evolved from (Ream et al., 2009). According to Ream et al. (2009) Pol IV has four non-Pol II subunits, Pol V has six non-Pol II subunits with the two polymerases having a difference of four subunits between them. Although subsequent work is in agreement with Ream et al. (2009) for the most part, there have been some discrepancies in the presence of some of the smaller subunits in Pol IV in work by others as discussed in Law et al. (2011). The largest two subunits of Pol IV are Nuclear RNA Polymerase (NRP)-D1 and NRPD2a which are

homologs of the Pol II subunit RPB1 (Huang et al., 2009, Ream et al., 2009). NRPD2a is also a subunit of Pol V whose largest subunit is NRPE1 (Huang et al., 2009, Ream et al., 2009).

The canonical RdDM pathway can be thought of in three parts; siRNA biogenesis, *de novo* DNA methylation and self-reinforcement of RdDM (Matzke & Mosher, 2014). Canonical RdDM is linked to TGS silencing described here in section 1.2.

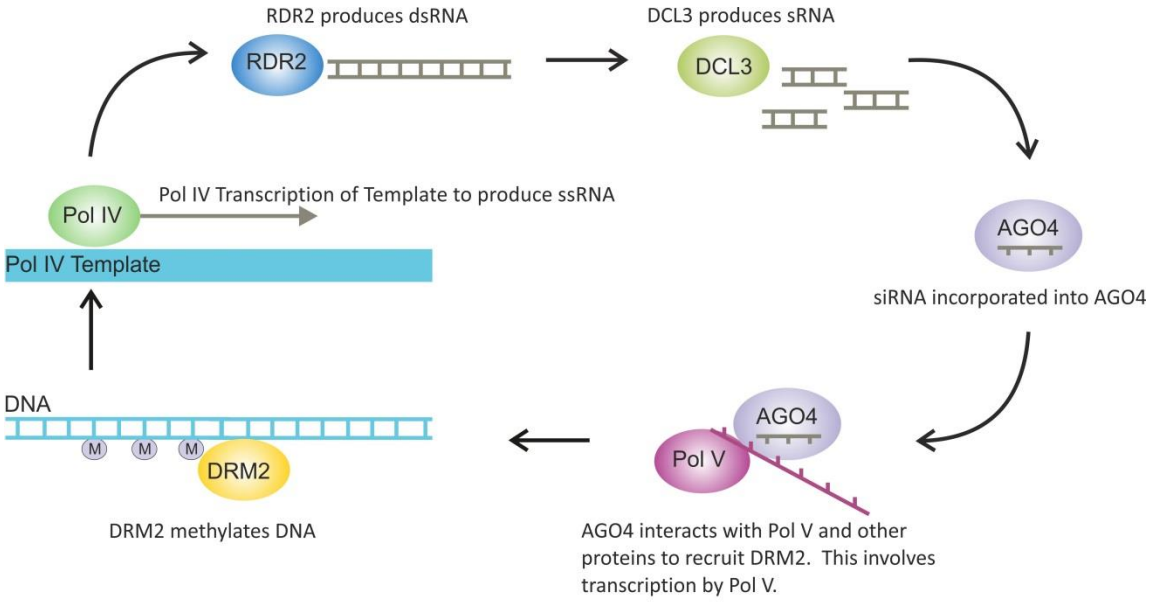


Figure 1.4. Key steps of RNA-directed DNA methylation

Adapted from Law and Jacobsen (2010). RDR2 produces double stranded RNA from the proposed Pol IV transcript. DCL3 converts the dsRNA into sRNAs, a strand of the resulting duplex is incorporated into AGO4. AGO4 interacts with Pol V and other proteins to form the effector complex. Pol V produces a DNA transcript which acts a scaffold to facilitate recruitment of proteins to the effector complex. However, much is unknown about the order of recruitment of the proteins into the effector complex DDR and how Pol V produces its transcript. The effector complex recruits DRM2 to the target DNA which performs the DNA methylation.

1.3.2.1 The biogenesis of siRNAs in RdDM

The siRNAs in the RdDM cycle are generated by the actions of Pol IV, RDR2 and DCL3 (Eamens et al., 2008, Herr et al., 2005, Xie et al., 2004) (Figure 1.4). The current model is that transcription by Pol IV provides a ssRNA template that is converted to dsRNA by RDR2 (Zhang et al., 2007, Matzke & Mosher, 2014, Law & Jacobsen, 2010). The dsRNA created though the action of Pol IV and RDR2 are then processed by DCL3 into siRNAs of 24nt in length (Henderson et al., 2006) (Figure 1.4). The sRNAs then associate predominantly with AGO4 and to a lesser extent with AGO6 or AGO9 (Havecker et al., 2010, Smith et al., 2007). It has been suggested that the sRNA stabilise the AGO protein itself (Xie et al., 2004, Li et al., 2006). Many other proteins are involved at this stage but these have been omitted from Figure 1.4 for simplicity. The most noteworthy proteins

include HEN1, the putative chromatin remodeller Classy 1 (CLSY1), SAWADEE HOMEDOMAIN HOMOLOGUE 1 (SHH1) and transcription factors such as RNA-DIRECTED DNA METHYLATION 4 (RDM4) (Greenberg et al., 2011, Law et al., 2011, Smith et al., 2007).

1.3.2.2 The recruitment of Pol IV

There are many questions surrounding the siRNA biogenesis part of the RdDM pathway. One of the biggest concerns is the issue of recruitment of Pol IV to its target loci. At the beginning of this PhD study there was much debate as to what the Pol IV template could be. For example, as RdDM is associated with transposons and repeat elements, this suggested that the Pol IV template could be methylated DNA. However, this raises the question of how the DNA became methylated in the first place. According to Figure 1.4, Pol IV is a DNA-dependent RNA polymerase. Prior to the start of this project it was not known if Pol IV used DNA as a template or acted as an RNA-dependent RNA polymerase (Pontes et al., 2006). Conversely it has been found that the production of many sRNAs still takes place in mutants defective for the methylation-step of RdDM, highlighting the possibility that it is not the methylated DNA that acts as the template (Zhang et al., 2007). Clearly answering the question of what is the triggering event for this form of silencing and what features Pol IV actually recognizes is an important one in understanding how RdDM is initiated and how it becomes maintained.

Since the start of this PhD project some light has been shed on the recruitment of Pol IV but the issue remains not fully resolved. Work by Haag et al. (2012) has demonstrated that *in vitro* Pol IV uses DNA as a template but requires RNA primers, therefore the depiction of Pol IV in Figure 1.4 is correct. Additionally they reported that Pol IV and RDR2 physical associate *in vivo* and that the action of RDR2 as an RNA-dependent RNA polymerase is dependent on its association with Pol IV. However, Pol IV can function in the absence of RDR2 (Haag et al., 2012).

Another finding that may shed some light on the issue of Pol IV recruitment is the discovery by Law et al. (2011) that SHH1 co-purifies with NRPD1 of Pol IV. This has led to the finding that SHH1 can recruit Pol IV to some of its genomic targets through the binding of its unique tandem Tudor-like fold to unmethylated H3K4 and methylated H3K9me2 (Law et al., 2013). Law et al. (2013) also found that SHH1 is also needed for Pol IV occupancy at those sites. However, SHH1 is not required for recruitment of Pol IV at all targets. According to Law et al (2013), 84.2% of wild type targets of silencing are dependent on 24nt siRNAs with 81.4% of these requiring Pol IV. Of the Pol-IV dependent loci, 44% of these also require SHH1. Additionally, there is a location dependent aspect to this, with SHH1 having a stronger involvement with Pol IV-dependent silencing in the euchromatic arms than at the pericentromeric heterochromatin. Therefore an alternative mechanism of Pol IV recruitment must exist.

1.3.2.3 de novo DNA methylation by RdDM

According to the canonical RdDM pathway, once the siRNA have formed complexes with one of the AGO proteins (most commonly AGO4) this complex associates with Pol V via the C-terminal domain of the NRPE1 subunit of Pol V (Li et al., 2006, Qi et al., 2006). Li et al. (2006) have provided evidence that sRNA, AGO4 and Pol V co-localize in the nuclear sub-organelle Cajal bodies along with SmD3, a small nuclear RNA (snRNA) binding protein. Pol V produces long non-coding RNA transcripts which are thought to act as scaffolds to guide the siRNA-AGO4 complexes to their targets to be methylated (Wierzbicki et al., 2008, Wierzbicki et al., 2009). Although *in vitro* studies revealed that Pol V uses DNA as a template and ChIP data demonstrated that Pol V associates with the genomic loci it transcribes from, at the beginning of this project little was known with regards to the recruitment of the enzyme to the DNA in the first place (Haag et al., 2012, Wierzbicki et al., 2008, Wierzbicki et al., 2009).

Pol V transcription is dependent on a group of proteins which together form what is known as the DDR complex. This complex consists of DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), DEFECTIVE MERISTEM SILENCING 3 (DMS3) and RNA-DIRECTED DNA METHYLATION 1 (RDM1). DRD1 is a chromatin-remodelling protein (Wierzbicki et al., 2008), DMS3 is a chromosome hinge domain protein (Greenberg et al., 2011) and RDM1 binds single-stranded methylated DNA (Wierzbicki et al., 2008, Greenberg et al., 2011, Gao et al., 2010). The transcription factor-like protein KTF1 also associates with Pol V and AGO4 and is thought to aid the recruitment of enzymes which modify chromatin (He et al., 2009, Huang et al., 2009). The microrchidia (MORC) ATPase MORC6 is also proposed to be involved with the DDR complex at some loci (Brabbs et al., 2013, Lorkovic et al., 2012, Moissiard et al., 2012). Together, the Pol V transcripts, siRNA-AGO4 and DDR complex trigger the recruitment of DRM2 to the target DNA which directs the *de novo* DNA methylation in all sequence contexts and completes the cycle (Naumann et al., 2011) (Figure 1.4).

1.3.2.4 The recruitment of Pol V

Before the start of this project a few transcription sites of Pol V had been identified but little was known about how Pol V was recruited to these sites. Most of these were intergenic sites which produced long RNA transcripts which could be used as scaffolds by the DDR complex (Wierzbicki et al., 2008). Genomic regions of Pol V transcription are adjacent to the area silenced by RdDM (Wierzbicki et al., 2008). During the time frame of this project, Zhong et al. (2012) performed ChIP-Seq with NRPE1. Their work revealed that NRPE1 is enriched at pericentromromeric heterochromatin regions but not at the centres of the centromeres. Additionally, gene promoters and evolutionarily young transposable elements had much higher levels of enrichment with 55%

of promoters that associated with NRPE1 corresponding to transposable elements (Zhong et al., 2012). They hypothesised that it is Pol V's evolutionally relationship with Pol II that has allowed Pol V to retain the ability to bind to promoters due to retained subunits composition.

Additionally, Pol V recruitment to some loci can be mediated by SU(VAR)3-9 HOMOLOG 2 (SUVH2), SUVH9 and SUVR2 (Johnson et al., 2014, Matzke & Mosher, 2014). SUVH2 and SUVH9 are SET-and RING-ASSOCITED (SRA) domain proteins which act redundantly to recruit Pol V to methylated DNA (Johnson et al., 2014). Although similar to KYP, SUVH5 and SUVH6 of the CMT3 pathway, SUVH2 and SUVH9 do not have histone methyltransferase activity (Johnson et al., 2014).

1.3.2.5 Self-reinforcement of RdDM

Figure 1.4 shows RdDM as a loop, which can become self-enforced. The previously discussed SUVH2 and SUVH9 proteins give an example as to how this might be achieved. At some loci these proteins bind methylated DNA and recruit Pol V to it (Johnson et al., 2014). This would enforce silencing of those loci by continually enlisting Pol V and the methylation machinery of RdDM to those sites (Johnson et al., 2014, Meyer, 2011).

RdDM has been linked to several histone-modifying proteins which implies that RdDM can mediate changes to the chromatin structure. This helps to maintain and reinforce silencing at the target loci. For example, the histone-modifying proteins JUMONJI C protein JMJ14 and histone deacetylase HDA6 both remove active marks from histones and are linked to RdDM as their mutants have a different methylation pattern to wild type plants (Aufsatz et al., 2002, Le Masson et al., 2012, Meyer, 2011). Additionally, the repressive histone mark H3K9me2 has an important relationship with RdDM as demonstrated by its binding with SHH1 (Law et al., 2013).

1.3.3 Non-canonical RdDM pathways

There are many examples where the RdDM pathway does not work in the canonical manner. One example of non-canonical RdDM is the involvement of Pol II in the place of Pol V. It has been shown that at some low copy number intergenic loci it is transcription by Pol II that recruits Pol IV and Pol V to perform RdDM (Zheng et al., 2009). Additionally there are some loci that become methylated in the absence of Pol IV mediated siRNA biogenesis and Pol V transcription as Pol II produces the scaffold RNA transcripts needed to trigger silencing (Zheng et al., 2009, Gao et al., 2010). Findings that support this include reports that RDM1 of the DDR complex can co-localise with Pol II in the nucleoplasm (Gao et al., 2010).

Another pathway that demonstrates a clear connection between PTGS and RdDM is the NEEDED FOR RDR2-INDEPENDENT DNA METHYLATION (NERD) pathway (Garcia et al., 2012, Pontier et al.,

2012). NERD is a GW repeat protein that has a PHD finger module which allows binding to H3K4me0. At a subset of loci NERD binds DNA as determined by ChIP and is proposed to interact with AGO2 and RDR6 to bring about DNA methylation via 21nt sRNAs and possibly the action of Pol V (Pontier et al., 2012). NERD mutants have reduced DNA methylation at their target loci, which can increase gene expression at a subset of these.

1.3.4 Silencing of Transposable elements

One of the main targets of the RdDM pathway are Transposable Elements (TEs). The mobilization of TEs tends to have a negative effect on the genome and DNA methylation is used to prevent this and therefore acts to protect the genome (Lisch & Bennetzen, 2011). Nearly all permutations of RNA silencing target TEs in some form, supporting the hypothesis that TEs are one of the main driving forces behind the evolution of the sRNA-directed epigenetic pathways including RdDM (Lisch & Bennetzen, 2011, Willmann et al., 2011).

Interestingly there are points in the life-cycles of angiosperms when TEs become un-silenced. This is in the vegetative nucleus of pollen and egg cell (central cell) during gametogenesis (Slotkin et al., 2009). Neither of these cells actually contributes to the germ-line but have an important supporting role for the germ cells and early embryo. The current hypothesis used to explain the activation of TE in these cells is that this causes high levels of siRNAs to be produced which are then transferred into the germ line cell, ensuring silencing of these TEs occurs in the next generation (Discussed by Slotkin et al., 2009, Lisch & Bennetzen, 2011). For the TEs to become unsilenced in the vegetative nucleus of pollen there is a reduction in the production of the 24nt siRNAs associated with RdDM (Creasey et al., 2014, Cui & Cao, 2014). Instead there is an increase in the levels of 21nt miRNA and RDR6 dependent sRNAs, and it is these which travel in to the sperm from the vegetative nucleus.

TE methylation is not conducted by RdDM exclusively. In euchromatin TE methylation is predominantly conducted by RdDM whereas in heterochromatic regions TE methylation requires Decrease in DNA METHYLATION 1 (DDM1) and the DNA methyltransferase CMT2 and is independent of RdDM (Zemach et al., 2013).

1.4 Overlap between PTGS and DNA methylation

Overlap between PTGS and DNA methylation has also been observed in the silencing of transgenes. Transgenes that undergo PTGS can also support RdDM, although this is in the form of gene body methylation instead of methylation of promoter regions as seen for TGS (English et al., 1996, Ingelbrecht et al., 1994, Jones et al., 1998, Jones et al., 1999, Morel et al., 2000, Sijen et al., 1996). Additionally, components of the RdDM pathway have been found to be required for the

maintenance of silencing for some transgenes (Eamens et al., 2008). How the overlap between PTGS and DNA methylation occurs is poorly understood and will be one of the main focuses of this work. Additionally, there are many examples of transgene silencing where PTGS does not overlap with DNA methylation so determining what triggers PTGS-dependent RdDM is also a question to address (Goodwin et al., 1996, Vanblokland et al., 1994).

Another way of viewing PTGS and DNA methylation is to ask how DNA methylation affects PTGS as opposed to how does PTGS support DNA methylation. The relationship between PTGS, TGS and maintenance of DNA methylation by methyltransferases described in section 1.3.1 has been investigated by Morel et al. (2000). Using mutants of *ddm1* and *met1* they observed a release from TGS of a GUS reporter gene under the 35S promoter. PTGS on the other hand underwent stochastic release of silencing.

There are other examples of overlap between PTGS and RdDM, which will be briefly described here. One of the ways that PTGS and RdDM overlap is in defence of the genome from invading DNA. The most studied example of this is the defence of the genome from virus invasion. The second, and least understood, is how the genome is protected from new TEs. The conundrum with both is how sequence specific silencing is triggered when the sequence of the invader is new and unknown to the plant.

The conundrum of recognising invading nucleic acids in the case of viruses can be partly explained due to the way in which many plant viruses replicate. RNA viruses can present a template for the action of DCLs by the presence of a dsRNA at some stage of the virus life cycle either by replicating via a dsRNA intermediate or by producing ssRNA with a stem loop structure (Matzke & Mosher, 2014, Willmann et al., 2011). All DCLs have been shown to play a role in the defence of the genome from different viruses (Willmann et al., 2011). The role of RDR6 was discussed earlier in this introduction. The dsRNA produced by the invading viruses is processed by a DCL which produced siRNA that can target virus RNA. The cleavage products produced can act as templates for RDRs so that secondary siRNAs are produced. Plants with mutations in RDR1, RDR2 and RDR6 have been shown to be more susceptible to some plant viruses (Willmann et al., 2011). Lastly systematic spreading of the silencing signal needs to occur for the plant to become resistant.

When it comes to the conundrum of recognising invading nucleic acids in the case of TEs, much is known about the suppression of existing TEs but less is known about the invasion of new TEs. Recent work by Marí-Ordonz *et al.* (2013) has reconstructed a TE "burst" for the endogenous retrotransposon *Evadé* (*EVD*) to simulate *de novo* invasion (Cui & Cao, 2014). *EVD* usually has a copy number of one in wild type plants. However, in lineages of epigenetic recombination inbred lines (epiRILs), *EVD* managed to acquire a high copy number (~40) over several generations before

it was ultimately silenced at generation F14, resulting in the discontinued proliferation of the TE. The RDR6 dependent PTGS pathway was initially raised as the defence against the TE, producing 21/22nt siRNAs, but was thwarted as binding of *EVD* RNA to the TE's GAG capsid protein protected it. However, as the copy number of *EVD* increased over the generations non- canonical RdDM began to produce 24nt siRNAs which lead to the silencing of the TE. Marí-Ordonz *et al.* (2013) report that the siRNA biogenesis in this case may occur via Pol II and RDR6 before being processed by DCL3 to produce 24nt siRNAs. A similar observation was made by Nuthikattu et al. (2013). They have reported on roles for both Pol IV-dependent RdDM and RDR6-dependent RdDM for the silencing of TEs. They found that Pol IV-dependent RdDM can maintain silencing of TEs whereas RDR6-dependent RdDM does not maintain TE silencing but can recognize active TEs and initiate silencing against them (Nuthikattu et al., 2013). Together these reports demonstrate an overlap between the PTGS and RdDM pathways.

PTGS and the proteins from the RdDM pathway also show some overlap when it comes to the mobility of the RNA silencing signal throughout the plant. The movement of the RNA silencing signal is an important part of defence against viruses and allows younger tissue to become resistant to a virus after the infection of older tissue. There is still much debate as to what the mobile elements of RNA silencing actually is. Proposed candidates have included dsRNA or small RNAs of different sizes. The most likely of these is the small RNAs of either 21 or 24nt in size, however this is still debated (Molnar et al., 2011, Melnyk et al., 2011). Systemic movement occurs at the cell-to-cell level via the plasmodesmata or at the long-distance level via the phloem with both being needed for full systemic silencing. The long distance systemic spread can be dependent on the action of NRPD1, RDR2 and DCL3 to produce 24nt siRNAs which may be the mobile signal (Brosnan et al., 2007, Melnyk et al., 2011, Molnar et al., 2011). In cases where this occurs the recipient cells need RDR6 to be able to effectively receive the signal and hence complete the spread of silencing (Schwach et al., 2005).

1.5 Factors influencing transgene silencing

Transgenes present a common target of RNA silencing, as previously alluded. This can affect the ease of production of genetically modified plants (Meyer, 2013). Clearly not all transgenes that are introduced become the targets of silencing, as evidenced by the plethora of experiments on transgenic plants. However, it remains unclear why some transgenes become targets of RNA silencing when others do not. Additionally, targets that undergo PTGS can become methylated (Jones et al., 1998). There are a few theories that aim to explain the variation in transgene silencing.

1.5.1 Locus specific effects

One possible explanation as to the variability of transgene expression is the effect of the genomic location it has integrated into. Originally it was thought that if a transgene had integrated into a non-transcribed region this would explain its lack of expression. Further work has demonstrated that RNA silencing plays a big part in preventing transgene expression. However, the ability of a transgene to support silencing may still be influenced by genome location (Jones et al., 1998).

Furthermore, Fischer et al. (2008) have observed that RNA-based silencing of a transgene can be affected by the DNA features closest to its site of integration, such as methylation or active gene transcription. Fischer et al. (2008) performed a systematic investigation of well-characterised single copy transgenes at known chromosomal position. They report that TGS and RdDM are affected by the type of DNA structures adjacent to the transgene, with near-by repeats correlating with silencing and neighbouring genes reducing the levels of gene silencing. They also observed that DNA methylation of the promoter region was necessary for TGS but not sufficient to trigger TGS alone in this example.

TEs and repeat elements are common targets of RNA silencing and RdDM. Is it possible that transgenes that insert next to, or within a region of, TEs more readily undergo RNA silencing and methylation as a result of their proximity to these? In maize spreading of methylation from some TEs from a subset of families to adjacent genes has been observed (Eichten et al., 2012, Gent et al., 2013).

Additionally, work on TEs by Zemach *et al*. (2013) has shown that there may be a location-dependent effect with regards to methylation of TEs. In euchromatin TE methylation is predominantly conducted by RdDM whereas in heterochromatic regions TE methylation requires DDM1 and CMT2 and is independent of RdDM (Zemach et al., 2013). Work in maize has shown that 24nt based silencing of TEs is more common when the TE in question is close to genes (Gent et al., 2013). Therefore transgenes that integrates into heterochromatic regions may have differences in silencing when compared to transgenes in euchromatic regions.

One of the key questions is does this methylation make these transgenes targets for the action of RDR6? Many transgenes can undergo RDR6 amplification but this does not always occur; a difference that could be explained by locus-specific effects. It raises the possibility that RDR6 may be interacting at the DNA level with the template producing-loci. Consistent with this idea it has been shown that RDR6-GFP fusion proteins localize in the nucleus which would means RDR6 could have access to DNA and DNA-associated proteins (Luo & Chen, 2007, Kumakura et al., 2009).

Furthermore, it has been observed that the DNA of genes that have undergone this amplification often become methylated via RdDM (Jones et al., 1999).

Another way in which locus specific effects can occur is due to properties of the transgene locus itself. These properties could include the complex integration of repeats of the transgene at a single integration site for example. Studies in a variety of plant species have shown that transgenes with multiple copies tend to undergo RNA silencing (Hobbs et al., 1993, Schubert et al., 2004) and may also undergo DNA methylation, especially when the integration structures are complex (reviewed by Butaye et al., 2005, Stam et al., 1998). Additionally, when transgenes are integrated into plants via *Agrobacterium* mediated transformation the T-DNA insert can carry part of the vector sequence (Kononov et al., 1997). The addition of this vector sequence in T-DNA has been connected to increased silencing of transgenes (Iglesias et al., 1997, Jakowitsch et al., 1999).

1.5.2 Transcript threshold

Alternative theories as to why transgenes have varying abilities to undergo silencing include the concept of a transcript threshold (Lechtenberg et al., 2003, Schubert et al., 2004). According to this theory there is a transcript threshold for transgenes or repeat elements which if exceeded results in silencing. In other words high levels of transgene expression could lead to silencing.

1.5.3 Matrix scaffold attachment regions

Matrix scaffold attachment regions (MARs) are conserved areas of eukaryotic genomes which are thought to attach to the proteinaceous matrix which may organise DNA within the nucleus. The proposed action of MARs is to cause the formation of DNA loops that stick out from the chromatin between the two MARs. MARs were proposed to protect the genes within the loop from chromosomal effects and there are reports that the inclusion of MARs in T-DNA cassettes increases the expression rates of transgenes and reduces the levels of transgene variation (reviewed by Allen et al., 2000). For example, Petersen et al. (2002) working in barley have reported that adding MARs to each side of the transgene cassette gave a two fold increase in expression of a transgene (Petersen et al., 2002).

1.6 Project aims

During this introduction I have highlighted some of the remaining questions in the field of RNA silencing and the silencing of transgenes. Questions of particular interest to me include, why some targets can support RDR6-dependent RNA silencing whereas others do not. Additionally, I was interested in why some transgenes are able to undergo RdDM when others do not. To answer these questions there were two main aspects I wished to investigate. Firstly, the locus-

specific effects of transgene insertion on RDR6 activity and RdDM, and secondly the recruitment of key proteins to targets of PTGS and/or RdDM.

To address the issue of the effect of genome location on silencing I planned to use a dual transgene system to investigate if the site of integration would affect the transgenes ability to undergo RDR6-dependent silencing. Multiple independent single locus GFP transgenic Arabidopsis lines were generated that were triggered to undergo silencing by crossing with an amplicon trigger line (AMP243)(Dalmay et al., 2000a). The lines were characterized in terms of transgene structure and integration site and the ability of these lines to undergo PTGS and/or RdDM was investigated in an attempt to correlate the characteristics of the transgene to silencing outcomes. Using a dual transgene system allowed differences in transgenes properties before and after silencing to be investigated. These properties included DNA methylation status, siRNA accumulation and histone modifications.

To address the issue of recruitment, immunoprecipitation experiments were planned that would allow the investigation of the involvement of key proteins in silencing. At the start of this project results from only a few immunoprecipitation experiments with key proteins had been published by other groups. This is an active area of research and consequently many results of immunoprecipitation experiments have been published over the past 4 years. A selection of these will be discussed in this thesis. To conduct immunoprecipitation experiments I planned to create epitope tagged versions of key proteins and to investigate whether some of these proteins are present at loci prior to the onset of silencing or whether they require siRNA biogenesis for their recruitment.

Chapter 2 Materials and Methods

2.1 Nucleic acid extraction and manipulation

2.1.1 Arabidopsis genomic DNA extraction

2.1.1.1 Commercial DNA extraction kits

To extract genomic DNA from *A. thaliana*, DNA extraction kits from Qiagen (http://www.qiagen.com/gb/) were used according to the manufacturer's instructions. When large amounts of DNA were required, for example in Southern Blotting, the DNeasy Plant Maxi kit was used with 1g of leaf material. For smaller tissue samples (100 mg) the DNeasy Plant Mini kit was used.

2.1.1.2 Cetrimonium bromide (CTAB) DNA extraction

Solutions required

1X CTAB buffer consists of: 2% (w/v) CTAB, 100 mM Tris pH 8, 20 mM EDTA pH 8, 1.4 M NaCl, 2% (w/v) Polyvinylpyrrolidinone (PVP)

Method

CTAB buffer was preheated to 65°C. Plant tissue (3 to 5 leaves) that had been snap frozen in an Eppendorf tube using liquid nitrogen was crushed with a pipette tip. 0.5 ml of preheated CTAB buffer was added and the sample incubated at 65°C for 3 hours. The sample was then centrifuged at 13,000 rpm for 10 minutes before the supernatant was transferred to a clean Eppendorf tube. 0.5 ml of Chloroform:isoamyl alcohol (24.1) was added and the sample mixed by inversion prior at 13, 000 rpm for 5 minutes at room temperature. The upper aqueous layer of the supernatant was transferred to a clean Eppendofr tube and an equal volume of isopropanol added to precipitate the DNA. This was centrifuged at 13,000 rpm for 15 minutes to pellet the DNA. The pellet was washed in 1 ml of 70% ethanol. The ethanol was then removed following centrifugation for 1 minute and the pellet allowed to air dry for 10 minutes. The pellet was then resuspended in 50 µl TE pH 8 buffer. One unit of RNase A (Qiagen) could be used at this stage to remove RNA in the sample.

2.1.2 Plasmid DNA extraction

The extraction of plasmid DNA was performed using the GeneJET plasmid mini prep kit from Thermo Scientific according to manufacturer's instruction.

2.1.3 DNA extraction from Agarose gel

To extract DNA from agarose gels the GeneJET gel extraction kit from Thermo Scientific was used in accordance with manufacturer's instruction.

2.1.4 RNA extraction

2.1.4.1 Commerical RNA extraction kits

For the extraction of mRNA the RNeasy Plant mini kit from Qiagen was used in accordance with the manufacturer's instructions.

2.1.4.2 Trizol-based siRNA extraction

To approximately 100 mg of frozen plant tissue that has been powered in the presence of liquid nitrogen 1ml of Trizol (LifeSciences) was added, thoroughly mixed and incubate for 10 minutes at room temperature. Samples were sometimes frozen at -80°C at this point. Then 0.2 ml of chloroform was added and the samples, mixed by inversion and incubated at room temperature for 5 minutes. The samples were then spun for 10 minutes at 13,000 rpm in a microcentrifuge at 4°C. The upper aqueous layer was then transferred to a clean Eppendorf and 0.5 ml of isopropanol added to precipitate the RNA, followed by centrifugation at 13,000 rpm for 25 minutes at 4°C to pellet the RNA. The supernatant was removed and the RNA pellet was washed by the addition of 1 ml of 80% ethanol and centrifugation at 13,000 rpm for 5 minutes at 4°C. All ethanol was removed and the pellet air-dried for 10 minutes. The pellet was then resuspended in 30 µl of 50% formamide.

2.1.5 Production of cDNA

To produce cDNA from total RNA, the first-stand cDNA synthesis kit from Thermo Scientific was used according to the manufacturer's instructions.

2.1.6 Precipitation of DNA

DNA samples that had undergone restriction digest, as per manufacturer's instructions, were concentrated by using Sodium Acetate and 100% ethanol to precipitate the DNA. After samples had been digested, one tenth volume of 3 M Sodium Acetate (pH 5.8) and 2.5 volumes of 100% EtOH were added to each samples and placed over night at -20°C. Once the DNA was precipitated it was pelleted by centrifugation in a microcentrifuge for 20 minutes at 13,000 rpm at 4°C. The pellet was washed with 1 ml of 70% EtOH before all ethanol was removed using centrifugation and then air dried. The pellet was then resuspended in 20 µl dH_2O

2.1.7 Agarose gel electrophoresis

2.1.7.1 Solutions required

0.5 X Tris-Borate EDTA (TBE) buffer

1 l of 5 X consists of 0.45 M Tris, 0.45 M Boric Acid and 0.01 M EDTA pH8. 9 l of H_2O was added to this to produce a 0.5 X buffer.

2.1.7.2 Method

DNA was separated using gel electrophoresis by making a 1% (w/v) agarose gel with 0.5 X TBE and running the gel in 0.5 X TBE. DNA could be imaged by soaking the gel in an ethidium bromide and 0.5 x TBE solution for 10 minutes before visualization on a transilluminator and imaged using GeneSnap (Syngene).

2.2 PCR

The PCR machine used for all PCR reactions performed here was the DNA Engine DYAD (MJ Research)

2.2.1 Primers

Primers were ordered from EuroFin MWG, Integrated DNA Technologies or Sigma- Aldrich.

Table 2.1 Primers used for PCR reactions in this project.

Primer	Sequence	Used for	Tm Used for PCR
TAILprimer 1	ATTTTCGCCTGCTGGGGCAAACCAGCGTGGA	TAIL PCR	Variable
TAILprimer 2	AACTCTCTCAGGGCCAGGCGGTGAAGGGCAAT	TAIL PCR	Variable
TAILprimer 3	GTTGCCCGTCTCACTGGTGAAAAGAAAAACCACC	TAIL PCR	Variable
AD1	NGTCGASWGANAWGAA	TAIL PCR	Variable
AD2	TGWGNAGSANCASAGA	TAIL PCR	Variable
AD3	AGWGNAGWANCAWAGG	TAIL PCR	Variable
AD6	WGTGNAGWANCANAGA	TAIL PCR	Variable
RDR6-Nterm	CACCATGGGGTCAGAGGGAAATATG	Construct generation	58
RDR6-STOP	TTAGAGACGCTGAGCAAGAAACTTAGC	Construct generation	58
RDR6-NO STOP	GAGACGCTGAGCAAGAAACTTAGCC	Construct generation	58
RDR6-Prom	CACCTGTTCTCTGTTGCTACCTTTAC	Construct generation	58
RDR6 5' Control	AGATTGGGACATTGGTTTCTCGGG	Control	55
RDR6 3' Control	TGGAGAGCTCAACTTCTGGGGG	Control	55
RDR2-Nterm	CACCATGGTGTCAGAGACGACG	Construct generation	55
RDR2-STOP	TCAAATGGATACAAGTCCACTTGTTTTCTC	Construct generation	55
RDR2-NO STOP	AATGGATACAAGTCCACTTGTTTTCTC	Construct generation	55
RDR2-Prom	CACCATGTGTTTGATTTTAAGGTACTC	Construct generation	55
RDR2 Control 5'	TGTTAGACATTTTCTCGGGC	Control	50
RDR2 Control 3'	AGAGAGAGCATTTGCGGG	Control	50
RDR2 NO STOP_2	AATGGATACAAGTCCACTTGTTTTCTCTTC	Generation of Construct number 6	57
Pol4-Nterm	CACCATGGAAGACGATTGTGAGGAGC	Construct generation	57
Pol4-STOP	TCACGGGTTTTCGGAGAAACCAC	Construct generation	57
Pol4-NO STOP	CGGGTTTTCGGAGAAACCACCG	Construct generation	57

Pol4 5'control	CTGCTGTTTGCCGTTCCG	Control	52
Pol4 3' Control	GGAATCCTCTTCCTTATCCTCC	Control	52
M13F	GTAAAACGACGGCCAGT	PCR and Sequencing	55
M13R	GGAAACAGCTATGACCATG	PCR and Sequencing	55
Bisulfite F primer	CCTTCRCAARACCCTTCCTCTATATAA	PCR and Sequencing	60
Bisulfite R primer	TATAAGAGAAAGTAGTGAYAAGTGTTGG.	PCR and Sequencing	60
rdr6-11 forward	TACTGTCCCTGGCGATCTCT	Genotyping rdr6-11 mutants	54
rdr6-11 Reverse	CCACCTCACACGTTCCTCTT	Genotyping rdr6-11 mutants	54
VA87 MLS	TCCAATACCTCTGGCTCTGC	MLS control for copy number qPCR	60
VA86 MLS	TGGATTGAACCCAAAACAGTG	MLS control for copy number qPCR	60
qGFP5 For	TCAAGGAGGACGGAAACATC	GFP copy number qPCR	60
qGFP5 Rev	AAAGGGCAGATTGTGTGGAC	GFP copy number qPCR	60
Tubulin new F	CCAACCACTTCGTCTTCTTCA	qPCR on ChIP data	60
Tubulin new R	CTTGGTATCTTCCCGTCGAA	qPCR on ChIP data	60
LTR 1 Matzke F	ATTAACCACACGCTGCATCTAAT	qPCR on ChIP data	60
LTR 1 Matzke R	GAGAGACGATCGTGCTGATAAC	qPCR on ChIP data	60
GFP3' New F	ACAGCTGCTGGGATTACACA	qPCR on ChIP data	60
GFP3' new R	TTGCGGGACTCTAATCATAAAAA	qPCR on ChIP data	60
GFP ChIP 5'R	TTGGCCATGGAACAGGTAGT	qPCR on ChIP data	60
GFP ChIP 5' F	CACGGGGGACTCTAGAGGAT	qPCR on ChIP data	60

2.2.2 Taq PCR

For a 20 µl reaction the following components were mixed on ice:

- X µl of dH_2O (depending of the volume of DNA used)
- 2 µl of 10 X ThermoPol Buffer (New England Biolabs)
- 0.4 µl of 10mM dNTPs (10mM of dATP, dTTP, dCTP & dGTP each from Fermentas)
- 0.2 µl of Taq DNA polymerase (New England Biolabs)
- 1 µl of Forward Primer at 10 µM
- 1 µl of Reverse Primer at 10 µM
- X µl DNA

The standard run conditions were

1. 95°C for 2 minutes
2. 95°C for 20 seconds
3. X°C for 20 seconds
4. 72°C for X seconds
5. Repeat steps 2 to 4 for 34 times
6. 12°C forever (until sample removed from machine)

The annealing temperature in step 2 depended on the primers used and the elongation time in step 3 depended on the size of the fragment generated. Taq polymerase takes 1 minute to amplify 1 Kb of DNA

2.2.3 Variations of the Taq PCR

For some PCR reactions, a DNA polymerase with a proof reading function was required. For this the PfuUltra IIF (Stratagene) was used. The conditions for the Taq PCR were altered so that the 0.4 µl of the enzyme was added per 20 µl and the appropriate PCR buffer was used. The run conditions were the same expect that a hot start was used.

For colony PCR the volume of DNA that would be added was replaced by water and the template DNA was provided by scraping a sample of the colony to be analysed into the PCR mix using a clean pipette tip.

2.2.4 qPCR

The iQ SYBR GreenSupermix (Bio-Rad Laboratories, www.bio-rad.com) and the MyiQ Real-Time PCR detection system was used to preform qPCR reactions in this project a according to the manufacturer's instructions (Bio-Rad Laboratories).

The standard reaction mix was;

- 10 µl SYBR GreenSuper mix
- 1.6 µl Forward Primer at 10 µM
- 1.6 µl Reverse Primer at 10 µM
- 4.8 µl dH_2O

- 2 µl DNA template

The reaction cycle was:

1. 95°C for 10 seconds
2. 60°C for 15 seconds
3. 72°C for 30 seconds
4. Repeat steps 2 to 4 for 39 times
5. 12°C forever (until sample removed from machine)

2.2.5 Thermal Asymmetric Interlaced PCR (TAIL PCR)

Three sets of PCR reactions were performed for each sample using one of the three TAIL-specific primers. Four arbitrary degenerate primers (AD) were used so that for each TAIL-specific primer four PCR reactions were performed per sample.

2.2.5.1 Primary reaction

4 reactions per sample were set up on ice – one for each of the AD primers, AD1, AD2, AD3 and AD6

- 12.6 µl of dH_2O
- 2 µl of 10 X ThermoPol Buffer (New England BioLabs)
- 0.2 µl of 10 mM dNTPs (10mM of dATP, dTTP, dCTP & dGTP each from Fermentas)
- 2 µl of TAIL-Specific primer 1 (to give a final concentration of 0.15 µM)
- 2 µl of AD Primer (to give a final concentration of 2 µM)
- 0.2 µl of Taq DNA polymerase (New England BioLabs)
- 1 µl Genomic DNA

The reactions were placed in the PCR cycler and ran as follows:

1. 92°C for 3 minutes then 95°C for 1 minute
2. 94°C for 30 seconds then 65°C for 1 minute and then 72°C for 2 minutes
3. Repeat step 2 four times
4. 95°C for 30 seconds then 25°C for 2 minutes then ramp to 72°C over 2 minutes then hold at 72°C for 2 minutes
5. 94°C for 30 seconds then 65°C for 1 minute then 72°C for 2 minutes
   94°C for 30 seconds then 65°C for 1 minute then 72°C for 2 minutes
   94°C for 30 seconds then 44°C for 1 minute then 72°C for 2 minutes
6. Repeat step 5, 14 times
7. 72°C for 5 minutes
8. 4°C forever

2.2.5.2 Secondary reaction

The PCR product from each primary reaction was diluted 1 in 40 with dH_2O. 4 reactions per sample were set up – one for each of the AD primers, AD1, AD2, AD3 and AD6

- 16.05 µl of dH_2O
- 2.5 µl of 10 X ThermoPol Buffer (New England BioLabs)
- 0.25 µl of 10mM dNTPs
- 2.5 µl of TAIL-Specific primer 2 (to give a final concentration of 0.2 µM)

- 2.5 µl of AD Primer with a final concentration of 2 µM
- 0.2 µl of Taq DNA polymerase (New England BioLabs)
- 1 µl 1/40 diluted DNA from primary PCR reaction

The reactions were placed in the PCR cycler and ran as follows:

1. 94°C for 3 seconds then 65°C for 1 minute then 72°C for 2 minutes
2. 94°C for 3 seconds then 65°C for 1 minute then 72°C for 2 minutes
3. 94°C for 3 seconds then 45°C for 1 minute then 72°C for 2 minutes
4. Repeat steps 1 to 3 11 times
5. 72°C for 5 minutes
6. 12°C forever

2.2.5.3 Tertiary reaction

The PCR product from the primary reaction was diluted 1 in 40 with dH_2O. Four reactions per sample were set up – one for each of the AD primers, AD1, AD2, AD3 and AD6

- 30.1 µl of dH_2O
- 5 µl of 10 X ThermoPol Buffer (New England BioLabs)
- 0.5 µl of 10mM dNTPs
- 5 µl of TAIL-Specific primer 3 with a final concentration of 0.2 µM
- 5 µl of AD Primer with a final concentration of 2 µM
- 0.4 µl of Taq DNA polymerase (New England BioLabs)
- 4 µl 1/40 diluted DNA from primary PCR reaction

The reactions were placed in the PCR cycler and ran as follows:

1. 94°C for 30 seconds
2. 45°C for 1 minute
3. 72°C of 2 minutes
4. Repeat steps 1 to 3 19 times
5. 72°C for 5 minutes
6. 4°C forever

2 µl of PCR product from each of the secondary and tertiary reactions were separated by gel electrophoresis using a 1.2% agarose gel. From this the AD reactions that gave clear bands were identified and the whole PCR reaction was run out on another 1.2% agarose gel. The band was then excised from the gel and DNA was extracted as in section 2.1.3. The DNA was then sent for Sanger sequencing with the TAIL-Specific primer 3.

Sequencing results were then used in BLAST analysis (Altschul et al., 1997) to locate the transgene insertion site. This site was then entered into the Epigenome browser ANNO-J to review the epigenetic landscape in that region (Tonti-Filippini, 2008).

2.3 Manipulation of Bacteria and plasmids

2.3.1 Bacterial growth

Bacteria were grown either as a culture in liquid Luria broth (LB) or on solid LB or were stored long term in LB containg 10% glycerol at -80^{o}C. Antibiotics were added as required. *E. coli* was grown at 37^{o}C over night and *Agrobacterium* at 28^{o}C for approximately 48 hours.

Liquid LB comprised 5g tryptone, 2.5 g Yeast Extract and 2.5 g Sodium chloride in 500 ml dH_2O. This was then autoclaved and subsequently handled in sterile laminar flow hoods.

To produce solid LB media 5 g of Bacto-Agar was included in the production of 500 ml liquid LB. This was then autoclaved and subsequently handled in sterile laminar flow hoods. The media was allowed to cool to approximately 40 ^{o}C before the addition of chemicals for selection. Media was poured into Petri dishes and set in a laminar flow hood.

To produce glycerol LB 50 ml of glycerol was included in the production of 500 ml liquid LB. This was then autoclaved and subsequently handled in sterile laminar flow hoods. Bacteria were grown on plates under selection at the appropriate temperature and length of time before being resuspended in 2 ml of glycerol LB. 1 ml aliquots were stored at -80^{o}C.

2.3.2 Cloning

Cloning of PCR fragments into the pENTR vector were conducted using the pENTR Directional TOPO cloning kits from Invitrogen according to manufacturer's instructions.

Recombination of the pENTR vector into the Binary destination vectors were conducted first by linearised the pENTR vector with the appropriate restriction enzyme. The Gene JET gel extraction kit from Thermo Scientific was used according to manufacturer's instructions to purify the linearized plasmid which was run on a 1% agarose gel to check that linearization had occurred. Nano drop was used to find the concentration of the DNA sample. The Gateway LR reaction could then be conducted using the Gateway LR clonase II Enzyme mix from Invitrogen. Manufacturer's instructions were followed but with the amendment that all reaction volumes were halved.

2.3.3 Bacterial transformation

2.3.3.1 Transformation of chemically competent E. coli cells

Two different strains of *E. coli* were used for transformation reactions. The pENTR vector was transformed into TOPO10 One Shot Chemically Competent *E. coli* from Invitrogen. DH5 α competent cells from Invitrogen were transformed with the completed binary vectors via LR clonase reactions.

