Identification and characterisation of MORC6 as a component of the RNA-directed DNA methylation pathway in *Arabidopsis thaliana*

Thomas Richard Brabbs BSc.

A thesis submitted for the degree of Doctor of Philosophy

University of York

Department of Biology

September 2012

Abstract

RNA silencing pathways control the expression of genes and other DNA loci by the action of small RNA molecules and are found in many eukaryotes. In plants there are a number of RNA silencing pathways, of which RNA-directed DNA methylation (RdDM) is one. In this pathway the small RNA molecules direct DNA methylation, resulting in the down regulation of expression of the target locus. In terms of the mechanism of the pathway it is mostly well characterised but several gaps exist in our knowledge. These relate to its initiation, where it is not known how RdDM targets the correct locus; methylation, where it is unclear how the action of small RNAs triggers methylation; and chromatin modification, where it is unclear how methylated DNA is converted into higher order chromatin modification. These gaps in the pathway raised the possibility of the involvement of novel proteins and so this project aimed to identify and characterise mutants in these proteins.

Screening of a library of putative RdDM mutants in *Arabidopsis thaliana* identified three alleles in *MORC6,* which encodes a GHKL ATPase containing protein associated with RdDM that is thought to form higher order chromatin in response to DNA methylation. Analysis of the three alleles revealed that *morc6* mutants have no effect on siRNA production but at certain loci do have an effect on DNA methylation and so would suggest that MORC6 is also involved in the DNA methylation process in RdDM at specific loci. It was also shown that silencing by RdDM can still occur in *morc6* mutants in a limited capacity and that this silencing is stochastic and cell autonomous in nature. These findings point to MORC6 also having further roles in RdDM other than higher order chromatin modification and so increase our understanding of the mechanism of RdDM.

Table of contents

List of figures

List of tables

List of equations

Acknowledgements

I would first like to thank Louise Jones for providing me with this opportunity to undertake a PhD and her help guidance throughout. I would also like to thank my TAP panel, Betsy Pownall, Ottoline Leyser and Richard Waites, for their advice and support during my PhD. This PhD was funded by the BBSRC and so I am also grateful to them for this financial support. The transgenic lines and EMS mutagenesis that this study is based on were created by Andrew Eamens so I wish to thank him for this. I would also like to thank the current and past members of the Louise Jones group for their help, advice and friendship and for making my time in the lab more enjoyable, particularly Tom Smith for helping with numerous experiments, making me envious of his lunch on a daily basis and convincing me that enforced fun is not fun; and Eleanor Walton for putting up with me working next to her and proving that melting LB media is far harder than it seems. The Pownall, Issac and Leyser groups have also helped out and provided advice with experiments on numerous occasions throughout my PhD and also made the L2 lab a fun and pleasant place to work. I would particularly like to thank: Simon Fellgett for being such a good housemate and for putting up with me as a housemate as well as his commitment to vigorously testing lab safety, Richard Maguire for being put upon as a IT and statistics help desk and for his services to your mum jokes, Simon Ramsbottom for showing that it is possible to finish a PhD without the need for lab books and for being a confocal guru, Jo Hepworth for the moral support when plants don't do what you want them to do and that tea solves many problems; and Di Quinn having helped two generations of the Brabbs family with their PhDs. Other L2 past and present lab members I would like to thank include: Sarah Wetherill, Maaike de Jong, Fiona Warrander, Nick Bland, Emily Winterbottom, Asma Bano, Gilu George, Hannah Brunsdon, Joe Vaughn, Vera Matser, Phil Garnett and Laura Faas. Finally I would also like to thank my parents for their encouragement and support during this PhD.

Dedication

This thesis is dedicated to John Frank Brabbs, Kathleen Brabbs, James Lincoln Thomas and Gunda Thomas.

Author's declaration

None of the work presented in this thesis has been previously published or submitted for a qualification either at the University of York or at any other institution. Some parts of this work have however been presented in posters at the PiCLS conference (Dundee 2010) and the GARNet conference (Cambridge 2011). All of the work presented here is that of the author except those stated in the acknowledgements section of each results chapter.

……..

Thomas Brabbs

1. General Introduction

1.1 RNA silencing in plants

1.1.1 Overview

RNA silencing is a term that refers to eukaryotic gene regulatory pathways that are directed by small RNA molecules. There are a number of such pathways in plants of which one of them, RNA-directed DNA methylation (RdDM), regulates gene expression by DNA methylation and it is this pathway that is the focus of this study (Wassenegger et al. 1994, Xie et al. 2004). Before discussing the mechanism of RdDM in detail, it is important to understand the basic structure of RNA silencing pathways and therefore this will be explained as well as a brief outline of the other pathways found in plants. The process through which RdDM represses genes, DNA methylation, will also be introduced in order to understand how the RdDM pathway functions. This will also introduce other pathways involved in methylation and the process by which these repressive marks are removed. The introduction will then move onto the RdDM pathway itself in terms of its mechanism and function. Finally this chapter will discuss the aims of this study and the rational for why it was undertaken.

1.1.2 Discovery of RNA silencing

In 1928 a study by Wingard noted that when Tobacco plants were infected with the tobacco ring spot virus, younger leaves further up the plant were asymptomatic and that the plants were resistant to further ring spot infections (Wingard 1928). At the time it was not known how the plants gained this immunity to the virus but it is now known to be due to an RNA silencing pathway acting in viral defence (Ratcliff et al. 1999). The actual discovery of RNA silencing occurred in the 1990s and came about as a result of research to identify the cause of unusual experimental findings from transgenic lines in *Petunia* (Napoli et al. 1990, van der Krol et al. 1990a, van der Krol et al. 1990b). The *Petunia* lines contained a transgenic copy of the endogenous *Chalcone synthase* (*CHS*) gene,

which is part of the floral pigment biosynthesis pathway, and the aim was to determine if the action of this enzyme was the rate limiting step of biosynthesis by increasing the level of CHS in petals. The introduction of the transgene was expected to either increase the purple flower pigmentation or have no effect, but instead the studies found that a significant percentage of transformant lines either totally or partially lacked pigmentation. Further investigation showed where there was no pigmentation both the transgenic and endogenous *CHS* genes had reduced expression and this affect was named co-suppression as a result of this dual repression. Co-suppression was later shown to be post transcriptional in nature as levels of RNA polymerase transcription were unaffected in silenced *Petunia* lines and instead the resulting mRNA transcripts were degraded (Van Blokland et al. 1994). Co-suppression was renamed post transcriptional gene silencing (PTGS) after studies in *Arabidopsis thaliana* and *Nicotiana tabacum* where a single transgene that was unrelated to any endogenous gene, also caused a post-transcriptional reduction in gene expression (Dehio and Schell 1994, Ingelbrecht et al. 1994, Elmayan and Vaucheret 1996). PTGS was initially thought to be a plant phenomenon until similar silencing mechanisms were identified in Fungi and Metazoan species which were termed quelling and RNA interference (RNAi), respectively (Cogoni et al. 1996, Fire et al. 1998, Kennerdell and Carthew 1998, Lohmann et al. 1999, Sánchez Alvarado and Newmark 1999, Wianny and Zernicka-Goetz 2000).

Investigation of the mechanism behind PTGS suggested that it involved double stranded RNA (dsRNA) as RNA corresponding to the antisense sequence of the *CHS* gene could be detected in silenced *Petunia* lines (Van Blokland et al. 1994). This was confirmed when it was shown that co-expression of two transgenes that produced sense and antisense mRNAs could successfully trigger RNA silencing as could the expression of an inverted repeat transgene (Hamilton et al. 1998, Waterhouse et al. 1998). These two studies showed that inverted repeats and antisense strand expression were both sources of dsRNA for PTGS. It was also suggested that another source of dsRNA was viral replication as several groups

showed that viral infection could trigger silencing of either an endogenous gene or transgene if the virus contained part of the target genes sequence (Lindbo et al. 1993, Kumagai et al. 1995, Ruiz et al. 1998). Later studies also found that dsRNA can be produced by RNA-dependent RNA polymerases (RDRs) that convert single stranded RNA (ssRNA) into dsRNA (Cogoni and Macino 1999, Dalmay et al. 2000b, Mourrain et al. 2000, Tang et al. 2003).

The next development in the investigation of the mechanism of PTGS was the identification, in four different PTGS systems, of small 25 nucleotide RNA molecules corresponding to the antisense sequence of the target gene (Hamilton and Baulcombe 1999). These small RNAs were subsequently shown to be created by cleavage of dsRNA by proteins with RNase III activity, known as Dicers in animals and Dicer-like (DCL) in plants, into short dsRNAs (Bernstein et al. 2001, Elbashir et al. 2001, Tang et al. 2003). These small RNA molecules were termed small interfering RNAs (siRNA) and were shown predominantly to be between twenty one to twenty four nucleotides in length in plants (Hamilton et al. 2002, Llave et al. 2002, Tang et al. 2003). It was discovered that the larger sizes of siRNA were involved in transcriptional gene silencing (TGS) rather than PTGS and so the term RNA silencing began to be used to refer to both PTGS and TGS that involved these siRNA molecules (Hamilton et al. 2002, Xie et al. 2004). The discovery of these siRNA molecules explained what happened to the dsRNA but raised the question of how silencing of targets is achieved.

In 2000 it was shown that in *Drosophila* cells primed with dsRNA, cellular extracts contained a protein complex termed the RNA induced silencing complex (RISC) that had endonuclease activity specific to the dsRNA target sequence and that its function was dependent on siRNAs produced from the dsRNA (Hammond et al. 2000, Zamore et al. 2000, Nykänen et al. 2001). It was assumed that the siRNA provides the RISC with target specificity through Watson and Crick base pairing but this requires the siRNA to be single stranded and further investigation revealed that one of the strands is removed in a ATP dependent manner after

association with the RISC complex (Nykänen et al. 2001). Analysis of RISC revealed one of the components to be an ARGONAUTE (AGO) protein that was later shown to be responsible for both the endonuclease and siRNA binding activities of RISC (Hammond et al. 2001, Liu et al. 2004, Ma et al. 2005, Rivas et al. 2005, Wang et al. 2008, Wang et al. 2009). AGO proteins were first identified in plants and the plant AGO1 protein has been shown to be required for PTGS, indicating that plant AGOs are also likely to function in RISCs (Bohmert et al. 1998, Fagard et al. 2000). With the identification of the RISC complex this completed the basic outline of RNA silencing pathways, which consists of the production of dsRNA by a variety of sources, cleavage of the dsRNA by a DCL enzyme into siRNAs and loading of one of the siRNA strands into an AGO protein to form a RISC (Figure 1.1).

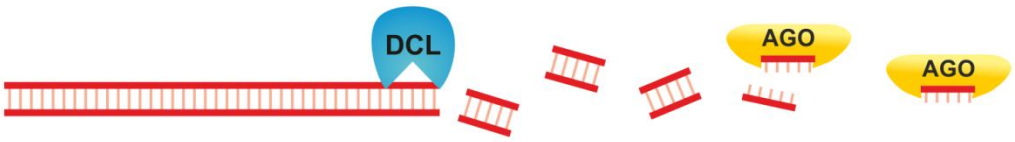


Figure 1.1: Basic outline of RNA silencing pathways

Diagram shows the key steps of RNA silencing pathways with RNA strands shown in red with bases in orange. A Dicer-like (DCL) RNase III enzyme is shown in blue and a AGO protein, which forms part of the RISC, is shown in yellow.

1.1.3 Divergence of RNA silencing pathways in plants

There are five distinct RNA silencing pathways in plants and although their basic structure (Figure 1.1) is the same there are key differences between them, namely how dsRNA is produced, the size of the small RNAs produced; and whether RISC silences post-transcriptionally or transcriptionally. The divergence of RNA silencing pathways in plants can be seen by the large number of silencing components in *Arabidopsis* as there are a total of six *RDR* genes, four *DCL* genes; and ten *AGO* genes (Finnegan et al. 2003, Yu et al. 2003, Havecker et al. 2010).

The five RNA silencing pathways are the: micro RNA (miRNA), trans-acting siRNA (tasiRNA), natural antisense siRNA (nat siRNA), viral defence siRNA and RNA-directed DNA methylation (RdDM) pathways (Mourrain et al. 2000, Hamilton et al. 2002, Park et al. 2002, Reinhart et al. 2002, Yu et al. 2003, Peragine et al. 2004, Vazquez et al. 2004, Xie et al. 2004, Wang et al. 2005, Yoshikawa et al. 2005, Axtell et al. 2006, Henz et al. 2007, Jin et al. 2008). This section will briefly explain how each pathway functions and highlight the differences between the pathways, which are also summarised in Table 1.1.

Pathway	dsRNA production	DCL	Main siRNA size (bp)	AGO	Mode of action
miRNA	Secondary structure	1	21	1	PTGS
tasiRNA	miRNA cleavage followed by RDR6	4	21	1 and 7	PTGS
nat siRNA	RDR6	1 and 2	21 and 24	?	PTGS
viral siRNA	RDR1 and RDR6	2 and 4	21 and 22	1 and 7	PTGS
RdDM	RDR2	3 and 4	24	4,6 and 9	TGS

Table 1.1: Differences between the five main RNA silencing pathways in plants

Summary of the key differences between RNA silencing pathways in plants including: how dsRNA is produced, which DCL is required, the size of the siRNAs produced, which AGO is required and whether the pathway acts post-transcriptionally (PTGS) or transcriptionally (TGS)

The miRNA pathway is a PTGS pathway where dsRNA is produced by the formation of a stem loop secondary structure by an RNA transcript and so does not require the activity of an RDR to produce dsRNA (Figure 1.2 A) (Park et al. 2002, Reinhart et al. 2002). The stem does not show full complementarity in terms of sequence resulting in bulges and it is this lack of complementarity that is the major difference between miRNAs and siRNAs as siRNAs are produced from full complementary dsRNA (Figure 1.2 A and B). The stem loops are known as pri-miRNA transcript and are cleaved by DCL1 to produce the mature miRNA from the stem, which are typically 21 nt in length, although some such as those associated with tasiRNAs (22 nt) are longer (Park et al. 2002, Chen et al. 2010, Cuperus et al. 2010). miRNAs are then loaded into the AGO1 protein and act in

trans to down regulate the level of its target mRNA (Rhoades et al. 2002). This down regulation can occur either through cleavage and resulting degradation of the target mRNA or by repression of ribosomal translation from this mRNA, however cleavage is the most common process through which miRNAs operate in plants (Aukerman and Sakai 2003).

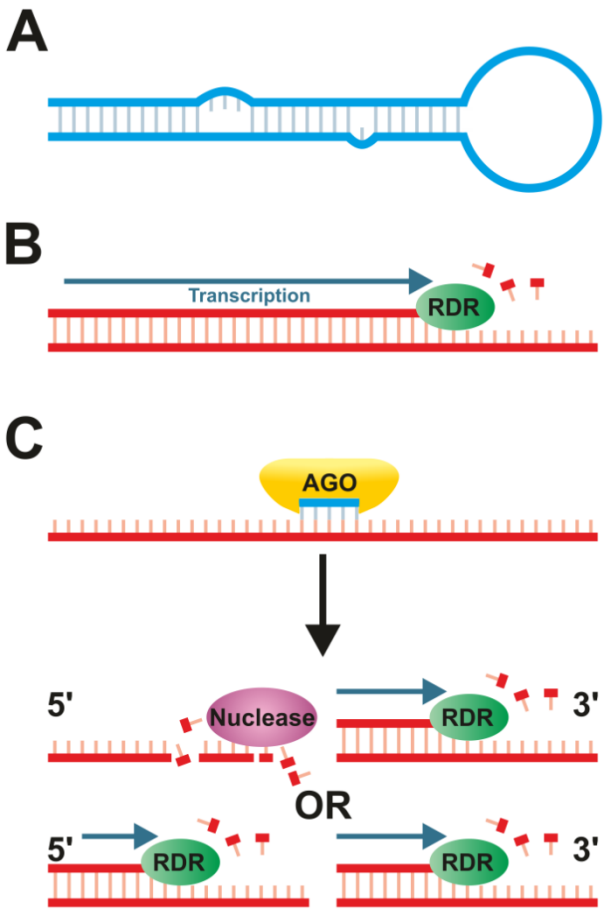


Figure 1.2: Differences in production of dsRNA by RNA silencing pathways

Diagram shows how dsRNA is produced for miRNAs (**A**), siRNA pathways including RdDM and viral defence (**B**) and tasiRNAs (**C**). miRNAs and pri miRNA transcript are shown in blue while siRNAs and associated dsRNA are shown in red. A RDR is shown in green while an AGO is shown in yellow and a nuclease is shown in purple. RNA nucleotides are shown as a red and orange T shape. The blue arrows show the direction of RDR transcription.

The tasiRNAs are produced from four non-coding RNA genes in *Arabidopsis* that are known as *TAS1-TAS4*, although there may be further as yet unidentified *TAS* genes (Allen and Howell 2010). siRNA production from these transcripts requires the cleavage of the transcript by a miRNA, which in the case of *TAS1, 2* and *3* is miR173 and for *TAS3* is miR390, and this miRNA-dependent initiation is one of the major difference between the tasiRNA pathway and other siRNA pathways (Yoshikawa et al. 2005, Axtell et al. 2006). The cleavage by these two miRNAs on their respective targets results in two cleaved transcripts of which the one to the 3' side of the cleavage site is turned into dsRNA by the action of RDR6 while the other transcript is either degraded or also turned into dsRNA, but at a far lower efficiency (Figure 1.2 C) (Peragine et al. 2004, Vazquez et al. 2004, Yoshikawa et al. 2005). The dsRNA is then cleaved by DCL4 into 21 nt siRNAs before being loaded into an AGO, possibly either AGO1 or AGO7 as both result in a reduction in tasiRNA silencing, although this may be down purely to the miRNAs involved in tasiRNA synthesis being loaded into these AGOs (Peragine et al. 2004, Vazquez et al. 2004, Adenot et al. 2006). The resulting complex then acts in *trans* to cleave target mRNA so is another PTGS pathway.

The nat siRNA pathway involves the silencing of a target gene by the production in cis of a corresponding antisense transcript and the resulting production of siRNAs from this transcript (Wang et al. 2005, Henz et al. 2007, Jin et al. 2008). Similarly to tasiRNAs, nat siRNAs are produced in a two-step process whereby initially 24 nt siRNAs are produced that direct cleavage of the sense gene's mRNA and the resulting cleaved RNA transcripts are used to produce 21 nt siRNAs that go on to cleave further mRNA transcripts of the target gene thereby amplifying the pathway (Borsani et al. 2005, Katiyar-Agarwal et al. 2006). The initial antisense transcript is transcribed by RNA polymerase IV (PolIV) and forms dsRNA by annealing to the sense transcript transcribed by RNA Polymerase II (PolII), which is then cleaved by DCL2 into 24 nt siRNAs. These 24 nt siRNAs cleave the sense transcripts, which are then themselves turned into dsRNA by RDR6 and 21 nt siRNAs are produced by DCL1 cleavage. The identity of the AGO

or AGOs involved in this pathway is unknown. Since the nat siRNAs result in cleavage of the sense mRNA target the pathway acts post-transcriptionally.

The viral defence pathway is a siRNA pathway that defends the plant against RNA virus infection through degradation of the viral RNA. In theory viral replication should create dsRNA that could be processed into siRNAs, however anti-viral defence often requires RDR1 and RDR6 thus suggesting that dsRNA can be produced by the action of these enzymes (Mourrain et al. 2000, Xie et al. 2001, Yu et al. 2003, Xie et al. 2004). The dsRNA is then processed into 21 and 22 nt siRNAs by DCL4 and DCL2 respectively. These are then loaded into AGO1 and AGO7 and go on to cleave the viral RNA resulting in its degradation (Xie et al. 2004, Deleris et al. 2006, Qu et al. 2008).

The final pathway is the RNA-directed DNA methylation (RdDM) pathway in which siRNAs direct TGS of targets through DNA methylation (Hamilton et al. 2002, Xie et al. 2004). The pathway acts in cis as initially the target DNA is transcribed by RNA polymerase IV and the resulting transcript is turned into dsRNA by RDR2 (Xie et al. 2004, Herr et al. 2005, Kanno et al. 2005, Onodera et al. 2005). The dsRNA is then processed by DCL3 and DCL4 into 24 nt siRNAs that are loaded into AGO4, 6 or 9 and directs DNA methylation of the target DNA (Hamilton et al. 2002, Park et al. 2002, Schauer et al. 2002, Finnegan et al. 2003, Xie et al. 2004, Havecker et al. 2010, Pélissier et al. 2011). RdDM is the only pathway of the five mentioned here that involves a transcriptional rather than post-transcriptional form of gene silencing. It is also the focus of this study and so will be discussed in further detail later on in this chapter, but before RdDM is discussed, the mechanism by which RdDM causes silencing, DNA methylation, must be first introduced. For this reason the next section will discuss DNA methylation in plants.

1.2 DNA methylation in plants

1.2.1 Types of DNA methylation found in plants

DNA methylation involves the modification of the base of a nucleic acid by the addition of a methyl group, but of the four DNA bases only adenine and cytosine have a functional significance in cells (Johnson and Coghill 1925, Wyatt 1951, Dunn and Smith 1958, Doskočil and Šormová 1965, Ehrlich et al. 1985). Cytosine can be modified at either the carbon atom at position five or the nitrogen side group of the carbon at position four, referred to as 5-methylcytosine and N^4-methylcytosine, respectively (Figure 1.3 A). Adenine can only be modified at the nitrogen side group of the carbon atom at position six and is referred to as N^6-methyladenine. The majority of DNA methylation in plants is comprised of 5-methylcytosine but low levels of N^6-methyladenine have been detected in the plant genome and from experiments using transgenes with adenine methylation it appears that this DNA modification acts as a repressive mark in plants and may have functional significance (Gruenbaum et al. 1981, Pintor-Toro 1987, Rogers and Rogers 1995, Van Blokland et al. 1998, Ashapkin et al. 2002, Sugimoto et al. 2003). By comparison to adenine methylation in plants, 5-methylcytosine methylation is considerably better understood and as it is this methylation mark that this study will concentrate on any future reference to DNA methylation refers to 5-methylcytosone only. Closer analysis of the sequence context of methylated cytosines reveals that in plants there are three sequence contexts in which methylated cytosines appear, these being CG, CHG and CHH, where H represent C, A or T with the majority of methylated cytosines being in the CG context (Figure 1.3 B) (Gruenbaum et al. 1981, Ingelbrecht et al. 1994, Meyer et al. 1994, Jacobsen and Meyerowitz 1997, Cokus et al. 2008, Lister et al. 2008). By comparison to plants, the majority of methylation in animals is in the CG sequence context with around 0.02% of methylation being non-CG in differentiated cells, although in stem cells this rises to around a quarter suggesting a difference in methylation mechanism between stem cells and

differentiated cells (Ramsahoye et al. 2000, Lister et al. 2009). The fact that non-CG methylation is more prevalent in plants than animals makes it a defining feature of plant DNA methylation. The nature of CG and CHG methylation means that there will also be a cytosine on the opposite strand that can be methylated and it has been found that if one cytosine is methylated the other tends to be as well, which has a relevance to how the methylation at these cytosines is maintained (Figure 1.3 B) (Cokus et al. 2008, Lister et al. 2008).

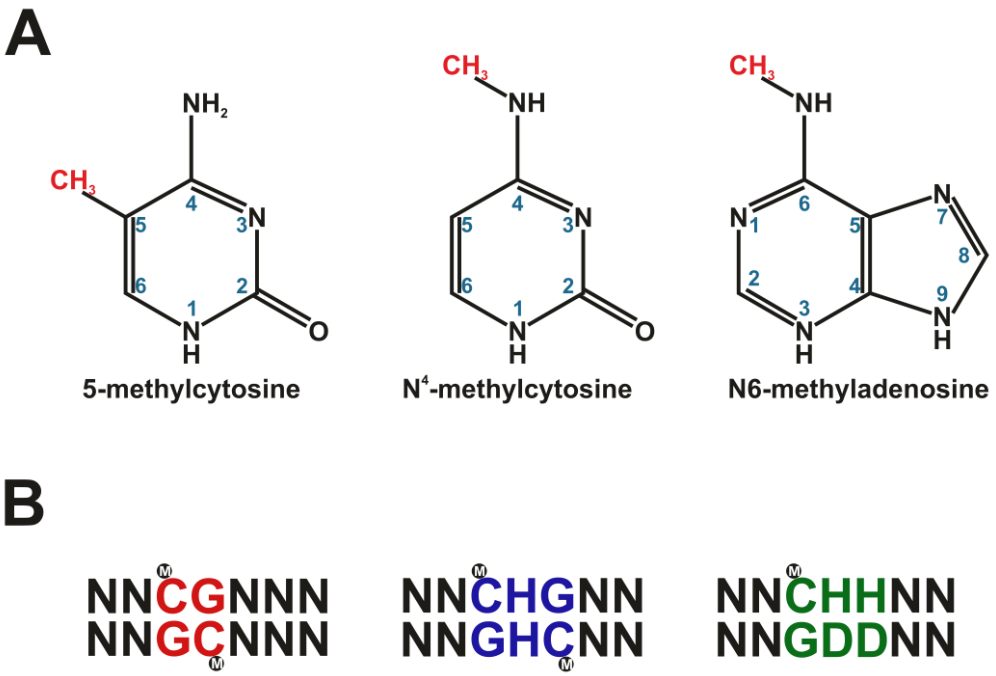


Figure 1.3: The different methylated bases found in plants and the cytosine sequence context that can be methylated

A: Chemical structures of the three types of DNA methylation found in eukaryotes and prokaryotes. Single lines represent single bonds while double lines represent double bonds. N is nitrogen while O is oxygen and H is hydrogen, and any junction between lines without a N or O is carbon. The CH_3 in red is the methyl group added to the respective base. The numbers inside the rings for each schematic show the position of each carbon and nitrogen in the rings. **B:** Shows the three sequence contexts for 5-methylcytosine methylation in plants. The methyl groups are shown as white Ms in a black circle. CG methylation is in red, CHG methylation in blue; and CHH methylation in

green. H represents cytosine (C), adenine (A) or thymine (T) and D represents guanine (G), adenine or thymine.

1.2.2 DNA methyltransferases in plants

The addition of methyl groups to DNA is catalysed by the methyltransferase enzymes, of which there are three families in plants: the Methyltransferase (MET) family, the Chromomethylase (CMT) family and the Domains rearranged methyltransferase (DRM) family (Finnegan and Dennis 1993, Henikoff and Comai 1998, Rose et al. 1998, Genger et al. 1999, Cao et al. 2000, McCallum et al. 2000, Henderson et al. 2010). Within these three families MET1, CMT3 and DRM2 are the principle methyltransferase enzymes from their respective families. MET1 and CMT3 are maintenance methyltransferases whereas DRM2 is required for *de novo* methylation, although for some loci CMT3 can also act as a *de novo* methyltransferase (Finnegan et al. 1996, Lindroth et al. 2001, Cao and Jacobsen 2002a, Cao and Jacobsen 2002b, Aufsatz et al. 2004). Interestingly in a transgene system mutation of either of the three principle methyltransferases resulted in a reduction in methylation in all cytosine contexts and so suggests a more complex relationship between the three methyltransferases at some loci (Singh et al. 2008). Only CG and CHG methylation can be maintained by maintenance methyltransferases whereas CHH requires continuous *de novo* methylation in order to be maintained (Finnegan et al. 1996, Lindroth et al. 2001, Cao and Jacobsen 2002a, Cao and Jacobsen 2002b, Aufsatz et al. 2004). The reason for this is that, for CG and CHG methylation, after DNA replication during S-phase the unmethylated daughter strand can be methylated using the parent strand methylation as a template since both CG and CHG are palindromic (Figure 1.4). Only one of the parental strands is methylated for CHH methylation, meaning that to methylate the other DNA helix *de novo* methylation must occur as there is no template to indicate where methylation should occur. The next three sections will discuss how each of these three methyltransferase enzymes functions.

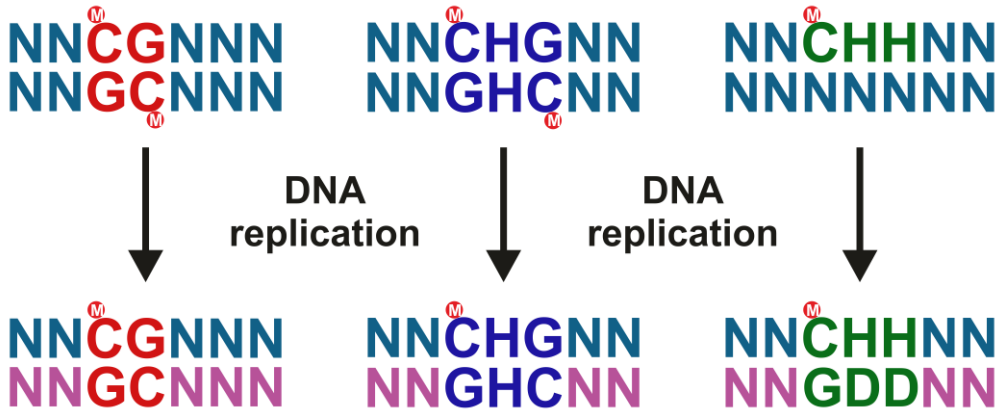


Figure 1.4: Effect on DNA replication on DNA methylation

Diagram showing what happens to CG, CHG and CHH methylation after replication has occurred. The parental strands are shown in blue while the daughter strands are shown in purple. DNA methylation is shown as a white M in a red circle.

1.2.3 MET1 methylation

MET1 is part of a family of four genes in *Arabidopsis* which also include MET2a, MET2b and MET3 (Genger et al. 1999). MET3 is a truncated protein so is unlikely to be functional and MET2a and MET2b are both expressed at lower levels compared to MET1, hence it can be considered that MET1 is the major methyltransferase in this family. This is supported by the fact that *met1* mutants have a near total loss of CG methylation, with some reduction in global levels of CHG and CHH methylation and would suggest that MET1 targets CG sites preferentially but may also influence the methylation of other sequence contexts (Finnegan et al. 1996, Ronemus et al. 1996, Kankel et al. 2003, Cokus et al. 2008, Lister et al. 2008). This loss of CG methylation in *met1* mutants is associated with the alteration of expression of a large number of genes and intergenic non-coding loci as well as pleiotropic developmental defects and so would suggest that MET1 and hence CG methylation is functionally important in plants (Finnegan et al. 1996, Ronemus et al. 1996, Kankel et al. 2003). MET1's primary function is to maintain CG methylation while *de novo* methylation is the

responsibility of DRM2 (Jones et al. 2001, Cao and Jacobsen 2002a, Cao and Jacobsen 2002b, Cao et al. 2003, Kankel et al. 2003). However, some loci may require MET1 for *de novo* methylation of CG sites as a study showed that a *met1* mutant was unable to establish *de novo* CG methylation of a transgene system (Aufsatz et al. 2004). Another study has also shown that reintroduction of MET1 into a *met1* mutant resulted in partial restoration of gene body CG methylation again suggesting *de novo* methylation activity (Zubko et al. 2012). In *Nicotiana tabacum* the *de novo* methyltransferase NtDRM1 was able to methylate CHG and CHH sites in vitro but was not as proficient at methylating CG sites, suggesting that NtDRM1 may not be the main CG *de novo* methyltransferase and so would support the idea that MET1 or in this case its *N. tabacum* homolog may carry out CG *de novo* methylation (Wada et al. 2003).

Due to their homology, the process of MET1 methylation is thought to be similar to that of the mammalian DNMT1 methyltransferase. MET1 recognises hemimethylated CG sites, where one of the cytosines is unmethylated and methylates this cytosine using S-adenosyl methionine (SAM) as a methyl donor (Figure 1.5) (Gruenbaum et al. 1982, Smith et al. 1992). The chromatin modifier DDM1 is required for MET1 function and is a SWI/SNF2 ATPase-containing protein that can translocate nucleosomes along DNA in an ATP-dependent manner (Jeddeloh et al. 1999, Brzeski and Jerzmanowski 2003). DDM1 may be required for MET1 to be able to access the DNA to methylate and similar to *met1* mutants, *ddm1* mutants result in a large reduction in CG methylation and also a reduction in CHG and CHH methylation at some specific loci (Vongs et al. 1993, Teixeira et al. 2009). This suggests that both MET1 and DDM1 may function in non-CG methylation in a locus specific manner. The activity of MET1 and DDM1 also results in the modification of histone three lysine 9 (H3K9) and histone four lysine 16 (H4K16) with the former being methylated and latter deacetylated (Soppe et al. 2002, Tariq et al. 2003). How DNA methylation results in histone modification is likely to be mediated by the methyl binding domain (MBD) proteins MBD6 and MBD7 which bind to methylated CG dinucleotides and also

associate with histone deacetylases (Zemach and Grafi 2003, Zemach et al. 2005). Both *met1* and *ddm1* mutants perturb the localisation of MBD6 and MBD7 thus suggesting that MBD and MET1 activities are indeed interlinked. The most likely candidate for the histone deacetylase is HDA6 as mutants in this protein result in a decrease in CG methylation (Aufsatz et al. 2002b, Aufsatz et al. 2004). The protein responsible for the methylation of the H3K9 residue is unknown but again could be directed by an MBD protein or alternatively a SRA-SET containing protein such as SUVH4/KRYPTONITE which is responsible for H3K9 methylation in relation to CMT3 methylation (Jackson et al. 2002, Malagnac et al. 2002, Zemach et al. 2005, Johnson et al. 2007).

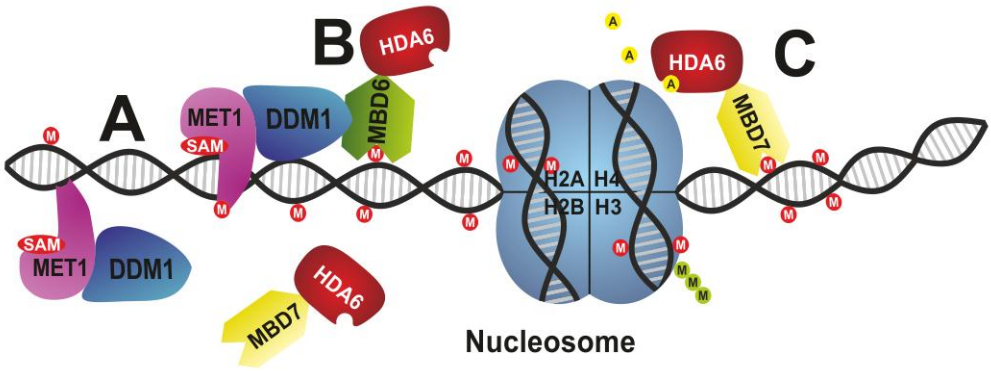


Figure 1.5: MET1 methylation and resulting histone modification

Diagram showing the linkage between CG methylation by MET1 and the resulting histone deacetylation that is mediated by MBD proteins. The DNA double helix is shown in black with grey lines representing base pairing. Methyl groups are shown as white Ms in a red circle. **A:** MET1 (purple) in conjunction with DDM1 (blue) recognises and methylates hemi-methylated CG sites using SAM (red) as a methyl group donor. **B:** The methylated cytosine binding proteins MBD6 and MBD7 (green and yellow respectively) bind to the methylated CG sites. **C:** The MBD proteins interact with HDA6 (red), which removes acetyle groups from the nucleosome component Histone 3. A single nucleosome is shown in blue and is labelled, including four of the eight histone subunits. Histone acetyl groups are shown as black As in yellow circles and histone methylation is shown as black Ms in green circles.

1.2.4 CMT3 methylation

There are three members of the CMT methyltransferase family in *Arabidopsis*, of which CMT1 function is non-essential as several ecotypes have a retrotransposon insertion in the *CMT1* gene that inactivates it (Henikoff and Comai 1998, Rose et al. 1998, McCallum et al. 2000). Of the other two family members, only CMT3 has so far been identified in mutant screens for DNA methylation mutants, so the exact function of CMT2 remains unknown (Bartee et al. 2001, Lindroth et al. 2001). In *cmt3* mutants there is a large scale decrease in CHG methylation throughout the genome suggesting that this protein is involved in CHG methylation maintenance since many of the loci tested only undergo maintenance methylation (Bartee et al. 2001, Lindroth et al. 2001, Tompa et al. 2002, Cao and Jacobsen 2002a). CMT3 has little effect on CG methylation but is involved in CHH methylation at some loci, thus it may also have *de novo* methyltransferase activity since CHH methylation only occurs by *de novo* methylation. CMT3 is also not the only methyltransferase involved in CHG methylation as at certain loci, DRM2 mutants have a greater effect on methylation than CMT3 (Cao et al. 2003). The fact that there is some level of redundancy between DRM2 and CMT3 in terms of CHG and CHH methylation means that CHG and CHH methylation is only totally lost in triple mutants of *cmt3, drm1* and *drm2* and these mutants exhibit developmental defects that are not observed in *cmt3* or *drm2* single mutants (Cao and Jacobsen 2002a, Cao et al. 2003, Chan et al. 2006, Lister et al. 2008).

The mechanism by which CMT3 methylation occurs is similar to MET1 in that CMT3 recognises hemimethylated DNA and methylates the cytosines on the daughter strand using the methylated cytosine on the parental strand as a template (Figure 1.6) (Cao and Jacobsen 2002a). CMT3 methylation does not require a nucleosome modifying enzyme such as DDM1, but does require LHP1, a protein that binds to di-methylated H3K9 and H3K27 residues, which are both repressive marks associated with heterochromatic regions (Jackson et al. 2002,

Exner et al. 2009). LHP1 may therefore target CMT3 to regions with H3K9 and H3K27 di-methylation and so would indicate that CHG methylation is controlled by histone methylation. CMT3 itself has also been shown to interact directly with the H3 histone but only when both H3K9 and H3K27 are methylated and so also supports the idea that histone methylation directs CHG methylation (Lindroth et al. 2004). There are a number of proteins that can perform H3K27 methylation but none of them have an effect on CHG methylation when mutated so the exact protein or proteins responsible for methylation of this residue in the CMT3 pathway is currently unknown. The protein responsible for H3K9 methylation is known and is the SRA and SET domain containing protein SUVH4, also known as KRYPTONITE (KYP) (Jackson et al. 2002, Malagnac et al. 2002). The SRA domain recognises and binds to methylated cytosine, with a preference for CHG and CHH, and mutations in this domain results in a loss of function of SUVH4 (Johnson et al. 2007). The fact that CHG DNA methylation requires histone modification to occur and vice versa means that methylation is cyclic and self-reinforcing in nature, whereby DNA methylation triggers histone methylation that triggers further DNA methylation and so on. This dual requirement for DNA and histone methylation explains the fact that only CHG methylation is affected in a partial mutant *in S-adenosylhomocysteine hydrolase* (*SAH*), which is responsible for the production of the SAM molecule that is the methyl group donor during methylation of both DNA and histones (Mull et al. 2006). There is also periodicity in CHG methylation with the average distance between methylated sites being around 167 base pairs (bp), which is similar to that of the length of DNA around a nucleosome(175-185 bp), and so CMT3 may be limited in terms of its access to site due to the presence of nucleosomes (Cokus et al. 2008, Lister et al. 2008). The reason for the disparity in sizes is believed to be due to the tight packing of nucleosomes.

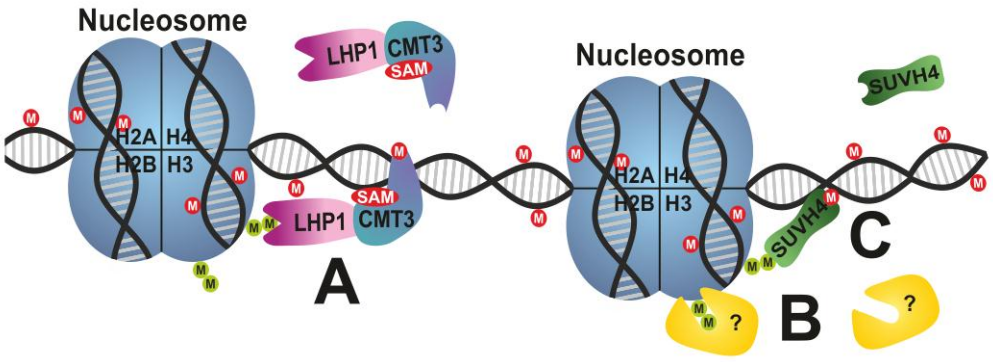


Figure 1.6: CMT3 DNA methylation and SUVH4 H3K9 di-methylation

Diagram showing the linkage between CHG methylation by CMT3 and H3K9 di-methylation by SUVH4. DNA methylation is shown as white Ms in a red circle while histone methylation is shown as black Ms in green circles. **A:** CMT3 (blue) is directed hemimethylated CHG sites by LHP1 (purple), which recognises histone H3K9 di-methylation. CMT3 then methylates the CHG site using SAM (red) as a methyl group donor. **B:** The H3K27 di-methylation occurs in response to CHG methylation, although the protein (yellow) or proteins involved are unknown. **C:** The SUVH2 protein (green) recognises CHG methylation and then di-methylates the H3K9 residue.

1.2.5 DRM2 methylation

The final group of methyltransferases in plants is the *de novo* methyltransferase DRM family comprising three proteins (Cao et al. 2000, Henderson et al. 2010). Of these three proteins, DRM2 is the main protein responsible for *de novo* methylation as *drm2* mutants are unable to carry out this function whereas DRM1 expression levels are low and *drm1* mutants do not affect methylation levels; and DRM3 lacks several of the ten conserved motifs that are important for methyltransferase function (Cao and Jacobsen 2002b, Henderson et al. 2010, Naumann et al. 2011).However, although unlikely to catalyse DNA methylation itself, DRM3 is still required for *de novo* DNA methylation and so may act as a co-factor for DRM2. It is not known if there is any role for DRM1 in DNA methylation either as a methyltransferase or as a co-factor to DRM2, but most studies have

used double mutants of *drm1* and *drm2* to investigate the role of *de novo* methyltransferases. The *drm1 drm2* double mutant results in a large decrease in CHH methylation and a more limited decrease in CHG methylation but does not affect CG methylation (Cao and Jacobsen 2002a, Cao and Jacobsen 2002b, Cao et al. 2003). As mentioned previously CHH and CHG methylation is only completely removed in triple mutants of *drm1, drm2* and *cmt3* and would suggest that CMT3 is required for *de novo* methylation at some loci but that DRM2 is the predominant methyltransferase involved (Cao and Jacobsen 2002a, Cao et al. 2003, Lister et al. 2008). The mechanism which triggers and controls DRM2 is the RdDM pathway and will be discussed in detail in section 1.3 so will not be covered here (page 39).

1.2.6 Location of DNA methylation within genome

In principle any cytosines can be methylated, however there does appear to be a preference as to which cytosines are methylated in terms of the surrounding sequence (Cokus et al. 2008, Lister et al. 2008). Certain sequences favour methylation with cytosines in all configurations more likely to be methylated if preceded by the sequence CN, with N being any nucleotide. Other sequences discourage methylation and for CG sites this is where the CG is directly preceded by an A and followed by a T, while for both CHG and CHH sites it is if the middle H is a cytosine. The likelihood of CHG and CHH methylation occurring also increases if a upstream CG site is methylated and in the case of CHH also if a upstream CHG site is methylated. The exact reason for this sequence preference is currently not known and neither is it known whether the methyltransferases themselves or a co-factor is responsible for this sequence recognition. Methylation levels are not uniform throughout the *Arabidopsis* genome with the majority of methylation located within the centromeric and pericentromeric regions which have a large concentration of repetitive sequences and transposable elements (Lippman et al. 2003, May et al. 2005, Zhang et al. 2006, Zilberman et al. 2007, Cokus et al. 2008, Lister et al. 2008). This methylation is

required for heterochromatin formation and silencing of these regions, which is important for genome stability in terms of preventing transposable element movement and centromere stability. Another characteristic of DNA methylation is that it overlaps with areas of the genome that have associated siRNAs and is to be expected considering that the RdDM pathway is responsible for *de novo* methylation (Chan et al. 2006, Cokus et al. 2008, Lister et al. 2008). Perhaps the most surprising finding is that CG but not CHG or CHH methylation is found within the gene bodies including those of vital plant genes and it has been suggested that this methylation is functionally significant in terms of correct expression or splicing, although its exact role has yet to be determined (Cokus et al. 2008, Lister et al. 2008, Takuno and Gaut 2011, Dalakouras et al. 2012, Zubko et al. 2012). It has also been shown that CHG gene body methylation also occurs but that this is subsequently removed by IBM1 (Miura et al. 2009, Rigal et al. 2012)

1.2.7 Demethylation

DNA methylation is a dynamic modification and can be removed either passively, through a lack of maintenance of the methylated cytosines after DNA replication, or actively by demethylation enzymes (Agius et al. 2006). One of these enzymes is ROS1, which has both DNA glycosylase and apurinic/apyrimidinic lyase activity and is specific to 5-methylcytosines (Gong et al. 2002, Agius et al. 2006, Morales-Ruiz et al. 2006). ROS1 functions by first removing the methylated cytosine base from the DNA backbone through DNA glycosylase activity, leaving an abasic site. Subsequently the DNA lyase domain removes the deoxyribose sugar from the backbone leaving a gap in the strand that can be repaired by the base excision repair pathway. ROS1 operates in a pathway with ROS3, which is an RNA binding protein that can interact with siRNAs, and so demethylation by ROS1 may be directed to specific targets by siRNAs (Zheng et al. 2008). ROS1 targets are also associated with 24 nt siRNAs from the RdDM pathway so it is possible that ROS3 interacts with siRNAs and mediates ROS1 demethylation (Penterman et al. 2007).

This would also suggest contradictory roles for siRNAs in that they may be required for both methylation and demethylation and so raises the question of if siRNAs are targeting the demethylation machinery to targets how it only affects a subset of loci with 24 nt siRNAs whereas methylation occurs at the others. One possibility could be the differences in the amount of siRNAs sequestered into the two pathways; however this requires further investigation. Another protein has also been shown to be required for ROS1 activity, IDM1, which is a histone acetyltransferase enzyme that acetylates the H3 histone of nucleosomes with repressive marks such as H3K9 di-methylation but not ones with H3K4 di-methylation, which are transcriptionally active marks (Qian et al. 2012). Since IDM1 targets nucleosomes with repressive marks and is required for ROS1 demethylation, it would suggest that IDM1 may modify nucleosomes in a transcriptionally inactive state in such a manner that allows ROS1 to demethylate the locus.

A second enzyme, DEMETER (DME), that is part of the same family of DNA glycosylases as ROS1 has also been shown to directly remove methylated cytosines (Choi et al. 2002, Gehring et al. 2006, Morales-Ruiz et al. 2006). Similarly to ROS1, DME also has both DNA glycosylase and apurinic/apyrimidinic lyase activity so is considered to function in a similar manner, although it is not known if DME also interacts with ROS3 and may not be directed to targets by siRNAs. DME activity is facilitated by the action of SSRP1, which is part of the FACT complex that modifies nucleosomes in order to allow transcription initiation and so may also similarity modify nucleosomes to allow DME demethylation activity (Ikeda et al. 2011). There are a further two proteins in the DME family in *Arabidopsis,* DML2 and DML3, which have also been shown to be required for DNA demethylation and again probably function in a similar manner to ROS1 (Ortega-Galisteo et al. 2008). DML2 and DML3 have also been shown to function in the same pathway as ROS1 and IDM1 but may target different subsets of loci as a triple mutant of *dml2, dml3* and *ros1* resulted in a larger number of loci with increased DNA methylation than a *ros1* single mutant (Qian

et al. 2012). Mutation of *ROS1, DML1* and *DML2* resulted in hypermethylation of loci that were previously unmethylated or had low levels of DNA methylation in wild-type plants, however single mutants of *ros1, dml1* and *dml2* and a triple mutant of these genes resulted in a loss of DNA methylation at loci with high levels of methylation in wild-type plants (Ortega-Galisteo et al. 2008, Qian et al. 2012). This would suggest a more complex relationship between DNA methylation and demethylation than merely an antagonistic one in which demethylation counteracts methylation and instead suggests that at specific loci that are highly methylated, demethylation is required in order to achieve these high levels. However, this finding could also be caused by down-regulation of DNA methylation genes in response to the loss of demethylation. Equally RdDM appears to be required for demethylation as in *rdr2* and *drm2* mutants, there was CG hypermethylation as a result of ROS1 down regulation (Penterman et al. 2007).

1.3 RNA-directed DNA methylation pathway

1.3.1 Discovery of RNA-directed DNA methylation pathway

The elucidation of the mechanism of the RdDM pathway has occurred over the past decade although the idea of an RNA silencing pathway carrying out transcriptional gene silencing through DNA methylation first appeared in the 1990s (Wassenegger et al. 1994). Previously to this suggestion of a role for RNA in DNA methylation the exact mechanism by which methylation occurred was not well understood and the only suggestion of a role for RNA in this process was a study that indicated that RNA inhibited methylation (Bolden et al. 1984). This changed after a study in 1994 showed that DNA methylation could occur as a result of RNA (Wassenegger et al. 1994). In this study it was shown that a T-DNA insertion in the genome of Tobacco plants that contained viroid cDNA sequence could be methylated upon infection with the viroid RNA and that methylation was localised to the viroid cDNA and persisted after the viroid infection had

abated. The authors therefore suggested that the viroid RNA guided Tobacco DNA methyltransferase enzymes to methylate the viroid cDNA and described the phenomenon as RNA-directed methylation. At this point it was still not known how the RNA caused DNA methylation and whether it was direct action of the viroid RNA or occurred through an intermediate as the viroid used in that study replicated in the nucleus. Another study showed that an intermediate was required by using an RNA virus to induce DNA methylation of a transgene containing part of the viral sequence (Jones et al. 1998). This RNA virus replicated in the cytoplasm thus the viral RNA molecule could not trigger DNA methylation directly but instead a signal must be transmitted from the cytoplasm to the nucleus. This raised the question as to the nature of the signal and it was hypothesised that it was small RNA molecules produced from aberrant RNA in the cytoplasm, although this had not been tested experimentally (Wassenegger and Pélissier 1998).

The identity of this signal was not identified until the early 2000s. Small RNA molecules had been shown to be associated with post transcriptional gene silencing in 1999 and a study in 2000 demonstrated that DNA methylation was also associated with small RNA molecules (Hamilton and Baulcombe 1999, Mette et al. 2000). In the study by Mette DNA methylation of a *NOS* transgene promoter was shown to occur upon introduction of a dsRNA molecule comprising of an inverted repeat of the *NOS* promoter and that the dsRNA was processed into 23-25 nucleotide (nt) RNA molecules. These small RNA molecules were suspected to be the signal that results in DNA methylation, however this needed to be proved experimentally. A later study identified two classes of small RNA molecules, now termed small interfering RNAs (siRNAs), a short class between 21-22 nt in length and a longer class between 24-26 nt in length (Hamilton et al. 2002). The study showed that the longer class of siRNAs were required for DNA methylation of both the *GFP* transgene used in the study and the AtSN1 retrotransposon and so would support the idea that small RNAs are the signal for DNA methylation. The study also found that the *sde4* mutant

resulted in a loss in both AtSN1 siRNAs and DNA methylation and as such was the first component of the RdDM pathway to be identified, although it was not until later that the identity of the mutant gene was discovered to be the large subunit of the *Arabidopsis* RNA polymerase IV (Hamilton et al. 2002, Herr et al. 2005). There is variation in the size of siRNAs associated with DNA methylation between studies, which was due to difficulties in obtaining an exact size from the northern blots used to identify the siRNAs. A later study was however able to produce an exact size of siRNAs associated with RdDM and that was 24 nt (Xie et al. 2004).

During this period as well as identifying 24 nt siRNAs as being required for RdDM several mutants defective in the RdDM pathway were also identified. As mentioned in a previous section *Arabidopsis* has multiple paralogs of key RNA silencing proteins including the six RDRs, four DCLs and ten AGOs and assessment of the phenotype of mutants in these proteins revealed RDR2, DCL3 and AGO4 to be required for RdDM (Zilberman et al. 2003, Xie et al. 2004). From this point onwards a large number of other proteins have been shown to be part of the RdDM pathway which have advanced our understanding of the pathway's mechanism. In the next sections I will describe the pathway itself in detail beginning with siRNA production

1.3.2 RdDM siRNA biogenesis

During the initial stages of RdDM, 24 nt siRNAs are produced. Early research into RdDM identified the largest and second largest subunits of a novel plant RNA Polymerase IV (PolIV), described as NRPD1 and NRPD2 respectively, as RdDM components (Herr et al. 2005, Kanno et al. 2005, Onodera et al. 2005). Further studies have also identified the fourth subunit of PolIV, NRPD4, as being required for RdDM (He et al. 2009a, Greenberg et al. 2011). No other subunits of PolIV have been identified in forward genetic screens for RdDM mutants. The reason for this may be that PolIV has evolved from RNA polymerase II (PolII) early in plant evolution, before the divergence of algae, and as such, of the twelve PolIV

subunits, eight are shared with PolII. These subunits are unlikely to appear in RdDM mutant screens as mutations in these subunits may be lethal due to the loss of PolII function (Luo and Hall 2007, Ream et al. 2009). It is not known why the fourth subunit that is specific to PolIV, NRPD7, has not to date been identified in an RdDM mutant screen. The subunits that differ between PolIV and PolII are located at the points of DNA template entry and RNA transcript exit and so would be expected to alter the function of PolIV (Figure 1.7) (Ream et al. 2009).

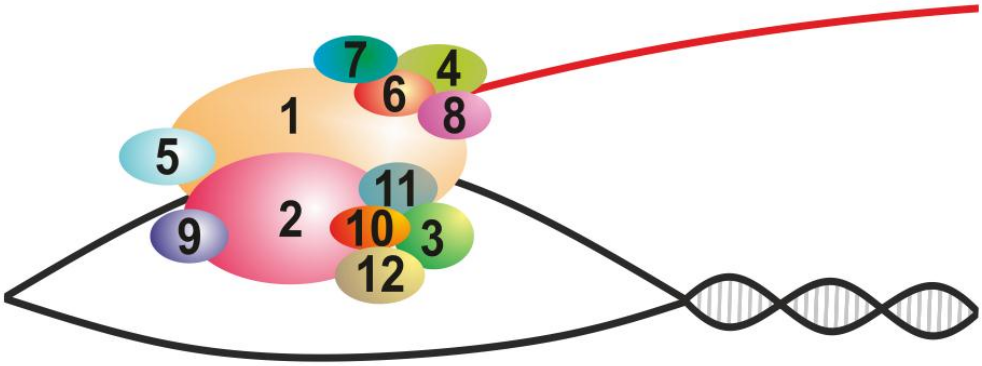


Figure 1.7: Subunit structure of RNA polymerases II, IV and V

Diagram shows the positions within the polymerase complex of the twelve subunits found in PolII, PolIV and PolV. Each subunit is labelled numerically with the 1 being the 1st subunit of the three polymerases etc. The black lines represent DNA while the red line is the RNA transcript. The entry point of the DNA template strand is found at the junction between the 1st, 2nd, 5th and 9th subunits. The point at which the RNA transcript exits the polymerase is found between the 1st, 8th and 4th subunit. The subunit sizes are not to scale and their positions are not exact in relation to other subunits. This diagram is adapted from Ream 2009.

In terms of the function of PolIV in RdDM, mutants in its subunits show loss of both DNA methylation and siRNA accumulation and so would suggest that PolIV

is required for siRNA production (Herr et al. 2005, Kanno et al. 2005, Onodera et al. 2005, He et al. 2009a, Greenberg et al. 2011). It was therefore concluded that PolIV most likely transcribes the target DNA and its transcript would then be used for siRNA biogenesis (Figure 1.8 B). However, no study to date has shown PolIV to have transcription activity in vivo or in vitro and PolIV transcripts have not been detected. Several groups have proposed that due to a lack of evidence for transcription of DNA, PolIV may actually transcribe RNA rather than DNA (Vaughn and Martienssen 2005, Pontes et al. 2006). However, this seems unlikely as both PolII and PolV use DNA as a template and PolIV is closely related to both (Luo and Hall 2007, Marcussen et al. 2010). Another possible explanation for the lack of detectable PolIV transcripts is the close association of PolIV with RDR2 (Law et al. 2011). RDR2 is a RNA-dependent RNA polymerase required for RdDM and as such is expected to convert single stranded RNA templates into double stranded RNA (dsRNA) (Dalmay et al. 2000b, Xie et al. 2004). RDR2 interacts strongly with the NRPD1 PolIV subunit, most likely at the point where the RNA transcript exits the polymerase and so it is possible that RDR2 converts the PolIV transcript immediately into dsRNA (Figure 1.8 B) (Law et al. 2011). This would explain the lack of a PolIV transcript as the dsRNA could be quickly targeted for processing into siRNAs. Whether this hypothesis is correct could be tested by searching for PolIV transcripts in *rdr2* mutants, which should in theory not produce dsRNA from these transcripts.

As with PolII transcription PolIV, also requires transcription factors and chromatin modifiers in order to function. The first such factor to be identified was CLASSY1 (CLSY1) a SNF2 protein which displayed a reduction in siRNA accumulation and DNA methylation and was believed to be involved in siRNA production (Smith et al. 2007, Greenberg et al. 2011). SNF2 proteins were first discovered in *S. cerevisiae* and modify chromatin in an ATP dependent manner through the translocation of nucleosomes, achieved by the disassociation of DNA from the histones and alterations to the histone octamers conformation (Hirschhorn et al. 1992, Whitehouse et al. 1999, Bruno et al. 2003, Kassabov et

al. 2003). These changes allow RNA polymerase complexes access to the DNA sequence and so in this case it was assumed that CLSY1 facilitates PolIV transcription (Figure 1.8 A). In support of this idea, it was recently shown that CLSY1 directly interacts with the largest subunit of PolIV, probably around the DNA template entry point, as the NRPD1 subunit spans both the DNA template entry point and RNA transcript exit point in the polymerase complex (Ream et al. 2009, Law et al. 2011). There are also three homologs of CLSY1 in *Arabidopsis* and these have been named CLSY2-4. These have also been shown to interact with the NRPD1 PolIV subunit, although it is not known whether the CLSY proteins act redundantly to each or are required for RdDM at different loci.

Two transcription factors have also been shown to be required for PolIV transcription, SHH1 and RDM4 (He et al. 2009b, Kanno et al. 2009, Law et al. 2011, Liu et al. 2011). RDM4, also referred to as DMS4, is a conserved IWR1-like transcription factor found in eukaryotes that associates with PolII, PolIV and PolV (Krogan et al. 2006, Collins et al. 2007, He et al. 2009b, Kanno et al. 2009). IWR1 transcription factors are normally associated with PolII so the association with PolIV and PolV is likely due to the close homology between the three polymerases (Luo and Hall 2007, Marcussen et al. 2010). Mutations in *RDM4* result in a loss of siRNAs and DNA methylation but also a reduction of expression at a number of PolII transcribed genes resulting in a pleiotropic developmental phenotype. RDM4 has also been shown to interact directly with the NRPD1 subunit and so is likely to target PolIV to RdDM targets, however like CLSY1, it is not known whether RDM4 is required for all RdDM targets or whether it is only required for a subset of these targets. The other transcription factor is SHH1, also known as DTF1, and is a member of the SAWADEE homeobox transcription factor family found in plants (Mukherjee et al. 2009, Law et al. 2011, Liu et al. 2011). SHH1 interacts directly with PolIV through the NRPD1 subunit and mutation of *SHH1* results in loss of siRNAs and DNA methylation at some RdDM target loci but not others. This would suggest that SHH1 is required for PolIV transcription at specific loci.

In terms of how these six proteins function in order to facilitate PolIV transcription, since they all interact with PolIV it could be considered that they form a complex that precedes PolIV (Figure 1.8 A). In this scenario SHH1 and RDM4 would bind to specific DNA sequence motifs and so bring in CLSY1 or one its homologs and PolIV, with the CLSY then allowing transcription to occur by removing nucleosomes from the site of transcription. The transcription factors would then disassociate upon transcription initiation. These proteins may also be target specific, such as different CLSYs being required for specific loci but not others. Whether these proteins function in such a complex is not known and would require further investigation to determine if this is the case. The six proteins so far identified that are required for PolIV transcription also highlight one of several gaps in the knowledge of the RdDM pathway and that is how PolIV is targeted to the correct locus. It is likely that combinations of CLSY1-4, SHH1 and RDM4 will facilitate PolIV transcription at different loci, however these six proteins cannot account for the myriad number of targets of RdDM in the genome that PolIV must be targeted to or provide enough scope for finer control of RdDM. This would suggest that there may be other transcription factors and chromatin modifiers involved in PolIV transcription.

Once PolIV transcription has occurred and the RNA transcript turned into dsRNA through the action of RDR2, the dsRNA is then targeted by Dicer-like 3(DCL3) and DCL4 (Figure 1.8 C)(Xie et al. 2004, Pélissier et al. 2011). DCL3 and DCL4 are two of the four plant homologs of the Dicer enzyme which cleave dsRNA into small RNA molecules between 21-24 nt in length. Both DCL enzymes specifically cleaves dsRNA in the RdDM pathway into 24 nt long segments (Hamilton et al. 2002, Park et al. 2002, Schauer et al. 2002, Finnegan et al. 2003, Xie et al. 2004, Pélissier et al. 2011). The two DCLs probably act redundantly in RdDM as the reduction in siRNA levels in single mutants of *dcl3* and *dcl4* are modest (Greenberg et al. 2011, Pélissier et al. 2011, Ausin et al. 2012a). A recent study has shown that the spliceosome component SR45 is an RdDM component and due to the reduction in both DNA methylation and siRNA levels in *sr45* mutants it

was concluded it was involved in siRNA production (Ausin et al. 2012a). In terms of SR45's role in siRNA production it was found that the *sr45* phenotype was enhanced in a double mutant with *dcl3,* which may suggest that SR45 facilitates both DCL3 and DCL4 function and so a *sr45 dcl3* double mutant that lacks DCL3 siRNA processing may also disrupt DCL4 activity through SR45. Splicing machinery has previously been implicated in the plant miRNA pathway as it was found that three proteins involved in the nuclear cap-binding complex, are required, along with DCL1, for the processing of long pri-miRNA transcripts into mature miRNAs (Laubinger et al. 2008). However, the exact function of SR45 in RdDM is not known and as no other spliceosome components tested in the study were found to result in reduction in DNA methylation or siRNA levels, the function may be different to its role in splicing. The double stranded RNA binding protein DRB2 is also involved in siRNA production but has an antagonistic role to DCL3 as in *drb2* mutants levels of 24 nt siRNAs are enhanced (Pélissier et al. 2011). This suggests that DRB2 is involved in the regulation of siRNA levels and prevents over accumulation. How DRB2 achieves this is unknown, although it is not through reduction in the levels of DCL3 itself, and could be either through sequestration of the dsRNA produced by RDR2 thus blocking access by DCL3 or DRB2 triggering degradation of the dsRNA. There is also a second DRB protein, DRB4, involved in RdDM, however it has an antagonistic role to DRB2 as in *drb4* mutants there is a reduction in siRNA levels, therefore suggesting that DRB4 either promotes or is directly involved in siRNA production. DRB4's function is likely to be in conjunction with DCL4 as DRB4 interacts directly with DCL4 but not DCL3; and in this case DRB4 through binding to the dsRNA would facilitate DCL4 cleavage of the RNA.