2 µl of cloning reaction was added to 50 µl of One Shot Chemically competent cells depending on the line of cells used. This was incubated on ice for 30 minutes before being subject to a heat shock for 30 seconds at 42°C and then rested on ice for 10 minutes. The cells were then grown in 250 µl of S.O.C medium from Invitrogen at 37°C for one hour whilst shaking. Aliquots of different volumes, typically 150 µl, 100 µl and 50 µl, were spread on selective LB Agar plates and incubated overnight at 37°C.

2.3.3.2 Transformation of plasmids into Agrobacterium tumefaciens via Electroporation

The completed Gateway binary vector constructs and the pBin:GFP binary vector were introduced into competent *Agrobacterium tumefaciens* using electroporation. Competent GV3101 *A. tumefaciens* were prepared by Louise Jones.

To perform the electroporation of the cells, a MicroPulser electroporator from BioRad was used according to the manufacturer's instructions. 1 µl of a 1:10 dilution of the plasmid solution was mixed with a 30 µl aliquot of electrocompetent cells and pipetted into an electroporator cuvette on ice. The cuvette was placed into the MicroPulser electroporator and the cells were shocked. 500 µl of LB liquid media was used to flush the cells out from the cuvette and into an Eppendorf tube to recover for 30 minutes at 28°C. Cells were then spread onto LB plates with selection and incubated at 28°C for 2 days.

2.4 Propagation and manipulation of Plants

2.4.1 *Arabidopsis thaliana* lines

Table 2.2 *Arabidopsis thaliana* lines used in this project

Line	Ecotype	Transgene	Reference
C24	C24	None	
Columbia (Col-0)	Col-0	None	
AMP243	C24	*35S-PVX:GFP*	(Dalmay et al., 2000a)
DAN2	C24	*35S:GFP* construct *and 35S-PVX:GFP* construct	(Dalmay et al., 2000a)
rdr6-11	Col	T-DNA insert	(Peragine et al., 2004)
rdr2-1	Col	T-DNA insert	(Xie et al., 2004)
SALK_059661	Col	T-DNA insert	(Arabidopsis Biological Resource Centre)

2.4.2 Growth media

Plants were either grown on an F2 soil mix or on Murashige and Skoog (MS) plates. Plates were made by adding 1.07 g of MS media and 5 g Agar to 500ml of dH_2O and adjusting the pH. The mixture was then autoclaved. If the plates were used to perform selection, then the necessary chemicals (either Kanamycin or PPT) were added immediately before pouring at a concentration of 50 μg/ml.

2.4.3 Chlorine gas sterilization of seeds

Surface sterilization of seeds was performed prior to growth on MS plates. To do this, open petri dishes or open Eppendorf tubes containing seeds were placed in an air tight container into which an open beaker of a mixture 3 ml concentrated HCl and 100 ml Haychlor Bleach (Brenntag) was placed. The seeds were exposed to the chlorine gas this mixture produced for 3 hours. Once sterilized the seeds were either sprinkled on to plates or placed on to plates with sterilized tooth picks.

2.4.4 Plant growth conditions

Seeds both on soil and on MS plates were stratified at 4°C for 2 to 3 days. Plants were either grown in glasses houses or in long day growth rooms (16/8 hour day/night). The day temperature of the growth room was approximately 21°C and the night temperature was around 19°C. The conditions in the glass house did vary more than those in the growth rooms due to the influence of the seasons however, the environmental control use in the glass house aimed to provide a minimum of a 16 hour day with a day temperature of 22°C and night temperature of 17°C.

2.4.5 Transforming Arabidopsis by floral dip

Arabidopsis was transformed by floral dip as described by (Bent & Clough, 1998). For each batch of transformation, 2 trays of soil grown plants were allowed to flower before being cut back to increase the number of inflorescences. 2 litres of liquid LB was prepared and then seeded with a 5 ml liquid LB culture of an *Agrobacterium tumefaciens* line carrying the desired binary vector. After the *A. tumefaciens* had grown for 2 days at 28°C the bacteria were pelleted using a benchtop centrifuge (13,000 rpm) and then resuspended in 5% sucrose and 500 µl per litre of Silwet L-77. The inflorescences were then dipped into this mixture and covered overnight in clear polythene. The plants were uncovered after 24 hours and then left to go to seed. Seeds from these plants could then undergo selection.

2.4.6 Crossing Plants

To cross different Arabidopsis lines together the immature unopened flowers of plants used as the female recipient were emasculated using fine tweezers to leave unpollenated stigmas. All inflorescence not involved in crossing were removed from the female plant. Mature pollen from the male parent was then used to pollenate the stigma which was covered in cling film overnight. The stem of the crosses were labelled with tape. If a discernible phenotype should have been present then this was checked prior to crossing. For example all 35S:GFP were visualized under a handheld UV light before crossing and the Zippy phenotype was looked for in *rdr6-11* mutant lines. The Zippy phenotype is the result of a shortened juvenile phase resulting in elongated, curled leafs, with an elevated number of trichomes when compared to the WT (Peragine et al., 2004).

2.4.7 Infiltration of *Nicotiana benthamiana*

5 ml of liquid LB was prepared and then seeded with an *A. tumefaciens* stain carrying the desired binary vector. After the *A. tumefaciens* had grown for 2 days at 28°C the bacteria were pelleted using a benchtop centrifuge and then resuspended in 10mM $MgCl_2$. Using a syringe, the *A. tumefaciens* mix was forced through the stomata of young *Nicotiana benthamiana* leafs. After 3-4 days expression of the protein of interest was screened for.

2.5 Blotting and detection of nucleic acids

2.5.1 Southern blot

2.5.1.1 Solutions required

20 X SSC

20 X SSC consists of 750 ml of dH_2O, 174.5 g of NaCl and 88.2 g of Na_3CIT which once dissolved was made up to a litre with dH_2O

Gel denaturing solution

Gel denaturing solution consists of 750 ml of dH_2O, 87.6 g of NaCl and 20 g of NaOH which once dissolved was made up to a litre with dH_2O

Gel neutralisation buffer

Gel neutralisation buffer consists of 750 ml of dH_2O, 175.3 g of NaCl and 60.6 g of Tris which once dissolved was pH to 6.5 with concentrated hydrochloric acid (HCl) and then made up to a litre with dH_2O

2.5.1.2 Method

DNA that had been digested with methylation-sensitive restriction enzymes (per manufacturer's instructions) and that had been precipitated by NaAc and EtOH was run on a 1% Agarose gel in 0.5 X TBE buffer. The gel was stained by soaking in a solution of Ethidium Bromide for 10 minutes to allow for visualization of the digested DNA. The DNA in the gel was denatured by placing the gel in an excess of denaturing solution for 20 minutes at room temperature whilst shaking. The gel was then washed twice in an excess of neutralisation buffer for 15 minutes. The DNA was transferred on to a Hybond Nx membrane (GE Healthcare) by capillary action in 20 X SSC. The transfer was allowed to happen overnight. The DNA was cross-linked to the membrane using the Stratalinker cross linker using the autocrosslink setting. Once dry the blot could be stored until it was probed.

2.5.2 Northern blot of mRNA

The concentration of the RNA samples was determined using a NanoDrop and 10 µl samples of RNA were prepared with equal amounts of RNA and sample buffer in each. The samples were heated to 80°C for 3 minutes prior to being loaded into a 1% agarose gel. This was run in 0.5 X TBE buffer as for a DNA gel. The gel was imaged following staining of the DNA with Ethidium Bromide for 10 minutes. The RNA was transferred onto hybond Nx membrane (GE Healthcare) by capillary action in 20 X SSC. The transfer was allowed to happen overnight. The RNA was cross

linked to the membrane using the Stratalinker cross linker with the auto setting. Once dry the membrane could be stored until it was probed.

2.5.3 Northern blot optimised for small RNA detection

The small RNA gel consisted of 30 ml of 30% Acrylamide/ bis-acrylamide, 25.2 g of Urea, 6 ml of 5 X TBE buffer and 5 ml of dH_2O heated to 65°C to dissolve the Urea. Once cooled, the polymerising agents were added. These were 30 µl TEMED and 120 µl of 25% ammonium persulfate. A vertical gel was cast using the gel. The wells were rinsed prior to sample loading to remove urea that leaches out of the gel. The gel was pre-run for an hour in 0.5 X buffer at 150V.

The concentration of the RNA samples was determined by using a NanoDrop so that equal amounts of RNA could be loaded. Each 20 µl sample of RNA was prepared by adding 4 µl of sample buffer. The samples were heated to 80°C for 3 minutes prior to being loaded into the gel.

The gel was initially run for 1 – 2 hours at 150 V in 0.5 X buffer before being turned down to 50 V and left to run over night. Once run, the gel was disassembled and then stained with ethidium bromide and visualized using a UV transilluminator.

RNA was transfer from the gel on to a Hybond –Nx membrane via electrophoretic transfer at 400 mA in 0.5 X for 2.5 hours. After this the RNA was cross linked to the membrane using the Stratalinker cross linker. Once dry the blot could be stored until it was probed.

2.5.4 Detection of nucleic acid by probing blots

2.5.4.1 Generation of probes

PCR fragments necessary for the probing of blots were generated by Taq PCR as previously described using Primers used are shown in Table 2.1 before being purified using the Quiquick PCR purification columns (Qiagen) per manufacturer's instructions. The probes were then radio-labelled by random primer based pairing using the Prime-It II Random Primer Labelling Kit from Agilent per manufacturer's instructions. The radioactive isotope used was α P^{32} –CTP nucleotide (Perkin Elmer).

Table 2.3 Primers Used to Generate Probes for Nucleotide Blots

Primer	Sequence	Used for	Tm Used for PCR
GFP 1	CAAGGAGATATAACAATGAAG	GFP Probe	50
GFP 8	CATGACGAACTCTAAGAGCTA	GFP Probe	50
5S F	GGATGCGATCATACCAG	5S Repeat probe	45
5S R	GAGGGATGCAMCACSAG	5S Repeat probe	45
Actin 5'	CATGGTTGGGATGAACCAGAAGGA	Actin probe	61
Actin 3'	GTCTCTTACAATTTCCCGCTCTGC	Actin probe	61
NPTII F	TGCTCCTGCCGAGAAAGTAT	Kanamycin resistance gene probe	55
NPTII R	AATATCACGGGTAGCCAACG	Kanamycin resistance gene probe	55

2.5.4.2 Church Buffer

Church buffer consists of 0.25 M Sodium Phosphate buffer (pH 7.2) and 7% SDS in dH_2O.

2.5.4.3 Pre-Hybridization

The membranes were placed in hybridisation bottles (Techne) with 10 ml of pre-heated Church buffer for an hour prior to the addition of the probe in a Techne Hybridiser HB-1D hybridisation oven. Southern blot and mRNA Northern blot membranes were treated at 65°C and small RNA northern blots at 42°C.

2.5.4.4 Hybridization

The P^{32}-labelled probes were heated to 100°C for 5 minutes and added directly to the hybridisation bottles (Techne) and allowed to hybridize overnight in the oven as for the pre-hybridisation step.

2.5.4.5 Washing excess probe from the blots

Two wash solutions were used; 2 X SSC 0.1% SDS and 0.1 X SSC 0.1% SDS. Small RNA Northern blots were washed twice in 2 X SSC 0.1% SDS for 20 minutes each with the solutions replaced between each wash. Southern blots and mRNA northern blots were washed twice in 2 X SSC 0.1% SDS for 20 minutes each and then washed twice in 0.1 X SSC 0.1% SDS for 20 minutes each. Again the solutions were replaced between each wash. Once washed the blots were wrapped in cling film.

2.5.4.6 Imaging P^{32}-labelled blots

The blots were either imaged using Hyperfilm autoradiograph film (GE Healthcare) or a Kodak phosphor screen (GE Healthcare). Blots exposed to the film were kept at -80°C for between 1 day

to 3 weeks depending on the strength of the signal. The films were developed in a compact X4 Xograph machine (Xograph Imaging Systems). Blots exposed to the phosphor screen were held at room temperature for 1 day to one week depending on the strength of the signal. These were then processed using the Personal Molecular Imager phosphorimager and software from BioRad.

2.6 Sequencing

2.6.1 Sanger sequencing

Sanger sequencing was conducted by the Bioscience Technology Facility at the University of York and by EuroFins MWG sequencing Services.

2.6.2 Bisulfite sequencing

For each sample, approximately 1 µg of genomic DNA was digested with 5 units of HindIII overnight at 37^{o}C in a volume of 200 µl. The digested DNA samples were cleaned up by the addition of 200 µl of phenol:chloroform and centrifuged for 5 minutes at 16,000 g in a microfuge. The supernatents were transferred to clean eppendorf tubes and 200 µl of chloroform added followed by centrifugation as for the previous step. One-tenth volume of 3 M Sodium Acetate and two volumes of EtOH were added to the supernatents and the samples placed at -20 ^{o}C for three hours. The samples were then centrifuged at 16,000 g for 30 minutes and pellets washed by the addition of 500 µl of 70% EtOH. Dried DNA pellets were dissolved in 20 ul of dH_2O and treated with sodium bisulfite using the EZ DNA Methylation- gold kit (Zymo Research) as per manufacturer's instructions. Following bisulfite treatment the samples were eluted into 20 µl dH_2O. Hot start PCR was then performed on these samples to amplify a portion of the *GFP* transgene. 1 µl of bisulfite treated DNA was used per 20 µl PCR reaction. Reaction was 1 x EpiMark reaction buffer, 200 µM dNTPs, 0.2 µM Bisulfite F primer, 0.2 µM Bisulfite R primer, 1 unit EpiMark Hot Start Taq polymerase (New England Biolabs).

The reactions were placed in the PCR cycler and ran as follows:

1. 95^{o}C for 30 seconds
2. 95^{o}C for 20 seconds
3. 60^{o}C for 30 seconds
4. 68^{o}C for 30 seconds
5. Repeat steps 1 to 4 39 times
6. 68^{o}C for 5 minutes

PCR products were purified using the Qiaquick PCR purification columns and either sent for sequencing directly or cloned into pGEM-T easy vector (Promega). Colonies were screened by colony PCR (M13 F and R primers) and positive products sent for sequencing. The DNA sequence was then analysed using the ClustalW2 tool from EMBL-EBI and CyMATE.

2.7 Chromatin Immunoprecipitation (ChIP)

The ChIP protocol used here is an adaption of the protocol from Werner Aufsatz on the Epigenome Network of Excellence Website (PROT12) (www.epigenome-noe.net/researchtools/protocol.php_protid=13.html).

2.7.1 Solutions required

Crosslinking Solution

The formaldehyde crosslinking solution final concentration of 2 M Sucrose, 1 M Tris-HCL (pH 8), 0.5 M EDTA, 37% formaldehyde and 200mM PMSF. This was prepared fresh for each experiment.

Extraction Buffer 1

Extraction buffer 1 had a final concentration of 0.4 M Sucrose, 10 mM Tris-HCL (pH 8), 10mM $MgCl_2$, 5 mM β-mercaptoethanol and 100 µl per 10 ml of protease inhibitors (Sigma P9599). This was prepared fresh for each experiment.

Extraction Buffer 2

Extraction buffer 2 had a final concentration of 0.25 M Sucrose, 10 mM Tris-HCL (pH 8), 10mM $MgCl_2$, 1% Triton X-100, 5 mM β-mercaptoethanol and 100 µl per 10 ml of protease inhibitors. This was prepared fresh for each experiment.

Extraction Buffer 3

Extraction buffer 3 had a final concentration of 1.7 M Sucrose, 10 mM Tris-HCL (pH 8), 2mM $MgCl_2$, 0.15% Triton X-100, 5 mM β-mercaptoethanol and 100 µl per 10 ml of protease inhibitors. This was prepared fresh for each experiment.

Nuclei Lysis Buffer

Nuclei lysis buffer had a final concentration of 50 mM Tris-HCL (pH 8), 10 mM EDTA, 1% SDS and 100 µl per 10 ml of protease inhibitors. This was prepared fresh for each experiment.

ChIP Dilution Buffer

ChIP Dilution buffer had a final concentration of 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCL (pH 8) and 167 mM NaCl. This was prepared fresh for each experiment.

Elution Buffer

Elution buffer had a final concentration of 1% SDS and 0.1 M $NaHCO_3$. This was prepared fresh for each experiment.

Low Salt Wash

Low salt wash buffer had a final concentration of 150 mM NaCl, 0.1 Triton X-100, 2 mM EDTA and 20 mM Tris-HCL (pH 8).

High Salt wash

High salt wash buffer had a final concentration of 500 mM NaCl, 0.1 Triton X-100, 2 mM EDTA and 20 mM Tris-HCL (pH 8).

LiCl Wash Buffer

LiCl wash buffer had a final concentration of 0.25 M LiCl, 1% Nondet P-40, 1% sodium deoxycholate, 1 mM EDTA and 10 mM Tris-HCL (pH 8).

TE Buffer

TE buffer used water as a solvent and had a final concentration of 10 mM Tris-HCL (pH 8) and 1 mM EDTA.

2.7.2 Chromatin crosslinking

The samples were crosslinked by immersing whole tissue, which had been rinsed in dH_2O, into 15 ml of Formaldehyde solution. This was left under vacuum for 15 minutes. To quench the crosslinking, 1.5 ml of 1 M glycine was added and samples incubated under vacuum for 10 minutes. Tissue was then washed again in dH_2O, dried lightly in blue roll, wrapped in tin foil and frozen in liquid nitrogen.

2.7.3 Chromatin preparation

Crosslinked tissue was ground to a fine powder in a pestle and mortar in the presence of liquid nitrogen. To this, 30 ml of cold extraction buffer 1 was added and the mixture was past though miracloth twice. The extract was centrifuged at 4000 rpm for 20 minutes at $4^{\circ}C$. The resulting pellet was resuspended 1 ml of extraction buffer 2. This was then transferred to an Eppendorf tube and at 13,000 rpm for 10 minutes at $4^{\circ}C$. The pellet was resuspended in 300 µl of extraction buffer 2 and floated on top of 300 µl of extraction buffer 3 in an Eppendorf tube. This was spun for 1 hour at 13000 rpm. The pellet was resuspended in 500 µl of cold nuclei lysis buffer and underwent sonication to break the DNA up. The sample was sonicated using the Bandelin Sonoplus HD 2070 with MS 73 probe with a setting of 4 X 10 seconds, 40% duty cycle and 20% power.

2.7.4 Pre-clearing and immune precipitation

The volume of the samples was increased to 1.5 ml using the ChIP dilution buffer. 40 µl of protein G agarose beads pre-absorbed with sheared salmon sperm DNA (Millipore) were equilibrated in ChIP dilution buffer for each sample. The Chromatin was pre-cleared with the agarose beads for 1 hour at 4°C. The beads were then removed by centrifugation at 5,000 rpm for 30 seconds. 60 µl of the pre-cleared chromatin was taken at this stage to serves as an input control for the later analysis after all of the antibody work was concluded. 600 µl of the chromatin solution was then incubated with the appropriate antibody overnight at 4°C with rotation. Antibodies used were Anti-acetyl H3 (Millipore) and Anti-H3K9me2 (Abcam) at 5 µl in 600 µl and Anti-FLAG (Sigma F3165-2MG) which was used at 1 µl per 600 µl. A no antibody control was also included for each sample which underwent all of the subsequent steps apart from the addition of the antibody.

2.7.5 Collection, washes and elution of immune complexes

The antibody-attached chromatin was bound to equilibrated Protein G agarose beads pre-absorbed with sheared salmon sperm DNA (Millipore) by incubating them together for 1 hour at 4°C with rotation. The beads where then washed in low salt wash buffer, high salt buffer and LiCl buffer once each in succession for 5 minutes each at 4°C. The final wash in TE buffer was then repeated twice. Each wash step consisted of 5 minute rotation with the wash solution, followed by 30 second centrifugation and removal of supernatant.

The DNA antibody complex was then eluted from the agarose beads by adding 250 µl of Elution buffer to the pelleted beads for 15 minutes and heating the mixture to 65°C. The beads were pelted and the DNA containing supernatant collected. The elution step was repeated twice.

2.7.6 Reverse crosslinking

To remove the crosslinking, 20 µl of 5 M NaCl was added to the samples which were then incubated overnight at 65°C.

2.7.7 DNA clean up

To remove the proteins from all samples including the input control 10 µl of 0.5 M EDTA, 20 µl of 1 M Tris-HCl (pH6.5) and 1 µl of proteinase K were added to the samples and incubated for 1 hour at 45°C. DNA clean-up was then performed on all samples using the PCR clean-up kit from Qiagen according to manufacturer's instructions.

Analysis of ChIP experiments was performed by qPCR (section 2.2.4).

2.8 Protein detection by western blot

In this project protein detection was conducted mainly by western blot. There were technical issues with western blots that were resolved over the course of the project. Whist that is not unusual for a project of this length it should be noted that results reported here are from two slightly different protocols. It is made clear which protocol has been used for which result in the body of the text and the figure legends as one protocol depends on the use of two antibodies whilst the other only requires one.

2.8.1 Solutions required for western blot

Tris-buffered saline (TBS)

1 litre of TBS consists of 8 g NaCl, 0.2 g KCl and 3 g Tris Base in 800 ml of dH_20. The pH was adjusted to 7.4 with HCl and the volume made up to 1 l using dH_2O before being autoclaved. To produce TBS-Tween (TBS-T) 1 ml of 10% Tween-20 was added per 100 ml of TBS.

5x Tris- glycine electrophoresis buffer – no SDS

1 litre of 5x Tris- glycine electrophoresis buffer – no SDS consisted of 15.1 g Tris base with 94 g glycine made up to 1 l with dH_20. This is diluted to 1x Tris- glycine electrophoresis running buffer before use by the addition of dH_2O.

5x Tris- glycine electrophoresis buffer – with SDS

1 litre of 5x Tris- glycine electrophoresis buffer – with SDS consisted of 15.1 g Tris base, 94 g glycine and 25 ml 20% SDS made up to 1 l with dH_20. This is converted into 1x Tris- glycine electrophoresis running buffer before used by the addition of dH_2O.

Sample buffer

To produce the sample buffer necessary for plant protein extraction 1 ml of Tris (pH 6.8), 2 ml 20% SDS, 2 ml Glycerol and bromophenol blue were added to water in a total volume of 8 ml. 200 µl of 1 M DDT per 1 ml was added immediately prior to use.

6% resolving gel

A 6% resolving gel consisted of 10.6 ml H_20, 4 ml of 30% acrylamide mix, 5 ml of 1.5M Tris (pH 8.8), 0.2 ml 10% SDS, 0.2 ml 10% ammonium persulfate and 16 µl of TEMED. Ammonium persulfate and TEMED are the polymerising agents and therefore were added quickly before pouring the gel into the handcast gel cast system from BioRad. Approximately 1 ml of 0.1% SDS was used to level the gel as it set. Once set the 0.1% SDS was removed so that the stacking gel could be cast on top of the resolving gel.

5% stacking gel

A 5% stacking gel consisted of 3.4 ml H_20, 0.83 ml of 30% acrylamide mix, 0.63 ml of 1.5M Tris (pH 6.8), 50 µl 10% SDS, 50 µl 10% ammonium persulfate and 5 µl of TEMED. Ammonium persulfate and TEMED are the polymerising agents and therefore were added quickly before pouring the gel into the handcast gel cast system from BioRad. A 10 well comb was added into the stacking gel and allowed to set. The comb was removed just prior to loading.

Coomassie staining solution

The coomassie stain consisted of 0.25 g of Coomassie Brilliant Blue, 45 ml dH_2O, 45 ml Methanol and 10 ml of Glacial acetic acid

Coomassie de-stain

The coomassie de-stain consisted 45 ml dH_2O, 45 ml Methanol and 10 ml of Glacial acetic acid

2.8.2 Protein extraction and separation

300 µl of the sample buffer with DDT was added to plant tissue samples which had been powdered in the presence on liquid nitrogen. For the *Xenopus laevis* controls the positive control was 5 *Xenopus laevis* embryos injected with 1 ng of Ror2:FLAG mRNA at the 2 cell stage and the negative control was 5 un-injected embryos. Protein was extracted from embryos using RIPA buffer. The protein samples were heated to 85°C for 10 minutes and then pluse spun in a centruged. 20 µl of the prepared protein samples were then loaded into a SDS-PAGE gel made from 5% stacking and 6% resolving sections. The gel was run in 1 X Tris- glycine electrophoresis buffer (with SDS) for ~40 minutes at 150 V. To check the quality of the protein samples, in some cases two identical gels were run concordantly. One of these gels was then subjected to coomassie staining. To stain a polyacrylamide gel it was placed in the coomassie stain solution with shaking for 2 hours and left in de-stain overnight. Further short washes in de-stain were then conducted as required.

2.8.3 Transfer

100% methanol was used to reactivate immune-blot PVDF membrane (BioRad). After being run the gel was equilibrated for 5 minutes in 1 X Tris- glycine electrophoresis buffer (without SDS) the proteins were transferred to the PVDF membrane in cooled 1 X Tris- glycine electrophoresis buffer (without SDS) for 1.5 hours at 550 mA in a Mini Trans-blot electrophoretic transfer cell (Biorad).

The less efficient transfer method that was originally used with the two antibody protocol was the same except that 1 X Tris- glycine electrophoresis buffer (with SDS) was used in the transfer cell and the gel was not equilibrated prior to transfer.

2.8.4 Blocking

The PVDF membrane was blocked in 5% milk (Marvel) in TBS-T for 1 hour at room temperature with shaking.

2.8.5 Primary antibody incubation

The membranes were incubated overnight in 5% milk (Marvel) in TBS-T in the presence of the primary antibody at a 1/1000 dilution 4°C. The primary antibodies used were anti-FLAG-HRP (Sigma A8592) for the one antibody protocol, and anti-FLAG (Sigma F 7425) for the two antibody protocol.

After primary antibody incubation the blot was then washed 3 times in excess TBS-T, 10 minutes per wash. HRP-conjugated antibodies were then imaged using a chemiluminescent detection system (2.8.7), whereas non- conjugated antibodies were followed with a secondary antibody.

2.8.6 Secondary antibody incubation

The membranes were incubated for 1 hour in 5% milk (Marvel) in TBS-T in the presence of the secondary antibody at 1 in 1000 dilution at room temperature. Secondary HRP-conjugated goat-anti-Rabbit was a polyclonal from Abcam was used.

2.8.7 Western blot detection using chemiluminescence

To image the blot it was covered in 1 ml of Enhanced chemiluminescence (ECL) mixture, which was made according to manufacturer's instructions, for 1 minute. Blots were exposed to the film for 30 seconds to 20 minutes depending on the strength of the signal. The films were developed in a compact X4 Xograph (Xograph Imaging Systems).

2.9 Imaging GFP

2.9.1 Hand-held UV lamp

The hand-held UV lamp used here to screen the 35S:GFP lines was the B100-AP Blak-Ray model from UVP (www.uvp.com).

2.9.2 Fluorescence microscopy

To image the GFP fluorescence in leaf tissue of Arabidopsis, a Leica MZFLIII UV fluorescence microscope was used with CoolSNAP camera and imaging software (RS Photometrics).

2.9.3 Confocal microscopy

To image the GFP fluorescence in leaf tissue of transiently transformed *N.benthamiana* and transformed Arabidopsis, confocal microscopy was used to take single plane images. A Zeiss LSM 510 Meta confocal microscope on an Axiovert 200M base (Zeiss) was used. A 488 nm laser was used to excite GFP and a 543 nm laser was used to excite chlorophyll and ZEN 2008 software was used. Images were processed using ImageJ (RSB).

Chapter 3 Genomic location does not affect amplicon-triggered RNA silencing or RdDM of a transgene

3.1 Introduction

Previous work has suggested that the ability of a transgene to undergo silencing may be influenced by genome location (Jones et al., 1998). For example, it has been observed that many transgenes can readily undergo RDR6-dependent PTGS but that this does not always occur; a difference that could be explained by locus-specific effects. It raised the possibility that RDR6 may be interacting at the DNA level with the template producing-loci. Consistent with this idea it had been shown that RDR6-GFP fusion proteins localize in the nucleus which would mean RDR6 could have access to DNA and DNA-associated proteins (Luo & Chen, 2007, Kumakura et al., 2009).

Furthermore, it has been observed that the DNA of genes that are undergoing RDR6-dependent PTGS often becomes methylated via RdDM (Jones et al., 1999). At the start of this project it was not clear where the cycle of DNA methylation caused by RNA silencing started. For example, there was evidence that being transcribed from methylated DNA may have marked the transcripts as being templates for RDR6 (Eamens et al., 2008). Additionally, whilst the involvement of RDR6 marks the pathway as non-canonical, it is not clear to what extent proteins involved with RdDM are also involved in RDR6-dependent DNA methylation. If genome location affects the ability of silenced genes to undergo RDR6-dependent maintenance of RNA silencing then genome location may also affect which of these genes becomes methylated. For example, recent work on transposable elements (TE) by Zemach *et al*. (2013) has shown that there may be a location-dependent effect with regard to methylation of TEs. In euchromatin TE methylation is predominantly conducted by RdDM whereas in heterochromatic regions TE methylation requires DECREASE IN DNA METHYLATION 1 (DDM1) and is independent of RdDM (Zemach et al., 2013). Therefore silenced genes closer to TEs methylated by RdDM may also be more likely to undergo DNA methylation due to the proximity of RdDM proteins already recruited to that area of the genome. This may affect the silencing of transgenes integrated into such regions more than endogenous genes as these tend to be resistant to the action of RDR6 and RdDM (discussed by Fischer et al., 2008).

In order to test the hypothesis that locus-specific effects are involved in supporting silencing events, multiple transgenic lines were generated that carry single-locus transgenes at different genomic locations. The GFP gene was used as it allowed for straightforward assessment of

silencing without any adverse phenotypic consequences for the plant. The first part of this chapter describes the creation of these homozygous single locus 35S:GFP transgenic lines.

Once created, the 35S:GFP lines were then crossed with a trigger of RNA silencing that targeted the GFP coding region. The lines were then screened for differences in silencing. The hope was to find a difference in either the presence of silencing or in the type of silencing, i.e either RdDM or PTGS. A schematic of the work conducted is shown in Figure 3.1.

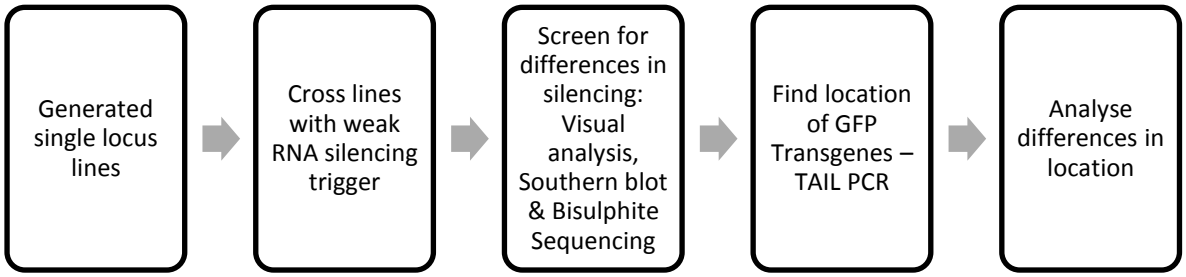


Figure 3.1. Schematic of the work discussed in this chapter

To initiate RNA silencing I used the trigger line AMP243 from the Dalmay et al. (2000a) paper. The AMP243 line contains an amplicon from potato virus X (PVX) which contains *GFP* cDNA. The system is under the control of the 35S promoter from cauliflower mosaic virus and its transcription is ended by the terminator sequence from the nopaline synthase gene. The viral section of the *35S-PVX:GFP* transgene in shown in Figure 3.2. The AMP243 line was generated by transforming *Arabidopsis thaliana* ecotype C24 plants using Agrobacterium with T-DNA containing the *35S-PVX:GFP* construct. Dalmay et al. (2000a) have shown that when crossed with a line containing a *GFP* transgene, AMP243 triggers RDR6-dependent post-transcriptional silencing of GFP (Dalmay et al., 2000b). They also demonstrated this was associated with DNA methylation of the *GFP* gene body. The AMP243 line produces low levels of viral GFP RNA which triggers PTGS towards *GFP* but does not display GFP fluorescence as the amplicon triggers self-silencing. When crossed with a repressor of RNA silencing AMP243 has been shown to display GFP fluorescence due to the release of GFP from self-silencing (Lewsey et al., 2007). AMP243 is considered to be a weak trigger of silencing in Arabidopsis due to non-host effects and produces only low levels of siRNA (Dalmay et al., 2000a).

Figure 3.2. Diagram of the viral section of the 35S-*PVX:GFP* transgene in trigger line AMP243.

The genes depicted in blue are all from PVX. RdRP is the RNA-dependent RNA polymerase. 25, 8 and 12 are the "triple block genes" which encode proteins of a size equivalent to their number. CP is the coat protein. Adapted from Dalmay et al. (2000a). Not to scale.

Silencing triggered by AMP243 is seen throughout the plant due to the constitutive activity of the 35S promoter, which means that any spreading of the silencing signal is masked. Therefore another line was also used in an attempt to trigger silencing that would allow examination of potential locus-specific differences in systemic movement of a silencing signal. Acquired from Oliver Voinnet, the line known as SUC-GF:IR expresses an inverted repeat of the first part of the *GFP* transgene driven by a phloem-specific promoter. This line is able to initiate PTGS of a *GFP* transgene in the phloem only companion cells which results in RDR6-dependent spreading of the silencing signal and reduction of GFP fluorescence away from the vascular tissue.

For the effects of genomic location to be assessed, the integration site of the *GFP* transgenes were identified through the use of thermal asymmetric interlaced PCR (TAIL PCR).

3.2 Results

3.2.1 Creating homozygous single locus GFP transgenic lines

In order to create homozygous single locus GFP transgenic lines approximately 650 Col-0 *A. thaliana* plants were transformed in two batches using the floral dip technique (Bent & Clough, 1998, Bent et al., 2000). A GFP construct under the expression of the 35S promoter from the cauliflower mosaic virus was used with a Kanamycin resistance marker gene. The T1 plants from a subset of these transformations were screened for Kanamycin resistance on MS plates and then assessed for GFP expression under a hand-held UV lamp. Of the ~600 T1 plants that were resistant to Kanamycin, only around 200 plants expressed GFP. The other plants either did not contain the *GFP* transgene or did carry it but were not able to express the protein to a detectable level. GFP expression was assessed visually to increase the speed of the screening process. This may have meant that lines that produced very low levels of GFP were excluded. Lines that did not express the protein at all may have been undergoing gene silencing events that often occur in the creation of transgenic lines. Plants that underwent spontaneous silencing were discounted as they were likely to have multiple insertions in complex configurations; on average this form of

Transfer DNA (T-DNA) transformation is reported to give ~1.5 integration sites per line (Alonso et al., 2003, McElver et al., 2001). Although these plants may have provided information regarding transgene silencing, it would have been labour intensive to identify the single locus GFP transgenic plants that were undergoing spontaneous silencing from the background of multiple loci, no-GFP protein or low expressing GFP plants. Furthermore, these lines could not be used with a trigger of silencing to look at loci before and after silencing. If these plants had been identified then the genomic location of their *GFP* transgene could have been identified to investigate any effects of genomic location of spontaneous silencing.

Between 50 and 100 plants of 130 T1 lines were assessed for a 3:1 ratio of GFP expression: no-expression as this is indicative of transgene integration at a single locus. Testing of 130 T1 lines resulted in sufficient material for further work and so the other 70 lines were not screened. Of the 130 lines tested, 74 displayed the 3:1 ratio of GFP expression indicative of single locus insertions. The ratio was analysed using the χ^2 test to indicate which lines are likely to have a true 3:1 ratio and therefore a possible single integration site. However, it should be noted that using a 3:1 ratio in this manner is not infallible and I am assuming that lines that meet the criteria of the χ^2 analysis have a single site of integration when this may not be the case. For example, this method of analysis would not be able to filter out lines with linked insertions. The raw data and the χ^2 analysis are displayed in a table as Appendix 1. A p-value cut off of 0.05 was used. The table shows data for only 99 lines as ones where there was 100% GFP expression were not included. The original numbers for the GFP lines are maintained throughout the thesis resulting in discontinuous numbers due to the removal of lines without the 3:1 ratio. For example line 4 was removed and not used again during this investigation but the remaining lines were not “shuffled up” into “gap” number 4.

Seeds from GFP positive T2 plants for each of the 74 lines were collected and sown out on soil so that the homozygous lines could be identified in the T3 generation. This produced 63 independent lines that were homozygous for the *35S:GFP* transgene. There were 5 homozygous 35S:GFP lines previously generated in the lab available to me (lines A-E). Therefore, I had 68 independent 35S:GFP lines to move forward with. An overview of the numbers of plants and plant lines screened to generate my final lines is shown in Figure 3.3.

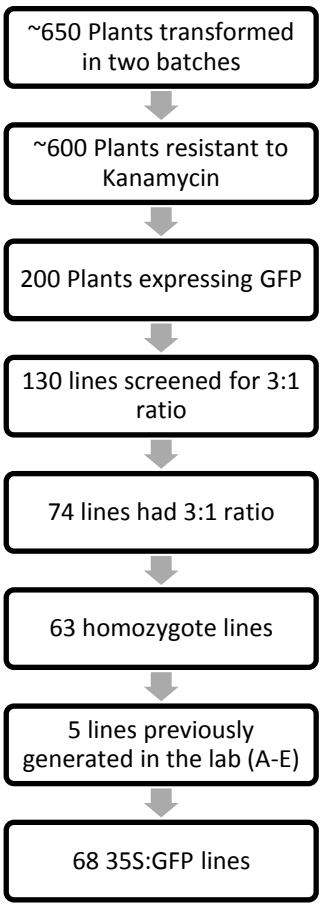


Figure 3.3. Overview of homozygous single locus GFP transgenic line production

3.2.2 Silencing was triggered by AMP243 in all 35S:GFP lines tested

52 of the 68 independent 35S:GFP lines were crossed successfully with trigger line AMP243, which was described in the chapter introduction. All 68 lines were set up for crossing but for 16 lines either the plants that were to be used as parents didn't produce pollen of sufficient quality for crossing or viable seeds were not produced by the cross. Although some of these lines would have been able to produce F1 seeds if the crosses had been repeated, the numbers and intensity of labour involved prohibited many multiple rounds of repeats. Initially reciprocal AMP243 crosses were performed, however as demonstrated later in this chapter there was no difference in silencing between the ♂ AMP243 X ♀GFP and the ♂ GFP X♀AMP243 crosses, therefore only the ♂ AMP243 X ♀ GFP cross were performed for all other crosses. This permutation of crossing was the most efficient as it ensured that all lines would definitely inherit the GFP gene from the parent. If silencing occurred then it could be assumed that the trigger had been successfully crossed into the line. If a line expressed GFP and did not undergo silencing then the presence on the AMP243 trigger would need to be confirmed by PCR to validate the result.

3.2.2.1 Visual Analysis of GFP fluorescence

The F1 progeny from all 52 crosses were analysed visually and all were observed to be silenced for GFP. The analysis was conducted on soil grown plants examined under a UV microscope. An example of a typical visual analysis is shown in Figure 3.4. GFP fluorescence was absent from the AMP243 X 35S:GFP12 cross but present in the GFP12 parent. This result was seen for all AMP243 crosses. A selection of lines were also crossed with a non-transgenic line as a negative control. The AMP243 line has a C24 background whereas the 35S:GFP lines have the Col-0 background. Therefore the C24 ecotype of Arabidopsis was used as the negative control to keep all genetic parameters the same. The visual analysis showed that the progeny of the C24 crosses still produced GFP whereas the AMP243 progeny did not (Figure 3.4). Therefore silencing occurred in the presence of the trigger and was not the result of crossing.

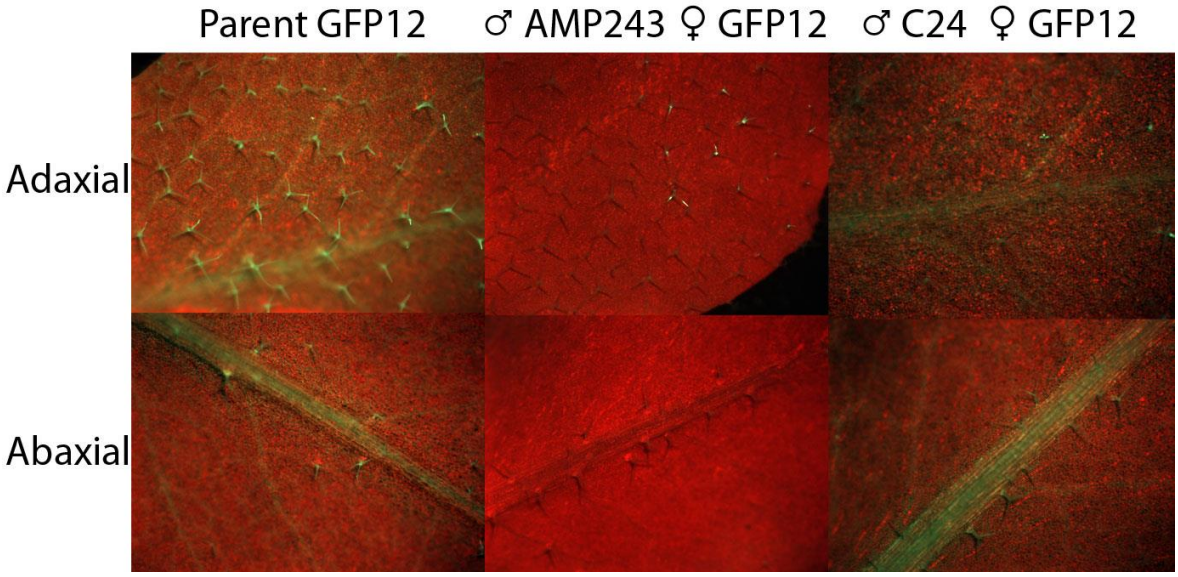


Figure 3.4. Typical result of visual analysis observed for all F1 AMP243 X 35S:GFP Crosses

Images of whole leaf samples visualized under UV microscope (X 2.5). Line shown is GFP12 with an uncrossed parent plant on the left, the F1 AMP243 cross in the middle and the control F1 C24 cross on the right. The top panels show the adaxial leaf surface and the bottom panels showing the abaxial surface. All plants were soil grown and pictured 24 days post germination. GFP expression is observable in the parent plants. This appears green against the background of chlorophyll, which appears red under UV light. No GFP expression is observed in the vascular tissue of the F1 crosses with AMP243, although the tips of a few trichomes appeared to be green. GFP fluorescence was observed in the crosses with C24, but at lower level to the parent plants. The trichomes were still green in these plants, but at a much lower level than in the parents.

3.2.2.2 GFP mRNA analysis by northern blot

In order to confirm that the visual loss of green fluorescence was due to loss of GFP mRNA, northern blot experiments were performed. Total RNA samples were prepared from pooled leaf tissue from F1 AMP243 crosses and from the corresponding “parent” 35S:GFP line (see method

section 2.1.4). The 35S:GFP lines are referred to as parent lines throughout this work. RNA was also extracted from the AMP243 line as a control. The northern blots were performed as described in methods section 2.5 using a GFP specific probe.

GFP mRNA levels were compared between the parent and AMP243 cross for 16 lines. This was deemed a sufficient number to confirm that examining visual fluorescence was a good indication of whether GFP mRNA was present or not. GFP mRNA is present in the parent 35S:GFP lines but absent in the AMP243 X 35S:GFP crosses (Figure 3.5). Therefore loss of visual fluorescence is associated with loss of GFP mRNA which is indicative of gene silencing although it is not possible to determine whether a transcriptional or post-transcriptional mechanism is responsible. Parent AMP243 did not produce GFP mRNA but it does produce a small amount of viral RNA carrying the GFP sequence which gives some banding at a higher molecular weight in the AMP243 sample on Figure 3.5A. This is consistent with observations from other groups (Dalmay et al., 2000a).

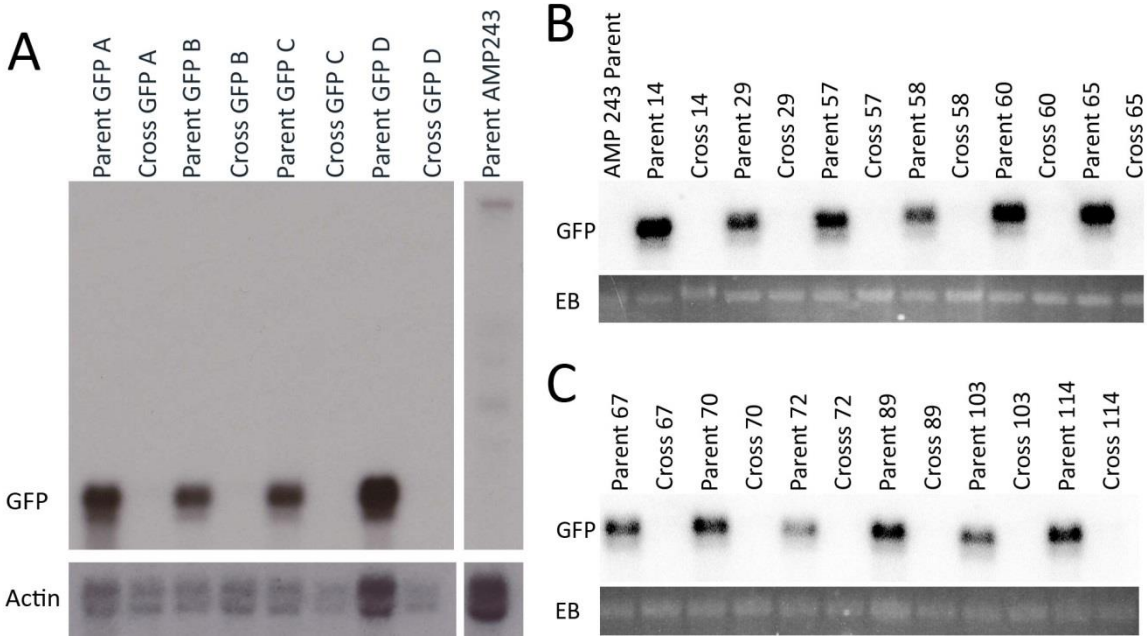


Figure 3.5. AMP243-triggered loss of GFP fluorescence is associated with loss of *GFP* mRNA

Parent refer to lines not crossed with AMP243 and Cross refer to F1 plants which were the result of crossing AMP243 with a 35S:GFP line. **Section A** shows the results of the 5 lines already available in the lab. The top panel shows the blot probed for *GFP* and the lower panel shows the blot probed for *Actin* as a loading control. The parent and cross for GFP E were also included on this gel; however, the RNA degraded and so the lanes were removed in Photoshop with a white gap left to indicate this. In **Sections B** and **C** the top panel shows the blot probed for *GFP* and the lower panels show ribosomal RNA stained with ethidium bromide prior to blotting as the loading control. These blots were imaged using the phosphorimager and therefore have a lighter background than Figure A which was imaged using X-ray film.

3.2.2.3 Small interfering RNA analysis

In order to investigate whether the AMP243 silencing was associated with the production of siRNAs, a variation of the northern blot technique which has been optimised for the lower molecular weight of siRNAs was performed as described in method section 2.5. Total RNA samples were prepared from pooled leaf tissue from F1 crosses and from the corresponding "parent" 35S:GFP line. The blot was probed using a P^{32} labelled GFP probe.

For this analysis the DAN2 line was used as a positive control. This line is the result of a cross between AMP243 and a 35S:GFP line known as GFP142 (Dalmay et al., 2000a). The DAN2 line had previously been shown to produce high levels of GFP siRNA and to support RdDM (Dalmay et al., 2000a).

GFP siRNAs accumulated in all of the AMP243 X 35S:GFP crosses analysed and was absent from the parent lines (Figure 3.6). This strongly indicates that siRNA-dependent post-transcriptional silencing is occurring in these lines. Only a selection of lines were analysed in this manner to connect the loss of visual fluorescence and GFP mRNA with the accumulation of GFP siRNAs. However, it should be noted that while this is highly suggestive of the occurrence of PTGS it does not wholly confirm this. Additionally, transcription silencing of the 35S promoter could be occurring in some lines which would lead to the reduction of GFP mRNA.

Although a siRNA ladder was included on the gel, the blots were not probed with a ladder specific probe. Therefore the size of the siRNAs present in Figure 3.6 cannot be confidently stated. However, it appears that only one species of siRNA are observed in Figure 3.6.

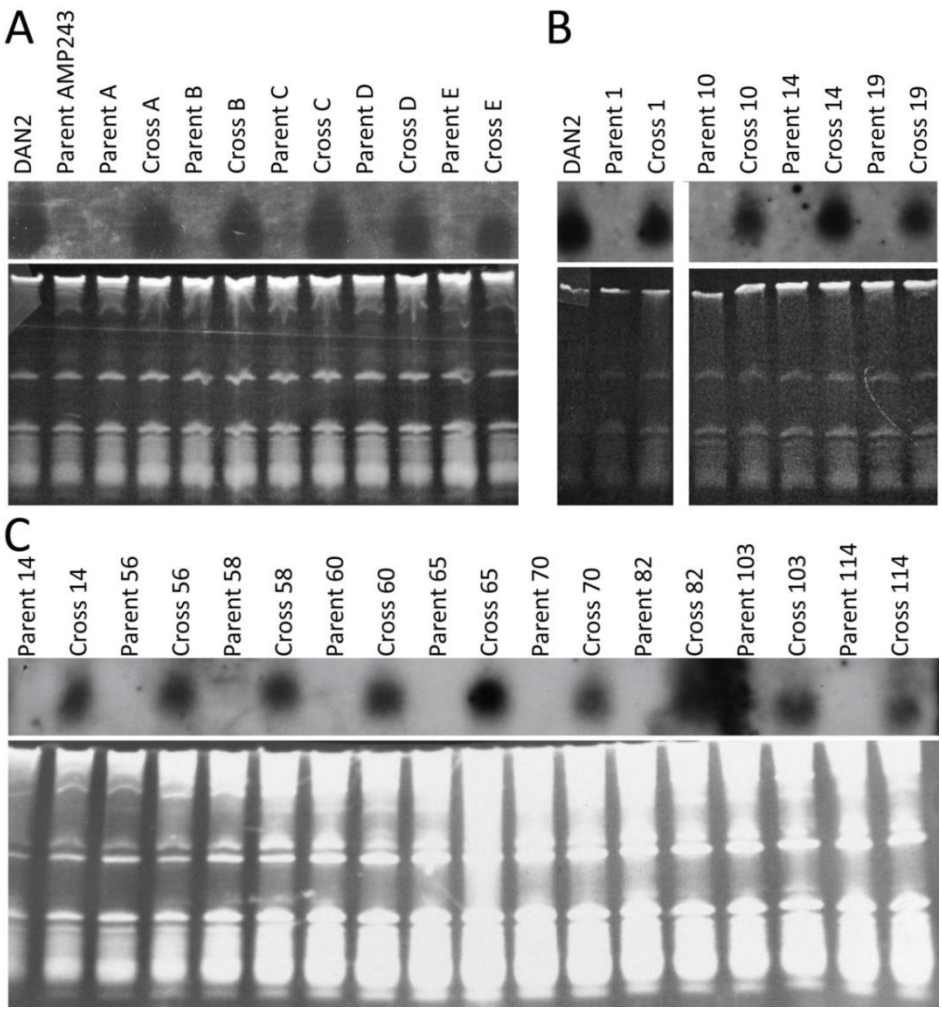


Figure 3.6. siRNAs accumulate in the AMP243 X 35S:GFP cross but not in the parent lines.

The top panel in each section is the siRNA blot and the lower panel is the gel stained with ethidium bromide prior to blotting as the loading control. **Section A** shows siRNA analysis of the five 35S:GFP lines available in the lab previously. **Sections B** and **C** show siRNA analysis for 12 other lines. Unfortunately there was a mark on the film in wells for samples cross 82 and parent 103 in section C that interferes with the results.

3.2.2.4 Analysis of the methylation status of AMP243 X 35S:GFP crosses using Southern blotting

It was originally thought that some lines would not undergo AMP243-triggered silencing and that differences between lines that underwent silencing and ones that did not could be investigated. This, however, was not the case as all 52 AMP243 X 35S:GFP crosses lost visual fluorescence. The results of the northern blot and siRNA analysis suggest that despite being reported as a weak silencing trigger, AMP243 was able to induce PTGS in all of the GFP lines. It was, however, unknown whether silencing was associated with RdDM in all lines. Therefore the next stage was to investigate whether or not lines could support RdDM and to explore why these differences may be occurring. My original hypothesis was that the genomic location of the *GFP* transgene may influence a transgenes susceptibility to undergo RdDM or RDR6-dependent PTGS. However, as all

lines are silencing in the presence of the AMP243 trigger, it is unlikely that genomic location affects the ability of a transgene to undergo RDR6-dependent PTGS.

To investigate whether the AMP243-triggered silencing is associated with DNA methylation in all lines, Southern blot analysis was used to assess the methylation status of the *GFP* transgene in both the parent and the AMP243 crosses. Southern blot was used as it provided an effective way of screening large numbers of lines compared to bisulfite sequencing. For the Southern blot, genomic DNA from each line was digested using either AluI or HaeIII, which are methylation-sensitive enzymes with sites within the GFP sequence. Figure 3.7 shows the restriction map of the *35S:GFP* transgene for AluI and HaeIII. These enzymes do not cut their restriction sites when cytosine is methylated. Therefore a methylated sample has higher molecular weight bands present compared to a non-methylated sample.

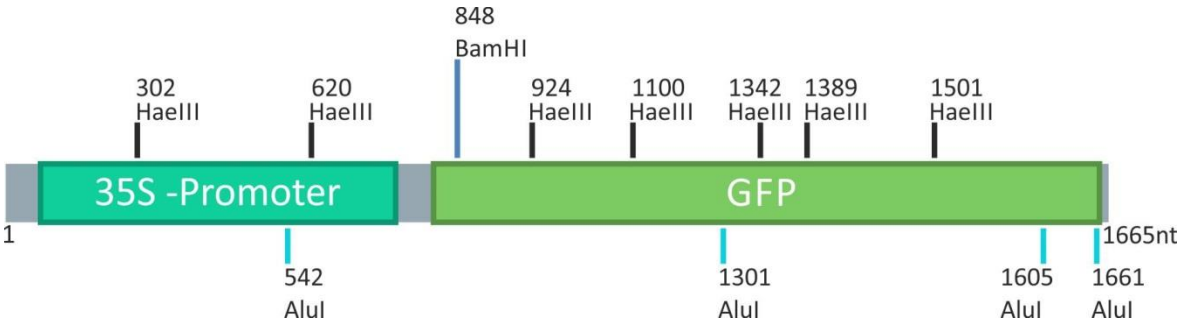
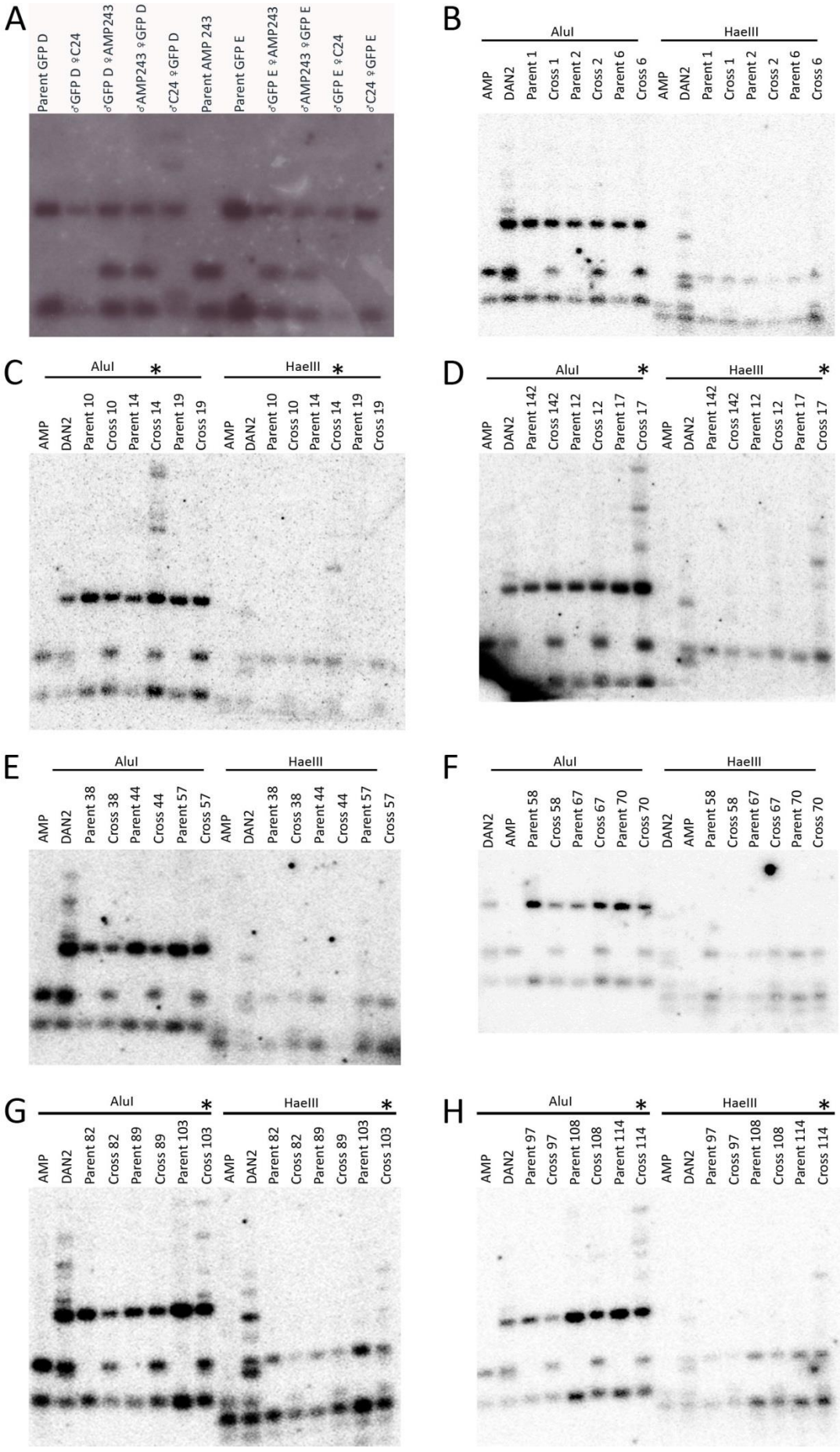


Figure 3.7. Restriction map of the *35S:GFP* transgene for AluI and HaeIII.

The AluI restriction sites are represented by blue lines at the bottom of the diagram and HaeIII ones by black bands at the top edge. There are five HaeIII sites within the GFP gene body and three AluI sites. Both restriction enzymes are sensitive to DNA methylation.

DNA samples from 40 of the AMP243 X 35S:GFP F1 crosses were analysed by Southern blot along with the parent lines, the results of which are shown in Figure 3.8. Out of the 40 lines that were analysed only 8 showed signs of methylation as indicated by the presence of higher molecular weight bands on the Southern blot (Figure 3.8). These are lines 14, 29, 50, 56, 65, 71, 103 and 114 (Figure 3.8). The line 17 cross also showed methylation (Figure 3.8D), however when sown out again the parent plants displayed either full silencing or a mosaic-like silencing phenotype and so line 17 was removed from further analysis.

The initial crosses performed included the reciprocal crosses for the AMP243 cross. However, both Southern blot and visual analysis demonstrated that there was no difference in silencing between the ♂ AMP243 X ♀GFP and the ♂ GFP X♀AMP243 crosses, therefore only the♂ AMP243 X ♀ GFP cross were performed for all other crosses (Figure 3.8D).


A
Parent GFP D
♂GFP D ♀C24
♂GFP D ♀AMP243
♂AMP243 ♀GFP D
♂C24 ♀GFP D
Parent AMP 243
Parent GFP E
♂GFP E ♀AMP243
♂AMP243 ♀GFP E
♂GFP E ♀C24
♂C24 ♀GFP E
B
AluI
HaeIII
AMP
DAN2
Parent 1
Cross 1
Parent 2
Cross 2
Parent 6
Cross 6
C
AluI
HaeIII
AMP
DAN2
Parent 10
Cross 10
Parent 14
Cross 14
Parent 19
Cross 19
D
AluI
HaeIII
AMP
DAN2
Parent 142
Cross 142
Parent 12
Cross 12
Parent 17
Cross 17
E
AluI
HaeIII
AMP
DAN2
Parent 38
Cross 38
Parent 44
Cross 44
Parent 57
Cross 57
F
AluI
HaeIII
DAN2
AMP
Parent 58
Cross 58
Parent 67
Cross 67
Parent 70
Cross 70
G
AluI
HaeIII
AMP
DAN2
Parent 82
Cross 82
Parent 89
Cross 89
Parent 103
Cross 103
H
AluI
HaeIII
AMP
DAN2
Parent 97
Cross 97
Parent 108
Cross 108
Parent 114
Cross 114

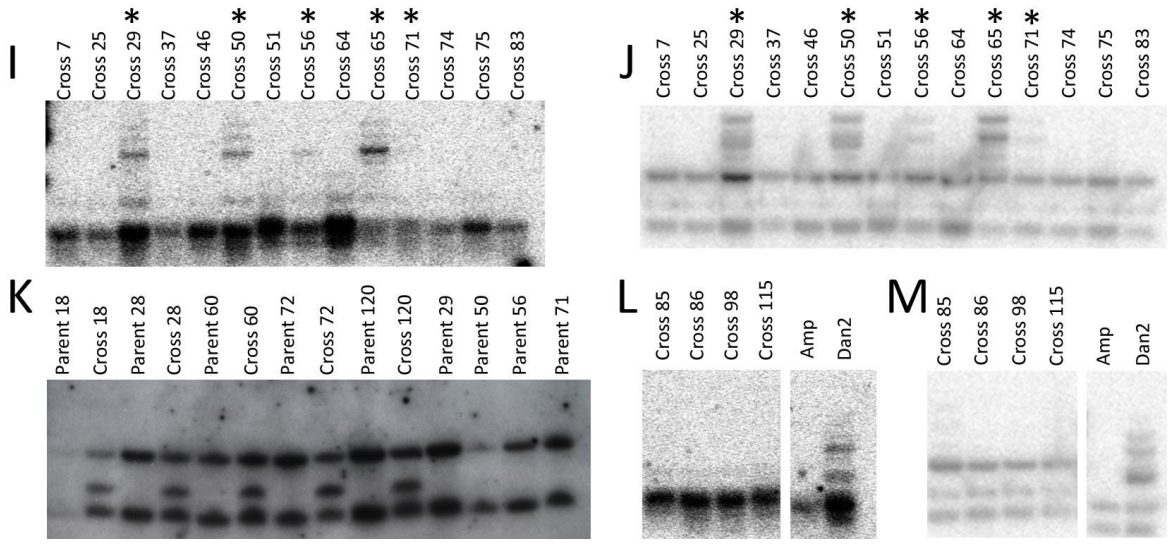


Figure 3.8. Analysis of methylation status of AMP243 X 35S:GFP crosses by Southern blot.

Samples have been digested with the methylation sensitive enzymes AluI or HaeIII. All blots have been probed with a GFP-specific probe. Samples where methylation is present are indicated with an asterix. **Panel A** F1 samples of Lines D and E reciprocal crosses with either AMP243 and the negative control C24. All samples have been digested with AluI. **Panels B to H** Samples from both parents and ♂ AMP243 X ♀ GFP F1 crosses digested with either AluI or HaeIII as indicated. **Panel I** ♂ AMP243 X ♀ GFP F1 crosses only. Digested with HaeIII. **Panel J** ♂ AMP243 X ♀ GFP F1 crosses only. Digested with AluI. **Panel K** ♂ AMP243 X ♀ GFP F1 Parents and crosses and Parent lines of F1 methylated crosses in Panels I and J. Digested with AluI. **Panel L** F1 ♂ AMP243 X ♀ GFP crosses. All digested with HaeIII. **Panel M** As for panel K but digested with AluI.

3.2.2.5 Confirmation of the methylation status of the GFP transgenes using bisulfite sequencing

The Southern blot analysis can only guarantee that there is methylation at the restriction sites analysed. Additionally, the presence of higher molecular weight bands could be the result of incomplete digestion by the enzyme and not a sign of methylation. Although it should be noted that the digestion of DNA for each blot was conducted concurrently with the same batch of reagents under the same conditions, so the fact that higher molecular weight bands were not seen in the parent lines which are not expected to undergo DNA methylation suggests that digestion was complete. Furthermore, lines that appear unmethylated in the Southern blot data may be methylated at the restriction sites tested but at levels too low to detect. The Southern blot method requires high amounts of DNA and may not detect low level methylation.

In order to validate the Southern blot data, bisulfite sequencing was performed. Bisulfite sequencing works by converting un-methylated cytosine residues to uracil whilst leaving methylated cytosines unconverted, thus allowing detection of methylation at a single-base resolution. Three lines that showed methylation in the F1 generation (14, 103 & 114) and four others that did not (10, 23, 57 & 70) were investigated.

The bisulfite sequencing protocol is described in method section 2.6.2. Briefly, genomic DNA from pooled leaf tissue for each line was digested with HindIII and then treated with sodium bisulfite using the EZ DNA Methylation- gold kit (Zymo Research). Hot start PCR was then performed on these samples to amplify a portion of the *GFP* transgene which could then be sequenced. Conventionally the PCR fragments would have been cloned into a plasmid vector and used to transform *E. coli* so that individual DNA strands could be sequenced. However, direct sequencing of the PCR fragments was sufficient to detect DNA methylation as shown in Figure 3.9. This resulted in a mixture of DNA strands being sequenced for each sample. In an unmethylated sample, all of the cytosine residues are converted to uracil; however, as methylation is never occurring at the 100% level, PCR products from a methylated sample will represent a mixture of converted and unconverted cytosines at a single locus. This is due to the fact that a mixture of different DNA strands have been sequenced.

A section of the chloroplast gene, PsaA, was used as a control as the chloroplast genome lacks DNA methylation and full cytosine conversion is expected. The control reaction was performed on 11 samples to confirm that the sodium bisulfite treatment was robust. 100% conversion was seen in all 11 samples (data summarized in Table 3.1).

To analyse the sequencing data, Sequence Scanner Software 2 from Applied Biosystems was used. The sequencing data returned from the Sanger sequencing was from the reverse strand so that the cytosine conversions are viewed as A to G conversions instead of C to T conversions. An example of the bisulfite sequencing data is shown in Figure 3.9 for lines 10 and 114. The Southern blot showed that the F1 AMP243 X 35S:GFP cross was methylated in Line 114 but not in Line 10. The image shows that in Line 10 there was full conversion of the cytosine residues in both the parent line and the cross by the absence of guanine residues in these sequences (Figure 3.9). Therefore this confirmed the results of the Southern blot. For Line 114 there was full conversion for the parent sample but guanine residues were present in the AMP243 X 35S:GFP cross indicating methylation was present all along the *GFP* transgene in this sample. Examples of the conversion sites are highlighted in Figure 3.9 with arrows.

For the remaining lines full conversion was seen in all parent lines and for the crosses that did not show methylation by Southern Blot in the F1, confirming that the GFP gene is unmethylated (data summarized inTable 3.1). However, the crosses 14, 103 and 114 showed mixed readouts consistent with the presence of methylated cytosine residues, confirming the Southern blot results (data summarized in Table 3.1).

There can be confidence in the methylation status for the lines that underwent full bisulfite sequencing. This is not true of the lines that didn't undergo full bisulfite sequencing. It must be

noted that without subjecting all of the lines to bisulfite sequencing of the *GFP* transgene it is not clear if the lines that did not display DNA methylation in the Southern blot are truly unmethylated. For example, the lines may be undergoing RdDM at a low level, they may be methylated at a level too low to detect by this method or they may be methylated in the region between the restriction sites used. However, due to the agreement between the Southern blot and bisulfite data it has been assumed for the remainder of this project lines that do not display methylation by when interrogated by Southern blot have a different methylation status to the lines that do. This is assumed to be a case of no or low methylation vs. high levels of DNA methylation and RdDM, but this is assumption which has limitations as previously outlined.

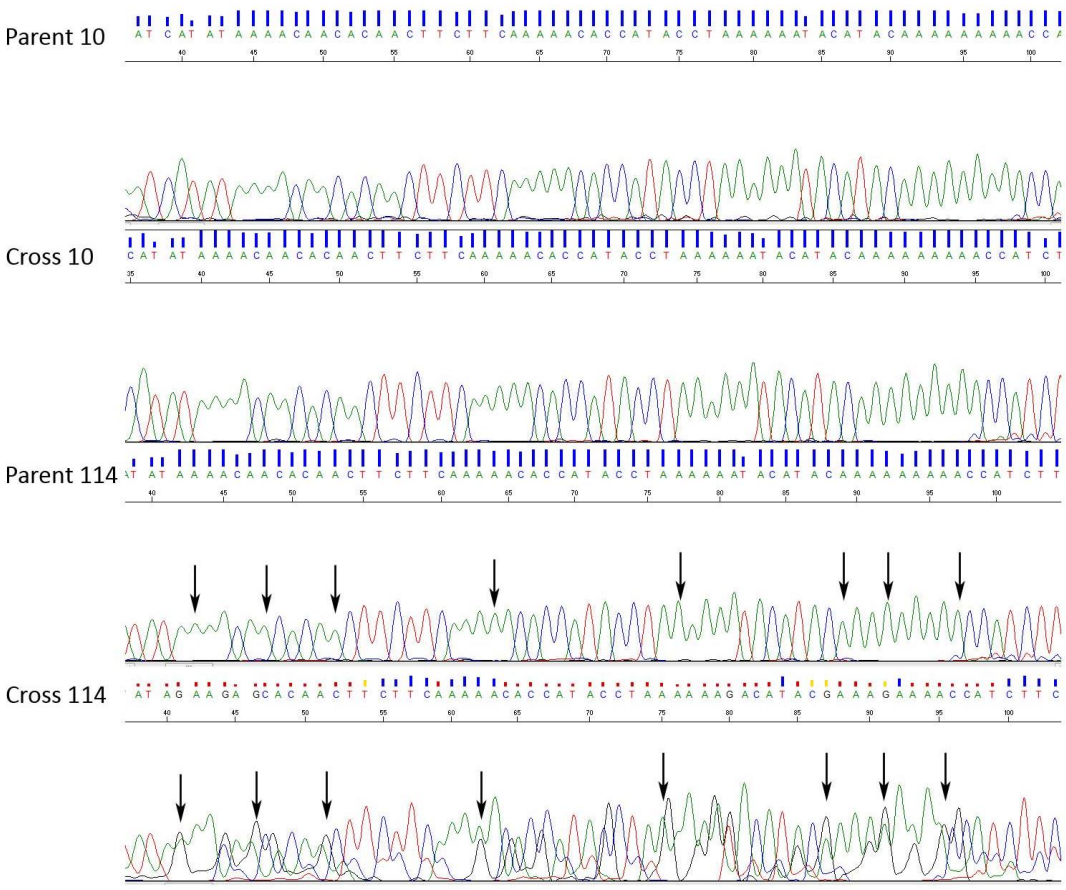


Figure 3.9. Bisulfite sequencing data confirm results from the Southern blots.

Examples of outputs from Sequence Scanner Software 2 from Applied Biosystems. The green line represents Adenine, red Thymine, blue Cytosine and black Guanine. **Top two panels** Line number 10 was not methylated in the F1 cross with AMP243 and full conversion is seen in both the parent and the cross. **Bottom two panels** Line 114 was methylated when silencing was triggered by AMP243 and full conversion is seen in the parent but a mixed peak profile is seen in the cross. Arrows indicate sites of A to G conversion with an unmixed profile seen in the parent and a mixed one in the cross.

Table 3.1 Summary of Bisulfite sequencing results which confirm Southern Blot data

35S:GFP line	Control PsaA region (for both parent and cross)	Parent	F1 Cross	Indicative of methylation in F1 Cross?	Match Southern blot data
10	Full conversion	Full conversion	Full conversion	No	Yes
14	Full conversion	Full conversion	Mixed peak profile	Yes	Yes
23	Full conversion	Full conversion	Full conversion	No	Yes
57	Full conversion	Full conversion	Full conversion	No	Yes
70	N/A	Full conversion	Full conversion	No	Yes
103	Full conversion	Full conversion	Mixed peak profile	Yes	Yes
114	N/A	Full conversion	Mixed peak profile	Yes	Yes
DAN2 Methylated control	Full conversion	N/A	Mixed peak profile	N/A	Yes

3.2.3 Deveoplment of lines homozygous for both the AMP243 trigger and the *35S:GFP* transgene

The DAN2 line could be used as a positive control for the Southern blot experiments in chapter 3 as its *GFP* transgene has previously been reported to be methylated (Dalmay et al., 2000a, Dalmay et al., 2001). However, when the cross that gave rise to DAN2 was recreated by crossing lines GFP142 and AMP243 no methylation of the *GFP* transgene was detected in the F1 generation (Figure 3.8D). This is likely to be the first time that the F1 generation's methylation status had been look at for this cross. When the F2 of the GFP142 and AMP243 cross was included on a Southern blot it displayed higher molecular weight bands indicative of DNA methylation. It suggested that the unmethylated F1 crosses may display methylation in future generations. This is investigated further in chapter 5 and the data for the F2 GFP142 and AMP243 cross is given.

To investigate methylation in subsequent generations, lines that are homozygous for both the PVX:GFP amplicon and the *35S:GFP* transgenes were required. These allowed the effects of zygosity to be investigated. Additionally, the lines were to be crossed with lines containing FLAG-tagged proteins from the RNA silencing pathways to investigate issues of recruitment, which will be discussed in chapter 4.

To generate the homozygous lines, F1 plants from lines 57, 58, 60, 65, 70, 72, 82, 89, 103 & 114 were allowed to self-fertilize and the resulting F2 plants were selected for resistance to Kanamycin and Phosphinothricin (PPT, the active ingredient in BASTA). Kanamycin is used to select for the *35S:GFP* transgene and PPT for the *PVX:GFP* amplicon transgene. Of the GFP lines analysed here 57, 58, 60, 70, 72, 82 & 89 appeared un-methylated in the F1 generation whereas 65, 103 & 114 were methylated. The seeds from individual plants from the F2s were then screened on MS plates for Kanamycin and PPT to find homozygous plants. For lines 57, 70, 72, & 89 no homozygous lines were found in the F3 population. Therefore, individual plants from the F3 were screened on plates as for the F2 population to find homozygous in these lines. The result of this work was the identification of homozygous lines for 58, 60, 65, 70, 72, 82, 89, 103 & 114, however none were found for line 57.

3.2.4 Attempting to trigger silencing using the SUC-GF:IR trigger line

The AMP243 system is only one way of triggering silencing and my initial aim was to investigate if the same pattern of silencing and DNA methylation was observed with another trigger. To do this, the SUC-GF:IR trigger line was obtained from Olivier Voinnet. The SUC-GF:IR trigger itself contains an inverted repeat of the first part of the *GFP* gene under a phloem specific promoter and is known to promote RDR6-dependent silencing. In theory if one of the 35S:GFP lines from this project were crossed with such a trigger then it would be possible to observe the spread of silencing of *GFP* outwards from the phloem where the initiation of silencing originates. This spread would be dependent on amplification of the RNA silencing signal, most likely by the action of RDR6.

The SUC-GF:IR line initially carried a *GFP* transgene as well as the trigger. In order to generate a line that only carries the trigger, the original line was crossed to a non-transgenic Col-0 parent and the F2 plants were assessed for the presence of the trigger and the loss of the GFP target by kanamycin and BASTA selection. Plants that were susceptible to Kanamycin and resistant to BASTA should contain the *SUC-GF:IR* trigger but not the *GFP* transgene. Two suitable lines were identified named SUC-GF:IR 22 and SUC-GF:IR 35. These were not true independent lines as they came from the same source and should be genetically identical; they are biological repeats of the selection process. SUC-GF:IR 22 and SUC-GF:IR 35 were crossed with 11 35S:GFP lines. However, this trigger did not lead to the silencing of GFP even in the phloem, as determined by visual analysis (Figure 3.10). Therefore work with this trigger line was not continued.

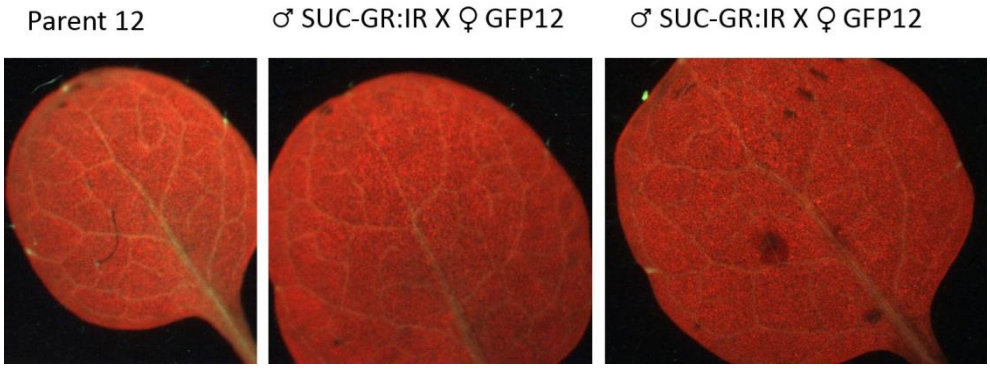


Figure 3.10. Typical result of visual analysis observed for all SUC-GF:IR x 35S:GFP crosses

Images of whole leaf samples visualized under UV microscope. **Left panel** is from Parent 12 which displays observable GFP expression. This appears green against the background of chlorophyll, which appears red under UV light, with green vascular tissue. The middle and right panels both show ♂ SUC-GF:IR X ♀GFP 12 crosses, with a different SUC-GF:IR line was used in each panel. SUC-GF:IR 22 was used for the **middle panel** and SUC-GF:IR 35 was used in the **right panel**. The crosses produced observable GFP with the same pattern as the parent line which implied that there was no silencing of GFP.

3.2.5 Ability to undergo RdDM did not seem to be influenced by genome location

Amplicon-triggered PTGS occurred in all of the independent lines suggesting that this form of silencing is not influenced by either genome location or locus structure. However, it was unclear why only 8 lines produced evidence of undergoing RdDM in the F1 AMP243 cross whereas the majority did not, despite accumulating high levels of GFP siRNA. It may be the case that only 8 lines underwent RdDM and the possible reasons for this will be investigated in this body of work. However, as previously stated, there are limitations to these kinds of assertions.

One of the original aims of the project was to determine whether genome location influences the ability of a transgene to undergo PTGS or RdDM. Therefore, the insertion sites of the *GFP* transgene were determined for a selection of lines. To do this thermal asymmetric interlaced (TAIL) PCR was used on the parent GFP lines to locate the transgene within the genome (Figure 3.11).

Figure 3.11. Primer plan for thermal asymmetric interlaced (TAIL) PCR

Three nested PCRs are carried out with sequence-specific primers located within the transgene close to the T-DNA left border and arbitrary degenerate primers (AD). The cycle details are outlined in the methods section 2.2.5. The resulting fragments are then separated on an agarose gel, extracted and sequenced. Adapted from (Liu et al., 1995).

The aim of TAIL PCR is to sequence out from a known sequence into the unknown part of the genome. Arbitrary degenerate primers (AD) are used to do this along with three nested sequence-specific primers at the end of the known sequence; a primer plan is shown in Figure 3.11. The TAIL PCR process is described in method section 2.2.5. In short, three PCRs are performed on each sample for each AD primer. Four AD primers were used (AD1, AD2, AD3 & AD6) resulting in 12 PCR reactions per sample. For each AD primer an initial PCR using sequence-specific Primer 1 was carried out. The PCR product was then diluted 1 in 40 and used as a template for the Primer 2 & AD reactions and the Primer 3 & AD reactions. The products from these two reactions were then separated on 1.2% agarose gels. An example of such a gel is shown in Figure 3.12 for TAIL PCR on lines 18, 19, 22, 23, 28 & 38. The reactions with sequence specific Primer 2 give a slightly higher molecular weight bands than the reactions with sequence specific Primer 3, as one would expect. The AD reaction with the strongest band was selected and the whole PCR mix was run out on an agarose gel again (Figure 3.12E). The most defined bands were then extracted from the gel and the DNA was Sanger sequenced. Initially samples were sequenced from both the 2nd and the 3rd reactions using the respective sequence-specific primer; however the sequencing data was found to be more reliable from the products from the 3rd reaction with the sequence-specific primer 3.

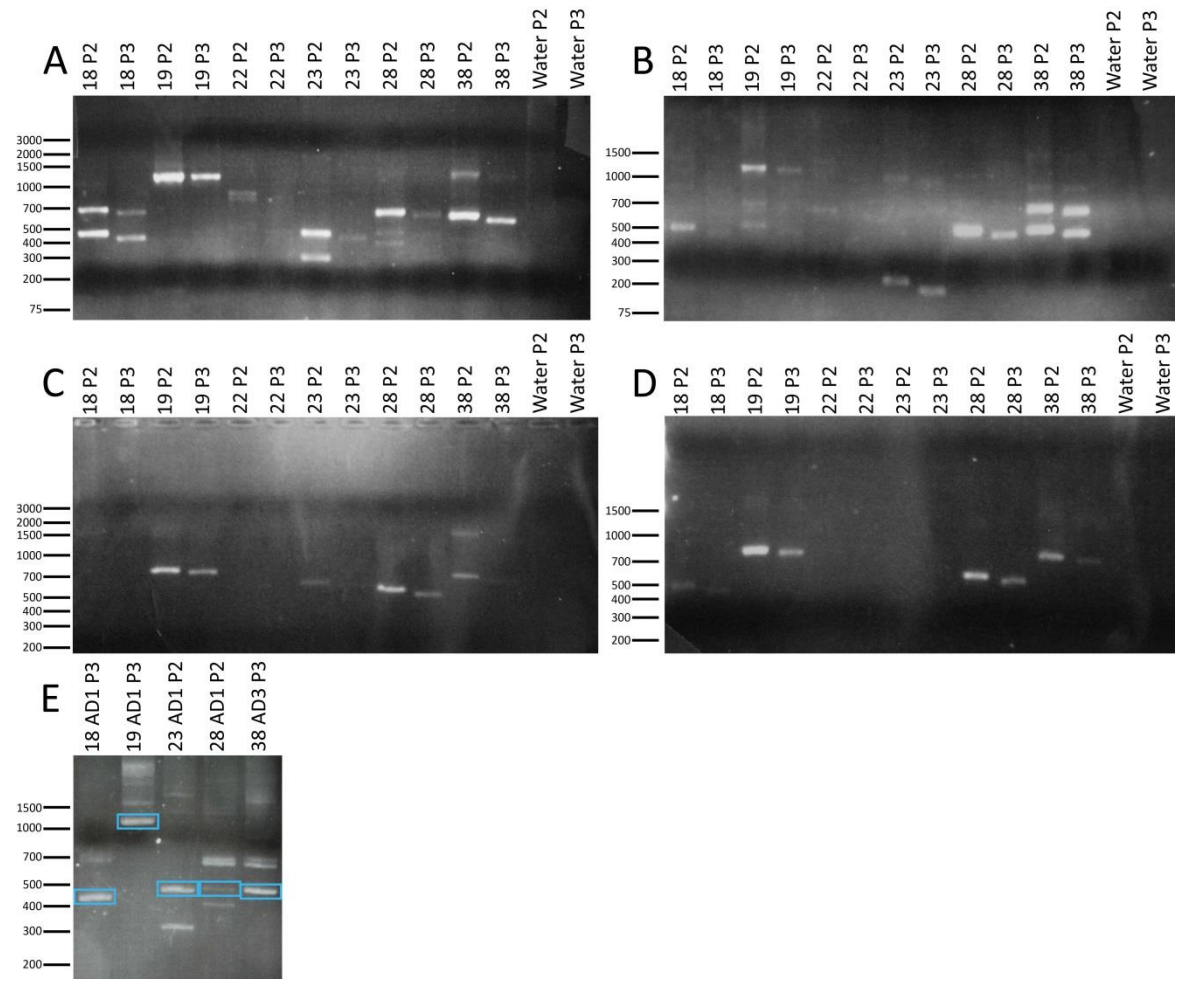
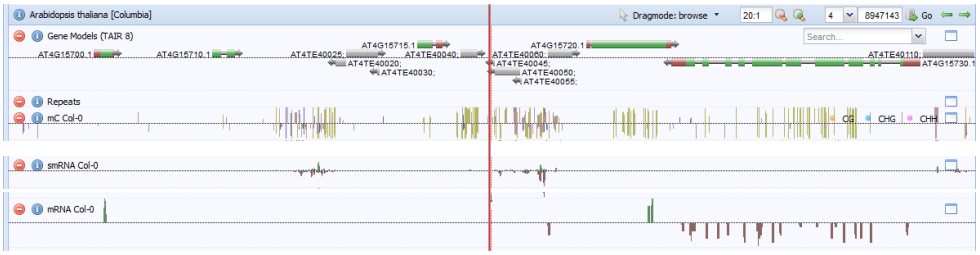


Figure 3.12. 1.2% Agarose gels for TAIL PCR products for lines 18, 19, 22, 23, 28 and 38

Panel A The PCR products from reactions with sequence specific primer 2 (P2) and sequence specific primers 3 (P3) with AD primer 1. **Panel B** Same as for A but for AD primer 2. **Panel C** Same as for A but for AD primer 3. **Panel D** Same as for A but for AD primer 6. **Panel E** Gel with samples for gel extraction. Boxes indicate the bands excised.

Once sequenced the results were run through NCBI's BLAST to locate the transgene insertion site (Altschul et al., 1997). This site was then entered into the Epigenome browser ANNO-J to review the epigenetic landscape in that region (Tonti-Filippini, 2008). An example of the output from ANNO-J is shown in Figure 3.13 for Lines 67 and 103. Gene annotation from The Arabidopsis Information Resource (TAIR) is given in the first track (TAIR8). Subsequent tracks show data on the Repeat Elements, DNA methylation, small RNA production and mRNA production for floral tissue from Col-0 (Figure 3.13).

67



103

Figure 3.13. Output from the Epigenome browser ANNO-J for lines 67 and 103.

The red vertical line in the centre of the image shows T-DNA integration site. This image covers approximately 22 kb. The top track shows which annotated genes are present at the loci. The second track shows any repeats elements present such as Transposable elements. The next track shows DNA methylation in that area in floral tissue from Col-0 wild type. The next track show small RNAs and the last track show messenger RNA.

I attempted TAIL PCR on parent lines for all of those that were methylated in the F1 AMP243 cross and a selection of other non-methylating lines including GFP142. The T-DNA integration sites have been located in 21 lines (Table 3.2). Not all integration sites could be located due to issues with the PCR not producing bands, the Sanger sequencing failing to give a sequence or binary vector sequence amplification (Table 3.2). All lines that could not be sequenced were repeated at least once. The table of results for the TAIL PCR gives the approximate distance to the next gene, repeat element, sites of moderate methylation and mRNA production in Col-0 floral tissue given in base pairs in both directions. Moderate methylation was defined as more than 2 methylated cytosines close together, as indicated by lines on the mC Col-0 track. The TAIL PCR results are consistent as shown by the repeat of the whole PCR process on parents 71 and 108 (Table 3.2).

From these 21 lines it does not appear that the ability the transgene to undergo RdDM is affected by location or by the methylation status of the genomic region into which it has inserted. Methylatable transgenes were found in both repetitive and non-repetitive regions of the genome as were the non-methylatable transgenes. For example the integration site for line 67 occurs within a transposable element but there was no methylation of the *GFP* transgene in the F1 AMP243 X 35S:GFP 67 cross (Table 3.2 & Figure 3.13). Additionally, line 103 was methylated in

the F1 cross but the integration site is not close to a methylated region or repeat elements (Table 3.2 & Figure 3.13).

In some lines the T-DNA inserted into genes. Not all genes found in the TAIL analysis were annotated in ANNOJ but the ones that were are given in Table 3.3. Information directly from the TAIR database is included in Table 3.3 to indicate the genes functions. No obvious phenotypic differences between the individual 35S:GFP lines were observed in this investigation.