The dsRNA molecules created by DCL3 and DCL4 have a two nucleotide overhang at the 3' ends, which can be degraded by endonucleases. In order to protect these ends from ribonuclease attack the 3' ends of both strands are modified by HEN1 by the addition of methyl groups to the end base (Figure 1.8 D) (Xie et al. 2004, Yu et al. 2005). However, only certain siRNAs require this modification as

in *hen1* mutants some siRNAs are reduced in abundance whereas levels of other siRNAs are unaffected. In the cases where HEN1 is not required for siRNA production it is not known how the dsRNA segments produced by DCL3 or DCL4 avoid degradation.

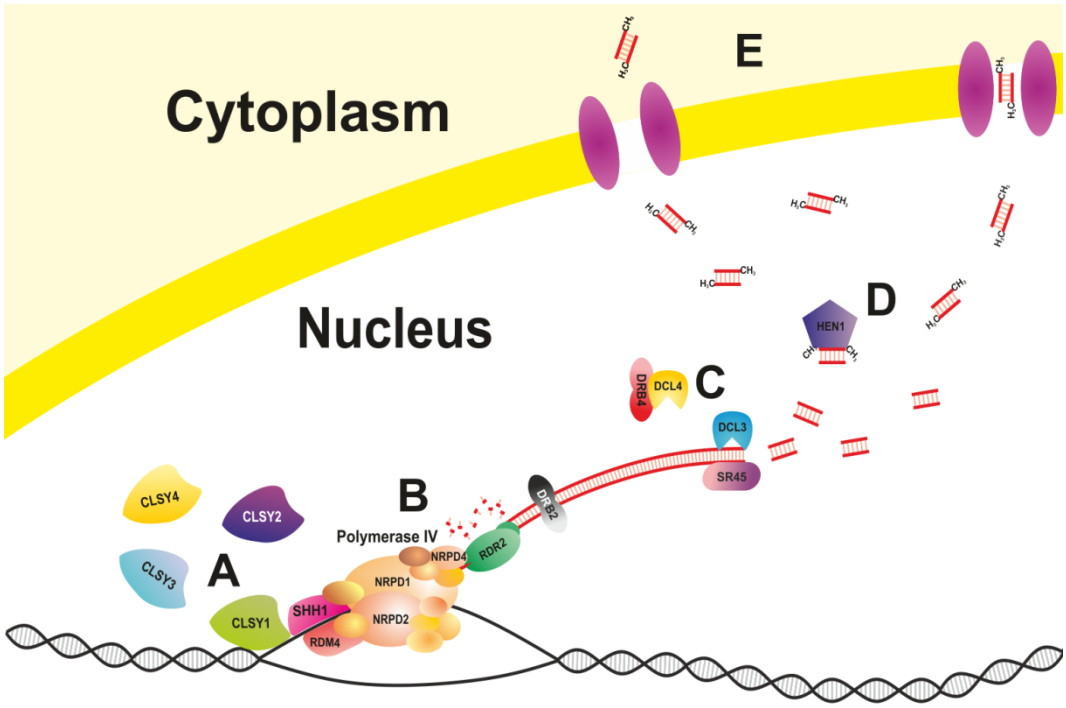


Figure 1.8: RdDM siRNA production

A: Transcription factors and nucleosome modifiers facilitate PolIV transcription. **B:** The RNA transcript produced by PolIV is turned into dsRNA by RDR2. **C:** DCL3 and DCL4 cleave the dsRNA into 24 nt segments. **D:** HEN1 methylates the 5' ends of the 24 nt dsRNA molecules. **E:** The 24 nt dsRNA molecules are then exported from the nucleus. The nuclear envelope is depicted as a yellow line with protein channels within this membrane shown as pairs of purple ovals. Areas with a white background are within the nucleus and areas in pale yellow are in the cytoplasm. RNA strands are shown as red lines with orange bars representing base pairing between strands. Methylation at the 3' end of siRNAs is shown by CH_3 labels and RNA nucleotides are shown as red and orange T shapes. DNA strands are shown as black lines with grey bars representing base pairing

between strands. All proteins in the pathway have been labelled except some PolIV subunits as these have not yet been shown to be required for RdDM.

1.3.3 Loading of siRNAs into AGO proteins in RdDM

After the dsRNAs are produced by DCL3 or DCL4, one of the strands is then loaded into ARGONAUTE (AGO) 4, 6 or 9 while the other is then likely degraded by ribonuclease activity (Zilberman et al. 2003, Xie et al. 2004, Zilberman et al. 2004, Qi et al. 2006). The strand that is loaded into AGO4 is known as the siRNA strand whereas the other strand is termed the passenger strand. As well as AGO4 there are also two other AGOs (AGO6 and AGO9) that have been shown to be involved in RdDM (Zheng et al. 2007, Havecker et al. 2010, Eun et al. 2011, Greenberg et al. 2011). These three AGO are closely related and with AGO8 form a separate clade from the other six *Arabidopsis* AGO proteins. Whether AGO8 is involved in RdDM is unknown but it lacks the PIWI catalytic domain and so could either have a divergent function or be a pseudogene (Takeda et al. 2008). The presence of two other AGOs in the RdDM pathway explains why *ago4* mutants have locus specific reductions in siRNA and DNA methylation levels and indeed AGO6 and AGO9 target different loci to AGO4, but AGO4 appears to be the predominant AGO in RdDM as it targets the majority of loci (Zilberman et al. 2003, Zilberman et al. 2004, Qi et al. 2006, Havecker et al. 2010). The target specificity is partly due to expression patterns between the AGOs as expressing AGO6 and AGO9 using a AGO4 promoter in a *ago4* mutant background resulted in rescue of the mutant phenotype at some loci. However, this did not occur at all loci and so suggested that factors other than expression pattern affect the specificity of the three AGOs.

AGO proteins are characterised by four domains: the N-terminus, PAZ, Mid and PIWI domains, of which the function of the PAZ, Mid and PIWI domains have been determined but that of the N-terminus domain is still unknown (Hutvagner and Simard 2008, Parker 2010). Of the three known domains, the PAZ and Mid

domains are important for siRNA strand loading whereas the PIWI domain is important in catalytic activity. The function of the PAZ domain is in the binding of the 3' end of the siRNA strand via the creation of a binding pocket in which the 3' end resides upon loading (Lingel et al. 2004, Ma et al. 2005). The Mid domain is also involved in siRNA binding and forms a pocket at the point where it borders the PIWI domain in which the 5' end of the siRNA strand binds. These two domains therefore are responsible for siRNA loading and hold the siRNA strand in place (Figure 1.9). The three AGOs involved in RdDM show a preference as to which of the strands from the dsRNA is loaded. This preference depends on the thermo-stability of the base pairing at the 5' end of each strand, with the strand with the least stable 5' end being preferentially loaded (Khvorova et al. 2003, Schwarz et al. 2003, Gyorgy 2005). The reason why the less stable 5' end is selected is thought to be that the base pairing will be inherently weaker and so is more likely to become single stranded and enter the Mid binding pocket. The AGOs also have a preference in terms of which base is at the 5' end, with adenosine being the preferred base (Mi et al. 2008, Havecker et al. 2010). In this case the preference is again likely to be caused by the Mid domain 5' binding pocket.

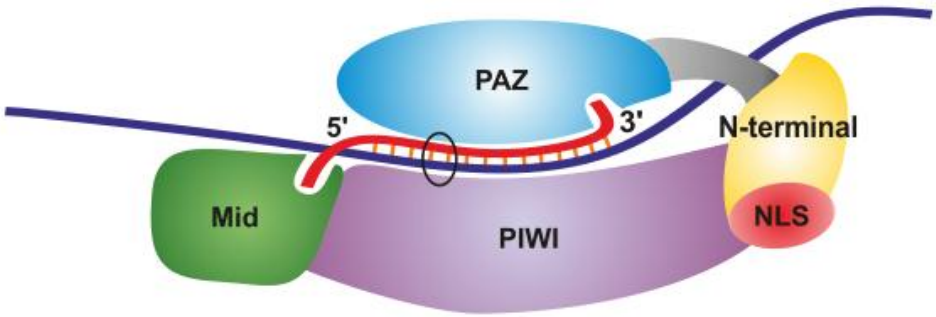


Figure 1.9: AGO protein domain structure

Schematic of the basic domain structure of AGO proteins and how it binds to a siRNA strand. The four domains are labelled and also shown in different colours. The N-

terminal domain also contains a nuclear localisation signal (NLS) shown in red. The PAZ domain is linked to the rest of the AGO protein by the grey linker region shown in the diagram. The siRNA strand is in red while the target transcript is in dark blue, with base pairing between the strands shown in orange. The orientation of the siRNA strand is given by the 5' and 3' labels and the black circle highlights the 13th and 14th bases from the 3' end of the siRNA, which is the point at which PIWI cleaves the target transcript. Diagram is adapted from Song 2004.

The loading of AGO proteins was initially believed to occur in cajal bodies within the nucleus (Li et al. 2006, Pontes et al. 2006, Pikaard et al. 2008). Cajal bodies are dense clusters of proteins and DNA that are associated with the nucleolus. These studies found that in isolated nuclei DCL3, RDR2, AGO4 and NRPE1 all localised strongly to the cajal bodies, although were present in the rest of the nucleus, and so it was concluded that due to the presence of these proteins that AGO4 loading must occur here. However, a recent study has contradicted this assumption and instead suggests the AGO loading occurs in the cytoplasm (Ye et al. 2012). This study found that the 24 nt siRNAs associated with RdDM were most prevalent in the cytoplasm rather than the nucleus and also found that although AGO4 was most predominant in the nucleus, it could also be detected in the cytoplasm. They showed that AGO4 proteins bind to the dsRNAs produced by DCL3 or DCL4 in the cytoplasm and require the PIWI domain for removal of the passenger strand. The PIWI domain catalyses the cleavage of RNA strands that are bound to the siRNA strand through the DDH motif, comprising two aspartic acid residues and histidine residues (Figure 1.9) (Liu et al. 2004, Rivas et al. 2005, Wang et al. 2008, Wang et al. 2009). This cleavage of the RNA strand bound to the siRNA occurs specifically between the 11th and 12th base from the 3' end of the siRNA and in this case it would appear that the passenger strand is cleaved in this manner, resulting in its release. Indeed mutants that lack the PIWI catalytic activity are shown to bind to siRNA duplex but are apparently unable to remove the passenger strand.

The binding of an siRNA with an AGO protein results in a conformational change whereby the PAZ domain is swung out from the main body of the protein so that it sits above the PIWI domain (Figure 1.9) (Song et al. 2004). There is also another conformational difference between when the siRNA is bound or unbound and this occurs when the passenger strand is released by AGO4 (Ye et al. 2012). This change exposes a nuclear localisation signal (NLS) in the N-terminal domain and it is thought that this results in the AGO4-siRNA complex being transported back into the nucleus (Figure 1.10 C and D). Mutants that lacked the NLS did not move back into the nucleus and mutants in the PIWI domain also were unable to move back into the nucleus as they were unable to remove the passenger strand. It should be noted that the mechanism by which dsRNAs and siRNA bound AGOs are exported and imported from the nucleus is not known and would raise the possibility of other proteins involvement in the pathway. The process of AGO4 loading with the siRNA strand has however been shown to require the HSP90 protein, which is a chaperone protein and its function in RdDM is ATP dependent (Figure 1.10 B). HSP90 has also been shown to be required for AGO loading for plant AGO1 and in Metazoan species suggesting that the requirement for HSP90 in AGO loading is universal (Iki et al. 2010, Iwasaki et al. 2010). It was therefore suggested that HSP90 binds to AGO4 proteins and facilitates the conformational change required for AGO4 to bind to the dsRNA molecules in a ATP dependent manner (Figure 1.10 E). The cleavage and resulting degradation of the passenger strand results in the release of HSP90. The removal of the passenger strand causes the NLS to be exposed and the AGO4 protein is then imported back into the nucleus. The study by Ye did not show if AGO6 and AGO9 were also loaded with siRNAs in the cytoplasm, although their close homology to each other would support the idea that they do (Ye et al. 2012). The fact that the rest of the pathway occurs within the nucleus also raises the question of why AGO4 loading occurs in the cytoplasm. One possible explanation is that as the prevalence of 24 nt siRNAs is higher in the cytoplasm than the nucleus that this process acts as a rate limiting step for RdDM and so prevents hypermethylation occurring.

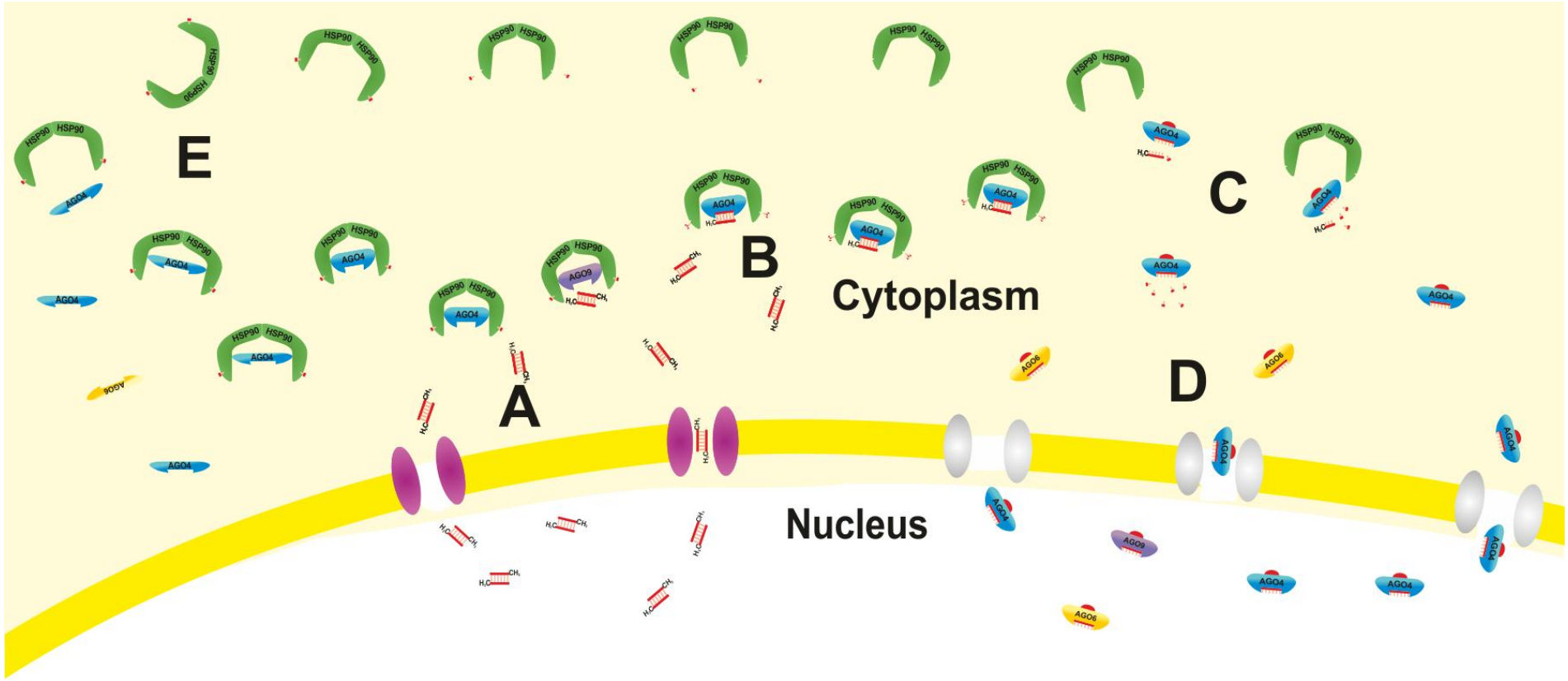


Figure 1.10: Loading of siRNAs into AGO proteins in RdDM

A: 24 nt dsRNA molecules are exported from the nucleus. **B:** dsRNA molecules are loaded into AGO4, 6 or 9 and the process is facilitated by HSP90. **C:** The passenger strand is cleaved by the AGO resulting in HSP90 releasing the AGO protein and the nuclear localisation signal (in red) becoming exposed. **D:** The siRNA loaded into a AGO protein is imported back into the nucleus. **E:** The HSP90 binds to a new AGO protein and facilitates a conformational change in the AGO protein that allows it to bind to a 24 nt dsRNA molecule. The nuclear envelope is depicted as a yellow line with protein channels within this membrane shown as pairs of purple ovals or silver ovals for exporters from or importers into the nucleus respectively. Areas with a white background are within the nucleus and areas in pale yellow are in the cytoplasm. RNA strands are shown as red lines with orange bars representing base pairing between strands. Methylation at the 3' end of siRNAs is shown by CH_3 labels and RNA nucleotides are shown as red and orange T shapes.

1.3.4 DNA methylation in RdDM

After the loading of the AGO proteins with 24 nt siRNAs, the next stage of the pathway is the methylation of the DNA target and this requires another plant specific RNA polymerase, PolV, which evolved from PolIV after the divergence of ferns from other land plants but before flowering plants diverged (Kanno et al. 2005, Pontier et al. 2005, Huettel et al. 2007, Luo and Hall 2007). Of the twelve PolV subunits, NRPE1, NRPE5 and NRPE7 are specific to PolV with NRPE2 and NRPE4 being shared with PolIV and the remaining seven subunits shared by PolV, PolIV and PolII (Ream et al. 2009). These PolV-specific subunits are located at the DNA template entry and RNA transcript exit points of the polymerase, which would suggest that the divergence in function between PolIV and PolV is due to changes at these two points (Figure 1.7). Of the subunits that are PolV specific or shared with only PolIV, mutants have been identified in NRPE1, NRPE2, NRPE4 and NRPE5, which means that only NRPE7 has yet to be shown to be required for RdDM (Kanno et al. 2005, Pontier et al. 2005, Huettel et al. 2007, Eamens et al. 2008, He et al. 2009a, Huang et al. 2009, Lahmy et al. 2009, Greenberg et al. 2011). The remaining seven subunits are unlikely to appear in RdDM mutant screens as they are also PolII subunits and so mutants in these subunits may have severe phenotypes due to loss of PolII function. NRPE9 has been identified in a mutant screen, however the reason for this may be that there are actually two versions of this PolII subunit and it is the second version (b) that is required for PolV activity in RdDM (Tan et al. 2012).

In terms of PolV's role in RdDM, it is believed to be responsible for producing transcripts from the target DNA locus that AGO proteins loaded with siRNAs then bind to. PolV transcriptional activity has not yet been demonstrated in vitro, however transcripts have been detected in vivo that are PolV-dependent which would suggest that PolV is transcriptionally active (Wierzbicki et al. 2008, Lee et al. 2012, Wierzbicki et al. 2012). This transcriptional activity does not require the presence of siRNAs and in PolV mutants the levels of the majority of siRNAs are

unaffected, however for some loci there is a reduction in siRNA levels and the exact reason for this is unknown, although it could be due to DNA methylation acting as a trigger for further siRNA production and so maintains RdDM at this locus (Kanno et al. 2005, Pontier et al. 2005, Wierzbicki et al. 2008, Rowley et al. 2011). PolV transcription has been shown to originate from outside the region targeted by 24 nt siRNAs and therefore produces transcripts that span across the target regions to which siRNA loaded AGOs bind (Wierzbicki et al. 2008, Wierzbicki et al. 2009). Further investigation of PolV transcription showed that 61% of the PolV transcription sites corresponded to sites with both methylation and 24 nt siRNAs, with a further 12% corresponding to sites with either 24 nt siRNAs or methylation. These sites included a high percentage of transposable elements. However, 27% of PolV transcription sites lacked both siRNAs and methylation and it was found that the majority of these sites corresponded to genes (Douet et al. 2009, Wierzbicki et al. 2012). This would suggest that PolV may have other roles outside of its role in RdDM. Such a role for PolV was found in higher order heterochromatin formation, as a mutant in the 1st PolV subunit, *nrpe1*, results in decondensation of the pericentromeric heterochromatin regions but this was shown not to occur in a PolIV mutant, *nrpd1* (Pontes et al. 2009). In the same way that PolV's role in plants appears to be not just restricted to RdDM, PolV is not the only polymerase involved in this stage of the pathway. PolII has also been shown to perform an identical role to PolV at a subset of RdDM loci where PolIV but not PolV is required (Pontier et al. 2005, Huettel et al. 2006, Zheng et al. 2009). These loci that are targeted by PolII tend to be intergenic or low copy number loci, whereas PolV targets are usually transposable elements, but the reason why PolII is used instead of PolV at certain loci is unknown (Zheng et al. 2009). The PolII-dependent RdDM pathway is likely to be similar to that of the PolV-dependent pathway as both create transcripts to which siRNA-loaded AGOs bind, however whether there are any differences between the pathways apart from the polymerases involved is currently unknown. Work into the RdDM mechanism has focused on the PolV-dependent

RdDM pathway so for this reason the rest of this section will refer only to the PolV-dependent pathway.

There is uncertainty about PolV transcription initiation as, similar to PolIV, it is not clear how PolV is targeted to the correct RdDM targets. The most likely explanation is through the action of transcription factors and so far PolV has been shown to require the RDM4 transcription factor to carry out transcription at a subset of loci (He et al. 2009b, Kanno et al. 2009, Greenberg et al. 2011). To date RDM4 is the only transcription factor known to be required for PolV transcription so does not explain how it is targeted to all RdDM target loci. This therefore raises the possibility of other undiscovered transcription factors being involved in PolV transcription. It has been shown, however that PolV transcription requires the DDR complex which comprises of RDM1, DRD1 and DMS3 and that this complex would appear to interact with the NRPE1 and NRPE2 subunits of PolV (Figure 1.11 B) (Law et al. 2010). DMS3 contains two coiled coil domains and a hinge domain, of which the hinge domain provides the protein with the flexibility to change its 3D structure and the coiled coil domains facilitate interactions with other proteins that also contain these domains (Kanno et al. 2008, Ausin et al. 2009, Greenberg et al. 2011). In this case it is thought that DMS3 forms homodimers that bind to DNA strands as a previous study has shown that a protein containing hinge and coiled coil domains was able to bind to DNA as a homodimer (Chiu et al. 2004, Kanno et al. 2008). The role in the DDR complex may therefore be to bind to the DNA strands and provide a platform for DRD1 and RDM1. DRD1 belongs to a plant-specific subfamily of SWI2/SNF2-like proteins that are defined by the SWI2/SNF2 ATPase domain (Kanno et al. 2004, Flaus et al. 2006, Chan et al. 2006, Greenberg et al. 2011). This is the same family as CLSY1 and would therefore suggest that DRD1 performs a similar function to CLSY1 in that it is responsible for nucleosome displacement in an ATP dependent manner, but is associated with PolV rather than PolIV (Hirschhorn et al. 1992, Whitehouse et al. 1999, Bruno et al. 2003, Kassabov et al. 2003). The final component of the DDR complex, RDM1, is a small protein that appears to bind to

single stranded DNA with a preference for methylated DNA, which suggests that it may target the DDR complex to RdDM targets that have already been methylated and so allow for maintenance of methylation of these sites by triggering further methylation (Gao et al. 2010). RDM1 binding to DNA is likely to occur at the transcription fork of PolV as RDM1 only binds to single stranded DNA and may also help localise the DDR complex to PolV.

Comparisons between the largest subunits of PolIV and PolV, NRPD1 and NRPE1, showed that the biggest difference between the two proteins is the presence of ten GW/WG motifs within the carboxy-terminal domain (CTD) of NRPE1 (Pontier et al. 2005, El-Shami et al. 2007). These motifs were shown to be required for PolV function as complementation of a *nrpe1* mutant with a NRPE1 protein that lacked these motifs was unable to restore DNA methylation (El-Shami et al. 2007). The GW/WG motifs allow proteins to interact and bind to other GW/WG containing proteins and previous studies in other eukaryotic species have shown that several AGO proteins have these repeats and bind to other proteins through them, hence the motifs are also known as an AGO hook (Behm-Ansmant et al. 2006, Ding and Han 2007, Partridge et al. 2007, Eulalio et al. 2008). This means that the function of the GW/WG repeat in the NRPE1 subunit could be to interact with AGO proteins and indeed it was shown that AGO4 has GW/WG repeats and interacts with NRPE1 (El-Shami et al. 2007). The study did not show if AGO6 and AGO9 can interact with NRPE1 in this way but due to their close homology to AGO4 and similar function in RdDM it is likely that they also do. The GW/WG repeats may therefore allow siRNA loaded AGO proteins to interact directly with PolV and so position them near the RNA transcript exit point of the polymerase thus making it easier for the siRNA loaded into the AGO to bind to the PolV transcript. AGO4 and NRPE1 are however not the only GW/WG repeat containing proteins in *Arabidopsis* and one of these, KTF1/SPT5-like, has been shown to be a RdDM component (El-Shami et al. 2007, Bies-Etheve et al. 2009, He et al. 2009c). This protein is a member of the SPT5 family of transcription elongation factors that are part of the DSIF complex which prevents PolII pausing or prematurely

terminating transcription (Sims et al. 2004, Bies-Etheve et al. 2009, He et al. 2009c). The main difference between KTF1/SPT5-like and the SPT5 *Arabidopsis* ortholog is the presence of forty GW/WG repeats at the C-terminal end of KTF1 that allows it to interact directly with AGO4 (Bies-Etheve et al. 2009). KTF1 has also been shown to interact directly with RNA, most likely the transcript produced by PolV, and these results would suggest that KTF1 acts as an elongation factor for PolV by binding to both PolV and the RNA transcript in order to stabilise the polymerase thus preventing transcription arrest or premature termination (Figure 1.11 C) (Parker 2010, He et al. 2009c). KTF1 may also have another function in RdDM, as a recent study has shown that although localisation of KTF1 to RdDM target sites is unaffected in *ago4* mutants, AGO4 localisation to targets is reduced in *ktf1* mutants (Rowley et al. 2011). This result and the fact that KTF1 can bind to AGO4 would suggest that KTF1, rather than NRPE1, may provide a platform for AGO binding via the GW/WG motifs and so allowing the siRNA loaded into the AGO to bind to the PolV transcript (Figure 1.11 C) (Bies-Etheve et al. 2009, He et al. 2009c). It should be noted that *ktf1* mutants lose DNA methylation at most RdDM targets so far tested, but some show no change in DNA methylation or AGO4 binding, which could therefore suggest that KTF1 is not required at certain target loci and that other transcription elongation factors may be involved, however this has yet to be proven (Bies-Etheve et al. 2009, He et al. 2009c, Rowley et al. 2011).

The function of KTF1 and AGO proteins has been further elucidated in a recent study. The study found that in *ago4,* there was an increase in the levels of PolV and KTF1 associated with some RdDM loci and suggests that AGO4 reduces the number of PolV-KTF1 complexes at these loci (Rowley et al. 2011). Mutation of *ktf1* also results in increased levels of PolV at some loci. Since KTF1 is a transcription elongation factor and so is responsible for preventing polymerase transcriptional arrest or premature termination, this would suggest that the reason for the increase in PolV complexes in *ktf1* and *ago4* mutants could be an increase in arrest of PolV transcription. Assuming that KTF1 behaves similarly to

SPT5, it would be associated with PolV after initiation of transcription and then throughout transcription elongation, however the binding of AGO is likely to be transient otherwise when the siRNA loaded in the AGO protein binds to the PolV transcript this would tether PolV preventing further transcription (Sims et al. 2004). For AGO binding to be transient it must be released from the complex with KTF1 and PolV and this could be triggered by the siRNA binding to the PolV transcript, since during siRNA loading the AGO undergoes conformational changes in response to the siRNA becoming single stranded, the opposite could occur when dsRNA is formed by the siRNA binding (Ye et al. 2012). The release of the AGO from PolV and KTF1 could then allow a second AGO protein to bind and stabilise PolV (Figure 1.11 C).This would result in multiple siRNA loaded AGO proteins being bound to the PolV transcript, however this hypothesis would need to be confirmed experimentally.

The next stage of the pathway, after the binding of siRNA loaded AGO proteins to the PolV transcript, is DNA methylation itself. Until recently it was unclear how the binding of siRNAs triggered this event. However, recent studies have identified the IDN2 protein and its homologs as RdDM components and these may provide the link between siRNA binding and DNA methylation (Ausin et al. 2009, Zheng et al. 2010, Ausin et al. 2012b, Xie et al. 2012, Zhang et al. 2012). The first of these components to be identified was IDN2, also known as RDM12, which was shown to be a homolog of SGS3, a component of the post-transcriptional gene silencing RNA silencing pathway that binds to dsRNA with a 5’ overhang (Mourrain et al. 2000, Ausin et al. 2009, Zheng et al. 2010). Both SGS3 and IDN2 are characterised by XS, XH and zinc finger domains, of which it is the RNA recognition motif of the XS domain that binds dsRNA with 5’ overhangs (Bateman 2002, Ausin et al. 2009, Zheng et al. 2010). There is some dispute as to whether IDN2 just binds to dsRNA with a 5’ overhang or whether it can bind to dsRNA lacking an overhang as two studies show in vitro binding to only dsRNA with 5’ overhangs whereas another study found it could also bind to dsRNA with blunt ends (Ausin et al. 2009, Ausin et al. 2012b, Zhang et al. 2012). In all three

cases binding was shown in vitro using different dsRNA constructs thus investigation of the in vivo binding properties of IDN2 may resolve this issue. Mutants in *idn2* show a reduction in DNA methylation but only a minor reduction in siRNA accumulation, which is characteristic of RdDM components acting downstream of siRNA production. Thus it was concluded that IDN2 may bind to the dsRNA formed when 24 nt siRNAs base pair with PolV transcript. Assuming that IDN2 can only bind to dsRNA with 5' overhangs, the overhangs in this case would be created by the fact that the PolV transcript is longer than the siRNA and the 5' end of the siRNA is held in the Mid domain binding pocket of the AGO protein it is loaded into (Lingel et al. 2004, Ma et al. 2005).

IDN2 and SGS3 have twelve homologs that also contain the XS, XH and zinc finger domain. Two homologs, IDNL1 and IDNL2 (which are also known as IDP1 and FDM1 in the case of IDNL1 and IDP2 and FDM2 in the case of IDNL2) have been shown to form heterodimers with IDN2 (Ausin et al. 2012b, Xie et al. 2012, Zhang et al. 2012). One of the studies that identified IDNL1 and IDNL2 also showed that three other homologs of IDN2, FDM3, FDM4 and FDM5, also function in RdDM, however they only exhibited a reduction in methylation in double mutants with *INDL1,* suggesting that they are redundant with IDNL1 (Xie et al. 2012). By comparison IDNL1 and IDNL2 are not redundant with each other as mutants in either both show a reduction in DNA methylation, although IDNL1 has a greater reduction than IDNL2, and this is increased in *idnl1 idnl2* double mutants (Ausin et al. 2012b, Zhang et al. 2012). It should also be noted that this double mutant has a similar decrease in methylation as the *idn2* mutants but the triple mutant of *idn2, idnl1* and *idnl2* showed no further decrease in DNA methylation. Yeast-2-hybrid data also shows that the strongest interactions amongst IDNL1, INDL2 and IDN2 occurred between IDNL1 and IDN2; and IDNL2 and IDN2.This would suggest that IDN2 is a constitutive part of the heterodimer whereas the dimer can contain either IDNL1 or IDNL2 and the composition of the heterodimer may vary between RdDM targets. The XH domain has been shown to be required for the formation of these heterodimers and mutation of this domain disrupts RdDM

function, suggesting that the formation of these dimers is a requirement for RdDM (Zhang et al. 2012). As with the binding affinity of IDN2 to dsRNA, there is also disagreement about the dsRNA binding of the XS domains of IDNL1 and IDNL2. One study found that neither the IDNL1 nor IDNL2 XS domains could bind any form of dsRNA, whereas another study showed that IDNL1 could bind dsRNA with 5' overhangs and interacts with AGO4, although IDNL2 was not tested (Xie et al. 2012, Zhang et al. 2012). This would require further investigation, preferably in vivo to determine whether or not IDNL1 and IDNL2 can bind to dsRNA and what their exact function in the complex is. In terms of what the heterodimers function is during RdDM, assuming that they do have dsRNA binding activity it was suggested that IDN2 and IDNL1 or IDNL2 bind to separate siRNA-PolV transcript duplexes and form a platform from which the zinc finger motifs of both proteins in the heterodimer can interact with the DNA of the RdDM target (Figure 1.11 D) (Ausin et al. 2012b). The interaction of the zinc finger domains may be the trigger for DRM2 to methylate the target DNA.

As previously discussed, the DRM2 methyltransferase is the main enzyme responsible for *de novo* methylation in plants, although CMT3 is also required at some loci and total loss of CHH methylation occurs only in *drm1 drm2 cmt3* triple mutants (Cao and Jacobsen 2002a, Cao and Jacobsen 2002b, Cao et al. 2003, Chan et al. 2006). For this reason it was concluded that DRM2 was the main methyltransferase of the RdDM pathway and that CMT3 may also have a minor role in this pathway at some loci. Whether DRM1 has a function in RdDM is unclear as most studies that used *drm1* mutants did so as *drm1 drm2* double mutants, however one study did test the *drm1* single mutant and found no change in DNA methylation levels in the transgene system used in the study (Cao and Jacobsen 2002b). This coupled with the fact that *DRM1* expression is lower than that of *DRM2* may suggest that DRM1 does not have a significant role in RdDM and requires further investigation to determine whether it has any involvement. A role has however been shown for DRM3, which although homologous to DRM2, lacks key functional residues in the methyltransferase

domain including the proline-cysteine residues of motif IV that are directly responsible for the methylation of the cytosine residue (Henderson et al. 2010). It should be noted that although DRM3 is not expected to have methyltransferase activity this was not directly tested. DRM3 is required for methylation of RdDM targets and so may either facilitate methylation by DRM2 or has redundancy with DRM2 depending on its catalytic activity.

DRM2 methylation shows a 10 bp periodicity in methylation of CHH sites although the reasons for why this occurs are currently unknown (Cokus et al. 2008, Lister et al. 2008). A recent study has shown that DRM2 activity does not require a functional RdDM pathway in order to carry out DNA methylation, however in this case methylation tends to be off target (Wierzbicki et al. 2012). It was found that in both *nrpd1* and *nrpe1* mutants, although there was no change in global CHH methylation, there was a 50-60% reduction in CHH methylation at RdDM targets and that methylation instead occurred at pericentromeric regions. There was also miss targeting of CHG methylation in *nrpd1* and *nrpe1* mutants, but not to the same extent as CHH methylation and there was no change in CG methylation. Why DRM2 would methylate pericentromeric regions in RdDM mutants is unclear but could be related to the presence of other chromatin modification responsible for the formation of heterochromatin at these regions.

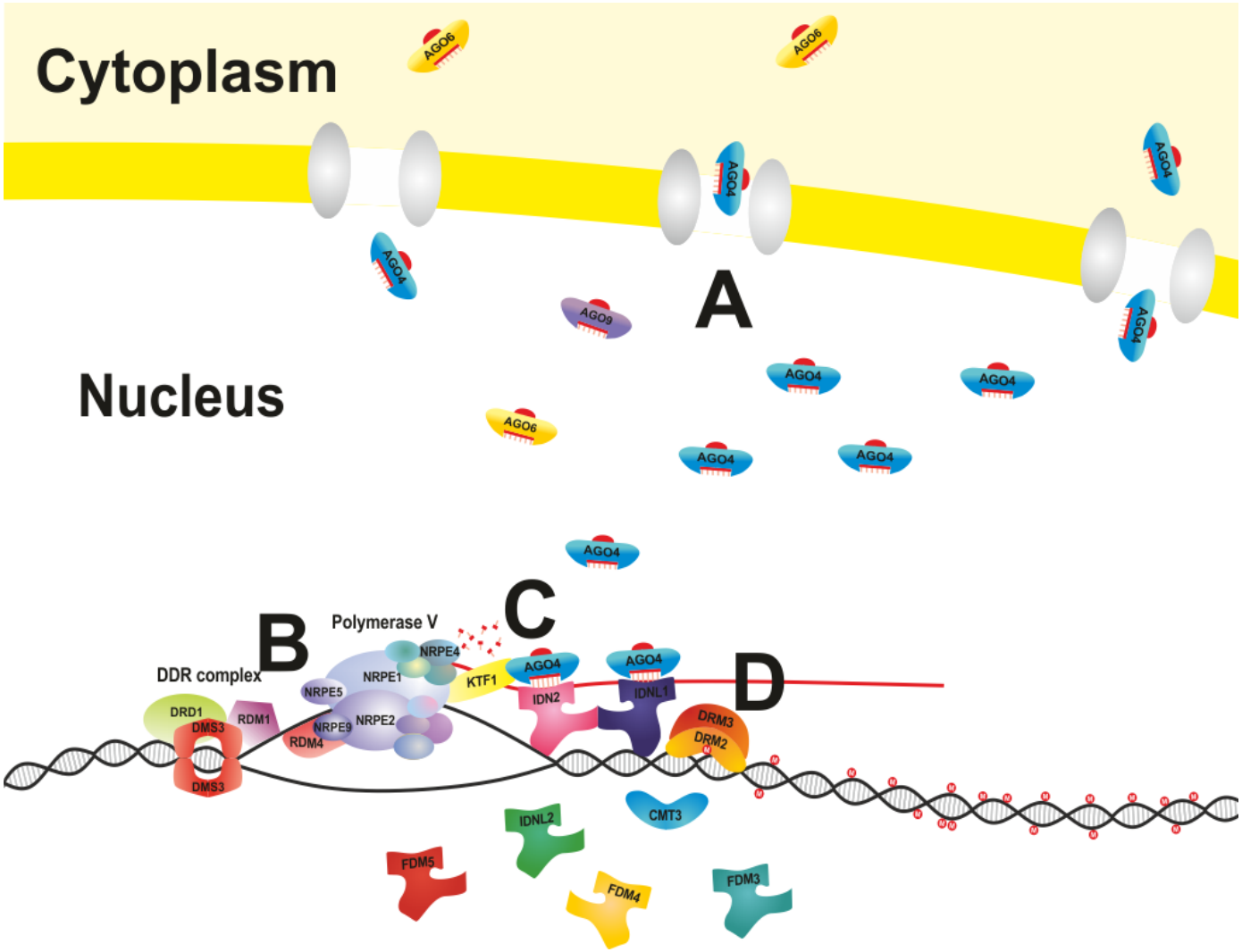

Cytoplasm
AGO6
AGO6
AGO4
AGO4
AGO4
A
AGO9
Nucleus
AGO4
AGO4
AGO6
AGO4
AGO4
AGO4
B
Polymerase V
C
D
DDR complex
NRPE4
NRPE1
NRPE5
NRPE2
NRPE9
KTF1
AGO4
AGO4
IDN2
IDNL1
DRD1
DMS3
RDM1
RDM4
DMS3
DRM3
DRM2
CMT3
IDNL2
FDM5
FDM4
FDM3

Figure 1.11: PolV transcription and DNA methylation in RdDM

A: siRNA loaded AGOs are imported into the nucleus. **B:** The DDR complex allows PolV transcription to occur. **C:** The siRNA loaded AGO binds to the PolV transcript and this binding is facilitated by KTF1. **D:** The duplex between the bound siRNA and PolV transcript is recognised by a IDN2 heterodimer and this triggers DRM2 to methylate the DNA. The nuclear envelope is depicted as a yellow line with protein channels within this membrane shown as pairs of silver ovals. Areas with a white background are within the nucleus and areas in pale yellow are in the cytoplasm. RNA strands are shown in red and DNA strands are shown in black. RNA nucleotides are shown as red and orange T shapes. All proteins have been labelled except some of the PolV subunits that have yet to be shown to be required for RdDM. DNA methylation is shown as a white M in a red circle on the DNA strands.

1.3.5 Reinforcement of DNA methylation by histone modification in RdDM

The next stage of RdDM is reinforcement of DNA methylation by histone modification. DNA methylation can disrupt the binding of transcription factors and other DNA binding proteins due to the presence of the methyl groups on the cytosine residues, however DNA methylation itself is not always sufficient to fully silence a DNA target and instead requires chromatin modifications, such as histone modification, in order to achieve full silencing (Buschhausen et al. 1987, Kass et al. 1993, Eden and Cedar 1994, Kass et al. 1997). DNA methylation can act as a signal for this histone modification and this was shown to be the case for the removal of acetyl groups from histone four (H4), which is a mark found at transcriptionally active sites (Hendrich and Bird 1998, Ballestar and Wolffe 2001, Berg et al. 2003, Zemach and Grafi 2003). The mechanism for this process involves the methyl binding domain (MBD) protein family, found in both plants and animals. These proteins recognise and bind to DNA methylation and then interact with histone deacetylases, triggering these enzymes to deacetylate the nucleosomes within that region thus changing it from transcriptionally active to inactive. In *Arabidopsis* the MBD6 and MBD10 proteins have been shown to have an involvement in RdDM silencing of rRNA loci (Figure 1.12 A), although it is not known if they are involved in silencing at other RdDM target sites or if any of the other eight *Arabidopsis* MBD proteins are also connected to RdDM silencing (Berg et al. 2003, Zemach and Grafi 2003, Preuss et al. 2008). MBD6 recognises CG methylation at the 5S rDNA locus but the preference of MBD10 is not known and the histone deacetylase responsible for the resulting histone deacetylation is thought to be HDA6 as it has also been shown to have a role in rRNA loci silencing (Aufsatz et al. 2002b, Preuss et al. 2008). HDA6 has also been implicated in RdDM and is responsible for the conversion of nucleosomes from a transcriptionally active state to an inactive state through the removal of acetyl groups from H4 (Figure 1.12 B) (Aufsatz et al. 2002b, Probst et al. 2004, Elmayan et al. 2005, He et al. 2009a). Mutation of *hda6* results in a reduction in DNA

methylation at several loci, with CG and CHG methylation being particularly affected and may be the result of a reinforcement loop between DNA methylation and HDA6 whereby DNA methylation recruits HDA6 activity which leads to further DNA methylation. The fact that *hda6* mutants cause a greater loss in CG and CHG methylation than CHH methylation may be related to the fact that the MBD proteins that trigger HDA activity have so far only been shown to bind to CG methylation (Preuss et al. 2008).

There is a second family of methylated cytosine binding proteins with an involvement in RdDM silencing. These are the SUVH family, of which there are nine in *Arabidopsis*, and are characterised by the SRA domain that binds to methylated DNA and the SET domain that has histone methyltransferase activity (Johnson et al. 2008). Of these proteins so far only SUVH2 and SUNVH9 have been implicated in RdDM and act redundantly in the pathway, with mutants in these two genes resulting in a reduction in DNA methylation at several RdDM target loci as well as minor reduction in siRNA levels. This could again be due to the loss of a reinforcement loop between DNA methylation and histone methylation whereby DNA methylation triggers SUVH methylation of histone residues which itself leads to further DNA methylation (Johnson et al. 2007). This has been shown to be the case for SUVH4 which is required for both CHG DNA methylation and di-methylation of the histone three lysine nine (H3K9) residue at sites targeted by CMT3 (Jackson et al. 2002, Malagnac et al. 2002, Lindroth et al. 2004, Johnson et al. 2007). Although both SUVH2 and SUVH9 are largely redundant to each other at most loci, there are some loci where SUVH2 but not SUVH9 is required. It was also shown that the binding preference of the two proteins is different, with SUVH2 preferentially binding to CG methylation whereas SUVH9 preferentially binding to CHH methylation (Figure 1.12 C) (Johnson et al. 2008). Biochemical analysis of SUVH2 and SUVH9 activity also raises questions as to the function of their SET domain as there is conflicting evidence as to whether this domain is capable of di-methylating the H3K9 residue (Ebbs and Bender 2006, Naumann et al. 2005, Johnson et al. 2008). This

raises the possibility that the SET domains of SUVH2 and SUVH9 may have a different function to that of histone modification, however this analysis was carried out in vitro, so it is possible that the SET domain of these proteins requires a co-factor in order to function that was missing in these in vitro studies.

SUP32 (also known as UBP36) is another histone modifier that's function is related to H3K9 di-methylation. It is a deubiquitination enzyme responsible for the removal of ubiquitin groups from the lysine 123 residue of histone 2B (H2BK123) (Figure 1.12 D) (Sridhar et al. 2007). Ubiquitin groups are found at transcriptionally active sites and so their removal results in a shift to an inactive state and it has been shown that their presence inhibits the di-methylation of H3K9 (Ng et al. 2002, Sun and Allis 2002, Sridhar et al. 2007). This means that SUP32 may interact with SUVH2 and SUVH9 in order to remove ubiquitin H4 marks and add H3K9 di-methylation marks in response to DNA methylation in RdDM, however no study has looked at an interaction between SUP32 and these proteins so would require further investigation to determine if this is indeed the case. The final histone modifier that has been associated with RdDM is JMJ14, which is also known as LSD1 and, like HDA6 and SUP32, removes histone modifications that are associated with transcriptionally active sites (Deleris et al. 2010, Searle et al. 2010). In this case the mark is the di-methylation of the lysine four residue of H3 (H3K4) (Figure 1.12 D). *jmj14* mutants result in an increase in di-methylation of H3K4 and a decrease in DNA methylation, with a minor decrease in siRNA levels. This would again suggest a feedback mechanism between DNA methylation and the activity of the JMJ14 protein, however it is not known if JM14 recognises DNA methylation itself or is activated by the action of another as yet undiscovered protein.

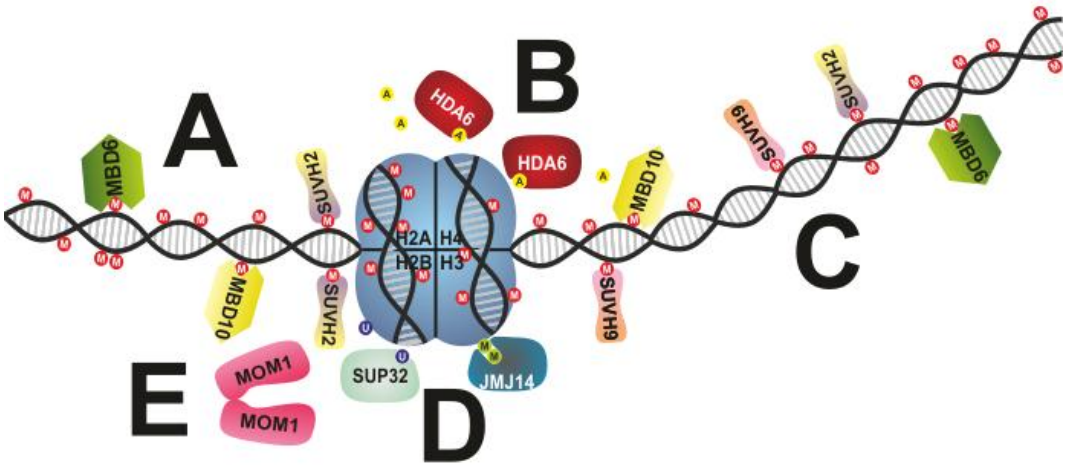


Figure 1.12: Histone modification in response to DNA methylation in RdDM

A: MBD6 and MBD10 bind to methylated cytosine residues. **B:** MBD binding triggers HDA6 to deacteylate nucleosomes. **C:** SUVH2 and SUVH9 also recognise and bind to methylated cytosines. **D:** SUP32 and JMJ14 remove the ubiquitin groups from H2BK123 residues and di-methylation from the H4K4 residues of nucleosomes respectively. **E:** MOM1 forms a homodimer when functioning in RdDM. DNA strands are shown as black lines with grey bars representing base pairing between strands. A nucleosome is depicted in blue with four of the eight histones shown. DNA methylation is shown as a white M in a red circle on the DNA strands while histone methylation is depicted as a black M in a green circle. Histone ubiquitination is shown as a white U in a blue circle and histone actelyation is shown as a black A in a yellow circle.

Another protein that is considered to function downstream of DNA methylation is MOM1. This protein was first identified as a being required for transcriptional silencing at a number of loci but mutants in *MOM1* did not result in a reduction in either DNA methylation or histone modification (Amedeo et al. 2000, Habu et al. 2006, Vaillant et al. 2006). Interestingly these loci targeted by MOM1 have an unusual intermediate chromatin state between that of hetero- an euchromatin and are characterised by the presence of both repressive state H3K9 di-methylation and active state H3K4 di-methylation (Habu et al. 2006). The exact reason for having intermediate chromatin at these loci is unknown but it is

suggested that it allows for the rapid change between active and repressed states in response to environmental factors. A screen for enhancers of loss of silencing in *mom1* mutants identified mutants in PolV and other RdDM components and therefore indicated that MOM1 and RdDM are linked (Yokthongwattana et al. 2010) Further analysis showed that this was loci dependent with some being silenced by both MOM1 and RdDM, others requiring only MOM1 or RdDM and some where MOM1 counteracted RdDM and therefore suggested a complex relationship between RdDM and MOM1. For loci that are targeted by both MOM1 and RdDM there is a modest decrease in H3K9 di-methylation, although not to the same extent as a *drm2 cmt3* double mutants, but no change in DNA methylation or siRNA levels (Numa et al. 2010). This decrease is not observed in loci that are targeted only by MOM1 and it was therefore suggested that MOM1 facilitates histone modification in response to RdDM and therefore works downstream of DNA methylation in the pathway. A study of the domain structure of MOM1 found that its activity requires only a nuclear localisation signal and the novel conserved MOM1 Motif 2 (CMM2) domain (Čaikovski et al. 2008). Subsequent biochemical analysis revealed that the CMM2 domain formed a hendecad-based coiled coil that could form homodimers (Nishimura et al. 2012). This domain is sufficient to maintain silencing at loci where MOM1 and RdDM are required, however the entire MOM1 protein is needed for loci that are silenced by MOM1 only. As no other MOM1 domain is required at RdDM targets it was suggested that the homodimer forms a binding platform for other proteins to bind but currently the identity of such proteins is unknown (Figure 1.12 E). These results would indicated that MOM1 has as yet unspecified role in RdDM during histone modification but only at specific loci and that this function requires dimerization.

1.3.6 Maintenance of DNA methylation by RdDM

Once DNA methylation and histone modification has occurred the target locus is fully silenced but these repressive marks must be maintained. This maintenance

is likely to involve in re-initiation of the entire RdDM pathway and is supported by the fact that 24 nt siRNAs are present in adult plant tissue for loci where initiation of silencing would be expected to occur in the ovule and pollen (Kanno et al. 2005, Brosnan et al. 2007, Mosher et al. 2009, Slotkin et al. 2009). There is also the evidence that one of the DDR complex's components, RDM1, binds preferentially to methylated single stranded DNA and so may target PolV to RdDM targets that have already been methylated (Gao et al. 2010). It is also likely that the maintenance DNA methyltransferase enzymes, MET1 and CMT3 are involved since both enzymes are able to maintain CG and CHG methylation, respectively, independently of RdDM (Jones et al. 2001, Lindroth et al. 2001, Cao and Jacobsen 2002a, Kankel et al. 2003, Aufsatz et al. 2004). CMT3 can also maintain CHH methylation at some loci but others require DRM2 and the RdDM pathway for maintenance (Cao and Jacobsen 2002a). Maintenance of histone modifications may occur through a cycle of reinforcement with DNA methylation whereby histone modification as a result of DNA methylation results in re-initiation of RdDM and further histone modification, such as occurs with CMT3 (Jackson et al. 2002, Malagnac et al. 2002, Lindroth et al. 2004, Johnson et al. 2007).

1.3.7 Summary of the RdDM pathway

The previous sections of the introduction have given a detailed account of the mechanism of the RdDM pathway and this section will now briefly summarise the aforementioned mechanism (Figure 1.13). The pathway is initiated by PolIV transcription that requires the nucleosome modifier, CLSY1, or its homologs and transcription factors such as SHH1 and RDM4 (Herr et al. 2005, Kanno et al. 2005, Onodera et al. 2005, Smith et al. 2007, He et al. 2009b, Kanno et al. 2009, Law et al. 2011, Liu et al. 2011) (Figure 1.13 A). The PolIV transcript is then converted by RDR2 into dsRNA (Figure 1.13 B) before being processed by DCL3 or DCL4 into 24 nt dsRNA segments (Xie et al. 2004) (Figure 1.13 C). These are then methylated at the 3' end of both strands by HEN1 to protect them from degradation before

being transported out of the nucleus (Xie et al. 2004, Ye et al. 2012) (Figure 1.13C and D). In the cytoplasm they are loaded into either AGO4, 6 or 9 and this is facilitated by HSP90 (Havecker et al. 2010, Ye et al. 2012) (Figure 1.13 D). Cleavage and subsequent removal of one of the two strands leaving just the siRNA strand results in a conformational change in the AGO protein that makes a nuclear localisation signal accessible (Ye et al. 2012) (Figure 1.13 E). This allows the siRNA loaded AGO to enter the nucleus and once in the nucleus it binds to a PolV transcript and this is facilitated by KTF1 (Bies-Etheve et al. 2009, He et al. 2009c, Rowley et al. 2011) (Figure 1.13 H). PolV transcription requires the DDR complex, consisting of DMS3, DRD1 and RDM1 as well as transcription factors such as RDM4 (He et al. 2009b, Kanno et al. 2009, Law et al. 2010) (Figure 1.13 G). The duplex created by the siRNA and PolV transcript upon siRNA loaded AGO binding is recognised and bound by a heterodimer of IDN2 and one of its five homologs (Ausin et al. 2009, Zheng et al. 2010, Ausin et al. 2012b, Xie et al. 2012, Zhang et al. 2012) (Figure 1.13 H). The resulting complex allows IDN2 and its homologs to interact with the target DNA through their zinc finger motifs and this is thought to trigger the methylation of the DNA by DRM2 and its non-catalytic co-factor DRM3 (Cao and Jacobsen 2002a, Cao and Jacobsen 2002b, Cao et al. 2003, Chan et al. 2006, Henderson et al. 2010, Ausin et al. 2012b, Xie et al. 2012, Zhang et al. 2012) (Figure 1.13 I). DNA methylation is then reinforced by histone modifiers and the methylated DNA is first recognised by MDB6 and 10; and SUVH2 and 9 (Figure 1.13 J). MDB6 and 10 binding results in HDA6 removing the acetyl groups from H4 while SUVH2 and 9 di-methylate the H3K9 residue (Aufsatz et al. 2002b, Johnson et al. 2008, Preuss et al. 2008) (Figure 1.13 K). This di-methylation is facilitated by the MOM1 protein and requires that SUP32 must first deubiquitinates the H2BK123 residue (Sridhar et al. 2007, Numa et al. 2010). JMJ14 also removes di-methylation of the H3K4 residue in response to DNA methylation (Deleris et al. 2010, Searle et al. 2010). The DNA methylation and histone modification is then maintained by re-initiation of the pathway and the action of the maintenance methyltransferases MET1 and CMT3 (Jones et al.

2001, Lindroth et al. 2001, Cao and Jacobsen 2002a, Kankel et al. 2003, Aufsatz et al. 2004).

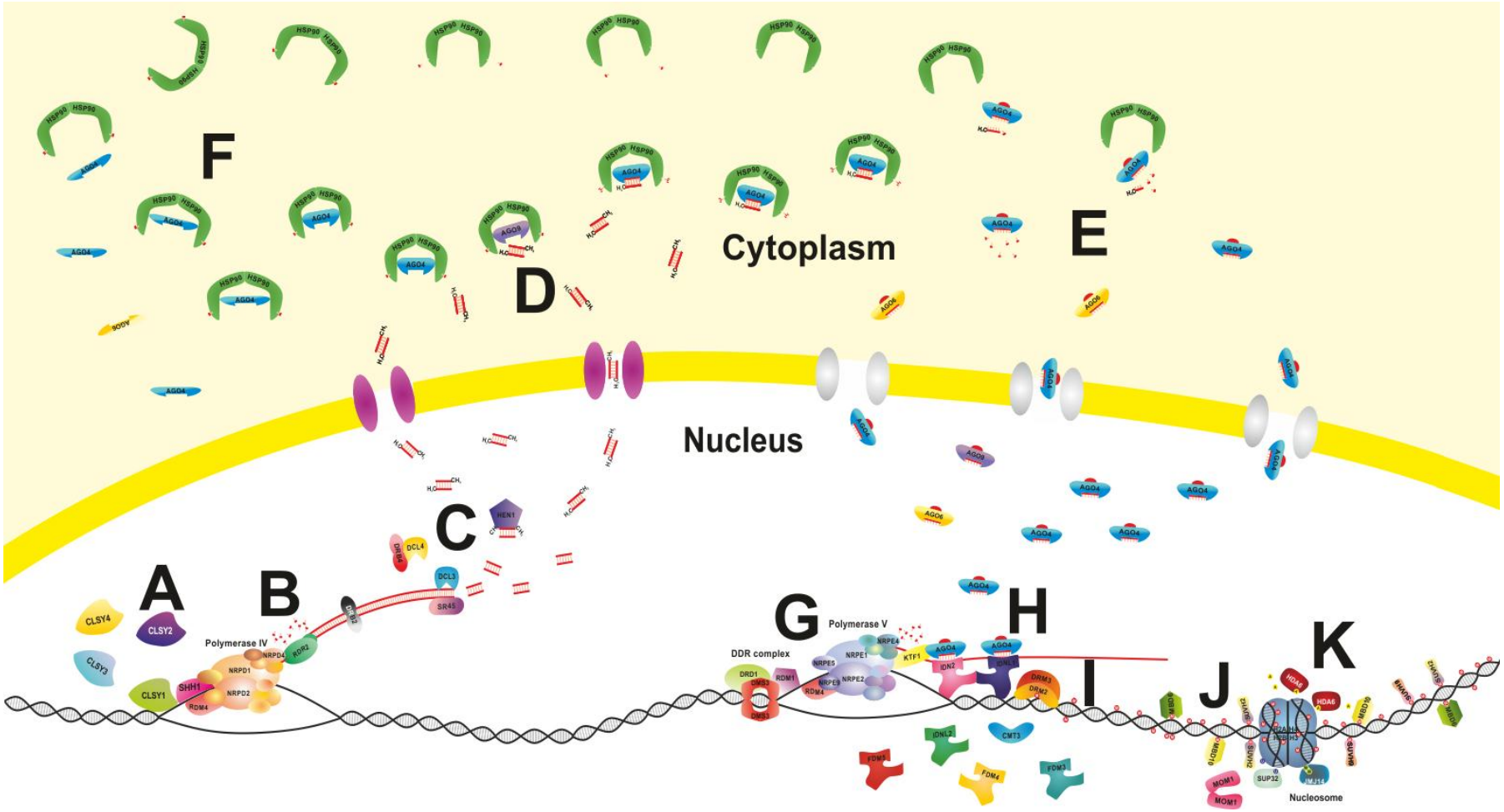

F
Cytoplasm
E
D
Nucleus
C
A
B
G
H
I
J
K
CLSY4
CLSY2
CLSY3
CLSY1
Polymerase IV
NRPD4
RDR2
NRPD1
NRPD2
SHH1
RDM4
DCL4
DCL3
SR45
HEN1
DDR complex
DRD1
DMS3
RDM1
Polymerase V
NRPE4
NRPE1
KTF1
NRPE5
NRPE2
NRPE3
RDM4
IDN2
DRM3
DRM2
CMT3
FDM2
FDM4
FDM3
HDA6
HDA9
SUP32
MOM1
Nucleosome

Figure 1.13: The RdDM pathway

A: Transcription factors and nucleosome modifiers allow PolIV to transcribe the target DNA locus. **B:** The PolIV transcript is turned into dsRNA by RDR2. **C:** The dsRNA is then cleaved into 24 nt segments by DCL3 and DCL4 and the 5' ends are then methylated by HEN1. **D:** The 24 nt dsRNA molecules are then exported from the nucleus and loaded into AGO4, 6 or 9. **E:** The passenger is cleaved by the AGO and is then degraded leaving the siRNA loaded into the AGO. This results in the release of the AGO from HSP90 and the exposure of the nuclear localisation signal (in red) that triggers the siRNA loaded AGO to be imported into the nucleus. **F:** The HSP90 binds to a new AGO and causes conformation changes in the AGO protein that allows it to bind to a 24 nt dsRNA molecule. **G:** The DDR complex facilitates PolV transcription of the target locus. **H:** The siRNA loaded AGO binds to the PolV transcript and this is aided by KTF1. The duplex formed between the siRNA and PolV transcript is recognised by a IDN2 heterodimer. **I:** The IDN2 heterodimer triggers DRM2 and its cofactor DRM3 to methylate the target DNA locus. **J:** The methylated DNA is recognised and bound to be DNA-binding proteins. **K:** These DNA binding proteins trigger the modification of histones with repressive marks. The nuclear envelope is depicted as a yellow line with protein channels within this membrane shown as pairs of purple ovals if they are exporters and silver ovals if they are importers. Areas with a white background are within the nucleus and areas in pale yellow are in the cytoplasm. RNA strands are shown in read and DNA strands in black. Methylation at the 3' end of siRNAs is shown by CH_3 labels and RNA nucleotides are shown as red and orange T shapes. The twelve subunits of RNA polymerase IV have been depicted in red, yellow and orange, but only subunits with a proven role in RdDM have been labelled and the twelve subunits of RNA polymerase V have been depicted in blue, green and purple, but only subunits with a proven role in RdDM have been labelled. DNA methylation is shown as a white M in a red circle on the DNA strands while histone methylation is depicted as a black M in a green circle. Histone ubiquitination is shown as a white U in a blue circle and histone actelyation is shown as a black A in a yellow circle.