Table 3.2. Analysis of TAIL PCR results using ANNOJ

Grey filled rows were methylated in the F1 AMP243 X 35S:GFP cross. Approximate distances are given for both upstream and downstream of the site

GFP Line	Chromosome	Integration site	Approximate distance to next gene (bp)		Approximate distance to repeat elements (bp)		Approximate distance to moderate methylation in Col-0 floral tissue (bp)		Approximate distance to mRNA producing region in Col-0 floral tissue (bp)	
1	At1	30307576	8,116	2,664	4,796	>10,000	>10,000	>10,000	>10,000	>10,000
2	At4	7467726	In gene (not annotated in TAIR)		>10,000	>10,000	286	1,254	>10,000	>10,000
6	At5	4990243	1,043	1,137	>10,000	>10,000	263	7,637	>10,000	5,097
10	Vector only									
12	Failed Sequencing									
14	At 1	25777399	6,319	>10,000	>10,000	5,941	>10,000	>10,000	>10,000	>10,000
17	Vector only									
18	Vector only									
19	At3	6609985	In gene (not annotated in TAIR)		>10,000	>10,000	>10,000	>10,000	765	>10,000
22	Poor PCR									
23	Failed Sequencing									
25	At4	10380471	1,891	2,389	>10,000	>10,000	2,951	>10,000	2,131	3,529
28	At1	12397494	AT1G34060.1		3,034	5,726 TE	~100bp over gene		>10,000	>10,000
29	Vector only									
38	At5	16366275	At5G40820.1		>10,000	7,325 TE	~100bp all over gene		6,375	>10,000
44	Failed Sequencing									
50	Vector only									
56	At1	10065820	2,340	1,560	380	>10,000	4,800	>10,000	>10,000	>10,000
57	At3	8631157	At3G23890.2		6,117	>10,000	7,117	1,777	In mRNA region	

58	At1	23527347	At1G63430.1		1,047	>10,000	In methylated region		In mRNA region	
65	Failed Sequencing									
67	At4	8947143	1,003	2,337	At4TE40045 - In region with other TEs		In methylated region		>10,000	3,857
70	At1	2163354	7,874	2,206	5,514	>10,000	1,714	8,466	8,374	7,366
71	At4	3660942	>10,000	4,838	At4TE16295 - In region with other TEs		In methylated region		>10,000	2,858
Repeat 71	At4	3660942	As before							
82	At2	12978162	862	1,798	1,122	9,478	>10,000	3,358	>10,000	1,798
89	At3	10448933	673	267	313	107	2,593	1,947	2,333	>10,000
97	At4	18168371	191	>10,000	>10,000	309	>10,0000	4,469	>10,000	>10,000
103	At4	17996864	444	2,596	>10,000	>10,000	2,704	4,376	864	2,736
108	At5	17525706	2,628	>10,000	846	>10,000	3,166	4,872	2,846	4,174
Repeat 108	At5	17525708	As before							
114	At1	19843794	In gene (not annotated in TAIR)		>10,000	>10,000	206	74	>10,000	4,706
GFP142	At3	2418107	In gene (not annotated in TAIR)		>10,000	>10,000	2,067	5,833	>10,000	5,213

Table 3.3. Information from TAIR database on genes in Table 3.2 that have been the site of T-DNA insertion in this study

Line	Gene Number	Information from TAIR
28	AT1G34060.1	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
38	At5G40820.1	Encodes an Arabidopsis ortholog of the ATR protein kinase that is involved in a wide range of responses to DNA damage and plays a central role in cell-cycle regulation.
57	ATG23890.2	Encodes a topoisomerase II that is highly expressed in young seedlings.
58	At1G63430.1	Leucine-rich repeat protein kinase family protein; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, ATP binding

3.3 Discussion

3.3.1 Summary

Here 68 potentially single locus 35S:GFP lines have been produced. Of these, 52 lines were crossed with the AMP243 trigger line and silencing among the progeny was assessed. A summary of the work conducted and the numbers of lines used is given in Figure 3.14. All lines underwent silencing but only 8 out of 40 lines supported RdDM. A comparison of the location of the *GFP* transgene gives no indication that the ability to support RdDM is dependent on genome location. This chapter also described the production of homozygous AMP243 X GFP lines which provide useful tools for further work.

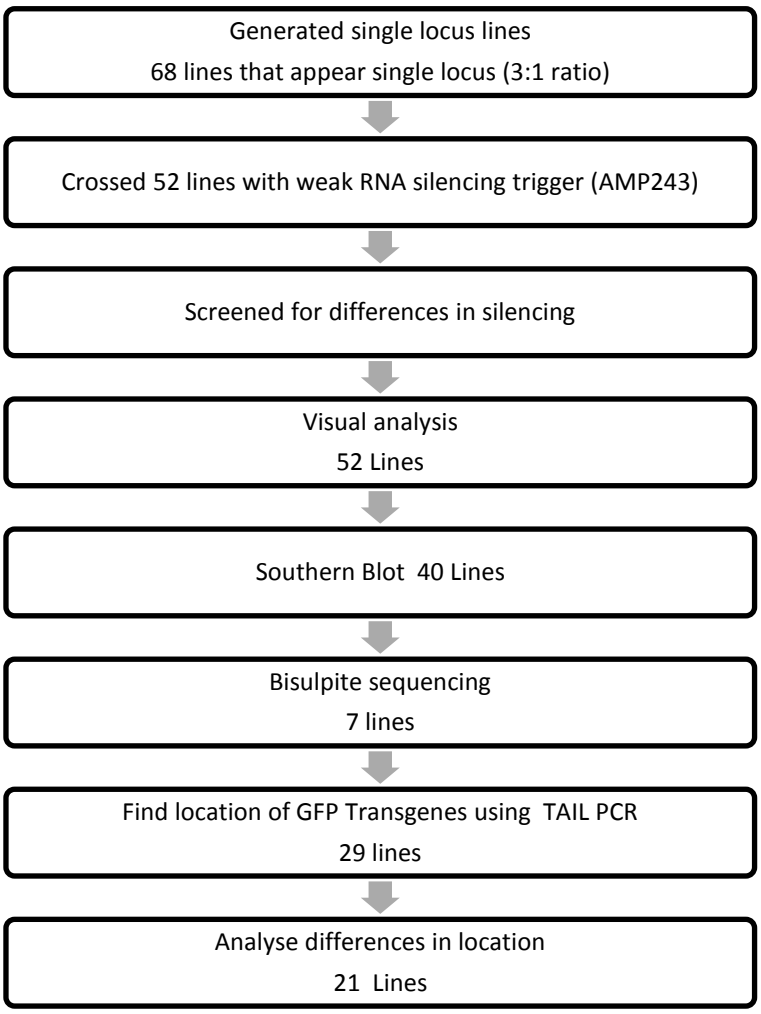


Figure 3.14. Overview of work performed in this chapter

3.3.2 Production of homozygous single locus GFP transgenic lines

The aim of creating multiple independent single locus 35S:GFP lines was to be able to compare a variety of location types; i.e, heterochromatic, euchromatic, within genes, intergenic regions, areas with many repeat etc. While *Agrobacterium* mediated T-DNA insertions remains one of the most efficient ways to transform plants with a gene of interest on a large scale, it does have some limitations. For example, genome-wide insertional mutagenesis studies have found that there is a bias as to where T-DNA will insert (Alonso et al., 2003, Schneeberger et al., 2005). For example, Alonso et al. (2003) found that the number of insertion events becomes considerably lower closer to centromeres and that insertion events can be correlated to gene density along chromosome arms. There is also a bias towards integration in 5'UTRs, 3'UTRs & promoters over introns (Alonso et al., 2003, Schneeberger et al., 2005). The reasons for these biases are unclear. One possible explanation is the T-DNA integration is dependent on Pol II and therefore sites that can be accessed by Pol II act more readily as T-DNA integration sites (Alonso et al., 2003). One of the hypotheses being tested in this project was that silencing may be more likely for transgenes that

had inserted into regions that had some sort of "sign" or structure that made them more of an obvious target of RDR6 or the RdDM machinery, such as existing DNA methylation for example. Perhaps insertions into repeat elements, heterochromatin or areas of heavy methylation would be more likely to support RdDM for example. However, this integration bias may have meant that achieving incorporation of the *35S:GFP* transgene into all of these types of DNA may have been more difficult and may have limited the number of lines of this type available to investigate.

In addition to the effects of T-DNA integration bias on the range of integration sites achieved here, there will also have been an effect caused by the screening method used to select 35S:GFP plants. If certain genomic locations do cause transgenes to undergo silencing then transgenes that have inserted into that region will be silenced and would not have been selected for GFP expression in the initial stages of the investigation. Therefore, the selection technique used here may exclude some of the lines that would be of high interest to a location based investigation. If either the *NPTII* gene or the *GFP* gene were silenced then GFP fluorescence would not be observed and the line excluded from the investigation. However, to generate any form of transgenic line selection criteria must be set and a screening method established. These are likely to have some form of drawback. Additionally, one of the major aims of this project was to look at loci before and after silencing, therefore GFP lines that did not undergo RNA silencing were required. Furthermore lines with multiple site insertions are likely to undergo silencing and these lines were excluded from this investigation.

The T-DNA integration bias and the selection criteria chosen may therefore have resulted in a limited range of integration sites to investigate. For example only one line, 89 had an integration site that was in the pericentromeric region (Table 3.4) (Kawabe et al., 2006). Of the 21 integration sites that were identified, 8 of those were within a gene and 13 were not. It is not clear however how many of the 13 integration sites were in promoters, as not all genes have annotated promoters in the TAIR database. The low number of pericentromeric region integration observed here is not unique in studies of transgene. For example, Schubert et al. (2004) created 107 single copy transgene lines using *Agrobacterium*-mediated T-DNA insertion, but only two of these transgenes inserted into pericentromeric heterochromatin.

Table 3.4. Centromeres in *Arabidopsis thaliana*.

The approximate position of centromere is from Kawabe et al. (2006) and the approximate position of the pericentromere is from Alonso et al. (2003) & Kawabe et al. (2006)

Chromosome	Approximate chromosome length (bp) (TAIR)	Approximate position of Centromere (Mb)	Approximate position of Pericentromere (Mb)
1	34,964,571	14.16 – 15.93	13.54 – 16.78
2	22,037,565	3.52 – 4.32	3.05 – 8.15
3	25,499,034	13.49 – 14.10	10.15 – 15.03
4	20,862,711	3.90 – 4.07	3.67 – 5.72
5	31,270,811	11.67 – 12.09	9.19 – 14.03

One alternative method of screening for GFP lines with a single locus insertion of the transgene would have been to screen for transgenic lines that had a 3:1 segregation pattern of Kanamycin resistance. PCR could then have been performed on these lines to screen for the *GFP* gene and then visual analysis could have been used to find lines with active GFP expression. RNA analysis would have also been used to find if the lack of GFP expression was the result of RNA silencing. This may have resulted in the identification of lines in which the *GFP* transgene was silenced but present at only a single integration site. This would have allowed investigation into how differences in location may have affected non-triggered RNA silencing. However, this would not have circumvented the issue of T-DNA integration bias. The reason this method was not deployed in this investigation was that it was more labour intensive than screening for GFP fluorescence. Additionally, the integration site analysis by TAIL PCR was not conducted until after the AMP243 work had been significantly invested in. If the TAIL PCR work had been done first and the limited number of heterchromatic inserts have been realised perhaps this method would have been used. However, if location does have a strong influence over silencing then the *NPTII* kanamycin resistance gene would also likely to have been silenced.

It should also be noted that there are limitations to using a 3:1 ratio of expression vs no expression to find single locus lines. The ratio was analysed using the χ^2 test to indicate which lines are likely to have a true 3:1 ratio and therefore a possible single integration site. The χ^2 test will have found lines that are statistically likely to have a true 3:1 ratio but, as with all statistics, there are limitations to this. Additionally a p-value of 0.05 was used which, although the most commonly used, it is not the most stringent cut off point. Some of the lines assumed to have a 3:1 ratio in this analysis would have been discounted using a different p-value. Additionally, assuming that lines that meet the criteria of the χ^2 analysis have a single site of integration has limitations. For example, this method of analysis would not be able to filter out lines with linked

multiple insertions. It would also not detect any second silenced copies of the *GFP* transgene. This will be discussed further in chapter 5.

3.3.3 Locus-specific effects

Section 3.1 and the general introduction outlined some of the observations that highlight the possibility that there may be locus-specific effects influencing a transgene's ability to support PTGS or RdDM. These locus-specific effects could either be in the form of the effect of genomic location or the effects of locus-specific transgene structure. In this chapter the former was investigated. In section 3.1 observations from the literature regarding RDR6 were outlined which suggest that the action of RDR6 may be responsible for these loci specific effects. The amplicon based silencing of line 142 required the action of RDR6 (Dalmay et al., 2000a, Herr et al., 2005). If all AMP243 based silencing is RDR6 dependent then the fact that amplicon-triggered PTGS occurred in all of the independent lines in this study when crossed with AMP243 would suggest that RDR6 activity is not influenced by either genome location or locus structure.

Although there seems to be no effect of genomic location on AMP243 triggered PTGS, the 40 lines screened differed in their ability to support DNA methylation in the F1 AMP243 crosses. To determine if these differences were caused by genomic location, the integration sites, as determined by TAIL PCR, were compared between the methylating and non-methylating lines. However, no discernible pattern could be determined suggesting that genomic location does not affect AMP243 triggered DNA methylation. For example, proximity to TEs or insertion into genes did not seem to influence AMP243-triggered DNA methylation.

It had been hypothesised that insertion into a heterochromatic region, such as that found at the pericentromere, would lead to RNA silencing. Although all lines here underwent AMP243 triggered PTGS, it could have been possible that insertion into a silenced region would also lead to DNA methylation. However, for reasons discussed above only one line was generated that had the *35S:GFP* transgene inserted into such a region. Line 89 had an integration site that was in the pericentromeric region (Table 3.4) but this did not undergo DNA methylation in the F1 AMP243 cross (Kawabe et al., 2006). However, in euchromatin TE methylation is predominantly conducted by RdDM whereas in heterochromatic regions TE methylation requires DECREASE IN DNA METHYLATION 1 (DDM1) and is independent of RdDM (Zemach et al., 2013). The pericentromeric region is more heterochromatic than euchromatic (Wang et al., 2006). Therefore, it could be reasoned that the *35S:GFP* transgenes inserted into that region may not be accessible to RdDM machinery, and therefore would not undergo AMP243-triggered RdDM. Additionally, the hypothesis that integration into a pericentromeric region would increase the likelihood of RdDM of a transgene relies on the bases that the pericentromeric region is methylated and that

neighbouring methylated DNA is able to recruit the relevant machinery to the transgene loci once RNA silencing has been triggered. However, according to the results from ANNOJ the integration site in line 89 is not in an area of DNA methylation, even though the site is reportedly in the pericentromeric region on chromosome 3 (Table 3.4). However, a sample size of one cannot allow firm conclusions to be made about the effect of integrating in to a pericentromeric region may have on methylation of a transgene.

TE are common targets of RdDM and tend to be methylated. Therefore another hypothesis was that if transgene integrated into regions of TEs then they would be more likely to undergo DNA methylation once silencing had been triggered due to the presence of RdDM machinery at neighbouring loci. The T-DNA in both lines 67 and 71 had integrated into an area with a high density of TEs however, only line 71 underwent DNA methylation in the F1 AMP243 cross. Spreading of DNA methylation from TEs on to neighbouring transgenes can occur from some TEs. Such an observation has been made in Maize where DNA methylation has been demonstrated to spread from a specific family of retrotransposons to silence neighbouring genes (Eichten et al., 2012, Gent et al., 2013). However, this cannot be the case here as DNA methylation was not detected by Southern blot in either parent. As previously discussed, some TEs are methylated by RdDM while others require DDM1 (Zemach et al., 2013). Perhaps in line 71 the transgene is next to TE that are methylated by the action of RdDM and therefore the machinery is also recruited to the transgene loci once silencing of GFP has been triggered. The TEs adjacent to the transgene in line 67 may be methylated by the DDM1-dependent pathway and so no RdDM machinery may be present at this loci. Again, it is hard to make conclusions from a sample size of two.

Of the lines whose integration sites are in areas of DNA methylation, 58, 67 and 71, only line 71 has DNA methylation present at the GFP loci in the F1 AMP243 cross. Therefore the hypothesis that DNA methylation at the integration site may lead to methylation of the transgene appears to be refuted. Additionally line 71 is not the only line where the *35S:GFP* transgene undergoes RdDM in the F1 AMP243. It should be noted that the ANNOJ Epigenome browser is limited to methylation patterns in flowers whereas all of DNA samples used for the Southern blot and the bisulfite sequencing were extracted from leaves.

The sample size of F1 AMP243 cross methylating lines vs non-methylating lines is quite small. The sample size of methylating lines with a known integration site is even smaller. All 8 lines that underwent DNA methylation were analysed by TAIL PCR. Of these only four were able to be sequenced and give an integration site for the transgene. An additional 17 lines that did not undergo methylation in the F1 had the T-DNA integration site located by TAIL PCR. Comparing four lines to 17 lines makes drawing firm conclusions hard. If all 8 F1 methylating lines could have

had their integration sites located then this may have provided more confidence in any patterns observed.

Although there was evidence in the literature available at the beginning of this project that supported the existence of genomic position effects, other work has been presented that does not. For example, Schubert et al. (2004) investigated genomic position effects on single copy transgenes that had been integrated via Agrobacterium. They concluded the positional effects did not affect the transgene expression, even for the two cases where the transgene had inserted into heterochromatin. This conclusion was based on the observation that all of the single copy lines generated did not undergo silencing, a similar result to the data presented here. However, the transgenes in their paper also relied on kanamycin selection and so the work has the same issues as faced here with regards to the necessary expression of selectable marker genes. This may also suggest the screening the 35S:GFP lines here using kanamycin resistance and not GFP to find a 3:1 of expression may have also given the same result already achieved here.

3.3.4 Overlap between PTGS and RdDM

It was surprising that only 8 out of the 40 lines underwent PTGS-associated methylation as previous work performed in the lab group and reports in the literature had led us to believe that there would be a strong overlap between PTGS and RdDM (Vaistij et al., 2002, Jones et al., 1999). The work performed here has demonstrated that this is not the case for AMP243 triggered silencing and that most of the lines only underwent PTGS.

Here detection of GFP mRNA and siRNA was performed on a subset of lines to establish if the loss of GFP fluorescence was associated with the loss of GFP mRNA and accumulation of GFP siRNAs. This was the case for the lines investigated here. It should be noted that it is possible that the loss of GFP fluorescence was not associated with the accumulation of siRNA in lines that were not tested. However, due to the GFP expression in the parent and the loss of GFP in the presence of the AMP243 trigger it has been assumed that all lines did accumulate siRNA in the F1 AMP243 crosses. The subset of crosses used for the siRNA work included lines that had been methylated in the F1 AMP243 cross indicating that the methylation was a result of RdDM. The sizes of the siRNA detected in the siRNA northern blots were not determined here. Therefore it cannot be determined if the DNA methylation observed in the F1 AMP243 crosses were of the 24nt class of siRNA that are commonly associated with RdDM or a smaller 21nt class of siRNA that can be linked to PTGS (reviewed by Matzke & Mosher, 2014). Work has also suggested that 22nt sRNA are more likely to act as templates for RDR6, although this may be due to the interaction of the sRNA with AGO1 (Cuperus et al., 2010, Chen et al., 2010).

Overlap between PTGS and RdDM has been seen by others (Eamens et al., 2008, Garcia et al., 2012, Jauvion et al., 2012, Pontier et al., 2012). Pontier et al. (2012) have proposed a new siRNA pathway that connects RDR6 to DNA methylation at some loci via NEEDED FOR RDR2-INDEPENDENT DNA METHYLATION (NERD). This is a GW repeat protein that has a PHD finger module which allows binding to H3K4me0. NERD binds DNA, as determined by ChIP, and is proposed to interact with AGO2 and RDR6 to bring about DNA methylation via 21nt sRNAs and may involve the action of Pol V (Pontier et al., 2012). NERD mutants have reduced DNA methylation at selected loci and increased gene expression at a subset of these. Although work by Garcia et al. (2012) and Pontier et al. (2012) on NERD focuses on endogenous genes, and not transgenes, the NERD pathway does demonstrate a clear connection between PTGS and RdDM which could potentially affect transgenes.

However, DNA methylation of a transgene does not always overlap with RNA silencing of the gene, especially when the gene body is methylated but not the promoter (Masclaux et al., 2005). Here only GFP gene body DNA methylation was investigated and not methylation of the 35S promoter. Methylation of the promoter is often associated with TGS (Otagaki et al., 2011).

3.3.5 Other areas of inquiry

The fact that all lines here underwent RNA silencing in the presence of the AMP243 trigger suggests that RDR6-dependent AMP243 silencing is not affected by the site of genomic integration. However, it would be prudent to check that the silencing observed here is RDR6-dependent, as the literature suggests (Dalmay et al., 2000a, Dalmay et al., 2001, Herr et al., 2005). This will be investigated in chapter 4. If the site of genomic location does not affect RDR6 activity then it remains unclear what signals a transcript to become a template for RDR6. There are many other schools of thought on the subject. For example, others have postulated that aberrant transcripts that lack the poly A tail or the 5' cap may be recognised by RDR6 as templates (Gazzani et al., 2004). However, none provide a satisfactory explanation as to why mRNA of endogenous genes remains recalcitrant to the effects of RDR6-mediated amplification, even when these are targets of RNA silencing.

The observation that some lines were able to undergo AMP243-triggered RdDM when most did not will become one of the new focuses of this investigation. There are other things that could affect the ability of a line to support transgene RdDM that have not been investigated in this chapter. For example the F1 of the DAN2 recreation cross between GFP142 and AMP243 showed no sign of DNA methylation but subsequent generations do. Differences in zygosity have been observed to effect gene silencing and RdDM and so this may be another factor to consider (James et al., 2002). This will be explored in chapter 5.

The possible influence of the integration structure of the transgene is discussed in section 3.3.2 with regards to line 17. This could be interesting to investigate because although the 35S:GFP lines should be single locus they are not necessarily single copy number. Chapter 5 will therefore also explore the effects of *GFP* copy number and integration complexity on transgene RdDM.

AMP243 is only one way for triggering RNA silencing of transgenes. The SUC-GF:IR trigger line was also used here to try and silence the *35S:GFP* transgene. However, GFP fluorescence was not reduced in lines that were crossed with this trigger, as determined by visual analysis, suggesting that different triggers of silencing may give different results.

Using a trigger allows differences in the loci before and after silencing to be investigated, which will be explored in later chapters. For example, RdDM has been connected to specific histone modifications (Law et al., 2013, Law et al., 2011). It is possible that the observed differences in AMP243 triggered PTGS and RdDM may be due to between line variations in histone modification present at the 35S:GFP loci. Furthermore, there has been much discussion in this chapter regarding the possible recruitment of RdDM machinery to transgenic loci due to the existing presence of the machine at neighbouring loci. It is possible that the RdDM machinery, such as Pol V which is known to interact with its targets, associates to certain loci before the triggering of silencing by the generation of siRNA. This will be explored and discussed further in subsequent chapters. Additionally, tools that will be useful in investigating recruitment of RdDM machinery, either prior or after silencing, will be described in chapter 4. These are FLAG fusion proteins of key RdDM and PTGS proteins such as RDR6, RDR2 and Pol V which will allow immunoprecipitation experiments to be conducted. The development of the homozygous AMP243 X 35S:GFP crosses described in section 3.2.3 will be a useful tool for crossing with such FLAG tagged lines as they provide stable expression of the *35S:GFP* transgene.

3.3.6 Conclusions

This investigation strongly suggests that genome location does not influence AMP243-induced RNA silencing. However, the results presented here also suggest that although RdDM may not be influenced by genome location, locus-specific differences are involved.

3.4 Acknowledgments

Olivier Voinnet provided the SUC-GF:IR trigger lines.

Chapter 4 Development of epitope-tagged RNA silencing components

4.1 Introduction

The issue of how key proteins in the PTGS and RdDM pathways become recruited to their targets is an ongoing area of research. In this project recruitment of key proteins was seen as part of the original question into locus specific affects. It was reasoned that if location did affect a targets ability to support RNA silencing then perhaps there was also a difference in that locations ability to recruit the proteins involved. The aim of this chapter was to generate material that could be used to test the hypothesis that differences in the recruitment of silencing proteins was responsible for locus-specific silencing differences. The planned experimental approach was to generate or obtain epitope-tagged proteins and to use these in Chromatin-immunoprecipitation reactions. These could also be potentially used to study protein-protein and RNA protein-interactions.

Chapter 3 described the generation of single-locus *35S:GFP* transgenic lines and the results of experiments designed to test whether these lines differed in their ability to support silencing and/or DNA methylation triggered by an amplicon transgene. Although it was demonstrated in Chapter 3 that genomic location of a *GFP* transgene does not affect AMP243 triggered PTGS or RdDM it still remains unclear why the lines varied in their ability to be methylated. It raises the question of what identifies a locus as a target for methylation. One possibility is that the ability of loci to recruit key silencing components of RdDM could be responsible. For example, a locus that is able to be methylated may be able to recruit Pol V whereas a locus that is not methylated may be unable. In the general introduction the self-maintaining feature of both transcriptional and post-transcriptional RNA silencing pathways was described. This feature makes the issue of recruitment of silencing components complicated to study in systems where silencing is already established, such as for complex transgenes or endogenous targets. The AMP243 system described in Chapter 3 involves multiple independent target lines and therefore loci can be studied in the presence and absence of the silencing trigger and also loci can be compared with each other.

The proteins selected to use in these experiments were the RNA dependent RNA polymerases RDR2 and RDR6, and the plant specific Polymerases Pol IV and Pol V. RDR6 was chosen due to its importance in PTGS as described in the general introduction. Additionally, AMP243 triggered silencing of a transgene has been demonstrated to be RDR6-dependent (Dalmay et al., 2000a,

Dalmay et al., 2001). RDR2 was selected as it is an important factor in canonical RdDM, as described in the general introduction. In *S. pombe* RNA-dependent RNA polymerase (Rdp1) has an important role in the localization of the RNA-induced initiation of Transcription Gene Silencing (RITS) complex to target loci (Sugiyama et al., 2005, Volpe et al., 2002). The roles of both RDR6 and RDR2 in AMP243 triggered silencing will also be investigated in this chapter briefly using mutant lines of each protein.

The plant specific polymerases Pol IV and Pol V have key roles in canonical RdDM, as described in the general introduction. AMP243 triggered silencing has been previously demonstrated to be Pol IV-dependent (Herr et al., 2005). It has also been demonstrated that Pol V is required for RdDM in plants that undergo PTGS (Eamens et al., 2008).

Prior to the start of this project other groups had generated epitope-tagged RDR2, RDR6, Pol IV and Pol V and had provided invaluable insight into the RNA silencing pathways. For example, Ream et al. (2009) used affinity purification of NRPD1:FLAG and NRPE1:FLAG fusion proteins to elucidate the subunit composition of both Pol IV and Pol V respectively. Huang et al. (2009) preformed similar work with NRPE1 fusion proteins to also resolve the subunit composition of Pol V and in the process found that KTF1 associates with Pol V. Additionally, Wierzbicki et al. (2008) used NRPE1:FLAG fusion proteins in ChIP experiments to demonstrate that NRPE1 occupies the site of a solo long terminal repeat (LTR), a target of RdDM. They also use the fusion protein to demonstrate the production of RNA transcription from this site, corroborating that transcription by Pol V produces scaffold RNA during RdDM.

Here the aim was to generate full-length RDR2 and RDR6 constructs that could be combined with the FLAG epitope using the Gateway Cloning System from Invitrogen. Additionally, creating a fusion protein between FLAG and the largest subunit of Pol IV, NRPD1, was also planned. A NRPE1:FLAG fusion protein containing line was acquired from Dominique Pontier. Once created or acquired the experimental approach planned was to use these epitope-tagged lines in ChIP reactions to investigate protein-DNA interactions at the sites of transgene silencing by crossing these lines with the homozygous AMP243 and 35S:GFP lines described in Chapter 3.

A binary vector system, which relies on the Gateway Cloning System (Invitrogen), was used where the gene of interest is cloned into an entry vector before being recombined into a destination vector to add the desired epitope or tag to the gene of interest. The system has the advantage of having a variety of destination vectors which should make adding different tags relatively straight forward. Therefore, in conjunction with the addition of FLAG to the key proteins investigated here, the addition of the GFP sequence to *RDR2*, *RDR6* and *NRPD1* was also planned. GFP fusion proteins would have allowed the cellular localisation of the proteins tagged to be assessed by

visualizing them using confocal microscopy. The cellular location of both RDR2 and RDR6 in Arabidopsis had already been ascertained by other groups and so this line of investigation was not the main thrust of this project (Hoffer et al., 2011, Kumakura et al., 2009, Luo & Chen, 2007). However, there is some disagreement in the literature as to whether RDR6 localises to the nucleus alone or to the nucleus and the cytoplasm. Creating a RDR6:GFP fusion protein in this investigation may have helped to address that issue. RDR2 has been reported to localize in the nucleus, forming a crescent around the perimeter of the nucleolus and also forms a "dot" within the nucleolus (Pontes et al., 2006). It has also been demonstrated that both Pol IV and Pol V can localize with chromosomal loci by Pontes et al. (2006). Additional, Pol V can also localize with other RdDM proteins in the nucleolus away from chromosomes (Pontes et al., 2006).

4.2 Results

4.2.1 AMP243 silencing is RDR6-dependent and is delayed in *rdr2* mutants

In order to determine whether RDR2 and RDR6 are required for AMP243 triggered silencing of the GFP lines generated in this work, five homozygous AMP243GFP lines (58, 60,65, 82, 103) were crossed with a *rdr2* mutant (SALK_059661) and three AMP243GFP lines (58, 60, 65) were crossed with the *rdr6-11* mutant line. The F1 progeny of these crosses remained silenced for GFP (data not shown). The F2 progeny were germinated on MS media containing kanamycin and PPT in order to select for the presence of both transgenes. Seedlings were then transferred to soil and assessed for production of GFP using UV microscopy.

F2 plants from all crosses showed segregation of the GFP silencing phenotype with approximately ¾ plants appearing fully silenced and ¼ plants having visible GFP for each line. In all of the *rdr6* X AMP243GFP crosses, approximately ¼ of the F2 progeny exhibited green fluorescence; representative individuals of the F2 generation are shown in Figure 4.1. These plants should contain both the AMP243 and *35S:GFP* transgenes and the proportion of plants observed is consistent with the GFP positive plants being homozygous for the *rdr6-11* mutation. This indicates that the AMP243 triggered silencing is RDR6-dependent, as expected from previous work (Dalmay et al., 2000a, Dalmay et al., 2001).

Like the *rdr6-11* crosses, F2 progeny of the AMP243GFP X *rdr2* crosses exhibited GFP fluorescence in approximately ¼ of plants. However, rather than being fully green fluorescent, the plants displayed a delay in the establishment of GFP silencing. In these plants the younger leaves were green fluorescent whereas older leaves were silenced (Figure 13). This suggests that RDR2 is involved in AMP243 triggered RNA silencing, perhaps at the initiation stage due to the delay of silencing observed. Further work would be required to establish the role of RDR2 in RDR6-

dependent PTGS. The work using the FLAG-tagged lines described in this chapter may be of some use to achieve this.

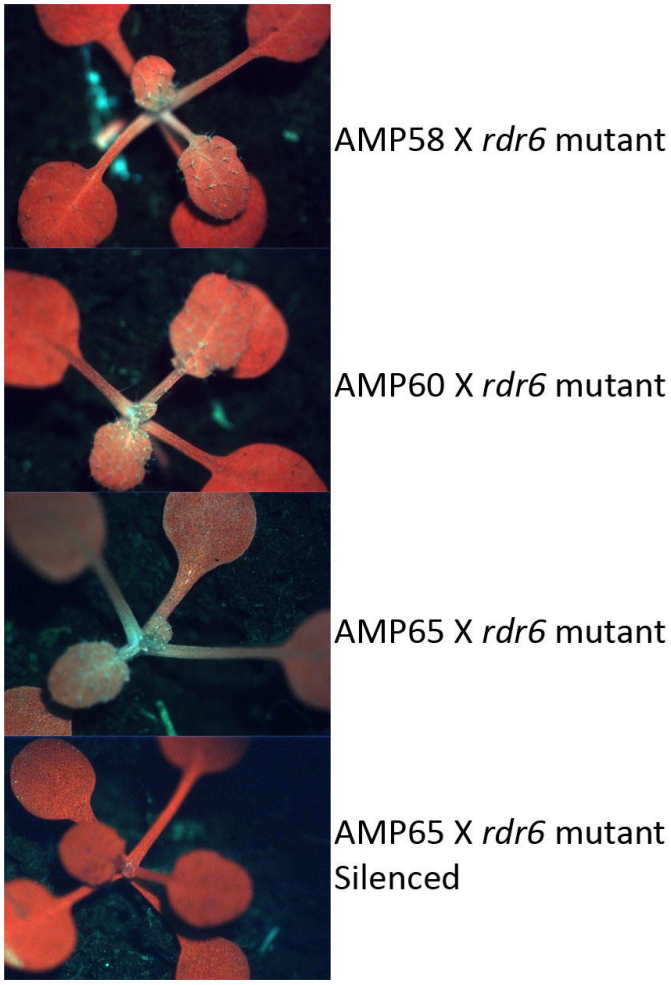


Figure 4.1. AMP243 triggered silencing of the *GFP* transgene is RDR6 dependent

Analysis of GFP fluorescence by UV microscopy in *rdr6* mutants crossed with 3 different AMP243GFP lines. Representative individuals of the F2 generation are shown. First three panels show the AMP243GFP X *rdr6-11* phenotypes and the last panel shows a representative individual of the red silenced AMP24365 X *rdr6* plants.

It was not possible to test whether NRPE1 is required for AMPXGFP silencing as the *nrpe1* mutant line used for the crosses resulted in silencing of the *35S:GFP* transgene in the F1 state when crossed in the absence of AMP243. This could be due to trans-silencing effects of the T-DNA insertion in the *nrpe1* mutant and the 35S promoter in the 35S:GFP lines (Daxinger et al., 2008). When *rdr2* and *rdr6* mutants were crossed with the 35S:GFP alone all of the progeny were green (data not shown).

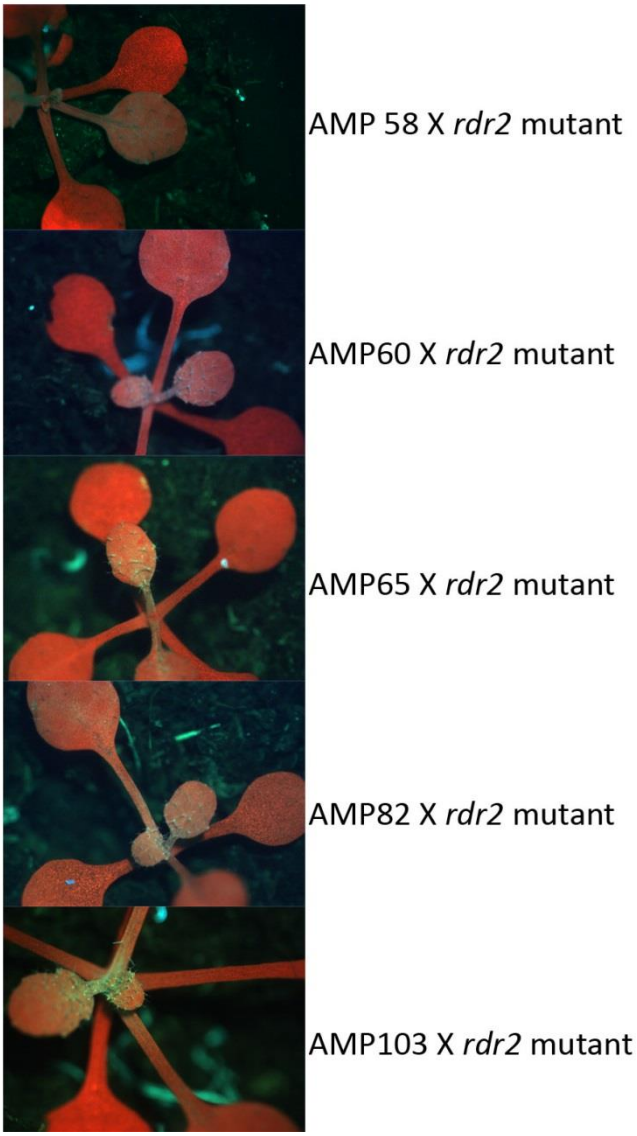


Figure 4.2. AMP243-triggered silencing is delayed in *rdr2* mutants

Analysis of GFP fluorescence by UV microscopy in *rdr2* mutants crossed with 5 different AMP243GFP lines. Representative individuals of the F2 generation are shown.

4.2.2 Using the Gateway cloning system

To produce full-length C-terminal FLAG-tagged RDR2 and RDR6 constructs that were under the control of the native promoter the Gateway cloning system from Invitrogen was used. Pol IV is a larger multi-subunit enzyme and so the aim was to create a C-terminal FLAG-tagged version of its largest subunit, NRPD1a also.

The gateway system has the advantage that once the gene of interest is in the pENTR vector it can be introduced into a wide variety of destination vectors. Therefore, GFP- tagged versions of these proteins under the control of the constitutive 35S promoter from Cauliflower Mosaic Virus were also attempted. A schematic of the work involved in this binary vector system is given in Figure 4.3.

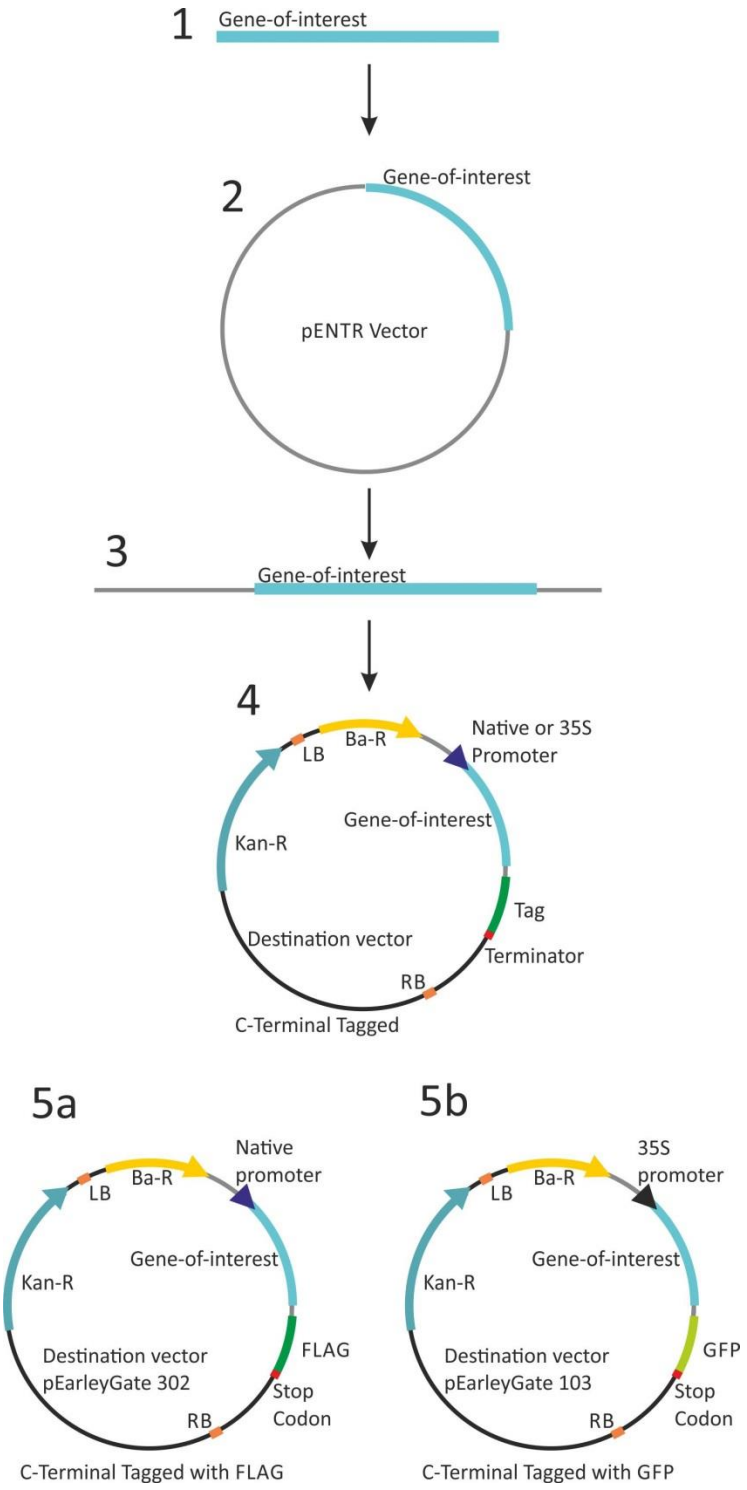


Figure 4.3. Schematic of work involved in creating epitope tagged constructs and desired outcomes

Step 1. PCR fragments were generated of gene with CACC sequences. **Step 2**. Clone PCR fragment into pENTR vector and transform this into One Shot Chemically Competent *E. coli*. **Step 3.** Linearize with Drd1. **Step 4.** Recombine with the appropriate destination vector to create a C-terminus construct. **Part 5**. Desired vectors carrying C-terminal tagged gene-of-interest. **5a)** FLAG tagged under native promoter. **5b)** GFP tagged under 35S promoter. Kan-R is a Kanamycin resistance gene and Ba-R is a BASTA resistance gene. LB and RB are the left and right border sequences for T-DNA integration.

The first stage of the Gateway Cloning System is to design primers that would amplify the target gene and include the sequence (CACC) necessary for directional cloning at the 5' end of the resultant PCR product (Invitrogen). The sequence of all primers used can be found in methods section 2.2 of Chapter 2. The *RDR2* and *RDR6* genes including their predicted promoter regions are 4224 bp and 4650 bp in length, respectively (TAIR). However, for NRPD1a, only 35S driven constructs were attempted due to the large size of the NRPD1a gene which would impact on the efficiency of Gateway cloning, and uncertainty regarding the extent of the promoter region. The genomic region which encodes the *NRPD1a* subunit of Pol IV is 7236 bp from start to stop codon (TAIR). Excluding introns *NRPD1a* is 4362 bp long and therefore cDNA was used as the template for PCR. The region upstream of the *NRPD1a* start codon has a 1,438bp 5' UTR which contains an intron.

Using these primers, PCRs were performed on genomic DNA using the proofreading polymerase Pfu Ultra II (Agilent). PCR fragments of the expected size were purified by gel extraction following agarose gel electrophoresis. The TOPO cloning reaction was used to introduce these fragments into the pENTR vector. The subsequent pENTR vector was then transformed into One Shot Chemically Competent *E. coli* (Invitrogen). Kanamycin selection on LB plates was conducted followed by colony PCR with the internal control primers to find colonies containing the gene fragment. Plasmids were extracted from PCR positive colonies and sequenced to determine whether the expected gene sequence was present in the plasmid and that it had no PCR-induced sequence changes.

Once a construct was in the pENTR vector and the sequence was confirmed, this was then recombined with the appropriate destination vector to create the desired tagged construct. Initially three different destination vectors were used. The pEarleyGate302 vector was used in conjunction with the pENTR vectors that contained the *RDR2* and *RDR6* constructs which included the native promoters to create a C-terminal FLAG-tagged constructs. To produce GFP tagged constructs the pENTR vectors with the 35S:RDR6 & 35S:RDR2 fragments were used with pEarleyGate103 vector to give C-terminus GFP-tagged proteins under the control of the 35S promoter. For selection in *E. coli* all of the pEarleyGate vectors had a Kanamycin resistance gene and for selection in plants they had a BASTA resistance gene.

Due to the fact that the destination vectors used and the pENTR vector all rely on Kanamycin resistance for selection, the pENTR vectors had to be linearised using the restriction enzyme DrdI before the recombination reaction. DrdI cuts within the pENTR part of the plasmid once only, leaving the inserted gene-of-interest intact. The linearized vectors were used in a LR

recombination reaction with the destination vector and then DH5 α competent cells (Invitrogen) were transformed with the recombination product.

The completed binary vectors were then grown-up in the DH5 α cells. The binary vector plasmids were extracted and the 5' and 3' junctions between the introduced gene sequence and the vector was sequenced (data not shown). The completed Gateway binary vector constructs were then introduced into *Agrobacterium tumefaciens* (GV3101) using electroporation and selected on kanamycin plates. The *A. tumefaciens* strain containing the completed Gateway binary vector constructs were used to test the constructs by transiently expressing them in *Nicotiana benthamiana* (*N. benthamiana*) by agroinfiltration before the constructs were introduced into Arabidopsis plants via the floral dip technique (Bent & Clough, 1998, Bent et al., 2000). Arabidopsis mutants of the gene-of-interest were used for this so that restoration of the wild-type phenotype could be used to assess the biological function of the fusion proteins.

Using the Gateway cloning system both C-terminally FLAG-tagged and GFP-tagged RDR2 and RDR6 binary vectors were created.

4.2.4 Attempts to generate NRPD1a tagged lines were unsuccessful

Initially the aim had been to create fusion proteins with the NRPD1a subunit from Pol IV. A pENTR vector containing construct 35S:NRPD1a was successfully created. This construct contained the *NRPD1a* open-reading frame generated with the N-term primer and the No Stop primer from the primer table in chapter 2 (Table 2.1). However, this pENTR vector could not be recombined into any of the destination vectors tried. After 5 attempts at recombination, work on construct NRPD1a was discontinued.

4.2.5 The production and characterization of GFP-tagged lines

It would have been desirable to create GFP fusion proteins of RDR6 and RDR2, which would have allowed the visualization of their locations *in planta*. An RDR6:GFP fusion protein may have helped to shed some light on the disagreement between work by other groups as to whether RDR6 localises to the nucleus or to the nucleus and the cytoplasm (Kumakura et al., 2009, Luo & Chen, 2007). However, as described in the chapter introduction this was not one of the main aims this project.