1.3.8 Other RNA directed transcriptional gene silencing pathways in plants

There is some evidence for other plant RNA silencing pathways that silence targets by DNA methylation. miRNAs are usually considered to function purely in PTGS but miRNAs have also been shown to be capable of directing DNA methylation. In *Arabidopsis* it was shown that two genes, *PHABULOSA (PHB)* and *PHAVOLUTA (PHV*), encoding transcription factors involved in early leaf development, ar*e* regulated by the miRNAs miR165 and miR166 and that this is associated with DNA methylation corresponding to the sequence downstream of the mRNA cleavage site of miR165 and miR166 (Bao et al. 2004). This methylation is lost in *phb* and *phv* mutants lacking the miRNA cleavage sites and is reduced in undifferentiated tissue where miRNA repression of these two genes does not occur, therefore suggesting that the DNA methylation is associated with silencing. However, it was not shown whether this methylation is the cause of the silencing or if it occurs as a consequence of PTGS by the miRNA; neither is the mechanism for how the miRNAs trigger methylation known and may still feature RdDM. miRNA-directed DNA methylation was however shown in the moss *Physcomitrella patens* where in *dcl1b* mutants that can still produce mature miRNAs but are unable to cleave target mRNA, a build-up of miRNA-target RNA duplexes resulting in DNA methylation of the target locus (Khraiwesh et al. 2010). DNA methylation of the promoter of a miRNA target gene, which is down regulated by abscisic acid signalling, occurred in wild-type *P. patens* upon treatment with the hormone, thus suggesting that DNA methylation by miRNA can occur naturally in this species. The ratio of miRNA and target RNA is believed to be important for whether silencing occurs through DNA methylation or PTGS. *These tw*o studies showed a possible role in DNA methylation for 21 nt miRNAs, which is the most prevalent length of miRNA, but it has also recently been shown that there are 24 nt miRNAs which are processed by DCL3 in *Arabidopsis*; but are at a considerably lower abundance to their 21 nt counterparts (Dunoyer et al. 2004, Vazquez et al. 2008). These 24 nt miRNAs are also found in *Oryza sativa*

and a study in this species showed that these miRNAs are loaded into the *O. sativa* AGO4 homolog and direct DNA methylation (Wu et al. 2010). It was suggested that this pathway may be similar to that RdDM, except, due to the nature of miRNA production, that it is not dependent on RDR2 and PolIV and its transcriptional co-factors, of which RDR2 was indeed shown not to be required for miRNA-directed DNA methylation. The 24 nt miRNAs are produced from the same transcript as 21 nt miRNAs for the same target, with the 21 nt and 24 nt miRNAs being encoded at different point on the transcript or overlapping each other, resulting in DCL1 and DCL3 competing to produce either a 21 or 24 nt miRNA respectively. This suggests that for the targets of these miRNA, genes are targeted for both PTGS and TGS and it is possible that these two systems are interlinked, although this has not so far been tested. A 24 nt miRNA has also been shown to be involved in DNA methylation of the *SP11* male self-incompatibility gene in *Brassica rapa,* indicating that the role of 24 nt miRNAs in DNA methylation is not just specific to *O. sativa,* although it has so far not been shown whether any of the *Arabidopsis* 24 nt miRNAs mediate DNA methylation (Tarutani et al. 2010, Finnegan et al. 2011). These four examples point to miRNAs being able to direct DNA methylation at specific loci, although the mechanism for this process is incomplete.

Two recent studies have also demonstrated that RNA silencing components that are traditionally associated with PTGS are required for DNA methylation but only at specific loci (Garcia et al. 2012, Pontier et al. 2012). The loci were originally identified as losing DNA methylation in a mutant for a novel chromatin modifier, NERD, which was shown to bind to unmethylated H3K4 residues that are found at transcriptionally inactive sites and that this methylation loss was associated with siRNA loss (Pontier et al. 2012). Analysis of these loci to determine which RDR is required for DNA methylation showed that RDR1 and RDR6 were required rather than RDR2. Interestingly neither *rdr1* nor *rdr6* mutants resulted in a decrease in siRNA levels associated with these loci whereas *rdr2* mutants did and so raises questions as to the exact function of RDR1 and RDR6 as they appear not

to produce dsRNA for siRNA production but are still required for DNA methylation. The DCL requirement of these loci is also unusual as for most loci single mutants in *dcl2, dcl3* or *dcl 4* were sufficient to lead to a reduction in DNA methylation but at one locus only a line with all four *DCL* genes mutated caused a reduction in DNA methylation and so this would suggest redundancy in DCL requirement at these loci. In terms of AGO requirement, neither AGO4 nor AGO6 were required for DNA methylation and instead AGO2 or its homolog AGO3 were required. The study by Garcia also demonstrated that for at least one of the loci the PTGS components SGS3 (a homolog of IDN2), SDE3 (a RNA helicase) and SDE5 were also required for DNA methylation (Garcia et al. 2012). These findings raise the possibility that PTGS components are involved in DNA methylation, however it remains to be determined as to whether they form an entirely novel pathway for DNA methylation or are a locus specific adaption of RdDM pathway. The fact that PolIV and PolV have also been showed to be required for DNA methylation at these loci as well as DCL3 and DCL4 may suggest RdDM or at least an RdDM like pathway is responsible (Xie et al. 2004, Pélissier et al. 2011, Garcia et al. 2012, Pontier et al. 2012).

1.3.9 RNA directed TGS silencing in other eukaryotes

RdDM is plant specific RNA silencing pathway as many of the key components, particularly RNA PolIV and V, are unique to plants and more specifically to higher plants as Pol V is only present in plant species that evolved after the divergence of ferns from other land plants (Luo and Hall 2007, Marcussen et al. 2010). This means that the pathway described in the previous sections applies only to higher plant species, while the mechanism in lower plant species is currently unknown but may be similar to the one described as PolIV is present in these species. Despite RdDM being plant specific there are RNAi pathways in other eukaryotic organisms that are involved in TGS. The RNAi pathway in fission yeast responsible for TGS is similar to RdDM in that it also involves the action of an RNA Polymerase, in this case PolII, and that siRNAs are produced from the PolII

transcript by the actions of an RDR and Dicer (Volpe et al. 2002, Verdel et al. 2004, Sugiyama et al. 2005, Irvine et al. 2006). There is also similarity between the two pathways in that once the siRNA is loaded into an AGO it triggers transcriptional silencing by binding to a PolII transcript, however unlike RdDM where binding to a PolV transcript triggers DNA methylation, the binding to PolII transcripts in fission yeast results in di-methylation of the H3K9 residue. Unlike the fission yeast pathway, the *Drosophila* and *Mus musculus* TGS RNAi pathways are highly dissimilar to RdDM. The RNAi pathway in these organisms responsible for TGS is mediated by PIWI RNAs (piRNA) that are between 25-30 nt in length and are produced by a mechanism that is different to that of RdDM (Aravin et al. 2001, Vagin et al. 2006, Aravin et al. 2007, Brennecke et al. 2007, Carmell et al. 2007, Gunawardane et al. 2007). This mechanism is described as “ping pong” and involves the PIWI proteins, a subset of AGO proteins and unlike RdDM does not feature RDR or Dicer enzymes. The mechanism involves the initial production of piRNAs from an antisense transcript, which are then loaded into a PIWI protein (PIWI in *Drosophila* and MIWI2 in *M. musculus*) that directs cleavage of a sense transcript, creating sense piRNAs which are then loaded into another PIWI protein (AGO3 in *Drosophila* and MILI in *M. musculus*) that then cleave an antisense transcript and create further antisense piRNAs. Another difference between piRNAs and RdDM is that piRNAs are specific to germ cells where they trigger histone modification or DNA methylation in *Drosophila* and *M. musculus* respectively, whereas RdDM is found in somatic cells as well as germ cells (Pal-Bhadra et al. 2004, Aravin et al. 2008, Aravin and Bourc’his 2008, Kuramochi-Miyagawa et al. 2008). Although the mechanism of piRNA production is different to RdDM, both silence transposable elements in order to prevent the deleterious movement of these elements and so could be considered to have similar roles (Huettel et al. 2006, Carmell et al. 2007, Kuramochi-Miyagawa et al. 2008, Slotkin et al. 2009).

1.4 RdDM function

1.4.1 Silencing of transposable elements and repetitive sequences

There are several proposed functions for RdDM including: silencing of transposable elements and repeat elements, control of development, stress responses and hybrid vigor. The best understood of these is the role in silencing transposable elements and repetitive sequences. It has been found that RdDM targets these genetic elements in areas of constitutive heterochromatin, such as the pericentromeric region, presumably to help maintain the heterochromatic state of the DNA and to prevent the proliferation of transposable elements in the genome which could have deleterious effects on the plant (Kato et al. 2003, Herr et al. 2005, Kanno et al. 2005, Onodera et al. 2005, Pontier et al. 2005, Lisch 2009, Lisch and Bennetzen 2011). In support of this, in RdDM mutants the expression level of transposable elements that are targeted is up regulated (He et al. 2009c, Pontes et al. 2009). However, with the exception of *nrpe1* mutants, no other RdDM mutant results in heterochromatin decondensation suggesting that the pathway is actually dispensable for heterochromatin maintenance and could suggest redundancy with other pathways (Pontes et al. 2009). This lack of an effect in RdDM mutants is also seen at telomeres where the pathway is responsible for CHH methylation of the CCCTAAA repeats that form the telomere (Cokus et al. 2008, Vrbsky et al. 2010). RdDM was therefore thought to have a role in maintaining the telomere heterochromatin and yet in an *rdr2* mutant where methylation and siRNA levels were reduced, the heterochromatin structure of the telomeres was unaffected. This would again suggest other epigenetic pathways involvement in telomere maintenance and so to demonstrate a role for RdDM in heterochromatin maintenance other pathways may also need to be inactivated.

RdDM also targets transposable element and repetitive sequences in euchromatic genome regions, particularly small transposable elements and this is again likely to be to prevent their movement within the genome (Tran et al. 2005, Zilberman et al. 2007). Another role for RdDM in targeting these euchromatin features is in controlling gene expression. A study looking at long terminal repeats (LTR) of retrotransposons found that, of those targeted by RdDM, only those in euchromatic regions were up regulated in mutants of the pathway whereas ones in heterochromatic regions were not (Huettel et al. 2006). This would suggest that unlike in heterochromatic regions, RdDM is the chief pathway responsible for silencing transposable elements and repetitive loci in euchromatic regions and so has a more significant role in the maintenance of repressive marks within euchromatin than heterochromatin. Further investigation also showed that these LTR regions also seemed to affect transcription from nearby genes as in one example a Solo LTR on chromosome five that was reactivated in RdDM mutants, triggered up regulation of two nearby genes (Huettel et al. 2006). The Solo LTR was shown to possess enhancer elements that were activated in RdDM mutants. This could suggest that RdDM and demethylation pathways may potentially be able to alter gene expression levels by alteration of the methylation status of such elements. This idea will be taken further in the section on stress response and RdDM.

The role of RdDM in control of transposable elements has particular relevance during gametogenesis. During pollen development each pollen grain produces two sperm cells and a vegetative nucleus, of which only the sperm cells go on fertilise the egg and central cell in the ovule resulting in the embryo and endosperm, respectively (Huh et al. 2008). Investigation into transposable element activity in pollen revealed that they were expressed and active in the vegetative nucleus but not the sperm cells and that 21 nt siRNAs associated with such elements were produced in the vegetative nucleus and then transported into the sperm cells (Slotkin et al. 2009). This would suggest that these siRNAs protect the sperm cells from transposable element activation and thus prevents

the deleterious effects of uncontrolled transposition being transferred to the plant's progeny. The reason why this is required is that during gametogenesis DNA methylation is reset and so methylation of transposable elements and repetitive loci must be re-established. In support of this, it was shown that DNA methylation of these elements had indeed occurred in mature sperm cells. Since RdDM is responsible for *de novo* methylation it is assumed that it is responsible for this methylation, although this has not been proven, particularly as 21 nt siRNAs are not normally associated with RdDM. There is stronger evidence of a function for RdDM in the endosperm. The endosperm is produced by the fertilisation of the central cell by a sperm cell and it was found that extensive demethylation by DME occurred in the endosperm tissue (Gehring et al. 2009, Hsieh et al. 2009). This demethylation of the endosperm correlates with changes in the embryo whereby hypermethylation of transposable elements and imprinted genes occurs and so it was suggested that the demethylation process in the endosperm triggers methylation in the embryo. This was supported by the discovery that 24 nt siRNAs expression is up regulated during female gametogenesis and in the endosperm and so these siRNAs, like those in pollen, could be transported from the endosperm to the embryo where they direct epigenetic reprogramming of the genome (Mosher et al. 2009). The up regulation during female gametogenesis could be a similar process to that in pollen. However, the exact mechanism by which this up regulation occurs is unknown as demethylation mutants did not affect siRNA levels and neither did mutations in other chromatin modifiers such as the histone deacetylases (Mosher et al. 2011). Another study looking at the effects of genome dosage on seed development in *Arabidopsis* found that if a tetraploid male was crossed with a diploid female the seed were larger than normal due to a delay in normal endosperm development whereas the reciprocal cross results in smaller seeds (Lu et al. 2012). In these crosses there is a decrease in 24 nt siRNA levels in the male tetraploid to female diploid cross indicating that RdDM may be responsible for the larger seed size. It was then shown that when a wild-type male diploid

plant is crossed with a *nrdp1* female diploid plant, this resulted in a similar seed size as when a male tetraploid is crossed with a female diploid and that the increased size was associated with up regulation of the endosperm specific *AGAMOUS-LIKE* transcription factor due to a reduction in 24 nt siRNAs targeting this gene. This supports the idea that RdDM is involved in gene imprinting in the endosperm as only the maternal copy of the *AGAMOUS-LIKE* gene is expressed. Taken together the studies appear to show that RdDM is involved in gene imprinting and reprogramming the epigenome in the embryo and endosperm, although apart from an increase in seed size RdDM mutants do not display any defects in seed development, meaning that further investigation is required to determine the significance of its role in these processes.

1.4.2 RdDM role in development

One of the biggest mysteries associated with RdDM is that mutants in the pathway do not show any strong phenotype apart from the up regulation of transposable elements as mentioned in the last section. This is exemplified by the contrasting *met1* mutant phenotype which shows severe pleotropic developmental defects whereas *drm2* mutants do not (Finnegan et al. 1996, Ronemus et al. 1996, Cao and Jacobsen 2002a, Cao and Jacobsen 2002b, Cao et al. 2003). As previously suggested this is likely down to redundancy with other epigenetic pathways and in support of this triple mutants in *cmt3, drm1* and *drm2* have developmental defects such as later flowering time, twisted leaves and dwarf stature that are not seen in the respective single mutants (Cao et al. 2003, Chan et al. 2006). This has also shown to be the case in double mutants of *cmt3* and other RdDM mutants such as either *nrpd1* or *drd1. Arabidopsis dcl3, nrpd1* and *nrpe1* mutants also show late flowering that is most obvious in short day growth conditions (8 hours) suggesting that components of the pathway have a role in the regulation of flowering time (Pontier et al. 2005, Greenberg et al. 2011). This effect on flowering is thought to be due to changes in the regulation of the *FLC* locus and will be discussed in more detail in the

introduction of chapter five (page 293) (Swiezewski et al. 2007). These results could suggest that RdDM is of low importance to plant growth and development due to the limited phenotype in single mutants in the pathway. However in maize, mutants in the homolog of *NRPD1* do show developmental defects in leaf and inflorescence tissue and would suggest that RdDM is biologically relevant in maize (Parkinson et al. 2007, Erhard Jr. et al. 2009). No *Arabidopsis* single mutant of *NRPD1* or *NRPD2* has shown such effects and so raises the question of why they are seen in maize but not *Arabidopsis.* One possibility is that there is a difference in function in RdDM between the two species and another is the fact that all investigations of RdDM mutants have so far only looked at mutants in ideal growth conditions so a RdDM phenotype may only appear in more natural conditions where the plants are subject to changes in growth conditions and resulting stresses. Investigation of RdDM *Arabidopsis* mutants under different conditions and over multiple generations may reveal both developmental and stress response defects.

1.4.3 RdDM role in stress response

The involvement of DNA methylation in stress response is well known but the evidence for the role of the RdDM pathway in stress responses is more limited. Microarray analysis of gene expression in response to stress found several genes involved in the RdDM pathway were up-regulated and although circumstantial, may suggest that RdDM is more active during stress conditions (Chinnusamy and Zhu 2009). Stronger evidence for a role in RdDM came from a study that looked at promoters of both biotic and abiotic stress response genes and whether these promoters were associated with 24 nt siRNAs, DNA methylation or transposable elements (Baev et al. 2010). They found that around 13 % of biotic stress response genes and 18 % of abiotic stress response genes were associated with 24 nt siRNAs and so could potentially be regulated by RdDM. Amongst these a subset also had associated DNA methylation and transposable elements within the promoter region and analysis of three of these promoters showed that two

lost DNA methylation in *nrpd1* and *nrpe1* mutants. This suggests that RdDM represses these stress response genes in normal conditions and that methylation could be removed by demethylation during stress conditions allowing gene expression to occurs. The reason that the study looked at promoters with transposable elements is that, as described previously, it has been shown that transposable elements that are silenced by RdDM can influence the expression of nearby genes, thus providing a route through which RdDM can control gene expression (Huettel et al. 2006). The findings of the study by Baev were validated by a recent study which analysed expression of *ONSEN* retrotransposons, which are targeted by RdDM for suppression and are found to be upregulated in *nrpd1* and *nrpe1* mutants, during heat stress (Ito et al. 2011). The study found that expression of *ONSEN* was upregulated during and immediately after heat stress before gradually decreasing back to pre-stress levels and that this up regulation of *ONSEN* is associated with the increased expression of nearby genes, two of which are stress response genes. The heat dependent release of *ONSEN* therefore provides an example of RdDM functioning in the control of stress response genes through the control of a transposable element.

Two studies have also provided evidence that RdDM is involved in the regulation of genes relevant to stress response, although in both cases they did not show RdDM altering expression levels in response to stress. One of these studies identified mutants in a enhancer element of *HKT1* that increased resistance to salt stress (Baek et al. 2011). HKT1 is responsible for the importation of Na^+ ions into cells and so down regulation results in a decrease in the number of ions entering the cell, which increases the plants resistance to salt stress. The mutants in the enhancer element reduced expression of *HKT1* and it was also shown that this element was likely controlled by RdDM as *rdr2* mutants lost methylation of the enhancer and resulted in up regulation of *HKT1* expression. Methylation of this enhancer differed between roots and leaves suggesting that RdDM is involved in tissue specific regulation, however it was not shown whether RdDM actually regulates *HKT1* expression in response to stress. Another

study found that *dcl3* mutants resulted in an increase in stress tolerance to the mutagenic agent methyl methane sulfonate, implicating a loss of RdDM control of certain loci was the cause of this tolerance (Yao et al. 2010). However, the genes involved were not identified and therefore it is also not known if RdDM regulation of these genes is altered in response to this mutagen in wild-type plants.

1.4.4 RdDM and hybrid vigor

Recent discoveries have suggested a role for RdDM in hybrid vigor, also known as heterosis, which is the phenomenon whereby the progeny of parents from different species, ecotypes, varieties or strains, display an improvement or increase in certain traits compared to its parents (Meyer et al. 2004, Chen 2010). For example crossing the Columbia and C24 *Arabidopsis* ecotypes produces progeny with significantly larger rosette size than either parent. A study investigating the differences between wild-type plants from the C24 and Landsberg *erecta* (Ler) ecotypes and hybrids of the two ecotypes revealed a decrease in 24 nt siRNAs in the progeny compared to the parental lines and that these changes correlated to regions of the *Arabidopsis* genome where the two ecotypes had differences in 24 nt siRNA levels (Groszmann et al. 2011). Further investigation showed that a majority of these loci correlated to genes or nearby intergenic loci and that the changes in siRNA levels were associated with DNA methylation changes. This would seem to indicate that the changes in siRNA levels for these genes and intergenic regions may account for some of the differences between parents and hybrids and indeed similar results showing a decrease in 24 nt siRNAs in hybrids has also been described in maize (Barber et al. 2012).

1.5 Project aims

1.5.1 Prospects of identifying novel RdDM mutants in a forward genetic screen

The RdDM pathway has been well characterised and large numbers of proteins have been shown to be involved, however gaps in our knowledge still remain. For instance there is uncertainty about how DNA methylation results in the formation of silenced chromatin. As mentioned previously DNA methylation may not be sufficient to achieve a fully silenced state and so chromatin modification is also required (Buschhausen et al. 1987, Kass et al. 1993, Eden and Cedar 1994, Kass et al. 1997). This includes the formation of higher order heterochromatin by the tight packing of nucleosomes into a coiled structure similar to a solenoid and the formation of loops of these coils, both of which will result in a fully silenced state. It is thought that such changes occur in response to RdDM, however no proteins have yet been identified in the RdDM pathway that cause these higher order chromatin. There are currently a total of eight proteins that are involved in histone modification in response to DNA methylation have so far been identified but there is still scope for other histone modifiers being involved. For example there several protein families, namely MBD, HDA and SUVH, which have members that are known to be a part of the RdDM pathway. Not all of the other genes in these families have been tested and so could also potentially be involved in the pathway as well.

Another gap in the mechanism of RdDM is how DNA methylation by DRM2 is triggered. Previous studies have shown that 24 nt siRNAs loaded into AGO4 bind to the PolV transcript and that it is this binding that triggers DRM2 to methylate the target DNA locus (El-Shami et al. 2007, Bies-Etheve et al. 2009, He et al. 2009c, Wang and Dennis 2009, Rowley et al. 2011). Recent studies have shown that IDN2 and its homologs recognise and bind to the siRNA and PolV transcript duplex and then interact with the target DNA through their zinc finger motif

(Ausin et al. 2009, Ausin et al. 2012b, Xie et al. 2012, Zhang et al. 2012). This was suggested to signal to DRM2 to methylate the DNA target, however no interaction between DRM2 and IDN2 and its homolog, nor any other RdDM protein associated with PolV or AGO4 has been demonstrated. Thus it is still unknown how the RdDM pathway signals DRM2 to methylate the target DNA locus and is therefore possible that other, as yet undiscovered, RdDM proteins may exist that sequester DRM2 to the correct target locus and trigger methylation.

Another aspect of the pathway where there is uncertainty is that PolIV is thought to transcribe the target DNA locus and this transcript is then turned into 24 nt siRNAs through the action of RDR2 and DCL3, however how PolIV is recruited to the correct DNA loci is unknown (Herr et al. 2005, Kanno et al. 2005, Onodera et al. 2005, Vaucheret 2005). There are several transcription factors and chromatin modifiers that interact with PolIV including RDM4, SHH1 and CLSY1, 2, 3 and 4, and these may guide PolIV to specific target sites, however no study has yet demonstrated that this is the case (Smith et al. 2007, He et al. 2009b, Kanno et al. 2009, Law et al. 2011, Liu et al. 2011). This modest number of transcription factors and chromatin modifiers would seem unlikely to be able to account for the targeting of PolIV to the myriad number of loci that RdDM is active at and would therefore raise the possibility that there are other proteins that interact with and guide PolIV. This problem of the targeting of PolIV also occurs with PolV, which has been shown to require RDM4 and the DDR complex, comprising of RDM1, DRD1 and DMS3, for transcription but no study has yet shown how PolV is recruited to the correct locus (Kanno et al. 2004, Huettel et al. 2007, Kanno et al. 2005, Kanno et al. 2008, Ausin et al. 2009, He et al. 2009b, Kanno et al. 2009, Gao et al. 2010, Law et al. 2010). There are an estimated to be between 1400 to 2000 transcription factors in *Arabidopsis*, most of which will only interact with PolII, but others may interact with PolIV or PolV, or as with RDM4, interact with PolII, PolIV and PolV (Davuluri et al. 2003, Guo et al. 2005, Iida et al. 2005, Riano-Pachon et al. 2007, He et al. 2009b, Kanno et al. 2009). It is also worth

mentioning that a large number of the proteins involved in the pathway have been discovered since this project began in 2008. Of the areas highlighted in this section where there is uncertainty in the mechanism this study will focus on chromatin modification and DNA methylation.

1.5.2 Strategy for the identification of novel mutants

The aim of this study was to identify novel RdDM components and investigate their function in the pathway. This was achieved using a forward genetic screen of an *Arabidopsis* ethyl ester methane sulfonic acid (EMS) mutant library for RdDM mutants. Screening mutants for a loss of RdDM silencing of endogenous targets would be time consuming, thus a transgene system that allows for easy and rapid RdDM mutant identification was used. The transgene system and how it works is described in the introduction of chapter three (page 124). Mutants that lost RdDM silencing of the transgene could then be characterised in terms of their effect on DNA methylation and siRNA production in order to sort them into phenotypic groups. Allelism tests between known mutants with a similar phenotype to the mutants would then be performed to identify any mutants that have already been identified. For those that were presumed to be novel the causal mutation would be identified by a dual mapping and sequencing strategy. If the mutant was indeed novel the phenotype would be fully characterised in order to determine a possible function in the RdDM pathway.

2. Material and Methods

2.1 Equipment and chemicals

2.1.1 List of chemicals

Below is a list of chemicals used to make solutions mentioned in the following sections. The abbreviation and manufacturer of each chemical are given. Distilled water dH_2O was obtained from a Purelab Ultra water purifier (ELGA).

Chemical	Abbreviation	Manufacturer
1-ethyl-3-(3-dimethylaminopropyl) carbodiimide	EDC	Sigma-Aldrich
1-methylimidazole		Sigma-Aldrich
2-(N-morpholino)ethanesulfonic acid	MES	Sigma-Aldrich
3-(N-morpholino)propanesulfonic acid	MOPS	Sigma-Aldrich
30 % Acrylamide		National Diagnostics
Agar		Sigma-Aldrich
Agarose		Melford
Ammonium persulfate	APS	Sigma-Aldrich
Bacto agar		Sigma-Aldrich
Boric acid		Fisher
Bovine serum albumin	BSA	Sigma-Aldrich
Bromophenol blue		Fisons indicator
Calcium di-chloride	$CaCl_2$	Sigma-Aldrich
Carbenicillin		Sigma-Aldrich
Cetyltrimethylammonium bromide	CTAB	Sigma-Aldrich
Chloroform: isomyl alcohol 24: 1		Sigma-Aldrich
Di-sodium hydrogen phosphate	Na_2HPO_4	Fisher
DL-phosphinothricin		Melford
Ethanol		Fisher
Ethidium bromide	EtBr	Sigma-Aldrich
Ethylenediaminetetraacetic acid	EDTA	Fisher
Formamide		Sigma-Aldrich
Glycerol		Fisher
Hydrochloric acid	HCl	Fisher
Isopropanol		Fisher
Kanamycin		Sigma-Aldrich
Mannitol		Sigma-Aldrich
Polyethylene glycol 8000	PEG (8000)	Fisher
Polyvinylpyrrolidone 40	PVP 40	Sigma-Aldrich
Potassium chloride	KCl	Sigma-Aldrich
Sodium acetate	NaAc	Fisher

Chemical	Abbreviation	Manufacturer
Sodium chloride	NaCl	VWR
Sodium di-hydrogen phosphate	NaH_2PO_4	Fisher
Sodium dodecyl sulfate	SDS	Sigma-Aldrich
Sodium hydroxide	NaOH	Fisher
Tetramethylethylenediamine	TEMED	Bio Rad
Tris		Invitrogen
Tri-sodium citrate	Na_3CIT	Fisher
Tryptone		Oxoid
Urea		Fisher
Xylene cyanol FF		BDH
Yeast extract		Oxoid
β-mercaptoethanol		Sigma-Aldrich

Table 2.1: Chemicals used in this study

2.1.2 Centrifuges

For samples in microcentrifuge tubes (Sarstedt) a S41SP microcentrifuge (Eppendorf) was used for centrifugation or if the samples needed to the centrifuged at 4 °C a 4515 R microcentrifuge (Eppendorf) was used. For samples in centrifuge tubes (15 ml or 50 ml) (Sarstedt) a CR312 centrifuge (Jouan) was used. For 96 well DNA extractions a 4K15C 96 well plate centrifuge (Sigma-Aldrich) was used.

2.1.3 Pipettes

A set of Gilson pipettes (P2, P10, P100, P200 and P1000) were used for all volumes smaller than 1 ml and a Pipetboy (Integra) and 10 ml or 25 ml pipettes (Sarstedt) were used for volumes up to 30 ml. For higher volumes measuring cylinders were used.

2.2 Chemical solutions

2.2.1 1X CTAB

To produce 200 ml of 1 X CTAB DNA extraction buffer, 4 g CTAB, 20 ml 1 M Tris pH 8, 0.25 M EDTA pH 8, 16.4 g NaCl and 4 g PVP 40 were added to 150 ml dH_2O and then the volume was brought up to 200 ml by the addition of dH_2O.

2.2.2 5 X TBE

1 l of 5 X tris-boric acid-EDTA buffer was made by the addition of 54 g tris, 27.5 g boric acid and 20 ml of 0.5 M EDTA pH8 to 980 ml dH_2O.

2.2.3 6 X Loading dye

To produce 30 ml of 6 X loading dye, 9 ml of glycerol, 75 µg bromophenol blue, 75 µg of xylene cyanol FF and 300 µl 0.5M EDTA pH8 was added to 15 ml dH_2O. The pH of the solution was then changed to pH7 by the addition of 10 M NaOH and the volume was then adjusted to 30 ml by the addition of dH_2O.

2.2.4 Southern denaturing solution

To produce 1 l of Southern denaturing solution, 87.6 g NaCl and 20 g NaOH were added to 800 ml dH_2O and the volume was then brought up to 1 l using dH_2O.

2.2.5 Southern neutralisation solution

To produce 1 l of Southern denaturing solution, 175.3 g NaCl and 60.6 g Tris were added to 800 ml dH_2O. The pH of the solution was then altered to pH 6.5 by the addition of concentrated HCl and the volume was then brought up to 1 l using dH_2O.

2.2.6 20 X SSC

To produce 1 l of saline sodium citrate, 174.5 g NaCl and 88.2 g Na_3ClT were added to 800 ml dH_2O and the volume was then brought up to 1 l using dH_2O.

2.2.7 1 M Sodium phosphate buffer

To produce approximately 450 ml of 1 M sodium phosphate buffer NaH_2PO_4 was added to 300 ml of Na_2HPO_4 until the pH was 7.2 (approximately 150 ml).

2.2.8 Hybridisation buffer

To produce 50 ml of hybridisation buffer 12.5 ml of 1 M sodium phosphate buffer and 17.5 ml 20 % SDS was added to 20 ml dH_2O and heated to 65 °C.

2.2.9 2 X SSC 0.1 % SDS

200 ml of 2 X SSC 0.1 % SDS solution was produced by adding 20 ml of 20 X SSC and 1 ml of 20 % SDS to 179 ml of dH_2O.

2.2.10 0.1 X SSC 0.1% SDS

200 ml of 0.1 X SSC 0.1 % SDS solution was produced by adding 1 ml of 20 X SSC and 1 ml of 20 % SDS to 198 ml of dH_2O.

2.2.11 Northern blot crosslinking solution

To produce 24 ml of northern blot crosslinking solution, 245 µl of 12.5 M 1-methylimidazole was added to 9 ml of dH2O and the pH adjusted to pH 8 using concentrated HCl. 0.753 g of EDC was then added to the solution and the volume increased to 24 ml using dH_2O.

2.2.12 Plasmolysis solution

To produce 50 ml of plasmolysis solution, 2.5 ml of 100 mM MES pH 5.7, 50 µl of 1 M $CaCl_2$ and 9.95 ml dH_2O were added to 37.5 ml of 1 M mannitol. The solution was then filter sterilised using a Corning 0.2 µm syringe filter (Corning) and a 20 ml syringe (BD).

2.2.13 Protoplast enzyme solution

15 ml of protoplast enzyme solution was produced by adding 0.225 g Cellulase RS (Yakult), 0.114 g Macerozyme (Yakult) and 2. 98 ml dH_2O to 9 ml of 1M mannitol. The solution was then incubated for 10 minutes at 55 °C to inactive proteases and then cooled to room temperature. 15 µl of $CaCl_2$, 1.5 ml of 10mg/ml BSA and 5.85 µl of β-mercaptoethanol were then added to the solution before it was then filter sterilised using a Corning 0.2 µm syringe filter (Corning) and a 20 ml syringe (BD).

2.2.14 Protoplast re-suspension solution

50 ml of protoplast re-suspension solution was produced by adding 2 ml of 100 mM MES pH 5.7, 2 ml of 100 mM KCl, 150 µl of 1 M $CaCl_2$ and 15.85 ml of dH_2O to 30 ml of 1 M mannitol The solution was then filter sterilised using a Corning 0.2 µm syringe filter (Corning) and a 20 ml syringe (BD).

2.3 *Arabidopsis thaliana* lines

Below are listed all plant lines used during this study along with the ecotype, transgenes, mutated loci and associated reference for each line. The ecotypes are C24, Landsberg *erecta* (Ler) and Columbia (Col).

Line	Ecotype	Transgene	Mutation	Reference
C24 WT	C24	None	Wild-type (WT)	
Ler WT	Ler		WT	
142	C24	*35S:GFP* construct	WT	(Dalmay et al. 2000a)
142S	C24	*NOS: 35S IR* silencer construct and a *35S:GFp* reporter construct	WT	(Eamens et al. 2008)
M1 (*morc6-5*)	C24	Same as 142S parent	EMS in At1G19100	
M2-M8	C24	Same as 142S parent	EMS in unknown locus	
M9 (*morc6-6*)	C24	Same as 142S parent	EMS in At1G19100	
M10-M24	C24	Same as 142S parent	EMS in unknown locus	
rmd1 (*nrpe1*)		Same as 142S parent	EMS in At2G40030	(Eamens et al. 2008)
rmd3 (*nrpd1*)	C24	Same as 142S parent	EMS in At1G63020	(Eamens et al. 2008)
SALK_059661 (*rdr2-1*)	C24	T-DNA insert from pROK2	T-DNA in At1G14790	(Alonso et al. 2003)
SALK_128428 *(nrpd1-2)*	Col	T-DNA insert from pROK2	T-DNA in At1G63020	(Alonso et al. 2003)
GABI-Kat 599B06 (*morc6-3*)	Col	T-DNA insert from pAC161	T-DNA	(Kleinboelting et al. 2012)
T5	Col	*35S:NOS* silencer construct and a *NOS: NPTII* reporter construct	WT	(Aufsatz et al. 2002a)

Line	Ecotype	Transgene	Mutation	Reference
hda6-1	Col	Same as T5 parent and also *MAS: PAT-OT* used as a T-DNA insertion	T-DNA in *hda6*	(Aufsatz et al. 2002b)
SAIL_610_G01 (mom1-2)	Col	T-DNA insert from either pCSA110 or pDAP101	T-DNA in *mom1*	(McElver et al. 2001)

Table 2.2: Plant lines

2.4 Propagation and manipulation of plants

2.4.1 Plant growth media

Murashige and Skoog (MS) medium was used to grow plants on plates. MS media was made by dissolving MS powder (Phytotechnology laboratories) in the appropriate volume of dH_2O and then adding 1 g of agar per 100 ml of media. This was then autoclaved before being poured into plates. If the media was to be used for the selection of lines containing the transgenes used in this study then kanamycin and DL-phosphinothricin were added to the media, at 50 µg/ml and 10 µg/ml respectively, before the media was poured into plates.

2.4.2 Surface sterilisation of seeds

Arabidopsis seeds were sterilised by chlorine gas. Seeds in open microcentrifuge tubes, or petri dishes if sterilising more than 400 µl of seed, were placed inside a sealable box. 5 ml of concentrated HCl was added to 100 ml of Haychlor bleach (Brenntag) in a beaker and this was placed inside the box, which was then sealed. The seeds were left to sterilise for 3 hours. Once completed seeds could then either be imbibed or stored.

2.4.3 Planting seeds

Plants grown on soil were either individually planted using a tooth pick or were loosely scattered over the soil, depending on what the plants would be used for. For plants grown on plates the seeds were first imbibed for 4 hours at room temperature or overnight at 4 °C in 0.1% agar. The seeds were then spread onto the plates by first suspending them in the agar through vortexing and then pipetted onto the plates.

2.4.4 Growth conditions

Both plants grown on soil and plates were first stratified for between three to seven days in a cold room before being placed in growth rooms or a greenhouse. For those placed in growth rooms they were placed in either a short day or long day growth room. Two long day growth rooms were used that operated a 16 hour day from 5 am to 9 pm. One of the growth rooms (L247) had a relative humidity of between 36-66% and day temperature for between 18.90-21.90 °C with the night temperature being between 19.20-20.07 °C. Light intensity was between 65-132 µmol photons m^{-2} s^{-1}. For the other growth room (L248) the relative humidity was between 40-70% and day temperature between 19.10-21.50 °C with the night temperature being between 18.30-20.9 °C. Light intensity was between 50-101 µmol photons m^{-2} s^{-1}. The short day growth room operated on an 8 hour day from 8 am to 4 pm and had a day temperature between 19.30-21.80 °C and a night temperature between 18.70-21.30 °C. The relative humidity was between 33-64% and the light intensity was between 52-82 µmol photons m^{-2} s^{-1}. For plants grown on plates the plates were placed horizontally in growth rooms, except for the root silencing experiment where the plates were near vertical in racks.

2.4.5 Crossing plant lines

Arabidopsis lines were crossed by using tweezers to remove the petals, sepals and stamen of flowers from the plant that would be the female parent. The flower was then pollinated by gently rubbing the anthers of a mature flower from the male parental line against the stigma. The pollinated flowers were then left to elongate and form mature siliques, from which the seeds were collected when the siliques turned yellow.

2.5 Plasmids

The only plasmid used in this study was the pGEM T-easy (Promega) vector that contains an ampicillin resistance gene and a *β-galactosidase* gene (Promega 2012). The cloning site for this vector is within the *β-galactosidase* gene so transformants that contain an insert can be selected for by blue/white screening.

2.6 Bacteria

2.6.1 Bacterial media and growth conditions

Luria broth (LB) was used for both bacterial liquid and solid media. 1 l LB medium was produced by adding: 10 g of tryptone, 5 g yeast extract and 10 g NaCl to 800 ml dH_2O. The pH was then adjusted to 7.5 using 10 M NaOH and dH_2O was added to get a final volume of 1 L. For solid medium 1 g of bacto-agar was added per 100 ml of media. The medium was then autoclaved.

If the media was to be used to select for bacteria containing the pGEM plasmid the antibiotic carbenicillin was added to the media to produce a concentration of 100 µg/ml. For liquid bacterial cultures a pipette tip was used to take a sample of a bacterial colony or glycerol stock and this was added to 5 ml of liquid LB media. These cultures were then incubated overnight at 37 °C in a shaker set at 250 rpm. When making LB plates the solid media was first melted in a 100 °C water bath and carbenicillin was added to the media before 25 ml of the media was added to each plate in a flow hood. If the plates were to be used for blue/white screening, 50 µl of 50 mg/ml X-Gal solution (Promega) was spread over each plate. When growing bacteria on plates the plates were placed in a 37 °C incubator overnight. All bacterial work was conducted under sterile conditions.

2.6.2 Bacterial strains

All cloning was carried out using either the Library Efficient DH5α (Invitrogen) or One Shot MAX Efficiency DH5α (Invitrogen) competent *E. coli* strains. The Library Efficient cells were used for the cloning of probes while the One Shot cells were used for bisulfite sequencing, bar the 21 dpg sequencing which used the Library Efficient cells.

2.6.3 Transformation

For bacterial transformation 50 µl of competent *E. coli* cells were mixed with 6 µl of plasmid in a microcentrifuge tube and incubated on ice for 30 minutes. The cells were then heat shocked for 45 s in a 42 °C water bath before being placed back on ice for 10 minutes. 500 µl of SOB media (Invitrogen) was added and the culture was then incubated for 30 minutes in a 37 °C incubator. The cultures were then spread onto plates containing a selective antibiotic, with 100 µl of culture added to each plate.

2.7 DNA purification

2.7.1 Plasmid extraction

The isolation of plasmids from *E. coli* was carried out using a QIAprep spin miniprep kit (QIAGEN) and was carried out according to manufacturer's instructions.

2.7.2 CTAB genomic DNA extraction

DNA was extracted from plant protoplasts using the CTAB extraction method. The protoplast samples, in 15 ml centrifuge tubes, were initially snap frozen in liquid nitrogen and then 1 ml of 1 X CTAB extraction buffer was added to each tube. The tubes were then incubated at 37 °C for 3 hours in a water bath. 1 ml of

phenol: chloroform: isomyl alcohol 25: 24: 1 (Sigma-Aldrich) was added and the solution was mixed by inversion before being centrifuged at room temperature and 1000 rcf for 30 minutes. The upper aqueous layer was transferred into a new 15 ml centrifuge tube and 1 ml of isopropanol was then added. The tubes were mixed by inversion before being centrifuged at room temperature and 1000 rcf for 30 minutes. The supernatant was removed and 1 ml 70% ethanol was then added. The tubes were centrifuged at room temperature and 1000 rcf for 20 minutes. The ethanol was removed and the pellet allowed to air dried for approximately 10 minutes. The pellet was then resuspended in 50 µl of dH_2O. The DNA samples were then treated with 1 µl of 10 mg/ml RNase A (QIAGEN) for half an hour to remove RNA in the sample.

2.7.3 Genomic DNA extraction using kits

The QIAGEN DNeasy 96 plant kit (QIAGEN) was used to extract DNA for CAPS and SSLP mapping and for all other DNA samples the DNeasy plant maxi kit (QIAGEN) was used. Both kits were used according to manufacturer's instructions.

2.7.4 Precipitation of DNA by sodium acetate

For precipitation of DNA using NaAc, 1/ 10 of the volume of 3 M NaAc pH 5.2 and 2 volumes of 100 % ethanol were added to a DNA solution in a microcentrifuge tube and left to precipitate overnight. The DNA was then pelleted by centrifugation at room temperature and 16100 x g for 20 minutes and the supernatant removed. 1 ml of 70% ethanol was then added and the tube was centrifuged at room temperature and 16100 x g for 10 minutes. The ethanol was then removed and the pellet allowed to air dry for 10 minutes before being re-suspended in the required volume of dH_2O.

2.7.5 Purification of DNA by phenol chloroform

For purification of DNA using phenol chloroform, initially 500 µl of phenol: chloroform: isomyl alcohol 25: 24: 1 (Sigma-Aldrich) were added to the DNA solution in a microcentrifuge tube and were then mixed by inversion before being centrifuged at room temperature and 16100 x g for 10 minutes. The upper aqueous layer was transferred into a new microcentrifuge tube and 500 µl of chloroform: isomyl alcohol 24: 1 (Sigma-Aldrich) was added and the solution was then mixed by inversion before being centrifuged at room temperature and 16100 x g for 10 minutes. The upper aqueous layer was transferred into a new microcentrifuge tube and 500 µl of isopropanol was then added. The tubes were mixed by inversion before being centrifuged at room temperature and 16100 x g for 20 minutes. The supernatant was removed and 1 ml 70% ethanol was then added. The tubes were centrifuged at room temperature and 16100 x g for 20 minutes. The ethanol was removed and the pellet was air dried for approximately 10 minutes before being re-suspended in the required volume of dH_2O.

2.7.6 Extraction of DNA from a electrophoresis gel

DNA bands were extracted and purified from gels using a QIAQuick gel extraction kit (QIAGEN) according to the manufacturer's instructions.

2.8 RNA purification

2.8.1 RNA extraction

Approximately one gram of plant leaf tissue was snap frozen in liquid nitrogen and ground into a power using a pestle and mortar before being transferred into a 15 ml centrifuge tube. 5 ml of TRIzol (Invitrogen) was then added to the tubes and the tubes were inverted before being incubated at room temperature for 10 minutes. 2 ml of chloroform: isomyl alcohol 24: 1 (Sigma-Aldrich) was added to

each tube and was mixed by inversion before being incubated at room temperature for 10 minutes. The tubes were then centrifuged at 4 °C and 1000 rcf for 30 minutes and the upper aqueous layer was transferred to a new 15 ml centrifuge tube. 5 ml of isopropanol was then added to each tube and then mixed by inversion, before being incubated for 10 minutes at room temperature. The tubes were then centrifuged at 4 °C and 1000 rcf for 30 minutes. The supernatant was then removed and the pellet washed with 2 ml of 70% ethanol before being centrifuged at 4°C and 1000 rcf for 20 minutes. The ethanol was then removed and the pellet air dried for 10 minutes. The RNA pellet was then resuspend in 200 µl of dH_2O.

2.8.2 Enrichment for small RNAs

RNA samples were enriched for small RNAs by PEG precipitation described previously by Lu (Lu et al. 2007). 50 % PEG and 5 M NaCl were added to the RNA sample to produce a final concentration of 5 % and 0.5 M respectively and was incubated on ice for 30 minutes. The samples were then centrifuged at 4 °C and 16100 x g for 10 minutes and the supernatant transferred to a new microcentrifuge tube. 2.5 volumes of 100 % ethanol was added to the sample and this was incubated at -20 °C overnight. The samples were then centrifuged at 4 °C and 16100 x g for 30 minutes and the supernatant removed. The pellets were then washed with 80 % ethanol and then allowed to air dry before being re-suspended in either dH_20 or 50 % formamide, if the RNA was to be used for northern blotting.

2.9 Quantification of nucleic acids

2.9.1 Non-denaturing agarose gel electrophoresis

Gel electrophoresis using agarose gels was used to visually assess the quantity and purity of DNA and RNA samples as well as for gel extraction purification of DNA. A 100 ml agarose gel was produced by the addition of 1g agarose and 10 µl

of 10 mg/ml ethidium bromide to 100 ml 0.5 X TBE buffer and these volumes could be scaled up of down depending on the size of the gel required. For gels that required separation of nucleic acids of similar sizes a higher percentage of agarose was used to a maximum of 3%. If the gel was to be used for extraction purification 10 µl of 10000 X SYBR safe (Invitrogen) was used instead of ethidium bromide. Before being run on the gel DNA samples were mixed with 6 X loading dye in a 5 DNA: 1 loading dye ratio. The 1 Kb plus DNA ladder (Invitrogen) was used as a molecular size marker. The gels were run in 0.5 X TBE buffer at between 50 V-150 V until a sufficient level of separation was achieved. The gels were then imaged on a UV light box for ethidium bromide or a blue light box (Invitrogen) for SYBR safe using the GENESnap camera and software (Syngene).

2.9.2 Denaturing acrylamide gel electrophoresis

Gel electrophoresis using a denaturing acrylamide gel was used for northern blotting to detect small RNAs. A single 24 % polyacrylamide gel was produced by adding 25.2 g urea, 6 ml 10 X MOPS and 1 ml dH_2O to 34 ml of 30 % acrylamide (National Diagnostics) and then heating this at 65 °C until it had melted. The gel mould apparatus was set up and the gel was left to cool to near room temperature. 30 µl of TEMED was then added and the gel was then mixed by inversion. 120 µl of 25 % APS was then added and again the gel was mixed by inversion. The gel was then poured and allowed to polymerize for 1 hour. Once polymerized the gel was placed in a gel tank filled with 1 X MOPS solution and the gel's wells were washed with 1 X MOPS using a 5 ml syringe (BD) and 25 G 0.5 x 1.6 mm needle (BD). The gel was then pre-run at 150 V for 1 hour and the RNA samples to be run on the gel were prepared by adding 2 µl of 6 X loading dye to 24 µl of RNA in 50 % formamide and heating to 80 °C for 3 minutes before being stored on ice. The wells were then washed again with 1 X MOPS and the RNA samples were loaded. For the first 30 minutes the gel was run at 250 V and then at 70 V overnight. Once the gel had been run it was removed from the gel apparatus and stained in 500 ml 1 X MOPS and 30 µl ethidium bromide for 10

minutes, before being imaged on a UV light box using the GENESnap camera and software (Syngene).

2.9.3 Spectrophotometer analysis

Nucleic acid quantification and purity was also assessed using a Nanodrop ND8000 8 sample spectrophotometer (Thermo scientific). The analysis was conducted as per the Nanodrop manufacturer's instructions and the concentrations and Absorbance (Ab) 260/230 and Ab 260/280 ratios, which are a measure of nucleic acid purity, were recorded.

2.10 PCR

2.10.1 Primers

Primers were produced by the Eurofins MWG and Sigm-Aldrich custom primer ordering services. The primers used in this study are listed in the tables below and the sequence, annealing temperature and product size is given for each primer. The annealing temperature was obtained by primer optimisation using the primer melting temperature (Tm) given by the primer manufacturer. The primers are arranged in terms of their use and this includes: mapping primers, sequencing primers, oligonucleotide probes and probe production primers.

Name	Primer Sequence	Target
AtREP2	GCGGGACGGGTTTGGCA GGACGTTACTTAAT	Oligonucloetide probe for the AtREP2 tranposon
miR167	TAGATCATGCTGGCAGCT TCA	Oligonucloetide probe for micrc RNA miR167 that is used as a loading control
$tRNA^{met}$	TCGAACTCTCGACCTCAG GAT	Oligonucloetide probe for RNA^{met} that is used as a loading control

Table 2.3: Oligonucleotide probes

Primers used as oligonucleotide probes for northern blotting.

Name	Primer Sequence	Annealling Temperature (°C)	Product size (bp)	Target
PVV45'	GTTTGAAAGTGTAGATGTAAC	56	1000	PVV4 Chromosome 1 CAPS marker
PVV4 3'	GGTTGTGTTTTGCTAGCATC	56	1000	PVV4 Chromosome 1 CAPS marker
21M1F	TTACTTTTTGCCTCTTGTCATTG	56	200	NF21M12 Chromosome 1 SSLP marker
21M1R	GGCTTTCTCGAAATCTGTCC	56	200	NF21M12 Chromosome 1 SSLP marker
M59F	GTGCATGATATTGATGTACGC	56	800	M59 Chromosome 1 CAPS marker
M59R	GAATGACATGAACACTTACACC	56	800	M59 Chromosome 1 CAPS marker
CAT3 5'	CGGTGGTGCTCCAGTCTCCAA	56	900	CAT3 Chromosome 1 CAPS marker
CAT3 3'	CAGATGCAATGGCATCGTGGA	56	900	CAT3 Chromosome 1 CAPS marker
GAPB 5'	GGCACTATGTTCAGTGCTG	55	1400	GAPB Chromosome 1 CAPS marker
GAPB 3'	TCTGATCAGTTGCAGCTATG	55	1400	GAPB Chromosome 1 CAPS marker
G4026 5'	GGGGTCACTTACATTACTAGC	51	800	G4026 Chromosome 1 CAPS marker
G4026 3'	GTACGGTTCTTCTTCCCTTA	51	800	G4026 Chromosome 1 CAPS marker
ADH 5'	GCGTGACCATCAAGACTAAT	61	1200	ADH Chromosome 1 CAPS marker
ADH 3'	AAAAATGGCAACACTTTGAC	61	1200	ADH Chromosome 1 CAPS marker
GPA1 5'	GGGATTTGATGAAGGAGAAC	61	1600	GPA1 Chromosome 2 CAPS marker
GPA1 3'	ATTCCTTGGTCTCCATCATC	61	1600	GPA1 Chromosome 2 CAPS marker
M249 5'	TGGTAACATGTTGGCTCTATAATTG	59	300	M429 Chromosome 2 CAPS marker
M249 3'	GGCAGTTATTAGAATGTCTGCATG	59	300	M429 Chromosome 2 CAPS marker
17D8LE 5'	CTCCTTTGTCATCTCCCGAATC	59	2000	17D8LE Chromosome 3 CAPS marker
17D8LE 3'	CCAACAACATGCATGATAGTTCA	59	2000	17D8LE Chromosome 3 CAPS marker
GL1 5'	ATATTGAGTACTGCCTTTAG	52	550	GL1 Chromosome 3 CAPS marker
GL1 3'	CCATGATCCGAAGAGACTAT	52	550	GL1 Chromosome 3 CAPS marker
F3H 5'	TACTCCTCCGTCACTTTCAC	55	900	F3H Chromosome 3 CAPS marker
F3H 3'	ATTTTCTCCACAGACCACAAG	55	900	F3H Chromosome 3 CAPS marker

Name	Primer Sequence	Annealling Temperature (°C)	Product size (bp)	Target
GA1 5'	CCGGAGAATCGTACGGTAC	59	1200	GA1 Chromosome 4 CAPS marker
GA1 3'	AAGCTTCGAACTCAAGGTTC	59	1200	GA1 Chromosome 4 CAPS marker
AG 5'	CAAACACCATTTAATCTTGACA	59	1300	AG Chromosome 4 CAPS marker
AG 3'	CAACAGGTTTCTTCTTCTTCTC	59	1300	AG Chromosome 4 CAPS marker
CAT2 5'	GACCAGTAAGAGATCCAGATACTGCG	54	900	CAT2 Chromosome 4 CAPS marker
CAT2 3'	CACAGTCATGCGACTCAAGACTTG	54	900	CAT2 Chromosome 4 CAPS marker
HAE1 5'	GAGTCGAGGGGGATGTATTCG	54	2000	HAE1 Chromosome 4 CAPS marker
HAE1 3'	GTGGTTTGCCATTCTTAACTTC	54	2000	HAE1 Chromosome 4 CAPS marker
NIT4 5'	CAACTCCACATCCGTCGGCG	65	3000	NIT4 Chromosome 5 CAPS marker
NIT4 3'	CGTTTCTTGTTGCATGGACATGAGAG	65	3000	NIT4 Chromosome 5 CAPS marker
DFR 5'	AGATCCTGAGGTGAGTTTTTC	55	1100	DFR Chromosome 5 CAPS marker
DFR 3'	TGTTACATGGCTTCATACCA	55	1100	DFR Chromosome 5 CAPS marker
10A10 5'	GCCTTATGGATTTTCTGGAGAAAG	61	2200	10A10 Chromosome 5 CAPS marker
10A10 3'	GGTCACAGGATCAAGATGTGG	61	2200	10A10 Chromosome 5 CAPS marker

Table 2.4: Mapping primers

Primers used for CAPS and SSLP mapping. The product sizes are estimated from an agarose gel

Name	Primer Sequence	Annealling Temperature (°C)	Product size (bp)	Target
191001F	AGCTTTGCATGAGCCTGATT	52	955	Sequencing the At1G19100 candidate gene
191001R	TTGCAGTAATGAACAAGCAGGT	52	955	Sequencing the At1G19100 candidate gene
191002F	TGCTCTGCATCTTCATCACC	52	984	Sequencing the At1G19100 candidate gene
191002R	AAAATGGGTTGCATGGAGAA	52	984	Sequencing the At1G19100 candidate gene
191003F	GCCCCTACATGTTGGTCAAT	52	1121	Sequencing the At1G19100 candidate gene

Name	Primer Sequence	Annealling Temperature (°C)	Product size (bp)	Target
191003R	GCCAGGTTCAGTGTAGAGAGG	52	1121	Sequencing the At1G19100 candidate gene
191004F	TCATTTGCTTTGCTCTTTTCA	51	833	Sequencing the At1G19100 candidate gene
191004R	TGGACCAATCCAGGTCAGA	51	833	Sequencing the At1G19100 candidate gene
191005F	TTTCCCCGAACTCTCTCTGA	52	1149	Sequencing the At1G19100 candidate gene
191005R	TTCCGAGACAGCCCATAAAC	52	1149	Sequencing the At1G19100 candidate gene
M13 For	GTAAAACGACGGCCAGT	55	300	For primer for pGEM sequencing
M13 Rev	GGAAACAGCTATGACCATG	55	300	Rev primer for pGEM sequencing
SR451F	TCGTTGCATCTAATGAGGAGA	52	1114	Sequencing the SR45 candidate gene
SR451R	TCTCCAATCCGTAGGCGTAG	52	1114	Sequencing the SR45 candidate gene
SR452F	GGGGTGAGATGGAACTTGAA	51	1153	Sequencing the SR45 candidate gene
SR452R	ATGTCGAAATTGCGATGGAT	51	1153	Sequencing the SR45 candidate gene
SR453F	GGAAAGAGACATACACCATCCA	52	1139	Sequencing the SR45 candidate gene
SR453R	CCTGCCACGTATCATCTTCG	52	1139	Sequencing the SR45 candidate gene

Table 2.5: Sequencing primers

Primers used for sequencing of candidate genes or inserts in the pGEM plasmid.

Name	Primer Sequence	Annealling Temperature (°C)	Product size (bp)	Target
35SsiRNAF	GTCTCAGAAGACCAAAGGGCTA	57	328	Amplifies part of the 35S promoter targeted by the *35S IR* transgene
35SsiRNAR	GGTCTTGCGAAGGATAGTGG	57	328	Amplifies part of the 35S promoter targeted by the *35S IR* transgene
5SF	GGATGCGATCATACCAG	45	628	Amplifies 5S rDNA probe

Name	Primer Sequence	Annealling Temperature (°C)	Product size (bp)	Target
5SR	GAGGGATGCAMCACSAG	45	628	Amplifies 5S rDNA probe
AtACTIN 5'	CATGGTTGGGATGAACCAGAAGGA	61	591	Amplifies the *Actin2* gene
AtACTIN 3'	GTCTCTTACAATTTCCCGCTCTGC	61	591	Amplifies the Actin2 gene
AtMU1F	GTGGATATACCAAAAACACAA	50	565	Amplifies the AtMu1 transposon
AtMU1R	CTTAGCCTTCTTTTCAATCTCA	50	565	Amplifies the AtMu1 transposon
AtSN1F	ACCAACGTGCTGTTGGCCCAGTGGT AAATC	65	170	Amplifies the AtSN1 retrotransposon
AtSN1R	TTTTTTTTTTTGAATATCTGGAAGTTC AAG	65	170	Amplifies the AtSN1 retrotransposon
BS14	CACACTTRTCTACTCCAAAAATATC	50	492	Amplifies the 35S promoter of the *35S:GFP* transgene for bisulfite sequencing
BS22	GAGGATAATGATAGGAGAAGTG	50	492	Amplifies the 35S promoter of the *35S:GFP* transgene for bisulfite sequencing
GFP1	CAAGGAGATATAACAATGAAG	50	813	Amplifies the *GFP* transgene
GFP8	CATGACGAACTCTAAGAGCTA	50	813	Amplifies the *GFP* transgene
MEA-ISRF	AAACCTTTCGTAAGCTACAGCCACTT TGTT	64	1091	Amplifies the MEA-ISR intergenic repeat element
MEA-ISRR	TCGGATTGGTTCTTCCTACCTCTTTAC CTT	64	1091	Amplifies the MEA-ISR intergenic repeat element
SoloLTRF	AATGCATTACAAAAACCTTCTGA	51	446	Amplifies the Solo LTR long terminal repeat
SoloLTRR	GGATTCACGATTAGAGAACGTAGA	51	446	Amplifies the Solo LTR long terminal repeat

Table 2.6: Probe primers

Primers used to produce probes for Southern and northern blotting

2.10.2 PCR machine

All PCRs were carried out using a DNA engine DYAD (MJ Research) PCR machine.

2.10.3 Taq PCR

Taq based PCR was used for amplification of sequences which would not be used for sequencing. PCR reactions that used the Taq thermo polymerase (New England Biolabs) to amplify DNA sequences were set up using the following volumes for a 20 µl reaction:

- X µl dH_2O (depending on DNA volume used)
- 2 µl 10 X Taq buffer
- 1 µl 10 µM Forward primer
- 1 µl 10 µM Reverse primer
- 0.4 µl 10mM dNTPs
- 0.2 µl Taq (1 Unit)
- X µl DNA

This reaction could be scaled up in order to produce more PCR product. The dNTPs consist of 10 mM ATP, 10 mM TTP, 10 mM CTP and 10 mM GTP (Fermentas). The basic PCR program used for Taq PCRs was:

1. 94°C for 2 minute
2. 94 °C for 30 s
3. X °C for 30 s
4. 72 °C for X minutes
5. Repeats steps 2 to 4 34 times
6. 72 °C for 10 minutes
7. 4 °C hold

The annealing temperature (3) and elongation time (4) varied between PCRs. The annealing temperature depends on the primers and the elongation time depends on the size of the insert; Taq amplifies 1 Kb of DNA a minute.

2.10.4 Colony PCR

Colony PCRs were used for screening bacterial colonies for plasmid inserts and used similar reaction volumes and conditions as normal Taq PCR. The differences were that instead of DNA a sample of a colony was added to the PCR reaction and that the initial denaturing step was extended to 4 minutes in order to lyse the bacteria to allow access to the DNA.

2.10.5 Pfu PCR

PfuUltra IIF Fusion (Strategene) was to amplify DNA sequences that would be sequenced. For a 20 µl reaction the following volumes were used:

- X µl dH_2O (depending on DNA volume used)
- 2 µl 10 X PfuUltra II Fusion buffer
- 1 µl 10 µM Forward primer
- 1 µl 10 µM Reverse primer
- 0.4 µl 10 mM dNTPs
- 0.2 µl Pfu
- X µl DNA

The reaction volumes could be scaled up in order to produce more PCR product. The reaction conditions were similar to that of a normal Taq PCR except that the amplifaction time is reduced as Taq amplifies 1 kb a minute whereas PfuUltra amplifies 4 Kb a minute

2.10.6 Bisulfite PCR

For bisulfite sequencing Epimark Hot Start Taq polymerase (New England Biolabs) was used as it is more effective than Taq at amplifying bisulfite treated DNA. The reaction volumes for a 20 µl PCR reaction are as follows:

- X µl dH_2O (depending on DNA volume used)
- 4 µl 5X Epimark buffer
- 1 µl 10 µM Forward primer

- 1 µl 10 µM Reverse primer
- 0.4 µl 10 mM dNTPs
- 0.2 µl Epimark Hot Start Taq
- X µl Bisulfite treated DNA

The reaction conditions were similar to that of a normal Taq PCR except that 40 amplification cycles were used and a hot start was performed. Hot start means that the PCR reactions were only placed in the PCR machine once it had reached the denaturing temperature (94 °C).

2.10.7 Reverse transcriptase PCR

For the detection of RNA levels reverse transcriptase (RT) PCR was used. Approximately 0.5 µg of RNA was used for each reaction and a no RT control was also carried out for each reaction. dH_2O was added to bring the volume of the RNA up to 10 µl and the RNA sample was then incubated at 75 °C for 10 minutes and 4 °C for 5 minutes. The following were then added to the reaction:

- 4 µl 5 x Superscript II RT buffer (invitrogen)
- 2 µl 10 mM dNTP (Fermentas)
- 1 µl Oligo dT (Invitrogen)
- 1 µl Superscript II Reverse Transcriptase(Invitrogen)

For the no Superscript II RT was added for the non RT control. The samples were then incubated for 1 hour at 37 °C and then the RT was inactivated by heating at 95 °C for 5 minutes. A normal Taq PCR using primers specific to the RNA transcript was then performed on the resulting cDNA to determine the relative levels of the transcript.

2.11 Manipulation of nucleic acids

2.11.1 Plasmid ligation

The pGEM T-easy (Promega) vector and ligation kit was used for all plasmid ligations and was used as per manufacturer's instruction.

2.11.2 Restriction enzyme digestion

DNA was digested using restriction enzymes (New England Biolabs and Promega) overnight at 37 °C. Below are listed the volumes required for a 20 µl restriction enzyme reaction but this can be scaled up if a larger reaction is required:

- X µl DNA
- 2 µl Manufacturer's Buffer
- 0.1 µl 10 mg/ml BSA
- 0.2 µl Restriction enzyme
- X µl dH_2O (depending on DNA volume used)

2.11.3 Bisulfite conversion of DNA

Bisulfite conversion of DNA was achieved using an EpiTect Bisulfite Kit (QIAGEN) that was used as per the manufacturer's instructions.

2.11.4 Global methylation analysis

Global DNA methylation was analysed using an Epigentek Methyl Flash kit (Epigentek) as per the manufacturer's instructions.

2.12 Blotting and detection of nucleic acids

2.12.1 Southern blotting

DNA samples that were to be analysed by Southern blotting were first digested using restriction enzymes and then run on an agarose gel. The DNA on the gel was first denatured in Southern denaturing solution for 20 minutes on a shaker set at 80 rpm. The denaturing solution was then replaced with Southern neutralisation solution and returned to the shaker for another 15 minutes. The neutralisation solution was replaced by fresh neutralisation solution and left on the shaker for a further 15 minutes. After this was completed the DNA was transferred overnight to a Hybond-NX (GE healthcare) or Hybond N^+ membrane (GE healthcare) by capillary action of 20X SSC solution. The DNA was then crosslinked to the membrane by UV radiation in a UV Stratalinker 1800 (Stratagene).

2.12.2 Northern blotting (mRNA)

RNA samples that were to be analysed for levels of mRNA were first mixed with glyoxal loading dye (Ambion) in a ratio of 5 µl of DNA to 1 µl of glyoxal. This RNA was then run on an agarose gel before being blotted to a Hybond N^+ membrane (GE healthcare) by capillary action of 20X SSC solution. The DNA was then crosslinked to the membrane by UV radiation in a UV Stratalinker 1800 (Stratagene).

2.12.3 Northern blotting (siRNA)

For the detection of small RNAs the RNA samples were run on polyacrylamide gel. The gel was then placed into the electroblotting apparatus with Hybond NX membrane (GE Healthcare) and placed in the blotting tank filled with 1 X MOPS. The gel was then electroblotted for two and a half hours at 5000 mA. The blotting apparatus was then disassembled and the RNA was chemically

crosslinked to the membrane by placing a piece of 3 mm blotting paper (Whatman) soaked in northern blot crosslinking solution underneath the membrane and wrapping this in cling film. This was then incubated at 60 °C for 2 hours and then the membrane was washed with dH_2O and allowed to dry.

2.12.4 Probes

Below are listed the probes used in this project along with whether they are an oligonucleotide or a DNA/RNA fragment.

Probe	Type of probe
5S rDNA	DNA fragment
AtMu1	DNA fragment
MEA-ISR	DNA fragment
Solo LTR	DNA fragment/RNA fragment
GFP	DNA fragment
35S	RNA fragment
At REP2	Oligo
miR167	Oligo
$tRNA^{met}$	Oligo

Table 2.7: Probes used in this study

2.12.5 Radio-labelling probes by random primer based amplification

For DNA probes a Prime-It II Random Primer Labelling kit (Agilent) was used to label the probe with the radioactive α P^{32}-CTP nucleotide (Perkin Elmer). The kit was used according to manufacturer's instructions. The α P^{32}-CTP had an activity of 370 MBq/ml.

2.12.6 Radio-labelling probes by end labelling

Oligonucleotide probes were end labelled with γ P^{32}-ATP (Perkin Elmer) using T4 polynucleotide kinase (New England Biolabs). The γ P^{32}-ATP had an activity of 370 MBq/ml. To probe a single blot a 16 µl reaction was set up consisting of:

- 10 µl dH_2O
- 2 µl 10 X T4 buffer
- 0.4 µl 10 µM oligonucleotide probe
- 5 µl γ P^{32} ATP
- 1 µl T4 polynucleotide kinase

The reaction is then incubated at 37 °C for 1 hour.

2.12.7 Radio-labelling probes by in vitro transcription

RNA probes were produced by in vitro transcription using the MAXIscripts kit (Ambion) and were used to detect siRNAs. The kit was used according to manufacturer's instructions and used the α P^{32} labelled UTP nucleotide (Perkin Elmer), which has an activity of 370 MBq/ml/

2.11.8 Hybridisation of blots with radio-labelled probe

Blots were placed in a Techne hybridisation bottle (Techne) with 20 ml of hybridisation buffer or Perfecthyb Plus Hybridisation Buffer (Sigma-Aldrich). The bottles were then placed in a Techne Hybridiser HB-1D hybridisation oven (Techne) set at 62 °C for Southern blots or 42 °C for northern blots for one hour to pre-hybridise. The P^{32}-labelled probe was then added to the bottle and the blot was left to hybridise overnight.

2.12.9 Washing blots to remove unbound probe

After hybridisation was complete the hybridisation buffer was then removed and replaced with 20 ml of either 2 X SSC 0.1 % SDS, for Southern blots, or 0.1 X SSC

0.1 % SDS, for northern blots, and returned to the hybridisation oven for a further 20 minutes. The solutions were then removed and replaced twice, except for Southern blots, the second wash was with 0.1 X SSC 0.1 % SDS. The radioactivity of the final wash was checked with a Geiger counter and if still high the blots were washed again.

2.12.10 Imaging blots using a phosphor screen

Blots that had been probed were wrapped in cling film and placed on a Kodak phosphor screen (GE Healthcare) and left to develop at room temperature for between 2 hours to 2 days, depending on the strength of the signal of the bound probe. The Quantity One (Bio rad) software and FX Molecular imager (Bio rad) machine were used to image the phosphor screen.

2.12.11 Imaging blots using a autoradiography film

Blots that had been probed were wrapped in cling film and placed with a Hyperfilm autoradiography MP 5 × 7" film (GE Healthcare) in an exposure cassette. The cassette was then placed in a -81 °C freezer to develop for 1 day to 1 month, depending on the strength of the signal of the bounds probe. The film was then developed using a Compact X4 xograph (Xograph Imaging Systems).

2.12.12 Stripping blots of radio-labelled probes

Blots were stripped of bound probe by incubation at 80 °C in 0.1 % SDS for 20 minutes in a hybridisation oven (Techne). The SDS solution was then removed and the blot checked with a Geiger counter and if there was still a detectable signal the blot was washed again in 0.1 % SDS. Once the probe had been removed the blot was placed on a phosphor screen (GE Healthcare) or in an exposure cassette with autoradiography film (GE Healthcare) and the screen or film was then developed to check that the probe had been removed totally.

2.13 Sequencing

2.13.1 Sanger sequencing

Sanger sequencing of plasmids and PCR products was carried out by the University of York Biology Department and Eurofins MWG sequencing services.

2.13.2 Next generation sequencing

Next generation sequencing was carried out by the Exeter University sequencing service.

2.14 Imaging

2.14.1 UV lamp

Images of plants under UV light were taken by Phil Roberts from University of York Biology Graphics department using a DSLR camera with a UV filter. The UV lamp used was a Blak Ray (UVP).

2.14.2 Fluorescence microscopy

Images of plants were taken under a Leica MZFLIII UV fluorescence microscope (Leica) using a CoolSNAP camera and imaging software (RS Photometrics). The GFP1 filter set up was used on this microscope, which consists of a: 395-455 nm excitation filter, 470 nm dichroic mirror and 480 nm barrier filter.

2.14.3 Confocal microscopy

Images of whole plant leaves and leaf cross-sections were taken using a 510 Zeiss confocal microscope. Whole leaves were mounted in 10 % glycerol between two 1.5 22 x 22 mm cover slips (SLS) using nail varnish (Boots) to allow for both sides of the leaf to be imaged and was then placed on 1.0-1.2 mm thick 76 x 26 mm

glass slide (Fisherbrand) for imaging. Cross-sections were mounted in 10 % glycerol between a 1 .0-1.2 mm thick 76 x 26 mm glass slide (Fisherbrand) and a 1.5 22 x 22 mm cover slip (SLS). All images were taken using a 20 X 0.5 objective and used an Argon laser emitting at 488 nm and a Helium-Neon laser that emits at 543 nm, which excite GFP and propidium iodide/chlorophyll respectively. The Argon laser was used at 16 % capacity while the Helium-Neon laser was used at 40 % capacity. For the whole leaf samples two channels were set up for GFP and propidium iodide using the following filter set: Main dichromatic beam splitter: HFT 488/453 nm, Dichromatic beam splitter 1: Mirror; and Dichromatic beam splitter 2: NFT 545 nm. A 505-530 nm band pass filter was also used for the propidium iodide channel. The only parameter altered between images was the master gain in order to account for differences in fluorescence levels between the samples. For the sectioning work a third channel was used, ChD, that detects the laser light that passed through the sample and so provides a bright field view of the object being imaged.

2.14.4 Propidium iodide staining

Leaves were stained with propidium iodide before being imaged under the confocal microscope. The leaves were placed in a microcentrifuge tube with 200 µl of 20 mg/ml propidium iodide (Sigma-Aldrich) and left on a roller for half an hour to two hours, depending on the size of the leaves. The leaves were then removed from the propidium iodide and washed in d H_2O for 10 minutes before being mounted in 10 % glycerol.

2.14.5 Leaf sectioning

Leaves were sectioned using a vibratome. The leaves were first set in blocks of 10 % porcine gelatine (Sigma-Aldrich) these were then fixed to the sectioning platform of a Leica VT1000S vibratome (Leica) so that the leaf tip was vertical and the abaxial side of the leaf was facing the vibratome's blade. The blade

vibration frequency was set at 90 Hz and the movement of the blade was set at 0.5 mm/s and the thickness of each section was set to 200 µm. The end point of each cut by the vibratome was set so that it did not go all the way through the gelatin block allowing for the sections to be kept in sequence. Once sectioning was completed the gelatin spine left by the vibratome could be remove using a scalpel and the desired section selected using tweezers and mounted on a slide for confocal microscopy.

2.15 Protoplasts

2.15.1 Production of mesophyll protoplasts

For the production of mesophyll protoplasts 30 adult rosette leaves between 1-2 cm in length were collected from 5 week old plants and placed in plasmolysis solution. A razor blade was used to cut 1 mm wide strips along the centrolateral leaf axis. The plasmolysis buffer was removed and the leaf strips were washed with fresh plasmolysis buffer. The buffer was again removed and replaced with protoplast enzyme solution and this was vacuum infiltrated into the leaf strips for 1 to 2 minutes and these were then incubated at 28 °C for two hours. The strips were then swirled gently to release the protoplasts and filtered through a 40 µm nylon mesh. The flow through was centrifuged at room temperature and 100 rcf for 5 minutes. The supernatant was then removed and the protoplasts re-suspended in protoplast re-suspension solution and kept at room temperature in the dark.

2.15.2 Cell sorting

Cell sorting was carried out by the University of York Imaging and Cytometry laboratory.

2.16 Bioinformatics

2.16.1 Sequence analysis

Wild-type DNA and protein sequences from the *Arabidopsis* Columbia ecotype were obtained from The *Arabidopsis* Information Resource (TAIR). For sequences that were produced by Sanger sequencing, the Sequence Scanner (ABI) and Lasergene EditSeq (DNAStar) programs were used to view and manipulate sequences. The NCBI BLAST (NCBI) and TAIR sequence viewer (TAIR) online tools were used to identify sequences and align them with known sequences. Clustal W (EMBL EBI) and Jalview were used to produce alignments of multiple sequences for presentation as figures and analysis of sequence homology (Waterhouse et al. 2009). For the analysis of bisulfite sequencing, the sequences were first aligned using Clustal W (EMBL EBI) and then cytosine methylation was determined using the CyMate online tool (Hetzl et al. 2007). Analysis of next generation sequencing was carried out by Zhesi He using the SHOREmap software (Schneeberger et al. 2009). The location in the *Arabidopsis* genome of single nucleotide polymorphisms (SNPs) identified from the SHOREmap analysis was found using the TAIR sequence viewer online tool (TAIR). TAIR was then used to determine whether the SNP was within a gene or in a intergenic region and whether the SNP resulted in a change in protein sequence.

2.16.2 Identification of protein orthologs

Orthologs of *Arabidopsis* proteins were identified using the *Arabidopsis* protein sequence in a BLAST (NCBI) search of protein sequences from specific Plantae, Metazoan and Bacteria species. Sequences with a 40 % homology to the *Arabidopsis* sequence were selected.

2.16.3 Domain identification and 3D domain structure

Protein domains were identified by searching for domains using the Pfam and SMART protein databases (Letunic et al. 2012, Punta et al. 2012). The 3D structure of domains were produced using the Phyre 2 3D structure prediction online tool (Kelley and Sternberg 2009).

2.16.3 Production of phylogenetic trees

Phylogenetic trees were produced from an alignment of protein sequences by Clustal W. The program used to produce the tree and bootstrap the tree was Clustal X2 and Dendroscope was used to view the tree and edit it for a figure (Huson et al. 2007, Larkin et al. 2007).

3. Identification of M1 and M9 as *morc6* mutants

3.1 Introduction

3.1.1 Dual transgene silencing system

The forward genetic screen carried out in this work used a dual transgene system to enable the identification of mutants in the RdDM pathway. The system consists of a *35S* promoter-driven *GFP* gene (*35Sp:GFP*) and a nopaline synthase (*NOS*) promoter driven inverted repeat of part of the *35S* promoter (*NOSp:35S IR*). The full sequence of both transgenes is given in Appendix 1.1 (page 346). The *35S* promoter provides near ubiquitous expression of *GFP* in the majority of plant tissue (Fang et al. 1989, Slater et al. 2003).The GFP protein in this case is targeted to the endoplasmic reticulum (ER) (Haseloff et al. 1997). The *NOS* promoter is expected to give constitutive expression of the *35S IR* transgene, although expression may be highest in roots (An et al. 1988, Aufsatz et al. 2004). The inverted repeats of the *35S* promoter are separated by an intron of the *Petunia Chalcone Synthase* A (*CHSa*) gene, which stabilises the inverted repeat during transcription (Kerschen et al. 2004, Eamens et al. 2008). The wild-type (WT) line that contains both transgenes is referred to as 142S throughout this work. 142S was produced from a WT transgenic plant line in the C24 ecotype. This line, GFP 142, contains the *35Sp:GFP* transgene (Dalmay et al. 2000a). GFP 142 is referred to as 142 hereafter. The *NOSp:35S IR* transgene was introduced into 142 to produce the 142S line (Eamens et al. 2008). Both transgenes have been shown to be single loci insertions through 3:1 Mendelian segregation.

In the dual transgene system the *35Sp:GFP* transgene acts as a reporter and the *NOSp:35S IR* acts as a silencer (Figure 3.1). The mRNA transcript of the *35S IR* transgene is expected to be produced by RNA polymerase II, although it is dependent upon PolIV for siRNA production (Figure 3.1 A) (Eamens et al. 2008). The transcript forms dsRNA via sequence complementarity (Figure 3.1 B). This acts as a trigger for the RdDM pathway, which produces 24nt siRNAs from this transcript. These siRNAs have homology to both the *35S IR* and the *35S* promoter

of the *GFP* transgene. These are then loaded into a AGO protein, likely to be AGO4, which then facilitate the methylation of the *35S* promoter of the *GFP* transgene (Figure 3.1 C and D). This represses transcription of *GFP,* thus resulting in a loss of GFP fluorescence in 142S, which contains both transgenes (Figure 3.1 E). In 142, which lacks the *35S IR* transgene, *GFP* is not silenced and therefore GFP fluorescence is detectable.

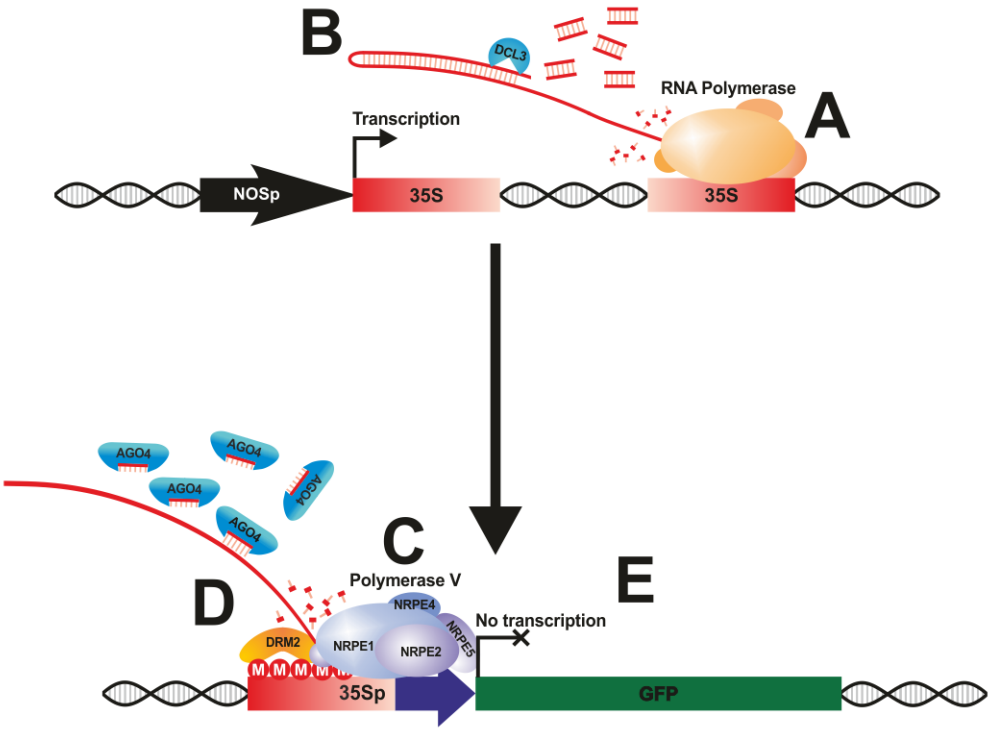


Figure 3.1: How the dual transgene silencing system functions

A: An RNA polymerase transcribes the *NOS: 35S IR* transgene. **B:** The 35S IR repeats form dsRNA through self-complementation and this is then cleaved by DCL3 into 24 nt siRNAs. **C:** PolV transcribes the corresponding sequence of the *35S* promoter of the *GFP* transgene and 35S siRNAs loaded into AGOs bind to the transcript. **D:** The binding of AGO loaded with siRNAs triggers DRM2 to methylate the *35S* promoter. **E:** The methylation prevents transcription of *GFP.* Promoters are shown as arrows with *NOS* in black and the *35S* in red and purple. The red section of the *35S* promoter is the region targeted by the *35S IR* transgene. Rectangles represent the CDS of each transgene with *GFP* being in green and the *35S IR* in red. There is an intron between the two inverted

repeats in the*35S IR* and this is represented as DNA double helix. The transcription activity of both transgenes is shown by the black line at the start of each transgene with an arrow meaning that is transcribed whereas a cross means that it is not.

An ethyl ester methane sulfonic acid (EMS) mutagenesis of the 142S line was completed by A. Eamens (Eamens et al. 2008). Progeny from this mutagenesis were then self-fertilized and M2 plants screened for defects in RdDM by visualising GFP fluorescence under UV-light. The M2 generation is the first generation which would contain homozygous EMS mutations. The initial study identified mutants in *nrpd1* and *nrpe1* (Eamens et al. 2008). Characterisation of these mutants revealed that as well as their established role in TGS both *NRPD1* and *NRPE1* are important for maintenance of PTGS. This was unexpected as both PolIV and PolV were thought to act exclusively in TGS pathways.

3.1.2 Aims of chapter

The aim of this chapter was to screen an EMS mutated *Arabidopsis* population for putative novel RdDM mutants and to identify the mutated gene. Bioinformatic tools would then be used to assess the gene's expression pattern, evolutionary history and protein structure. This was achieved by:

A) Identification of putative novel RdDM mutants. The library was screened for *GFP* fluorescence and putative mutants were characterised and sorted into phenotypic groups according to their patterns of DNA methylation and siRNA accumulation. Complementation assays were then used to identify alleles in the same gene and assess if the mutants are in a known RdDM gene.

B) Identify mutant locus in novel RdDM mutants. A rough map position for the mutant was obtained using cleaved amplified polymorphic sequences (CAPS) and simple sequence length polymorphisms (SSLP) mapping and then genome sequencing was done to find the candidate RdDM gene. This was confirmed by

sequencing other alleles of the RdDM mutant and complementation assays with known mutants in that gene.

C) Characterise the RdDM gene using bioinformatics. The gene expression pattern, phylogenetic tree and protein structure of the RdDM gene was assessed using bioinformatic tools.

3.2 Chapter-specific methods

3.2.1 Mutation identification strategy

The two most common strategies for the production of mutants for forward genetic screens in *Arabidopsis thaliana* are mutagenesis using a mutagen, usually EMS, or T-DNA insertion (Krysan et al. 1999, Greene et al. 2003). Whereas T-DNA insertion mutations can be easily identified by sequencing out from the T-DNA insertion into the genome, EMS mutations are harder to identify. Before the advent of next generation sequencing technology, EMS mutants were identified by mapping the mutant locus to a specific region of the genome and sequencing candidate genes within this region. However, with modern sequencing technology it is now possible to use re-sequencing of the entire genome and identify candidate single nucleotide polymorphisms (SNPs) from this sequence (Ossowski et al. 2008, Schneeberger et al. 2009, Schneeberger and Weigel 2011). Both strategies can also be used in conjunction with each other, with mapping first identifying the region of the genome where the mutation resides and this can be used to direct the choice of candidate SNPs in the re-sequencing. Initially a purely mapping based strategy was used but it was realised that it would be more cost effective to instead use re-sequencing to identify mutants and so for this reason a dual strategy that has been used for this study. Although unintentional this strategy should provide the added benefit of a dual strategy described previously.

3.2.2 Mapping of RdDM mutant gene locus using CAPS and SSLP markers

To identify the region of the genome where the mutant locus was located mapping was employed for this purpose. Mapping relies upon the principle that two features on a chromosome are more likely to segregate together if they are close to one another than if they are further apart (Morgan 1911a, Morgan 1911b). Features that segregate together are said to be linked and it was shown that the frequency of co-segregation did not only provide a measure of linkage between the two features, but that it also is a direct measure of the distance between the two features (Sturtevant 1913). This frequency of segregation and hence distance can be calculated for the EMS mutation and markers. As the location of the on the chromosome is known this can therefore provide an indication of the location of the EMS mutation. However, for this to work there must be at least two alleles of each marker and in *Arabidopsis* researchers have taken advantage of the fact that genetic variation between different ecotypes of *Arabidopsis* results in several loci that vary between ecotypes and therefore can be used as markers. Two of these marker types have been used in this study and they are cleaved amplified polymorphic sequences (CAPS) and simple sequence length polymorphisms (SSLP). CAPS markers are sections of genomic DNA that when amplified by PCR and cleaved with a restriction enzyme, produce products of specific sizes (Konieczny and Ausubel 1993). These products are ecotype specific so different ecotypes produce products of different sizes. SSLP markers are microsatellites that are of a specific length in each ecotype, but vary in length between ecotypes (Bell and Ecker 1994).

3.2.3 Producing pooled DNA samples for Illumina sequencing

In RdDM mutant lines produced using EMS mutagenesis, identifying a causal mutation is difficult due to the presence of a large number of other mutations

carried in that particular line. In order to reduce the number of EMS mutations backcrossing and sample pooling was employed. Initially the mutant line was backcrossed twice to the parental 142S line and plants that were homozygous for the RdDM mutation were selected in the following F2 generation using GFP fluorescence (Figure 3.2 A). In the F1 generation's germline recombination will occur between the RdDM mutant and 142S parental line chromosomes, producing chromosomes with sections from each line (Figure 3.2 B). Due to the selection for homozygote mutants in the RdDM locus in the F2 generation both the RdDM locus and surrounding sequence must come from the RdDM mutant. However, there is no selection on any other section of the genome, meaning that they can originate from either the mutant or 142S line. In the F2 generation each plant should have undergone a different set of recombination events so will also carry a different set of EMS mutations, except for the causal mutation that has been selected for and those that are closely linked. To take advantage of this variation between individuals the DNA sample from the RdDM mutant was produced from plant tissue collected from one hundred of the F2 GFP fluorescent plants from the 2nd backcross. Only the young rosette leaves, typically around 5 mm long, were collected. The reference sample of 142S did not require pooling so adult rosette leaves were collected from five 142S plants.

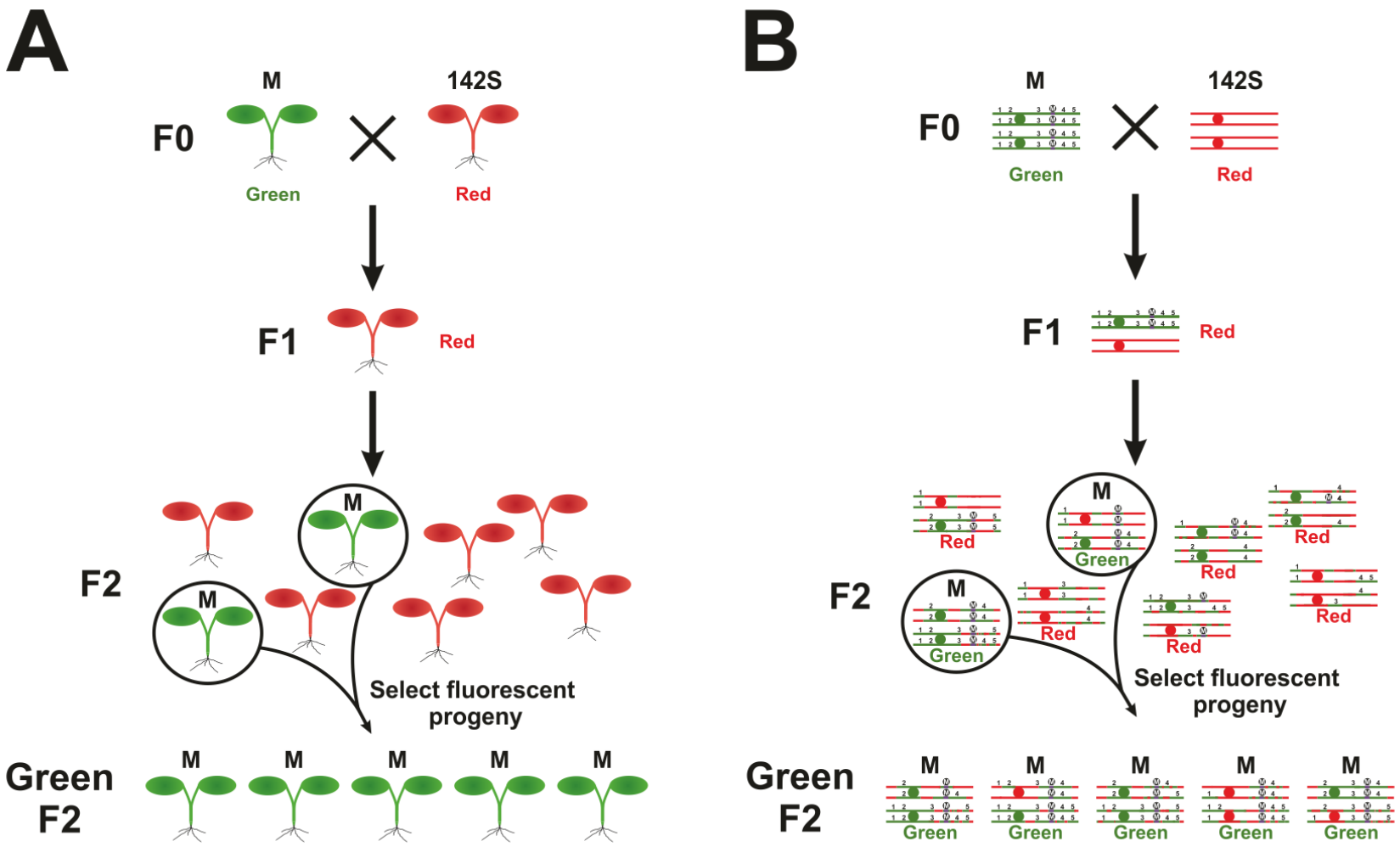

A
M
142S
F0
Green
Red
F1
Red
F2
M
M
Select fluorescent progeny
Green F2
M
M
M
M
M
B
M
142S
F0
Green
Red
F1
Red
F2
Red
M
Green
Red
Red
M
Green
Red
Red
Select fluorescent progeny
Green F2
M
M
M
M
M
Green
Green
Green
Green
Green

Figure 3.2: Removal by backcrossing of unlinked mutations segregating with the causal mutation

Diagrams showing how back crossing a mutant line (M) to 142S can reduce the number of non-causative mutations in the mutant line. **A:** Diagram showing how the back crossing and selection process works. Plants that are coloured green are GFP fluorescent and therefore a mutant (M) while plants that are red are not and are wild-type (142S). The generation number is shown to the left of the figure, with the M and 142S lines designated the F0 generation. **B:** Diagram showing the same process as **A** but shows the chromosomal pair, on which the causal M mutation resides, at each stage. The M chromosomes are coloured green while the 142S chromosomes are coloured red. The numbers (1-5) on the M chromosome show homozygous mutations carried by the M line that are not causative of the RdDM phenotype. The causative mutation is represented by a white M in a black circle.

3.3 Results

3.3.1 Identification of M1 and M9 as alleles in a potentially novel RdDM mutant

To identify mutants in the RdDM pathway the dual transgene system, as described in the introduction, was used. The 142S line carries both transgenes whereas line 142 carries only the *35S:GFP* transgene. An EMS mutagenesis of 142S was performed during a previous study to identify RdDM mutants and the M2 generation were screened for GFP fluorescence as these are likely to be a mutation in the RdDM pathway (Eamens et al. 2008). From this initial screen there were twenty four uncharacterised mutants that were GFP fluorescent, these were numbered M1 to M24.

The mutants were then separated into two phenotypic groups through assessment of 5S rDNA methylation, those being mutants with no change from WT in DNA methylation and those with reduced DNA methylation. The 5S rDNA repeats are an endogenous target of RdDM and are commonly used to characterise RdDM mutants (Aufsatz et al. 2002b, Vaillant et al. 2007, Ausin et al. 2009). DNA methylation was assessed using methylation sensitive Southern blotting whereby the DNA is digested with a methylation sensitive restriction enzyme that creates a banding pattern, which is altered by presence or absence of DNA methylation. Cytosine methylation was assessed in the CG, CHG and CHH sequence contexts, where H represents A, T or C, using the restriction enzymes: *Hpa*II which is sensitive to CG methylation and *Hae*III which is sensitive to CHG and CHH methylation. When DNA methylation is present the restriction enzymes will not cleave the DNA and higher molecular weight (HMW) bands are formed, which is seen in the WT lines 142 and 142 (Figure 3.3). There were four lines with a banding pattern similar to WT, indicating no loss in methylation, which comprised M1, M6, M9 and M19. The remaining twenty lines all showed an increase in lower molecular weight (LMW) bands compared to WT indicating a

loss in DNA methylation in these lines. In the initial blots differences in the amount of DNA loaded for 142S (Figure 3.3 A) and 142 (Figure 3.3 C) meant that the banding pattern for these samples was similar to that of the lines that lost DNA methylation. For this reason the blots were repeated for the group of mutants that retained methylation and are shown in Figure 3.3 B and D; and in chapter five (page 298). These repeats again show that the mutants retain DNA methylation. This study concentrated on the group of mutants that did not lose DNA methylation due to the fact that at the time the only known RdDM mutant that did not show a change is MOM1 and therefore would suggest that these mutants are also involved downstream of DNA methylation either in histone modification or higher order heterochromatin modification (Amedeo et al. 2000, Habu et al. 2006, Vaillant et al. 2006, Numa et al. 2010). This is one of the areas of the RdDM pathway where there are gaps in our knowledge and so there was the potential for the involvement of novel components. Re-assessment of GFP fluorescence revealed that M19 had been wrongly identified as fluorescent, indicating that it was not an RdDM mutant and so was discarded. M6 was shown to be a mutant in the *35S IR* transgene, therefore only M1 and M9 have been studied in detail. Both M1 and M9 are GFP fluorescent, however not to the same extent as 142, indicating that silencing must still occur in the mutants but at reduced levels (Figure 3.4).

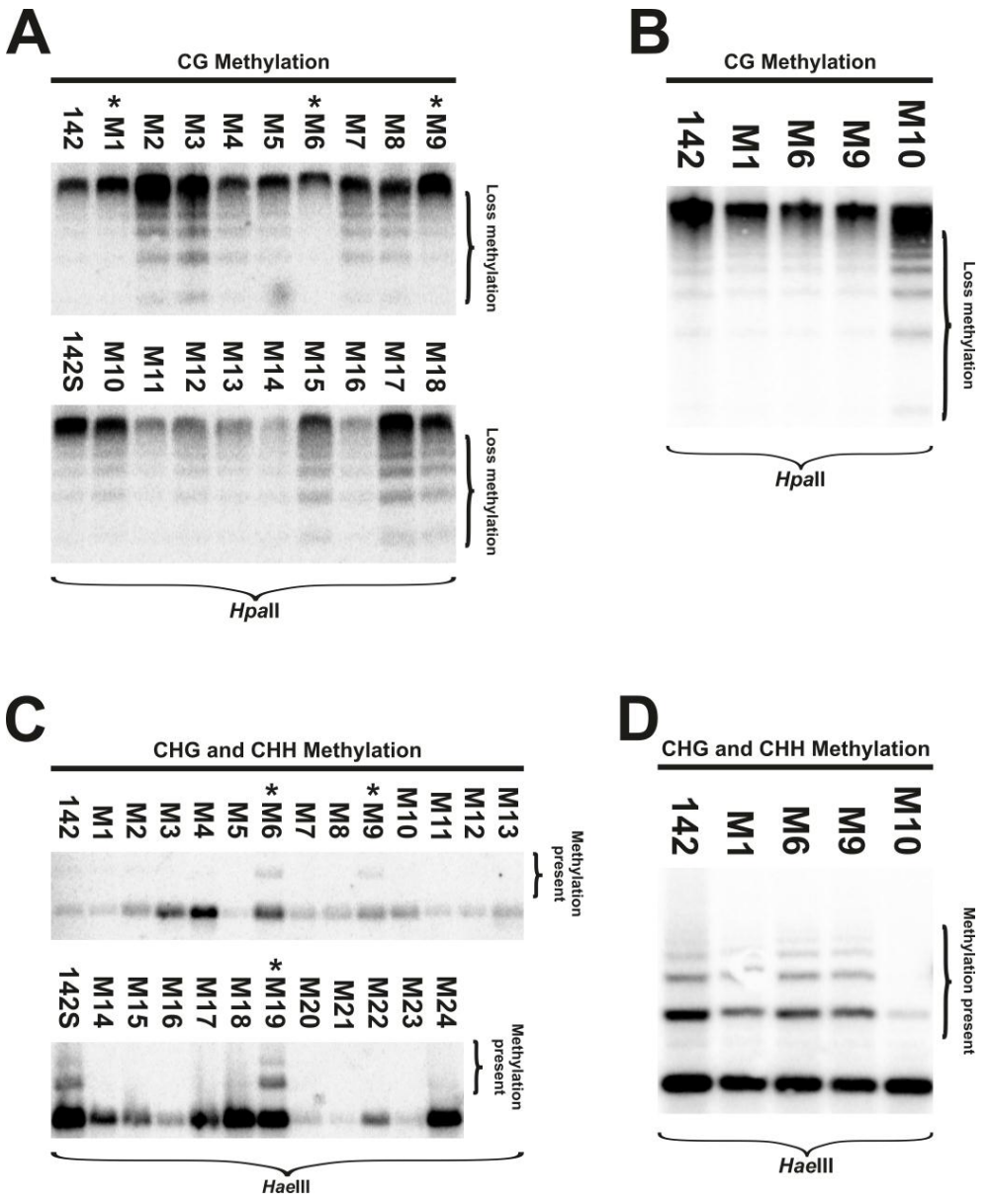


Figure 3.3: Screening EMS mutants for loss of 5S rDNA methylation

Images of Southern blots used to assess the methylation status of the 5S rDNA repeats in all twenty four EMS mutant lines (**A** and **C**), with mutants that maintain methylation highlighted by an asterisk, and three mutants (M1, M6 and M9) that show no decrease in DNA methylation (**B** and **D**). Mutants were tested for CG methylation (**A** and **B**) and CHG and CHH methylation (**C** and **D**). For **B** and **D** line M10 was used to show the banding pattern of an RdDM mutant that loses DNA methylation, although the mutant gene in M10 is unknown. The restriction enzyme used to digest the DNA is shown at the bottom of each blot while the cytosine configurations that the restriction enzymes are

sensitive to are shown above each blot. The order of samples is shown above each lane, with 142 and 142S being the WT samples.

Figure 3.4: Differing levels of GFP fluorescence between 142, 142S, M1 and M9

Images of lines 142 (**A**), 142S (**B**), M1 (**C**) and M9 (**D**) taken under UV light at 32 dpg. Images are representative for each lines. The autofluorescence on the top-middle leaf of 142S is due to necrosis and has no relevance to the silencing phenotypes. The area of fluroscence on the right hand leaf of M1 is soil on top of the leaf that fluoresces.

Allelism testing showed that M1 and M9 were allelic as they failed to complement each other and therefore were alleles in the same RdDM mutant locus (Figure 3.5). The phenotype of M1/M9, discussed in detail in the next two chapters, was similar to another mutant named rmd6, identified in the previous screen of the EMS mutant library by A. Eamens, but whose causative mutation had not been identified. Allelism testing revealed that rmd6 was also an allele in the same gene as M1 and M9 (Figure 3.5 G and H). All characterisation in this

study, apart from sequencing, has only been carried out on M1 and M9 but not rmd6. The reason for this is that rmd6 has a stunted growth phenotype that is most likely unrelated to the RdDM mutation but has not yet been removed by back crossing and may interfere with phenotypic analysis. Allelism tests were then performed between M1 and M9 and a previously identified *mom1* mutant to determine if they were mutant alleles of *MOM1* (Amedeo et al. 2000). The mutants were also tested against a *hda6* mutant as at the point when these tests were being carried out MOM1 and HDA6 were the only two proteins known to act downstream of DNA methylation (Amedeo et al. 2000, Aufsatz et al. 2002b). However, *hda6* mutants cause a reduction in DNA methylation so unless M1 and M9 were partial loss of function mutants in *HDA6* they were unlikely to be mutants in this gene (Aufsatz et al. 2002b). The allelism tests between M1 and M9; and the known RdDM mutants showed complementation, thus indicating that that the mutant locus in M1 and M9 is neither *hda6* or *mom1* (Appendix 1.2 page 352). Since the allelism testing showed no complementation this suggested that M1 and M9 are alleles in a novel RdDM mutant, however the M1/M9 locus needed to be identified in order to determine whether this is the case or not. To achieve this, mapping and genome re-sequencing were employed.

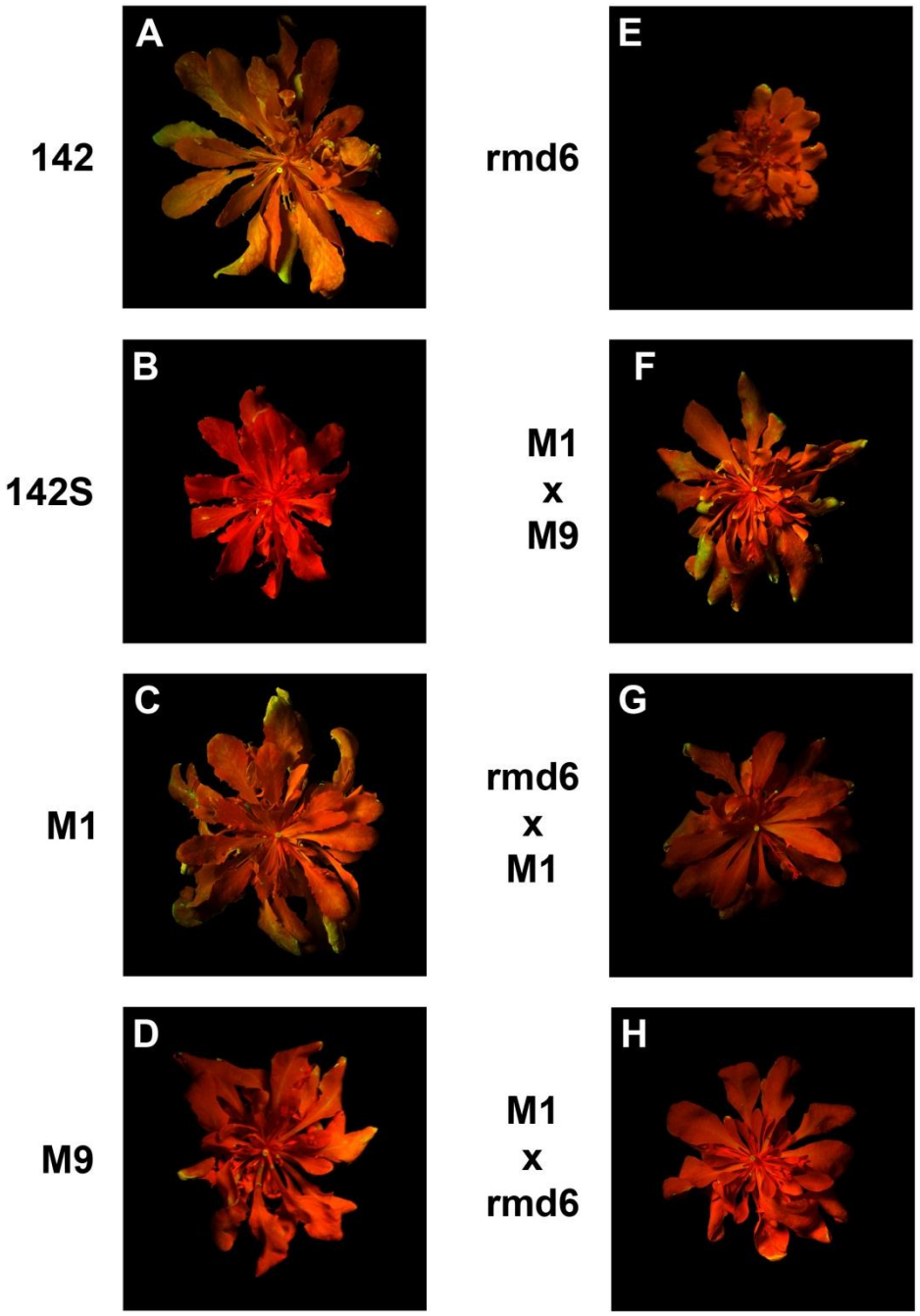


Figure 3.5: Allelism tests identifying M1, M9 and rmd6 as allelic

Images taken under UV light of parents and crosses at 46 dpg. The parental lines shown are 142, 142S, M1, M9 and rmd6. For all crosses the names of the parents are given, the first name is the male and the second is the female. The bright fluorescence seen on the older leaves of lines shown are due to necrosis and are not linked to the silencing phenotypes. Images are representative for each line.

3.3.2 Obtaining a rough map position for the M1/M9 locus

In order to identify the location of the M1/M9 locus within the genome, mapping with SSLP and CAPS markers was used. The distance between these markers and the M1/M9 locus can be determined using linkage analysis and this principle is explained further in the methods section of this chapter (page 128). However, for the markers to work they require the M1 and M9 mutants to be in a mixed ecotype background so that there are two segregating ecotype specific alleles for each marker. Both M1 and M9 mutants are in the C24 ecotype background so were crossed with WT Landsberg *erecta* (Ler) and the M1/M9 mutants identified in the F2 generation by GFP fluorescence. As Ler lacks the two transgenes not all F2 plants will have both transgenes, therefore in order to correctly identify M1 and M9 the two transgenes must also be selected for. The *GFP* transgene insertion includes a kanamycin resistance gene and the *35S IR* transgene insertion includes a DL-phosphinothricin resistance gene, therefore to select for only the plants with both transgenes the F2 generation was germinated on MS plates containing kanamycin and DL-phosphinothricin. The plants that were resistant to both compounds were then screened for GFP fluorescence and those that were fluorescent were transferred to soil as these will be the M1/M9 mutants. A number of plants that were not fluorescent were also transferred to soil as these will act as controls for the mutant plants.

The F2 progeny will have undergone crossing over events between C24 and Ler chromosomes resulting in heterogeneous chromosomes with sections from both ecotypes, however as the M1 and M9 mutation have been selected for, the origin of markers close to the M1/M9 locus should be mostly from C24 rather than Ler. The proportion of markers that only come from C24 should decrease as the distance from the M1/M9 locus increases. One problem with this technique is that as well as the M1/M9 locus the two transgenes will also be selected for so the plants will also be either homozygous or heterozygous for the two transgene loci. If by chance a large number of plants that are homozygous for the

transgenes are selected this will results in there being three loci with strong selection for C24. This will not be a problem if the transgenes and M1/M9 locus are not linked to each other as three distinct areas of linkage to C24 will be detected. However, if the M1/M9 mutant is linked to either transgene this will mean that the area to which the mutant can be mapped to is increased as there is selection for a larger area of the C24 chromosome due to the presence of both the M1/M9 locus and one of the transgenes. This would consequently make identifying candidate genes from genome re-sequencing more difficult due to the larger area in which the locus could reside.

In order to produce a mapping position for the loci, the map distance of the mutant loci from each marker needed to be calculated. This is also known as the recombination frequency. The map distance is the maximum distance either side of the marker for the mutant locus' location and the locus can be found anywhere between these two points. To narrow down the location further multiple markers and their respective map distances to the mutant locus are compared and the region where these map distances overlap will be the location of the mutant locus. To calculate a map distance the parental status of markers are tested in a large number of DNA samples, each sample coming from a single plant that will have different crossover events and the markers are scored for being: homozygous for C24; homozygous for Ler; or a heterozygous. The number of DNA samples that are heterozygous and homozygous for Ler are recombination events and can be used to calculate the map distance using the formula shown in Equation 1.

$$\text{Mapping distance} = \frac{((2 \times L)+H)}{2 \times N}$$

Equation 1: Mapping distance equation

Equation to calculate the mapping distance between a marker and the linked locus of interest. It is measured by the total number of recombinant chromosomes divided by the total number of chromosomes. The number of recombinant chromosomes is the sum of all chromosomes in plants that are homozygous for Ler (L) and one of the chromosomes for each heterozygous (H) plant. N is the total number of DNA samples/plants screened, so is doubled to get the total chromosome number.

Initially CAPS and SSLP markers had to be identified that showed a difference between C24 and Ler. Seventeen CAPS markers with such a difference had previously been identified and these covered all five chromosomes (Table 3.1). Screening of the M1 and M9 DNA samples with these markers showed linkage with markers on the left arm of chromosome one. Linkage was not observed for the control DNA samples, suggesting that the markers were linked to the M1/M9 locus rather than the *GFP* or *35S IR* transgene (Table 3.2 and Figure 3.6). The markers on the left arm of chromosome one are PVV4 and CAT3. This gave a rough map position for the M1/M9 locus, however to produce a finer map position for the locus, more markers on the left arm of chromosome one were required. Using the TAIR marker database search tool NF21M12 and M59 were identified for this purpose and showed a difference between C24 and Ler (Table 3.1). These two markers also showed linkage to the M1/M9 locus (Table 3.2 and Figure 3.6).

Marker	Type	Chromosome	Position of chromosome (bp)	Restriction enzyme	Bands in C24 homozygote (bp)	Bands in Ler homozygote 9bp)
PVV4	CAPS	1	174459	*Afl*III	650 and 280	280 and 250
NF21M12	SSLP	1	3212189	NA	200	180
M59	CAPS	1	5855075	*Bstu*I	520, 130 and 110	520 and 240
CAT3	CAPS	1	7144251	*Hinc*II	970	790 and 180
GAPB	CAPS	1	16127765	*Dde*I	605, 284, 225 and 174	284, 255, 225 and 174
G4026	CAPS	1	22276033	*Rsa*I	650	800
ADH	CAPS	1	28975141	*Hae*II	1200	1400
GPA1	CAPS	2	11196996	*Afl*III	705, 680 and 209	1385 and 209
M249	CAPS	2	14454000	*Hpa*II	250 and 100	350 and 100
17D8LE	CAPS	3	589969	*Hinc*II	13000, 1500 and 600	1500, 1000, 650 and 600
GL1	CAPS	3	10360633	*Rsa*I	400 and 110	220, 160 and 110
F3H	CAPS	3	19025417	*Bcl*I	1200	700 and 350
GA1	CAPS	4	1242594	*Bfa*I	1200	3000, 1200, 900, 700, 500 and 200
AG	CAPS	4	10384141	*Dde*I	700, 230 and 220	700, 600, 230 and 220
CAT2	CAPS	4	16701115	*Dde*I	420, 260, 160 and 150	420, 280 and 260
HAE1	CAPS	5	980447	*Hae*III	850, 600 and 350	850, 620 and 350
NIT4	CAPS	5	7377397	*Hae*III	1750 and 650	1000 and 850
DFR	CAPS	5	17164364	*Bsa*AI	609 and 534	609, 318 and 216
10A10	CAPS	5	24572772	*Hph*I	1300, 310, 290 and 100	800, 500, 310, 290 and 100

Table 3.1: CAPS and SSLP markers that can be used differentiate between C24 and Ler

Table of all CAPS and SSLP markers used for mapping of the M1/M9 locus and their location on the chromosomes. The restriction enzyme used for each CAPS marker is also given and the size of all bands produced from each marker is also given.

Name	Chromosome	Position of chromosome (bp)	Genetic map position (cM)	Number C24	Number Ler	Number Heterozygote	Map distance (cM)
PVV4	1	174459	0.77	62 (89.86%)	0 (0.00%)	7 (10.14%)	5.07
NF21M12	1	3212189	14.24	83 (69.75%)	1 (0.84%)	35 (29.41%)	15.55
M59	1	5855075	25.99	131 (93.57%)	1 (0.72%)	8 (5.71%)	3.57
CAT3	1	7144251	31.69	150 (88.76%)	3 (1.77%)	16 (9.47%)	6.51

Table 3.2: Markers linked to the M1/M9 locus and their respective distances from the locus

CAPS and SSLP markers that are linked to the M1/M9 gene on chromosome one and their physical and genetic map positions. The combined number of M1 and M9 plants that were either C24 homozygotes, Ler homozygotes or heterozygotes are shown for each marker and the percentage of the total is given in brackets for each catergory. The map distance for each maker is given, this is the distance of the M1/M9 gene from the marker in centiMorgran (cM) map units.

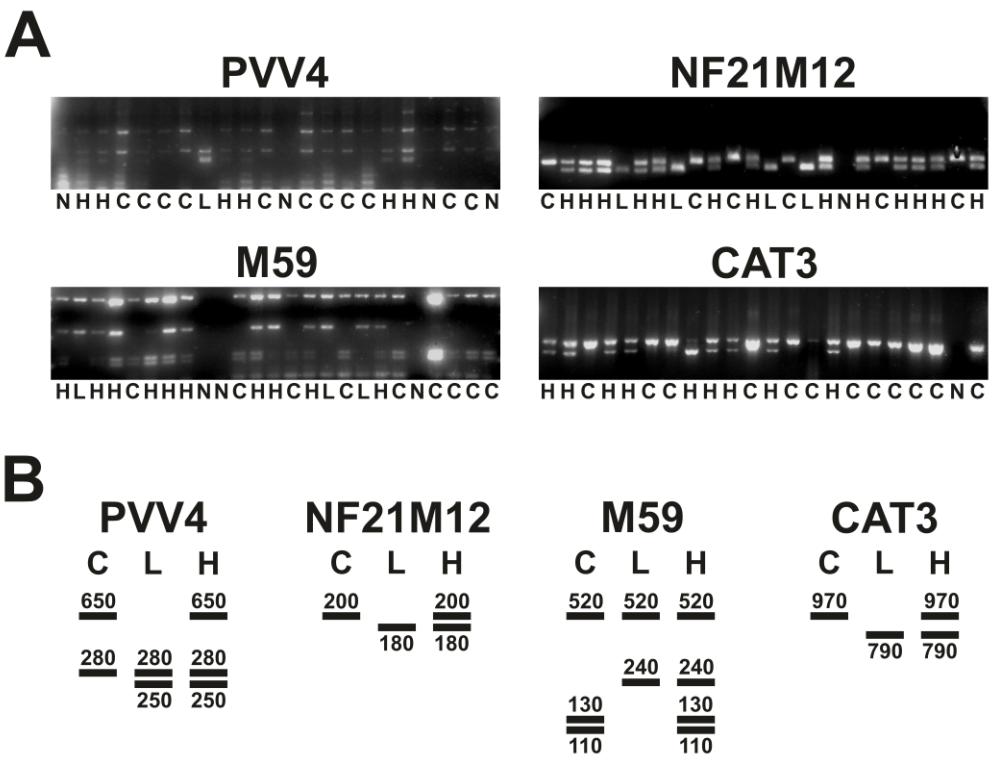


Figure 3.6: Examples of banding pattern of the linked CAPS and SSLP markers

A: Example images of gels of the PVV4, NF21M12, M59 and CAT3 markers in control DNA samples. The marker being tested for each gel is shown above the image. The status of the marker in each DNA samples is given below each image; with C being homozygous for C24, L being homozygous for Landsberg *erecta* and H being heterozygous. For PVV4, NF21M12 and M59 the gels show at least one DNA sample in each category, but CAT3 lacks a sample that is homozygous for Landsberg *erecta*. Lanes on the gel marked with N denote those where the marker failed to amplify. **B:** Diagrams showing the banding pattern for each marker when it is homozygous for C24 (C), homozygous for Landsberg (L) or heterozygous (H). The black lines represent the bands on the gels and the numbers above give their approximate size in base pairs (bp). The bands position are not to scale.

Having identified four markers on the left arm of chromosome one that showed linkage to the M1/M9 locus, the map distance of the markers to the locus were compared. The marker M59 showed the strongest linkage with 93.57% of plants

tested being a homozygote for the C24 variant of that marker, thus suggesting that the M1/M9 locus is closest to this marker. The second most strongly linked marker is PVV4 with 89.86% of plants tested being homozygous for the C24 variant. However, PVV4 and M59 are far apart from each other with PVV4 being at the end of the chromosome whereas M59 is nearly half way along the chromosome arm and so the mapping distances from these two markers do not overlap. It should also be noted that the NF21M12 marker, which is in between these two markers, has lower linkage than either PVV4 or M59. Both PVV4 and M59 do not show linkage in the control DNA samples, which would suggest that linkage to the one of the transgenes is not the cause of this discrepancy. One possible explanation is that double crossover events can cause stronger linkage and consequently a shorter distance between a marker and gene than there is in reality. Crossover events interfere with other nearby crossover events thus making it less likely that double crossover events will occur, however this interference decreases with distance, therefore the distance between PVV4 and M59 may allow for two crossover events to occur (Foss et al. 1993). Another possibility is that the markers for some DNA samples were mis-scored, for example if there was poor separation of the 280 bp and 250 bp bands when testing the PVV4 marker, thus leading to an incorrect mapping distance.

The map distance for three of the four markers overlapped with each other and so it was assumed that these markers were correct and the PVV4 marker was ignored when the region of interest (ROI) for the M1/M9 locus was determined. The region where the map distances of the NF21M12, M59 and CAT3 markers overlap defines the ROI for the M1/M9 locus and is located between 25.18 cM (5.68 Mb) and 29.56 cM (6.66 Mb), spanning 0.98 Mb (Figure 3.7). However, due to the comparatively low number of plants tested for each marker the M1/M9 locus could fall outside this region, so during sequencing mutants that are within 1 Mb of this region were also considered.

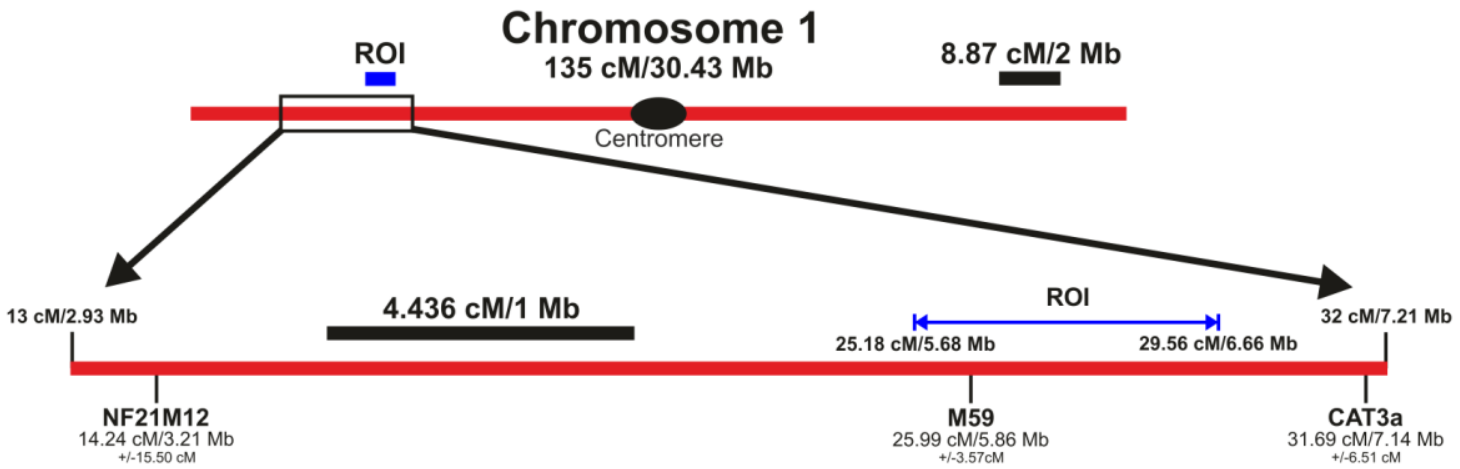


Figure 3.7: Location of the region of interest for the M1/M9 locus on chromosome one

Schematic showing the ROI for the M1/M9 locus on chromosome one both in terms of its relative position on the chromosome and its position in relation to CAPS and SSLP markers. The red lines represent the chromosome with the black oval being the centromere and the blue lines indicating the region of interest (ROI) of the M1/M9 locus. On the enlargement of a section of chromosome one, CAPS and SSLP marker locations are shown below the red line. The name of each marker is given along with its position (middle) and the maximum distance away from that marker that the M1/M9 locus is (bottom). A scale bar (black line) is given for both the whole chromosome and enlargement. In all cases positions are given in centiMorgans (cM) and Megabases (Mb).

3.3.3 Identification of candidate genes using Illumina sequencing of M1

Having identified a region of interest where the M1/M9 locus is located, Illumina genome re-sequencing of the M1 and 142S genome was then used to identify candidate genes for this locus. The M1 sequence was compared against the parental line sequence in order to identify the EMS mutations in M1. EMS mutagenesis predominantly produces G to A and C to T transitions, with >99% of all mutations being of this type, however G/C to C/G; G/C to T/A; and A/T to G/C can also occur at low frequency (Krieg 1963, Greene et al. 2003). This means that the differences between M1 and 142S should predominantly be G/C to A/T changes with other base substitutions being the result of either EMS mutagenesis or the natural rate of mutation between generations. However, the M1 sample is likely to carry a large number of EMS mutations other than the causal mutation, making its identification difficult. The mutation rate for EMS is between 10^{-6} to 10^{-7} events per base pair, which over the entire *Arabidopsis* genome equates to between 12.5 to 125 base changes in each mutant line (Kovalchuk et al. 2000). For this reason, steps were taken to reduce the number of EMS mutations carried in M1 by backcrossing and producing a pooled DNA sample, the strategy and principle behind this is explained in the methods section of this chapter (page 128). This process will not however, remove all other mutations, particularly those mutations that are closely linked to the causative mutation. This means that several candidate genes may be identified, although this number will be greatly reduced compared to that of an M1 sample that has not undergone the backcrossing and pooling process.