4.2.5.1 Using transient expression in Nicotiana benthamiana to test the expression of the GFP-tagged constructs

Initially the GFP constructs were tested by transiently expressing them in *N. benthamiana* via infiltration using *A. tumefaciens*. Confocal microscopy was used to image *N. benthamiana* leaves 3 days after the infiltration of leaves with the GFP tagged constructs (Figure 4.4). All images

showed bleed though of chlorophyll into the GFP channel. This is because of the wide emission spectrum chlorophyll fluoresces at when excited with the 488nm laser. Chlorophyll auto-fluorescence is also exacerbated by the infiltration procedure itself, which can cause cell damage. To negate any difference between the control and the test samples in terms of cell damage, the control tissue had also been infiltrated with *Agrobacteria* that lacks a GFP-expressing binary vector. It is possible to filter out the chlorophyll autofluorescence using spectral unmixing. However, the GFP levels were low and required high laser power to visualise them. This was too high to perform spectral unmixing efficiently as the GFP spectral trace was too poor to use.

Unfortunately, the leaves infiltrated with the RDR2:GFP containing *Agrobacteria* looked the same as the control leaf. The RDR6:GFP sample showed a clear difference from the control (Figure 4.4c &d). The RDR6:GFP appeared to be localising to the cytoplasm of the cells and possibly in the nucleus as highlighted by an arrow (Figure 4.4c), however a nuclear stain, such as DAPI which stains nuclear DNA, would need to be included to validate this. I did not use such a stain in these *N. benthamiana* experiments as I was optimistic that the construct would work in Arabidopsis and I planned to include a nuclear stain at that point.

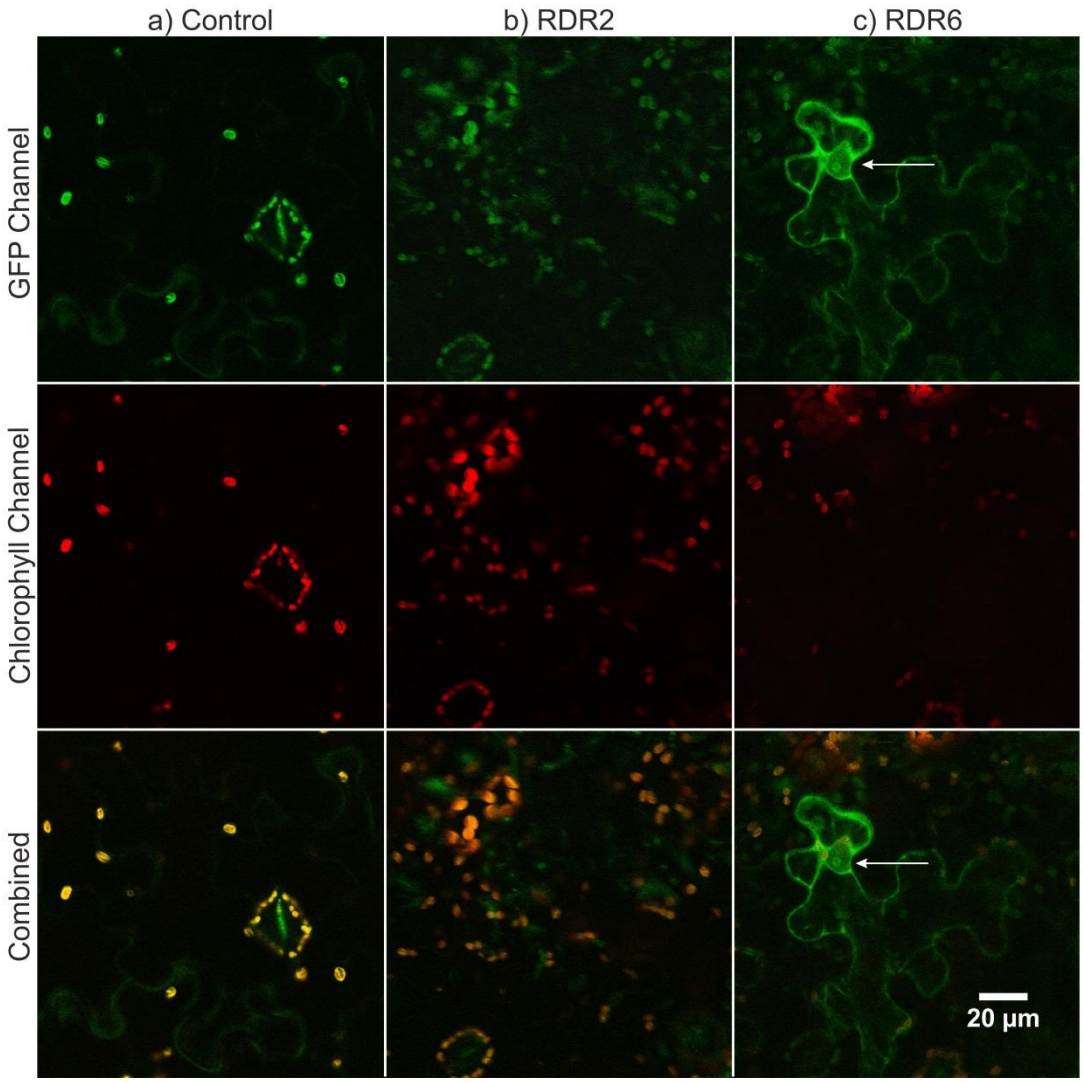


Figure 4.4. RDR6:GFP showed GFP fluorescence when transiently expressed in *Nicotiana benthamina* as determined by confocal microscopy.

Confocal images of epidermal cells from the abaxial surface of *Nicotiana bethamina*. The top row shows the GFP channel, the second row shows the Chlorophyll channel and the third row shows the GFP and the Chlorophyll channel combined. The columns show **a**) the control sample which was infiltrated with *A. tumefaciens* lacking the GFP gene, **b**) RDR2:GFP, **c**) RDR6:GFP from different leaf samples. The white arrow indicates the possible location of the nucleus, but a nuclear stain would be needed to confirm this.

4.2.5.2 Transformation of the GFP tagged constructs into Arabidopsis

An *rdr6 A. thaliana* Col-0 mutant, *rdr6-11*, was available (Kunova et al., 2012). The use of a mutant line allowed for complementation of the mutant phenotype to be screened for to confirm the biological functionality of the construct. The *rdr6-11* line was transformed via floral dip with the RDR6:GFP binary vector. The T1 generation was selected on soil using BASTA spraying and eight transgenic plants in this generation were analysed using confocal microscopy. However, of

the eight T1 RDR6:GFP plants analysed none produced visual signs of fluorescence (data not shown).

RDR2:GFP was transformed into Arabidopsis Col-0 as no suitable BASTA-sensitive *rdr2* mutant was initially available. Again no signs of GFP expression were seen in these lines even though they had successfully been selected with BASTA.

Since GFP could not be visualised when fused to either RDR2 or RDR6 in Arabidopsis this work was not pursued further.

4.2.6 Validating a mutant *rdr2* Arabidopsis line

The *rdr2-1* mutant line (Xie et al., 2004) was initially available in the lab, however it was found to be resistant to BASTA. This is the plant selection marker used in the Gateway destination vector used here and therefore an *rdr2* mutant line that was not BASTA resistant was required. The SALK_059661 mutant line was acquired and was confirmed as BASTA sensitive (data not shown). The mutation is caused by a T-DNA insertion into the *RDR2* gene (Arabidopsis Biological Resource Centre) and plants homozygous for the insertion were selected via PCR (data not shown). The mutant phenotype of the SALK_059661 mutant needed to be validated so that it could be used in phenotype recovery analysis later to check the biological functionality of the constructs. To do this a Southern blot was performed using DNA extracted from leaves of Col-0 and both *rdr2* mutants. The available BASTA resistant *rdr2-1* mutant acted as a control for validating the new mutant. RDR2 is needed for DNA methylation of the 5s rDNA repeat and therefore *rdr2* mutants should have lower levels of 5s rDNA methylation (Douet et al., 2009). The DNA samples were digested with HaeIII and analysed via Southern blotting using a probe specific for the 5s rDNA repeats (Figure 4.5). The blot shows methylation of this region in the wild type, indicated by the higher molecular weight bands (Figure 4.5). In the *rdr2* mutants the higher molecular weight bands are absent in both the *rdr2* and SALK_059661 indicating a lack of methylation in this region consistent with expectations.

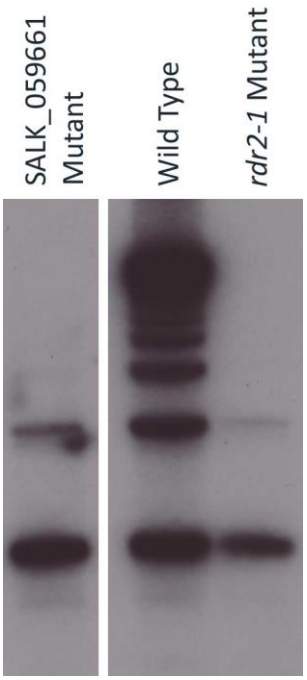


Figure 4.5. Validation of the SALK_059661 *rdr2* mutant by Southern Blot.

A Southern blot of the SALK_059661 mutant with a wild type and *rdr2-1* mutant controls. Genomic DNA from leaves was digested with the methylation sensitive restriction enzyme HaeIII. A P^{32} labelled 5S repeat probe was used.

4.2.7 Generating and characterising RDR2:FLAG and RDR6:FLAG transgenic Arabidopsis lines.

4.2.7.1 Using transient expression in N. benthamiana to test the functionality of the FLAG-tagged constructs

Both the RDR2-FLAG and RDR6-FLAG constructs were introduced into *N. benthamiana* leaves via *A. tumefaciens* infiltration, as the GFP tagged constructs had been. A western blot was performed using an anti-FLAG antibody to test protein production/ tag detection. The anti-FLAG antibody used cross-reacted with a *N. benthamiana* protein and therefore produced a band even in control samples (data not shown). However, when RDR2:FLAG and RDR6:FLAG infiltrated *N. benthamiana* were tested no additional bands were detected, and both samples looked the same as the control (data not shown). Expression of proteins by infiltration of *N. benthamiana* does not always result in detectable levels of the protein or the epitope tag, even for constructs successful in Arabidopsis. Additionally, these constructs contained the native promoters which are likely to affect protein expression in *N. benthamiana* leaves. Furthermore, only one anti-FLAG antibody was used to perform the western blots and it is possible that an alternative anti-FLAG antibody may have given a different result. Further work on these fusion proteins in Arabidopsis, which will be described later, demonstrated that although the antibody used here does work, an Horseradish peroxidase (HRP)-conjugated anti-FLAG anti-body works more effectively.

4.2.7.2 Transformation of the FLAG-tagged genes into Arabidopsis mutants resulted in phenotype recovery

The SALK_059661 *rdr2* and the *rdr6-11* mutants were transformed with the RDR2:FLAG and the RDR6:FLAG binary vectors respectively, using the floral dip technique (Bent & Clough, 1998, Bent et al., 2000). Selection with the herbicide BASTA on soil was performed on the F1 and F2 generation to find plants containing the vector. Phenotype rescue was then looked for in the resulting lines.

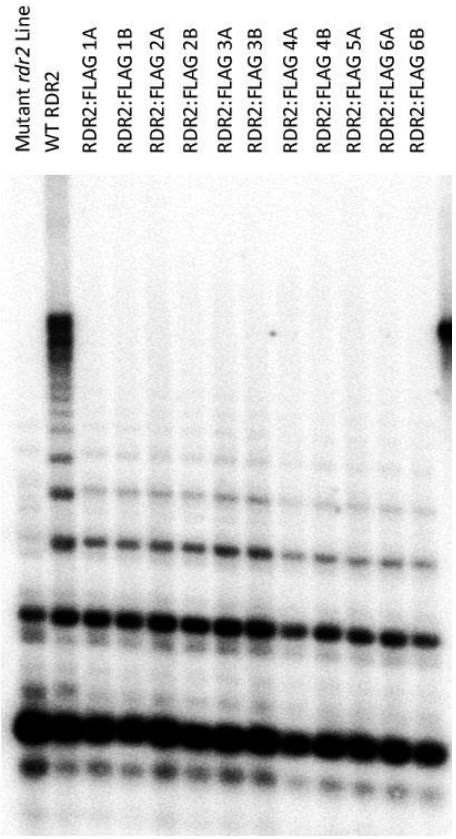


Figure 4.6. The introduction of RDR2:FLAG to *rdr2* plants results in partial restoration of wild type 5S rDNA DNA methylation levels as detected by Southern blot.

Genomic DNA extracted from leaves of F2 individual plants which had been BASTA selected in both generations. Two plants had been taken from each line which stems from individual F1 plants (labelled as A & B). Samples had been digested with HaeIII. P^{32} labelled 5S repeat probe used.

The introduction of the RDR2:FLAG constructs to the *rdr2* plants partially recovers the wild type phenotype (Figure 4.6). DNA samples extracted from the leaves of F2 plants were digested with HaeIII and the methylation status of the 5S rDNA repeats analysed by Southern blotting as described above. These are F2 plants and so it is not known if they are homozygous or not for the *RDR2:FLAG* gene. Figure 4.6 shows that there were more higher molecular weight bands in the RDR2:FLAG samples than in the *rdr2* mutant, which has impaired methylation. However, although

the RDR2:FLAG lines did have higher molecular weight bands, they did not have the same level as the wild type. Therefore the addition of RDR2:FLAG only partially rescued the 5S rDNA methylation phenotype. This highlights the possibility that the FLAG epitope may affect the protein's function to some extent, although it was not possible to assess by how much using this blot alone.

The addition of the RDR6:FLAG construct or the RDR6:GFP construct to the *rdr6-11* plants recovers the wild type visual phenotype. The visual mutant phenotype is known as Zippy and is the result of a shortened juvenile phase resulting in elongated, curled leafs, with an elevated number of trichomes when compared to the WT (Figure 4.7B)(Peragine et al., 2004). This visual phenotype is therefore easy to assess. T2 plants were grown on soil without selection to see if the *rdr6* mutant visual phenotype was complemented by the addition of the RDR6:FLAG or RDR6:GFP construct. It was not possible to assess phenotype rescue in the T1 generation as treatment with the BASTA herbicide during the selection process interferes with visualizing the *rdr6-11* phenotype.

Figure 4.7. Phenotype rescue of the RDR6 wild type phenotype was seen in T2 progeny from transformations of the *rdr6-11* mutant with the RDR6:FLAG construct.

Here a representation of the rescue is shown with one plant for each of the lines. **Panel A.** A Col-0 wild type plant demonstrating the WT phenotype. **Panel B**. A *rdr6-11* plant displaying the Zippy phenotype of elongated, curled leafs, with an elevated number of trichomes. **Panel C**. A T2 RDR6-FLAG plant which appears with a WT phenotype. **Panel D**. A T2 RDR6-GFP plant which also has a WT phenotype. Not to scale. Picture taken ~3 weeks post germination

Although the work on the RDR6:GFP lines was not pursued due to the unsuccessful confocal microscopy analysis, phenotype rescue was observed in these lines. 10 lines of T2 RDR6:GFP were

sown out with 10 plants for each line. In all lines complementation of the *rdr6* mutant phenotype was observed, although not in all plants (Figure 4.7A & D). The lack of complementation in all plants is to be expected as the transgene will still be segregating in this generation and the plants were grown in the absence of selection.

For the RDR6:FLAG lines, 12 lines of T2 were sown out on soil with no BASTA selection, 10 plants for each line. Again in all lines the wild type phenotype was restored, but not all plants in each line had the wild type phenotype due to the expected transgene segregation and the lack of selection (Figure 4.7). Figure 4.8 shows a repeat of the T2 complementation experiment with one of the RDR6:FLAG lines and the *rdr6-11* mutant at three different time points to show the stability of the phenotypes. The plants in the top rows of the RDR6:FLAG tray show a more wild type phenotype than the *rdr6-11* mutant plants, whereas the two plants in the bottom two wells of the RDR6:FLAG tray appear to still have the mutant phenotype (Outlined in blue box RDR6:FLAG sample on Figure 4.8). There was a wild type plant in the *rdr6-11* sample, which may be the result of seed contamination (outlined in purple on Figure 4.8). All of the *rdr6-11* plants had been visually assessed for the presence of the zippy phenotype before undergoing the floral dip with any of the constructs used. However, to make sure the RDR6:FLAG plants with a "wild type" phenotype were not the result of contamination, PCR analysis was carried out on a selection of lines (Figure 4.9 & Figure 4.10).

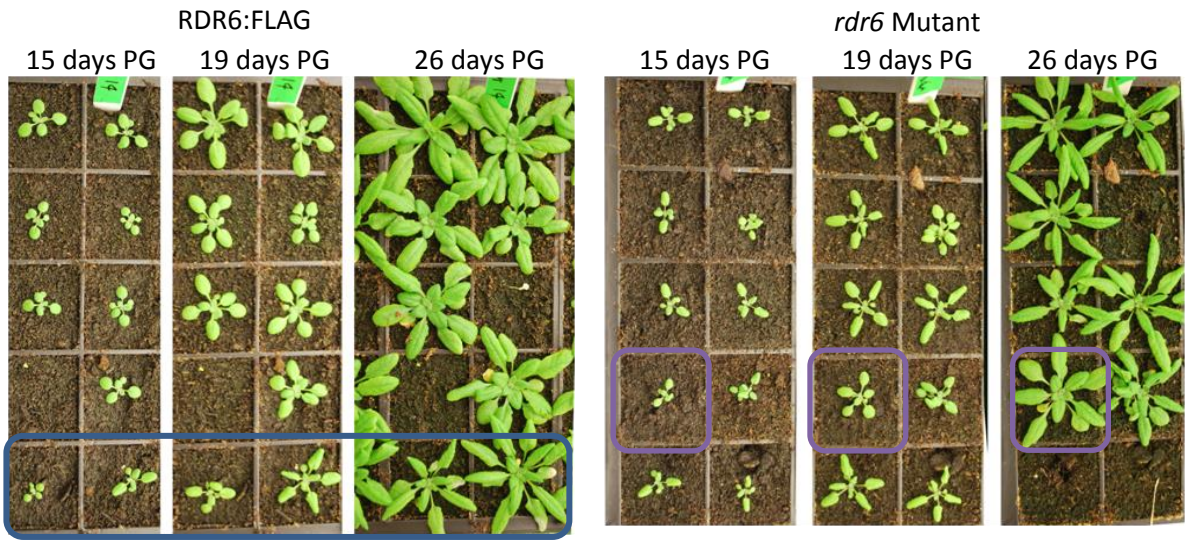


Figure 4.8 Complementation of the *rdr6* phenotype.

Images of an RDR6:FLAG line and the *rdr6* mutant taken at 15, 19 and 26 days post germination. Plants had been removed from the *rdr6* mutant tray between the 19 day and 26 day images to be used for tissue. The blue box highlights RDR6:FLAG plants that have maintained the zippy phenotype. The purple boxes highlight plants with WT phenotype in the *rdr6-11* samples

Two forms of PCR analysis were performed on genomic DNA extracts from the T2 RDR6:FLAG plants from the phenotype study. One PCR analysis was used to confirm the presence of the transgene and the other was to genotype the *rdr-11* mutant background in the transformed lines. Plants of six individual lines were tested (numbered 5, 6, 7, 8, 9 & 10) and included plants which had the Zippy phenotype (Zip) or a wild type phenotype (WT). The results for the first PCR are shown in Figure 4.9. In this PCR the primers were designed to generate a PCR product which spanned the junction between *RDR6* and *FLAG* in the vector to demonstrate that the complementation is associated with the presence of the *RDR6:FLAG* transgene. The transgene was present in plants referred to as 6 WT, 7 WT, 8 WT and 10 WT demonstrating that the gene was associated with the complementation in some instances. The transgene was also present in a plant from line 6 which had a Zippy phenotype suggesting that the transgene could be present but not be expressed. Plant 9 WT did not appear to have the transgene present (Figure 4.9). This could either be the result of contamination in line 9 or a failed PCR reaction for this sample. Only one RDR6:FLAG line was required for further work therefore analysis of this line was not continued with.

Figure 4.9. The RDR6:FLAG transgene is present in T2 plants.

Primers used produced a PCR fragment that spanned the DNA sequence between the RDR6 gene and the FLAG tag. The plasmid positive control was the one introduced into the *A. tumefaciens* used for the floral dip. Size given in base pairs.

Figure 4.10. Genotyping of T2 plants reveals that the epitope tagged construct is present in plants with the *rdr6-11* mutant background

A PCR fragment of the *rdr6-11* gene was digested with TaqI and the produces separated on a 1% agarose gel. The DNA used for the *rdr6* mutant, Col-0 and RDR6:GFP 1 & 2 samples had been pooled from two plants. The other samples are all from individual plants. Size scale is given in base pairs. A gel was run for all samples of the PCR fragment before digestion which showed a single band for each sample of uniform size (data not shown).

The second PCR performed demonstrated that complementation had occurred in the *rdr6-11* mutant background for plants 6 WT, 7 WT and 8 WT (Figure 4.10). For this analysis a fragment of *rdr6-11* was amplified which contains a point mutation that affects the size of fragments produced by digestion with the TaqI restriction enzyme, as demonstrated by the *rdr6* mutant and Col-0 samples in Figure 4.10. Plants 6 WT, 7 WT and 8 WT lines had a banding pattern consistent with the presence of both the *rdr6-11* mutant (top and bottom bands) and the RDR6 gene (middle and bottom bands). Again there were plants with the Zippy phenotype that had the DNA for both the *rdr6-11* mutation and the RDR6:FLAG transgene (5 Zip) and there were WT plants that seemed to be the result of contamination (5 WT).

Four of the T2 RDR6:GFP plants from two lines also underwent this PCR genotyping analysis (Figure 4.10). At this stage in characterization it was clear that the GFP lines would not be continued with and so the PCR was performed to confirm that some of the lines tested did have the RDR6 gene in an *rdr6-11* background. Therefore, as an individual line was not required to take forward, the tissue for the DNA extraction was pooled into two samples (one per line) with two plants in each. One of the samples gave a banding pattern consistent with the presence of the RDR6:GFP construct (Figure 4.10). Therefore the *rdr6-11* plants treated with the RDR6:GFP vectors were transformed but this did not result in visible GFP production.

4.2.7.3 Detection of the FLAG fusion proteins by western blot

The western blot technique was used to assess if the FLAG tagged fusion proteins were detectable. There had been problems in finding a positive FLAG control in Arabidopsis but a positive control was acquired from the laboratory of Dr Betsy Pownall (University of York) in the form of a ROR2:FLAG fusion protein expressed in *Xenopus laevis*.

The RDR6:FLAG lines from the PCR analysis in Figure 4.10 were analysed using western blot. RDR6 has a molecular weight of 137 kDa. A band with molecular weight over 100 kDa was detected in plants 6 WT, 7 WT and 8 WT (Figure 4.11). No equivalent band was observed in any of the other samples. This indicates that plants 6 WT, 7 WT and 8 WT produce detectable levels of RDR6:FLAG protein. The PCR analysis showed that these plants have the *RDR6:FLAG* gene in the *rdr6-11* mutant background. The results of the phenotype rescue, the western blot and PCR data together confirm that three lines had been produced in an *rdr6* mutant background that had detectable levels of the RDR6:FLAG fusion protein.

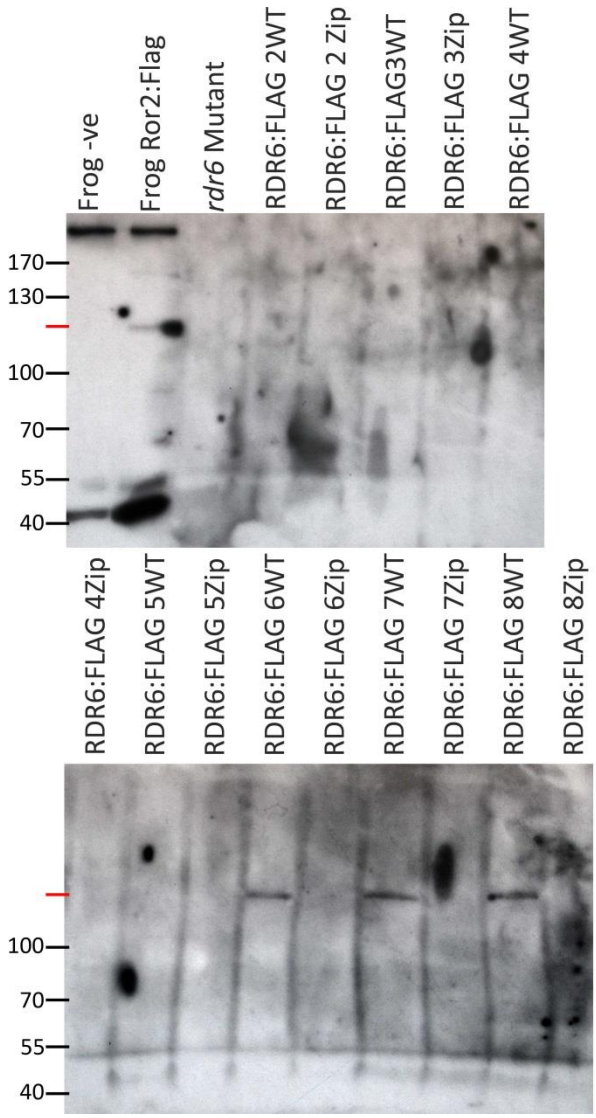


Figure 4.11. The RDR6:FLAG fusion protein is detectable by western blot in some RDR6:FLAG lines using an anti-FLAG primary antibody.

Protein extracts from rosette leaves of T2 plants of lines 2- 8 were analysed by western blotting using the Sigma anti-FLAG polyclonal antibody (F7425). The secondary antibody used was the HRP-conjugated goat anti-rabbit from Abcam. WT refers to plants displaying the Wild type phenotype and ZIP refers to plants showing the *rdr6* mutant "Zippy" phenotype. Plants 6 WT, 7 WT and 8 WT give a positive signal as indicated by the red line. Plant tissue is from rosette leaves. The Frog Positive control was 5 *Xenopus laevis* embryos injected with 1 ng of Ror2:FLAG mRNA at the 2 cell stage. Negative is 5 un-injected embryos. Both were kindly donated by the Pownall group. Protein size ladder given in kDa. The 130 and 170 kDa protein weight marks did not transfer in the blot in the lower panel.

Western blots using the *Xenopus laevis* positive control were performed using primary and secondary antibodies as in Figure 4.11 for a selection of RDR2:FLAG lines to find lines with detectable levels of the fusion protein (data not shown). One line was found that gave detectable levels of the RDR2:FLAG fusion protein. This was line 4 from Figure 4.6, which had shown partial

restoration of the wild type 5S rDNA DNA methylation phenotype. There had been issues with the quality of the western blots with regards to the signal strength using the primary polyclonal and secondary antibodies initially available. These issues were improved when a direct conjugated monoclonal anti-FLAG antibody was used with an adjusted transfer protocol. The changes resulted in higher quality western blots with clearer bands. This may have been due to the increased specificity given by monoclonal antibodies. Figure 4.12 shows that the RDR2:FLAG fusion protein is detectable in floral tissue. RDR2 has a molecular weight of 129 kDa (Figure 4.12). The RDR6:FLAG line in Figure 4.12 is from line 6 WT from before.

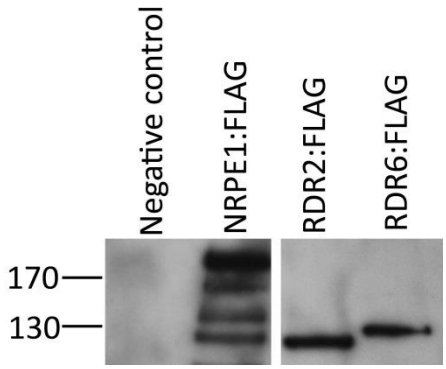


Figure 4.12. The FLAG tag is detectable by Western Blot in NRPE1:FLAG, RDR2:FLAG and RDR6:FLAG plants.

The expected size of the NRPE1 subunit is approximately 218 kDa, 129 kDa for RDR2 and 137 kDa RDR6. RDR2:FLAG is line 4 from the 5S rDNA work and the RDR6:FLAG line is line 6 WT from the previous RDR6:FLAG western blot. The NRPE1:FLAG line is from Dominique Pontier. Proteins extracted from floral tissue. Scale given in kDa. Monoclonal anti-FLAG-HRP from Sigma A8592

4.2.8 Characterising NRPE1: FLAG lines

The largest subunit of Pol V is NRPE1. A FLAG-tagged NRPE1 line was generously donated by Dominique Pontier. The NRPE1:FLAG fusion protein was detectable by western blot providing me with a useful tool for further analysis (Figure 4.12). The NRPE1 subunit is 218 kDa in size and the strongest band on the western blot is over 170 kDa in size (TAIR). However, there were many other bands present on this western blot. These bands are not present in the RDR2:FLAG or RDR6:FLAG samples or non-transgenic negative control indicating that they are not the result of non-specific binding to other Arabidopsis proteins. It is therefore possible that these extra bands may be the result of protein breakdown products of the *NRPE1* gene which could results in different size proteins with a detectable FLAG epitope.

It should be noted that before the NRPE1:FLAG line was acquired an attempt was made to generate a peptide antibody for NRPE1. The NRPE1 peptide antibody was ordered from Genosphere Biotechnologies (www.genosphere-biotech.com). The peptide used for immunization was EEESTSEILDGEIVC which is from the N-terminal region of the NRPE1 protein. However, when western blots on wild type Col-0 flowers were attempted using this antibody the western blot failed (data not shown). This was using the western blot transfer protocol that proved to be somewhat inefficient. By the time this was realised the NRPE1:FLAG line have been acquired and so this was not repeated.

4.3 Discussion

4.3.1 AMP243 triggered silencing is affected by both *rdr6* and *rdr2* mutants

Work here has demonstrated that AMP243 triggered silencing is RDR6-dependent, as expected from the literature (Dalmay et al., 2000a, Dalmay et al., 2001). Previous work investigated the results of crossing only one homozygous AMP243GFP line with an *rdr6* mutant (AMP243 X GFP142). In this body of work the effect of *rdr6-11* on GFP silencing by AMP243 has been observed in three independent lines, expanding the numbers of lines tested. However, as AMP243 triggered silencing was shown not to be affected by genomic location in Chapter 3, the number of homozygous AMP243 GFP crossed with *rdr6-11* was not expanded further as the only variable would have been the location of the *GFP* transgene.

Interestingly *rdr*2 mutants cause a delay in AMP243 triggered silencing suggesting that RDR2 has a role in the RDR6-dependent PTGS triggered here. The reasons for this are unclear. RDR2 is required for canonical-RdDM and the effect of the *rdr2* mutant may be related to RdDM. However, the methylation status of the lines when crossed to either mutant is unknown and therefore all discussion of this is speculative. Here the crosses to investigate the requirement for RDR6 and RDR2 were between the mutants and plants that were homozygous for both the *35S:GFP* transgene and the AMP243 trigger. Work with the DAN2 recreation cross in Chapter 3 suggests that there may be a difference in RdDM between generations, perhaps due to increased zygosity of the transgene (de Carvalho et al., 1992, De Wilde et al., 2001, James et al., 2002, Velten et al., 2012). It is therefore possible that all of the homozygous AMPGFP parents underwent RdDM of the *GFP* transgene before crossing to the mutant lines. If there is a generational effect, opposed to one of zygosity, then these lines may also be undergoing RdDM. Alternatively, if zygosity is a factor then lines that underwent RdDM in the hemizygous F1 AMP243 cross may still have the potential to undergo RdDM in these crosses, whereas lines that were not part of the F1 methylating set may not. Of the lines crossed with mutants here, lines 65

and 103 underwent RdDM in the F1 AMP243 cross preformed in Chapter 3 and lines 58, 60 & 82 did not. Delayed silencing was observed for all lines crossed with *rdr2*.

However, the role of RDR2 in AMP-triggered RDR6-dependent PTGS may not be related to the methylation of the transgene. Jauvion et al. (2012) found that RDR6 and not RDR2 is required for the methylation in PTGS of a transgene. The Jauvion et al. (2012) paper used a GUS transgene under the control of the 35S promoter and the L1 trigger of silencing to explore the effects of mutants on PTGS of the GUS transgene. They reported that there was lower GUS-specific siRNA accumulation in *rdr2* mutants which were wild type for RDR6. However, using a weaker trigger of GUS silencing, they found that *rdr2* mutants increased PTGS efficiency in plants wild type for RDR6. They therefore concluded that RDR6 and RDR2 have an antagonistic relationship and that RDR2-produced siRNAs counteract RDR6 during the initially stages of PTGS. The observation here that *rdr2* mutation caused a delay in silencing may suggest RDR2 acts in the initial stages of silencing to increase PTGS, which is at odds with the conclusions of Jauvion et al. (2012).

4.3.2 The generation of epitope tagged constructs and potential future uses

Here the proteins RDR2 and RDR6 have been successfully fused to the FLAG epitope and expressed in Arabidopsis. This has been confirmed by Western Blot and rescue of the wild type phenotype in the respective mutant background. Additionally, a NRPE1:FLAG line was acquired and successfully used in a Western blot. These FLAG-tagged lines can be used in future immunoprecipitation experiments to look at issues surrounding their recruitment during RDR6-dependent AMP243 triggered PTGS and RdDM.

To use the FLAG-tagged lines in immunoprecipitation experiments they will be crossed with the homozygous AMPGFP lines developed in Chapter 3. This will allow associations between the RDR6, RDR2 & Pol V and the *GFP* transgene loci to be investigated. This will be discussed further in Chapter 5. The duel transgene system used to create the AMP243 crosses has the added advantage that the recruitment of proteins before and after silencing can be assessed. The FLAG-tagged lines could be crossed to the parent 35S:GFP line in the absence of the AMP243 trigger and used in immunoprecipitation experiments. The results could be compared with the investigation of the homozygous lines.

The FLAG-tagged lines could be used to investigate possible reasons for the variation in RdDM observed in Chapter 3. For example, Pol V is known to bind directly to its targets, but it is not clear when this recruitment occurs. It is possible that such a protein could be recruited to its targets before the triggering of silencing, perhaps by the presence of specific DNA methylation or histone marks or by a particular transgene structure.

The FLAG tagged lines could also potentially be used to study protein-protein and RNA-protein interactions. Protein complex immunoprecipitation experiments using FLAG-tagged NRPD1a and NRPE1 have been conducted to great effect by other groups to delineate the protein complexes and associations they form in RdDM (Huang et al., 2009, Law et al., 2011, Ream et al., 2009, Wierzbicki et al., 2008). Haag et al. (2012) have performed co-immunoprecipition reactions and report that NRPD1 of Pol IV and RDR2 precipitate together. Prior to the start of this work little was understood about RDR6's interactions with the target loci.

RNA immunoprecipitation (RIP) could also be conducted with the FLAG tagged lines described here. This would allow the interaction between the key proteins chosen here and any GFP RNAs they associate with to be investigated.

Unfortunately not all of the planned constructs were generated during this project. Work on an NRPD1a fusion line was stopped as there was difficulty in cloning this into the pENTR vector and then recombining this with a destination vector. This meant that no investigation in to Pol IV was conducted as planned.

This project also attempted to generate GFP fusion proteins of RDR2 and RDR6. However, although the RDR6:GFP fusion protein was transiently expressed in *N. benthamiana* successfully, GFP fluorescence was not detected in Arabidopsis for either RDR6 or RDR2. For reasons previously described this was not the main thrust of this PhD project and so it was not considered necessary to invest large amounts of time into troubleshooting this issue. The reading frames of both constructs were checked and both have been maintained (data not shown). There are a few possible reasons why GFP fluorescence was not observed in Arabidopsis – even for the RDR6:GFP which worked well in *N. benthamiana*. The GFP tagged genes are under the expression of the 35S promoter which should cause overexpression. This may be detrimental to the tagged gene as overexpression may cause deleterious effects and lead to selecting plants in which the transgene is silenced. It is also possible that the GFP tag may interfere with the stability of the protein. For the RDR6:GFP lines there may be some tertiary folding issues with the protein that result in functioning RDR6, resulting in complementation, but to not produce fluorescent GFP. Without conducting further investigation into these issues no firm conclusions can be drawn. Further PCR analysis would have been informative to ascertain if the whole *RDR:GFP* transgenes were present in Arabidopsis. Northern blots could also have been conducted to investigate if GFP mRNA is produced to but not expressed which may indicate a protein folding issue.

There were problems in this project with the western blot protocol used. This was resolved to be an issue with the transfer buffer which contained SDS being used; this was a surprise as the buffer recipe used was the one recommended by many manufactures of antibodies for the size of proteins investigated here. This slowed progress on this chapter down considerably and may have affected the outcome on the work with the peptide antibody for NRPE1. However, the protocol did work somewhat, and the lines that expressed RDR6:FLAG and RDR2:FLAG were found using this technique. Therefore in inability of the first Western blot protocol to detect the peptide antibody suggests that the results may have not been improved with the modified transfer protocol.

4.4 Acknowledgments

Dominique Pontier donated the FLAG tagged NRPE1 line and offered some technical advice for the western blot procedure.

The Pownall group (University of York) provided a Flag positive control in the form of a ROR2:FLAG fusion protein expressed in *Xenopus laevis*.

Chapter 5 RdDM may be influenced by transgene structure and copy number

5.1 Introduction

Chapter 3 described the results of crossing multiple independent 35S:GFP lines with the trigger of RNA silencing AMP243. AMP243 had triggered silencing in all 52 lines however, out of the 40 crosses analysed only 8 lines displayed evidence of DNA methylation in the F1 cross. The possibility that the genomic location of the T-DNA insert may affect the ability of the *GFP* transgene to support DNA methylation was investigated. Using TAIL PCR, the T-DNA integration sites were located within the Arabidopsis genome. However, no evidence could be found to support the hypothesis that genomic location influenced AMP243-triggered methylation. There are other possible reasons why these 8 lines clearly supported DNA methylation when the others did not; this will be the focus of this chapter.

Multiple independent 35S:GFP lines that are likely to have a single integration site were obtained by screening for lines which displayed a 3:1 ratio of GFP expression: no-expression. However, this method of selection does not provide any information on the copy number of the T-DNA at the site of integration. For example, a line may contain one copy of the T-DNA but another may contain multiple copies at a single site. Studies in a variety of plant species have shown that transgenes with multiple copies tend to undergo RNA silencing (Hobbs et al., 1993, Schubert et al., 2004) and may also undergo DNA methylation, especially when the integration structures are complex (reviewed by Butaye et al., 2005, Stam et al., 1998). The term complex integration structure can include direct or inverted repeats for example, as well as integration of large expanses of the vector backbone. It is therefore possible that when silencing of a transgene is triggered by AMP243, lines that contain more than one copy of the transgene in a complex arrangement may be more likely to also undergo DNA methylation compared to single copy number lines. Therefore, the copy number and integration complexity of the GFP T-DNA was assessed to look for differences between lines that clearly undergo DNA methylation in the F1 crosses and those that do not.

This chapter will also investigate the possibility that the zygosity of the transgene may alter its ability to support methylation. It has been observed that transgenes are more likely to undergo silencing when in a homozygous state compared to the hemizygous state (de Carvalho et al., 1992, De Wilde et al., 2001, James et al., 2002, Velten et al., 2012). Work on the DAN2 recreation cross in chapter 3 highlighted this as a possible line of enquiry when methylation was not seen in

the F1 AMP243 X GFP142 cross despite being present in the DAN2 homozygous line (Dalmay et al., 2000a).

One of the key questions in the field of RNA silencing and RdDM is how do the different components become recruited to their targets. These questions were discussed in the introduction to chapter 4. Was the reason for the difference in ability to undergo DNA methylation when crossed with AMP243 the result of differential recruitment of RNA silencing proteins which can lead to DNA methylation? One way to address this question is to investigate recruitment of silencing components to the different transgene loci in the presence and absence of the silencing trigger. It is possible that some RdDM components already associate with loci before silencing occurs. These could then facilitate the recruitment of further proteins needed for DNA methylation once the production of siRNAs had been triggered. This could occur if RdDM components recognize features of some loci and not others. A key component of the RdDM machinery is the NRPE1 subunit of Pol V as this is reported to associate with the genomic loci of its targets (Wierzbicki et al., 2008, Zhong et al., 2012). However, there are still uncertainties about how it is recruited to targets. Does Pol V associate with some of the *35S:GFP* loci before the addition of the AMP243 trigger and is it this difference that causes only a proportion of the lines to undergo DNA methylation in the F1? Alternatively, Pol V may only be recruited once silencing has been triggered. One way to address these questions would be to perform chromatin immunoprecipitation (ChIP) experiments on the progeny of crosses between the NRPE1:FLAG line described in chapter 4 and a selection of parent 35S:GFP lines. Firstly, this would answer the question as to whether or not Pol V occupies loci in the absence of a silencing trigger. Secondly, screening parents from lines that underwent DNA methylation in the F1 AMP243 cross and ones that did not would allow for differences in Pol V occupation to be assessed. This may also provide an explanation as to why there is a difference in the DNA methylation triggered by AMP243.

In recent years RdDM has been associated with specific histone modifications. For example methylated Histone 3 K9 (H3K9me) is an important repressive modification in the plant epigenome. It has a strong association with CMT3-based maintenance of CHG DNA methylation (Du et al., 2012, Law & Jacobsen, 2010, Meyer, 2011). Additionally, H3K9me2 is a main driver of transgene silencing and heterochromatin formation in *S. pombe* with its addition to histones being mediated by the RNAi machinery (reviewed by Goto & Nakayama, 2012). This mark is also associated with siRNA based silencing in plants (Bernatavichute et al., 2008). The protein SHH1, which interacts with Pol IV and is required for RdDM at some targets, has been found to predominantly bind to methylated H3K9 (Law et al., 2010, Law et al., 2013). However, it is unclear if the connection between histone modification and RdDM is a cause of, or a result of, the

recruitment of RNA silencing machinery to a target. Further to the addition of repressive histone modifications, RNA silencing can also be associated with loss of active histone modifications. For example, deacetylation has been shown to be involved with gene silencing in some instances through the action of Arabidopsis histone deacetylase HDA6 (Blevins et al., 2014). Histone acetylation can work in opposition to methylation in some instances. For example, when H3K14 is acetylated H3K9 tends not to be methylated and vice versa (Johnson et al., 2004). It would therefore be interesting to investigate if there are observable difference in covalent histone modifications between the parent and AMP243 cross GFP lines and whether this correlates with the ability to support DNA methylation. For example, H3K9me2 is linked to SHH1 binding which may aid Pol V recruitment. Here ChIP will be used to investigate differences in histone modifications at the GFP loci between the 35S:GFP lines. Two modifications to histone H3 will be investigated. H3K9me2 because it is a repressive mark connected to RNA silencing and acetylation of H3 because these are active marks that may work in opposition to H3 methylation in some instances.

5.2 Results

5.2.1 Analysis of transgene copy number in the multiple independent 35S:GFP lines

In chapter 3 it was described how the multiple independent 35S:GFP lines were selected which were likely to have a single integration site as indicated by the 3:1 ratio of GFP expression: no-expression. However, this method of selection does not provide any information on the copy number of the T-DNA at the site of integration. In the general introduction the possibility of locus specific effects was explained as a term that could apply to both the genomic site of T-DNA integration, as investigated in chapter 3, or could also refer to the structure of the insertion site. Therefore, the effects of copy number and integration complexity of the GFP T-DNA on the ability to support methylation were investigated. To estimate the relative copy number of the *GFP* transgene in the independent 35S:GFP lines SYBR green qPCR was performed on the parents of the F1 methylated lines and a selection of other parent lines. 33 parent lines were tested in total (Figure 5.1).

To quantify the results the comparative $2^{-\Delta\Delta ct}$ method was used (reviewed by Bubner & Baldwin, 2004, Livak & Schmittgen, 2001). This method normalises the Ct data from the transgene, in this case *GFP*, to an endogenous control. Here the *MALATE SYNTHASE* (*MLS*) gene was used as the endogenous control and the Ct values were calculated by the MyiQ Real-Time PCR detection system software (Bio-Rad). The data was then standardized to a calibrator sample as a ratio

(Livak & Schmittgen, 2001). Ideally the calibrator would be a line with a known copy number that everything else could be measured against, however such a line was not available. The investigation into integration complexity, which will be described in section 5.2.2, suggests that line number 10 has a copy number of one and therefore this line was used as the calibrator, automatically giving it a relative copy number of one. The parents were all homozygous. Therefore a line with a copy number of one would in fact have two copies of the *GFP* gene due to the zygosity.