The 142S and M1 DNA samples were sequenced using a Illumina HiSeq 2000 platform. The M1 and 142S DNA samples were run in separate columns on a Illumina flowcell and this produced a total of 129,533,529 reads for M1 and 66,098,454 reads for 142S, which provides an approximate 100 times coverage of the genome for M1 and 51 times coverage of the genome for 142S. However,

for 142S this number of reads was lower than the 100 million reads per column that the HiSeq 2000 was expected to produce. A second sequencing run was therefore completed this time using an older platform, the Genome Analyser IIx, which produced at total of 187,244,483 read for M1 and 85,214,126 reads for 142S. When combined this gave a total of 316,778,012 reads for M1 with a 244 times coverage of the genome and a total of 151,312,580 reads for 142S with a 116 time coverage of the genome and so the combined data set was used for sequence analysis to compensate for the reduced number of reads in the first sequencing run.

To identify and map the mutation sites in M1 compared to 142S, the SHORE mapping and analysis pipeline was used. SHORE is a sequencing analysis tool that can both assemble a genome from a reference genome and identify single nucleotide polymorphism (SNPs) differences between the reference genome and the data set being tested (Ossowski et al. 2008, Schneeberger et al. 2009). The SHORE analysis was carried out by Xhesi He at the University of York. To analyse the sequence data a SHORE-specific predefined directory structure was created to store raw data and alignment results. Initially the C24 reference genome, obtained from the 1001 genomes website, was preprossed into a SHORE format for mapping the sequencing data to it (Cao et al. 2011, Schneeberger et al. 2011). The sequence data from the Illumina sequencing was in the FastQ format so was first converted into a SHORE flat file format. For 142S the reads from the first and second sequencing runs were mapped to the C24 reference genome separately before being merged together and the same process was carried out for M1. Consensus analysis was then performed between 142S and M1 to identify SNPs, indels and CNVs (Copy number variation). In this analysis, as the C24 reference genome is not complete, the results are expected to be less informative than comparing to a complete genome because of the incontinuity and lack of definite location on the chromosomes. This indeed was the case as for example, no insertions were detected, but SNP differences between M1 and 142S could be identified. SHORE identifies all SNP differences between the reference, 142S and

M1 so the SNPs had to be filtered to select for homozygous SNPs and ones that are present in M1 but not in 142S. To provide further information, such as indels or CNVs, the analysis was repeated with the Columbia genome as the reference genome, as this is fully annotated. However, this resulted in a dramatic increase in the number of homozygous SNPs identified from one hundred and seventy nine using C24 to eighteen thousand nine hundred and fifty nine using Columbia. This discordance, although unexpected, between the two SHORE analyses is due to the difference between C24 and Columbia and therefore only the analysis using C24 was used.

Analysis of the sequencing data using SHOREmap identified one hundred and seventy nine SNPs where there was a difference between M1 and the parental 142S line and the change was homozygous, of which sixty nine are located on the left arm of chromosome one. A table of all one hundred and seventy nine SNPs is shown in Appendix 1.3 (page 353). For the majority of the SNPs identified on the left arm of chromosome one the base change from 142S to M1 is either C to T or G to A, which is the predominant mutation caused by EMS mutagenesis, however for SNPs identified on other chromosomes or the right arm of chromosome one, C to T and G to A changes were not the predominant mutation. This shows that homozygous EMS-induced SNPs were enriched within or near the region of interest. As to the nature of the SNPs not on chromosome one's left arm, some will also be the result of EMS mutagenesis and others could either be due to the natural mutation rate between generations or artefacts of the sequencing analysis. The natural mutation rate for *Arabidopsis* is 7×10^{-9} base substitutions per generation, equating to a change of around 0.88 bp per generation, so some of these SNPs identified will be due to the natural mutation rate (Ossowski et al. 2010).

From the mapping analysis the causal mutation was expected to be amongst the sixty nine SNPs on the left arm of chromosome one. Therefore in order to identify the mutation, the location of each SNP and whether it is within a gene or

intergenic region was determined for those SNPs on the left arm of chromosome one. The C24 genome has not been fully annotated and is in sections called scaffolds. Although a rough position for each scaffold is known, where each scaffold corresponds to on the fully annotated Columbia genome is not known. This means that despite knowing the exact location of each SNP on the scaffolds, it was not known if these locations correspond to a gene or intergenic region nor where on the left arm they are located. To identify the exact location of each SNP in the genome the upstream and downstream sequence for each SNP was identified from the C24 genome and this sequence was then used in a BLAST search of the Columbia genome. This gave the exact location of the SNP on the left arm and whether it was in genic or intergenic sequence. For those SNPs that were within genes, the function of the gene was recorded, as well as whether the SNP was in a UTR, exon or intron and, if it was within an exon, whether the mutation caused an amino acid change or not. For intergenic mutations the distance to the nearest gene and the function of the gene were recorded. There were a total of nineteen intergenic SNPs, ten intronic SNPs, seven located in UTRs and thirty three exonic SNPs. A table showing the location and nature of all chromosome one left arm SNPs is shown in Appendix 1.4 (page 358). Of the thirty three exon SNPs, seventeen caused a change in amino acid sequence (Table 3.3). Further analysis concentrated on the exon mutations that caused an amino acid change as, although other mutations may also impact on gene function, amino acid changes were the most likely to cause loss of function.

SNP position (bp)	TAIR ID	Gene name	Exon	Codon	Base change	Amino acid change	Proportion of concordance in M1
272314	AT1G01740	AT1G01740	10	417	G to A	Glutamine (CAA) to Stop codon (UAA)	0.87 (111)
448847	AT1G02280	TOC33	6	264	C to T	Glycine (GGA) to Arginine (AGA)	0.86 (127)
613895	AT1G02810	AT1G02810	1	38	C to T	Serine (UCC) to Phenylalanine (UUC)	0.85 (113)
989786	AT1G03890	AT1G03890	2	145	C to T	Phenylalanine (UUC) to Leucine (UUA)	0.89 (89)
1893527	AT1G06190	AT1G06190	1	49	C to T	Glycine (GGG) to Glutamic acid (GAG)	0.87 (109)
2543135	AT1G08130	LIG1	16	756	C to T	GAA (Glutamic acid) to Lysine (AAA)	0.90 (96)
3753966	AT1G11200	AT1G11200	2	24	C to T	Serine (UCG) to Stop codon (UAG)	0.91 (125)
3816744	AT1G11340	AT1G11340	1	226	C to T	Glycine (GGA) to Glutamic acid (GAA)	0.93 (132)
4798406	AT1G14000	VIK	3	146	C to T	Proline (CCA) to Serine (UCA)	0.95 (120)
5326847	AT1G15500	NTT2	1	142	C to T	Alanine (GCU) to Valine (GUU)	0.96 (127)
5419895	AT1G15750	TPL	4	98	C to T	Valine (GUG) to Methionine (AUG)	0.98 (106)
5677535	AT1G16610	SR45	6	195	C to T	Aspartic acid (GAU) to Asparagine (AAU)	0.96 (103)
6595680	AT1G19100	AT1G19100	2	41	C to T	Glutamine (CAA) to Stop codon (UAA)	0.97 (109)
6651377	AT1G19250	FMO1	4	360	C to T	Glycine (GGG) to Glutamic acid (GAG)	0.96 (126)
7280563	AT1G20910	AT1G20910	2	25	C to T	Glutamic acid (GAA) to Lysine (AAA)	0.98 (106)
7445662	AT1G21270	WAK2	1	222	C to T	Proline (CCU) to Serine (UCU)	0.93 (124)
13085698	AT1G3552	ARF15	2	19	T to C	Arginine (AGA) to Glycine (AGA)	0.81 (189)

Table 3.3: SNPs on the left arm of chromosome one that result in an amino acid change in M1

Table of SNPs found within exons that cause a change in the amino acid sequence. The position of each SNP in the genome is given, as is the TAIR ID and gene name of the gene it is located in. For each SNP the exon and codon it is located in is given as is the base change and amino acid change. The

proportion of concordance in M1 is also given, this is the proportion of reads that contain this mutation, the exact number of reads that have this base change is shown in brackets in each case.

Of the seventeen SNPs with amino acid changes, there are three SNPs that resulted in a premature stop codon, which are likely to lead to a loss of function of that gene and they are At1G01740, At1G11200 and At1G19100 (Table 3.3). These were therefore good candidates for the M1/M9 locus. However, two of these, At1G01740 and At1G11200, are over 1.5 Mb away from the region of interest (ROI) so may not be the causal mutation, although At1G01740 is close to the marker PVV4 so cannot be discounted. The other gene with a premature stop codon, At1G19100, is within the ROI so was a strong candidate. The function of all sixteen genes was also ascertained and this identified *SR45*, which has a aspartic acid to asparagine amino acid change, as another strong candidate for the M1/M9 RdDM mutant. This is because it has been previously identified as a RdDM mutant, is within the ROI, and the amino acid change is not conservative as it is a change from an acidic amino acid to a non-acidic amino acid (Table 3.4) (Ausin et al. 2012a).

Gene name	Function
AT1G01740	Protein kinase containing tetratricopeptide repeat domain
TOC33	GTPase involved in chloroplast importation machinary
AT1G02810	Plant inverstase/pectin methyltransferase inhibitor, inhibits cell wall modification
AT1G03890	RmlC-like cupins superfamily protein, involved in metabolism
AT1G06190	Rho termination factor, involved in transciption termination
LIG1	DNA ligase involved in DNA replication and base excision repair
AT1G11200	Unknown function
AT1G11340	S-locus protein kinase, involved in pollen recognition
VIK	MAP Kinase involved in auxin signalling
TPL	Transcription factor involved in repressing root development in aerial tissues
SR45	Involved in RdDM and spliceosome
AT1G19100	MORC1 homolog, GHKL ATPase
FMO1	Response to viral infection, promotes cell death
AT1G20910	DNA binding proteins which contains ARID/BRIGHT domains
WAK2	Serine/threonine protein kinase involved in cell expansion
ARF15	Auxin response transcription factor

Table 3.4: Function of genes with an amino acid change

Table of gene function for genes that have an amino acid change in M1 on chromosome one. Each gene function was determined from the TAIR annotation of each gene.

As well as *SR45* and At1G19100 there are five other genes with mutations that are either within the ROI or are 1 Mb either side (between 4.59 Mb to 7.71 Mb) (Table 3.5). Although, *SR45* and At1G19100 were tested first, the other five mutations were also good candidates for the causal mutation due to their location in the genome, so would be tested next if neither *SR45* nor At1G19100 were the causal mutant. During the completion of this work At1G19100 was also identified as a RdDM mutant and was shown to be a ATPase that is involved in both DNA methylation and higher order chromatin modification in the RdDM pathway (Lorković et al. 2012, Moissiard et al. 2012). Due to the nature of the mutation in M1 and the fact that At1G19100 is a RdDM component this made At1G19100 the top candidate for the causal mutation in M1 and M9. The two studies also showed that At1G19100 is a ortholog of *Mircorchidia 1* (*MmMORC1*) in *Mus musculus* so was named *AtMORC6* or *MORC6* and will be referred to as such from henceforth in this work. The reason it is called *MORC6* is that it is the sixth of seven paralogs in *Arabidopsis* to *MmMORC1.*

SNP position (bp)	Gene name	Exon	Codon	Base change	Amino acid change
4798406	VIK	3	146	C to T	Proline to Serine
5419895	TPL	4	98	C to T	Valine to Methionine
5677535	SR45	6	195	C to T	Aspartic acid to Asparagine
6595680	At1G19100	2	41	C to T	Glutamine to Stop codon
6651377	FMO1	4	360	C to T	Glycine to Glutamic acid
7280563	At1G20910	2	25	C to T	Glutamic acid to Lysine
7445662	WAK2	1	222	C to T	Proline to Serine

Table 3.5: Mutations in M1 that are within the ROI for the M1/M9 locus

Table of mutations within or 1 Mb either side of the ROI in M1. The position of each mutant in the genome is given. The gene, exon and codon the SNP is located in is also given, as is the base and amino acid change for that SNP.

3.3.4 Sequencing of M9 and rmd6 to identify *MORC6* as the mutated gene

To determine which of the candidate genes is the causal mutation the candidates were sequenced by Sanger sequencing in the other two alleles, M9 and rmd6. The RdDM gene is expected to have a mutation in all three lines whereas it is unlikely that a gene that has no effect on the silencing phenotype will be mutated in all three lines. The top two candidate genes from SHOREmap analysis were *SR45* and *MORC6* so these two were initially sequenced. No exon or intron mutations were detected in M9 for *SR45,* but premature stop codon mutations were detected in *MORC6* for both M9 and rmd6 (Figure 3.8).

The *MORC6* gene is 6.43 Kb in length and is predicted to encode a protein 663 amino acids in length (Figure 3.8 and Figure 3.10). The coding sequence (CDS) of *MORC6* is split between nineteen exons and the 1st exon is non-coding, as it is part of the 5'UTR (Figure 3.8 A). A full genomic sequence for *MORC6* can be found in Appendix 1.5 (page 366) and this also shows ecotype specific differences between the Columbia and C24 ecotypes, most of which are in introns and the 5'UTR. There is however one exon SNP found in *MORC6,* but this does not cause an alteration in amino acid sequence as it is a change from the Leucine codon CTC to the Leucine codon CTT and can therefore not affect MORC6 function.

The premature stop codon in M9 is in codon 267 in the 9th exon and is a change from the glutamine CAA codon to the stop codon TAA, whereas rmd6 has a premature stop codon in codon 392 in the 12th exon and is a change from the glutamine CAG codon to the stop codon TAG (Figure 3.8 B). The mutation in M1 is a mutation of the 41st codon in exon two from a CAA glutamine codon to the TAA stop codon. The premature stop codons will either lead to the production of a truncated *MORC6* protein or will trigger nonsense mediated decay and therefore no protein will be produced from *MORC6*. Preliminary results indicate

that a transcript is produced in all three lines and would therefore produce a truncated form of the MORC6 protein (Jones, L. personal communication). The presence of a premature stop codon in all three cases means it is therefore likely that the mutations in *MORC6* are indeed the causative mutations. However, to conclusively show this an allelism test between M1 and a GABI-KAT line that has a T-DNA insertion in the *MORC6* gene, *morc6-3,* was carried out. The T-DNA insertion was previously characterised in the study by Moissiard that identified MORC6 as an RdDM component and was shown to lack expression of *MORC6* (Moissiard et al. 2012). M1 and wilditype (WT) Columbia were also crossed with each other to show that the *35S IR* transgene could silence the *GFP* transgene when both are in a heterozygous state, which is indeed the case (Figure 3.9 A). The progeny of the cross between *morc6-3* and M1 showed a lack of complementation, seen by GFP fluorescence, which confirms that *MORC6* is the causative mutation in M1, M9 and rmd6 (Figure 3.9 A).

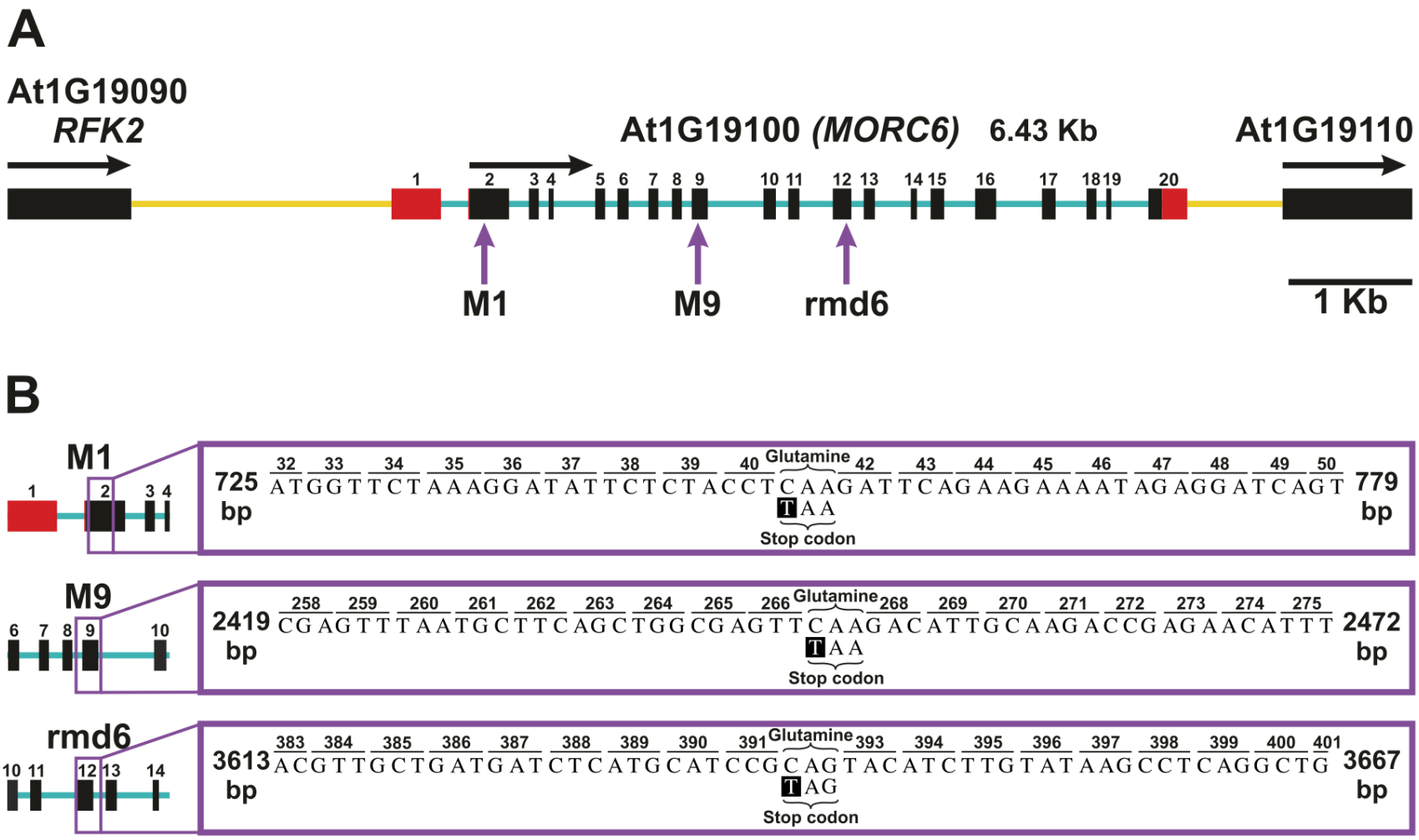

A
At1G19090
RFK2
At1G19100 (MORC6) 6.43 Kb
At1G19110
M1
M9
rmd6
1 Kb
B
M1
725 bp
ATGGTTCTAAAGGATATTCTCTACCTCAAGATTCAGAAGAAAATAGAGGATCAGT
Glutamine
TAA
Stop codon
779 bp
M9
2419 bp
CGAGTTTAATGCTTCAGCTGGCGAGTTCAAGACATTGCAAGACCGAGAACATTTT
Glutamine
TAA
Stop codon
2472 bp
rmd6
3613 bp
ACGTTGCTGATGATCTCATGCATCCGCAGTACATCTTGTATAAGCCTCAGGCTG
Glutamine
TAG
Stop codon
3667 bp

Figure 3.8: The location within the *MORC6* gene of the mutations in M1, M9 and rmd6

Schematic showing the structure of the At1G19100 (*MORC6*) gene and the location of the mutations in M1, M9 and rmd6. **A:** Is a scale diagram of the structure of the *MORC6* gene and its position in relation to the upstream and downstream genes. The yellow lines represent intergenic DNA and the upstream and downstream genes of At1G19100 are shown in black, with the direction of transcription shown by the black arrow above each gene. The UTRs of At1G19100 are shown in red, with exons in black and introns in blue. The exon number is given above each exon and purple lines below the gene highlight where the M1, M9 and rmd6 mutations are located. A scale bar is given in the bottom right. **B:** Shows the location of each mutation in terms of exon and codon and gives the sequence for each mutation. The right hand purple box shows the sequence of the area highlighted to the left of the diagram. For each box the top sequence is the C24 WT sequence with the change in the respective mutant line shown below, with the substituted base highlighted in black. The codons are represented by a black line above the sequences and the codon number for each one is also given. For the mutated codon the original amino acid encoded is given above the sequence. The position within the *MORC6* sequence for the start and end of each sequence shown is given at the beginning and end of each sequence.

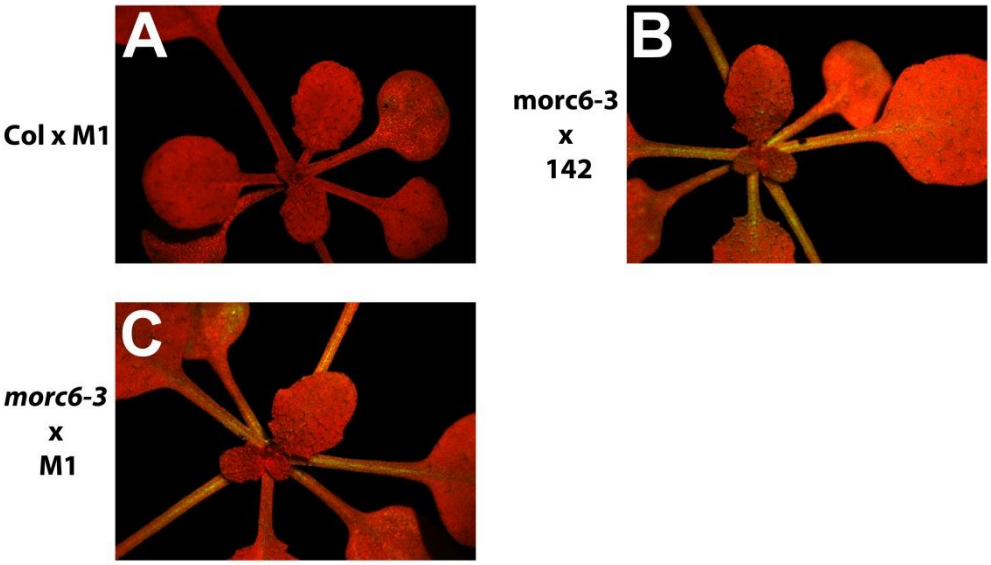


Figure 3.9: Allelism test demonstrating M1 is a *morc6* mutant

Images of the progeny of crosses between WT Columbia and M1 (**A**), *morc6-3* and 142 (**B**), and *morc6-3* with M1 (**C**) under UV light. In each case the first name is the male and the second is the female. The images were taken at 18 dpg and are representative for each line.

3.3.5 Truncation of the MORC6 protein in the mutants

If *MORC6* is expressed in M1, M9 and rmd6 it will result in truncated forms of the MORC6 protein that are 40, 266 and 391 amino acids long respectively (Figure 3.10). To predict what affect truncation of MORC6 may have, the protein domains of MORC6 first needed to be identified using the protein databases Pfam and SMART, both of which can identify putative protein domains (Letunic et al. 2012, Punta et al. 2012). This analysis identified two functional domains in MORC6, these being a GHKL ATPase domain and a coiled coil domain (Figure 3.10).

The GHKL domain contains four conserved sequence motifs responsible for ATP binding, the sequences for which are shown in Table 3.6, and all four could be identified in MORC6 (Figure 3.10) (Bergerat et al. 1997, Mushegian et al. 1997,

Ban and Yang 1998, Lorković et al. 2012). There is however sequence divergence of the conserved motifs in MORC6, for example in motif I the highly conserved alanine residue at the end of the motif is instead a glutamic acid residue in MORC6, which is a change from a small hydrophobic amino acid to a larger acidic amino acid (Table 3.6). Of the motifs, motif III shows the greatest difference between the consensus sequence and MORC6 sequence as there has been an addition of four amino acids to the motif and a conversion of two highly conserved threonine residues to serine residues (Table 3.6). However, the substitution of threonine for serine is a relatively conservative change as both are hydrophilic and differ by a single methyl group. It should be noted that motif IV is outside of the ATPase domain defined by both Pfam and SMART so the ATPase domain will extend beyond the area identified by these two databases and the previous study by Lorković suggests the domain extends to the 500th amino acid (Figure 3.10) (Lorković et al. 2012). The reason why this was not identified is that motif IV contains only three conserved residues and these can be separated by any number of random amino acids, hence the failure of the databases to identify this as part of the ATPase domain.

	Consensus Sequence	**MORC6 Sequence**
Motif I (N box)	uubEuuaNouDA	ouoEuuaNouDE
Motif II (G1 box)	uxuxDNGxGuxbaauxxuu	uxuxDDGxGuxuxxuxxuu
Motif III (G2 box)	uGxxGxouxSxxxuoxbuTuxT	uGxxGuouxxSxxxuoxauuuuSuxS
Motif IV (G3 box)	Tx_nGT	Tx_nGT

Table 3.6: Conserved sequence motifs in GHKL ATPases

Table showing the consensus sequence and MORC6 sequence for the four conserved sequence motifs of GHKL ATPases. The amino acids in capitals are specific highly conserved amino acid residues whereas the lower case can be any amino acid with a particular set of properties, those being: u are amino acids which have a bulky side group and are hydrophobic; b are amino acids that are basic; a are amino acids that are acidic; o are amino acids with small side chains; and x can be any amino acids. For motif IV the x_n can be any number of amino acids.

The coiled coil domain was identified by the SMART database and is found at the C terminal end of the protein (Figure 3.10). This coiled coil domain allows for binding and interaction with other proteins and MORC6 has been shown to interact with DMS3 through this domain (Crick 1952, Mason and Arndt 2004, Lorković et al. 2012). DMS3 is a RdDM component and has a hinge domain found in structural maintenance of chromosome (SMC) proteins, but lacks the ATPase domains of SMC proteins; MORC6, through its interaction with DMS3, therefore provides the missing ATPase domain (Hirano 2005, Kanno et al. 2008, Ausin et al. 2009, Lorković et al. 2012). It is therefore likely that MORC6 and DMS3 function together in the RdDM pathway as a SMC protein, although MORC6 functioning separately to DMS3 cannot be discounted.

Having identified the protein domains in MORC6, the three mutants were then analysed to determine what affect they would have on the domains. The M1 MORC6 protein lacks all functional domains identified and therefore can be assumed to totally disrupt MORC6 function. The M9 protein still has the majority of the ATPase domain identified by Pfam and SMART, however it lacks the fourth motif of the GHKL ATPase and is also lacking the coiled coil domain. This means that it should not be able to interact with DMS3 so any function in RdDM that requires both proteins interaction should be inhibited. The ATPase domain in M9 will be disrupted to some extent as the fourth motif is missing so the protein will not fold in the same way as the WT protein; whether this would affect RdDM function is unknown. rmd6 produces the most complete MORC6 protein and contains all four conserved GHKL ATPase motifs. As with M9 it lacks the coiled coil domain so interaction with DMS3 should again be prohibited. It should also be noted that the study by Lorković suggests the ATPase domain extends to the 500th amino acid and therefore rmd6 would lack 109 amino acids of the ATPase domain, this would again affect protein folding.

3. Identification of M1 and M9 as morc6 *mutants*

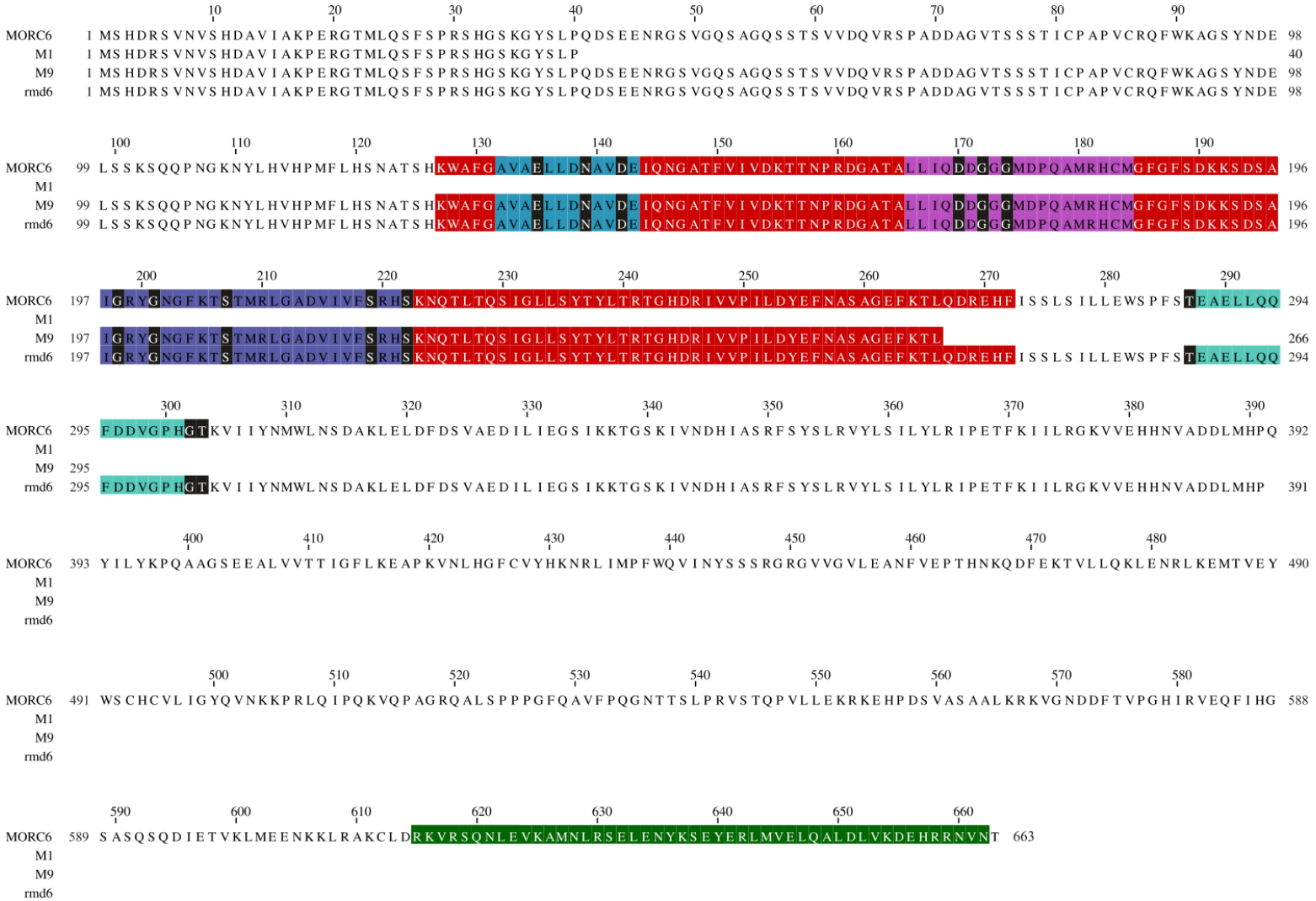


Figure 3.10: Truncation of the MORC6 protein in the three *morc6* mutants compared to WT

Alignment of the M1, M9 and rmd6 protein sequences against the WT MORC6 protein sequence, the sequence named MORC6 being the WT sequence. The domains of MORC6 have been highlighted; the GHKL ATPase domain is highlighted in red while the coiled coil domain is highlighted in green. The conserved sequence motifs of the GHKL domain are highlighted: motif I is in light blue; motif II is in light purple; motif III is in dark blue; and motif IV is in turquoise. The highly conserved residues within these motifs are highlighted in black.

Although M1 clearly lacks all identified functional domains of MORC6, both M9 and rmd6 contain at least part of the GHKL ATPase domain. In order to determine whether this affects protein folding Phyre 2 was used to predict the 3D structure of the WT MORC6, M9 MORC6 and rmd6 MORC6 GHKL ATPase domains (Figure 3.11). Phyre 2 identifies 3D structure through homology with other proteins whose 3D structure is known and then uses this known structure to predict that of the protein being tested (Kelley and Sternberg 2009). All predicted 3D structures show the strongest homology to the HSP90 GHKL ATPase domain and start near the beginning of the predicted ATPase domain but end at different points (Figure 3.10 and Table 3.7). The M9 3D structure ends at amino acid 226, this being the last amino acid in the M9 MORC6 protein, whereas for rmd6 the structure ends at amino acid 285, which is not at the end of the rmd6 protein. In both cases, although both proteins form a cleft where ATP binding and hydrolysis occurs, they lack the larger extraneous structure of WT MORC6 (Figure 3.11). In WT MORC6 the cleft is deeper and partly enclosed and this larger structure may be required for the correct function of the GHKL ATPase domain, therefore both M9 and rmd6 would not function normally as the 3D structure of the GHKL ATPase domain is disrupted. However, it cannot be concluded that they are a null mutants, like M1, as both the M9 and rmd6 proteins have at least part of the ATPase domain and therefore could possess some residual function.

MORC6	Predicted 3D structure between amino acids	Strongest homology
WT	132-490	HSP90 Chain A
M9	129-226	HSP90 ATPase domain
rmd6	129-285	HSP90 Chain A

Table 3.7: Size of the predicted structure of the WT, M9 and rmd6 MORC6 GHKL ATPase domain

Table showing the amino acids where a predicted 3D structure has been found for each protein tested and which protein with a known 3D structure it is most homologous to.

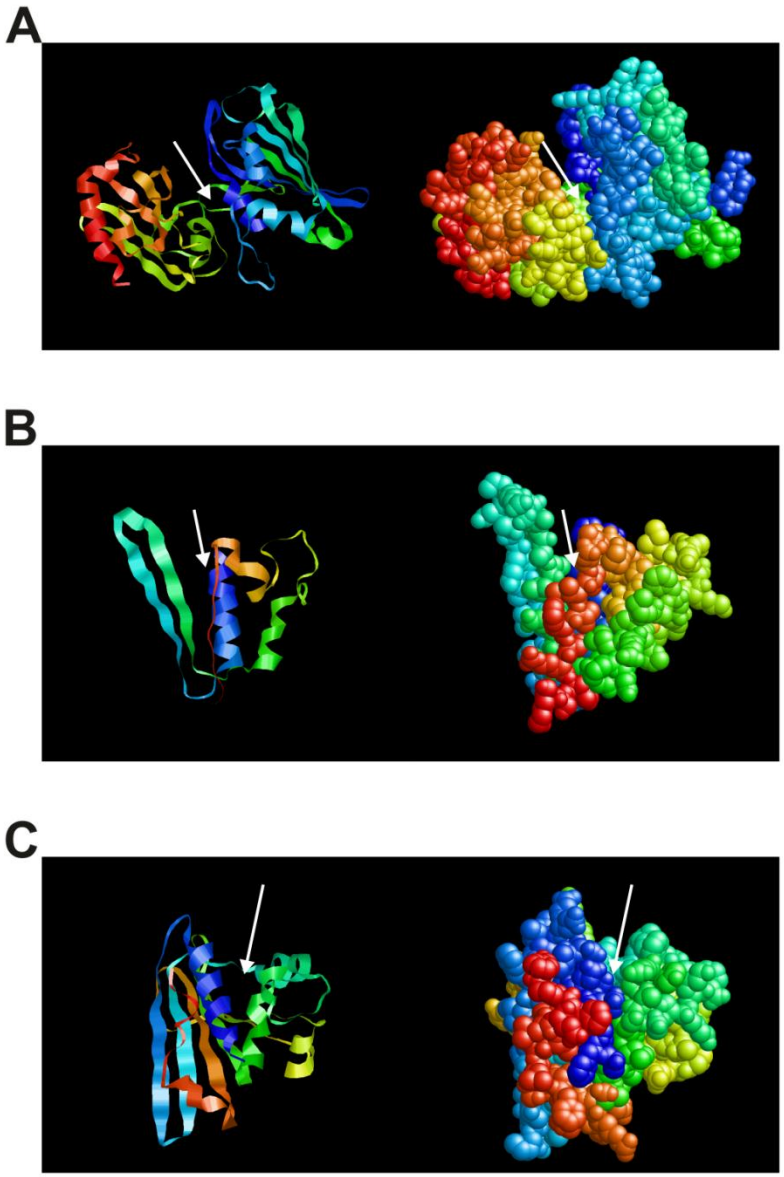


Figure 3.11: Differences in the predicted 3D structures of the GHKL ATPase domain between the WT, M9 and rmd6 MORC6 proteins

3D structures of the MORC6 GHKL ATPase domain for the WT protein (**A**) M9 protein (**B**) and rmd6 protein (**C**). For each protein a ribbon and space fill model of the protein is given. For the ribbon model α helices are shown as spirals; β sheets are flat sections and unstructured section are thin strands. The colour represents where in the sequence each section is located, with the N terminus of the structure being red and then going through the spectrum to dark blue at the C terminus. The catalytic domain and cleft where ATP binding and hydrolysis occurs is highlighted in each case by a white arrow

3.3.6 *Arabidopsis* MORC protein family

BLAST searches of the *MORC6* DNA and protein sequence revealed that *MORC6* is one of a family of seven *MORC* genes in *Arabidopsis thaliana* that are found on chromosomes one, four and five. A list of the seven *MORC*s and their location in the genome can be found in Appendix 1.6 (page 372). Three of the *MORC* genes*, MORC1, MORC2* and *MORC3,* are located next to each other and could therefore be the consequence of recent gene duplication events. Gene size, exon number and protein length are similar amongst the *MORC* gene family with most *MORC* genes being around 4 to 4.5 Kb long containing eighteen exons and producing a protein around 600-700 amino acids in length. There are exceptions to this, for instance *MORC6* is 1.49 Kb larger than the next largest *MORC* gene although it produces a protein 663 amino acids long. As well as *MORC6, MORC1* has also been shown to be involved in RdDM, however whether any of the other MORC proteins are involved in RdDM is unknown (Moissiard et al. 2012). As well as a role in RdDM, MORC1 and MORC6 also have a more general role in formation of heterochromatin and its localisation to the nuclear membrane, seen by the loss of chromocentres in mutants of both genes. The only other function of MORC proteins so far determined experimentally is that both MORC1 and MORC2 are involved in pathogen response in the cytoplasm (Kang et al. 2008, Kang et al. 2010).

3.3.7 Difference in gene expression of *Arabidopsis MORC*s

Having identified that there are multiple MORC proteins within *Arabidopsis,* the expression patterns of the *MORC* genes were then analysed using Genevestigator. Genevestigator has a large database of microarray data that can be used to assess gene expression patterns in different mutant backgrounds, stages of development and tissue types (Hruz et al. 2008). For this analysis only microarray data from WT Columbia and C24 plants were used, except for RdDM mutant analysis, where *rdr2-1* and *dcl3-1* Columbia mutant microarray data was used. The reason that both Columbia and C24 used is that there is very little C24 microarray data available compared to Columbia and so this could skew the results. Expression levels between Columbia and C24 when looked at separately show similar patterns in terms of whether a *MORC* gene is up or down regulated, but the exact level of gene expression is different between the two ecotypes (Appendix 1.7 page 374). Again this is likely to be due to the low number of C24 microarray data available, but could also be a ecotype difference. Data was available on gene expression for *MORC*s 1, 2, 4, 5, 6 and 7 but no data was available for *MORC3.* The data used in this analysis were produced produce by the normalisation of absolute expression levels, which are measured by the intensity of the signal for the *MORC* gene probes on the array. Absolute values will vary between different microarrays so the data provided by Genevestigator is normalised using a trimmed mean to provide a relative intensity level. In this case the 90% trimmed mean is used, meaning that the top and bottom 5% of expression values are ignored and the mean calculated from the remaining values. This mean is then given an expression value that is the same for all microarray data stored in Genevestigator and the values for specific probes are calculated using the ratio between the trimmed mean and the absolute value.

Initially *MORC* expression was assessed during development to determine if any of the *MORC* genes are developmentally regulated. Genevestigator has eight broad classifications for development of *Arabidopsis*, which are: germinated

seed, seedling, young rosette, developed rosette, bolting plant, young flowers, developed flower, flowers and siliques, mature siliques and plant senescence. The reason for these broad classifications is that compared to more specific classifications, more microarrays can be included into each classification, which will reduce the effects of specific conditions in the microarray that affect gene expression but are not linked to development. From the developmental expression data there is variation in expression between the *MORC* genes (Figure 3.12 A). *MORC5* expression is low at all developmental stages and is low enough that the relative intensity may be artefacts from the microarrays rather than actual expression. Both *MORC2* and *MORC7* have a similar expression pattern in that both have a generally low level of expression but expression increases from germination, peaking during bolting, before decreasing again. This could suggest that both *MORC2* and *MORC7* are involved in activities related to rosette development and bolting. *MORC4* is relatively stable during plant growth but shows high levels of expression during seed development and plant senescence. However, as there are only six microarrays for plant senescence the large increase in *MORC4* expression may not be connected to development but rather due to other factors. There is however a larger number of microarrays supporting the high expression levels of *MORC4* in germinated seeds, so *MORC4* could have a role during germination. *MORC1* expression oscillates during development, but is generally higher than all other *MORC* proteins, except during flowering. This would suggest that *MORC1* is active throughout plant development. *MORC6,* like *MORC1,* is highly expressed, compared to the other *MORC* genes, but is up regulated during flowering and siliques formation and therefore could have a role during this developmental stage, for instance in the establishment of silencing of transposable elements in pollen and ovules (Mosher et al. 2009, Slotkin et al. 2009).

As well as *MORC* expression during development the expression of *MORC*s in different plant tissues was also assessed and again *MORC* genes showed a difference in expression patterns (Figure 3.12 C). The expression pattern of

MORC6 would suggest that its role in flowering is specific to the stamen and pistil, as expression is high in these tissues whereas in siliques, sepals and petals *MORC6* expression is similar to that of the vegetative tissues. This would suggest that MORC6 has a specialised role in pollen and ovule development. However, the number of microarrays for some of these tissues is low so the mean values may have been skewed by the small sample size. *MORC5* again has a low level of gene expression and this is consistent across all tissues except pollen, so *MORC5* may also be involved in pollen development. The expression of the *MORC*s was also assessed in RdDM mutant lines for *rdr2-1* and *dcl3-1* in the Columbia background (Figure 3.12 B). Only a limited number of microarray data sets are available for the two mutants so other factors could skew the results and expression data is only available for the flowers and siliques developmental category. The expression pattern of *MORC2* and *MORC7* do not show much variation between the WT and mutant arrays. *MORC1, MORC4, MORC5* and *MORC6* do however show a change in expression. Expression is reduced for *MORC1, MORC4* and *MORC5* in both mutants and *MORC6* in the *dcl3-1* mutant, but there is increased expression of *MORC6* in the *rdr2-1* mutant. This change in expression, if genuine, could suggest that the *MORC* genes are regulated by RdDM or that the genes are up or down regulated in response to perturbation of the RdDM pathway.

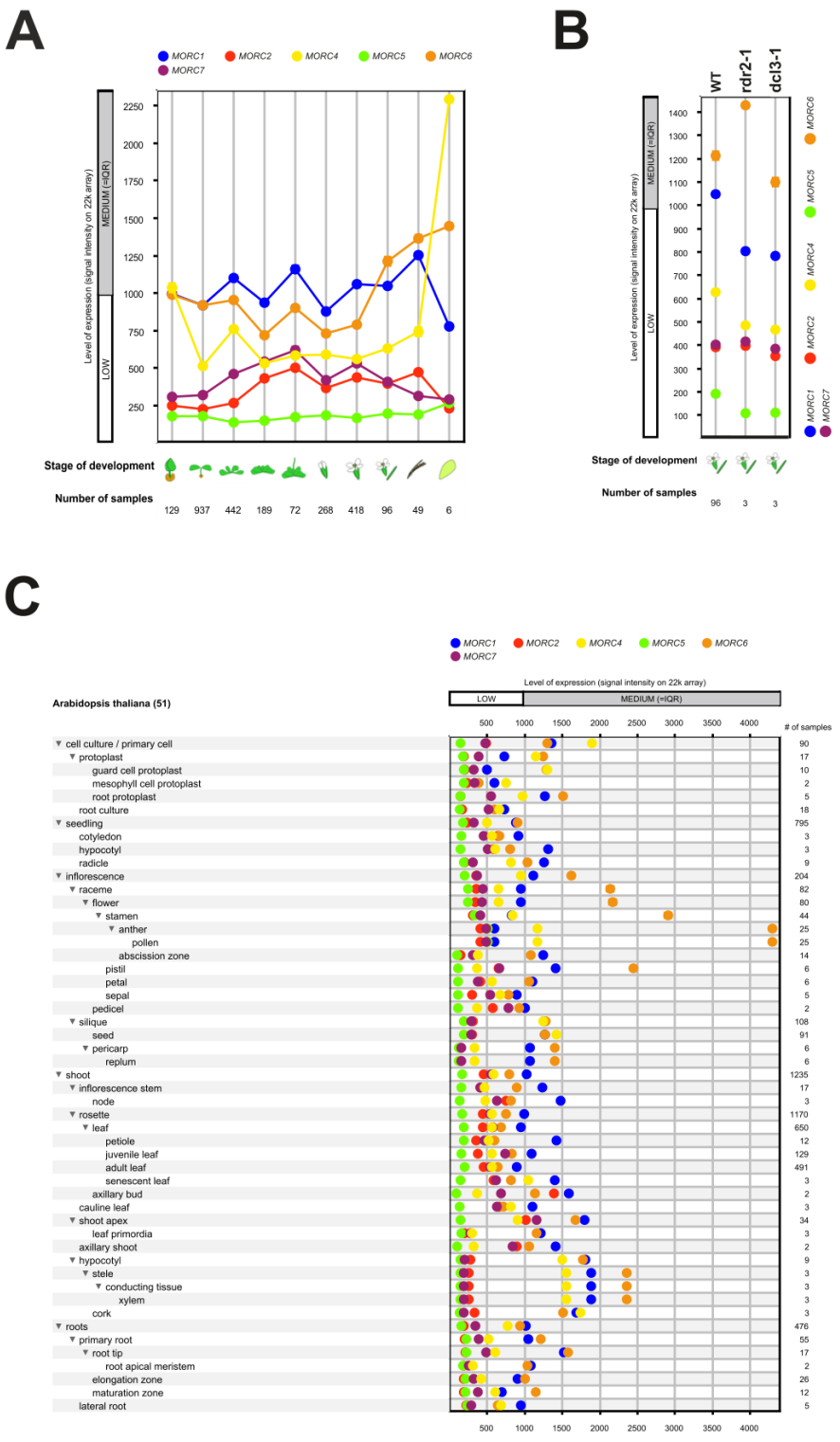


Figure 3.12: Gene expression of the *MORC* gene in *Arabidopsis* during development, in RdDM mutant backgrounds and in different tissues

Graphs produced by Genevestigator of microarray data on *MORC* gene expression for different developmental stages (**A**), RdDM mutant backgrounds (**B**) and tissues (**C**). For each graph a key shows which MORC corresponds to each coloured circle; the circles themselves represent the mean level of gene expression for that *MORC* gene. The error bars are two standard errors of the mean but are mostly obscured by the circles themselves. Gene expression is measured using a relative intensity scale; low levels of gene expression are considered to be genes with a relative intensity among the 25% lowest intensities for all genes and medium levels of gene expression are considered to be those within the interquartile range (IQR) of relative intensity. **A:** Shows gene expression at different developmental stage, the stages are shown as pictures. The number of microarrays from which the mean of each *MORC* is derived is shown below each developmental stage. **B:** Shows gene expression in flowers and siliques in the *rdr2-1* and *dcl3-1* mutants. Again the number of microarrays from which the means are produced for each mutant line is shown at the bottom of the graph. **C:** Shows gene expression in different tissues and the tissue type for each data point is shown to the left of the graph. The number of microarrays used to produces the means for each tissue type are shown to the right of the graph.

3.3.8 Conserved protein structure of MORC protein family

Having identified that there are seven MORC proteins in *Arabidopsis*, their protein sequence was compared against each other in order to identify differences in protein structure. All MORCs in *Arabidopsis* contain the core GHKL ATPase domain and a single coiled coil domain, with the exception being AtMORC3 that lacks part of the core GHKL domain (Figure 3.13). Five of the MORCs also have areas of low sequence complexity, these are areas with a bias towards a specific amino acid or acids and are known as low-complexity regions (LCRs) (Coletta et al. 2010). Their exact role is unclear but proteins with LCRs bind to a larger number of other proteins than other proteins in the same family that lack LCRs (Ekman et al. 2006). It is also suggested that LCR found at the ends of proteins are associated with proteins involved in transcription whereas proteins with LCRs found in the centre are involved in translation or stress response

(Coletta et al. 2010). The five MORC proteins with LRCs have LCRs at both terminal and central regions of the protein, except MORC3 which only has a central LCR (Figure 3.13). This would suggest, along with the truncated GHKL domain, that MORC3 has a more specialised function than the other MORCs, perhaps in either translation or stress response. The fact that the other MORCs have both terminal and central LCRs could suggest they are involved in transcription or stress response, although further work would be required to prove this. The lack of LCRs for MORC5 and MORC6 may suggest that these two proteins have a reduced number of binding partners than the other MORCs, although again this would need to be tested experimentally.

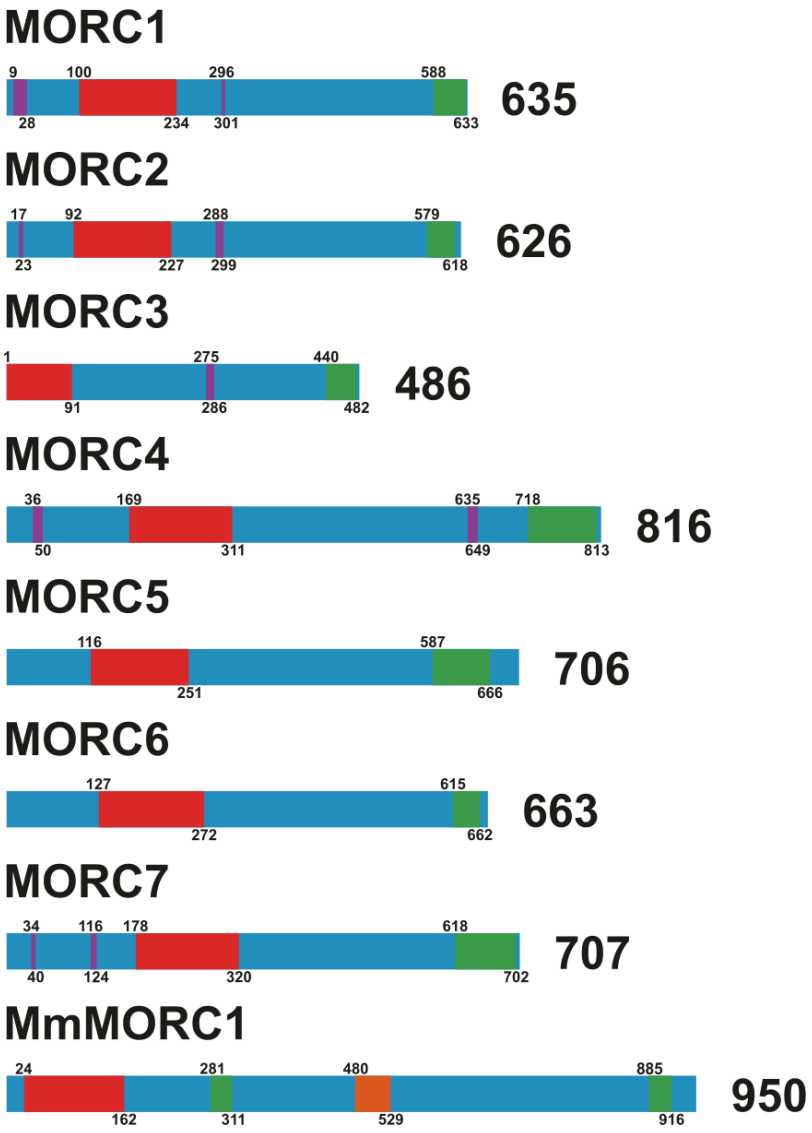


Figure 3.13: Conserved domain structure of the *Arabidopsis* MORC protein family

Diagrams showing the domain structure of the *Arabidopsis* MORC proteins and *Mus musculus* MORC1 protein. For each protein the total size is given on the right and the start and end amino acid for each domain is given above and below each schematic. The GHKL ATPase domain is shown in red; the coiled coil domain is shown in green; and the zinc finger domain is shown in orange. Low complexity regions (LCRs) are highlighted in purple. The MmMORC1 protein is representative of Metazoa MORCs while the *Arabidopsis* MORCs are representative of Plantae MORCs. For MmMORC1 there is LCR

not shown as it overlaps with the 2nd coiled coil domain. Its position is between the 904th and 917th amino acids.

The *Arabidopsis* MORC proteins were also compared against orthologs from Plantae and Metazoa species in order to identify differences in protein structure between species. The species chosen were: *Medicago truncatula*, *Populus trichocarpa, Vitis vinifera, Oryza sativa*, *Zea mays*, *Physcomitrella patens*, *Selaginella moellendorffii, Mus musculus*, *Homo Sapiens, Bos taurus*, *Xenopus tropicalis, Danio rerio, Caenorhabditis elegans*, *Nematostella vectensis* and *Myroides odoratimimus. M. odoratimimus* is a prokaryote with a MORC homolog and acts as an outgroup to the other MORCs. Seven Plantae species, excluding *Arabidopsis,* were chosen and include three dicots (*M. truncatula, P. trichocarpa* and *V. vinifera*), two monocots (*O. sativa* and *Z. mays*), *P. patens* as an example of a lower plant and *S. moellendorffii* which is a member of the oldest extant vascular plant lineage. Seven Metazoan species were also chosen and includes two invertebrate species (*C. elegans* and *N.* vectensis), three mammalian species (*H. Sapiens, M. musculus* and *B. taurus*), an amphibian species (*X. tropicalis*) and a fish species (*D. rerio*). These species were chosen in order to provide information on MORC protein structure in a broad range of Plantae and Metazoan lineages representing different evolutionary stages. The MORCs were identified by BLAST searches using the MORC1 sequence from *Arabidopsis* and selection of hits with more than 40% sequence homology to the MORC1 sequence. This method may not identify all MORCs within these species and a more exhaustive search would be required to identify any other MORCs, but should be sufficient for comparisons of MORC protein structure. The number of MORCs identified in each species varied, with the highest being *Physcomitrella patens* with nine MORCs and the lowest *M. musculus, D. rerio* and C. *elegans* with one MORC identified in each species (Table 3.8). However, generally Plantae species have more MORC proteins than Metazoan with an average of 5 MORCs per species compared to 2.1 MORCs per species in Metazoa. This could suggest greater diversification of function in Plantae than Metazoa. A total of fifty six

3. Identification of M1 and M9 as morc6 *mutants*

MORC proteins were identified, of which forty are from Plantae, fifteen are from Metazoa and one is from prokaryotes.

Both Plantae and Metazoan MORC proteins are characterised by the presence of the GHKL ATPase domain (Table 3.8 and Appendix 1.8 page 375). Most MORCs also contain at least one coiled coil domain, allowing interaction with other proteins. There are however exceptions, with *M. truncatula, P. patens, P. trichocarpa, B. taurus* and *N. vectensis* each having one MORC that lacks a coiled coil domain and all *S. moellendorffii* MORCs lack the coiled coil domain. The lack of coiled coil domains in *S. moellendorffii* is likely to be due to incomplete sequences as the length of the sequences were noticeably shorter compared to other species and the fact that both the higher plants and *P. patens,* which diverged from other plant species before *S. moellendorffii,* have MORC proteins with coiled coil domains. However, for the species that have a single MORC protein lacking a coiled coil domain the sequences appear to be full length and so would suggest that these proteins may have a different function from other MORCs. The fact that the Plantae (*P.patens*)and Metazoan (*N. vectensis*) species that were first to diverge from the other species in their respective kingdoms both have a MORC protein lacking a coiled coil domain could suggest that the last common ancestor to plant and animal species contained a MORC protein lacking this domain. This has subsequently been lost in most higher Plantae and Metazoan species but some species have retained it and so this MORC may have species specific functions. The bacterial MORC homolog in *M. odoratimimus* contains a coiled coil domain and therefore would suggest the prokaryotic ancestor to MORC proteins had a coiled coil domain meaning that the MORCs lacking this domain found in lower eukaryotic species may have lost the domain rather than the other MORCs gaining the coiled coil domain. However more prokaryotic MORC homologs would need to be assessed to determine if this indeed true.

3. Identification of M1 and M9 as morc6 *mutants*

Name	Organism	Kingdom	Protein length (aa)	GHKL ATPase Domain (aa)	Coiled Coil Domains (aa)	Zinc Finger Domain (aa)
AtMORC1	*Arabidopsis thaliana*	Plantae	635	100-234	688-633	None
AtMORC2	*Arabidopsis thaliana*	Plantae	626	92-227	579-618	None
AtMORC3	*Arabidopsis thaliana*	Plantae	486	001 -91	440-482	None
AtMORC4	*Arabidopsis thaliana*	Plantae	816	169-311	718-813	None
AtMORC5	*Arabidopsis thaliana*	Plantae	706	116-251	587-666	None
AtMORC6	*Arabidopsis thaliana*	Plantae	663	127-272	615-662	None
AtMORC7	*Arabidopsis thaliana*	Plantae	707	178-320	618-702	None
MtMORC1	*Medicago truncatula*	Plantae	943	346-463	889-943	None
MtMORC2	*Medicago truncatula*	Plantae	577	69-206	None	458-503
PtMORC1	*Populus trichocarpa*	Plantae	627	95-225	573-617	None
PtMORC2	*Populus trichocarpa*	Plantae	653	123-253	None	None
PtMORC3	*Populus trichocarpa*	Plantae	518	132-268	29-50	None
PtMORC4	*Populus trichocarpa*	Plantae	862	171-311	728-812	None
VvMORC1	*Vitis vinifera*	Plantae	653	99-231	601-644	None
VvMORC2	*Vitis vinifera*	Plantae	641	87-219	589-632	None
VvMORC3	*Vitis vinifera*	Plantae	709	143-281	648-704	None
VvMORC4	*Vitis vinifera*	Plantae	631	65-199	670-627	None
VvMORC5	*Vitis vinifera*	Plantae	830	170-310	735-822	None
OsMORC1	*Oryza sativa*	Plantae	629	64-190	575-618	None
OsMORC2	*Oryza sativa*	Plantae	682	154-289	622-682	None
OsMORC3	*Oryza sativa*	Plantae	819	270-408	737-792	None
OsMORC4	*Oryza sativa*	Plantae	788	139-281	679-787	None
OsMORC5	*Oryza sativa*	Plantae	715	143-280	658-709	None

3. Identification of M1 and M9 as morc6 *mutants*

Name	Organism	Kingdom	Protein length (aa)	GHKL ATPase Domain (aa)	Coiled Coil Domains (aa)	Zinc Finger Domain (aa)
ZmMORC1	*Zea Mays*	Plantae	771	134-276	667-771	None
ZmMORC2	*Zea Mays*	Plantae	798	154-297	688-797	None
PpMORC1	*Physcomitrella patens*	Plantae	790	158-292	689-756	None
PpMORC2	*Physcomitrella patens*	Plantae	800	157-292	688-795	None
PpMORC3	*Physcomitrella patens*	Plantae	752	112-245	641-748	None
PpMORC4	*Physcomitrella patens*	Plantae	848	196-318	737-766 and 794-846	None
PpMORC5	*Physcomitrella patens*	Plantae	666	158-302	None	564-609
PpMORC6	*Physcomitrella patens*	Plantae	801	156-288	688-796	None
PpMORC7	*Physcomitrella patens*	Plantae	785	159-277	670-781	None
PpMORC8	*Physcomitrella patens*	Plantae	386	14-114	337-363	None
PpMORC9	*Physcomitrella patens*	Plantae	1226	502-579	1109-1139 and 1162-1189	None
SmMORC1	*Selaginella moellendorffii*	Plantae	394*	25-160	None	None
SmMORC2	*Selaginella moellendorffii*	Plantae	417*	44-188	None	None
SmMORC3	*Selaginella moellendorffii*	Plantae	390	19-145	None	None
SmMORC4	*Selaginella moellendorffii*	Plantae	387*	53-188	None	None
SmMORC5	*Selaginella moellendorffii*	Plantae	477	31-155	None	419-468
SmMORC6	*Selaginella moellendorffii*	Plantae	364	59-157	None	None
MmMORC1	*Mus musculus*	Metazoa	950	24-162	281-311 and 885-916	480-529
HsMORC1	*Homo sapiens*	Metazoa	984	25-162	283-353 and 900-934	481-530
HsMORC2	*Homo sapiens*	Metazoa	1032	26-164	285-362, 547-583 and 963-1013	494-543
HsMORC3	*Homo sapiens*	Metazoa	939	26-159	686-738 and 767-873	408-453
HsMORC4	*Homo sapiens*	Metazoa	937	48-179	762-878	424-471
BtMORC1	*Bos taurus*	Metazoa	981	25-163	283-353 and 900-934	482-531

Name	Organism	Kingdom	Protein length (aa)	GHKL ATPase Domain (aa)	Coiled Coil Domains (aa)	Zinc Finger Domain (aa)
BtMORC2	*Bos taurus*	Metazoa	1038	26-164	285-322, 547-583, 746-763 and 974-1019	494-543
BtMORC3	*Bos taurus*	Metazoa	713	26-163	None	409-454
XtMORC1	*Xenopus tropicalis*	Metazoa	890*	69-206	840-863	530-577
XtMORC2	*Xenopus tropicalis*	Metazoa	943	26-164	284-324 and 547-583	494-543
XtMORC3	*Xenopus tropicalis*	Metazoa	902	26-158	674-759 and 781-837	406-451
DrMORC2	*Danio rerio*	Metazoa	1035	26-168	285-322, 548-584, 720-757 and 870-1016	494-543
CeMORC1	*Caenorhabditis elegans*	Metazoa	845	31-145	284-311	None
NvMORC1	*Nematostella vectensis*	Metazoa	436*	13-142	None	382-431
NvMORC2	*Nematostella vectensis*	Metazoa	689	28-160	293-322 and 556-594	503-552
MoMORCH	*Myroides odoratimimus*	Bacteria	668	20-161	448-480	None

Table 3.8: Comparison of domain structure of Plantae and Metazoa MORC proteins

Table listing MORC proteins in Plantae and Metazoa and also a MORC homolog (MORCH) from the bacteria species *Myroides odoratimimus* (Mo). The Plantae species are: *Arabidopsis thaliana* (At), *Medicago truncatula* (Mt), *Populus trichocarpa* (Pt), *Vitis vinifera* (Vv), *Oryza sativa* (Os), *Zea mays* (Zm), *Physcomitrella patens* (Pp) and *Selaginella moellendorffii* (Sm). The Metazoa species are: *Mus musculus* (Mm), *Homo Sapiens* (Hs), *Bos taurus* (Bt), *Xenopus tropicalis* (Xt), *Danio rerio* (Dr), *Caenorhabditis elegans* (Ce) and *Nematostella vectensis* (Nv). The kingdom that each species belongs to is shown in the table. For all species there are multiple MORCs in each species bar *M. Musculus*, *D. rerio* and *C. elegans* which have only one each. For each MORC the length of the protein is given in amino acids (aa). The position of the start and end of the GHKL ATPase domains, coiled coil domains and zinc finger domains are given as amino acid positions in the sequence. For those sequences with an asterisk next to the protein length, these sequences are incomplete as they lack the methionine start codon.