Of the 33 lines tested 16 lines had a relative copy number of 1.5 or less which has been taken to be indicative of a copy number of one. Ten lines had a relative copy number of between 1.5 and 2.5 indicating a copy number of 2. Six lines had a relative copy number between 2.5 and 3.5 indicating a copy number of 3. One line had a relative copy number of 4. However, this assumes that the 35S:GFP lines tested did have a single site of integration and did not have second non-transcribed transgenes. This will be discussed in detail later.

Lines that clearly underwent methylation the F1 AMP243 cross, as denoted by an asterisk, tended to have a high relative copy number of 2 or 3 (Figure 5.1). However, line 71 was an exception to this as it was the only F1 methylating line with a copy number of one. Many lines that did not appear to methylate in the F1 also had a high copy number such as line 120 that has a copy number of 4 yet is unmethylated in the F1 cross. 14 of the 25 non-methylating lines had a relative copy number of one, six had a relative copy number of two, and two had a relative copy number of three.

The copy number analysis did not give the clear discreet values which may be expected from qPCR analysis. Here the results have been rounded to estimate relative copy numbers which may be inaccurate. Investigating the integration structure will add some clarity to this data.

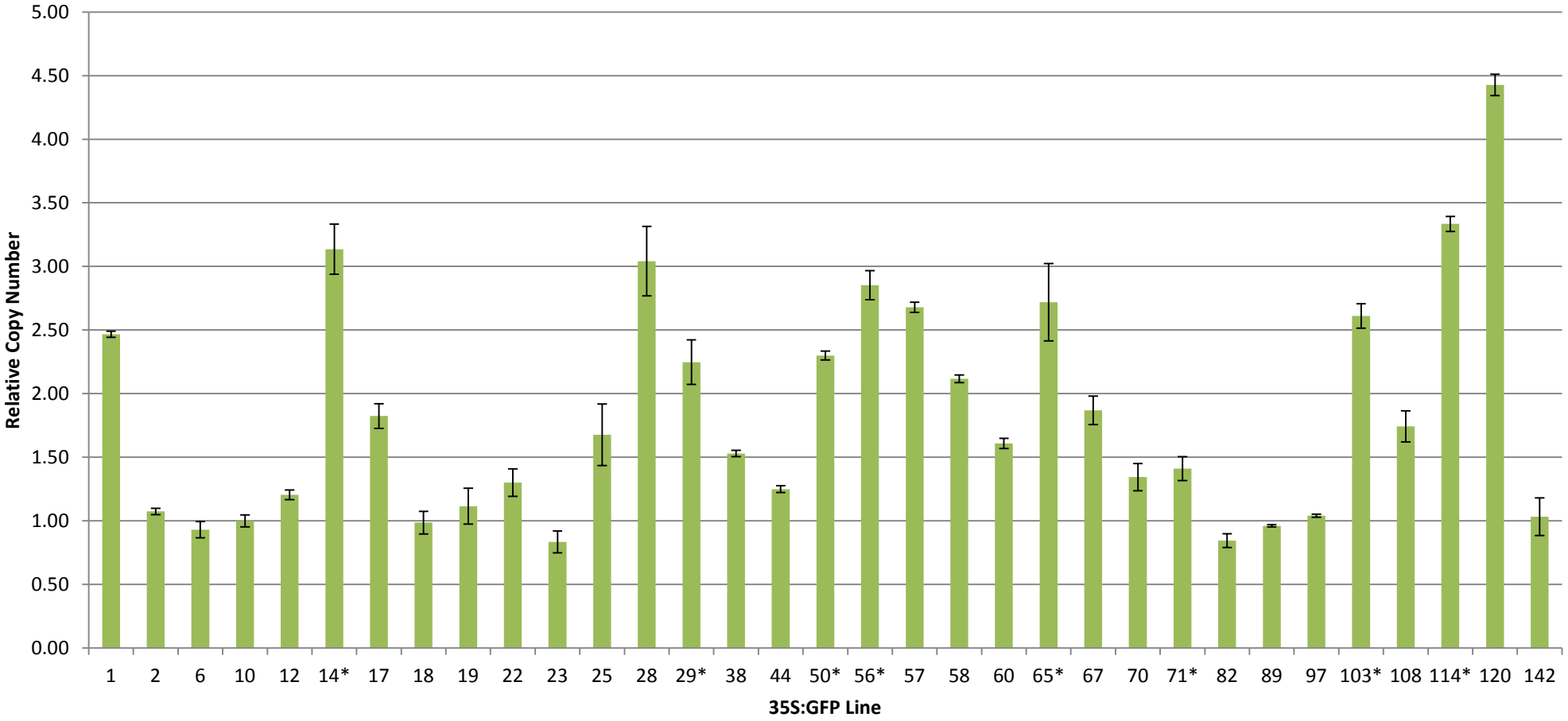


Figure 5.1. Copy number may affect the ability of F1 crosses to undergo methylation

Relative GFP copy number was calculated using the comparative $2^{-\Delta\Delta ct}$ method. MLS was used as the endogenous control. DNA was extracted from pooled homozygous parent plants. 3 technical repeats per sample were performed. Standard error is given. Lines that methylated in F1 cross marked with *

5.2.2 Analysis of integration complexity in the multiple independent 35S:GFP lines

The integration structure of the T-DNA inserts was investigated to see if the complexity of the structure correlated with methylation. Transgenes that insert with a multiple copy numbers in the same location tend to undergo silencing more readily than lines that have a single copy number, especially if the repeats are organized in an inverted configuration (Skarn et al., 2006, Wang & Waterhouse, 2000). To assess the transgene integration structure, a Southern blot was performed using a unique BamHI restriction site located at one end of the *GFP* transgene that does not occur within the rest of the T-DNA (Figure 5.2). DNA samples from a selection of the parent lines were digested with BamHI and subjected to Southern blot analysis. The lines were chosen to include lines that were able to support methylation in the F1 AMP243 cross and lines that did not appear to. The blot was initially probed with a GFP-specific probe before being stripped and then re-probed with the Kanamycin resistance gene (*NPTII*). Figure 5.2 shows examples of possible integration structures and how these would appear on the blot. For example, if the T-DNA had integrated as a single copy, single integration event then there would be one band on the Southern blot for both GFP and Kanamycin resistance probes, although these should be of different sizes (Figure 5.2B). If two copies of the T-DNA had integrated at a single locus the banding pattern would depend on the orientation of the T-DNA copies relative to each other; examples of this are given in Figure 5.2C, D & E. For example, if the copies had inserted in the same orientation, as a direct repeat, then there would be two bands for each probe (Figure 5.2E). If the T-DNA had inserted in opposite orientations, as an inverted repeat, then there would be one band for one probe and two bands for the other depending on which gene was in the middle of the complex (Figure 5.2C & D). For additional copies additional bands will be present. It should be noted that the band sizes could not be predicted as the distance to the next BamHI site in the genome is not known. Often vector sequence is included along with the T-DNA and this can affect the length of a fragment. Additionally, there may be bands on the Southern blot of the same size that cannot easily be distinguished, although these instances may be represented with bands of higher intensity. The BamHI Southern Blot results for both the GFP and Kanamycin genes are shown in Figure 5.3.

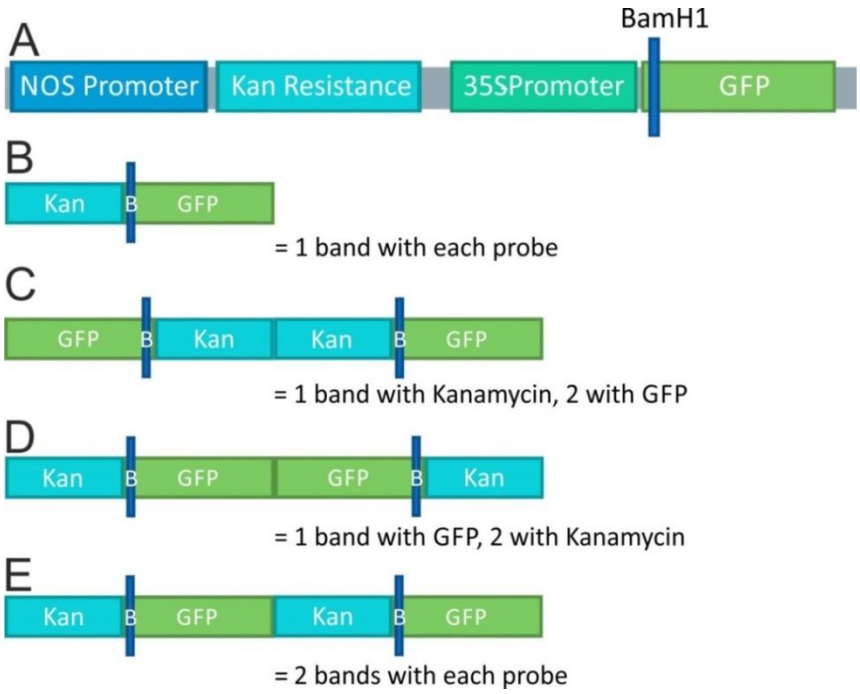


Figure 5.2. Schematic of the GFP T-DNA insert and possible outcomes of the BamHI Southern Blot

A Schematic of the GFP T-DNA insert. **B** Outcome of a single copy of the T-DNA. **C & D** Outcome of inverted repeats. **E** Outcome of direct repeats

The BamHI Southern blot suggests that a complex integration structure may be associated with the ability to undergo methylation in the F1 AMP243 crosses. Lines 14, 29, 50, 56, 65, 103 and 114 were all methylated in the F1 generation and have multiple bands present for either one of both of the probes used which is indicative of more than one insertion being present. These will be described in more detail later. However, this is not the case for line 71, which also methylated in the F1 cross. For this line there were single bands present for each probe which is indicative of a single insertion. Additionally, some lines that did not methylate in the F1 AMP243 cross also appeared to have complex integration structures, for example line 70.

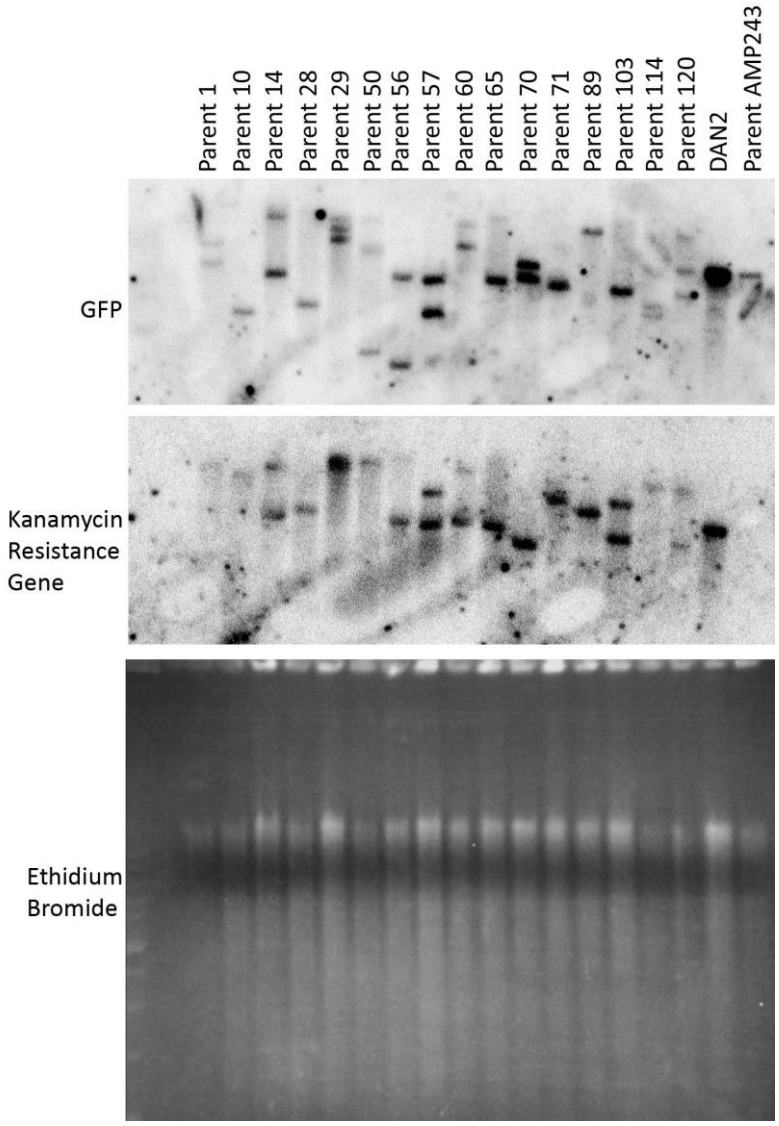


Figure 5.3. Analysis of transgene integration structure in a selection of 35S:GFP transgenic lines

Samples were genomic DNA digested with BamHI. **Top panel** results of a GFP probe. **Middle panel** membrane was stripped and then re-probed with a Kanamycin resistance gene probe. **Bottom panel** the ethidium bromide stained gel as a proxy for a loading control

The results of the BamHI Southern blot and the copy number qPCR have been collated in Table 5.1 with possible patterns of integration suggested. These suggestions are included to provide an illustration of the possible complexity and do not represent the definitive structure. As previously stated the distance to the next BamHI site in the genome is not known meaning that bands sizes could not be predicted. Additionally a line with multiple copies may have some of these copies hidden by producing fragments of the same size, which cannot be distinguished by gel electrophoresis. These could result in bands of high intensity but this can be a fairly subjective measure. Software is available that can compare the intensity of the bands but this was not implemented here. To do this the blot would need re-probing with a loading control. Line 28 may

represent a case of a bands on the gel masking multiple fragments as according to the qPCR data it has a copy number of 3, yet the Southern blot of this line has produced a banding pattern which would be consistent with a copy number of one (Table 5.1). Line 29 also may represent another case of masking as it has 3 faint bands for GFP and one intense band for the Kanamycin resistance gene and a copy number of 2 according to the qPCR.

The integration analysis has led to a few revisions to the data from the qPCR. The relative copy number for line 70 is 1.34 which was rounded down to one originally. However, the BamHI Southern blot suggests that this line has a copy number of 2. Additionally, the relative copy number for line 103 is 2.61, which had been rounded up to 3. Conversely, the BamHI Southern blot suggests that this line has a copy number of 2. This indicates that there may be some rounding errors in the data, as predicted earlier. It is also possible that a complex banding pattern and non-discreet qPCR value could be the result of a line having multiple loci insertion. This may not have been detected in the screen using a 3:1 ratio of expression vs no expression if the gene at one of the loci was not expressed or was silenced.

Table 5.1. Summary of the copy number and integration data

The grey rows indicate lines that underwent DNA methylation in the F1 AMP243 crosses

Line	Number of bands		Relative Copy number (qPCR)	Possible copy number	Possible structure of the T-DNA inserts based on copy number and integration data							
	GFP	Kanamycin										
1	2	1	2.47	2	G	K	K	G				
10	1	1	1	1	G	K						
14	2	2	3.14	3	G	K	K	G	G	K		
28	1	1	3.04	?								
29	3	1	2.25	>1								
50	3	1	2.30	>1								
56	2	2	2.85	3	G	K	K	G	G	K		
57	2	2	2.68	>1								
60	2	2	1.61	2	G	K	G	K				
65	2	1	2.72	3								
70	2	1	1.34	2	G	K	K	G				
71	1	1	1.41	1	G	K						
89	1	1	0.96	1	G	K						
103	1	2	2.61	2	K	G	G	K				
114	2+	1	3.33	3								
120	3	2	4.43	4	G	K	K	G	G	K	K	G
DAN2	1	1										
Amp243	1		N/A									

In agreement with the copy number analysis the BamHI Southern blot suggests that lines that underwent methylation in the F1 crosses were likely to have a higher copy number and a more complex integration structure than lines that did not display signs of F1 DNA methylation. However, this cannot be the only reason as some lines with a more complex integration structure did not undergo methylation in the F1, such as line 70 which has a copy number of two in an inverted repeat. Additionally line 71, which had a simple structure, did undergo methylation.

5.2.3 Analysis of DNA methylation in the F2 generation of the AMP243 X 35S:GFP crosses

The DAN2 line was used as a positive control for the Southern blot experiments described in chapter 3 as its *GFP* transgene has previously been reported to be methylated (Dalmay et al., 2000a). However, when the DAN2 line was recreated by crossing GFP142 and AMP243 together, no DNA methylation of the *GFP* gene was detected by Southern blot analysis in the F1 generation. There are no reports describing the methylation status of the *GFP* gene in the DAN2 F1 and the published work on DAN2 is based on homozygous lines of subsequent generations. This suggested that the seemingly unmethylated F1 crosses described in chapter 3 may show methylation in future generations and that zygosity may be a factor affecting a transgene's ability to support RdDM. Other groups have reported on transgene silencing events where lines are more likely to undergo methylation of a transgene when it is in the homozygous state but tend not to when the transgene is hemizygous (James et al., 2002, Velten et al., 2012, De Wilde et al., 2001).

To investigate this, the DAN2 recreation F1 cross along with 7 other AMP243 X 35S:GFP F1 lines, were allowed to self-fertilize and the resulting F2 plants were selected for resistance to Kanamycin and Phosphinothricin (PPT, the active ingredient in BASTA). Kanamycin is used to select for the *35S:GFP* transgene and PPT for the *PVX:GFP* amplicon transgene. Of the GFP lines analysed here 57, 58, 70, 82 & 89 were un-methylated in the F1 generation whereas 103 & 114 were methylated. DNA was extracted from approximately 40 pooled F2 seedlings and analysed by Southern blot as for the F1 generation (Figure 5.4). The methylation sensitive enzymes AluI and HaeIII were used to digest the DNA and the blot was probed with a P^{32}-labelled GFP-specific probe. DNA methylation of the *GFP* transgene was observed in all of the pooled F2 samples as evidenced by the higher molecular weight bands on the Southern blot (Figure 5.4). In Figure 5.4 the bands observed in a non-methylated sample, as seen in chapter 3, are highlighted with a red band. The DNA used for the blot was pooled from individuals with unknown zygosity; pooling the tissue meant that homozygous and hemizygous plants would be represented in the samples. However, DNA methylation was observed in all lines even those that had not been methylated in

the F1 (Figure 5.4). This suggests the possibility that the ability of the transgenes to support RdDM is influenced either by zygosity and/or the continued presence of the silencing trigger over multiple generations.

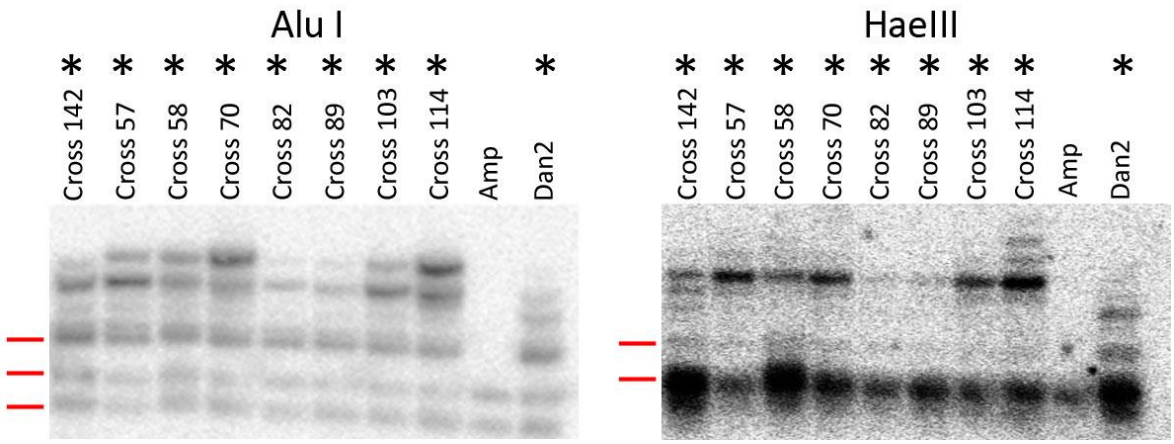


Figure 5.4. GFP DNA methylation is prevalent in the F2 AMP243 X 35S:GFP crosses.

Southern blot of pooled tissue from "F2" seedlings. F1 lines were self-fertilized and the seeds were selected on Kanamycin and PPT plates. DNA extracted and then digested with methylation sensitive enzymes. P^{32} tagged GFP probe used. Bands that are found in seemingly unmethylated samples are highlighted with a red line. Bands of a higher molecular weight, without the red line, are indicative of methylation. See chapter 3 for examples. The presence of methylation in a sample is denoted with an *. Only lines 103 and 114 had detectable levels of DNA methylation in the F1 AMP243 crosses. **Left panel** samples digested with AluI. **Right panel** samples digested with HaeIII.

To explore the effect of zygosity further, the homozygous AMP243 X 35S:GFP lines for 58, 65, 82 and 103 described in chapter 3 section 3.2.3 were used. These lines represent 2 unmethylated and 2 methylated F1 lines. The four homozygous AMP243 X 35S:GFP lines were crossed back to the wild-type ecotype Col-0 to produce hemizygous plants. These hemizygous plants are referred to as AMPLine number X Col-0 throughout, so for example the hemizygous (AMP243X 58) X Col-0 would be known as AMP58 X Col-0. These out-crossed lines did not produce GFP fluorescence (data not shown) and so were silenced, as expected from the presence of the *PVX:GFP* amplicon transgene. The methylation status of the *GFP* transgene in the non-silenced homozygous parents, the homozygous silenced AMP243 crosses and the hemizygous Col-0 out crosses (silenced) were then analysed by bisulfite sequencing (method section 2.6.2). Bisulfite treatment was used here so that levels of GFP DNA methylation at individual cytosines could be assessed. In this analysis PCR primers were used in which the forward primer was located at the 3' end of the 35S promoter region and the reverse primer in the GFP coding region. This ensured that only the methylation status of the *35S:GFP* transgene was assessed and not the GFP sequence in AMP243. The PCR products were cloned into the pGemT-easy vector (Promega) and individual clones sequenced.

The sequencing data from the bisulfite work was analysed using the CyMATE tool (Hetzl et al., 2007). Table 5.2 gives a breakdown of the percentage of each type of DNA methylation as calculated from the number of converted and un-converted cytosines. The 289 bp sequence analysed contains 52 cytosines. Of these 52 positions, 3 are in the CG context (5.77%), 8 are CHG (15.38%) and 41 in the CHH context (78.85%). In Table 5.2 the overall level of methylation is given in the second column. For example, all of the homozygous GFP parent lines exhibited low levels of DNA methylation with most having an overall level of cytosine methylation of under 3%. A minimum of four clones were sequenced for each line but as the number of clones sequenced varied the overall number of positions analysed was different for each line. The number of positions tested is given in column 3. A breakdown of the percentage of methylation in each context is also given. The percentage of CG methylation appears very high in each sample compared with the other context types, but this may be due to the small number of CG sites in the length of sequence analysed, with only 3 potential CG sites per clone. Figure 5.5 gives the same data but in pictorial form.

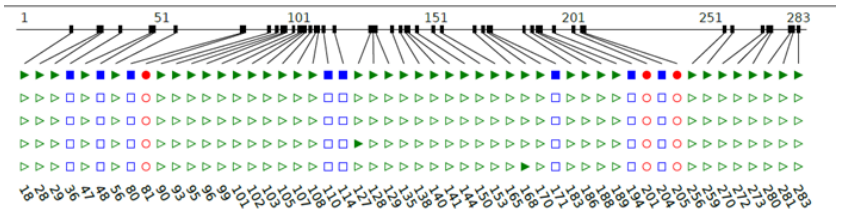
The non-silenced GFP parental samples for each line had a very high level of cytosine conversion which is indicative of a low level of DNA methylation (0.77% to 2.69%) whereas the homozygous samples had much higher levels of cytosine methylation (34.3% to 66.04%) irrespective of whether the F1 AMP243XGFP cross was methylated (Table 5.2). This is consistent with the results in chapter 3 and Figure 5.4. Changes in DNA methylation occurred in all three sequence contexts in all lines. For the hemizygous Col-0 crosses in lines 65 & 103, which were methylated in the F1 AMP243 cross, methylation is also present at a similar level to the homozygous cross samples. There was a 9.9% and 8.73% reduction in methylation between the homozygous and hemizygous Col-0 crosses in lines 65 and 103, respectively. Zygosity therefore may be unlikely to affect the methylation status of lines that readily undergo DNA methylation in the hemizygous F1 AMP243 cross. For line 58, which was not methylated in the F1 AMP243 cross, there was a 44.94% reduction in DNA methylation in the hemizgous Col-0 cross compared with the homozygous sample. This would support the hypothesis that increased zygosity can influence methylation of T-DNA; however, there is very little difference (1.85%) the level of DNA methylation between the homozygous and hemizgous samples for line 82 which was also not methylated in the F1 AMP243 cross, dispelling that hypothesis. It should be noted that line 82 had a much lower percentage of overall DNA methylation in the homozygous sample than the other lines did (34.3% vs. 56.52% & 65.94%). However, with only two lines representing non F1 methylating plants it is hard to draw firm conclusions.

Table 5.2 Percentage of *GFP* transgene DNA methylation in parent, homozygous and hemizygous samples for 4 lines.

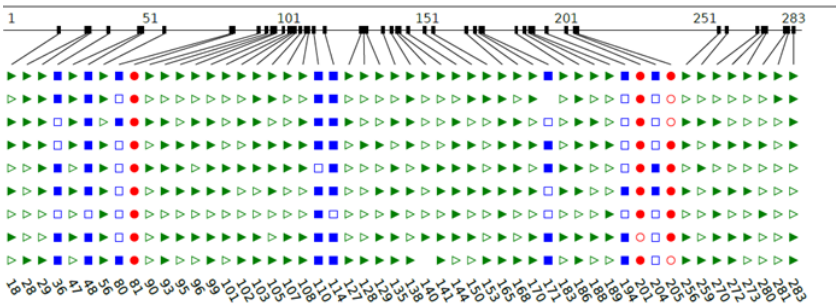
Grey rows represent samples that were methylated in the F1 AMP243 crosses. Of the total possible methylated site in the segment tested, 5.7% of the methylation could be in the CG context, 15.38% in the CHG context and 78.85% in the CHH context.

Sample	Overall methylation (%)	Sites tested	CG sites methylated (%)	CHG sites methylated (%)	CHH sites methylated (%)
Parent 58	0.96	208	0	0	1.22
Homozygous AMP243 X 58	56.52	414	83.33	58.73	54.13
AMP58 X Col-0	11.58	311	38.89	10.42	9.8
Difference between Homozygous 58 - Parent 58	55.56	206	83.33	58.73	52.91
Difference between AMP58 X Col-0 - Parent 58	10.62	103	38.89	10.42	8.58
Difference between Homozygous 58 - AMP58 X Col-0	44.94	103	44.44	48.31	44.33
Parent 65	2.69	260	13.33	2.5	1.95
Homozygous AMP243 X 65	66.04	466	92.59	73.2	62.77
AMP65 X Col-0	56.14	415	79.17	62.5	53.71
Difference between Homozygous 65 - Parent 65	63.35	206	79.26	70.7	60.82
Difference between AMP65 X Col-0 - Parent 65	53.45	155	65.84	60	51.76
Difference between Homozygous 65 -AMP65 X Col-0	9.9	51	13.42	10.7	9.06
Parent 82	0.77	260	0	0	0.98
Homozygous AMP243 X 82	34.3	414	79.17	39.06	30.06
AMP82 X Col-0	32.45	413	66.67	29.69	30.46
Difference between Homozygous 82 - Parent 82	33.53	154	79.17	39.06	29.08
Difference between AMP82 X Col-0 - Parent 82	31.68	153	66.67	29.69	29.48
Difference between Homozygous 82 - AMP82 X Col-0	1.85	1	12.5	9.37	-0.4
Parent 103	0.77	260	6.67	0	0.49
Homozygous AMP243 X 103	65.94	414	79.17	71.88	63.8
AMP103XCol-0	57.21	416	70.83	60.94	55.49
Difference between Homozygous 103 - Parent 103	65.17	154	72.5	71.88	63.31
Difference between AMP103 X Col-0 - Parent 103	56.44	156	64.16	60.94	55
Difference between Homozygous 103 - AMP103 X Col-0	8.73	-2	8.34	10.94	8.31

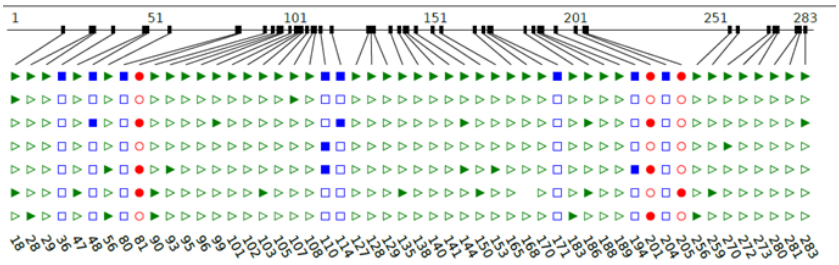
Parent 58



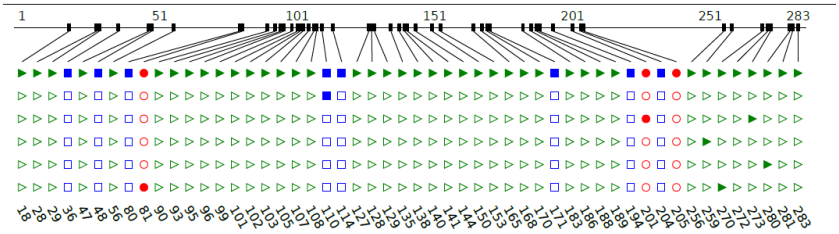
Homozygous AMP243 X 58



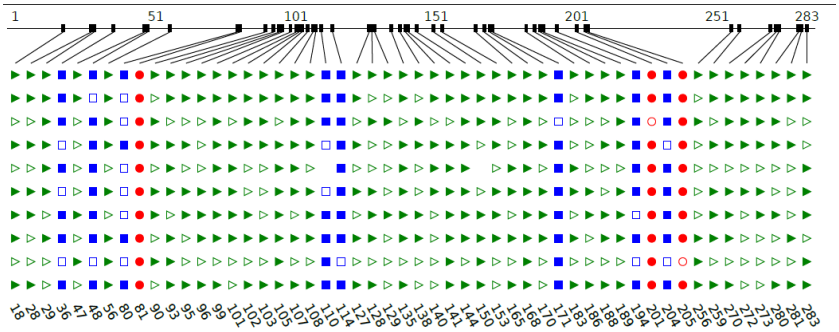
AMP58 X Col-0



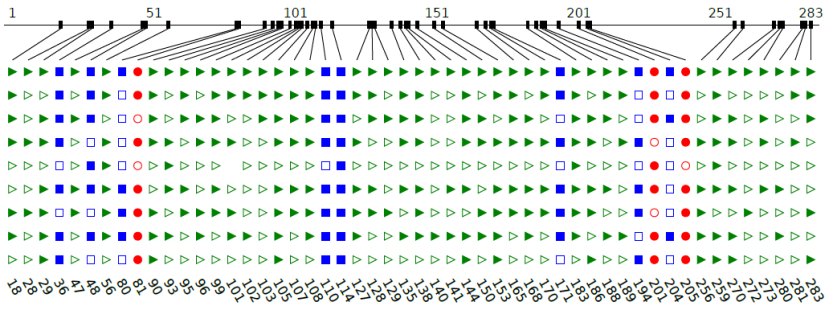
Parent 65



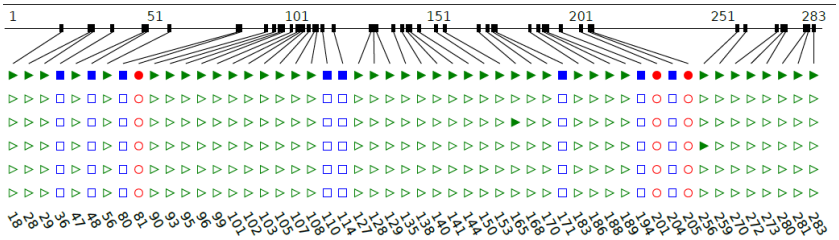
Homozygous AMP243 X 65



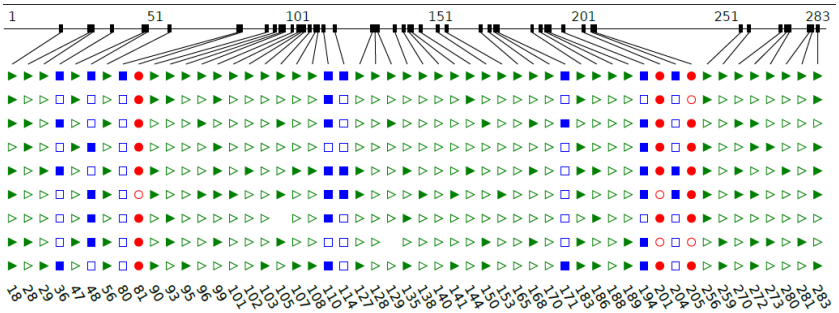
AMP65 X Col-0



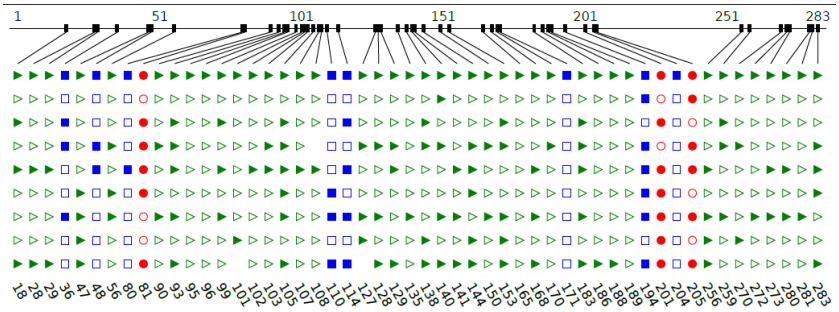
Parent 82



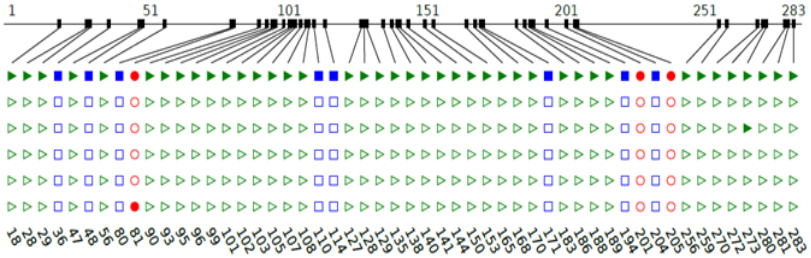
Homozygous AMP243 X 82



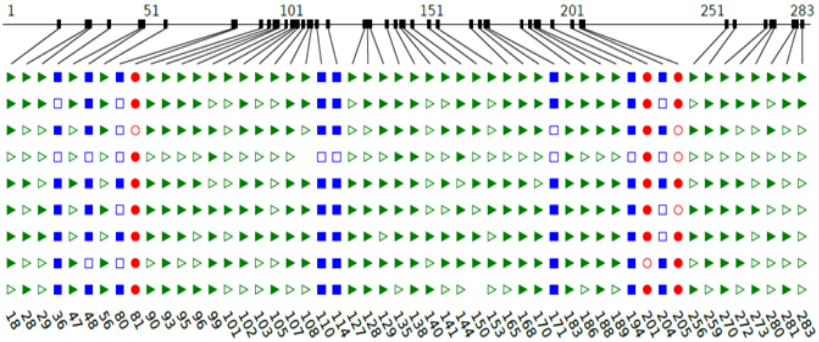
AMP82 X Col-0



Parent 103



Homozygous AMP243 X 103



AMP103 X Col-0

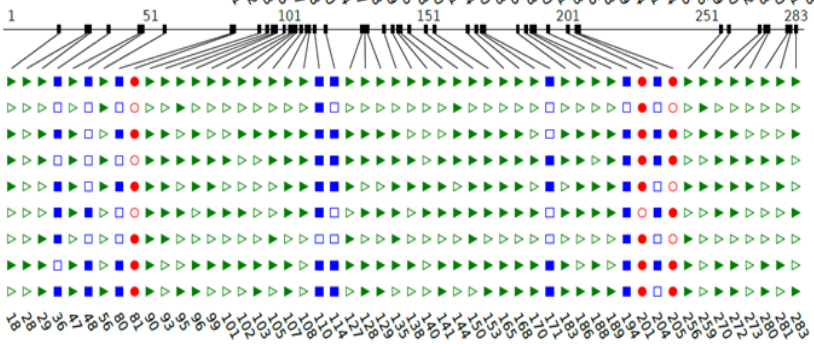


Figure 5.5 Analysis of DNA methylation status in parent, homozygous and hemizgous samples for four lines

Results of bisulfite sequencing of a segment of the 35S:GFP gene. Lines 58, 65, 82 and 103 are presented here with methylation status analysed in the homozygous parent, homozygous AMP243 cross and the hemizgous AMP X Col-0 cross for each line. A minimum of four clones per sample were sequenced and analysed using CyMATE (Hetzl et al., 2007) with more clones analysed for the crosses than for the parent lines. Open symbols are indicative of unmethylated cytosines and closed symbols indicative of methylated cytosines. Circles, squares and triangles represent cytosines in CGH, CHG, CHH methylation configurations respectively. For each panel the top line shows a master sequence and is not generated by bisulfite sequencing.

5.2.4 Spontaneous silencing in the 35S:GFP parent lines

Towards the end of this project it was noted that some of the parent plants for lines 70, 103 and 114 underwent low levels of spontaneous silencing when sown out again later for further experiments during the course of the investigation. Lines 103 and 114 undergo DNA methylation in the F1 AMP243 cross whereas line 70 does not. A number of these 3 lines were sown out in order to count numbers of silencing plants. Line 70 had 1 silenced plant out of 84, Line 103 had 1 red plant in 150 and Line 114 had 15 red plants out of 180. For Lines 70 and 103 this was considered to be a low level of background spontaneous silencing. However, there were a higher number of plants that underwent spontaneous silencing in Line 114. Line 114 and 103 were both methylated when crossed with AMP243 in the F1 generation.

Seeds were collected from the silenced line 70 and 103 plants and from 3 silenced line 114 plants. As a control seeds were also collected from a comparative number of non-silenced plants. The seeds were sown in soil and the percentage of plants exhibiting silencing was counted after 5 weeks (Table 5.3). The aim was to determine whether silencing was more likely in plants that had exhibited silencing in the previous generations. The progeny of lines that were silenced are still a mixture of silenced and non-silenced lines. There does not appear to be a disenable pattern; which would suggest that spontaneous silencing is no more common in progeny of plants that were spontaneously silenced before. For example, the percentage of silenced plants in the next generation of line 70 plants was approximately the same regardless of whether the parent was silenced or not. The results from line 114 are variable but the average of the silencing of plants with non-silenced parents was 24.4% and the average of the silencing of plants with silenced parents was 24.0%.

Table 5.3. Spontaneous silencing in next generation of 35S:GFP plants

Numbers and percentage of silenced progeny of 35S:GFP plants that were either silenced or non-silenced for GFP. Silencing status of the parent plants is given. Lines that underwent DNA methylation in F1 cross with AMP243 indicated with an *.

Line	Parent Silenced or not	Numbers red silenced plants	Total number of plants	Percentage of silencing plants (%)
70	Non-Silenced	7	240	2.9
	Silenced	2	83	2.4
103*	Non-Silenced	13	205	6.3
	Silenced	2	186	1.1
114*	Non Silenced$_1$	52	210	24.8
	Non Silenced$_2$	0	179	0.0
	Non Silenced$_3$	98	203	48.3
	Silenced $_1$	21	192	10.9
	Silenced $_2$	76	224	33.9
	Silenced $_3$	54	198	27.3

5.2.5 Two Histone modifications are not altered by AMP243 triggered silencing

Some histone modifications have been linked to RdDM. One such modification is methylated Histone 3 K9 (H3K9) (Law et al., 2010, Law et al., 2013). It is possible that the reason that some lines can support RdDM may be that the appropriate histone modifications are already present at the GFP loci before silencing but absent at the locus in lines that do not support F1 DNA methylation. Alternatively, RdDM may lead to changes in the histone code to add those marks after silencing. Additionally, active marks may be removed by RdDM.

To investigate if differences in histone marks between parents accounts for the disparities in ability to undergo DNA methylation, chromatin immunoprecipitation (ChIP) was performed on a selection of parent lines. This selection included lines that had supported DNA methylation in the F1 and lines that had not appeared to. Additionally, to investigate if RdDM affected the histone marks at loci after silencing had occurred ChIP was also performed on the corresponding homozygous AMP243 X 35S:GFP crosses for the parent lines used. These crosses all underwent silencing and are homozygous for both the trigger and the target. Lines 58, 60, 65, 70, 72, 82, 103 and 114 were used. Lines 65, 103 and 114 displayed DNA methylation in the F1 whereas the others did not. Methylation in the homozygous AMP243 X 35S:GFP crosses has been clearly demonstrated in lines 58, 65, 70, 82, 103 and 114 but untested in line 60 and 72.

ChIP was performed using whole seedlings grown on MS plates (methods section 2.7). Two antibodies were used for the ChIP; an antibody for anti-histone H3 dimethyl K9 (Abcam)

(H3K9me2) and an anti-acetyl H3 antibody (Millipore) as acetylated H3 histone marks are associated with gene activation. Both an input and a no-antibody sample were used as controls for each DNA sample.

To analyse the results from the ChIP, SYBR green qPCR was performed. Four different primer sets were used; Tubulin which was a control for active gene expression, soloLTR which is a target of RdDM (Wierzbicki et al., 2008), GFP5 which amplifies a fragment from the 5' end of the GFP gene, and GFP3 which amplifies the 3' end of GFP. To quantify the results the comparative $2^{-\Delta\Delta ct}$ method was used (reviewed by Bubner & Baldwin, 2004, Livak & Schmittgen, 2001). The data was normalised against the input Ct values. The data for the anti-acetyl H3 ChIP was calibrated to Tubulin, giving this a value of 1 (Figure 5.6). The anti-H3K9me2 ChIP was calibrated to soloLTR (Figure 5.7). The whole ChIP process was repeated for lines 65, 70 and 114 (data not shown). These gave the same results pattern as the first experiment, providing confidence in the data.

The ChIP using the anti- H3 acetyl antibody was calibrated to Tubulin for each sample, giving it a value of one. Tubulin was chosen as a positive control for H3 acetyl marks as it is an actively expressed gene and H3 acetyl mark is associated with gene activity. The Tubulin reactions were highly enriched compared to the no antibody reactions (data not shown), giving confidence in the data. Additionally, the soloLTR had a low relative enrichment which would be expected from an inactive locus, providing further confidence in the data. The ChIP using the anti- H3 acetyl antibody shows that this mark was present at the *GFP* locus in all parent and homozygous AMP243 X 35S:GFP lines (Figure 5.6). This is demonstrated by the increase seen for all lines in the GFP5 and GFP3 reactions. However, there is no difference in the level of H3 acetyl marks between the parental lines and their homozygous AMP243 X 35S:GFP line counterparts (Figure 5.6). If AMP243 triggered silencing and DNA methylation had led to changes in this histone mark then a reduction in enrichment may have been expected, assuming silencing removes active marks. This data strongly suggests that this does not occur for this mark.

The analysis of the H3K9me2 ChIP demonstrates that this mark was not present at the GFP or tubulin loci for any of the parent lines as there was no enrichment (Figure 5.7). Enrichment was observed for the soloLTR primers, as would be expected by the selection of this locus as a positive control. This, in addition to the lack of tubulin, provides confidence in the data. Therefore this mark is unlikely to contribute to the recruitment of methylation machinery to RNA silencing targets. Additionally, there was no difference between the parents and homozygous AMP243 X 35S:GFP lines (Figure 5.7). This indicates that AMP243 triggered silencing does not lead to the addition of the H3K9me2 to its targets, even in homozygous lines that undergo DNA methylation.

Originally the potential for differential histone modifications between the parents, the homozygous crosses and the F1 crosses was to be investigated. However, looking at the parents and the homozygous only, it is apparent that AMP243-triggered silencing does not result in a detectable change in either H3AC or H3K9me2 levels of the *35S:GFP* regions examined. Additionally, the difference in the ability to undergo methylation in these triggered lines is not cause by pre-existing H3K9me2 modifications. Therefore the ChIP experiments were not repeated on the F1 crosses.