There are two clear differences between Plantae and Metazoa MORCs. One is that Plantae MORC proteins have a single coiled coil domain whereas Metazoa MORC proteins can have multiple coiled coil domains, which may suggest a divergence in function between Plantae and Metazoa MORCs (Table 3.8 and Figure 3.13). The other difference between Plantae and Metazoa is the presence of a CW zinc finger domain in Metazoa MORCs and absence in Plantae MORCs, with the exception *M. truncatula, P. patens* and *S. moellendorffii* which each have a single MORC that has a zinc finger motif. This raises the question of whether the Plantae MORCs have lost the zinc finger domain or the Metazoa and three Plantae MORCs have gained the zinc finger domain. The bacterial outgroup used in the alignment also lacks the zinc finger motif and would therefore suggest that the Metazoa and three Plantae MORCs have gained the zinc finger domain rather than a loss of the zinc finger domain in Plantae MORCs. This would require further analysis to prove as only one bacterial MORC homolog was used and more Plantae and Metazoa species would need to be tested.

There are four conserved sequence motifs found in the GHKL ATPase domain that are characteristic of this specific ATPase (Table 3.6) (Bergerat et al. 1997). The four motifs could be identified in all MORC proteins, except AtMORC3 and PpMORC8 that both lack motif I and in the case of AtMORC3 half of motif II (Figure 3.13 and Appendix 1.8 page 375). Both these proteins are truncated at the N-terminus but AtMORC3 is unlikely to be a evolutionary conserved form of PpMORC8 as it would appear to be the product of a recent gene duplication in the *Arabidopsis* genus. These two proteins may have a specialised function due to the large scale changes in the main functional domains or are in the process of becoming pseudogenes.

3.3.9 Evolution of the MORC protein family in eukaryotes

From the assessment of the domain structure of MORC proteins in both Plantae and Metazoa there appears to be a clear distinction between the two kingdoms

in terms of domain structure of MORCs. To assess the evolutionary relationship between the MORCs both between and within the two kingdoms a phylogenetic tree was produced using the fifty six protein sequences used in the protein sequence alignments. The tree produced is a phylogram so the length of the branches represents the percentage change in amino acid sequence from the last common ancestor with other sequences and can be a used as a measure of evolutionary separation, but does not indicate exact time in years since divergence. The prokaryote MORC sequence from *M. odoratimimus* was again used as an outgroup for the phylogram, as it will be the most evolutionary distant from the eukaryote sequences.

As expected the Plantae and Metazoa MORCs form separate clades, except MtMORC2, PpMORC6 and SmMORC5. Both MtMORC2 and PpMORC6 separate from all other MORC proteins and SmMORC5 shows closest similarity to Metazoa MORCs, rather than Plantae MORCs (Figure 3.14). These three MORC proteins contain a zinc finger motif domain, which is not seen in any other Plantae species and therefore would suggest that this protein has evolved separately from other Plantae MORCs. The fact that *S. moellendorffii* and *P. patens* are lower plant species may suggest that early Plantae species did have a MORC protein with a zinc finger domain, but this has been lost in higher plants, except *M. truncatula*. The reason for *M. truncatula* maintaining the zinc finger domain may be related to the fact that it is the only legume species tested in the phylogenetic tree and therefore the zinc finger containing MORC protein may be required for a function specific to legumes. Attempts were made to identify MORCs in the legumes *Gylcine max* and *Lotus japonicas* to test this hypothesis but did not identify any full length sequences that could be used to analyse the domain structure and construct a phylogenetic. Further testing is therefore required in order to determine if legumes require this zinc finger containing MORC and if higher order Plantae species have lost this zinc finger containing MORC.

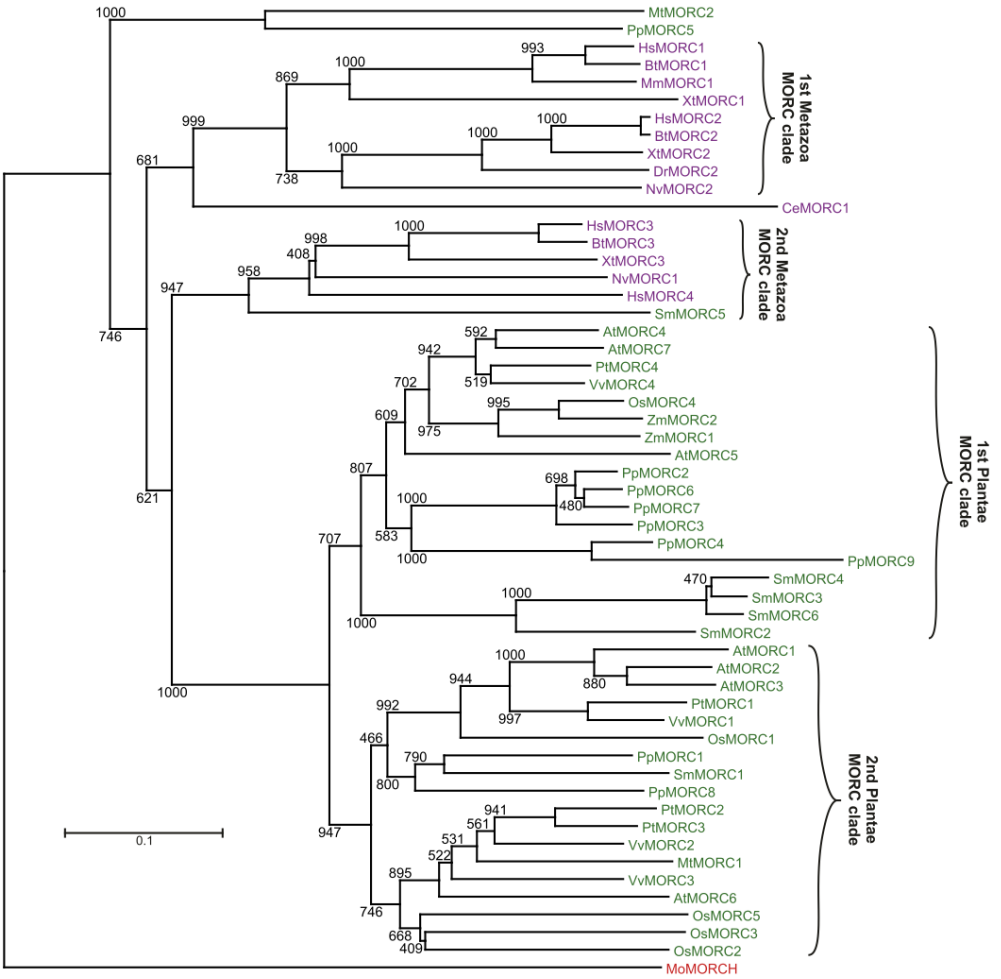


Figure 3.14: Phylogenetic tree of MORCs in the Plantae and Metazoa kingdoms

Phylogram of Plantae and Metazoa MORC proteins from the following species: *Myroides odoratimimus* (Mo). *Arabidopsis thaliana* (At), *Medicago truncatula* (Mt), *Populus trichocarpa* (Pt), *Vitis vinifera* (Vv), *Oryza sativa* (OZ), *Zea mays* (Zm), *Physcomitrella patens* (Pp), *Selaginella moellendorffii* (Sm), *Mus musculus* (Mm), *Homo Sapiens* (Hs), *Bos taurus* (Bt), *Xenopus tropicalis* (Xt), *Danio rerio* (Dr), *Caenorhabditis elegans* (Ce) and *Nematostella vectensis* (Nv). The prokaryote *M. odoratimimus.*is used as a outgroup for this tree and is shown at the bottom of the phylogram in red. Plantae species are shown in green while Metazoa species are shown in purple. A scale for the phylogram is shown at the bottom right and shows the length on the phylogram representing a 0.1 (10 %) change in the amino acid sequence. The brackets to the right of the phylogram denote MORC evolutionary clades.

Within the two kingdoms the other MORC proteins form distinct evolutionary clades. There are two clades of MORC protein in Metazoa, described hence forth as Metazoan (MZ) 1 and MZ 2. The two clades would suggest that there has been divergence in the function of MORC proteins into two distinct classes, although the nature of these functions is unknown due to the lack of data on MORC function. This separation into the two clades occurs in Anthozoa, Amphibia and Mammalia MORCs so would therefore suggest that the two types of MORC existed early in Metazoa evolution. The *C. elegans* MORC has not been assigned a clade in Figure 3.14 due to the large difference between it and other Metazoan MORCs, although it is most similar to the MZ 1 clade. The MZ 1 clade can be subdivided into two sub-clades, MZ 1A and MZ 1B, MZ 1A comprising of the mammalian and amphibian MORC1 orthologs and MZ 1B comprising of Metazoan MORC2 orthologs (Figure 3.14). Two of the MZ 1A sub-clade have been previously shown to be expressed specifically in the male germ cells whereas HsMORC2, which is part of the MZ 1B sub-clade is expressed in somatic tissue as well as germ cells (Nagase et al. 1998, Inoue et al. 1999, Wang et al. 2010). This means that the appearance of two sub-clades could be the results of the MORCs in the MZ 1A sub-clade having a specialised function in vertebrate male germ cells.

There are also two clades in Plantae, described as Plantae (PL) 1 and PL 2 from henceforth, and all plant species except *M. truncatula* have at least one MORC protein in each clade, suggesting that two clades must have existed early in the plant evolution (Figure 3.14). Interestingly both lower Plantae species have only one sequence in the PL2 clade whereas higher plants have a number of paralogs in this clade. This could suggest that there has been a divergence in function of MORCs within this clade in higher plants and it should be noted that MORC6 is within this clade so it is possible this divergence is in relation to RdDM. In the PL1 clade the two lower plant species separate both from each other and from the higher plant species to the point that they could form two sub-clades. This would again indicates a divergence in function between the higher plant and the lower

plant MORCs. Within the PL2 clade there are two sub-clades PL2A and PL2B of which PL2A consists of AtMORC1 homologs and PL2B consists of AtMORC6 homologs. This again could suggest functional divergence and it should be noted that AtMORC1 and AtMORC6 both have a function in RdDM, but AtMORC1 also has a function in pathogen response so it could be related to this difference in function (Kang et al. 2008, Kang et al. 2010, Moissiard et al. 2012). Within each clade different species have undergone expansion in the number of MORC proteins, for instance in *Arabidopsis* only AtMORC6 resides within the PL 2B clade, but both *Populus* and *Vitis* have two AtMORC6 orthologs within this sub-clade. Assuming that the phylogram is correct in when the last common ancestor for the MORCs are, it would also appear both Plantae clades are most similar to the MZ 2 clade of Metazoa, which may suggest that Plantae MORCs are more functionally similar to Metazoan MORCs in the 2nd clade.

Looking specifically at MORC evolution with respect to *Arabidopsis,* the phylogram shows that AtMORC6 is distinct from the six other MORCs and that the other MORCs fall into two groups, these groups comprising AtMORC4, 5 and 7 (PL 1 clade) and AtMORC1, 2 and 3 (PL 2A clade) (Figure 3.14). Comparisons between these two groups show that AtMORC1 and its homologs are more similar to each other than AtMORC4 and its homologs. The percentage difference between AtMORC1, 2 and 3 is small indicating a recent last common ancestor and this idea is supported by the fact that these MORCs are found together on chromosome four, hence are most likely due to recent gene duplication events (Figure 3.14). This similarity is supported by the fact that both AtMORC1 and AtMORC2 are involved in pathogen response in *Arabidopsis* and have similar domain structure (Figure 3.13) (Kang et al. 2008, Kang et al. 2010). However, AtMORC3 has not been reported to be involved in pathogen response and appears to have undergone a truncation of the GHKL ATPase domain, so may have an entirely different role to the other two PL 2A MORCs. The other group shows more divergence with AtMORC4 being the most distantly related of the three MORCs. This can be seen by the fact that it is nearly one hundred amino

acids longer than the other two proteins (Figure 3.13). There is no information on the functions of AtMORC4, 5 and 7 so it cannot be determined if the differences affect function.

3.4 Discussion

3.4.1 Identification of *MORC6* as the mutant RdDM gene

Premature stop codon mutations in the *MORC6* gene have been identified in lines M1, M9 and rmd6, suggesting that *MORC6* is responsible for the reactivation of the *GFP* transgene and this was confirmed by a complementation assay with a known *morc6* line. Two groups have identified *MORC6* as a component of the RdDM pathway in *Arabidopsis* (Lorković et al. 2012, Moissiard et al. 2012). The previously identified *morc6* mutant lines have been designated *atmorc6-1, atmorc6-2, atmorc6-3* and *atmorc6-4,* therefore the three mutants identified in this study have been designated *atmorc6-5* (M1), *atmorc6-6* (M9) and *atmorc6-7* (rmd6) (Table 3.9). The T-DNA insertion line (*atmorc6-3*) has been shown to be a full knockout of *MORC6* as no mRNA transcript can be detected for this gene, however both previous studies did not show whether *MORC6* is expressed as a truncated protein or whether the transcript is targeted by non-sense mediated decay in the three EMS *morc6* mutants (Lorković et al. 2012, Moissiard et al. 2012). This is also true for the three mutants identified in this study; however assuming the mutants express a truncated form of the MORC6 protein, there is predicted to be a difference in the severity of the mutations. In all EMS mutant lines the coiled coil domain of MORC6 is missing so the protein should not be able to interact with other proteins, such as DMS3, or dimerise with itself or other MORCs. This would be expected to inhibit any MORC6 function that requires MORC6 to bind to other proteins, but may not prevent any function that does not require interaction with other proteins. M1 lacks the GHKL ATPase domain and so could be considered a full *MORC6* knockout, whereas the other five EMS *morc6* lines contain a truncated form of the GHKL

domain. It has been shown that such a truncation reduces, but does not eliminate, ATPase activity in *atmorc6-4,* so it is possible that the other four MORC6 mutant proteins also have residual ATPase activity and may not be full knockouts of *MORC6* (Lorković et al. 2012). However, *atmorc6-4* produces a longer truncated protein than the other four *morc6* lines and so would be expected to be the weakest allele in terms of loss-of-function. To determine whether the other MORC proteins have ATPase activity would require biochemical testing, but an indication of their activity can be derived from comparing the phenotype of M1, a full *MORC6* knockout, against the phenotypes of M9 and the other *morc6* lines to determine if they are full knockouts.

Mutant	Mutation type	Exon affected	Codon affected	Identified by
atmorc6-1	EMS	9	267	Moissiard
atmorc6-2	EMS	9	293	Moissiard
atmorc6-3	T-DNA	4	NA	GABI-KAT line
atmorc6-4	EMS	13	439	Lorković
atmorc6-5	EMS	2	41	This study
atmorc6-6	EMS	9	267	This study
atmorc6-7	EMS	12	392	This study

Table 3.9: *morc6* mutant lines in *Arabidopsis*

Table showing all seven *morc6* lines in *Arabidopsis thaliana* and which study they were discovered by. The type of mutation is shown for each line, either an EMS base substitution mutation or T-DNA insertion mutation. The exon and codon where the mutation is located is given for each mutant, apart from the T-DNA insertion line where only the affected exon is given.

The *MORC6* gene is homologous to the mouse *Microrchidia 1* (*MORC1*) gene, hence was given the name *AtMORC6* or *MORC6,* but is also referred to as *defective in meristem silencing 11* (*dms11*) (Lorković et al. 2012, Moissiard et al. 2012). MORC proteins are characterised by the presence of a GHKL ATPase and coiled coil domain and are found in both eukaryotes and prokaryotes (Perry and Zhao 2003, Iyer et al. 2008). The GHKL ATPase domain is one of several ATPase domain families that catalyses the hydrolysis of ATP providing energy for protein function and this particular family has a unique ATP binding fold, known as the

Bergerat fold, which distinguishes the GHKL family from other ATPases (Bergerat et al. 1997). The Bergerat fold consists of four βsheets and three α helices that form a unique cone shaped structure and contains four highly conserved sequence motifs, known as the N box and G1-G3 boxes, that bind the ATP molecule (Bergerat et al. 1997, Mushegian et al. 1997, Ban and Yang 1998, Dutta and Inouye 2000). This domain was initially identified in four protein families: DNA gyrases, histidine kinases, mismatch repair enzymes and chaperone proteins; hence the domain is known as the Gyrase, HSP90, histidine kinase, MutL ATPase (GHKL) domain, HSP90 being a chaperone and MutL being a mismatch repair enzyme (Bergerat et al. 1997, Ban and Yang 1998, Tanaka et al. 1998, Ban et al. 1999, Bilwes et al. 1999, Dutta and Inouye 2000). The GHKL ATPase domain is thought to have evolved from a bacterial restriction enzyme system and to be mainly associated with enzymes that carry out chromatin modification (Iyer et al. 2008). The coiled coil domain allows for protein-protein interaction (Crick 1952, Mason and Arndt 2004).

Despite *MORC6* been previously identified as an RdDM component and therefore no longer a novel mutant in the pathway, the mutants identified in this study can be used to expand on the phenotype and role of MORC6 in RdDM, especially as the two previous studies give differing phenotypes for *morc6* mutants and as such have drawn different conclusions regarding the exact role of the MORC6 protein in the pathway. The study by Moissiard shows that there is large scale up regulation of RNA transcripts in the *morc6* mutants and that these are associated with loci that are silenced by the RdDM pathway (Moissiard et al. 2012). However, there are no global changes in either siRNA or DNA methylation levels in the *morc6* mutant, although there are localised increases and decreases in methylation. The *morc6* mutants do not have changes in the repressive histone mark H3K9me2, but do show loss of higher order chromatin structure and this is seen by decondensation of pericentromeric higher order chromatin structures known as chromocentres. It was also shown that MORC6 localised to the nuclear bodies adjacent to these chromocentres and so this study concluded that

MORC6 was likely to be involved in the formation of higher order chromatin structure in response to DNA methylation and repressive histone marks. It is thought that this higher order structure reinforces silencing by RdDM and that perturbation of MORC6 results in the loss of these higher order structures.

The study by Lorković also showed local changes in DNA methylation, although no global measure of DNA methylation was carried out (Lorković et al. 2012). DNA methylation is lost in the *GFP* reporter and at three intergenic (IGN) RdDM targets , but other targets such as the transposable elements AtSN1 and SoloLTR do not lose methylation. This loss in methylation at the *GFP* reporter transgene was also associated with a loss in H3K9me2 repressive histone marks and a gain in histone acetylation, which is a mark of euchromatin. This loss of DNA methylation and histone repressive marks was not associated with a decrease in siRNAs but there was a decrease in the transcripts produced by PolV. It was therefore concluded that MORC6 is involved in the production of the PolV transcript and the resulting methylation of the DNA. To support this idea it was shown that MORC6 interacts with, and has a similar expression pattern to DMS3, a hinge domain protein previously identified as a RdDM component (Kanno et al. 2008, Ausin et al. 2009, Lorković et al. 2012). DMS3 mutants also cause a reduction in DNA methylation and loss of PolV transcripts. The MORC6-DMS3 complex resembles a structural maintenance of chromosomes (SMC) protein, which have ATPase and hinge domains, but are usually encoded by one gene rather than two separate genes as is the case here (Hirano 2002, Hirano 2005). MORC6 contains two domains, a GHKL ATPase and coiled coil domain and it is through the coiled coil domain that it interacts with DMS3, but can also dimerise with itself to form homeodimers (Lorković et al. 2012). This suggests that MORC6 may function in the RdDM pathway in a complex with DMS3, but may also have other functions in RdDM that do not require DMS3. The MORC6-DMS3 complex is unusual as there is only one other known SMC with a GHKL domain, the mouse SmcHD1, which is involved in X chromosome inactivation through CG methylation (Blewitt et al. 2008, Lorković et al. 2012).

3.4.2 *Arabidopsis MORC* family

MORC6 is one of a family of seven MORC proteins found in *Arabidopsis thaliana* and a literature search reveals that the *MORC4, MORC5* and *MORC7* genes have not previously been characterised, but *MORC1, MORC2* and *MORC3* have been. The three *MORC* genes that have been previously characterised were first identified in a screen for mutants with reduced pathogen response to the Turnip Crinkle Virus (TCV) and it is because of this that *MORC1* is also named *Compromised recognition of TCV-1* (*CRT1*) and *MORC 2* and *MORC3* are named *Compromised recognition of TCV homolog 1* (*CRH1)* and *CRH2* respectively (Kang et al. 2008, Kang et al. 2010). The initial study identified *MORC1* as a pathogen response gene but did not demonstrate a role for the close homologs, *MORC2* and *MORC3* (Kang et al. 2008). A later study showed that both *MORC1* and *MORC2* are involved in general pathogen response and are not specific to TCV infection, but was unable to report a function for *MORC3* as homozygous *morc3* mutants are embryonically lethal. This suggests that MORC3 has a vital role in the embryo (Kang et al. 2010). Both MORC1 and MORC2 act early on in pathogen response and interact with plant resistance (R) proteins and mutation in these two genes result in a failure to control viral replication or systemic spread in TCV infections.

MORC1 has also been shown to be an RdDM component, thus has at least two cellular functions. This dual role is unusual as it is uncommon for proteins involved in chromatin modification to also have another function in a cytoplasmic pathway, although the RdDM components PolV and AGO4 have also been previously reported to function in pathogen response (Agorio and Vera 2007, López et al. 2011, Moissiard et al. 2012). It is not known whether PolV, AGO4 and MORC1 interact with each other in pathogen response. This dual role for MORC1 raises the question of how it is involved in two such different pathways. The presence of both the coiled coil domains and LCRs in MORC1 may allow it to interact with a large number of different proteins and so its function

could be as a scaffold protein for a number of diverse pathways. However, this would require testing through interaction studies and biochemical analysis to see if MORC1 has any specific activity in the two pathways or acts as a general scaffold protein. In terms of its functions in RdDM, *morc1* mutants are reported to have a similar phenotype to *morc6* mutants and double mutants of the two genes do not result in a significant increase in the severity of the phenotype, therefore indicating that MORC1 and MORC6 act redundantly in RdDM (Moissiard et al. 2012). It is possible that other proteins in this family are also involved in RdDM and this has been suggested as a reason why loss of DNA methylation is loci specific in *morc6* (Lorković et al. 2012). Of the five MORCs that have no known RdDM function, MORC2 is the most likely to have a role in RdDM due to its close homology to MORC1 and the fact that both it and MORC1 have previously been shown to act redundantly in pathogen response (Kang et al. 2008, Kang et al. 2010). This hypothesis would need to be tested experimentally through DNA methylation analysis of mutants in the five untested *MORC* genes.

From bioinformatics analysis of protein sequence and domain structure there has clearly been divergence in the roles of the MORC proteins such that not all MORCs may be involved in RdDM. Two studies have failed to isolate homozygous mutants in *morc3*, which is suggestive that mutations in this gene are lethal, at least during embryogenesis or seed germination (Kang et al. 2008, Kang et al. 2010). This lethality of *morc3* mutants would suggest a divergence in its role from that of MORC1, MORC2 and MORC6, which when mutated are not lethal, and this is supported by the fact that its GHKL ATPase domain is significantly altered compared to the other MORCs. MORC3 lacks the first and half of the second highly conserved motifs of GHKL ATPase domains and the domain itself is smaller than any other Plantae or Metazoa MORC. It is not known if any other *Arabidopsis* MORC has a similar role to MORC3 as although *morc1*, *morc2* and *morc6* mutants are not lethal, *morc4, morc5* and *morc7* mutants have yet to be reported on. None of the other six MORC proteins have such a large scale alteration to the GHKL domain as MORC3 but there are sequence differences

between the MORCs that may alter 3D structure and function. Experimental testing would be required to determine the effects of these sequence differences.

The other main difference in terms of protein sequence between the MORC proteins is that both MORC5 and MORC6 lack the low complexity regions (LCR) found in the other MORC proteins. These regions are associated with proteins that act as hubs for protein binding and LCRs are therefore thought to increase the number of proteins that can bind (Ekman et al. 2006). This would therefore suggest that MORC1, MORC2, MORC3, MORC4 and MORC7 have an increased number of protein binding partners than MORC5 and MORC6 and therefore act as hub proteins, which could translate into these proteins being involved in a greater number of cellular pathways. MORC5 and MORC6 would therefore be more specific in their binding partners and hence be involved in a more limited number of pathways. This difference could explain why MORC1 has been shown to act in pathogen response and RdDM, which require MORC1 to bind to an entirely different set of proteins. However, the significance of the LCR differences between the MORC proteins would need to be verified by interaction studies for all MORCs.

Gene expression data on the MORC proteins also shows differences between the MORC proteins. Microarray data for MORC5 expression is low both throughout development and in different tissue types, to the extent that it could be considered background noise from the microarray rather than expression of the *MORC5* gene. This could suggest that *MORC5* is actually a pseudogene or in the process of becoming one, but this would require further analysis of its expression pattern. The expression pattern of *MORC6* is also of interest as it shows *MORC6* expression increases during flowering and is highest in pollen and therefore may have a similar function to that of the *Mus musculus* MORC1 protein, which is required for spermatogenesis, although its exact function in this process is unknown (Watson et al. 1998, Inoue et al. 1999). However, unlike *mmmorc1*

mutants where male mice are sterile, *atmorc6* mutants are not sterile, which would suggest that there is either redundancy with other MORC proteins or that AtMORC6's function in pollen is different to MmMORC1. There is also high expression in the female floral organs so MORC6 may instead be involved in the re-establishment of heterochromatin in the embryo and pollen and this could be linked to the increase in RdDM associated siRNAs during embryogenesis (Mosher et al. 2009, Slotkin et al. 2009).

3.4.3 MORC function in other species

The first *MORC* or *Microrchidia* gene to be identified and characterised was discovered in a forward genetic screen for mutants with defective spermatogenesis in mice (*Mus musculus*) (Watson et al. 1998). Homozygous male mice had a reduction in size of the testes, hence the gene name, but no phenotype was observed in homozygous female mice, indicating that the gene had a male specific function. Later studies showed this to be the case as *MORC1* is only expressed in male germ cells (Inoue et al. 1999). It was shown that *morc1* mutants caused an arrest of spermatogenesis at prophase I during either the leptotene or zygotene stages and that MORC localises to the nucleus in germ cells, thus suggesting that MORC1 has a role in chromatin organisation during meiosis. One of the human homologs of MORC also localises to the male germ cells, but other human homologs are found in somatic cells.

MORC proteins have been studied in greatest detail in humans in which there are four *MORC* genes (Inoue et al. 1999, Liggins et al. 2007, Takahashi et al. 2007, Mimura et al. 2010, Shao et al. 2010, Wang et al. 2010). *MORC1* is expressed in human male germ cells so is thought to have a similar function to the mouse *MORC1* gene, but the other three human *MORC* genes are somatically expressed (Inoue et al. 1999), *MORC2* is highly expressed in the testis, ovary and brain tissue and the protein is found predominantly in the nucleus but is also found in the cytoplasm at lower levels (Nagase et al. 1998, Wang et al. 2010). The MORC2

protein is a tumour suppressor gene and has been shown to repress gene expression, including the *Carbonic Anhydrase IX* gene which promotes growth and survival in tumour cells (Shao et al. 2010, Wang et al. 2010). Only the coiled coil domain and a proline rich domain, which mediates DNA binding, are required for this transcriptional repressor activity, suggesting that MORC2 may have other activities that require the GHKL ATPase and zinc finger domains. MORC4 has also been implicated in cancer but its exact function is unknown (Liggins et al. 2007). The other human MORC protein, MORC3, is required for p53 activity and localisation into the promyelocytic leukemia (PML) nuclear bodies (NB) during cell senescence and the mutation of *MORC3* results in reduced p53 function (Takahashi et al. 2007). This activity of MORC3 requires the formation of MORC3 homodimers through the coiled coil domains of the two MORC3 proteins and it is thought that the dimer acts as a clamp (Mimura et al. 2010). Through conformational changes to the dimer, caused by ATP binding and hydrolysis, the dimer can close upon ATP binding and open upon ATP hydrolysis; and this closing and opening of the dimer results in the association and disassociation of MORC3 with the PML NBs respectively.

In terms of MORC proteins in other species having a role in RNA silencing and RNA interference, as is seen in *Arabidopsis,* previous studies have shown that MORC proteins are active predominantly in the nucleus and are involved in chromatin organisation or gene repression (Watson et al. 1998, Takahashi et al. 2007, Iyer et al. 2008, Mimura et al. 2010, Shao et al. 2010, Wang et al. 2010, Moissiard et al. 2012). However although these activities are similar to that of the *Arabidopsis MORC1* and *MORC6* genes, no other study has shown a direct link between a MORC protein and RNA silencing or RNA. One of the studies that identified MORC6 as a RdDM component did show that in *Caenorhabditis elegans,* mutation of its sole *MORC* gene resulted in a loss of silencing of a *GFP* reporter and that this loss is phenotypically similar to the mutant in the RNA intereference component *rde4* (Tabara et al. 2002, Moissiard et al. 2012). However, no direct interaction was shown between the MORC protein and *C.*

elegans RNA interference machinery. Further study of MORC function in relation to chromatin modification and RNA silencing in other species is therefore required to demonstrate whether MORC function in *Arabidopsis* is conserved in other species.

3.4.4 Evolution of Plantae and Metazoa MORCs

In *Arabidopsis* there has been a number of recent gene duplication events, seen in the close relationship between *MORC1, MORC2 and MORC3* in terms of genome location and sequence similarity. These *MORC*s have undergone diversification of function, particularly MORC3, which has a truncated GHKL ATPase domain and homozygous mutations are lethal. However, both MORC1 and MORC2 function in pathogen response, suggesting that there has not been as much divergence in terms of function between these two proteins as has occurred for MORC3. No other Plantae species used in this study had a gene duplication of *MORC1* homologs, therefore suggesting that this duplication is recent and may have occurred only in *Arabidopsis thaliana,* although this would require a search for *MORC1, MORC2* and *MORC3* homologs in other species within the *Arabidopsis* genus. *AtMORC6* has not undergone any gene duplications, although gene duplication events have occurred in *MORC6* homologs in both *Populus* and *Vitis.* The final clade, *MORC4* homologs, has undergone gene duplication events as there are three *Arabidopsis MORC* genes within this clade and they are *MORC4, MORC5* and *MORC7.* These three *MORC* genes are not as closely linked as the *MORC1* homologs so this duplication may have occurred earlier in evolution than the duplication of *MORC1, MORC2* and *MORC3.* Exactly when is not known, but it appears to have occurred after the last common ancestor with *Populus* and *Vitis.*

MORC proteins are found in both Plantae and Metazoa and have orthologs in prokaryotes (Iyer et al. 2008). The main difference between MORCs in the two kingdoms is the presence of a zinc finger domain in Metazoan MORC proteins

but the lack of such a domain in the majority of Plantae MORCs. This could suggest divergence in function, although it is possible that Plantae MORCs interact with another protein that provides the zinc finger motif function. It is likely that this difference arose after the loss of the zinc finger motif in Plantae as zinc finger domains are found in lower plant species. However, as the prokaryotic homolog also lacks a zinc finger domain further investigation is required to support this conclusion. Plantae MORCs also only have one coiled coil domain whereas Metazoan MORCs can have one or more and this would result in either increased binding through these domains or alteration in protein folding due to interactions between the domains. This again may suggest a change in function, but this would require testing experimentally. Amongst the Plantae species there are two exceptions one in *M. truncatula* and the other in *P. patens* that contain a zinc finger protein but no coiled coil domain and could therefore have an entirely different function to other Plantae MORCs. This MORC may have been required in lower plants but not higher plants except *M. truncatula* which is the only legume analysed so may have retained this MORC for a legume specific function. These two MORCs do not fall within either a Plantae or Metazoan MORC clade and would suggest a very different evolutionary history to other MORCs, thus would be an interesting prospect for further investigation.

In both Plantae and Metazoa there has been an expansion in MORC proteins with each kingdom containing two clades of MORC and in both cases one of these clades can be sub-divided into two sub-clades. One of these sub-clades is associated with higher plant and animal species respectively and therefore may suggest the emergence of a third clade in higher organisms. Since both Plantae and Metazoa have two clades it is possible that the common ancestor of these two kingdoms had two MORC proteins. Although the phylogram shown in this study does not support this in terms of similarity between the MORC clades it should be noted that the phylogram represents differences between MORC sequence rather than time since divergence, hence it is possible that Metazoan

MORC proteins have undergone more rapid evolution than their Plantae counterparts.

3.5 Acknowledgments

The identification and optimisation of CAPS markers that have a difference between C24 and Landsberg *erecta*, except M59, was completed by A. Eamens, who also produced the EMS mutant lines and 142S transgenic line used in this research. The Illumina sequencing and sample preparation was carried out by Exeter sequencing service and Zhesi He, of the Ian Graham group at the University of York, analysed the sequence data using SHOREmap. The imaging of the plants under UV light was carried out with the assistance of Phil Roberts. Sarah Wetherill carried out the characterisation of the rmd6 mutant which led to its identification as a *MORC6* mutant. James Llyod from the University of Leeds provided assistance with the identification of full length *P. patens* sequences.

4. Silencing phenotype of the *35Sp:GFP* transgene system in the *MORC6* mutants

4.1 Introduction

4.1.1 Dual transgene silencing system

The screen of the EMS mutant library for novel RdDM mutants identified M1 and M9 as alleles in *morc6*. The EMS mutant library used a dual transgene system to identify mutants in the RdDM pathway, consisting of a *35S* promoter driven *GFP* gene (*35Sp:GFP*) and a nopaline synthase (*NOS*) promoter driven inverted repeat of part of the *35S* promoter (*NOSp:35S IR*). How this system works has been described in detail in the introduction of chapter three (page 124); but briefly, the *35S IR* transgene, through self-complementarityproduces a dsRNA transcript that is recognised by the RdDM pathway which, due to sequence homology, will target and methylate the *35S* promoter of the *GFP* transgene and therefore silence the transgene. RdDM mutants lose this silencing of the *GFP* transgene and so are fluorescent. Having used this fluorescence to first identify the M1 and M9 mutants, their effect on the silencing of the *GFP* transgene will be studied in detail in this chapter. Before discussing this I will briefly cover other transgene silencing systems and the advantages and disadvantages for each system.

This *GFP* dual transgene system, used in this study, has several advantages. Firstly identification of RdDM mutants is straightforward as putative mutants can be screened for GFP fluorescence. The system can also detect RdDM mutants with a weak phenotype as partial release of GFP can still be detected in this system. GFP fluorescence can also be used to assess the effect of the RdDM mutant on different tissue types. The main disadvantage of this system is that both transgenes are required for the system to work. It is a lengthy process transforming both transgenes into a plant line and demonstrating that they are both functional. This makes it prohibitive to test RdDM mutants, which currently lack the transgenes system.

4.1.2 Other transgene silencing systems

As well as the transgene system used in this study there are several other systems that have been used to identify RdDM mutants each with their own strengths and weaknesses. In this section I will briefly go over how each system works and the advantages and disadvantages for each system. There are several systems that are similar to the one used in this work. In each case a reporter transgene is silenced by an inverted repeat of the promoter used to drive the reporter gene. In one case the seed specific *α'* promoter was used to drive a *GFP* reporter transgene while the inverted repeat of the *α'* promoter is driven by the *NOS* promoter (Kanno et al. 2004). As the *α'* promoter is seed specific it allows for detection of RdDM mutants in seeds speeding up the process of detecting mutants. However as the *GFP* transgene is only expressed in seeds it cannot be used to look at tissue specific effects of RdDM mutants. The other similar transgene system uses a *NOS* promoter driven kanamycin resistance gene (*NPTII*) transgene as a reporter and a *35S* promoter driven inverted repeat of the *NOSp* transgene as a silencer (Aufsatz et al. 2002b). RdDM mutants are then selected by resistance to kanamycin from release of silencing of the *NPTII* transgene. This makes identification of RdDM mutants straightforward as only RdDM mutants will survive on kanamycin plates. It does however mean that RdDM components with a small effect on silencing may not be detected as their partial reduction in silencing of the *NPTII* gene may not be enough to confer resistance to kanamycin. Tissue specific effects can also not be detected in this system.

One system utilises the silencing of the *FWA* flowering time gene (Cao and Jacobsen 2002b, Chan et al. 2004). In WT plants the expression of the *FWA* gene is silenced by methylation of two SINE related direct repeats at the 5' end of the *FWA* gene (Soppe et al. 2000, Yuki et al. 2007). Methylation of *FWA* is normally stably transmitted from parent to progeny. Loss of methylation at the repeats causes *FWA* expression and leads to a delay in flowering time. If a *FWA* transgene is transformed into WT plants it is methylated by RdDM and does not

affect flowering time. In RdDM mutants this *de novo* methylation of the transgenic *FWA* does not occur and the transgenic *FWA* is expressed, resulting in late flowering. The *FWA* gene can therefore be introduced into mutant lines, both known and unknown to test if they are involved in the RdDM pathway (Cao and Jacobsen 2002b, Chan et al. 2004). As only a single transgene is required this system is effective at testing putative RdDM mutants, which currently lack a transgene system. The delay in flowering time also provides an indication of RdDM mutant's effect on the pathway, whether it has a large or small effect on methylation. The disadvantage with this system is that identification of mutants takes longer than other systems as the screen is based upon flowering time. Also as the system is specific to flowering it cannot be used to identify tissue specific effects.

In the *RD29Ap:LUC* transgene system rescue of bioluminescence is used to identify RdDM mutants. *LUC* is a firefly luciferase gene which results in bioluminescence when expressed. The *RD29A* promoter responds to either cold, salt, drought or abscisic acid stress (Yamaguchi-Shinozaki and Shinozaki 1994). The WT transgenic line of *RD29Ap:LUC* stably expresses *LUC* in stress conditions, but in the *ros1* mutant the transgene is silenced by hypermethylation (Gong et al. 2002). *ROS1* is a DNA glycosylase involved in demethylation of DNA. So in *ros1* mutants DNA methylation is not removed and *de novo* hypermethylation can occur. The hypermethylation of *RD29Ap:LUC* can be reversed by mutating RdDM components (He et al. 2009a). Putative RdDM mutants can thus be identified by bioluminescence. The main advantages of this system are that mutants can easily be identified by bioluminescence and that RdDM mutants with a partial effect can be identified. The system also allows for identification of tissue specific effects. The disadvantage with this system is that any RdDM mutants identified will be in a *ros1* mutant background. As the demethylation pathway is inactive in these plants this could conceivably influence the phenotype the RdDM mutant displays. This system cannot be easily transferred to existing mutants as the system requires the plants to be homozygous for a *ros1* mutation.

The *SDC:GFP* transgene has recently been developed to identify mutants in both the RdDM and CMT3 pathways (Moissiard et al. 2012). In most plant tissues the *SDC* gene is silenced by both DRM2 and CMT3, which act redundantly (Zhang et al. 2006). Methylation is targeted to seven tandem repeats in the *SDC* promoter and these repeats are present in the promoter of the transgene (Henderson and Jacobsen 2008, Moissiard et al. 2012) This means that in WT plants the *SDC* promoter is methylated and *GFP* is not expressed but silencing can be released in *drm2 cmt3* double mutants, although not in single mutants (Moissiard et al. 2012). RdDM mutants can be identified by using the *SDC:GFP* transgene in a *cmt3* background or for mutants that inhibit both CMT3 methylation and RdDM the *SDC:GFP* transgene can instead be used in a WT background. The advantages of this system are that mutants can easily be identified by GFP fluorescence and the level of fluorescence can give an indication of the mutant's effect on the RdDM pathway. The disadvantage of this system is that to transfer the *SDC:GFP* transgene into a RdDM mutant line currently the transgene would also require the creation of a double mutant of the RdDM mutant and *cmt3* in order for the system to work.

The final transgene system that will be discussed is the JAP transgene system. The JAP line contains a single transgene. It comprises an inverted repeat of the *PHYTOENE DESATURASE (PDS)* gene driven by the phloem specific *SUC2* promoter (Smith et al. 2007). The inverted repeat triggers PTGS silencing of endogenous *PDS* gene resulting in photo bleaching of cells where *PDS* is silenced. The silencing of *PDS* is initially restricted to the phloem, but transmission of siRNAs out of the phloem trigger silencing in surrounding mesophyll tissue. Although strictly a PTGS process, RdDM is involved in both producing dsRNA from the transgene and also regulation of the transgene. RdDM proteins involved in the production of dsRNA, such as NPRD1, RDR2 and CLSY1, are required for *PDS* PTGS silencing and mutants in these proteins do not silence the *PDS* gene. As well as this role in dsRNA production, RdDM also acts as a regulator of PTGS from this transgene. Mutants in *dcl3* and *ago4* result in PTGS of *PDS*

spreading across the entire leaves. This indicates that RdDM triggers DNA methylation of the transgene, which reduces the extent of PTGS silencing. The main advantages of this system are that the phenotype of RdDM mutants are easily identifiable and the phenotype can inform as to whether the mutant is involved in dsRNA production or siRNA production and DNA methylation. The disadvantage of this system is that some RdDM mutants do not affect *PDS* silencing, such as *DRD1* and *NRPD2*. The system may therefore not identify all RdDM mutants.

4.1.3 Aims of chapter

The aim of this chapter is to characterise the silencing phenotype of the 35Sp:GFP transgene in the M1 and M9 mutants. This will be done by studying:

A) Visible phenotype. Visualise the GFP silencing pattern in different plant tissues.

B) The molecular phenotype. Ascertain siRNA levels and DNA methylation status of the 35S promoter.

4.2 Results

4.2.1 Delay in the silencing of *GFP* in M1 and M9

In the two alleles of *MORC6*, M1 and M9, expression of the *GFP* transgene is enabled. Although it was observed that in seedlings *GFP* expression in M1 and M9 is comparable to 142, in adult plants M1 and M9 have lower levels of GFP than 142. This indicates that both M1 and M9 still have some capacity to silence the GFP transgene. To assess this plants from lines 142, 142S, M1 and M9 were observed and imaged using a fluorescence microscope at weekly intervals from germination. At 7 days post germination (dpg) GFP fluorescence in M1 and M9 is similar to that of 142, whereas for the parental line 142S, plants appear fully

silenced at 7 dpg (Figure 4.1 A-D). At 14 dpg GFP fluorescence in both M1 and M9 is reduced and is at an intermediate level between that of 142 and 142S (Figure 4.1 E-H). The intermediate level of fluorescence at 14 dpg is comparable to that previously seen in adult plants (Figure 3.4). This indicates that both M1 and M9 undergo reduced levels of silencing compared to the parental 142S and that establishment of silencing is delayed in both mutants.

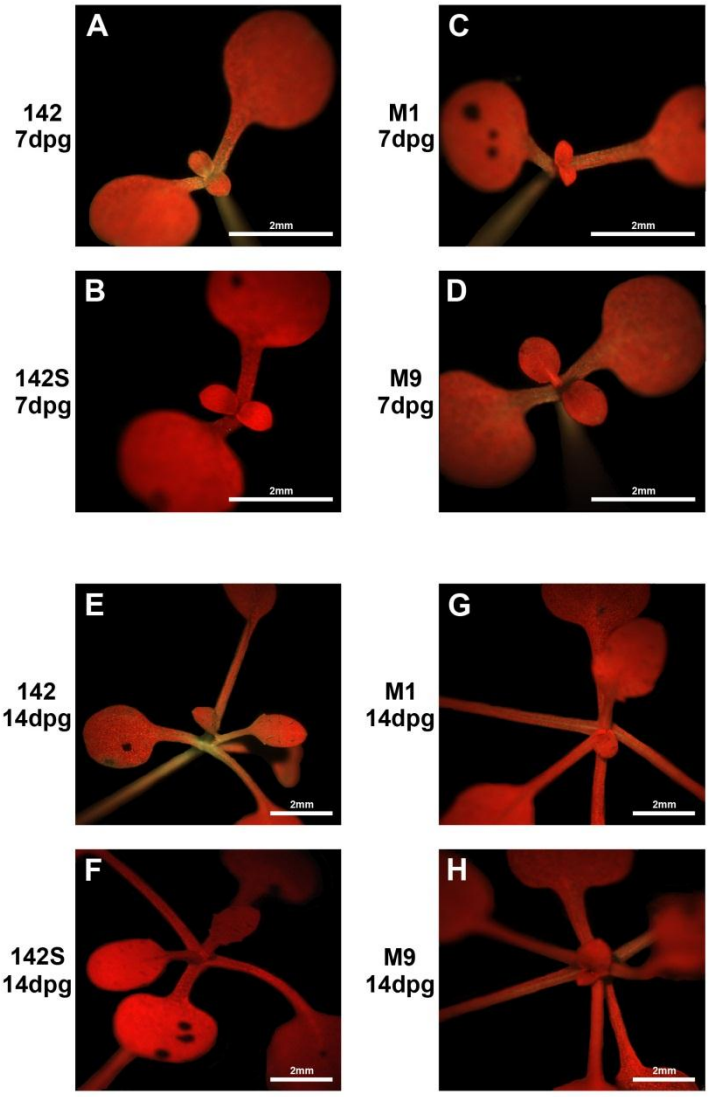


Figure 4.1: Change in *GFP* silencing between 7 dpg and 21 dpg in M1 and M9

Seedlings were grown on soil in long day growth rooms and were imaged using a fluorescence microscope. Images of seedlings were taken of lines 142, 142S, M1 and M9 at 7 dpg and 14 dpg. A scale bar is shown at the bottom of each image.

In order to confirm this delay in silencing establishment a northern blot for *GFP* mRNA was completed. The aim of this analysis was to see if the onset of visible silencing that occurs between 7 dpg and 14 dpg correlates with a reduction in *GFP* mRNA in M1 and M9. RNA samples from aerial tissue of plants that were pre visual silencing onset (7 dpg) and post visual silencing onset (21 dpg) were taken from lines 142, 142S, M1, M9, *rdr2-1* and *rmd3*. *rdr2-1* lacks the *GFP* transgene so acts as a negative control for *GFP* mRNA detection (Xie et al. 2004). *rmd3* is a mutant in the large subunit of PolIV (*nrpd1*) identified in the previous screen of the EMS mutant library used in this study (Eamens et al. 2008). At 7 dpg *GFP* mRNA levels in lines M1, M9 and *rmd3* are similar to 142 (Figure 4.2). *GFP* mRNA in 142S is present at 7 dpg but at very low levels, compared with 142 or the three RdDM mutants. There is a marked decrease in *GFP* mRNA in all three RdDM mutant lines at 21 dpg. This provides molecular evidence of a delay in silencing onset in M1 and M9. The delay in silencing in *rmd3* correlates with previous research, that silencing is delayed in *nrpd1* mutants (Herr et al. 2005). There is also a decrease in *GFP* mRNA in 142S at 21 dpg compared to 7 dpg. However, the decrease is small and could be due to the proportion of tissue types in each sample. At 7 dpg the majority of tissue collected will be from cotelydon leaves, whereas at 21 dpg a larger proportion will be rosette leaves. It has been observed that silencing in 142S is comparably weaker in the cotelydon leaves than rosette leaves. This would explain the decrease in *GFP* mRNA between 7 dpg and 21 dpg. The difference in the proportions of tissue types may also affect the M1 and M9 *GFP* mRNA levels, however as the change in 142S is small it could be argued that the effect of tissue types will be minor in M1 and M9.

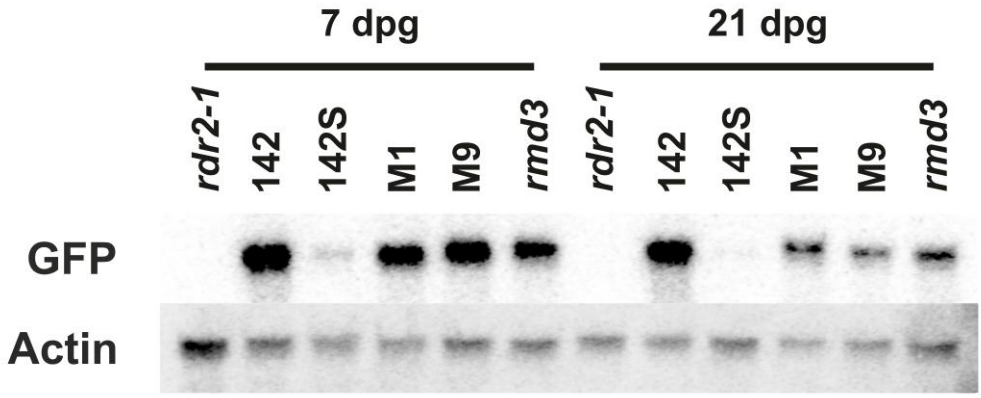


Figure 4.2: Changes in the expression of *GFP* mRNA between 7dpg and 21 dpg

Image of a northern blot probed with radio-labelled probes for *GFP* and *ACTIN 2*. 7 dpg and 21 dpg RNA samples of lines *rdr2-1,* 142, 142S, M1, M9 and *rmd3* were used. Approximately 3 µg of total RNA was run for each sample. After being probed for *GFP* mRNA the blot was stripped and re-probed for *ACTIN 2* mRNA as a loading control. *rdr2-1* does not possess the GFP transgene so acts as a negative control. 142 has ubiquitous GFP expression so acts as the positive control.

4.2.2 Mosaic silencing pattern of *GFP* in the epidermal layers of leaves

In order to further characterise the silencing phenotype and establish the exact time when silencing occurs in M1 and M9 a time course of GFP fluorescence was completed. Silencing was assessed daily on whole leaf samples from 7 dpg onwards, until silencing was established in the mutants, using confocal microscopy. The time course was repeated three times. GFP fluorescence was assessed in both the upper and lower epidermal layers of plant leaves. Only these two layers were used in the time course due to the difficulty in imaging the mesophyll layers in whole leaves. Due to absorption and refraction of the laser beam and light emitted by GFP the confocal microscope cannot image further than the first layer of mesophyll cells. In order to see if silencing in the epidermal layers varied between leaves of different ages, all odd numbered adult rosette

leaves (Figure 4.3) were imaged. Leaves were imaged from the parental line 142S and the two mutants M1 and M9. Leaves from line 142 were also imaged in order to show levels of GFP fluorescence without silencing. All plant lines were grown under a 16 hour long day growth regi`me. For time points where the hours post dawn are given, dawn is considered to be when lights turn on, which is at 5am in the growth rooms used for the time course. Before imaging under the confocal microscope leaves were first stained with propidium iodide. This stains plant cell walls, so any silenced cells which lack GFP, can be identified. All images taken of the epidermal layers were from a single focal plane as the curvature of the epidermal layers was low enough to allow visualisation of large areas of the leaf surfaces.

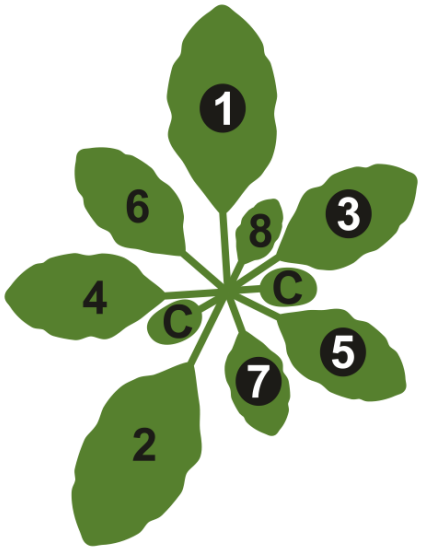


Figure 4.3: Choice of leaves for confocal time course

Diagram of *Arabidopsis* rosette from above. Each leaf is numbered according to when it appeared. The leaves marked with C are cotyledon leaves. Leaves where numbers are in white and are in a black circle are those used in the confocal time course.

For the time points before the onset of silencing in M1 and M9 only the 10 dpg time point is shown as it is representative of the observed pre silencing phenotype. Before the onset of silencing GFP fluorescence in both M1 and M9 is

comparable to 142 (Figure 4.4 A, B and E to H). In these lines GFP and propidium iodide fluorescence co-localise. GFP is targeted to the endoplasmic reticulum while propidium iodide is a cell wall marker (Haseloff et al. 1997, Moreno et al. 2006). The reason they co-localise is that the large vacuoles in pavement cells push the cytoplasm to the plasma membranes. Due to the resolution of the images this results in the apparent co-localisation. In 142S there is silencing of the GFP transgene in the majority of cells as seen by the lack of GFP fluorescence (Figure 4.4 C and D).

The pavement cells in the 142 upper epidermis appear larger than those in the other lines (Figure 4.4 B). This is due to where the image was taken on the leaf. As leaves age and reach the second morphogenesis stage of leaf development, cell replication halts and cells undergo cell expansion to increase the size of the leaf (Andriankaja et al. 2012). Cell expansion does not occur at the same time in all cells but instead initially occurs at the leaf tip and then moves basipetally (Donnelly et al. 1999, Kazama et al. 2010). This occurs before the 10 dpg time point and will result in different sizes of pavement cells throughout the leaf. Thus the cells in the 142 upper epidermis image are likely to be nearer the leaf tip than those in the other images. In several images there are circles of fluorescence in the propidium iodide channel (Figure 4.4 A, B and H). These circles are chloroplasts from the mesophyll layer below the epidermal layers being imaged. Due to the setup of the confocal microscope both propidium iodide and chlorophyll fluorescence will be detected in the same channel. The reason why fluorescence from the mesophyll layer is detected is due to the size of the pin hole. In order to achieve excitation and emission by both propidium iodide and GFP in the epidermal layer, the size of the pin hole had to be large. Increasing the pin hole results in the excitation of fluorophores, such as chlorophyll, above and below the focal plane.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

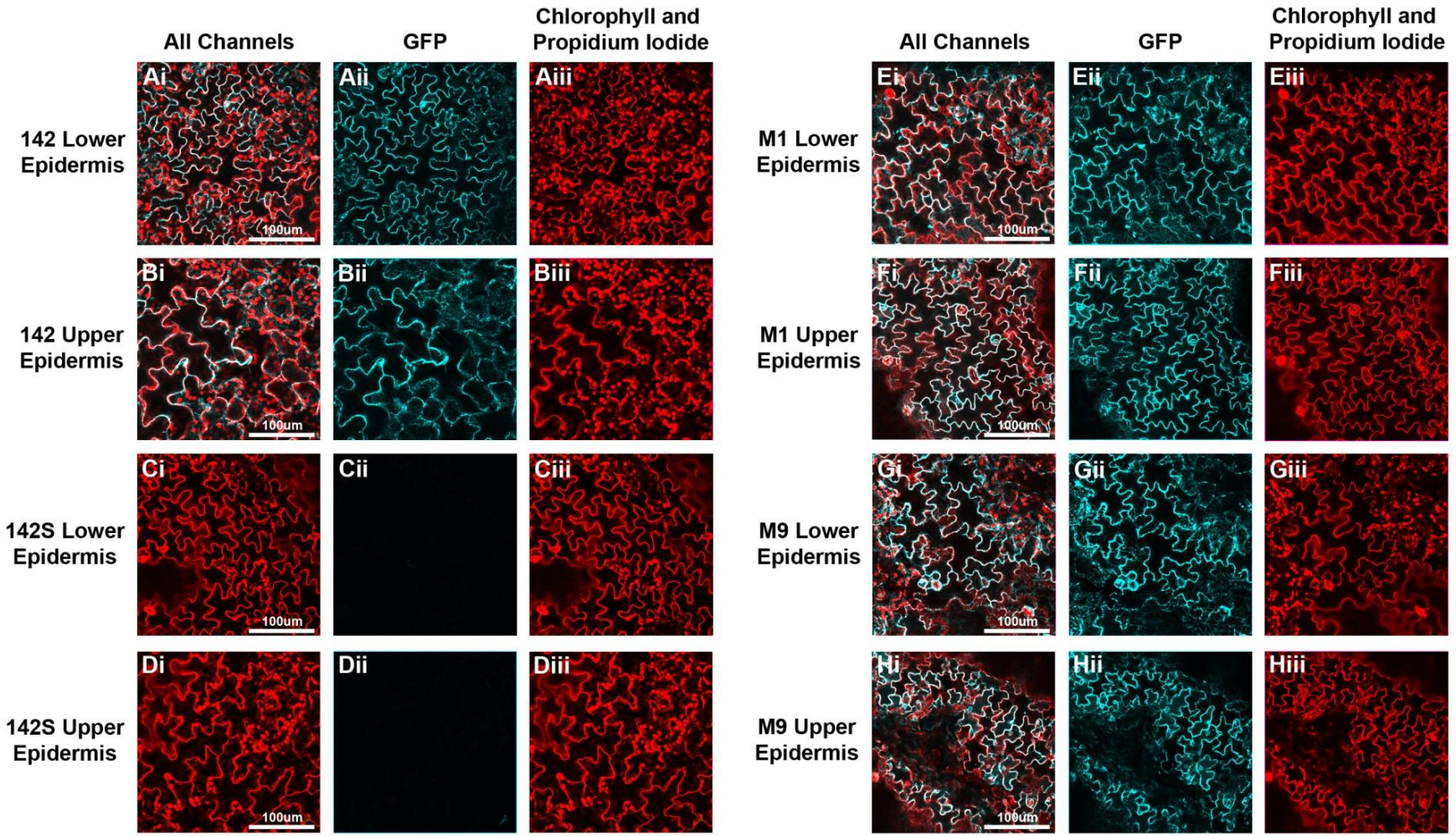


Figure 4.4: GFP expression in the epidermal layers prior to the onset of silencing in M1 and M9

The 1st leaves of 10 dpg seedlings were imaged under a confocal microscope to observe GFP expression. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**), M1 (**E** and **F**) and M9 (**G** and **H**). A scale bar is shown for each image in the combined image. Images are representative of the three leaves visualised.

At 13 dpg the M9 mutant line establishes silencing in the lower epidermis of the 1st leaf. This onset is rapid and occurs within a six hour period. However, this is only true for the lower epidermis, the upper epidermis does not establish silencing and the mesophyll layers have not been assessed in this time course. At 7 hours post dawn (hpd) M9 and 142 have similar levels of GFP fluorescence (Figure 4.5). However, by 13 hpd the lower epidermis has established silencing while the upper epidermis is unaffected (Figure 4.6). This silencing onset was also observed in the 3rd leaf, demonstrating that this delay is not leaf specific (Appendix 2.1 page 387). M1 does not silence at 13 dpg. As only the lower epidermal layer silences at this time point, this suggests a tissue specific nature to the silencing pattern. Once established the silencing pattern is mosaic in nature with some areas showing mostly GFP fluorescent/unsilenced cells (Figure 4.6 E) and others mostly silenced cells (Figure 4.6 F). Within these areas there are lone or small groups of silenced or unsilenced cells respectively. As silencing amongst the pavement cells in the lower epidermis is mosaic in nature this suggests that silencing is stochastic. The mosaic silencing pattern is also seen in the 3rd leaf (Appendix 2.1 page 387). This stochastic silencing is unusual as it has been shown that the silencing signal can be transmitted from cell to cell (Smith et al. 2007, Molnar et al. 2010, Melnyk et al. 2011). As silenced and unsilenced cells can reside next to each other this suggest that either the silencing signal cannot be transmitted or that the ability of cells to respond to this signal is affected.