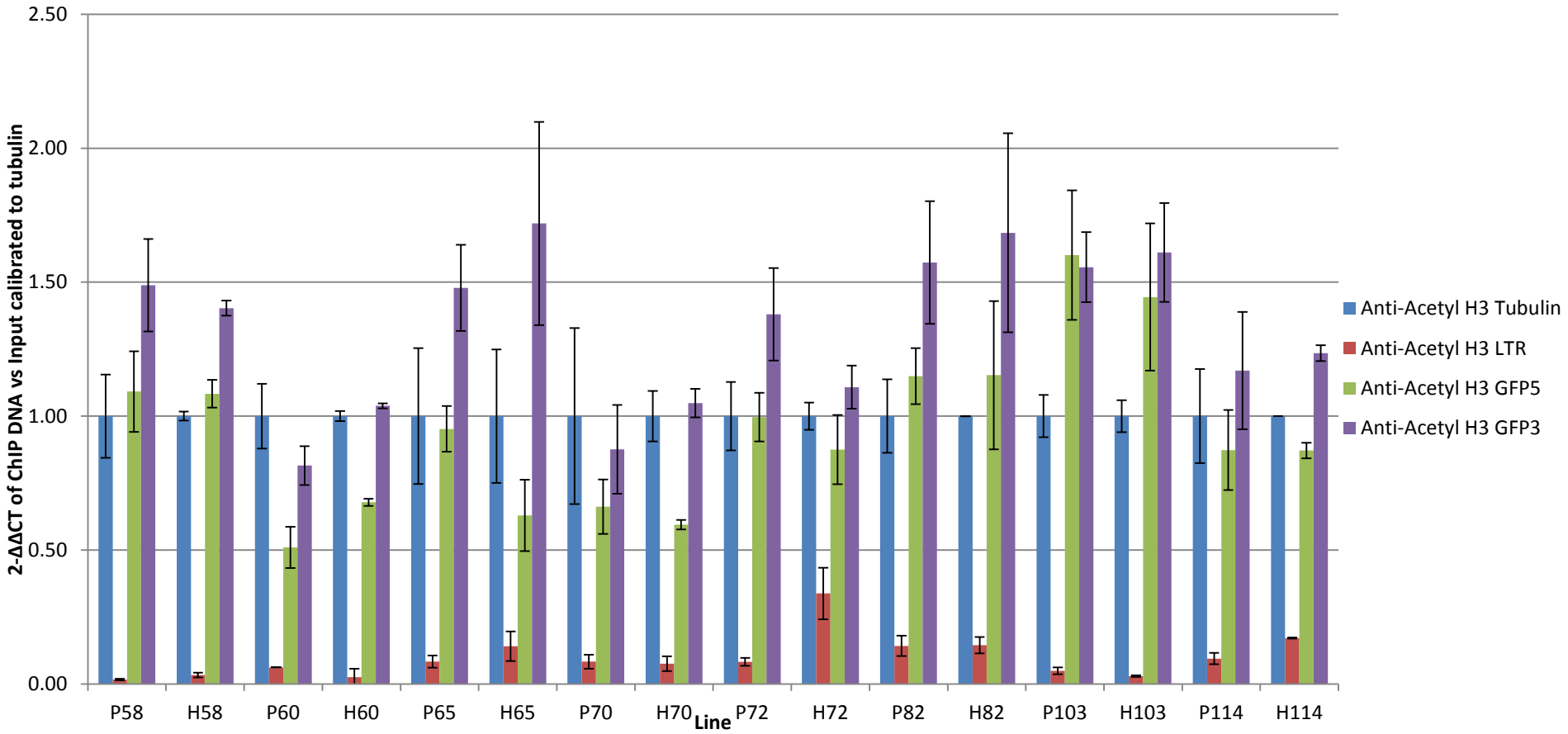


Figure 5.6. Amplicon-mediated silencing does not lead to a reduction in H3 acetylation levels at target loci

The result of qPCR analysis using comparative $2^{-\Delta\Delta ct}$ method. Data calibrated to tubulin sample for each line. qPCR was performed with 3 technical repeats for each sample. ChIP was performed on formaldehyde crosslinked seedling. P = parent and H = AMP243 X GFP homozygous lines. Antibody used was Anti-Acetyl (Millipore 06-599). Standard error is given

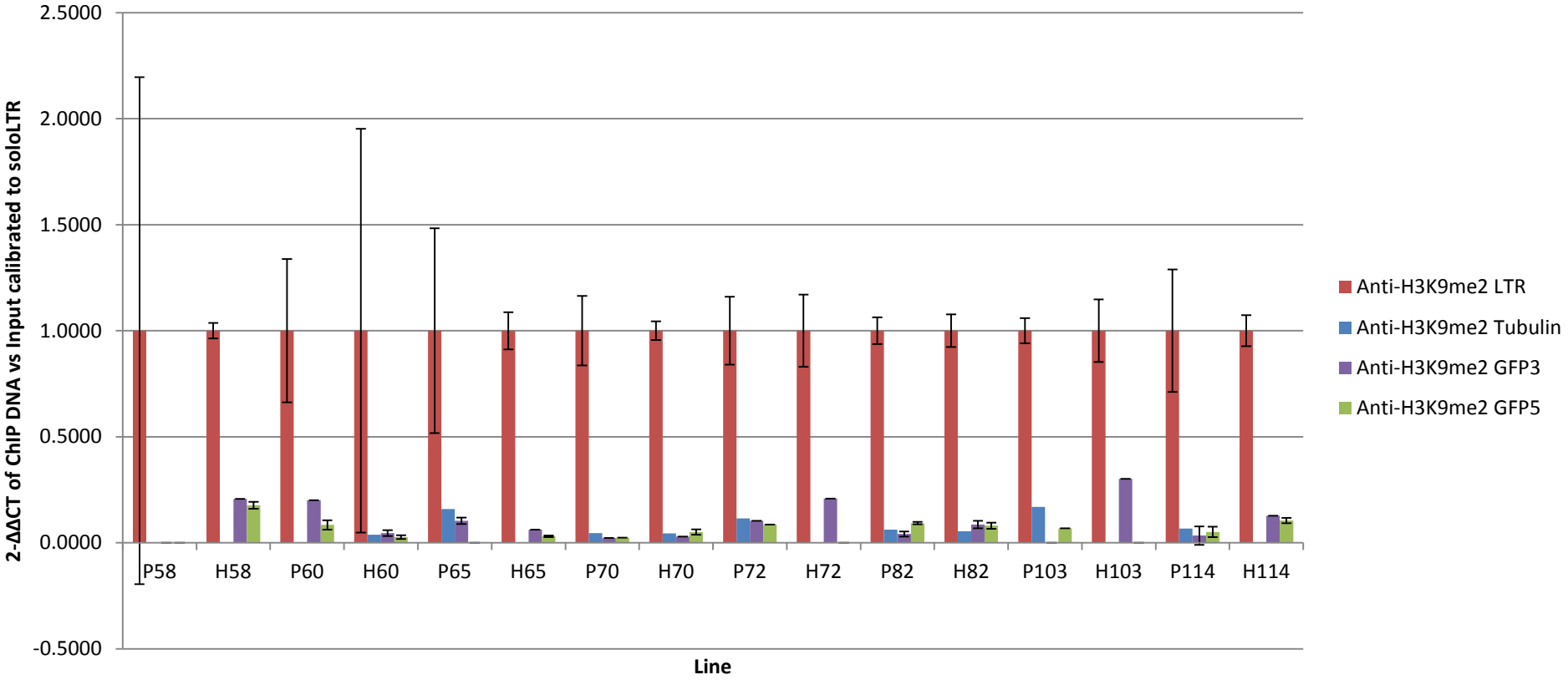


Figure 5.7. Amplicon-mediated silencing does not lead to an increase in H3K9me2 levels at target loci

The result of qPCR analysis using comparative $2^{-\Delta\Delta ct}$ method. Data calibrated to soloLTR sample for each line. qPCR was performed with 3 technical repeats for each sample. ChIP was performed on formaldehyde crosslinked seedling. P = parent and H = AMP243 X GFP homozygous lines. Antibody used was Anti-Acetyl (Millipore 06-599). Standard error is given.

5.2.6 Attempts to use the FLAG fusion proteins to investigate recruitment of RNA silencing machinery to the GFP loci were unsuccessful

One of the key questions in the field of RNA silencing is how do the different components of the silencing machinery become recruited to target. These questions were discussed in chapter 4. For example, is it possible that components that can induce DNA methylation and associate with DNA, such as Pol V, are already present at some loci before silencing occurs? Is it this that causes the difference in DNA methylation? Alternatively, are these components recruited to the loci after silencing has been triggered by AMP243? To answer this the NRPE1:FLAG, RDR2:FLAG and RDR6:FLAG lines described in chapter 4 were crossed to a selection of parent 35S:GFP lines and their respective homozygous AMP243 X 35S:GFP crosses from chapter 3. These lines were 58, 60, 65, 70, 82 and 103. ChIP analysis was then performed on the F1 generation for the RDR6:FLAG, RDR2:FLAG and NRPE1:FLAG with line 103 as an initial study (data not shown). However, there where technical issues with the experiments that meant that no conclusions could be formed. This will be described below.

ChIP was performed using whole seedlings grown on MS plates (methods section 2.7). An anti-FLAG antibody (Sigma F3165) was used in the ChIP and both an input and a no-antibody sample were used as controls for each DNA sample. To analyse the results from the ChIP, SYBR green qPCR was performed. The qPCR was carried out with primers for the 5' end of GFP, the 3' end of GFP, tubulin as a non-silencing negative control and the solo long terminal repeat (LTR). SoloLTR was used as a positive control for Pol V as this locus has been reported as being occupied by NRPE1 (Wierzbicki et al., 2008). However, none of the ChIP samples gave enrichment above the no antibody control. The failure of the positive control is indicative of technical problems with the ChIP process and means that no confidence can be had in the results. Therefore, the outcomes of these experiments are not presented here.

There are many reasons why the ChIP may not have worked. One possibility is that protein expression in the seedlings of NRPE1, RDR2 and RDR6 is not high enough to produce detectable levels of FLAG whereas the protein expression may be higher in the floral tissue. NRPE1:FLAG is the only line used here with a definitive positive control for the ChIP experiments so efforts to troubleshoot this ChIP work were performed first. However, time was limited at this stage and so only initial progress in troubleshooting could be made.

To investigate if using floral tissue improved the ChIP data flowers from soil grown parents and crosses from lines 70 and 103 were used. However, qPCR using the soloLTR primers, which should act as a positive control for the NRPE1:FLAG lines, did not give any enrichment above background as determined from the no-antibody negative control (data not shown).

It was surprising that the ChIP protocol used had been unsuccessful here as the protocol had been used successfully for the histone work. This could indicate an issue with the antibody used. All of the FLAG fusion proteins were detectable by western blot in the uncrossed lines (chapter 4). However, when western blots were performed on the crosses of NRPE1:FLAG, RDR6:FLAG and RDR2:FLAG used for ChIP technical problems prevented accurate detection of the FLAG fusion proteins due to issues with the controls (data not shown). Due to limited time both the ChIP experiments and the western blots using the FLAG tagged lines were not taken any further. However, this line of enquiry would be one to expand upon in the future.

5.3 Discussion

5.3.1 Summary of results

In chapter 3 the AMP243 trigger line was used to cause de novo silencing of a *GFP* transgene in multiple independent single locus lines. Whilst silencing occurred in all lines, only 8 out of the 40 lines screened displayed evidence of DNA methylation as determined by Southern blot. In chapter 3 the effects of the genomic site of transgene insertion on the ability to undergo DNA methylation were investigated. However, no support for possible genomic site effects was found. This chapter has investigated other possible causes of the difference in DNA methylation between lines. These included examining differences in copy number, integration structure and histone modifications between the different lines. Additionally, ChIP analysis was attempted to explore possible differences in Pol V recruitment; however this was unsuccessful. This chapter has demonstrated that the levels of DNA methylation of the *GFP* transgene increases over generations in the presence of the AMP243 trigger line.

5.3.2 Copy number

The copy number analysis, as conducted by qPCR, suggests that lines that supported DNA methylation in the F1 AMP243 X 35S:GFP cross tend to have a higher relative copy number than those that did not appear to. So how might copy number influence the likelihood that the amplicon-trigged silencing will result in DNA methylation of the target? One aforementioned idea is that having multiple copies is likely to correlate with complex integration structures which may make them more susceptible to DNA methylation. This will be discussed in detail later. However, another copy number-dependent hypothesis for the reasons behind the generally observed variation in transgene expression is the idea of a threshold concentration (Lechtenberg et al., 2003, Schubert et al., 2004). This suggests that if the expression of a transgene causes the number of its transcripts to breach a gene specific threshold value then that transgene will become a target for silencing. In other words high levels of transgene expression could lead to

silencing. This makes sense from the plants perspective as a way to protect the plant from high expression of a potentially harmfully non-endogenous gene. Although, how the plant distinguishes a highly expressing endogenous gene from a transgene is not clear in this model.

Having a high copy number may lead to higher expression of the transgene, resulting in a breach of the threshold leading to silencing. However, in the system used here silencing was triggered by AMP243. This is RDR6-dependent, as demonstrated in chapter 4, which suggests that RNA is being produced and that expression of the transgene is occurring. This would give any plant-based system that monitors transcript level transcripts to detect. However, if a copy number of 2, 3 or 4 caused sufficient expression for the threshold to be breached then in theory silencing of the *GFP* gene should have occurred before the addition of the AMP243 trigger. Schubert et al. (2004) demonstrated in their work that plants that had 5 or more copies of a *GFP* transgene underwent silencing in the absence of a trigger. None of the lines here had a *GFP* copy number of 5. However, the work by Schubert et al. (2004) included lines with *GFP* inserted at multiple loci. It is possible that once silencing has been triggered then a transgene that produces a high level of transcripts due to a large copy number may present a target for the RdDM machinery. Whereas lines with low copy numbers and low expression may not, perhaps because they do not meet the threshold concentration to act as a target. However, a difference in GFP mRNA transcript levels between lines has not been fully investigated in this project. In chapter 3 it was established that GFP mRNA is detectable by northern blot in the parents but not in the F1 AMP243 crosses, as would be expected from silenced lines. However, what was not looked at was the variation in expression levels of GFP mRNA between the parent lines before silencing was triggered. This would be useful as copy number may not correlate with transcription due to the possibility of partial insertion of part of the *GFP* transgene cassette in which may result in multiple copies of the *GFP* gene bur not of its promoter 35S.

Visual inspection of the northern blot of mRNA levels presented in chapter 3 does suggest that there is a slight variation line-to-line in the levels of GFP mRNA expressed. For example parent 14 gives a very strong band, indicative of high GFP mRNA levels whereas line 103 has a much weaker band, even though the ethidium bromide gel image suggests they both have similar levels of total RNA. An actin loading control was only present for one of the blots. Whilst an ethidium bromide gel image does give a good indication of the relative amounts of RNA loaded on to the gel, it does not give an idea of how effective the transfer was. Both lines 14 and 103 undergo methylation in the F1 AMP243 cross. Here line 14 has a copy number of 3 and 103 has a copy number of 2. Line 65, another F1 methylating line, has a similar intensity band to line 14 and it too has a copy number of 3. However, parent of lines 57 and 60 also have a similar intensity bands to 14 and 65

but these lines do not appear to undergo methylation in the F1. This does cast doubt on the possibility that differences in mRNA transcription levels, as a result of varying copy numbers, affect the ability of RNA silencing to cause transgene DNA methylation. However, this assertion is made using visual inspection of northern blot data. A more accurate way to measure transcription levels of GFP mRNA would be to use qPCR on cDNA made from total RNA extracted from the parent lines. This would provide data on any differences in GFP mRNA accumulation.

There are other aspects of the results from this investigation that suggest that copy number variation alone does not affect RNA silencing outcomes. For example, line 71 was an exception to the idea that high copy number leads to DNA methylation as it has a relative copy number of 1.41, which is indicative of a copy number of only one, but underwent DNA methylation in the F1 AMP243 cross. The relative copy number in line 71 was supported by the integration structure analysis. The idea that a high copy number alone would lead to DNA methylation was also negated by the fact that many lines with high copy numbers did not support DNA methylation in their F1 cross. For example, line 120, which had a copy number of four, did not undergo DNA methylation in the F1 cross. Again this highlights the possibility that copy number alone is unlikely to be the driving cause of RdDM.

The qPCR results for the copy number analysis did not give discrete values that may have been expected from a PCR designed to assess copy number. Amplification of a fragment is dependent on a pair of primers and so in theory PCR should occur between the two annealing sites or not at all. Therefore even if the T-DNA integration of one of the transgene cassettes had been incomplete PCR amplification would have occurred if the fragment it copies was present, even if the rest of the gene was absent. If the fragment was absent then PCR should not occur. The continuous nature of the relative copy numbers given is therefore unexpected. Initially melting curves were included in the qPCR run to check for multiple products caused by the primers (data not shown). This was not performed for all lines however.

Another reason as to why the data from the qPCR was not discrete may be due to the lines having multiple sites of integration, for the reasons outlined in chapter 3. There are limitations to using a 3:1 ratio of expression vs no expression to find single locus lines. The ratio was analysed using the χ^2 test to indicate which lines are likely to have a true 3:1 ratio and therefore a possible single integration site. The χ^2 test will have found lines that are statistically likely to have a true 3:1 ratio but, as with all statistics, there are limitations to this. Additionally, assuming that lines that meet the criteria of the χ^2 analysis have a single site of integration has limitations. For example, this method of analysis would not be able to filter out lines with linked multiple insertions. It would

also not detect any second silenced copies of the *GFP* transgene. If these additional copies of the *GFP* transgene are present it may affect the outcome of the copy number analysis.

The comparative $2^{-\Delta\Delta ct}$ method used to analyse the qPCR data requires a calibrator line to standardise the data to a known value after it has been normalized to the endogenous control (reviewed by Bubner & Baldwin, 2004, Livak & Schmittgen, 2001). In this case a line with a known GFP copy number of one would have been helpful, but such a line was not available at the start of the investigation. The qPCR and the integration analysis were performed concurrently so once it was established that line number 10 had a copy number of one then it was used as the calibrator. Line GFP142, which had been used to create the DAN2 line, reportedly had a copy number of one (Dalmay et al., 2000a). Although segregation analysis on this line was reported it was decided to use a line generated in this body of work with its copy number determined by qPCR and Southern blot. The complete qPCR and integration analysis presented here do show that line GFP142 does indeed have a copy number of one.

5.3.3 Transgene integration structure

The analysis of the transgene integration structure was determined by a combination of the BamH1 Southern blot and the qPCR copy number work. It appears that lines that undergo methylation in the F1 AMP243 crosses tend to not only have a high copy number and therefore a complex integration structure, with the exception of line 71. Of the 8 methylating lines a possible combination of T-DNA inserts could be reasonably determined for only four of them (Table 5.1). Line 71 had a copy number of one and therefore had a simple integration structure. Line 103 appears to have two transgene cassettes in an inverted repeat formation. Lines 14 and 56 have a series of inverted repeats made up of three cassettes. The integration structure of the remaining four lines, 29, 50, 65 and 114, could not be easily determined by the Southern blot and qPCR analysis. These had a relative copy number of 2 or 3 but a banding pattern on the Southern blot which did not give a clearly explainable structure. For example, lines 29 and 50 gave a relative copy number of 2 but had three bands for the GFP sample and only one for the Kanamycin resistance gene sample on the Southern blot. This makes it unclear as to what the integration structure is in these samples as there are a few possible explanations. For example, there may have been fragments generated by the BamH1 digest that are the same size and therefore do not separate by gel electrophoresis. The kanamycin-resistance gene band for line 29 may represent a good case of this masking as it is quite an intense band. Secondly, it may be possible that the whole GFP gene has not been incorporated and that incomplete copies of the T-DNA cassette have been integrated. The GFP probe used for the Southern contained the whole length of the GFP gene whereas the qPCR primer only amplifies the 5' end of GFP. Therefore it is possible that

if only the 3' end of the *GFP* transgene had been added then the Southern blot would detect it but the qPCR would not.

When T-DNA is inserted into a plant genome part of the vector sequence is very often inserted with it (Kononov et al., 1997). The addition of vector sequence in T-DNA has been connected to silencing of transgenes (Iglesias et al., 1997, Jakowitsch et al., 1999). This presents the possibility that the non-plant origin of the vector sequence is detected by the plant and influences its capacity to undergo DNA methylation. It would be interesting if large additions of the vector sequence correlated with the 8 methylating lines. It is hard to determine from either the qPCR data or the Southern blot performed here how much of the vector sequence has also been integrated with the *35S:GFP* transgene; this has not been directly investigated. The results from the TAIL PCR in chapter 3 may shed some light on the matter. Of the 8 lines that underwent DNA methylation in the F1 AMP243 crosses an integration site for 5 lines was determined by the TAIL PCR. The remaining 3 lines were 29, 50, and 65. All three of these gave complex integration structure that could not be ascertained by the Southern blot. Sequencing after TAIL PCR only produced vector sequence for lines 29 and 50 demonstrating that for these lines a complex pattern of vector sequence and transgene is present. The TAIL PCR did produce a PCR fragment for the line 65 sample, however the sequencing failed. Consequently it is unclear if large amounts of vector sequence are present in the T-DNA inserted into that line or not. It should be noted that the TAIL PCR was not designed to evaluate the integration of the vector sequence and so this is just an interesting observation. Consequently lines where the integration site was ascertained by TAIL PCR may have also had vector sequence integrated with the T-DNA. Therefore no strong conclusions can be drawn from this. For example the structure of line 114 could also not be determined by the Southern blot analysis, but the integration site for this line was found by the TAIL PCR work. This does not exclude the possibility of repeats of the vector sequence in this line, it just doesn't confirm it.

Inclusion of vector sequence may not be what sets the methylating lines apart from the other lines. For example there were other lines that only returned vector sequence from the TAIL work, which were not part of the set of F1 methylating lines. These were lines 10, 17, and 18. Line 17 did undergo methylation in the F1 but spontaneous silencing was observed in the parent lines, therefore the line was not included as an F1 methylating line as the silencing that occurs in this line may not be AMP243 dependent. Lines 10 and 18 demonstrated no signs of methylation in their F1 AMP243 crosses and have a copy number of one. Meza et al. (2002) conclude that inclusion of vector sequence does not trigger silencing of single copy transgene in Arabidopsis. Additionally, work on tobacco has demonstrated that up to 75% of transgenes integrated into

plant genomes via Agrobacterium-mediated transformation contain vector sequence (Kononov et al., 1997). This is a high percentage which may suggest that the presence of vector sequence alone is unlikely to initiate RdDM as 75% of the lines here did not appear to undergo DNA methylation.

It should be noted that, even for lines where the structure appears to have been solved, these are not the definitive structures. To ascertain these full sequencing of the transgene could be performed. To do this PCR reactions could be designed that could identify inverted repeat structures for example. Additionally, direct sequencing of the 35S:GFP lines may help to undercover any potential multiple insertion sites.

5.3.4 Zygosity

It has been demonstrated that transgenes are more likely to undergo silencing when in a homozygous state than when in a hemizygous one (de Carvalho et al., 1992, De Wilde et al., 2001, James et al., 2002, Velten et al., 2012). However, none of the referenced work investigates the effects of zygosity on DNA methylation. Here all of the hemizygous F1 AMP243 X 35S:GFP crosses were silenced due to the presence of the AMP243 trigger. However, only 8 out of 40 lines underwent detectable DNA methylation in the hemizygous cross. Yet in the F2 generation, as generated by selfing, DNA methylation was observed in all 9 lines screened even if they had not undergone detectable methylation in the F1. How does this occur? The observation that silencing occurs more readily when the transgene is in a homozygous state compared to the hemizygous one has been attributed to the increase in gene dosage. Additionally, there may be DNA-DNA interactions may account for the increase in RdDM.

In order to establish if zygosity was a factor in RdDM four homozygous AMP243 X 35S:GFP lines were out crossed to the wild type Col-0 and underwent bisulfite treatment. As expected the lines that had undergone DNA methylation in the hemizygous F1 AMP cross also underwent DNA methylation in the hemizygous cross Col-0 (Lines 65 and 103). However, outcrossing to the hemizygous state did not restore levels of DNA methylation to the levels observed in the parents of the line for any of the lines investigated. Line 58, which was not methylated in the F1 had the biggest reduction in DNA methylation between the homozygous AMP243 X 35S:GFP 58 and the hemizygous AMP58 X Col-0 which hints at the possibility of an effect of zygosity. However, the results from line 82 are not consistent with this hypothesis. Line 82 has the smallest difference in DNA methylation between any of the AMP homozygous line and the hemizygous out crossed Col-0 lines with only a change of 1.85%. However, with such a small sample size it is hard to draw any firm conclusions. Both 58 and 82 do not undergo DNA methylation in the F1. Line 58 has a relative copy number of two whereas line 82 has a relative copy number of 1.

The bisulfite sequencing work for the out cross highlights that the DNA methylation observed in Figure 5.4 is not equal for each line. Line 82 has a lower level of overall DNA methylation in the homozygous AMP243 cross at 34.3% than any of the other lines (56.52 to 66.04%). This may explain why there was no reduction in this line between the homozygous AMP243 cross and AMP82 X Col-0; the difference was low because the increase DNA methylation between the crosses and the parent was low to start with. Further bisulfite sequencing on other lines would be required to establish whether variation in the levels of F2 methylation existed. This would require the generation of more homozygous lines.

The alternative explanation to the increase in DNA methylation observed in Figure 5.4 could be a result of the lines passing through generations. Work by Kuhlmann et al. (2014) in *A. thaliana* using a transgene system suggested that DNA methylation in the CG context was heritable, even in *drm2* mutants. They therefore conclude that there are two stages to the establishment of DNA at a locus. The first stage would be the full silencing of the region by RdDM which would be followed by a switch to maintenance of DNA methylation in the CG context in subsequent generations. They report that for the more generations a locus undergoes RdDM the more persistent DNA methylation is through subsequent generations.

Working on inverted-repeat triggered transgene silencing in tobacco Lunerová-Bedřichová et al. (2008) have also reported an increase in CG context methylation through up to 3 generations after an initial round of RdDM. They report that the CG methylation is still present after the transgene locus has been segregated away from the trigger of silencing but that the transgene is released from silencing without the trigger. Kuhlmann et al. (2014) had reported that silencing still occurred in lines that maintained the RdDM induced CG methylation in the absence of the trigger, but their system also involved methylation of the promoter and TGS.

The observations made in both reports explains how DNA methylation caused by RdDM can be maintained through generations, but would not explain how moving through generation would trigger RdDM, whereas the changes in zygosity might. However, maintenance of RdDM triggered CG methylation through generation may have affected the results of outcrossing the homozygous AMPGFP lines to Col-0, as they have all undergone RdDM previous to crossing. However, in the out-crosses made here, methylation in all contexts was altered seemly equally or not at all (Table 5.2).

5.3.5 Spontaneous silencing in the parent lines

Plants of lines 70, 103 and 114 displayed low levels of spontaneous silencing. This was only noticed when a larger number of plant of these lines were sown out again later in the project.

These lines had been sown out many times throughout this project to generate tissue for DNA & RNA extraction as well as parents for crosses, so the fact that spontaneous silencing was noticed quite late in the project does suggest that the levels are very low. This is corroborated by data gathered to investigate the issue of spontaneous silencing. Out of 84 plants there was only one silenced plant for line 70 and for line 103 there was only one silenced plant out of 150. So for this particular experiment only 1.2% and 0.6% of plants were silenced for lines 70 and 103, respectively. These results suggest a low level of spontaneous silencing in these lines, which would be unlikely to affect the results of the chi-square analysis conducted in chapter 3. A 3:1 ratio of GFP positive to GFP negative was used to determine plants with single locus insertions of the *35S:GFP* transgene.

For line 114 however, 8.3% of the plants tested displayed silencing which is higher than in line 70 and 103. It should be noted that the silenced plants in all three lines here are not the result of contamination as their progeny produced GFP expressing plants. The qPCR data gave line 114 a relative copy number of 3.33 and the integration structure was inconclusive. Both lines 70 and 103 each have a relative copy number on 2 and both lines contain an inverted repeat according to the analysis preformed here. The TAIL PCR placed the integration site of the T-DNA in line 114 into an un-named gene whereas line 70 and 103 were not with in a genic region. Here line 114 is reported to have a higher level of spontaneous silencing than 70 and 103, which is borne out in the data.

The low level of spontaneous silencing seen in lines 70 and 103 was also seen in the next generation of these lines. This data suggests that silencing in one generation does not lead to a heritable increase in silencing in the next as the similar levels of silencing are seen in both lines regardless of the silencing status of the parents. The subsequent generation of line 114 plants have a highly variable amount of spontaneous silencing. Again this was regardless of the silencing status of the parent plant suggesting that although the rate of spontaneous silencing may be heritable within a line, the rate itself is not affected by the previous generation.

5.3.6 The use of FLAG tagged proteins

It was originally hoped that ChIP could be performed using the FLAG-tagged lines described in chapter 4 to investigate differences in protein recruitment to the GFP loci before and after silencing had been triggered. Here ChIP was performed on Parent and AMP103 lines for the 35S:GFP 103 line that had been crossed with FLAG-tagged NRPE1, RDR2 and RDR6. However, there was a technical problem with the ChIP which meant that no data from this area of investigation could be gathered. However, the positive control used was only for NRPE1 and therefore it is unclear if the ChIP for RDR6 and RDR2 had worked or not. It is tempting to

conclude that they haven't. It was unknown if RDR6 and RDR2 would be detectable at the GFP loci from the outset of this project. The generation of lines containing FLAG fusion proteins had been fairly time consuming as had crossing the lines with the 35S:GFP lines and their AMP243 crosses. This resulted in limited time available to troubleshoot the technical issues. Initially only line 103 whole seedlings had been investigated. According to TAIR NRPE1 is highly expressed in floral tissue. Therefore, the ChIP was repeated on flowers from soil grown plants. To check it was not an issue with line 103 the floral repeat included line 70 as well. Again the soloLTR positive control did not work here either. The ChIP protocol used here worked when investigating histone modifications which suggests that, at least for the whole seedlings, there is not a problem with the crosslinking used to prepare the plant tissue. Additionally, it suggests that the ChIP and qPCR protocols used for the soloLTR primers do work. Therefore there is either a problem with the antibody used or that protein expression of RDR6, RDR2 and NRPE1 FLAG fusion proteins has been reduced and is not detectable. Histones proteins are abundant making their detection easier therefore the protocol may not be sensitive enough to detect proteins with a low abundance.

To overcome any potential issues with the antibody other anti-FLAG antibodies could be used. The FLAG tag is very commonly used and many other antibodies are available. With regards to the level of protein production the option of using other tissue has already been examined. It is possible that the act of crossing the FLAG lines with the 35S:GFP and AMP lines has affected the expression of the fusion proteins. In all three lines the FLAG fusion proteins could be detected by western blot before crossing with the GFP lines. Western blot could therefore be used to investigate if the FLAG fusion proteins are detectable after crossing. This western blot was conducted but there was a technical issue that meant it did not produce meaningful data. Unfortunately being able to detect a protein by western blot does not always translate to ChIP. Another way of addressing what is detectable by western blot and what is detectable by ChIP would be to use the FLAG fusion lines alone in a ChIP experiment. The 35S:GFP and AMP lines were not mutant for *NRPE1*, *RDR2* or *RDR6* meaning that they will produce non-tagged functional version of all of these proteins. It is possible that the FLAG tagged versions of the proteins don't work as well as the non-tagged versions. Evidence of this was seen for RDR2 in chapter 4 as the RDR2:FLAG fusion protein only partial rescued the *rdr*2 5S rDNA phenotype. Perhaps the FLAG tag has caused some protein folding issues, which either affects the expression of the FLAG fusion proteins or how the epitope is presented to the antibody. All of these issue highlight why using antibodies raised against the endogenous genes can simplify matters.

5.3.7 Histone modifications

The main advantage of using the AMP243 system to trigger RNA silencing in a transgene system is that it allows differences in the transgene's protein associations before and after silencing to be investigated. Here the differences in two histone modifications were examined, Acetyl H3 and H3K9me2. Acetyl H3 is associated with active chromatin and H3K9me2 is a mark associated with gene silencing. The controls used for both sets of ChIP worked well, providing confidence in the ChIP data. At least for the two sets of GFP primers used here, the *35S:GFP* transgene was associated with the active acetyl H3 mark both before and after silencing for all lines tested. The *35S:GFP* transgene was not associated with the H3K9me2 mark before or after silencing for any of the lines investigated. Together these results demonstrate that for the two marks investigated here AMP243-triggered silencing does not alter their presences or absence. Additional there was no detectable differences between the parent lines for these two marks that could account for the observed variation in F1 cross DNA methylation.

The lack of the H3K9me2 modification to the *GFP* transgenes in this study may suggest that the protein SHH1 is not involved with the AMP243-triggered silencing of this transgene. The SHH1 protein has been linked to recruitment of Pol IV to around 44% of Pol IV-dependent silencing targets (Law et al., 2013). It has not been established here if AMP243 trigger DNA methylation is dependent on Pol IV but it has been reported that silencing of the DAN2 line, as triggered by AMP243 is influenced by Pol IV (Herr et al., 2005). It had been originally hoped to use NRPD1:FLAG constructs to look at this issue, however this did not come to fruition for reasons described in chapter 4. Therefore it is not clear if the DNA methylation observed as a result of AMP243 triggered silencing occurs in a manner consistent with the canonical RdDM pathway involving Pol IV.

A genome wide study has suggested that H3K9me2 predominately occurs at pericentromeric regions and has a high correlation with DNA methylation, mainly in the CG context (Bernatavichute et al., 2008). H3K9me2 has also been connected to the CMT3 pathway that predominately maintains CHG methylation. The bisulfite sequencing of the *GFP* transgene in Figure 5.5 revealed that the DNA methylation present was mainly in the CHG context.

The ChIP experiments conducted for this project only explore differences between two histone modifications. The histone code consists of many other modifications. Without repeating the ChIP for other histone marks it is not clear if AMP243 triggered silencing does have some effect on other histone modifications. For example, H3K27me1 is a histone modification that is associated with heterochromatin formation and its addition to some RNA silencing loci, such as soloLTR, has been shown to be dependent on Pol V (Wierzbicki et al., 2008). Additionally, H3K4me2 and

H3K4me3 histone marks are associated with active chromatin (Greenberg et al., 2013). Greenberg et al have demonstrated that mutations in the demethylases *jumonji 14* (*jmj14*) and *lysine-specific demethylase 1-like 1 & 2* (*LDL1/LDL2*) decrease the RdDM due to increased presence of H3K4me2/me3 modifications that have not been removed.

It should be noted that the methodology used in this project would mean that lines that initially associate with inactive histone modifications that induce silencing before the addition of the trigger would be discounted from the investigation. However, to be able to investigate modifications before and after silencing a target that does not silence is required, and there is no way around this using the AMP243 trigger.

5.3.8 Conclusion

It would appear from this body of work that from the areas investigated here no one single reason could be attributed to the variation of DNA methylation observed in the F1 AMP243 crosses. Complex integration structure and high copy number seem the most likely source of variation, although there are exceptions to this meaning that these alone do not provide a complete explanation. This leaves two possible explanations. Either, there is another reason that has not been explored here that is responsible for the variation. Alternatively, a combination of the factors examined here, and perhaps others, lead to the silencing of transgenes.

Two histone modifications were investigated here that were not affected by AMP243-triggered silencing. This was a surprise as the H3K9me2 mark has been strongly linked to DNA methylation. This suggests that this non-canonical pathway of DNA methylation may not be linked with changes in histone modification, although further testing will be required to establish if this is true histone-independent DNA methylation of a transgene.

Additionally, the results presented in this chapter clearly demonstrate that the levels of RdDM increase over a generation. The zygosity of the transgene was explored as a possible explanation for this observation. However, it would appear that investigation into the effects of passing through generations on AMP243 triggered RdDM may be more informative.

5.4 Acknowledgments

Louise Jones preformed the bisulfite treatment for Figure 5.5. The data analysis is my own work.

Fabian Vaistij gave technical advice for the qPCR and ChIP protocol

Sally James from the Technology Facility gave technical advice regarding qPCR.

Sequencing performed by the Technology Facility, Department of Biology, York

Chapter 6 Discussion

The variability of transgene expression has been a widely researched question due to its importance as a barrier to easily producing transgenic plants and crops (Butaye et al., 2005, Meyer, 2013). Although many of the ideas explored here have been established by others, such as effects of genomic location of transgene integration on silencing, the AMP243 system of triggering RDR6-dependent PTGS and (potentially) RdDM has allowed differences before and after silencing to be investigated (Dalmay et al., 2000a, Fischer et al., 2008, Herr et al., 2005, Jones et al., 1998). Here *de novo* silencing events have been looked at in multiple independent 35S:GFP lines both before and after silencing, which to the authors knowledge has not been conducted elsewhere to this extent using the AMP243 trigger. The AMP243 line has been a useful tool to investigate silencing in many other investigations, the work on line GFP142 is testament to this, however this body of work presents the largest number of lines crossed using the trigger (Dalmay et al., 2000a, Herr et al., 2005). The multiple independent 35S:GFP lines have also allowed investigation into the effects of genomic location and transgene structure on both silencing and RdDM.

6.1 AMP243-triggered PTGS was not affected by the genomic location

The data presented here strongly suggests that the genomic location of a transgene does not affect AMP243-trigged PTGS. The finding answers one of the original aims of this piece of work which was to establish if the genomic location of a transgene affected its ability to undergo RDR6-dependent PTGS. This was based on a previous study that found that transgenic lines carrying identical transgenes at different genomic locations differed in their ability to support the maintenance phase of PTGS (Jones et al., 1998). Using the AMP243 trigger all 35S:GFP lines tested underwent silencing. AMP243 silencing has been demonstrated here to be dependent on RDR6 and therefore, this work suggests that genomic location does not affect the action of RDR6-dependent silencing. Unfortunately, this does not give a definitive answer to the question of RDR6 target specificity.

There are alternative theories as to what labels a transcript as a target of RDR6. For example, transcripts that are “aberrant” in some way, for example missing the poly A tail or lacking the 5’ cap, have been reported to be used by RDR6 as a template (Gazzani et al., 2004, Luo & Chen, 2007). Luo and Chen (2007) used the GUS reporter gene and various RNA silencing mutants, including *rdr6-11*, to investigate the effects of aberrant transcripts on RDR6 activity. They

concluded that truncated, unpolyadenylated transcripts were likely candidates for RDR6 templates as these accumulated in *rdr6* mutants and GUS transgene that lacked promoters triggered higher levels of GUS transgene silencing. They also reported that GUS transgenes that contained multiple copies at a single integration site caused silencing with high efficiency. Subsequent work by Curaba and Chen (2008) has suggested that RDR6 activity is not affected by the presence or absence of a cap or a poly-A tail on a RNA transcript *in vitro* which could potentially mean that although aberrant transcripts can be processed by RDR6 they are not the preferred substrate. However, this does not exclude a role for aberrant transcripts in RDR6-dependent silencing it just suggests that the factors that make the RNA aberrant are not the main signals to RDR6. Other proteins may affect the activity of RDR6 *in vivo*.

The quality of the mRNA of the *35S:GFP* transgene used here was not investigated and so it is unknown if any of the lines produced such aberrant GFP mRNAs and therefore if this would have an effect on AMP243 triggered silencing. However, due to the strong GFP expression in all non-crossed 35S:GFP lines, it is unlikely that aberrant RNAs were present as this would have resulted in the silencing in the absence of the trigger, assuming aberrant transcripts can cause silencing. The trigger used here, AMP243, produces low levels of GFP siRNAs that can target GFP mRNA when expressed from the transgene. The cleavage products from these targets can then become templates for RDR6 activity. In the absence of RDR6 the low levels of AMP243 derived siRNAs are presumably not sufficient to ensure GFP silencing.

However, aberrant transcripts does not explain why cleaved fragments from RNA silencing of endogenous genes do not often become RDR6 templates, suggesting that RDR6 is able to distinguish between cleaved transcripts from different sources (Vaistij et al., 2002). For example, it has been suggested that the presence of gene related structure, such as the presence of introns, may protect endogenous genes from the action of RDR6. Genes without introns are more likely to undergo silencing and RDR6 amplification than genes with introns. It is thought that the presence of introns may act as a signal that a gene is endogenous. Christie et al. (2011) found that efficient splicing of introns from transgenes was required to limit the silencing of the transgene, not just the presence of an intron. Their work suggests that the act of splicing disrupts RDR6 function which prevents amplification of silencing.

Work has suggested that targets cleaved by a 22nt miRNA instead of a 21nt miRNA tend to act as templates for RDR6, possibly due to the interaction of the miRNA with AGO1 (Chen et al., 2010, Cuperus et al., 2010). However, these reports focus on cleavage by endogenous miRNAs and so it is not clear how these would work for transgene targets.

6.2 AMP243-triggered RdDM may be influenced by multiple-factors

One of the most interesting aspects of this work has been the observation that of the 40 lines that underwent AMP243 triggered PTGS, only 8 clearly underwent DNA methylation in the F1 generation. The other lines either did not undergo DNA methylation, or did so at a level that was not detectable by Southern blot. The possible reasons for this difference investigated here include; copy number, T-DNA integration structure, integration site and the presence of two histone H3 modifications. Additionally, the F2 AMP243 X 35S:GFP crosses displayed high levels of DNA methylation regardless of the methylation status in the F1 cross. The results of this body of work have not been straight forward to interpret due to the complex nature of transgene silencing. There has not been one key element discovered here that clearly labels a transgene for RdDM. However, this work, in conjunction with the literature, suggest that there may be several aspects that contribute to the observed variation in transgene silencing that may also contribute to the variation in AMP243-triggered DNA methylation witnessed here.

The work here suggests that some factors are more likely to affect RdDM of a transgene than others. For example, a high copy number and complex integration structure would account for 7 of the 8 lines that undergo DNA methylation and RNA silencing, whereas genomic location does not affect RdDM of transgenes in this study. These effects in a variety of plant species in systems where silencing had been triggered, have been observed by others. Studies have shown that transgenes with multiple copies tend to undergo RNA silencing (Schubert et al., 2004, Hobbs et al., 1993, Luo & Chen, 2007) and may also undergo DNA methylation, especially when the integration structures are complex (reviewed by Butaye et al., 2005, Stam et al., 1998). However, as discussed in chapter 5, line 71 does not fit with the connection between copy number and complex integration structure with AMP243 triggered DNA methylation as it has a copy number of one yet undergoes DNA methylation in the F1 cross.

Many of the features that have been investigated here in connection to AMP243-triggered RdDM are discussed in the literature with regards to spontaneous (or constitutive) transgene silencing. If any of the factors investigated here were strong enough to cause RdDM of the *GFP* transgene then we would have expected to observe methylation in the absence of the AMP243 trigger.

The effect of DNA methylation on PTGS was not investigated in this study to see if the methylation at the GFP loci in the 8 lines was influencing silencing, or if PTGS would have still occurred in these lines regardless of the methylation status. Wang and Waterhouse (2000) used demethylation treatment with 5-azacylidne on silencing lines to investigate if the loss of DNA

methylation resulted in the loss of PTGS. In their experiment it did not, providing confidence in their data that what they had observed was DNA methylation caused by RNA silencing and that RNA silencing could occur without the DNA methylation. However, work on mutants deficient for *met1* resulted in release of transgene PTGS for a portion of the progeny suggesting that DNA methylation may be required to aid the maintenance phase of PTGS at some loci (Morel et al., 2000). Additionally, it has been observed that mutants of *nrpe1* that fail to undergo DNA methylation are defective in the maintenance phase of PTGS (Eamens et al., 2008).

Another key observation in this body of work is that DNA methylation was increased in the F2 generation of AMP243 X 35S:GFP crosses. This led to the investigation of zygosity as a factor affecting AMP243-triggered RdDM. Due to the small sample size of the lines tested it was hard to draw any firm conclusions around this. Increased zygosity could influence RNA silencing due to doubled gene dosage. Additionally, there may be DNA-DNA interactions that could account for the increase in RdDM (de Carvalho et al., 1992, De Wilde et al., 2001, James et al., 2002, Velten et al., 2012). However, the increased levels of RdDM in the F2 may be the result of the lines passing through additional generations (Lunerova-Bedrichova et al., 2008, Kuhlmann et al., 2014). This explanation would explain why DNA methylation persisted to some extent in the lines out crossed to the Col-0 wild type. Further investigation of the effect of passing through generations on RdDM is a good candidate for future research generated from this project.

6.3 Histone modifications

Work into the differences in histone modifications before and after silencing demonstrates that, for the two histone modifications looked at here, there does not appear to be any changes correlating with AMP243 triggered RNA silencing. This was slightly surprising as the relationship between histone modifications, such as H3K9me2, and DNA methylation are well documented (Bernatavichute et al., 2008). It was assumed here that the relationship between H3K9me2 and RdDM may also be witnessed in AMP243-triggered RdDM (Law et al., 2013). However, the work by Law et al. (2013) links the H3K9me2 mark to Pol IV-dependent RdDM. It has not been established here if AMP243 trigger DNA methylation is dependent on Pol IV but it has been reported that silencing of the DAN2 line, as triggered by AMP243 is influenced by Pol IV (Herr et al., 2005). An association between H3K9me2 and AMP243 silencing was not seen here which makes it tempting to suggest that Pol IV is not involved. Further experimentation would be needed to corroborate this idea. A FLAG-tagged NRPD1 line could have been used in immunoprecipitation experiments to investigate this, which is why the generation of such a line was attempted in Chapter 4.