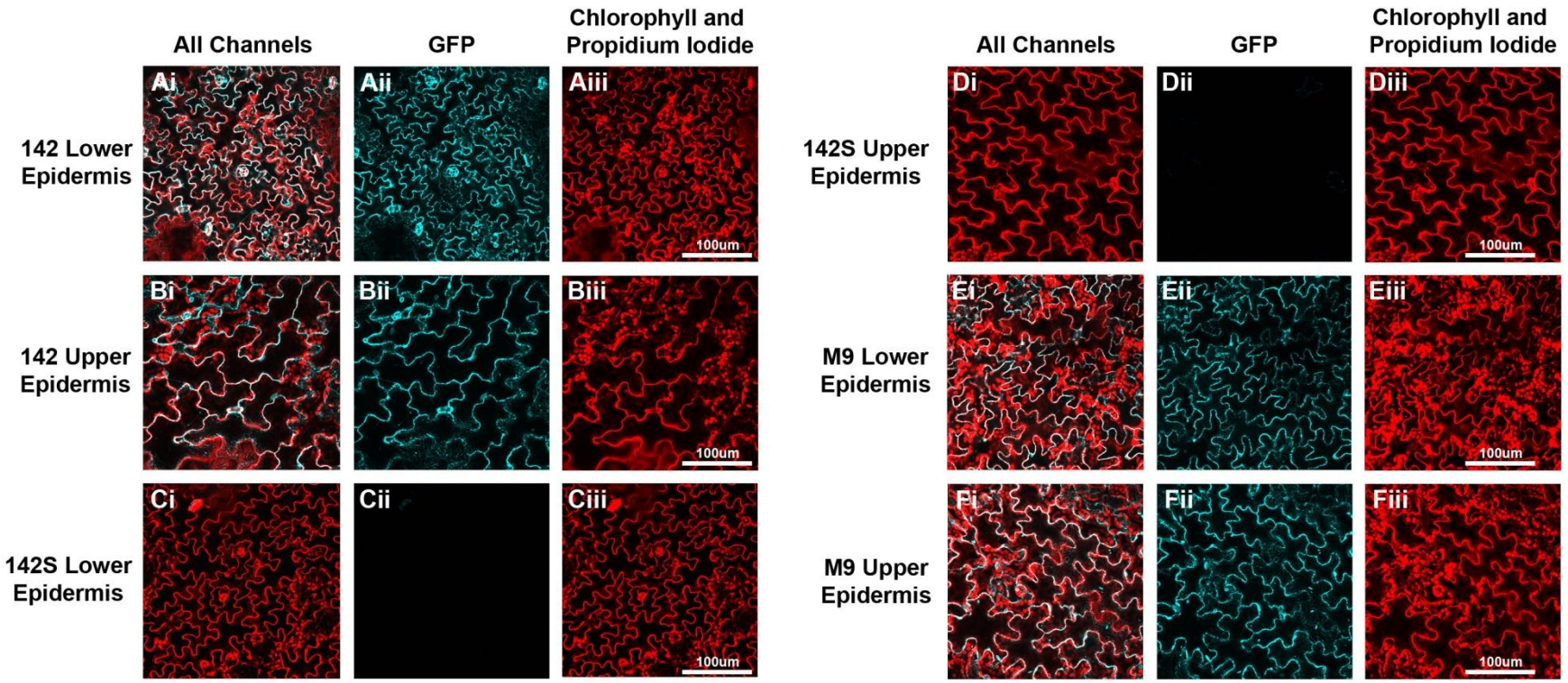


Figure 4.5: M9 does not silence the *GFP* transgene at 7 hpd on the 13th dpg in either epidermal layer

The 1st leaves of 13 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 7 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan)

and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**) and M9 (**E** and **F**). A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

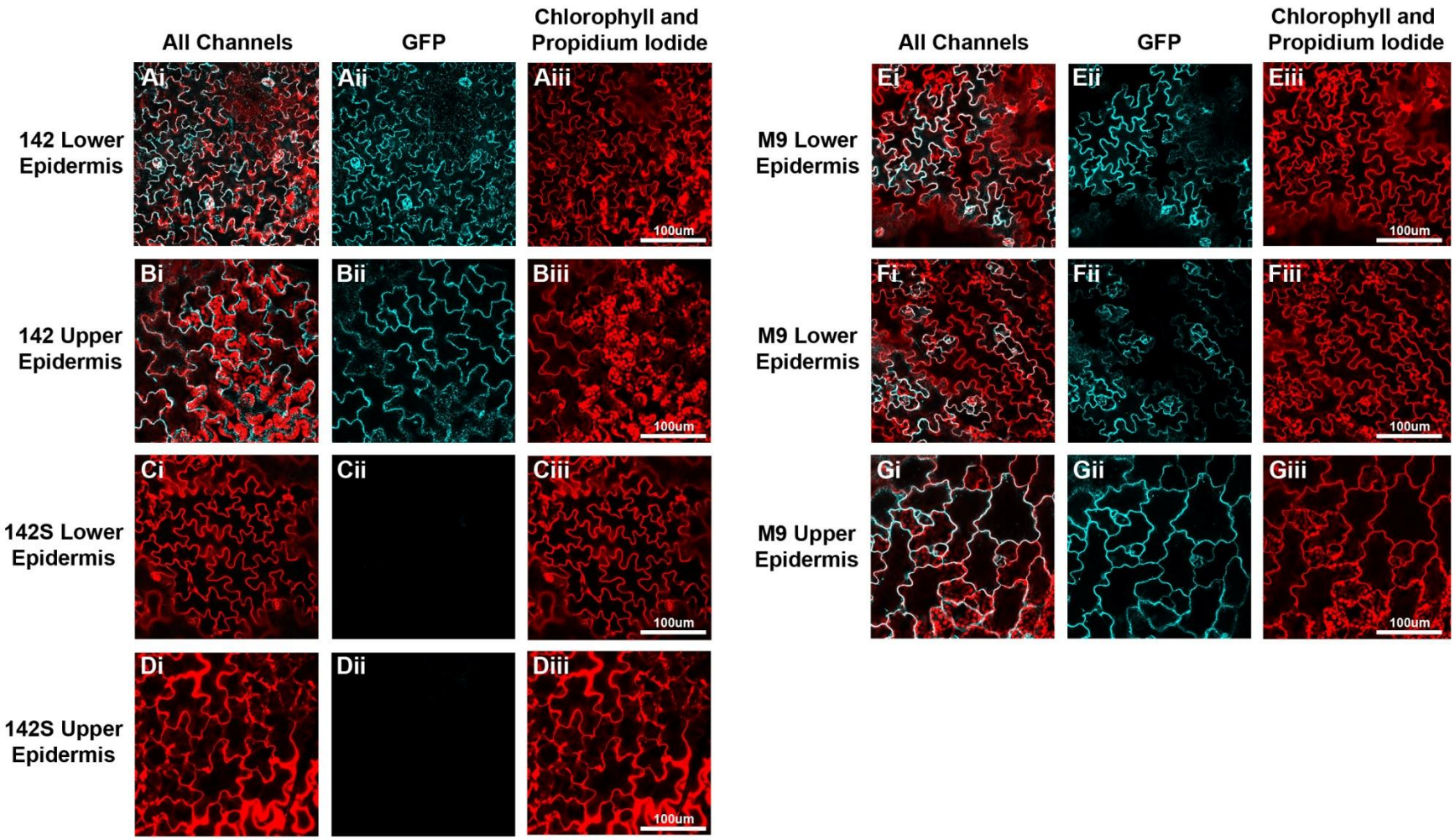


Figure 4.6: Silencing of *GFP* has occurred in the lower epidermal layer of M9 leaves by 13 hpd on the 13th dpg

The 1st leaves of 13 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 13 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**) and M9 (**E**, **F** and **G**). **E** and **F** show areas of the leaf where silencing is less (**E**) or more (**F**) prevalent. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

M1 establishes silencing at 14 dpg in the lower epidermis but not the upper epidermis. Like M9, the lower epidermis of M1 goes from unsilenced at 7 hpd (Figure 4.7) to partially silenced by 13 hpd (Figure 4.8). As both M1 and M9 silence in the lower epidermis at such specific time points this could indicate that a specific developmental event leads to the onset of silencing in the mutants. This event may also be tissue specific as the upper epidermis is unaffected at this time point and observation of the mesophyll layer using a fluorescence microscope suggests that this layer has an earlier onset of silencing. The single day difference in onset of silencing in the lower epidermis between M1 and M9 is suggestive that the M1 allele of *MORC6* has a greater effect on silencing and so is a more severe allele. Again, once silencing is established, silencing is mosaic, as there are areas of unsilenced cells with lone or small groups of silenced cells amongst them (Figure 4.8 E) and areas of silenced cells with lone or small groups of unsilenced cells amongst them (Figure 4.8 F). This delay and mosaic silencing pattern also occurs in both 3rd and 5th leaves (Appendix 2.2 page 391).

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

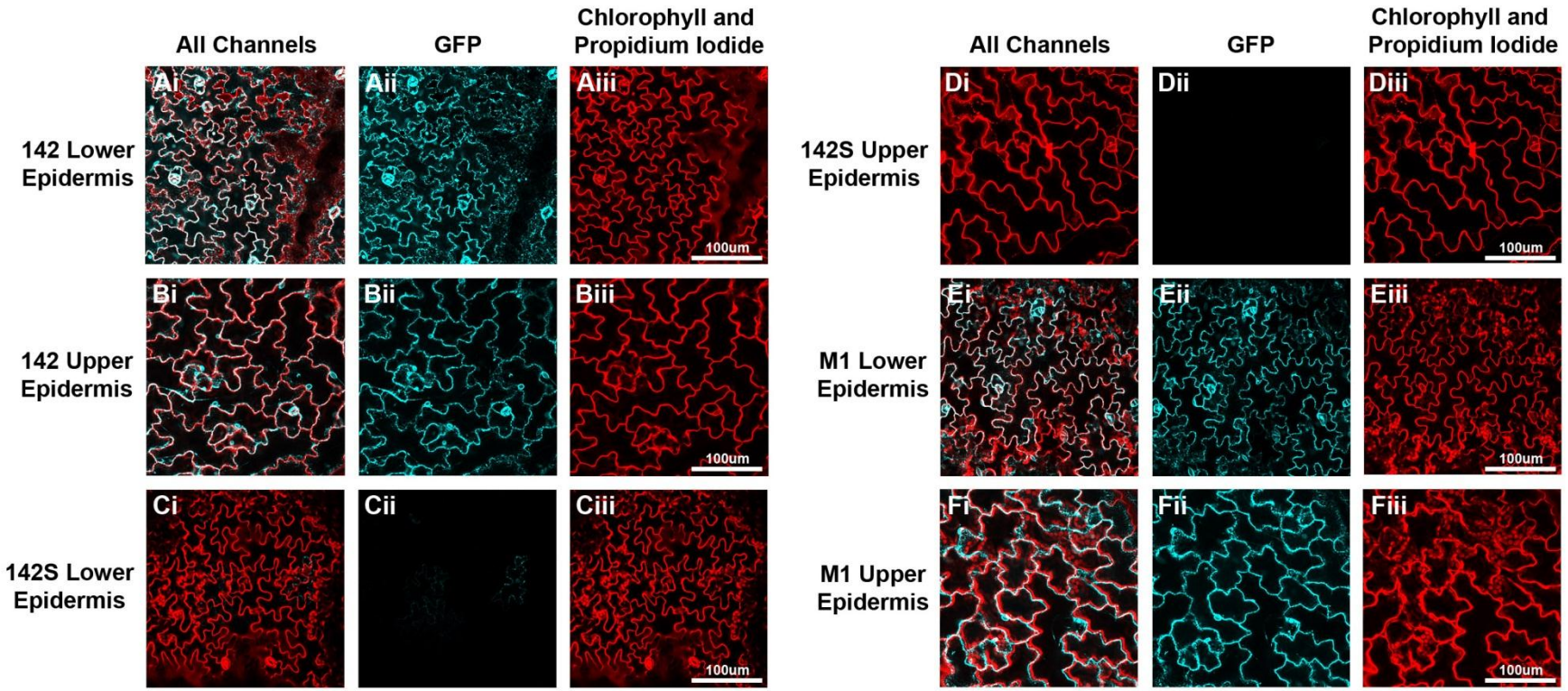


Figure 4.7: M1 does not silence the *GFP* transgene at 7 hpd on the 14th dpg in either epidermal layer

The 1st leaves of 14 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 7 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan)

and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**) and M1 (**E** and **F**). A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

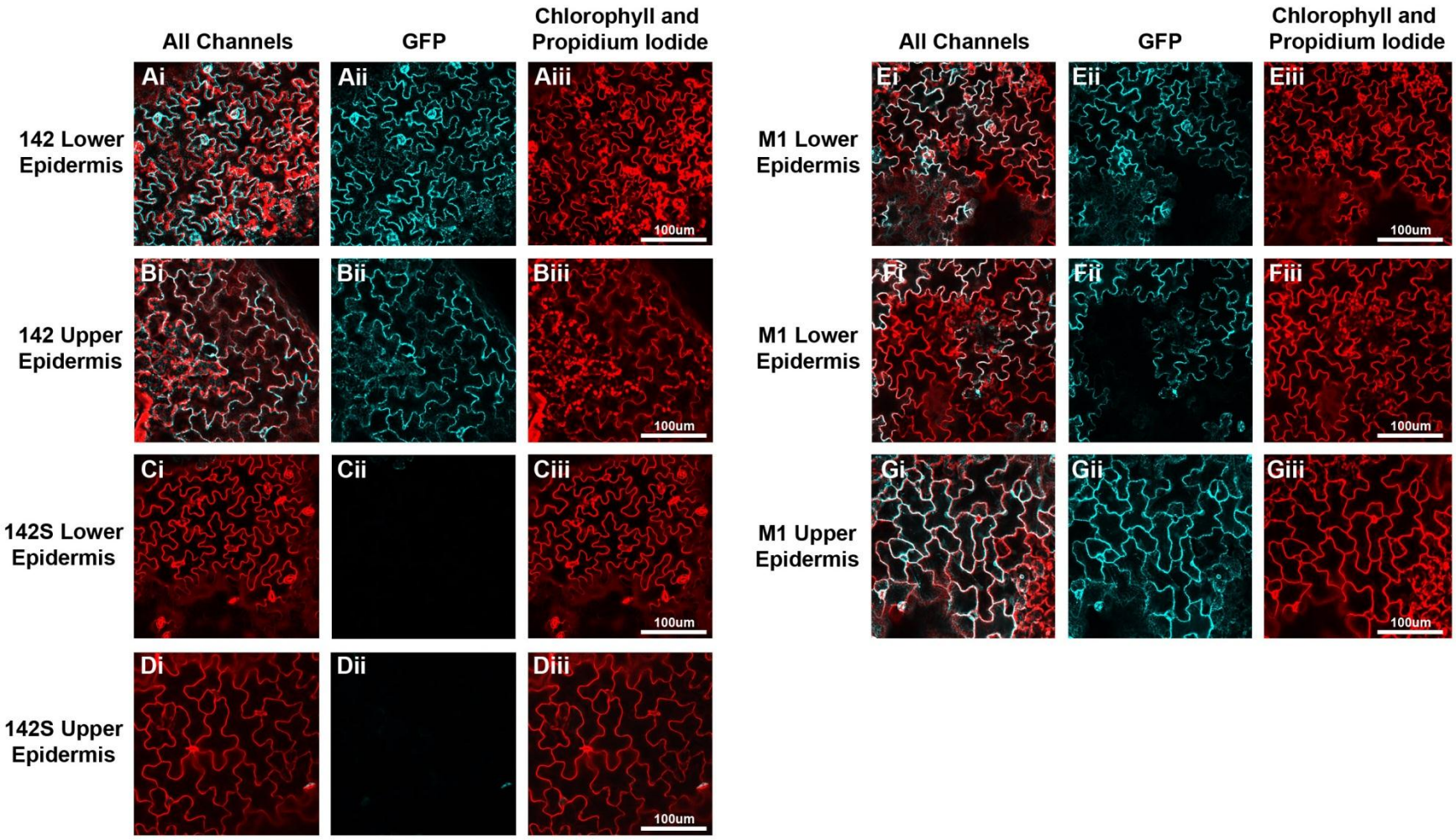


Figure 4.8: Silencing of *GFP* has occurred in the lower epidermal layer of M1 leaves by 13 hpd on the 14th dpg

The 1st leaves of 14 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 13 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**) and M1 (**E**, **F** and **G**). **E** and **F** show areas of the leaf where silencing is less (**E**) or more (**F**) prevalent. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

To see if the silencing pattern established at 13 dpg and 14 dpg for M9 and M1 respectively remained stable or if levels of silencing increased, silencing was assessed in 20 dpg leaves. Although the mosaic silencing pattern of the lower epidermis, seen at 13 dpg and 14 dpg, is still apparent, newly emerging leaves show a progressive increase in the level of silencing (Figure 4.9-Figure 4.12). This is particularly obvious when comparing the oldest leaf to emerge (1st leaf) against the youngest leaf to emerge (7th leaf) (Figure 4.9 and Figure 4.12). The areas of silenced cells are both larger and more prevalent in the 7th leaf than the 1st leaf. This may be linked to the onset of silencing as the 1st leaf was present when silencing occurred whereas the 7th leaf was not. It would suggest that cells present before silencing onset do not respond to silencing to the same extent as cells that appeared after silencing onset. As well as this difference in silencing according to when leaves emerge there is also a difference between M1 and M9. In M9 there are fewer GFP expressing cells than in M1 thus indicating that silencing is more prevalent in M9. This provides further evidence that M1 is a more severe allele of *MORC6*.

Silencing can also occur in the upper epidermis but is dependent when the leaf emerged. There is no silencing in the 1st leaf for either M1 or M9 (Figure 4.9 G and J). Small areas of silenced cells are present in M9 in the 3rd leaf but not in M1 (Figure 4.10 G and J). Both M1 and M9 have areas of silencing in the 5th leaf, although these are more prevalent in M9 (Figure 4.11 G and J). In the 7th leaf silencing of the upper epidermis has reached similar levels to the lower epidermis in M9 while in M1 the areas of silencing are still small but have increased in prevalence compared to the 5th leaf (Figure 4.12 G and J). The exact time point when silencing is initiated in the upper epidermis of M1 and M9 is not known, but occurs between 15 dpg and 20 dpg. The difference in the strength of silencing of the upper epidermis between M1 and M9 also supports the suggestion that M1 is a more severe allele than M9.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

Differences in both the silencing pattern and time of onset between the upper and lower epidermal layers demonstrate tissue specificity in silencing pattern. The delay in silencing also suggests that silencing may be dependent on a developmental event. This could suggest that there is both a tissue and developmental component to *MORC6* expression. MORC6 expression would therefore be highest at earlier time points than later time points and higher in the upper epidermis than the lower epidermis. Redundancy with other proteins, possibly other MORCs, with different expression patterns would cause the development of silencing observed.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

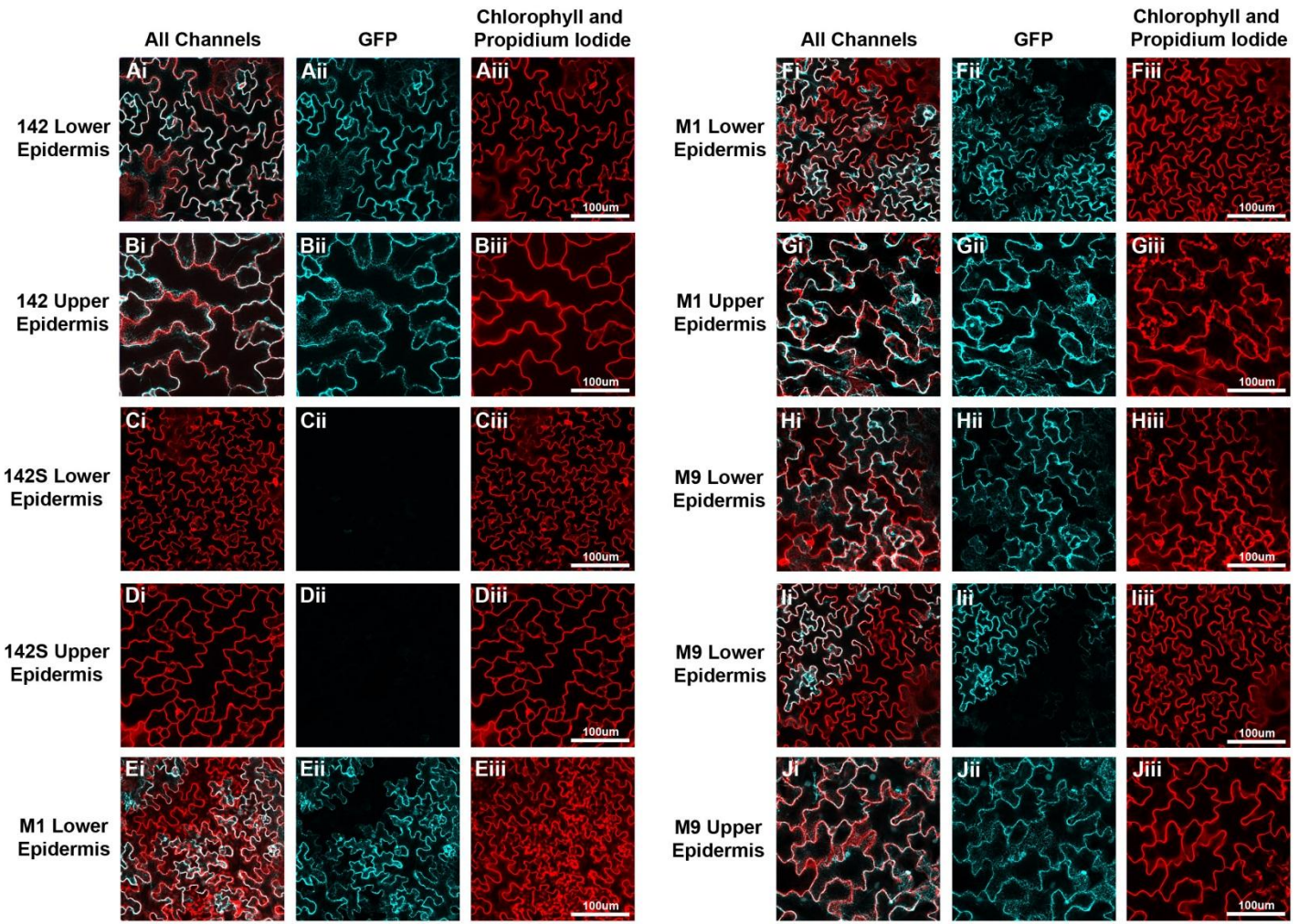


Figure 4.9: Mosaic silencing pattern of the lower epidermis in 1st leaves of 20 dpg M1 and M9 plants

The 1st leaves of 20 dpg seedlings were imaged under a confocal microscope to observe GFP expression. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**), M1 (**E**, **F** and **G**) and M9 (**H**, **I** and **J**). **E** and **F** and **H** and **I** show areas of the leaf where silencing is less (**E** and **H**) or more (**F** and **I**) prevalent for M1 and M9 respectively. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

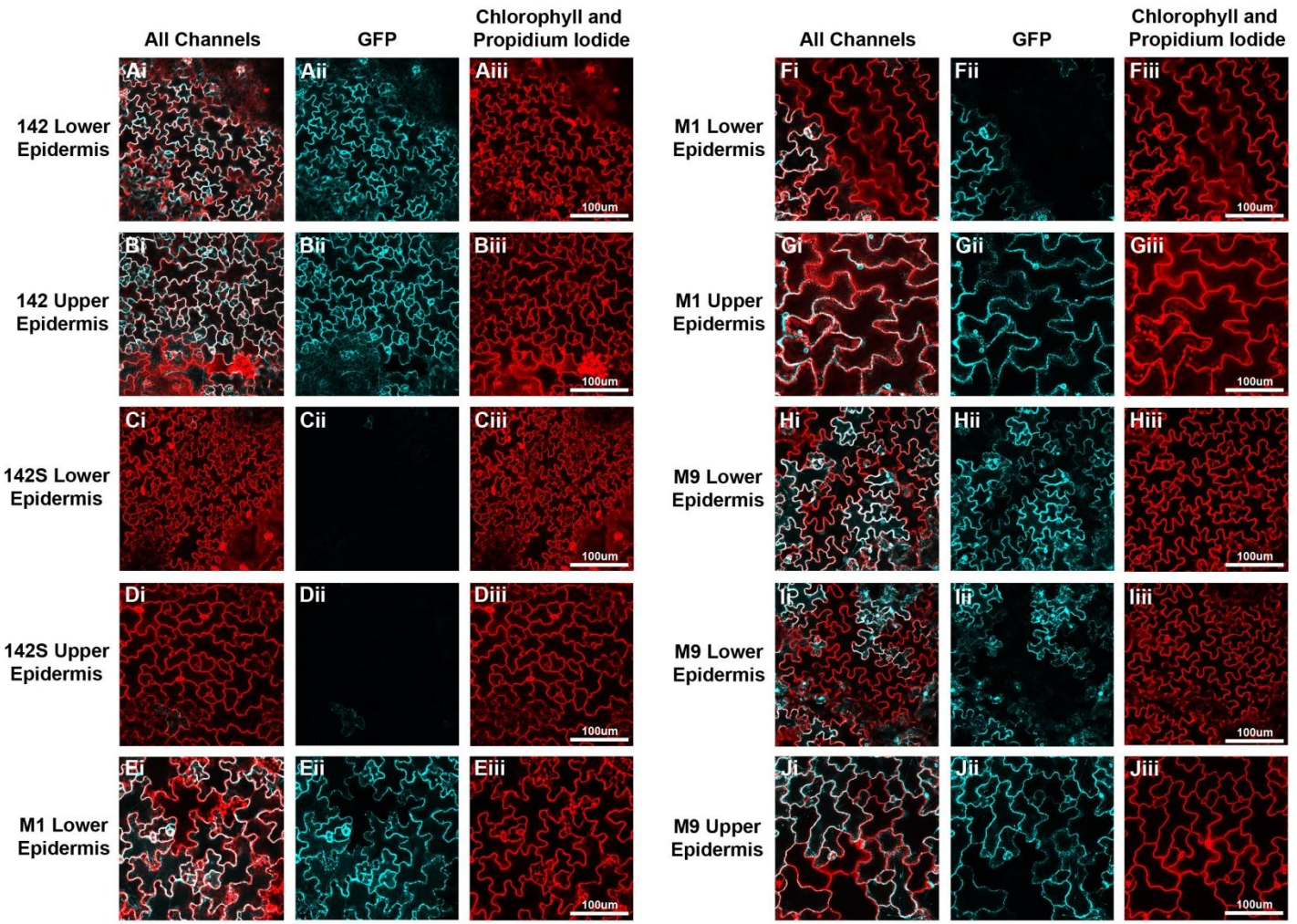


Figure 4.10: Mosaic silencing pattern of the lower epidermis in 3rd leaves of 20 dpg M1 and M9 plants and silencing of the upper epidermis in M9

The 3rd leaves of 20 dpg seedlings were imaged under a confocal microscope to observe GFP expression. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**), M1 (**E**, **F** and **G**) and M9 (**H**, **I** and **J**). **E** and **F** and **H** and **I** show areas of the leaf where silencing is less (**E** and **H**) or more (**F** and **I**) prevalent for M1 and M9 respectively. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

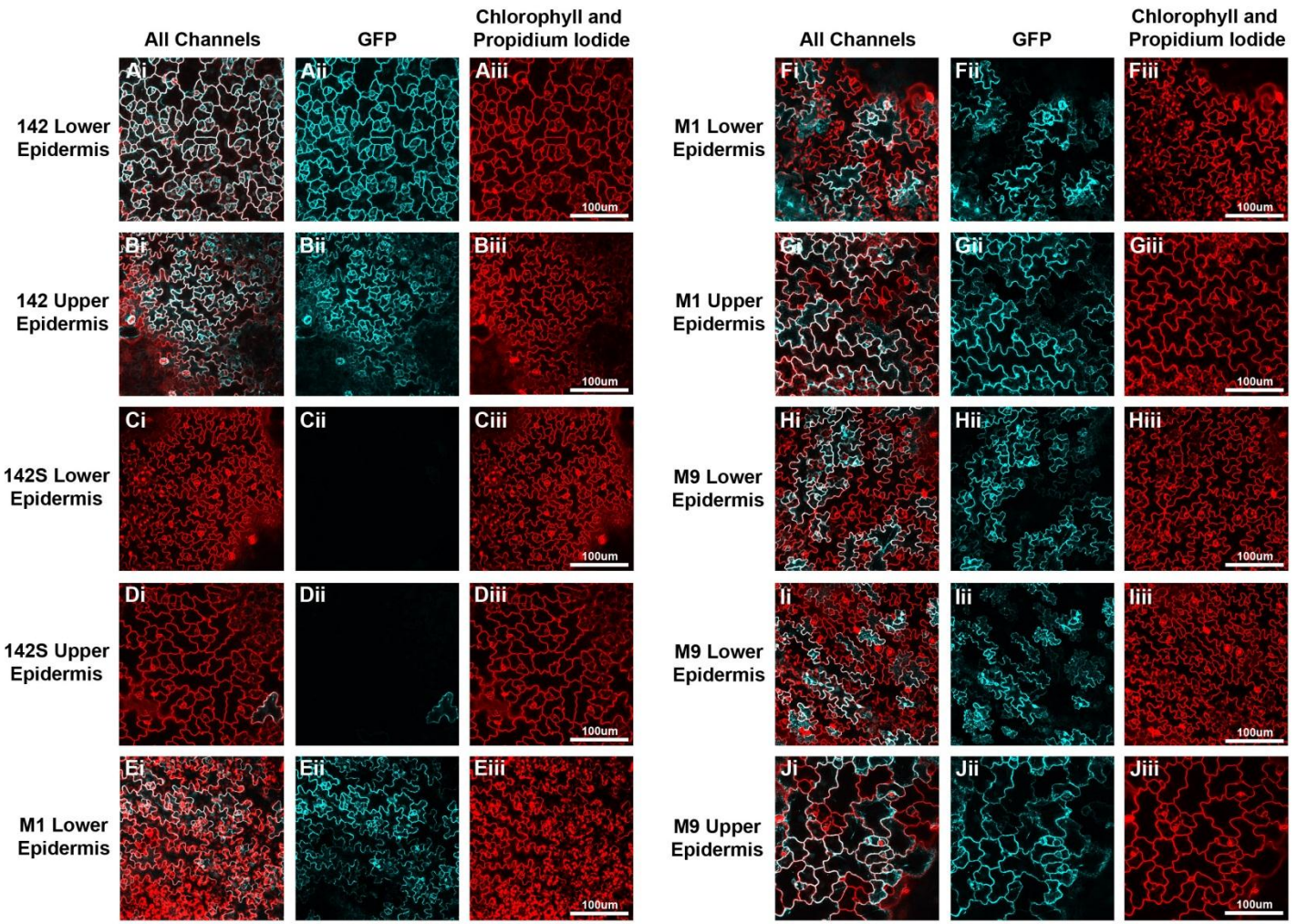


Figure 4.11: Mosaic silencing pattern of the lower epidermis and silencing of the upper epidermis in 5th leaves of 20 dpg M1 and M9 plants

The 5th leaves of 20 dpg seedlings were imaged under a confocal microscope to observe GFP expression. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**), M1 (**E**, **F** and **G**) and M9 (**H**, **I** and **J**). E and F and H and I show areas of the leaf where silencing is less (**E** and **H**) or more (**F** and **I**) prevalent for M1 and M9 respectively. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

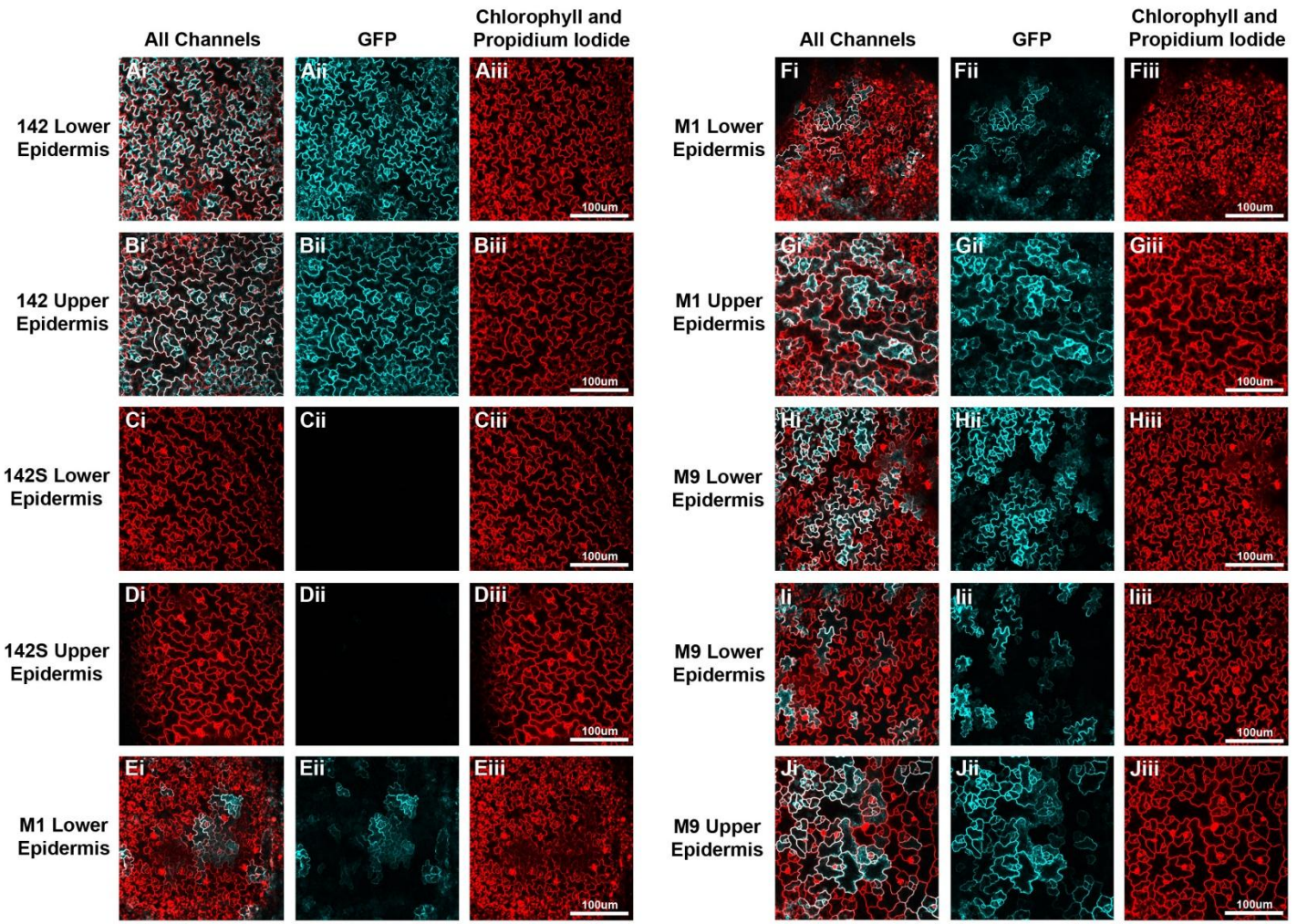


Figure 4.12: Mosaic silencing pattern of the upper and lower epidermis in 7th leaves of 20 dpg M1 and M9 plants

The 7th leaves of 20 dpg seedlings were imaged under a confocal microscope to observe GFP expression. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A** and **B**), 142S (**C** and **D**), M1 (**E**, **F** and **G**) and M9 (**H**, **I** and **J**). **E** and **F** and **H** and **I** show areas of the leaf where silencing is less (**E** and **H**) or more (**F** and **I**) prevalent for M1 and M9 respectively. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

4.2.3 Is the onset of silencing linked to developmental stage?

As discussed previously, as silencing of the lower epidermis occurs at such a specific point, it is possible that MORC6 is required for silencing during early development. To investigate whether there is any link to development, plants were grown in short day (8 hour day length) conditions. In the previous time course, plants were grown in 16 hour day length conditions. Due to the reduced hours of light the development of the plants will be retarded. This means that both M1 and M9 will be at an earlier stage of plant growth at 13 dpg and 14 dpg than in long day conditions. Plants have 4 rosette leaves at 13 dpg and 5 rosette leaves at 14 dpg in long day growth rooms. But in short day conditions plants only have 2 rosette leaves at both 13 dpg and 14 dpg. The 1st leaf itself will also be at a earlier stage of leaf development at 13 dpg and 14 dpg in short day conditions than in long day conditions. If silencing is linked to development theoretically both M1 and M9 will silence later in the short day conditions.

Both M1 and M9 silence at the same time in short day growth rooms as seen in long day growth rooms. Silencing of the 1st leaf lower epidermis was assessed in M1 and M9 at 15 dpg when silencing has occurred in both M1 and M9. At 15 dpg plants in short day growth rooms are still at the 2 rosette leaf stage. Plants were initially observed using a fluorescence microscope and representative images taken by confocal microscopy (Figure 4.13). Only the lower epidermis is shown as the upper epidermis does not silence at this stage in long day conditions. Both M1 and M9 show a mosaic silencing pattern (Figure 4.13 C-F). It is clear, from this experiment, that the onset of silencing is not linked to a specific plant growth stage or point in leaf developmental but instead on time after germination. However, the exact point at which M1 and M9 silence in short day conditions was not determined.

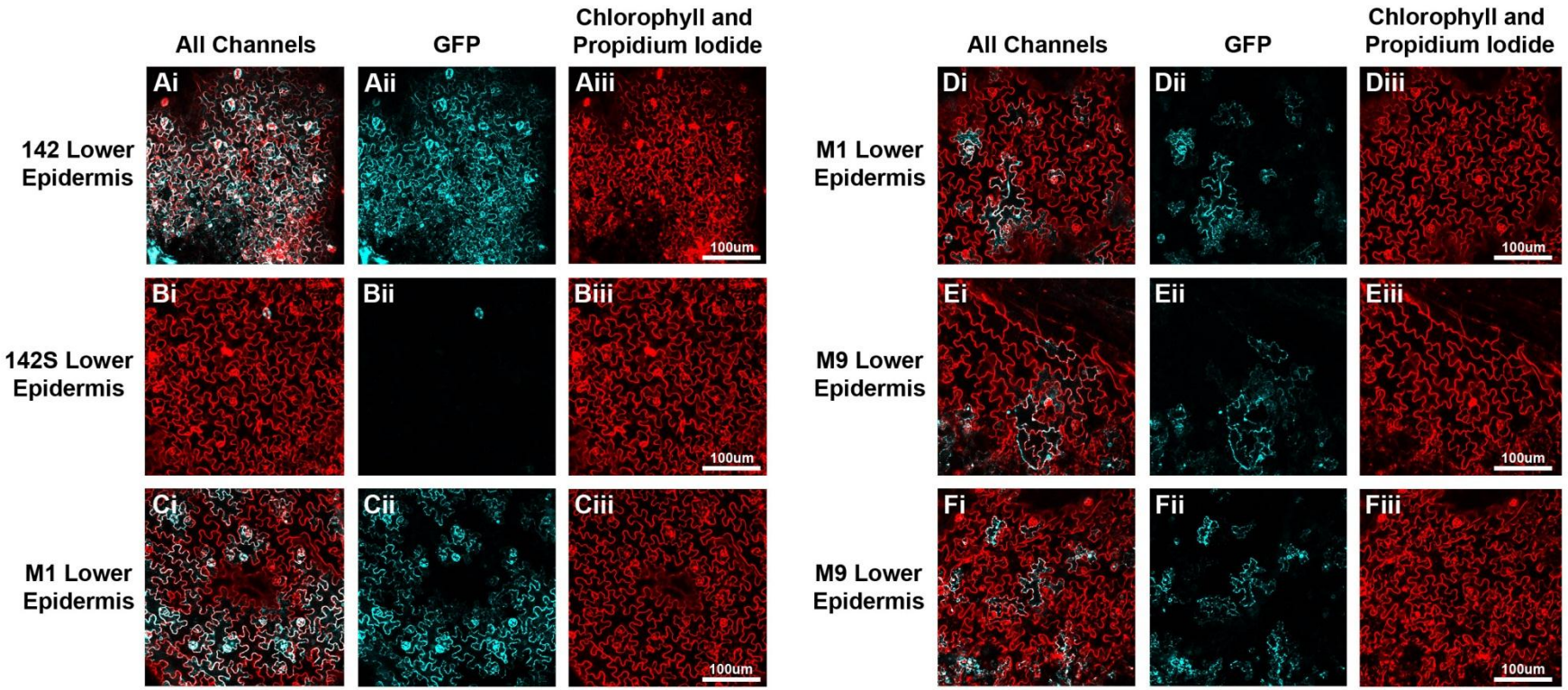


Figure 4.13: Mosaic silencing is still established in both M1 and M9 by 15 dpg in the lower epidermis

The 1st leaves of 15 dpg seedlings grown in short day conditions were imaged under a confocal microscope to observe GFP expression. The leaves were first stained with propidium iodide which binds to the cell walls. All images are of the lower epidermal layer. Two separate channels were used,

one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were imaged: 142 (**A**), 142S (**B**), M1 (**C** and **D**) and M9 (**E** and **F**). A scale bar is shown for each image in the propidium channel only image. Images are representative of the approximately 100 plants from each line visualised under UV fluorescence microscope.

4.2.4 Mosaic silencing pattern in the mesophyll layers

The time course experiment showed a mosaic silencing pattern in the upper and lower epidermal layer. It also showed that the severity of the loss of silencing is negatively correlated with the time at which a leaf emerged. In order to investigate whether a similar silencing pattern occurs in the mesophyll layers, leaves were imaged using UV fluorescence microscopy. The adaxial and abaxial sides of odd numbered rosette leaves from 21 dpg plants were imaged (Figure 4.3). Leaves were taken from lines 142, 142S, M1 and M9. GFP fluorescence seen under the fluorescence microscope is mainly from the mesophyll layers. This is due to both the greater number of mesophyll cells compared to epidermal pavement cells and that GFP fluorescence is comparatively lower in the epidermal layers than the mesophyll layer. The difference in fluorescence level is because mesophyll cells have a larger cytoplasm than pavement cells, due to the large vacuole in pavement cells.

GFP fluorescence in 142 is ubiquitous on both the adaxial and abaxial side of the leaves. Fluorescence is particularly strong in the leaf veins. In 142S the majority of cells are silenced. Images for 142 and 142S can be found in Appendix 4.3 (page 399). In both M1 and M9, silencing is similar to that seen under the confocal microscope. The abaxial side of the leaves in M1 and M9 show a mosaic silencing pattern, with areas of silencing and areas of GFP expression (Figure 4.14). There appears to also be a relationship between the level of silencing and leaf emergence time. Prevalence and size of the areas of silencing increase from the 1st leaf, where levels of silencing are low, to the 7th leaf, which has higher levels of silencing in both M1 and M9 (Figure 4.14 C, D, G, H, K, L, O and P). This suggests that silencing in the spongy mesophyll layer is similar in nature to that seen in the lower epidermis. The silencing patterns also clearly show a difference between M1 and M9. Levels of silencing appear to be greater in M9 than M1. Again this indicates that M1 is the stronger of the two alleles.

On the abaxial side of leaves there is strong fluorescence in the veins, although this is reduced compared to 142 (Figure 4.14 and Appendix 4.3 page 399). Silencing of the palisade mesophyll layer is dependent on the time of leaf emergence. There is no silencing on the adaxial side of the 1st and 3rd leaves of M1 and M9, but there are small areas of silencing on the 5th leaf and larger areas of silencing on the and 7th leaf (Figure 4.14 A, B, E, F, I, J, M and N). This pattern is similar to that of the upper epidermis, indicating that palisade layer silencing is comparable to the upper epidermis. As both the palisade and spongy mesophyll layers mosaic silencing pattern closely resembles that of the upper and lower epidermal layers respectively, this would suggest that the final silencing pattern is dependent on whether the cells are adaxial (upper epidermis and palisade mesophyll layer) or abaxial (lower epidermis and spongy mesophyll layer) rather than a specific single tissue type. However, it should be noted that time of onset of silencing is different between the epidermal layers and mesophyll layers so there still may be some level of tissue specificity in silencing pattern.

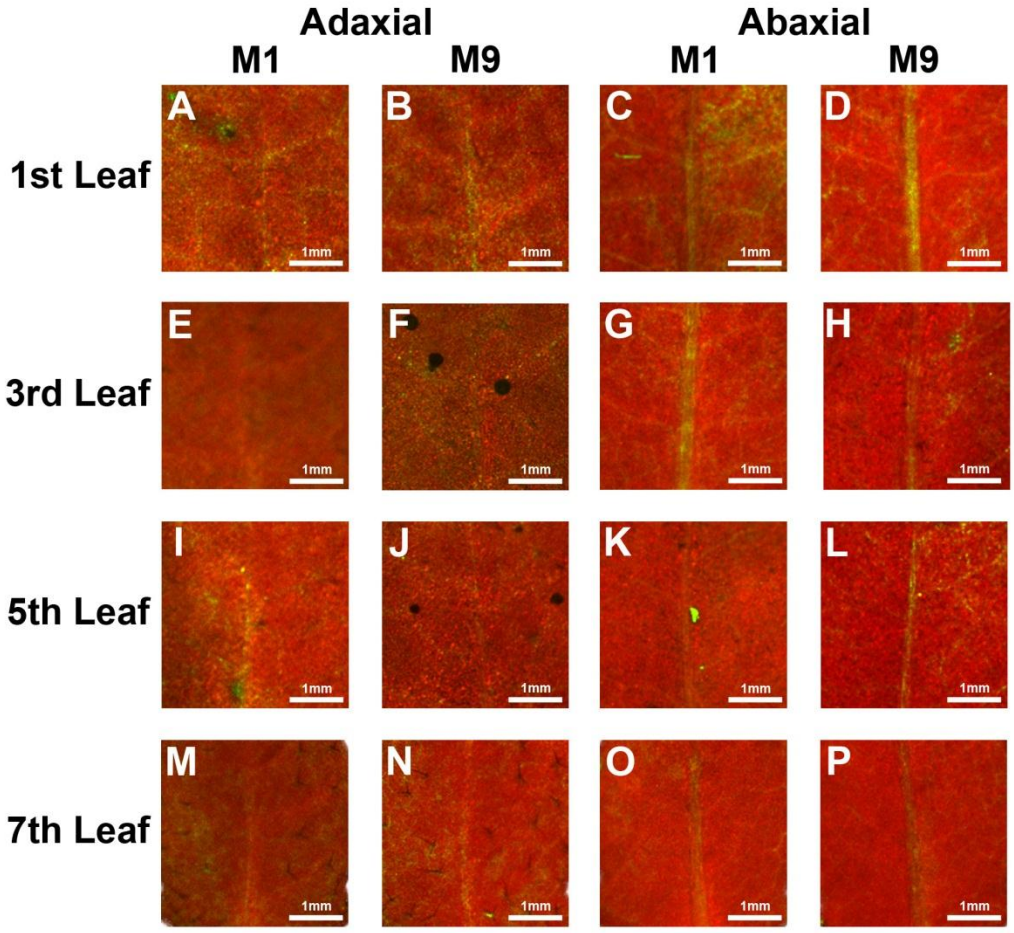


Figure 4.14: Mosaic silencing pattern of *GFP* on the adaxial and abaxial sides of M1 and M9 leaves at 21 dpg

Images of the adaxial and abaxial side of odd numbered leaves from lines M1 and M9, taken under a fluorescence microscope at 21 dpg. Each image shows an enlargement of the mid-section of each leaf. Images **A**, **B**, **C** and **D** are of the 1st leaf, images **E**, **F**, **G** and **H** are of the 3rd leaf, images **I**, **J**, **K** and **L** are of the 5th leaf and images of **M**, **N**, **O** and **P** are of the 7th leaf. The first column (**A**, **E**, **I** and **M**) are image of the adaxial side of M1 leaves. The second column (**B**, **F**, **J** and **N**) are images of the adaxial side of M9 leaves. The third column (**C**, **G**, **K** and **O**) are image of the abaxial side of M1 leaves. The fourth column (**D**, **H**, **L** and **P**) are images of the abaxial side of M9 leaves. A scale bar is shown at the bottom right of each image.

The fluorescence microscopy imaging on the mesophyll layer suggests that the palisade mesophyll layer has a silencing pattern similar to the upper epidermis while the spongy mesophyll layer has a pattern similar to that of the lower epidermis. To further investigate this observation the mesophyll layers were imaged using confocal microscopy. Since confocal microscopy on whole leaves could not image beyond the first layer of the mesophyll layer, sectioning was used. Sections in the adaxial to abaxial plane were taken from the mid-leaf of 40 dpg 1st leaves, from lines 142, 142S, M1 and M9. Images were then taken of the main vein and the mesophyll layer using a confocal microscope. Previously the epidermal layers were imaged in one focal plane. However, for the mesophyll layers it was not possible to image in a single focal plane. This is due to the curvature of the leaf sections causing different cells in the first layer of the section to be visible in different focal planes (Figure 4.15). To produce an image of the entire mesophyll layer the sections were Z-stacked. Z-stacking takes multiple images at different focal planes. The images can then be combined into a composite image. In this case maximum intensity was used. This is where each pixel in the final image comes from the image in the Z-stack with the highest intensity of fluorescence at that pixel.

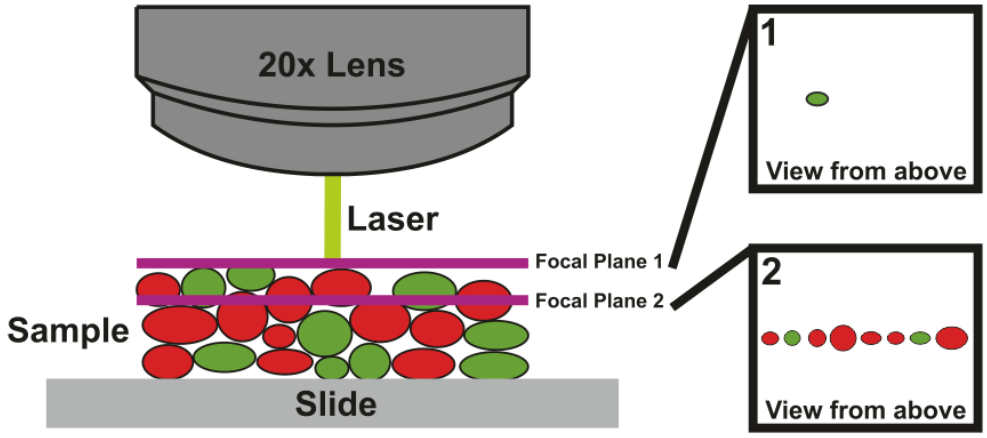


Figure 4.15: Explanation of use of z-stacking.

Shows a 2D diagrammatic representation of imaging a leaf section on a slide under a confocal microscope. Depth and length are shown and for simplicity the section in this diagram is also one cell wide. The red and green ovals represent silenced and unsilenced mesophyll cells respectively. The purple lines represent different focal planes and the two boxes on the right show the resulting images from these two focal planes. Only a single cell is visible in the first focal plane whereas multiple cells are visible in the second.

In line 142, GFP fluorescence can be seen throughout the sections and is particularly high in the vein itself (Figure 4.16 A and B). Expression is higher in the spongy mesophyll layer than the palisade layer (Figure 4.16 I and J). This could indicate that there is tissue specific expression of the GFP transgene. It should be noted that cells in the sections are ill-defined in the images. There are two reasons for this. Firstly the Z-stacks may, as well as the first layer of cells, also extend into the layer of cells below. When the composite image is produced overlapping cells in the top layer and the layer below will be superimposed onto each other, leading to blurred boundaries between cells. The second reason is for each image in the Z-stack, there is bleed through of fluorescence from layers above and below the focal plane of that image. The pin hole of the laser, which controls the focus of the laser, is set high in order to detect fluorescence from the samples. Larger pin holes means the laser is less focused and will excite fluorophores above and below the focal plane. Therefore as the pin hole is large there will be more fluorescence from these layers. This results in a loss of definition of cells in the plane being imaged. The silencing pattern in the mesophyll layer of line 142S is similar to that seen in the epidermal layers, in that the majority of cells are silenced but there are isolated unsilenced, GFP fluorescent, cells (Figure 4.16 K and L). The unsilenced cells appear in both the palisade and spongy mesophyll layers, so tissue specificity does not appear to be the cause of this lack of silencing. There is also no GFP fluorescence in the main vein in the leaf (Figure 4.16 C and D).

In M1 there is a clear decrease in GFP fluorescence compared to 142 (Figure 4.16). There is however still strong GFP fluorescence in the main vein (Figure 4.16 E and F). In the palisade and spongy mesophyll layers, silencing of GFP is mosaic as there are areas of silenced and unsilenced mesophyll cells (Figure 4.16 M and N). The mosaic silencing pattern is stochastic in nature as within each mesophyll layer cells can be either silenced or unsilenced. Compared to 142, GFP expression in the palisade layer in M1 is not reduced to the same extent as in the spongy mesophyll layer. The loss of silencing in M1 is therefore more pronounced in the palisade layer than the spongy mesophyll layer, suggesting that there is cell type specificity in silencing. This is similar to what was observed from fluorescence microscopy, in that the palisade mesophyll cells appeared to have a smaller reduction in GFP expression than the spongy mesophyll cells. In the epidermal layers there is more GFP expression in the upper epidermis than the lower epidermis. As the two upper layers (palisade mesophyll cells and upper epidermal pavement cells) show a greater loss in silencing to the two lower cell layers (spongy mesophyll cells and lower epidermal pavement cells), this could suggest a adaxial to abaxial gradient in silencing response in the leaves of the mutants.

Like M1, M9 also shows a marked decrease in GFP expression from that seen in 142 (Figure 4.16). This decrease is however greater in M9 than M1, indicating that levels of silencing are higher in M9 than M1. This can be seen from the reduction in fluorescence in the main vein compared to M1 Figure 4.16 F and H). This correlates with what is seen in the epidermal layers in that M1 has a greater loss of silencing than M9, again indicating that M9 is the weaker of the two alleles. As in M1, silencing is mosaic in both mesophyll layers (Figure 4.16 O and P). There is also a difference between the two mesophyll layers. Compared to 142, the loss of GFP fluorescence is greater in the spongy mesophyll than the palisade mesophyll layer. M9 also shows greater levels of GFP fluorescence in the upper epidermis than the lower epidermis. This suggests that, like M1, there is a possible adaxial to abaxial gradient in silencing in the leaves.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

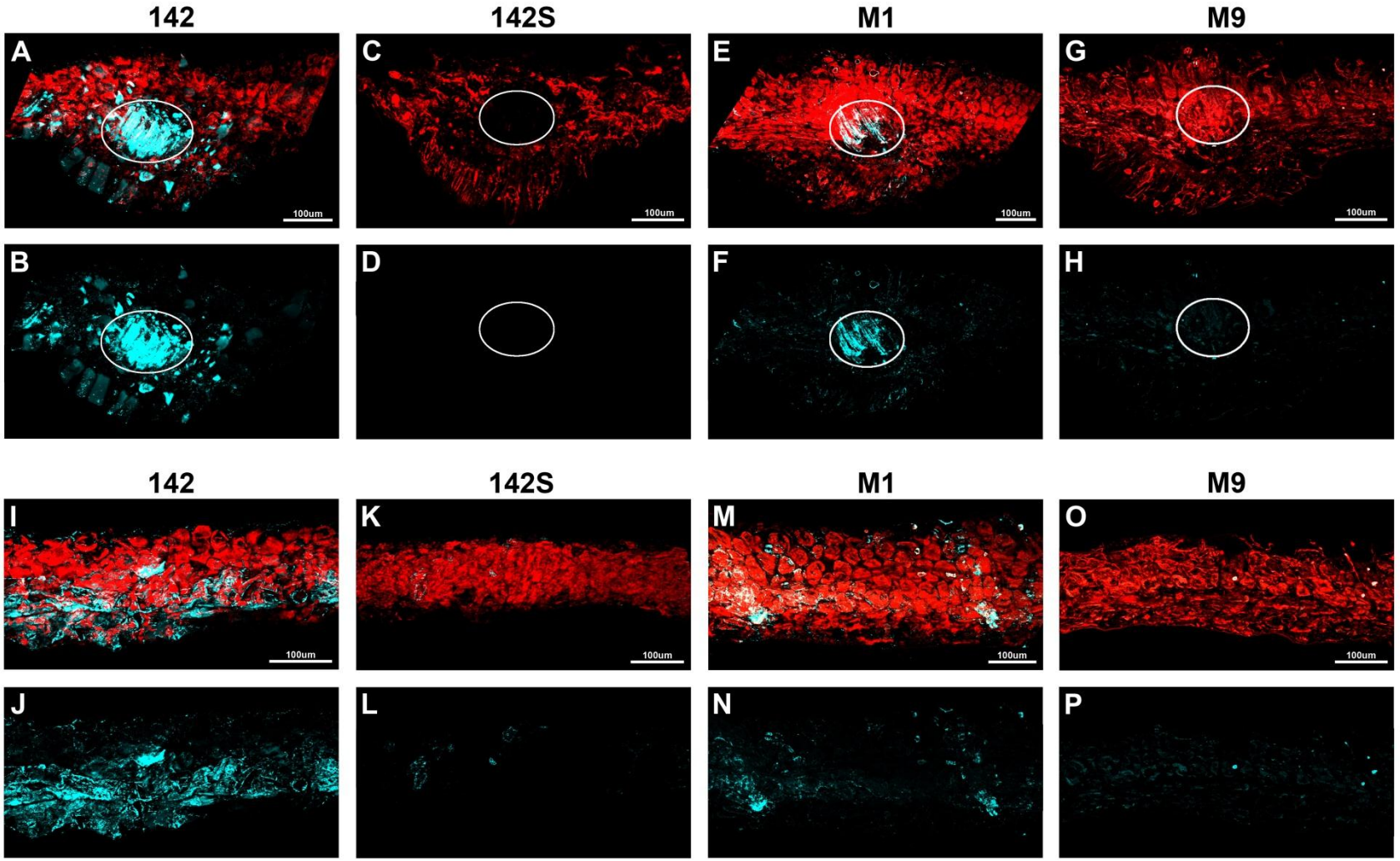


Figure 4.16: Cross sections of a 142, 142S, M1 and M9 leaves at 40dpg showing mosaic silencing in the mesophyll layers

Confocal microscopy images of cross sections from the 1st adult rosette leaf of 40 dpg old plants from lines 142, 142S, M1 and M9. All sections were taken from the mid-leaf. Two channels were used for imaging GFP (cyan) and chlorophyll (red). For each separate image a combined image of both channels (**A**, **C**, **E**, **F**, **I**, **K**, **M** and **O**) and a image of just the GFP channel (**B**, **D**, **F**, **H**, **J**, **L**, **N** and **P**) are shown. All images are maximum intensity projections from the z stacks of each section. **A-H:** are images of the main vein and surrounding tissue. The white circle in images **A-H** highlights the main vein in each image. **I-P:** are images of the mesophyll layer. **A**, **B**, **I** and **J** are images from line 142. **C**, **D**, **K** and **L** are images from line 142S. **E**, **F**, **M** and **N** are images from line M1. **G**, **H**, **O** and **P** are images from line M9. In all cases the leaf cross sections are orientated so that the adaxial side of the cross section is at the top of each image. A scale bar is shown for each combined image.

4.2.6 Root specific *GFP* silencing pattern

From the different silencing pattern of the upper and lower epidermal layers and mesophyll layers, it was concluded that there may be tissue or cell type specific differences in silencing for M1 and M9. The silencing pattern in roots was observed to see if there was a tissue or cell type specific effect on silencing in roots. Roots from plant lines 142, 142S, M1 and M9 were imaged at 33 dpg under a UV fluorescence microscope. Background autofluorescence, shown in red in the images, is present in the main roots and can be seen in 142S, M1 and M9, but no autofluorescence can be detected in 142 (Figure 4.17). This is likely to be due to the camera used to take these images being unable to detect the background autofluorescence above the GFP fluorescence in 142.

142 had ubiquitous GFP expression in both the main and lateral roots (Figure 4.17 A). However, silencing in 142S differed from that seen in leaves as there was only partial silencing of the *GFP* transgene (Figure 4.17 B). Silencing in 142S is strongest in the main roots, while in the lateral roots GFP fluorescence is closer to that of 142. Although compared to 142, GFP fluorescence was reduced, it was not to the same extent as it is in leaves and would suggest that there is reduced expression of the *35S IR* transgene in root tissue compared to leaf tissue. However, this is unexpected as the *NOS* promoter has been previously shown to be more strongly expressed in root tissue than leaf tissue (An et al. 1988). The probable cause of this is that regulatory elements in the genome where the 3*5S IR* transgene is inserted are repressing its transcription in roots. It is also possible that RdDM silencing of the *GFP* transgene is reduced in roots rather than a reduction in the expression of the *35S IR* transgene. In both M1 and M9 GFP expression has increased compared to 142S, with GFP fluorescence in the lateral roots being similar to that seen in 142 (Figure 4.17C and D). Interestingly M9 shows higher levels of GFP fluorescence in the main root than M1, which is contrary to what is seen in leaves where the opposite is true. The reversal in levels of silencing between M1 and M9 could be due to one of the other EMS

mutations carried in the M1 and M9 lines affecting silencing levels, rather than being a feature of the *morc6* mutations themselves.

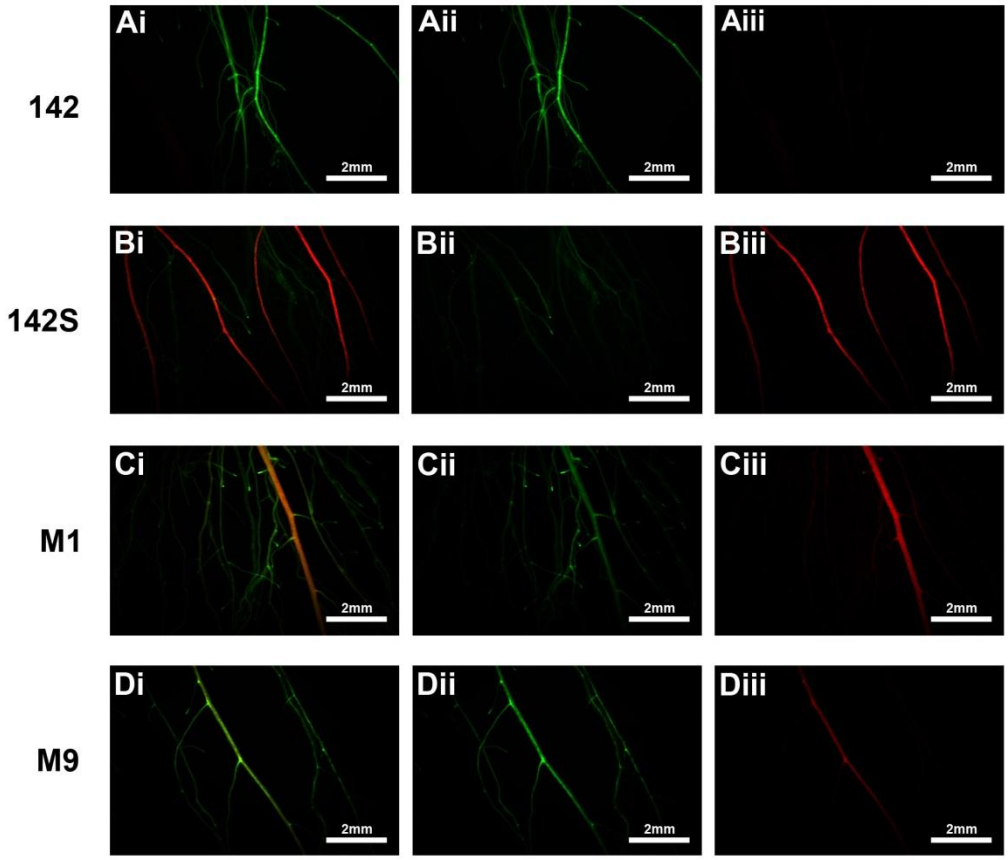


Figure 4.17: Differences in silencing pattern of *GFP* in roots of 142, 142S, M1 and M9 33 dpg plants

Plants were grown on vertical MS plates in long day growth rooms until 33 dpg then imaged *in situ* using a fluorescence microscope. Images of roots from lines 142 (**A**), 142S (**B**), M1 (**C**) and M9 (**D**) are shown. For each line there is a combined image of GFP (green) and plant autofluorescence (red) (**i**) and seperate images of GFP (**ii**) and plant autofluorescence (**iii**). A scale bar is shown as the bottom of each image. Images are representative of 8 plants for 142, M1 and M9 and 10 plants for 142S.