The absence of changes to histone modifications observed here raises the possibility that AMP243 triggered silencing may be a histone-independent form of RdDM. However, only two histones were investigated and so assertions into changes in to other histone modifications cannot be made. Therefore AMP243-triggered PTGS/ RdDM may involve the addition or removal of other histone marks. As discussed in Chapter 5, work by others has linked other histone modifications with RdDM which were not investigated here. For example, efficient RdDM can involve the removal of active histone marks such as H3K4me2 and H3K4me3 (Greenberg et al., 2013). The addition of repressive histone modifications, such as H3K27me1, as a result of RdDM has also been observed (Wierzbicki et al., 2008).

6.4 Attempting to investigate recruitment of RNA silencing proteins to targets using FLAG tagged lines.

This project also aimed to investigate the recruitment of RDR2, RDR6 and Pol V to target loci. Pol V is known to associate with target loci whereas RDR2 and RDR6 may not (Wierzbicki et al., 2008). Therefore it was always an unknown if work with RDR2:FLAG or RDR6:FLAG would give positive data. Sometimes negative data can be just as beneficial as positive data but is also harder to corroborate. For example, RDR2 may not associate with chromatin directly, although it does associate with proteins that do (Haag et al., 2012). Therefore ChIP experiments with RDR2 may not produce a positive result, but it would be hard to conclude that RDR2 definitely does not associate with DNA due to the lack of a positive control to validate the ChIP experiment with.

Unfortunately, work on RDR2:FLAG, RDR6:FLAG and NRPE1:FLAG remains incomplete due to the length of time taken to generate the RDR fusion proteins and further technical complications with the ChIP analysis. The FLAG lines were therefore not used in further experiments such as protein-protein based IP or protein-RNA experiments.

6.5 Limitations of the work and alternative techniques

With any project there are some limitations of the techniques utilized. The sample size of the lines that underwent DNA methylation in the F1 AMP243 cross was small. This often meant that firm conclusions could not be made as to the possible reasons for the differences in methylation statuses. However, if a clear pattern had been apparent then perhaps the numbers here would have been sufficient to validate these. Instead no clear patterns were seen for any of the areas of research and so a small sample number limited what could be concluded. A larger sample number may have provided clearer correlation between DNA methylation and the factors tested. For example the TAIL PCR did not give integration sites for all of the lines investigated and there

was a limited range of types of integration sites. If the order of work had been reversed, with the TAIL PCR conducted before the AMP243 work, then perhaps the limited range of integration sites would have been spotted and could have been corrected for. This may not have increased the numbers of lines that underwent DNA methylation though. However, the AMP243 work was conducted before the TAIL work due to the length of time crossing can take.

The speed at which plants can be crossed was a limiting factor to this project. Another trigger line was utilized in this project but this too depended on crossing as did the work with the FLAG lines. Other triggers of RNA silencing are available that do not depend on crossing. For example, some triggers, such as inverted repeats, can be transformed into a plant using the floral dip technique. These could be used in a dual transgene system with a transgene to investigate RNA silencing of this. However, to use this system to investigate the changes to transgene before and after silencing would be more complex than with crossing with a trigger. The transgene line would need to be established and characterised and then transformed again with the trigger. However, it may be complex to screen for plants containing both transgenes at single integrations sites for each transgene. This may not be strictly necessary as long as silencing was observed in the target transgene. Alternatively, there are inducible triggers of RNA silencing that would also allow investigation of transgene loci before and after silencing to be investigated. For example, the heat-shock inducible promoter HSP18.2 can be used in Arabidopsis to induce a trigger of silencing by exposing plants to heat (Masclaux et al., 2004). This would still involve a dual-transgene system but once a line with the trigger was established this could then be directly transform with the target transgene. Characterisation of the line with regards to the target transgene could be performed before and after silencing was triggered without the need for crossing. Other promoters that are induced by chemicals are also available (Gatz & Lenk).

One of the advantages of using GFP as a reporter is that it allows for easy screening of expression in living plants by visualization under UV-light. However, this has led to GFP expression being observed in this experiment as a binary factor, with expression either being absent or present. Many other papers have used reporter genes such as GFP, GUS and NPTII and have been able to investigate variation of the percentage changes in protein expression caused by RNA silencing events (Lunerova-Bedrichova et al., 2008, Schubert et al., 2004). This can lead to a more detailed analysis of data and therefore the mechanism of the action of silencing.

6.6 Other areas of research into increasing expression of transgenes and reducing variability

Work into silencing of transgenes is important due to its effect on producing genetically modified plants for both basic research and crop development. There are many other areas of research into reducing the variability of transgene silencing and increasing expression levels that have not been described here. One such area is that of Matrix Attachment Regions (MARs). MARs are conserved areas of eukaryotic genomes which are thought to attach to the proteinaceous matrix which may organise DNA within the nucleus (reviewed by Allen et al., 2000). The proposed action of MARs is to cause the formation of DNA loops that stick out from the chromatin between the two MARs. MARs were proposed to protect the genes within the loop from chromosomal effects and there are reports that the inclusion of MARs in T-DNA cassettes increases the expression rates of transgenes and reduces the levels of transgene variation (reviewed by Allen et al., 2000). There has also been evidence that MARs may protect regions from the action of RNA silencing in *cis*. One of the reasons MARs were not investigated here was due to the trans-acting nature of the AMP243 trigger.

Many attempts to increase transgenes expression in this manner involve the use of MARs from non-plant species. However, the conserved nature of MARs mean that it is possible that Arabidopsis also has its own MARs. The fine chromatin structure of the area surrounding the *35S:GFP* transgene insertions were not investigated here. It is possible that some of the transgenes here were in areas that were looped out away from condensed chromatin due to the presence of endogenous MARs upstream and downstream of the insertion site. However, without exactly knowing where MARs are in the Arabidopsis genome relative to the *35S:GFP* transgenes, any discussion here about the effects of MARs is purely speculative. All transgenic lines here underwent PTGS when crossed with AMP243 so if any transgenes had inserted into DNA loops formed by MARs then this did not prevent amplicon triggered RNA silencing, which would be expected from the literature (reviewed by Allen et al., 2000). Lines where the *35S:GFP* transgene was perhaps not looped out from the chromatin may have been more susceptible to DNA methylation after RNA silencing had been triggered by AMP243. This may have been due to proximity of other silenced elements which may have increased the ease of recruitment of RdDM proteins to the transgene. However, as with this theory regarding proximity to TEs, DNA methylation was prevalent in the F2 generation regardless of the F1 methylation status, which may lend more credence to transcript level or zygosity based theory of silencing variability.

In barley adding MARs to each side of the transgene cassette gave a two fold increase in expression of a transgene (Petersen et al., 2002). Butaye et al. (2004) have demonstrated that

inclusion of MARs from Chicken in a transgene cassette increases the expression of a GUS reported gene by eightfold when transformed into an *rdr6* mutant background. They demonstrated that although the MAR had no effect on transgene expression when transformed into a wild type background, there was an increase in GUS expression when transformed into *rdr6* mutants that surpassed transgene expression without MARs in *rdr6*. Butaye et al. (2004) propose that MARs act as enhancers of transgene expression and that the fact this is only witnessed in PTGS mutants also lends support to the theory of an expression threshold triggering RNA silencing.

6.7 Concluding remarks

Work investigating silencing of transgenes in higher plants has revealed many interesting aspects of the mechanisms of RNA silencing. It is becoming apparent that there are many areas of overlap between different RNA silencing pathways and perhaps the boundaries between the pathways are not as clearly defined as first thought. Here work investigating variation in a transgene's ability to support PTGS or RdDM has been presented. The use of the AMP243 trigger line has been advantageous as it has allowed differences before and after RNA silencing to be investigated. The results from this investigation strongly suggest that location does not influence either AMP243-triggered RDR6-dependent PTGS or RdDM, answering the original aim of the project. This result led to other areas of investigation with the view to explaining the variation in transgenes to support either PTGS or RdDM. Whilst complex integration structure and high copy number seem the most likely source of variation, exceptions to both rules were present in this data.

This work has laid the foundations for further research into the effects of generation on AMP243 triggered RdDM. The project has also hinted at the possibility of a histone-independent form of silencing and suggests a path of further experimentation to validate this.

This work clearly demonstrates the complexity of issues surrounding RNA silencing pathways and variation of transgene expression that still need to be explained.

Appendix 1. Chi square analysis of T1 35S:GFP lines

Chi square analysis used to find 35S:GFP lines that are likely to have a single site of insertion by using a 3:1 ratio of GFP expression: no-expression.

To be under p=0.05 the $\chi 2$ value had to be under 3.841 (df=1) for the ratio to be 3:1

π	Class	Observed	Expected	(O-E)^2	(O-E)^2/E	χ2	df	3:1?
1	Red	31	30.25	0.5625	0.018595	0.024793	1	Yes
	Green	90	90.75	0.5625	0.006198			
2	Red	15	18	9	0.5	0.666667	1	Yes
	Green	57	54	9	0.166667			
3	Red	15	13.5	2.25	0.166667	0.222222	1	Yes
	Green	39	40.5	2.25	0.055556			
6	Red	20	28	64	2.285714	3.047619	1	Yes
	Green	92	84	64	0.761905			
7	Red	15	17.25	5.0625	0.293478	0.391304	1	Yes
	Green	54	51.75	5.0625	0.097826			
9	Red	20	16.25	14.0625	0.865385	1.153846	1	Yes
	Green	45	48.75	14.0625	0.288462			
10	Red	14	17.25	10.5625	0.612319	0.816425	1	Yes
	Green	55	51.75	10.5625	0.204106			
11	Red	20	11.5	72.25	6.282609	8.376812	1	No
	Green	26	34.5	72.25	2.094203			
12	Red	10	12.25	5.0625	0.413265	0.55102	1	Yes
	Green	39	36.75	5.0625	0.137755			
13	Red	17	18.25	1.5625	0.085616	0.114155	1	Yes
	Green	56	54.75	1.5625	0.028539			
14	Red	21	16.75	18.0625	1.078358	1.437811	1	Yes
	Green	46	50.25	18.0625	0.359453			
15	Red	45	25.5	380.25	14.91176	19.88235	1	No
	Green	57	76.5	380.25	4.970588			
16	Red	33	34.25	1.5625	0.04562	0.060827	1	Yes
	Green	104	102.75	1.5625	0.015207			
17	Red	29	31.5	6.25	0.198413	0.26455	1	Yes
	Green	97	94.5	6.25	0.066138			
18	Red	24	29	25	0.862069	1.149425	1	Yes
	Green	92	87	25	0.287356			
19	Red	19	22.75	14.0625	0.618132	0.824176	1	Yes
	Green	72	68.25	14.0625	0.206044			
20	Red	40	20.75	370.5625	17.85843	23.81124	1	No
	Green	43	62.25	370.5625	5.952811			

22	Red	17	14.5	6.25	0.431034	0.574713	1	Yes
	Green	41	43.5	6.25	0.143678			
23	Red	14	14.75	0.5625	0.038136	0.050847	1	Yes
	Green	45	44.25	0.5625	0.012712			
25	Red	24	22.25	3.0625	0.13764	0.183521	1	Yes
	Green	65	66.75	3.0625	0.04588			
26	Red	26	23	9	0.391304	0.521739	1	Yes
	Green	66	69	9	0.130435			
28	Red	7	14.25	52.5625	3.688596	4.918129	1	No
	Green	50	42.75	52.5625	1.229532			
29	Red	18	15.25	7.5625	0.495902	0.661202	1	Yes
	Green	43	45.75	7.5625	0.165301			
31	Red	44	19.75	588.0625	29.77532	39.70042	1	No
	Green	35	59.25	588.0625	9.925105			
32	Red	10	11	1	0.090909	0.121212	1	Yes
	Green	34	33	1	0.030303			
33	Red	8	8	0	0	0	1	Yes
	Green	24	24	0	0			
34	Red	16	10.5	30.25	2.880952	3.84127	1	Yes
	Green	26	31.5	30.25	0.960317			
35	Red	29	19.75	85.5625	4.332278	5.776371	1	No
	Green	50	59.25	85.5625	1.444093			
36	Red	28	23.25	22.5625	0.97043	1.293907	1	Yes
	Green	65	69.75	22.5625	0.323477			
37	Red	17	11.5	30.25	2.630435	3.507246	1	Yes
	Green	29	34.5	30.25	0.876812			
38	Red	21	18	9	0.5	0.666667	1	Yes
	Green	51	54	9	0.166667			
39	Red	16	13.5	6.25	0.462963	0.617284	1	Yes
	Green	38	40.5	6.25	0.154321			
40	Red	15	10.75	18.0625	1.680233	2.24031	1	Yes
	Green	28	32.25	18.0625	0.560078			
42	Red	26	13.75	150.0625	10.91364	14.55152	1	No
	Green	29	41.25	150.0625	3.637879			
43	Red	16	13	9	0.692308	0.923077	1	Yes
	Green	36	39	9	0.230769			
44	Red	7	5.75	1.5625	0.271739	0.362319	1	Yes
	Green	16	17.25	1.5625	0.09058			
45	Red	15	14	1	0.071429	0.095238	1	Yes
	Green	41	42	1	0.02381			
47	Red	20	15.5	20.25	1.306452	1.741935	1	Yes
	Green	42	46.5	20.25	0.435484			

49	Red	70	20	2500	125	166.6667	1	No
	Green	10	60	2500	41.66667			
50	Red	11	10	1	0.1	0.133333	1	Yes
	Green	29	30	1	0.033333			
51	Red	23	21.75	1.5625	0.071839	0.095785	1	Yes
	Green	64	65.25	1.5625	0.023946			
52	Red	10	8.25	3.0625	0.371212	0.494949	1	Yes
	Green	23	24.75	3.0625	0.123737			
53	Red	13	9	16	1.777778	2.37037	1	Yes
	Green	23	27	16	0.592593			
55	Red	19	14.5	20.25	1.396552	1.862069	1	Yes
	Green	39	43.5	20.25	0.465517			
56	Red	20	14.75	27.5625	1.868644	2.491525	1	Yes
	Green	39	44.25	27.5625	0.622881			
57	Red	16	11.25	22.5625	2.005556	2.674074	1	Yes
	Green	29	33.75	22.5625	0.668519			
58	Red	43	38.5	20.25	0.525974	0.701299	1	Yes
	Green	111	115.5	20.25	0.175325			
60	Red	34	30.25	14.0625	0.464876	0.619835	1	Yes
	Green	87	90.75	14.0625	0.154959			
63	Red	36	25.5	110.25	4.323529	5.764706	1	No
	Green	66	76.5	110.25	1.441176			
64	Red	29	22.75	39.0625	1.717033	2.289377	1	Yes
	Green	62	68.25	39.0625	0.572344			
65	Red	28	24.5	12.25	0.5	0.666667	1	Yes
	Green	70	73.5	12.25	0.166667			
66	Red	26	25.25	0.5625	0.022277	0.029703	1	Yes
	Green	75	75.75	0.5625	0.007426			
67	Red	21	19	4	0.210526	0.280702	1	Yes
	Green	55	57	4	0.070175			
68	Red	58	26.5	992.25	37.4434	49.92453	1	No
	Green	48	79.5	992.25	12.48113			
69	Red	19	16.25	7.5625	0.465385	0.620513	1	Yes
	Green	46	48.75	7.5625	0.155128			
70	Red	25	22	9	0.409091	0.545455	1	Yes
	Green	63	66	9	0.136364			
71	Red	23	19.75	10.5625	0.53481	0.71308	1	Yes
	Green	56	59.25	10.5625	0.17827			
72	Red	18	19.5	2.25	0.115385	0.153846	1	Yes
	Green	60	58.5	2.25	0.038462			
73	Red	31	20	121	6.05	8.066667	1	No
	Green	49	60	121	2.016667			

74	Red	14	17	9	0.529412	0.705882	1	Yes
	Green	54	51	9	0.176471			
75	Red	14	14.75	0.5625	0.038136	0.050847	1	Yes
	Green	45	44.25	0.5625	0.012712			
76	Red	15	14.5	0.25	0.017241	0.022989	1	Yes
	Green	43	43.5	0.25	0.005747			
77	Red	11	35	576	16.45714	21.94286	1	No
	Green	129	105	576	5.485714			
78	Red	41	48.25	52.5625	1.089378	1.452504	1	Yes
	Green	152	144.75	52.5625	0.363126			
79	Red	36	29.25	45.5625	1.557692	2.076923	1	Yes
	Green	81	87.75	45.5625	0.519231			
80	Red	33	33	0	0	0	1	Yes
	Green	99	99	0	0			
82	Red	17	21.5	20.25	0.94186	1.255814	1	Yes
	Green	69	64.5	20.25	0.313953			
83	Red	14	17.75	14.0625	0.792254	1.056338	1	Yes
	Green	57	53.25	14.0625	0.264085			
84	Red	7	14.5	56.25	3.87931	5.172414	1	No
	Green	51	43.5	56.25	1.293103			
85	Red	32	29	9	0.310345	0.413793	1	Yes
	Green	84	87	9	0.103448			
86	Red	25	30.25	27.5625	0.911157	1.214876	1	Yes
	Green	96	90.75	27.5625	0.303719			
87	Red	37	28.75	68.0625	2.367391	3.156522	1	Yes
	Green	78	86.25	68.0625	0.78913			
88	Red	32	31.75	0.0625	0.001969	0.002625	1	Yes
	Green	95	95.25	0.0625	0.000656			
89	Red	36	28.5	56.25	1.973684	2.631579	1	Yes
	Green	78	85.5	56.25	0.657895			
93	Red	26	24	4	0.166667	0.222222	1	Yes
	Green	70	72	4	0.055556			
95	Red	35	29.25	33.0625	1.130342	1.507123	1	Yes
	Green	82	87.75	33.0625	0.376781			
97	Red	19	16.25	7.5625	0.465385	0.620513	1	Yes
	Green	46	48.75	7.5625	0.155128			
98	Red	14	20.25	39.0625	1.929012	2.572016	1	Yes
	Green	67	60.75	39.0625	0.643004			
99	Red	9	16.25	52.5625	3.234615	4.312821	1	No
	Green	56	48.75	52.5625	1.078205			
100	Red	23	21	4	0.190476	0.253968	1	Yes
	Green	61	63	4	0.063492			

101	Red	5	17.75	162.5625	9.158451	12.21127	1	No
	Green	66	53.25	162.5625	3.052817			
103	Red	40	37.25	7.5625	0.20302	0.270694	1	Yes
	Green	109	111.75	7.5625	0.067673			
106	Red	41	39.5	2.25	0.056962	0.075949	1	Yes
	Green	117	118.5	2.25	0.018987			
107	Red	20	29	81	2.793103	3.724138	1	Yes
	Green	96	87	81	0.931034			
108	Red	46	41.5	20.25	0.487952	0.650602	1	Yes
	Green	120	124.5	20.25	0.162651			
109	Red	62	41.25	430.5625	10.43788	13.91717	1	No
	Green	103	123.75	430.5625	3.479293			
110	Red	5	17.75	162.5625	9.158451	12.21127	1	No
	Green	66	53.25	162.5625	3.052817			
113	Red	60	38	484	12.73684	16.98246	1	No
	Green	92	114	484	4.245614			
114	Red	31	29.25	3.0625	0.104701	0.139601	1	Yes
	Green	86	87.75	3.0625	0.0349			
115	Red	20	22	4	0.181818	0.242424	1	Yes
	Green	68	66	4	0.060606			
116	Red	21	23.25	5.0625	0.217742	0.290323	1	Yes
	Green	72	69.75	5.0625	0.072581			
117	Red	21	25.5	20.25	0.794118	1.058824	1	No
	Green	81	76.5	20.25	0.264706			
118	Red	28	27	1	0.037037	0.049383	1	Yes
	Green	80	81	1	0.012346			
120	Red	22	23.75	3.0625	0.128947	0.17193	1	Yes
	Green	73	71.25	3.0625	0.042982			
122	Red	9	12.25	10.5625	0.862245	1.14966	1	Yes
	Green	40	36.75	10.5625	0.287415			
128	Red	15	11.75	10.5625	0.898936	1.198582	1	Yes
	Green	32	35.25	10.5625	0.299645			
129	Red	22	20	4	0.2	0.266667	1	Yes
	Green	58	60	4	0.066667			

List of Abbreviations

AGO	Argonaute
ChIP	Chromatin Immunoprecipitation
CLSY1	Classy 1
CMT3	Chromomethylase 3
DCL	DICER-like enzymes
DDM1	Decrease in DNA Methylation 1
DMS3	Defective Meristem Silencing 3
DNA	Deoxyribonucleic acid
DRD1	Defective in RNA-Directed DNA Methylation
DRM2	Domains Rearranged Methyltransferase 2
dsDNA	Double-stranded DNA
dsRNAs	Double-stranded RNA
epiRILs	epigenetic recombination inbred lines
EVD	Evadé
GFP	Green Florescence Protein
HEN1	Hua enhancer 1
KTF1	Kow Domain-Containing Transcription Factor 1
KYP	Kryptonite
LB	Luria Broth
MAR	Matrix Attachment Region
MET1	Methyltransferase 1
miRNAs	Micro RNAs
MORC	Member of the microrchidia (MORC) family of ATPases.
mRNA	Messenger RNA
MS	Murashige and Shoog
nat – si RNAs	Natural-antisense siRNAs
NERD	Needed for RDR2 independent DNA Methylation
NRP	Nuclear RNA Polymerase
nt	Nucleotide
PCR	Polymerase Chain Reaction
phasiRNA	Phased, secondary, small interfering RNAs
Pol II, IV, V	RNAs Polymerases

PTGS	Post-transcriptional gene silencing
RdDM	RNA-directed DNA methylation
RDM4	RNA-directed DNA Methylation 4
RDRs	RNA-dependent RNA polymerases
RISC	RNA-induced silencing complexes
RITS	RNA-induced transcriptional silencing
RNA	Ribonucleic acid
SDS3	SUPPRESSOR OF GENE SILENCING 3
SHH1	Sawadee Homo Domain Homologue
siRNAs	small-interfering RNAs
sRNAs	Small RNAs
ssRNA	Single stranded RNA
SUVH	SU(VAR) 3 -9 HOMOLOG 2
TAIL PCR	Thermal Asymmetric Interlaced PCR
TAS gene	Trans-acting siRNA gene
ta-si RNAs	Trans-acting siRNAs
TE	Transposable Elements
VIM	Variant in Methylation

References

Allen GC, Spiker S, Thompson WF, 2000. Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Molecular Biology* **43**, 361-76.

Alonso JM, Stepanova AN, Leisse TJ*, et al.*, 2003. Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* **301**, 653-7.

Altschul SF, Madden TL, Schaffer AA*, et al.*, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402.

Aufsatz W, Mette MF, Van Der Winden J, Matzke M, Matzke AJM, 2002. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *Embo Journal* **21**, 6832-41.

Axtell MJ, Jan C, Rajagopalan R, Bartel DP, 2006. A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**, 565-77.

Bent AF, Clough SJ, 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal* **16**, 735-43.

Bent AF, Desfeux C, Clough SJ, 2000. Female reproductive tissues are the primary target of Agrobacterium-mediated transformation by the Arabidopsis floral-dip method. *Plant Physiology* **123**, 895-904.

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE, 2008. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS One* **3**, e3156.

Blevins T, Pontvianne F, Cocklin R*, et al.*, 2014. A two-step process for epigenetic inheritance in Arabidopsis. *Mol Cell* **54**, 30-42.

Brabbs TR, He Z, Hogg K*, et al.*, 2013. The stochastic silencing phenotype of Arabidopsis morc6 mutants reveals a role in efficient RNA-directed DNA methylation. *Plant J* **75**, 836-46.

Brosnan CA, Mitter N, Christie M, Smith NA, Waterhouse PM, Carroll BJ, 2007. Nuclear gene silencing directs reception of long-distance mRNA silencing in Arabidopsis. *Proc Natl Acad Sci U S A* **104**, 14741-6.

Bubner B, Baldwin IT, 2004. Use of real-time PCR for determining copy number and zygosity in transgenic plants. *Plant Cell Reports* **23**, 263-71.

Butaye KMJ, Cammue BPA, Delaure SL, De Bolle MFC, 2005. Approaches to minimize variation of transgene expression in plants. *Molecular Breeding* **16**, 79-91.

Butaye KMJ, Goderis IJWM, Wouters PFJ*, et al.*, 2004. Stable high-level transgene expression in Arabidopsis thaliana using gene silencing mutants and matrix attachment regions. *Plant Journal* **39**, 440-9.

Cao X, Aufsatz W, Zilberman D*, et al.*, 2003. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol* **13**, 2212-7.

Chen HM, Chen LT, Patel K, Li YH, Baulcombe DC, Wu SH, 2010. 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci U S A* **107**, 15269-74.

Chen X, 2004. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science* **303**, 2022-5.

Chen X, 2009. Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol* **25**, 21-44.

Christie M, Croft LJ, Carroll BJ, 2011. Intron splicing suppresses RNA silencing in Arabidopsis. *Plant J* **68**, 159-67.

Cokus SJ, Feng SH, Zhang XY*, et al.*, 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215-9.

Creasey KM, Zhai J, Borges F*, et al.*, 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature* **508**, 411-5.

Cui X, Cao X, 2014. Epigenetic regulation and functional exaptation of transposable elements in higher plants. *Curr Opin Plant Biol* **21C**, 83-8.

Cuperus JT, Carbonell A, Fahlgren N*, et al.*, 2010. Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat Struct Mol Biol* **17**, 997-1003.

Curaba J, Chen X, 2008. Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J Biol Chem* **283**, 3059-66.

Dalmay T, Hamilton A, Mueller E, Baulcombe DC, 2000a. Potato virus X amplicons in arabidopsis mediate genetic and epigenetic gene silencing. *Plant Cell* **12**, 369-79.

Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC, 2000b. An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**, 543-53.

Dalmay T, Horsefield R, Braunstein TH, Baulcombe DC, 2001. SDE3 encodes an RNA helicase required for posttranscriptional gene silencing in Arabidopsis. *Embo Journal* **20**, 2069-77.

Daxinger L, Hunter B, Sheikh M*, et al.*, 2008. Unexpected silencing effects from T-DNA tags in Arabidopsis. *Trends Plant Sci* **13**, 4-6.

De Carvalho F, Gheysen G, Kushnir S, Van Montagu M, Inze D, Castresana C, 1992. Suppression of beta-1,3-glucanase transgene expression in homozygous plants. *EMBO J* **11**, 2595-602.

De Wilde C, Podevin N, Windels P, Depicker A, 2001. Silencing of antibody genes in plants with single-copy transgene inserts as a result of gene dosage effects. *Mol Genet Genomics* **265**, 647-53.

Ding SW, 2010. RNA-based antiviral immunity. *Nat Rev Immunol* **10**, 632-44.

Douet J, Tutois S, Tourmente S, 2009. A Pol V-mediated silencing, independent of RNA-directed DNA methylation, applies to 5S rDNA. *PLoS Genet* **5**, e1000690.

Du J, Zhong X, Bernatavichute YV*, et al.*, 2012. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**, 167-80.

Eamens A, Vaistij FE, Jones L, 2008. NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in Arabidopsis. *Plant J* **55**, 596-606.

Eichten SR, Ellis NA, Makarevitch I*, et al.*, 2012. Spreading of heterochromatin is limited to specific families of maize retrotransposons. *PLoS Genet* **8**, e1003127.

English JJ, Mueller E, Baulcombe DC, 1996. Suppression of Virus Accumulation in Transgenic Plants Exhibiting Silencing of Nuclear Genes. *Plant Cell* **8**, 179-88.

Fei Q, Xia R, Meyers BC, 2013. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* **25**, 2400-15.

Fischer U, Kuhlmann M, Pecinka A, Schmidt R, Mette MF, 2008. Local DNA features affect RNA-directed transcriptional gene silencing and DNA methylation. *Plant Journal* **53**, 1-10.

Gao Z, Liu HL, Daxinger L*, et al.*, 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**, 106-9.

Garcia-Ruiz H, Takeda A, Chapman EJ*, et al.*, 2010. Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. *Plant Cell* **22**, 481-96.

Garcia D, Garcia S, Pontier D*, et al.*, 2012. Ago Hook and RNA Helicase Motifs Underpin Dual Roles for SDE3 in Antiviral Defense and Silencing of Nonconserved Intergenic Regions. *Molecular Cell* **48**, 109-20.

Gatz C, Lenk I. Promoters that respond to chemical inducers. *Trends in Plant Science* **3**, 352-8.

Gazzani S, Lawrenson T, Woodward C, Headon D, Sablowski R, 2004. A link between mRNA turnover and RNA interference in Arabidopsis. *Science* **306**, 1046-8.

Gent JI, Ellis NA, Guo L*, et al.*, 2013. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res* **23**, 628-37.

Ghildiyal M, Zamore PD, 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**, 94-108.

Goodwin J, Chapman K, Swaney S, Parks TD, Wernsman EA, Dougherty WG, 1996. Genetic and biochemical dissection of transgenic RNA-mediated virus resistance. *Plant Cell* **8**, 95-105.

Goto DB, Nakayama J, 2012. RNA and epigenetic silencing: insight from fission yeast. *Dev Growth Differ* **54**, 129-41.

Greenberg MV, Ausin I, Chan SW*, et al.*, 2011. Identification of genes required for de novo DNA methylation in Arabidopsis. *Epigenetics* **6**.

Greenberg MVC, Deleris A, Hale CJ, Liu A, Feng SH, Jacobsen SE, 2013. Interplay between Active Chromatin Marks and RNA-Directed DNA Methylation in Arabidopsis thaliana. *Plos Genetics* **9**.

Haag JR, Ream TS, Marasco M*, et al.*, 2012. In Vitro Transcription Activities of Pol IV, Pol V, and RDR2 Reveal Coupling of Pol IV and RDR2 for dsRNA Synthesis in Plant RNA Silencing. *Molecular Cell* **48**, 811-8.

Havecker ER, Wallbridge LM, Hardcastle TJ*, et al.*, 2010. The Arabidopsis RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. *Plant Cell* **22**, 321-34.

He XJ, Hsu YF, Zhu S*, et al.*, 2009. An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell* **137**, 498-508.

Henderson IR, Zhang X, Lu C*, et al.*, 2006. Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* **38**, 721-5.

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC, 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118-20.

Hetzl J, Foerster AM, Raidl G, Mittelsten Scheid O, 2007. CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J* **51**, 526-36.

Hirochika H, Okamoto H, Kakutani T, 2000. Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation. *Plant Cell* **12**, 357-69.

Hobbs SLA, Warkentin TD, Delong CMO, 1993. Transgene Copy Number Can Be Positively or Negatively Associated with Transgene Expression. *Plant Molecular Biology* **21**, 17-26.

Hoffer P, Ivashuta S, Pontes O*, et al.*, 2011. Posttranscriptional gene silencing in nuclei. *Proc Natl Acad Sci U S A* **108**, 409-14.

Huang L, Jones AM, Searle I*, et al.*, 2009. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol* **16**, 91-3.

Iglesias VA, Moscone EA, Papp I*, et al.*, 1997. Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. *Plant Cell* **9**, 1251-64.

Ingelbrecht I, Van Houdt H, Van Montagu M, Depicker A, 1994. Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc Natl Acad Sci U S A* **91**, 10502-6.

Jakowitsch J, Papp I, Moscone EA, Van Der Winden J, Matzke M, Matzke AJM, 1999. Molecular and cytogenetic characterization of a transgene locus that induces silencing and methylation of homologous promoters in trans. *Plant Journal* **17**, 131-40.

James VA, Avart C, Worland B, Snape JW, Vain P, 2002. The relationship between homozygous and hemizygous transgene expression levels over generations in populations of transgenic rice plants. *Theor Appl Genet* **104**, 553-61.

Jauvion V, Rivard M, Bouteiller N, Elmayan T, Vaucheret H, 2012. RDR2 partially antagonizes the production of RDR6-dependent siRNA in sense transgene-mediated PTGS. *PLoS One* **7**, e29785.

Johnson L, Mollah S, Garcia BA*, et al.*, 2004. Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res* **32**, 6511-8.

Johnson LM, Du J, Hale CJ*, et al.*, 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**, 124-8.

Jones-Rhoades MW, Bartel DP, 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* **14**, 787-99.

Jones AL, Thomas CL, Maule AJ, 1998. De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus. *EMBO J* **17**, 6385-93.

Jones L, Hamilton AJ, Voinnet O, Thomas CL, Maule AJ, Baulcombe DC, 1999. RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell* **11**, 2291-301.

Kawabe A, Hansson B, Hagenblad J, Forrest A, Charlesworth D, 2006. Centromere locations and associated chromosome rearrangements in Arabidopsis lyrata and A. thaliana. *Genetics* **173**, 1613-9.

Khvorova A, Reynolds A, Jayasena SD, 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209-16.

Kononov ME, Bassuner B, Gelvin SB, 1997. Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: Evidence for multiple complex patterns of integration. *Plant Journal* **11**, 945-57.

Kuhlmann M, Finke A, Mascher M, Mette MF, 2014. DNA methylation maintenance consolidates RNA-directed DNA methylation and transcriptional gene silencing over generations in Arabidopsis thaliana. *Plant Journal* **80**, 269-81.

Kumakura N, Takeda A, Fujioka Y, Motose H, Takano R, Watanabe Y, 2009. SGS3 and RDR6 interact and colocalize in cytoplasmic SGS3/RDR6-bodies. *FEBS Lett* **583**, 1261-6.

Kunova A, Zubko E, Meyer P, 2012. A pair of partially overlapping Arabidopsis genes with antagonistic circadian expression. *Int J Plant Genomics* **2012**, 349527.

Law JA, Ausin I, Johnson LM*, et al.*, 2010. A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in Arabidopsis. *Curr Biol* **20**, 951-6.

Law JA, Du J, Hale CJ*, et al.*, 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385-9.

Law JA, Jacobsen SE, 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**, 204-20.

Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE, 2011. SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV. *PLoS Genet* **7**, e1002195.

Le Masson I, Jauvion V, Bouteiller N, Rivard M, Elmayan T, Vaucheret H, 2012. Mutations in the Arabidopsis H3K4me2/3 Demethylase JMJ14 Suppress Posttranscriptional Gene Silencing by Decreasing Transgene Transcription. *Plant Cell* **24**, 3603-12.

Lechtenberg B, Schubert D, Forsbach A, Gils M, Schmidt R, 2003. Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. *Plant Journal* **34**, 507-17.

Lewsey M, Robertson FC, Canto T, Palukaitis P, Carr JP, 2007. Selective targeting of miRNA-regulated plant development by a viral counter-silencing protein. *Plant J* **50**, 240-52.

Li CF, Pontes O, El-Shami M*, et al.*, 2006. An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. *Cell* **126**, 93-106.

Lindroth AM, Shultis D, Jasencakova Z*, et al.*, 2004. Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *Embo Journal* **23**, 4146-55.

Lisch D, Bennetzen JL, 2011. Transposable element origins of epigenetic gene regulation. *Curr Opin Plant Biol* **14**, 156-61.

Liu YG, Mitsukawa N, Oosumi T, Whittier RF, 1995. Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J* **8**, 457-63.

Livak KJ, Schmittgen TD, 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**, 402-8.

Lorkovic ZJ, Naumann U, Matzke AJM, Matzke M, 2012. Involvement of a GHKL ATPase in RNA-Directed DNA Methylation in Arabidopsis thaliana. *Current Biology* **22**, 933-8.

Lunerova-Bedrichova J, Bleys A, Fojtova M, Khaitova L, Depicker A, Kovarik A, 2008. Trans-generation inheritance of methylation patterns in a tobacco transgene following a post-transcriptional silencing event. *Plant J* **54**, 1049-62.

Luo Z, Chen Z, 2007. Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis. *Plant Cell* **19**, 943-58.

Mallory A, Vaucheret H, 2010. Form, function, and regulation of ARGONAUTE proteins. *Plant Cell* **22**, 3879-89.

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-+.

Masclaux F, Charpenteau M, Takahashi T, Pont-Lezica R, Galaud JP, 2004. Gene silencing using a heat-inducible RNAi system in Arabidopsis. *Biochemical and Biophysical Research Communications* **321**, 364-9.

Masclaux FG, Pont-Lezica R, Galaud JP, 2005. Relationship between allelic state of T-DNA and DNA methylation of chromosomal integration region in transformed Arabidopsis thaliana plants. *Plant Molecular Biology* **58**, 295-303.

Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J, 2007. Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell* **130**, 851-62.

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

Mcelver J, Tzafrir I, Aux G*, et al.*, 2001. Insertional mutagenesis of genes required for seed development in Arabidopsis thaliana. *Genetics* **159**, 1751-63.

Melnyk CW, Molnar A, Baulcombe DC, 2011. Intercellular and systemic movement of RNA silencing signals. *EMBO J* **30**, 3553-63.

Meyer P, 2011. DNA methylation systems and targets in plants. *FEBS Lett* **585**, 2008-15.

Meyer P, 2013. Transgenes and their contributions to epigenetic research. *Int J Dev Biol* **57**, 509-15.

Meza TJ, Stangeland B, Mercy IS*, et al.*, 2002. Analyses of single-copy Arabidopsis T-DNA-transformed lines show that the presence of vector backbone sequences, short inverted repeats and DNA methylation is not sufficient or necessary for the induction of transgene silencing. *Nucleic Acids Research* **30**, 4556-66.

Moissiard G, Cokus SJ, Cary J*, et al.*, 2012. MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing. *Science* **336**, 1448-51.

Molnar A, Melnyk C, Baulcombe DC, 2011. Silencing signals in plants: a long journey for small RNAs. *Genome Biol* **12**, 215.

Morel JB, Mourrain P, Beclin C, Vaucheret H, 2000. DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Curr Biol* **10**, 1591-4.

Mosher RA, Schwach F, Studholme D, Baulcombe DC, 2008. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci U S A* **105**, 3145-50.

Mourrain P, Beclin C, Elmayan T*, et al.*, 2000. Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533-42.

Mourrain P, Van Blokland R, Kooter JM, Vaucheret H, 2007. A single transgene locus triggers both transcriptional and post-transcriptional silencing through double-stranded RNA production. *Planta* **225**, 365-79.

Naumann U, Daxinger L, Kanno T*, et al.*, 2011. Genetic evidence that DNA methyltransferase DRM2 has a direct catalytic role in RNA-directed DNA methylation in Arabidopsis thaliana. *Genetics* **187**, 977-9.

Nuthikattu S, Mccue AD, Panda K*, et al.*, 2013. The Initiation of Epigenetic Silencing of Active Transposable Elements Is Triggered by RDR6 and 21-22 Nucleotide Small Interfering RNAs. *Plant Physiology* **162**, 116-31.

Otagaki S, Kawai M, Masuta C, Kanazawa A, 2011. Size and positional effects of promoter RNA segments on virus-induced RNA-directed DNA methylation and transcriptional gene silencing. *Epigenetics* **6**, 681-91.

Park W, Li J, Song R, Messing J, Chen X, 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol* **12**, 1484-95.

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS, 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev* **18**, 2368-79.

Petersen K, Leah R, Knudsen S, Cameron-Mills V, 2002. Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance of transgene expression in barley. *Plant Molecular Biology* **49**, 45-58.

Pontes O, Li CF, Nunes PC*, et al.*, 2006. The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79-92.

Pontier D, Picart C, Roudier F*, et al.*, 2012. NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in Arabidopsis. *Mol Cell* **48**, 121-32.

Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ, 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**, 1008-12.

Ream TS, Haag JR, Wierzbicki AT*, et al.*, 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell* **33**, 192-203.

Ruiz MT, Voinnet O, Baulcombe DC, 1998. Initiation and maintenance of virus-induced gene silencing. *Plant Cell* **10**, 937-46.

Schneeberger RG, Zhang K, Tatarinova T*, et al.*, 2005. Agrobacterium T-DNA integration in Arabidopsis is correlated with DNA sequence compositions that occur frequently in gene promoter regions. *Funct Integr Genomics* **5**, 240-53.

Schubert D, Lechtenberg B, Forsbach A, Gils M, Bahadur S, Schmidt R, 2004. Silencing in Arabidopsis T-DNA transformants: The predominant role of a gene-specific RNA sensing mechanism versus position effects. *Plant Cell* **16**, 2561-72.

Schwach F, Vaistij FE, Jones L, Baulcombe DC, 2005. An RNA-dependent RNA polymerase prevents meristem invasion by potato virus X and is required for the activity but not the production of a systemic silencing signal. *Plant Physiol* **138**, 1842-52.

Sijen T, Wellink J, Hiriart JB, Van Kammen A, 1996. RNA-Mediated Virus Resistance: Role of Repeated Transgenes and Delineation of Targeted Regions. *Plant Cell* **8**, 2277-94.

Skarn M, Eike MC, Meza TJ, Mercy IS, Jakobsen KS, Aalen RB, 2006. An inverted repeat transgene with a structure that cannot generate double-stranded RNA, suffers silencing independent of DNA methylation. *Transgenic Research* **15**, 489-500.

Slotkin RK, Vaughn M, Borges F*, et al.*, 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461-72.

Smith LM, Pontes O, Searle I*, et al.*, 2007. An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell* **19**, 1507-21.

Stam M, Viterbo A, Mol JNM, Kooter JM, 1998. Position-dependent methylation and transcriptional silencing of transgenes in inverted T-DNA repeats: Implications for posttranscriptional silencing of homologous host genes in plants. *Molecular and Cellular Biology* **18**, 6165-77.

Sugiyama T, Cam H, Verdel A, Moazed D, Grewal SIS, 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 152-7.

Tonti-Filippini J, 2008. http://www.annoj.org. In. (2014.)

Vaistij FE, Jones L, Baulcombe DC, 2002. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell* **14**, 857-67.

Vanblokland R, Vandergeest N, Mol JNM, Kooter JM, 1994. Transgene-Mediated Suppression of Chalcone Synthase Expression in Petunia-Hybrida Results from an Increase in Rna Turnover. *Plant Journal* **6**, 861-77.

Vaucheret H, Vazquez F, Crete P, Bartel DP, 2004. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev* **18**, 1187-97.

Vazquez F, Legrand S, Windels D, 2010. The biosynthetic pathways and biological scopes of plant small RNAs. *Trends Plant Sci* **15**, 337-45.

Vazquez F, Vaucheret H, Rajagopalan R*, et al.*, 2004. Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell* **16**, 69-79.

Velten J, Cakir C, Youn E, Chen J, Cazzonelli CI, 2012. Transgene silencing and transgene-derived siRNA production in tobacco plants homozygous for an introduced AtMYB90 construct. *PLoS One* **7**, e30141.

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA, 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833-7.

Wang MB, Waterhouse PM, 2000. High-efficiency silencing of a beta-glucuronidase gene in rice is correlated with repetitive transgene structure but is independent of DNA methylation. *Plant Mol Biol* **43**, 67-82.

Wang XB, Wu Q, Ito T*, et al.*, 2010. RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **107**, 484-9.

Wang Y, Tang X, Cheng Z, Mueller L, Giovannoni J, Tanksley SD, 2006. Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. *Genetics* **172**, 2529-40.

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

Wierzbicki AT, Haag JR, Pikaard CS, 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**, 635-48.

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS, 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**, 630-4.

Willmann MR, Endres MW, Cook RT, Gregory BD, 2011. The Functions of RNA-Dependent RNA Polymerases in Arabidopsis. *Arabidopsis Book* **9**, e0146.

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

Xie Z, Johansen LK, Gustafson AM*, et al.*, 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* **2**, E104.

Yang Z, Ebright YW, Yu B, Chen X, 2006. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res* **34**, 667-75.

Yoshikawa M, Peragine A, Park MY, Poethig RS, 2005. A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev* **19**, 2164-75.

Zemach A, Kim MY, Hsieh PH*, et al.*, 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193-205.

Zemach A, Mcdaniel IE, Silva P, Zilberman D, 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916-9.

Zhang H, Zhu JK, 2011. RNA-directed DNA methylation. *Curr Opin Plant Biol* **14**, 142-7.

Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE, 2007. Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci U S A* **104**, 4536-41.

Zhang X, Yazaki J, Sundaresan A*, et al.*, 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189-201.

Zheng B, Wang Z, Li S, Yu B, Liu JY, Chen X, 2009. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev* **23**, 2850-60.

Zhong X, Hale CJ, Law JA*, et al.*, 2012. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat Struct Mol Biol*.