4.2.7 Difference in the proportion of silenced to unsilenced mesophyll cells between the M1 and M9 mutants

Observation of the silencing patterns in M1 and M9 by confocal and fluorescence microscopy suggest that levels of silencing are higher in M9 than M1. This suggests that the mutation in M1 has a more severe effect on the MORC6 protein's function than M9. To get an actual quantitative value of the extent of silencing in both M1 and M9 in leaves, cell sorting and flow cytometry was used to separate unsilenced cells from silenced cells. The main aim of the cell sorting was to extract DNA from these two cell populations in order to look at DNA methylation, which is discussed later in this chapter (page 278). However, the sorting process also allowed for flow cytometry data on numbers of silenced and unsilenced cells to be obtained. Cell sorting uses the scattering of a laser beam to identify and sort different cells. The cell properties that can be deduced from the scattering of the laser beam include: cell size, cell shape, cell viability, granularity of the cytoplasm and fluorescence of different flurophores (Rieseberg et al. 2001). The cell sorter can then pick cells, referred to as events, which match specific properties and separate these from the other cells. Multiple cell types can be separated into different collection tubes in this way. Information of the properties of each cell that passes through the cell sorter can, as in this case, also be used to determine the proportion of a specific cell type within an entire population.

For plant cells to be sorted they must first be unicellular and so protoplasts were produced from adult rosette leaves from five week old (35 dpg) plants from lines 142, 142S, M1 and M9. The majority of these protoplasts will be derived from mesophyll cells. Initially the exact properties for unsilenced and silenced protoplasts were not known. To identify the properties that denote silenced and unsilenced protoplasts, cell sorting of protoplasts from 142 and 142S was completed. In 142 all cells are GFP fluorescent (GFP positive) so will give the properties of unsilenced protoplasts, while in 142S the majority of cells are

silenced so lack GFP (GFP negative) and will give the properties for silenced protoplasts.

For cell sorting of the protoplast samples four properties were measured: frontal scatter (fsc), side scatter (ssc), GFP fluorescence and chlorophyll fluorescence. fsc is a measure of cell size whereas ssc is a measure of the granularity of cytoplasm. Granularity is a measure of 'complexity' of the cells or how many organelles are found in the cytoplasm, the more organelles the higher the ssc measurement. GFP fluorescence is measured by looking at the intensity of emissions from each cell at a wavelength of 530 nm. GFP expression and fluorescence will increase the intensity of the emissions at this wavelength. Chlorophyll fluorescence is measured in the same way as GFP except emissions at a wavelength of 670 nm are measured. GFP and chlorophyll fluorescence were used to identify silenced and unsilenced protoplasts. The fsc and ssc measurements were used to identify protoplasts since after protoplast production, as well as plant protoplasts there will also be cellular debris and undigested plant tissue present in the solution. The protoplasts can be separated from this debris using cell size and granularity. Protoplasts will be larger and more complex than cellular debris but smaller and less complex than undigested plant tissue.

First the properties of the unsilenced protoplast population were identified using 142. In order to select protoplasts from cellular debris the fsc and ssc were measured (Figure 4.18 A). The fsc and ssc of cellular debris will be lower than that of protoplasts. Therefore events with a high fsc and ssc were considered protoplasts. However, to avoid undigested tissue or droplets that contain several protoplasts, events with a very high fsc or ssc were also not selected. The fraction of the sample that contains the protoplast population is highlighted in Figure 4.18 A by the R1 box. All graphs in Figure 4.18 are gated using these properties. This means that only objects where the fsc and ssc measurements are within the area highlighted by R1 are shown in the other graphs. Within the R1 region there will still be debris as well as protoplasts, however these cannot

be separated from protoplasts using just size and granularity. In order to identify the protoplasts within this region, GFP and chlorophyll fluorescence intensity was then analysed. When histograms of intensity of GFP and chlorophyll fluorescence were plotted against each other in a dot plot a large population was identified with specific chlorophyll and GFP intensities (Figure 4.18 B). This is represented in Figure 4.18 B by the green R2 box in the enlarged section of the graph. This population is likely to be the GFP positive/unsilenced protoplast population and so gives the properties of unsilenced protoplasts.

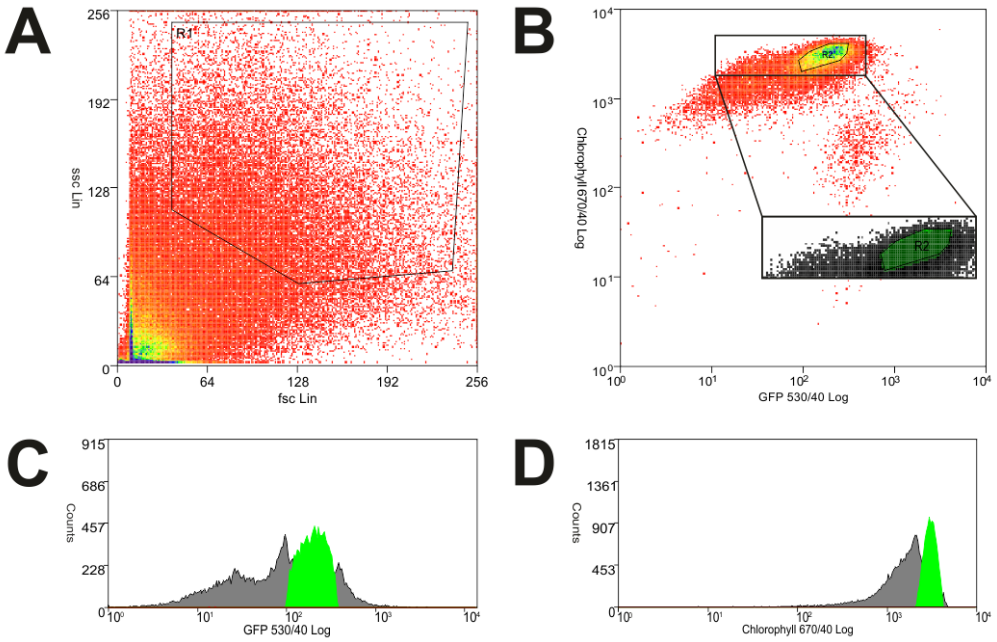


Figure 4.18: Identification of GFP positive protoplasts in line 142

Graphs showing results of flow cytometry analysis of the 142 protoplast sample. **A:** Dot plot of ssc against fsc. The values for the y axis and x axis are relative values of ssc and fsc respectively. The dots colour indicates how many events are found at this point. It ranges from red (low number events) to purple (high number events). The R1 box highlights the area of the graph where protoplasts are found. Anything below or to the left of the R1 box is likely to be cell debris, while anything above or to the right of the R1 box is likely to be multiple protoplasts still attached to each other. The R1 box acts as a gate to the other 3 graphs, so only points found within the R1 box are shown on the

other graphs. **B:** Dot plot of chlorophyll fluorescence intensity against GFP fluorescence intensity. Both axes give relative values of intensity for the two flurophores. The R2 box highlights the GFP positive protoplast population. The R2 box and surrounding area of the graph has been blown up and the R2 box is highlighted in green in this enlargement. **C:** Histogram of intensity of GFP fluorescence. **D:** Histogram of intensity of chlorophyll fluorescence. For both **C** and **D** the area of the histogram corresponding to the GFP positive protoplasts is highlighted in green. The x axes for both **C** and **D** are relative measurements of intensity.

After identifying a population of protoplasts, which is likely to be derived from GFP positive cells in 142, the GFP negative/silenced population was identified in 142S. As in 142 the protoplast population was first sorted using fsc and ssc to separate the protoplasts from cellular debris (Figure 4.19 A). The same fsc and ssc parameters as 142 were used. All other graphs were then gated with these parameters to show only events that are large and complex enough to be protoplasts. Once completed, a dot plot of GFP and chlorophyll fluorescence was plotted (Figure 4.19 B). From this graph a large protoplast population was identified, which had a different GFP intensity to the population identified in 142. This new population is represented in Figure 4.19 B by the red R3 box in the enlarged area of the graph. The R3 population made up 97.09% of the total population of protoplasts (Table 4.1). There is a smaller population of protoplasts, 2.91% total population, within the R2 box (Figure 4.19 B Table 4.1). This is to be expected as 142S has a small number of unsilenced cells. When chlorophyll and GFP intensity is compared separately between the R2 and R3 populations there is a clear difference for GFP (Figure 4.19 C and D). Chlorophyll intensity does not change between R2 and R3. This indicates a loss of GFP fluorescence in the R3 population. Thus R3 is likely to be the GFP negative protoplast population and so provides the parameters for the silenced protoplast population.

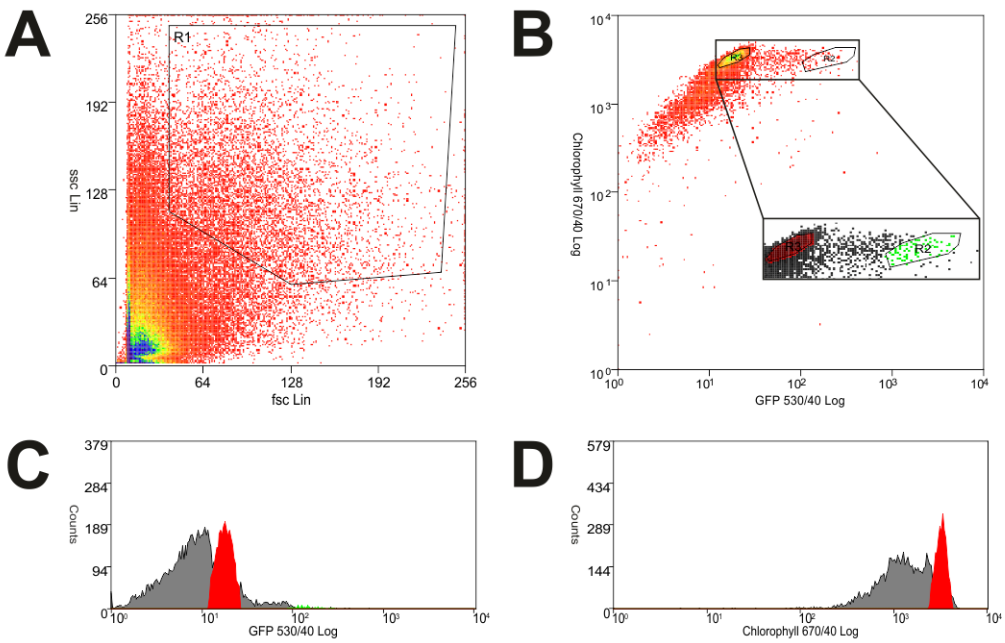


Figure 4.19: Identification of GFP negative protoplasts in line 142S

Graphs showing results of flow cytometry analysis of the 142S protoplast sample. **A:** Dot plot of ssc against fsc. The values for the y axis and x axis are relative values of ssc and fsc respectively. The dots colour indicates how many events are found at this point. It ranges from red (low number events) to purple (high number events). The R1 box highlights the area of the graph where protoplasts are found. Anything below or to the left of the R1 box is likely to be cell debris, while anything above or to the right of the R1 box is likely to be multiple protoplasts still attached to each other. The R1 box acts as a gate to the other 3 graphs, so only points found within the R1 box are shown on the other graphs. **B:** Dot plot of chlorophyll fluorescence intensity against GFP fluorescence intensity. Both axes give relative values of intensity for the two flurophores. The R2 box highlights the GFP positive protoplast population. The R3 box highlights the GFP negative protoplast population. This area of the graph has been blown up. **C:** Histogram of intensity of GFP fluorescence. **D:** Histogram of intensity of chlorophyll fluorescence. For both **C** and **D** the areas of the histograms corresponding to the GFP positive protoplasts is highlighted in green while the areas of the histograms corresponding to the GFP negative protoplasts is highlighted in red. The x axes for both **C** and **D** are relative measurements of intensity. In **D** the GFP positive and negative protoplast

populations overlap each other so the GFP positive population is hidden by the negative population.

In order to confirm that R2 is the GFP positive population and R3 is the GFP negative population, a mixed protoplast population derived from both 142 and 142S was used. The mixed population of 142 and 142S was made by mixing 1 ml of the 142 protoplast solution with 1 ml of the 142S protoplast solution. This should give a protoplast population where approximately half are GFP positive and half are GFP negative. The sample was first sorted and gated using fsc and ssc and the then a dot plot of GFP and chlorophyll fluorescence intensity was plotted (Figure 4.20 B). This produced two large populations of protoplasts in the R2 and R3 boxes. R3 again had reduced GFP intensity compared to R2, but chlorophyll intensity was unchanged between the two populations. Out of the total number protoplasts in R2 and R3, 48.27% were from the R2 population and 51.73% were from the R3 population (Table 4.1). The two populations are therefore roughly equal to each other. This suggests that the GFP positive and GFP negative populations have been correctly identified.

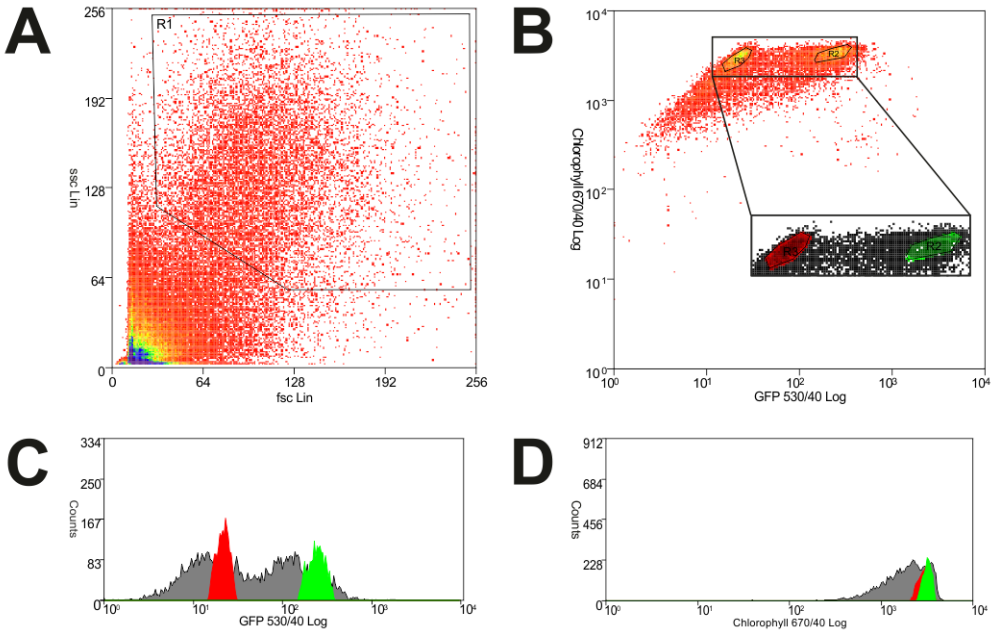


Figure 4.20: Confirmation of identification of GFP positive and negative protoplasts using mixed 142 and 142S sample

Graphs showing results of flow cytometry analysis of the 142S protoplast sample. **A**: Dot plot of ssc against fsc. The values for the y axis and x axis are relative values of ssc and fsc respectively. The dots colour indicates how many events are found at this point. It ranges from red (low number events) to purple (high number events). The R1 box highlights the area of the graph where protoplasts are found. Anything below or to the left of the R1 box is likely to be cell debris, while anything above or to the right of the R1 box is likely to be multiple protoplasts still attached to each other. The R1 box acts as a gate to the other 3 graphs, so only points found within the R1 box are shown on the other graphs. **B**: Dot plot of chlorophyll fluorescence intensity against GFP fluorescence intensity. Both axes give relative values of intensity for the two flurophores. The R2 box highlights the GFP positive protoplast population. The R3 box highlights the GFP negative protoplast population. This area of the graph has been blown up. **C**: Histogram of intensity of GFP fluorescence. **D**: Histogram of intensity of chlorophyll fluorescence. For both **C** and **D** the areas of the histograms corresponding to the GFP positive protoplasts is highlighted in green while the areas of the histograms corresponding to the GFP negative protoplasts is highlighted in red. The x axes for both **C** and **D** are

relative measurements of intensity. In D the GFP positive and negative protoplast populations overlap each other.

For final proof that the R2 and R3 populations were indeed the GFP negative and GFP positive protoplast populations the R2 and R3 populations were separated using the cell sorter. Protoplasts from the R2 and R3 fractions were assessed under a confocal microscope for GFP fluorescence. Protoplasts in the R3 population had chlorophyll fluorescence but lacked GFP fluorescence (Figure 4.21 A). Protoplasts from the R2 population had both GFP and chlorophyll fluorescence Figure 4.21 B). This confirms that the R2 population is the GFP positive protoplast population and the R3 population is the GFP negative protoplast population.

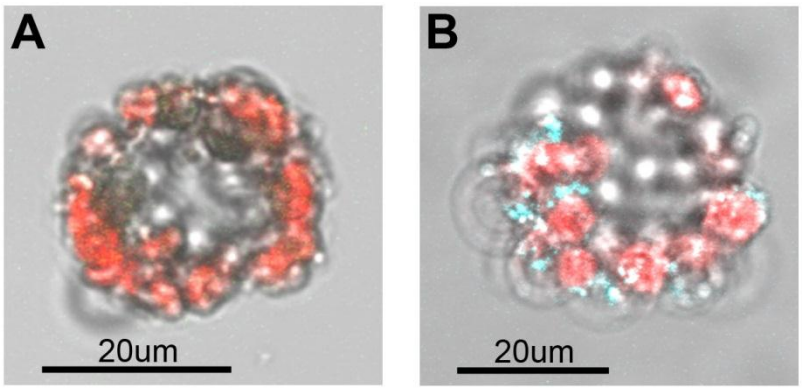


Figure 4.21: Confocal images confirming the identification of GFP positive and negative protoplasts in a mixed population

Confocal images of a single GFP negative (**A**) and GFP positive (**B**) protoplast. Images are a composite of a bright field image of a protoplast overlayed with a chlorophyll channel (red) and GFP channel (cyan). The scale bars for each image are shown. Both **A** and **B** come from a mixed population of 142 and 142S protoplasts. Images are representative for each fraction.

Having proven that GFP positive and GFP negative protoplasts could be correctly identified by cell sorting the results from the cell sorter could then be used to

assess the ratio of GFP positive (unsilenced) to GFP negative (silenced) protoplasts in the M1 and M9 mutants. Graphs showing fsc, ssc and chlorophll and GFP fluorescence for M1 and M9 are shown in Appendix 2.4 (page 400). The number of GFP positive and GFP negative protoplasts and their relative proportions are given in Table 4.1. There is a clear difference between M1 and M9. In M9 53.03% of the protoplasts are GFP positive whereas for M1 the value is 82.64%. These values do not match with what is observed in the upper epidermis, where the majority of cells are GFP positive, but are a closer match to what is observed in the lower epidermis where there are similar numbers of GFP negative and GFP positive cells in M9 and more GFP positive cells than negative cells in M1. However, as stated previously, the majority of the protoplasts in the M1 and M9 samples will be derived from mesophyll cells, so the percentages of GFP positive and negative protoplasts should be similar to that of levels of silencing in the mesophyll layers.

From confocal and fluorescence microscope observations the majority of cells in the palisade mesophyll layer are GFP positive whereas the spongy mesophyll layer has both GFP positive and GFP negative cells. As an average of both these cell layers the percentage of GFP negative and GFP positive protoplasts are similar to the observed phenotype. As each protoplast sample was derived from forty leaves from different plants it should also average out the difference in levels of silencing between leaves that emerged at different times. This again shows that the M1 is a more severe allele of the *MORC6* than M9.

Sample	Number of GFP positive protoplasts	Number of GFP negative protoplasts	Total number of protoplasts	Percentage protoplasts GFP positive (%)	Percentage protoplasts GFP negative (%)
Mixed sample 142 and 142S	1777	1904	3681	48.27	51.73
142	11001	9	11010	99.92	0.08
142S	8106	270300	278406	2.91	97.09
M1	810300	170200	980500	82.64	17.36
M9	227500	201500	429000	53.03	46.97

Table 4.1: Differences in the proportion of silenced and unsilenced protoplasts in M1 and M9

Table shows the total number of GFP positive and negative protoplasts for a mixed sample of 142 and 142S, 142, 142S, M1 and M9. The mixed sample of 142 and 142S was made up of approximately equal numbers of 142 and 142S protoplasts. The percentage of the total number of protoplasts that are GFP positive and negative is also given.

4.2.8 No reduction in 35S siRNA levels in mutants

Both M1 and M9 disrupt silencing of the *GFP* transgene but it was not known which point in the RdDM pathway was affected. To determine if the mutants affected siRNA production a siRNA northern blot to assess 35S siRNAs levels was completed on RNA samples from 7 dpg and 21 dpg leaves and floral tissue. Samples were taken from lines 142, 142S, M1 and M9. The blot was first probed for 35S siRNAs and then miR167, which acts as a loading control, using radio-labelled probes. In the 142 samples the levels of 35S siRNAs were so low that they were hard to distinguish from background unspecific binding (Figure 4.22). This could suggest that 142 does produce 35S siRNAs at a very low level but not enough to trigger RdDM. For 142S, 35S siRNAs could be detected in the 21 dpg sample at low levels and at higher levels in the floral sample, but for the 7 dpg sample they were similar to 142, which may suggest that this and the 142 signal is background noise rather than a genuine band. Comparisons of the M1 and M9 21 dpg and floral samples to 142S showed no difference between 142S and M9, but there was a decrease in 35S siRNAs in the M1 floral sample. However, the amount of M1 RNA on the initial gel is lower than the other floral samples so could suggest that this is the cause of the apparent reduction in 35S siRNA levels. The miR167 loading control shows that the RNA levels were similar between the 7 dpg and 21 dpg samples but that there is a difference in the floral samples. The miR167 signal in 142 is comparable to the 7 dpg and 21 dpg samples but 142S, M1 and M9 are all lower. The reason for this could be poor transfer of RNA for these three samples or that during the stripping process to remove the 35S

probe and the Solo-LTR probe (mentioned in the next chapter page 312), before the blot was probed for miR167, also inadvertently removed RNA from these samples. For this reason a second blot with just floral samples of lines 142, 142S, M1, M9 and *rmd3* was completed. *rmd3* is a *nrpd1* mutant so should be defective in siRNA production and was indeed shown to lack 35S siRNAs (Figure 4.23). siRNAs were also detectable in 142 but at low levels so again this could be either the result of non-specific hybridisation or that 142 does produce low levels of 35S siRNAs that are not abundant enough to trigger silencing. As with the previous blot, there was a slight reduction in 35S siRNAs in M1 compared to 142S, but whereas previously 142S and M9 had similar siRNA levels, in this blot 35S siRNA levels are higher in M9. One possible reason for this difference in siRNA levels may be the fact that 24 nt siRNA levels vary depending on the stage of floral development so the stage of the flowers collected may affect abundance accordingly (Mosher et al. 2009). However, there was also a problem with the miR167 loading control as both M1 and M9 showed reduced levels compared to 142 and 142S. The fact that the miR167 loading control is not consistent for all samples in either blot makes comparisons of 35S siRNAs difficult and suggests that it may not be ideal as a control. However, since 21 dpg levels in M1 and M9 are similar to 142S and there is not a large reduction in 35S siRNAs for either M1 or M9 in floral samples it would suggest that MORC6 is not involved in siRNA production.

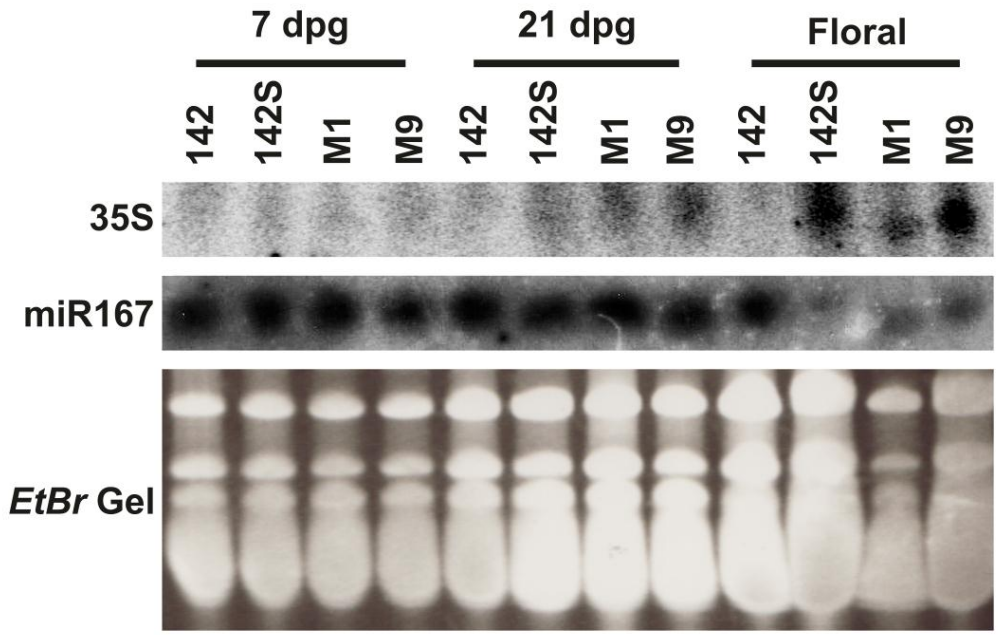


Figure 4.22: Detection of 35S siRNAs at different developmental stages in M1 and M9

Image of a northern blot probed with a ^{32}P-labelled probes for the *35S* promoter (top) and miR167 (middle). 7 dpg, 21 dpg and floral RNA samples of lines 142, 142S, M1 and M9 were used, the order of these is indicated. The bottom image is of the polyacrylamide gel stained with ethidium bromide before being blotted. This shows loading of RNA onto the gel. The miR167 is a loading control for the blot.

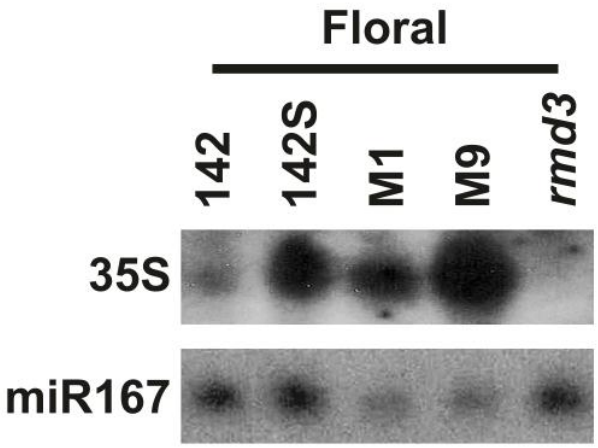


Figure 4.23: Detection of 35S siRNAs in M1 and M9 floral tissue

Image of a northern blot probed with a ^{32}P-labelled probes for the *35S* promoter (top) and miR167 (bottom). Floral RNA samples of lines 142, 142S, M1, M9 and *rmd3* were used, the order of these is indicated. The miR167 is a loading control for the blot. *rmd3* acts as a negative control as it is a mutant in RdDM siRNA production.

4.2.9 Reduction in DNA methylation of the *35S* promoter in M1 and M9

Having determined the effect of *MORC6* on abundance of 35S siRNAs, its effect on DNA methylation of the *35S* promoter was also assessed in order to resolve *MORC6* function in RdDM of the *GFP* transgene. To ascertain if there is any change in the DNA methylation profile of the *35S* promoter in M1 and M9, bisulfite sequencing was used. Bisulfite treatment of DNA changes unmethylated cytosines into uracils but does not convert methylated cytosines (Shapiro et al. 1970, Wang et al. 1980). These changes can then be detected by sequencing the *35S* promoter of bisulfite treated DNA and then comparing it against the original sequence to identify cytosine residues that have not been converted and therefore are methylated.

As both M1 and M9 have a delay in silencing establishment, bisulfite sequencing was carried out on DNA samples taken at 7 dpg and 28 dpg. However, as the time point for when silencing occurs in the mesophyll layer is unknown there may be silencing in the 7 dpg sample. DNA from rosette leaves of lines 142, 142S, M1 and M9 were bisulfite treated and the *35S* promoter sequence was amplified by PCR from the treated DNA. These PCR fragments were then cloned into the pGEM vector before being transformed into bacteria. The plasmids from at least ten individual colonies were extracted and sequenced. Plasmids from each colony should contain a *35S* promoter fragment that originates from different DNA strands. As RdDM is cell autonomous with respect to which cytosines are methylated, each DNA strand will have a different DNA methylation pattern. The methylation pattern from at least 10 DNA strands must be sequenced in order to get a consensus of DNA methylation in the rosette leaves. 142 does not silencing the *GFP* transgene so should lack DNA methylation of the *35S* promoter, so all cytosines should be converted by bisulfite treatment. As bisulfite conversion was carried out on samples from all four lines concurrently, 142 should therefore act as a control for the efficiency of bisulfite conversion. 142S will give the WT DNA methylation pattern of the *35S* promoter.

Only part of the 35S promoter was sequenced (Figure 4.24), although this section does cover all of the area targeted by the *35S IR* transgene. In order to prevent amplification of the *35S IR* transgene, the region being sequenced also extends into the CDS of the *GFP* gene so only the *35Sp:GFP* transgene will be amplified and not the *35S IR* transgene. The primers have also been designed to amplify only the lower strand. This means that only methylation on the lower strand and not the upper strand will be detected in this assay. There are ninety seven cytosine residues on the lower strand within the region amplified for bisulfite sequencing (Figure 4.24). Of these seventy seven cytosine residues lie within the section targeted by the *35S IR*. There are a further eighteen cytosine residues within the *35S* promoter sequence that have not been targeted and two within the *GFP* CDS. It is possible that methylation will spread from the area targeted by

the *35S IR* and cause the methylation of these cytosines. Several studies have shown evidence of spread of methylation in both a transgene and endogenous environment (Daxinger et al. 2009, Ahmed et al. 2011).

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

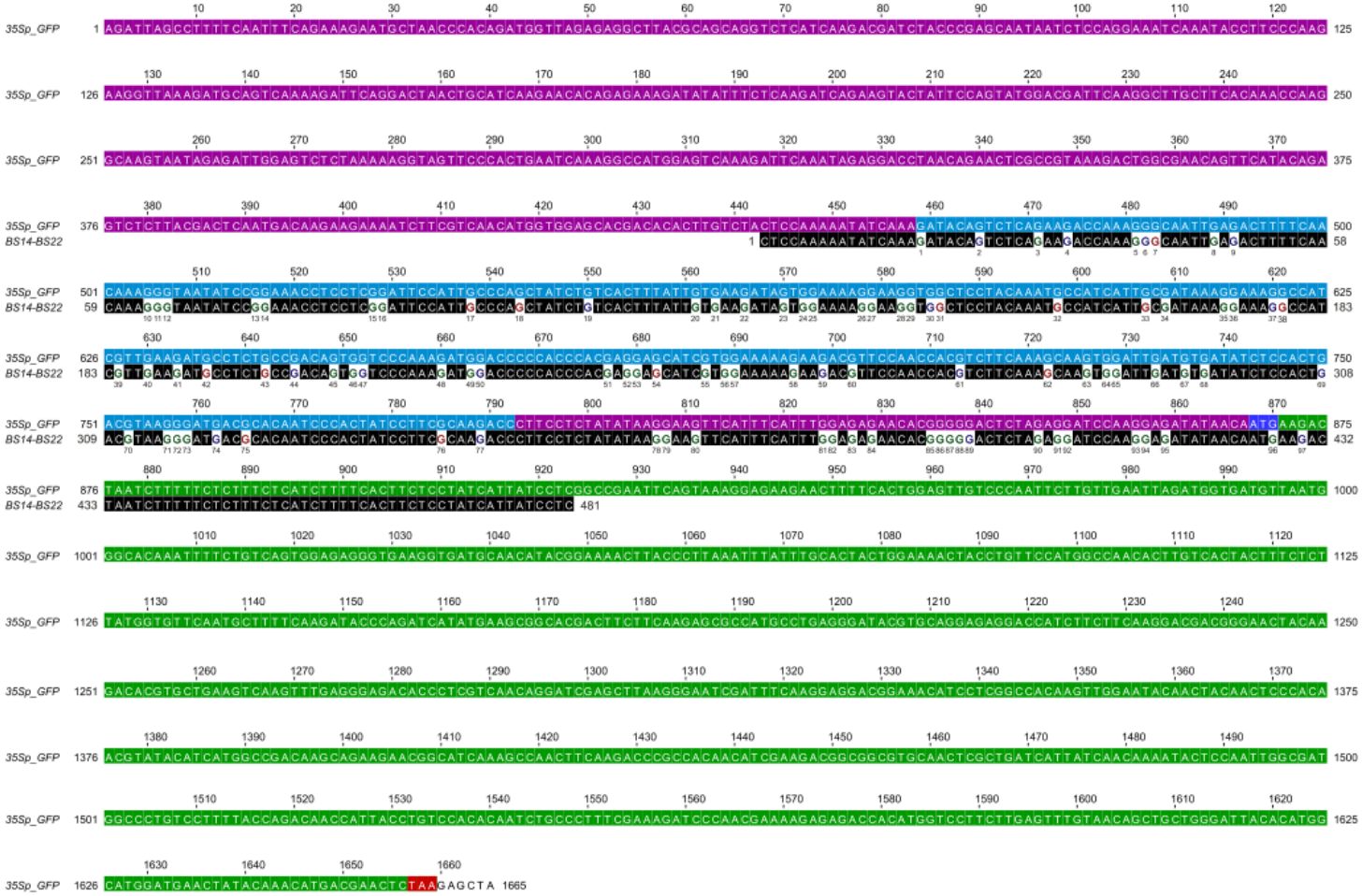


Figure 4.24: Region of 35Sp:GFP transgene being analysed by bisulfite sequencing

Figure shows the complete sequence of the 35Sp:GFP transgene (top line, named 35Sp_GFP) and the region of the transgene being sequenced (bottom line, named BS14-BS22). This diagram shows the sequence of the upper strand. The position in the sequence for start and end base for each line is given and every 10 bps of the 35Sp:GFP transgene is marked above the sequence. The 35S promoter is shown in purple and cyan. The area highlighted in cyan is the region of the 35S promoter targeted by the *35S IR* transgene. The GFP CDS is shown in green with the start codon in blue and the stop codon in red. The region being sequenced is in black and spans both the 35S promoter and part of the GFP CDS. Guanine residues, in the region being sequenced have been highlighted. Their sequence contexts are given, red means CG, blue means CHG and green means CHH, where H stands for C, T or A. The guanine residues are numbered from 1 to 97 below each residue.

35S promoter sequences were analysed using the online CyMATE tool, which presents results diagrammatically (Figure 4.25) (Hetzl et al. 2007). This tool is able to identify methylated and unmethylated cytosines. The DNA methylation pattern of 7 dpg samples was assessed first. In the 7 dpg 142 sample there was a maximum of three unconverted cytosines per sequence, indicating a good rate of bisulfite conversion (Figure 4.25 A). The maximum number of unconverted cytosine residues per sequence in 142 was used to determine if sequences in 142S, M1 and M9 were methylated or unmethylated. In this case sequences with more than three unconverted cytosine residues were considered to be methylated and sequences with three or less were considered to be unmethylated.

Methylation in the 142S at 7 dpg is restricted to the seventy seven cytosine residues within the region targeted by the *35S IR* (Figure 4.25 B and Figure 4.26 B). This indicates that in this system DNA methylation does not spread from the region targeted by siRNAs. Within the methylated region there is variation in levels of DNA methylation between sequences. In some sequences nearly all cytosines are methylated while in other sequences the majority of cytosine residues are unmethylated. One sequence is unmethylated. The presence of an unmethylated sequence matches data from the cell sorting and confocal microscopy, in that there is a small subset of unsilenced cells in 142S. As the population of unsilenced cells is small it is likely that the sequences with low levels of methylation are silenced and therefore that small numbers of methylated cytosines are sufficient for silencing to occur in this transgene system. Both M1 and M9 have two groups of sequences (Figure 4.25 C and D). One group lacks methylation and the other maintains methylation. The number of methylated sequences is smaller in both M1 and M9 than the unmethylated sequences, although for M1 as well as the three methylated sequences there are two further sequences with just four methylated cytosine residues (Figure 4.25 C). It is unclear whether these two sequences are methylated at a very low level or the cytosines have failed to be converted during bisulfite treatment. Amongst

the group of methylated sequences, levels of methylation are low, as for all sequences there were more unmethylated than methylated cytosine residues. In most cases methylation of cytosines residues was unique to one sequence, with very few cases of more than two sequences methylating the same cytosine residue (Figure 4.26 C and D).

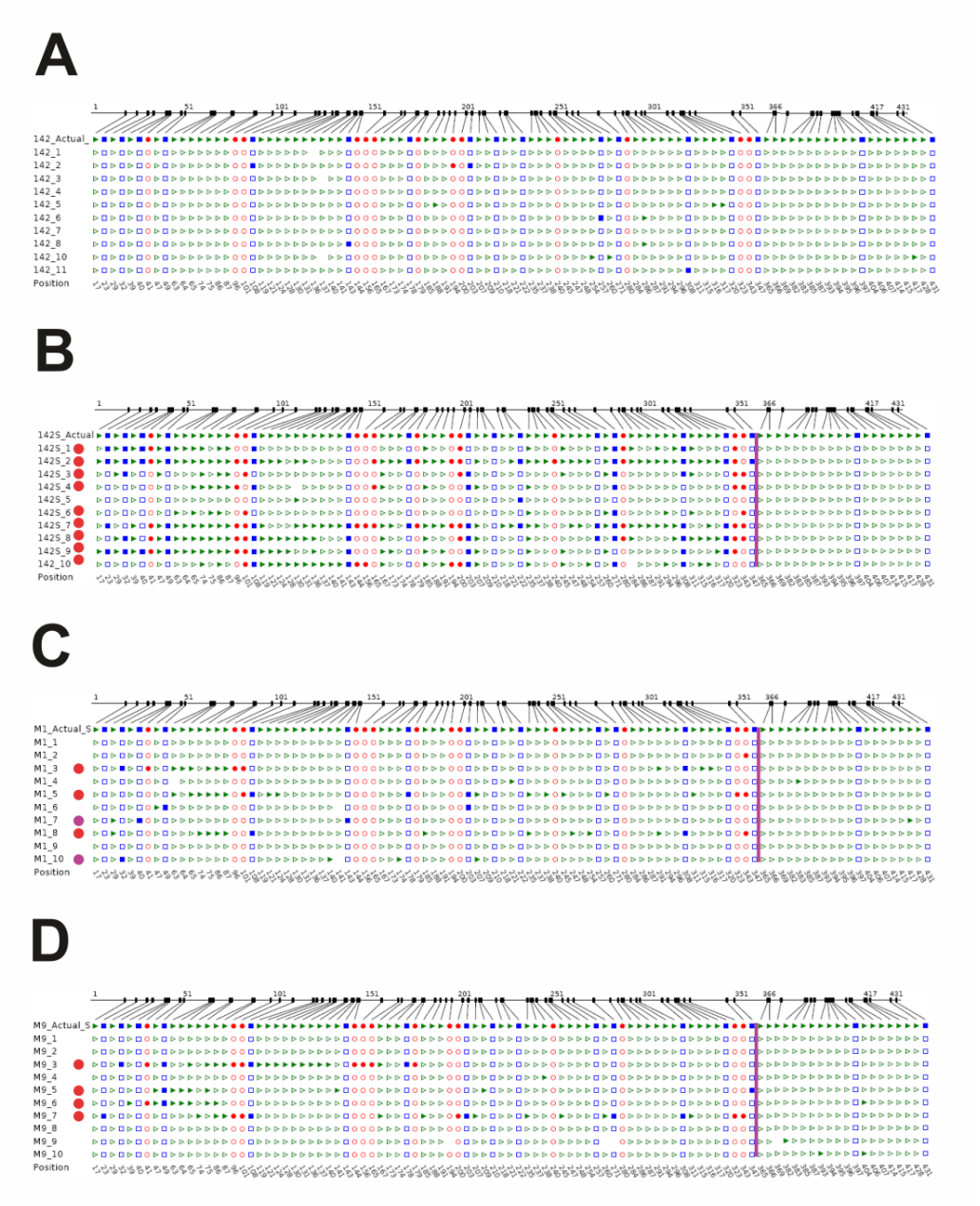


Figure 4.25: Presence of methylated and unmethylated sequence in M1 and M9 at 7 dpg

Alignment of bisulfite sequences to the unconverted sequence for 7 dpg samples from lines 142 (**A**), 142S (**B**), M1 (**C**) and M9 (**D**). The top sequence, where all shapes are filled in, is the unconverted sequence. The black lines show where each cytosine residue is found in the sequence. Red circles are CG sites, the blue squares are CHG sites and green triangles are CHH sites. If the shapes are filled this indicate that cytosine is methylated. At the beginning of the 142S (**B**) M1 (**C**) and M9 (**D**) sequences red circles indicate that this sequence is methylated while purple circles indicate that only four cytosines are methylated. Any cytosine residue to the left of the purple vertical line in the 142S (**B**) M1 (**C**) and M9 (**D**) alignments is targeted by the *35S IR* transgene while cytosines to the right are not targeted.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

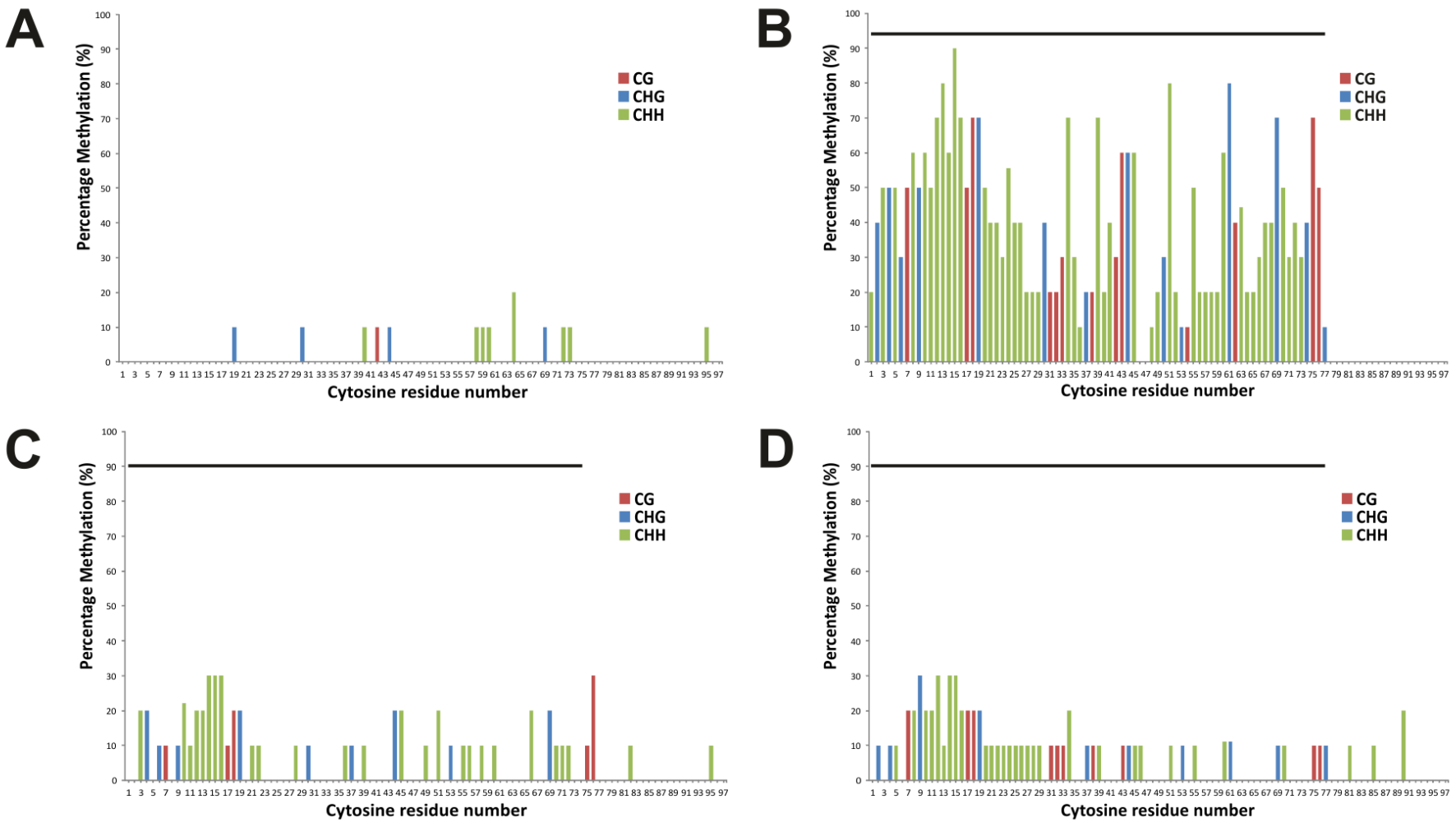


Figure 4.26: Reduction in methylation of the 35S promoter methylation at 7 dpg in M1 and M9

Bar charts of percentage methylation at each cytosine residue in 7 dpg samples from lines 142 (**A**), 142S (**B**), M1 (**C**) and M9 (**D**). The x shows the position of the cytosine residues in the sequence. The y axis shows percentage methylation. The colour of the bar corresponds to the sequence context of the cytosine, red is CG, blue is CHG and green is CHH. The black line above the bars represents the region targeted by the *35S IR* transgene.

Further analysis of methylation at 7 dpg shows that although the percentage methylation of CG, CHG and CHH is reduced, the proportion of each type of methylation is not significantly altered in M1 or M9 compared to 142S (Figure 4.27 A and B). Multi-dimensional scaling (MDS) analysis has been used to demonstrate that the sequences considered methylated are significantly different from those considered to be unmethylated and that the methylated cytosines are mostly unique to single methylated sequences. MDS calculates how different each sequence is from another sequence in two dimensions. This was done by looking at whether the methylation status for each individual cytosine residue is different or the same in the two sequences. MDS then calculates the differences in the two dimensions for every pair of sequences. The sequences can then be placed on a scatter plot where the x and y axis distance between each sequence represents the calculated difference between the two sequences in dimension 1 and dimension 2, respectively. These calculated differences are relative, so different MDS analyses cannot be compared directly to each other. In 142S the sequences are widely distributed (Figure 4.27 C). This is because there is variation in the methylation pattern between each sequence and this translates into large differences in dimensions 1 and 2. However, for both M1 and M9 there are two clear groups, as the sequences considered unmethylated cluster together while the sequences that maintain methylation do not cluster together but are clearly separated from the unmethylated sequences. The fact that the methylated sequences do not cluster shows that cytosine methylation in each sequence is mostly unique to that sequence. MDS provides statistical evidence that the identification of two groups is robust.

The presence of methylated and unmethylated sequences is unexpected, as the *morc6* mutants previously characterised have either no effect on DNA methylation or a minor decrease in methylation levels and therefore that DNA methylation in M1 and M9 would be unaffected or show a modest reduction (Lorković et al. 2012, Moissiard et al. 2012). The loss of DNA methylation in the majority of sequences for M1 and M9 would therefore suggests that MORC6 in

fact, does have a role in DNA methylation, although this role could be specific to the transgene locus. The presence of methylated sequences is also unusual, as at 7 dpg, GFP fluorescence and GFP mRNA levels in M1 and M9 are comparable to 142, suggesting that silencing does not occur at this stage. The cells with methylated sequences are therefore not responding to the methylation of the *35S* promoter, but could be the cells that silence at a later stage. As methylation is present at this point the delay could be due to a failure of MORC6's role in higher order heterochromatin formation in the M1 and M9 mutants. These results would therefore suggest a dual role for MORC6 in RdDM, as both a chromatin modifier and a role in DNA methylation itself. However, as the onset of silencing has not been visualised in all tissues it is possible that some level of silencing does occur at this stage and that the methylated sequences may belong to cells that are already silencing.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

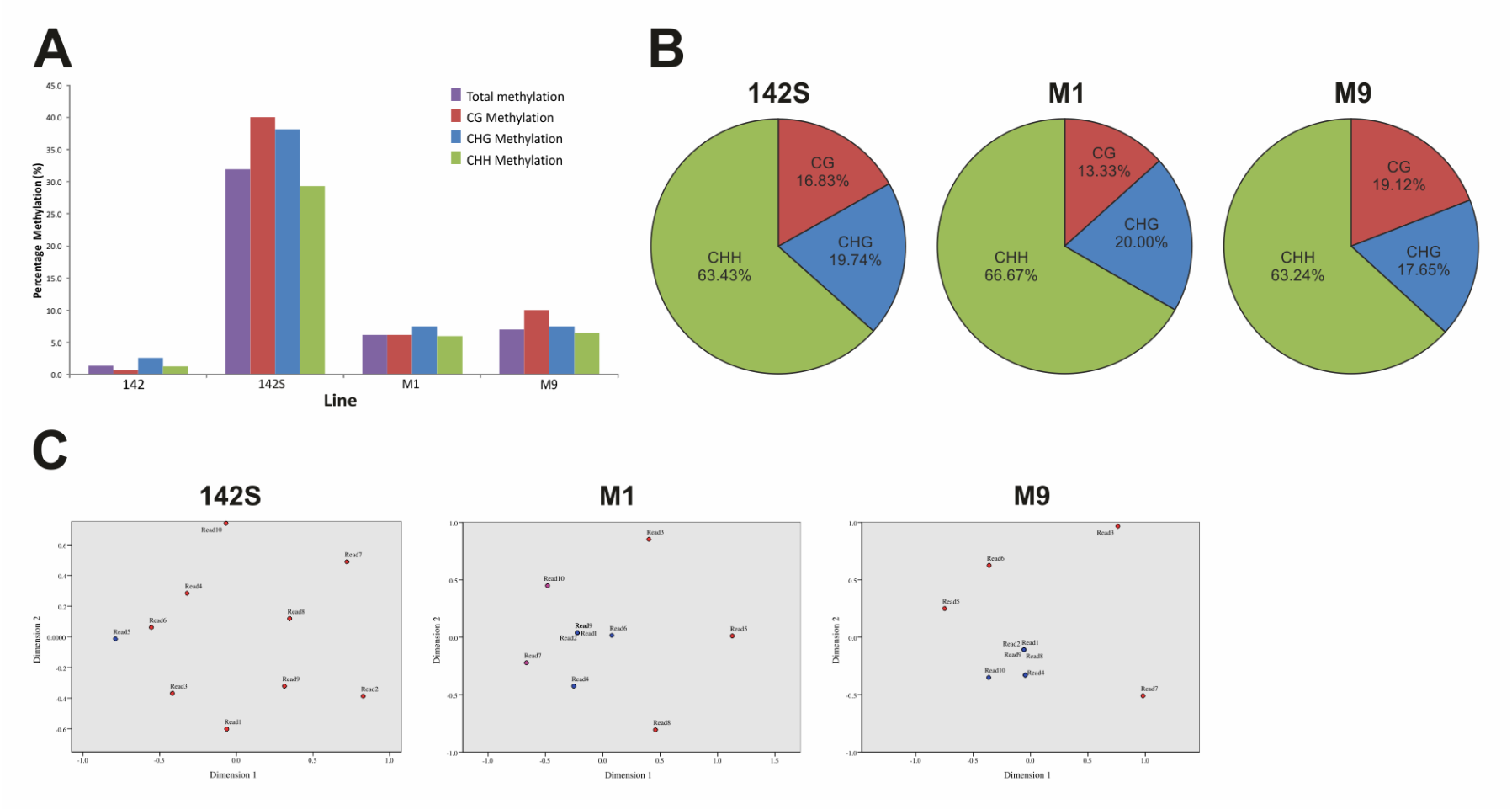


Figure 4.27: Further analysis of the loss of methylation in M1 and M9 at 7 dpg

A: Bar chart giving the proportions of total cytosines in all (purple), CG (red), CHG (blue) and CHH (green) configurations that are methylated. The x axis shows the plant line the sequences come from. The y axis is percentage methylation. **B:** Pie charts for lines 142S, M1 and M9 that show the proportion of the total methylated cytosines that are CG (red), CHG (blue) and CHH (green). The percentage of the total methylated cytosines is also given for each segment of the chart. **C:** Scatter plots of MDS analysis results for lines 142S, M1 and M9. The x axis is dimension 1 and the y axis is dimension 2, both these scales are relative values for difference between sequences. Red dots are methylated sequences, purple dots are sequences with only four methylated cytosines and blue dots are those that are assigned as being unmethylated sequences. The read numbers correspond to those in Figure 4.25 for 142S, M1 and M9. The positions of several read on the plots in M1 and M9 overlap, this is indicated by several labels surrounding a dot.

DNA methylation of 28 dpg DNA samples were also assessed. At this point silencing has occurred in all leaf tissues. The 142 control showed good conversion of cytosines with a maximum of one unconverted cytosine per sequence (Figure 4.28 A). This value was used as the cut off for methylated and unmethylated sequences for the 142S, M1 and M9 28 dpg samples. As seen in the 7 dpg sample, levels of methylation in 142S vary between sequences and there are a small number of unmethylated sequences (Figure 4.28 B). Unlike in the 7 dpg sample, there was a hypermethylated sequence where methylation extends into the region of the *35S* promoter not targeted by the *35S IR* transgene. As the 142 control shows good conversion this is unlikely to be caused by failure of cytosine conversion but instead it suggests that in a small number of cells in this transgene system, methylation can spread from the siRNA target site. DNA methylation in 142S may also not be stochastic, but instead specific cytosine residues are favourably methylated. For cytosines where 50% or more of the sequences are methylated at that residue, there were twenty four residues in the 28 dpg 142S sample and thirty one residues in the 7 dpg 142S sample, of which, twenty three are found in both samples (Figure 4.26 B, Figure 4.29 B and Table 4.2). The reason why these cytosines are strongly methylated is unclear, but could be connected to sequence preference by DRM2 or related to siRNA levels. Methylation in the 142S line would also appear to be higher for cytosines located within the first 140 bp of the part of the *35S* promoter that has been sequenced (Figure 4.26 B Figure 4.29 B). Again the reason for why there is this preference is unknown.

Site	Cytosine residue configuration	Percentage Methylation 7 dpg (%)	Percentage Methylation 28 dpg (%)
9	CHG	50.00	53.85
10	CHH	60.00	76.92
11	CHH	50.00	61.54
12	CHH	70.00	69.23
13	CHH	80.00	69.23
14	CHH	60.00	69.23

Site	Cytosine residue configuration	Percentage Methylation 7 dpg (%)	Percentage Methylation 28 dpg (%)
15	CHH	90.00	76.92
16	CHH	70.00	61.54
17	CG	50.00	61.54
18	CG	70.00	61.54
19	CHG	70.00	69.23
34	CHH	70.00	69.23
39	CHH	79.00	61.54
44	CHG	60.00	69.23
45	CHH	60.00	53.85
51	CHH	80.00	53.85
55	CHH	50.00	69.23
60	CHH	60.00	69.23
61	CHG	80.00	61.54
69	CHG	70.00	69.23
70	CHH	50.00	61.54
75	CG	70.00	53.85
76	CG	50.00	61.54

Table 4.2: Cytosine residues with 50% or more methylation in 142S

Table shows the cytosine residues with 50% or more methylation that are found in both the 7 dpg and 28 dpg 142S methylation profiles. The site number refers to the cytosine number given to each cytosine in Figure 4.24. The configuration of each cytosine is given. The percentage methylation in the 7 dpg and 28 dpg sample is given.

As seen at 7 dpg, there are two groups of sequences in M1 and M9, the largest being unmethylated sequences with a smaller group of methylated sequences (Figure 4.28 C and D). There are four methylated sequences in M1 and two in M9, but there are also three sequences in M1 and two in M9 where there are two or three methylated cytosine residues. It is unclear whether these sequences are actually methylated or whether it is cytosine conversion failure, particularly as they have the same level of cytosine methylation as 7 dpg 142 control sequences. Of the sequences that are clearly methylated, both M9 methylated sequences have low levels of methylation, at the lower end of what is seen in methylated 142S sequences (Figure 4.28 D and Figure 4.29 D) and in M1 four of the five methylated sequences also have low levels of methylation, while the

fifth sequence has high levels of methylation, comparable to highly methylated 142S sequences (Figure 4.28 C and Figure 4.29 C). As methylation in both the 7 dpg and 28 dpg M1 and M9 samples is low, bar one sequence, this would suggest that DNA methylation, in cells that do methylate the *35S* promoter, is reduced and therefore that DNA methylation by RdDM is also impaired in these cells. If WT methylation was maintained in this subsection of cells it would be expected that more cells would have higher levels of DNA methylation.

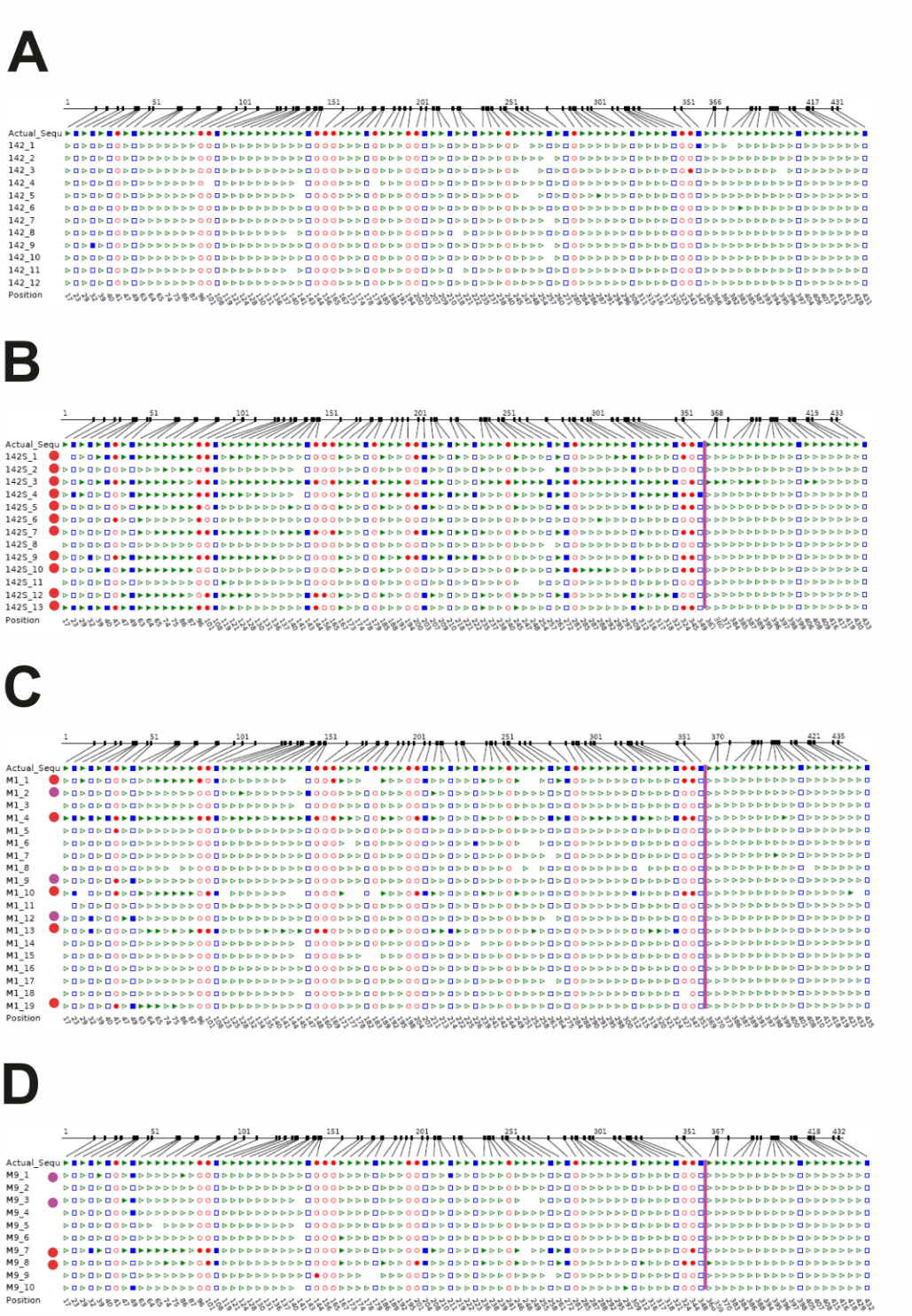


Figure 4.28: Presence of methylated and unmethylated sequence in M1 and M9 at 28 dpg

Alignment of bisulfite sequences to the unconverted sequence for 28 dpg samples from lines 142 (**A**), 142S (**B**), M1 (**C**) and M9 (**D**). The top sequence, where all shapes are filled in, is the unconverted sequence. The black lines show where each cytosine residue is found in the sequence. Red circles are CG sites, the blue squares are CHG sites and green triangles are CHH sites. If the shapes are filled this indicate that cytosine is methylated. At the beginning of the 142S (**B**) M1 (**C**) and M9 (**D**) sequences red circles indicate that this sequence is methylated while purple circles indicate that only two or three cytosines are methylated. Any cytosine residue to the left of the purple vertical line in the 142S (**B**) M1 (**C**) and M9 (**D**) alignments is targeted by the *35S IR* transgene while cytosines to the right are not targeted.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

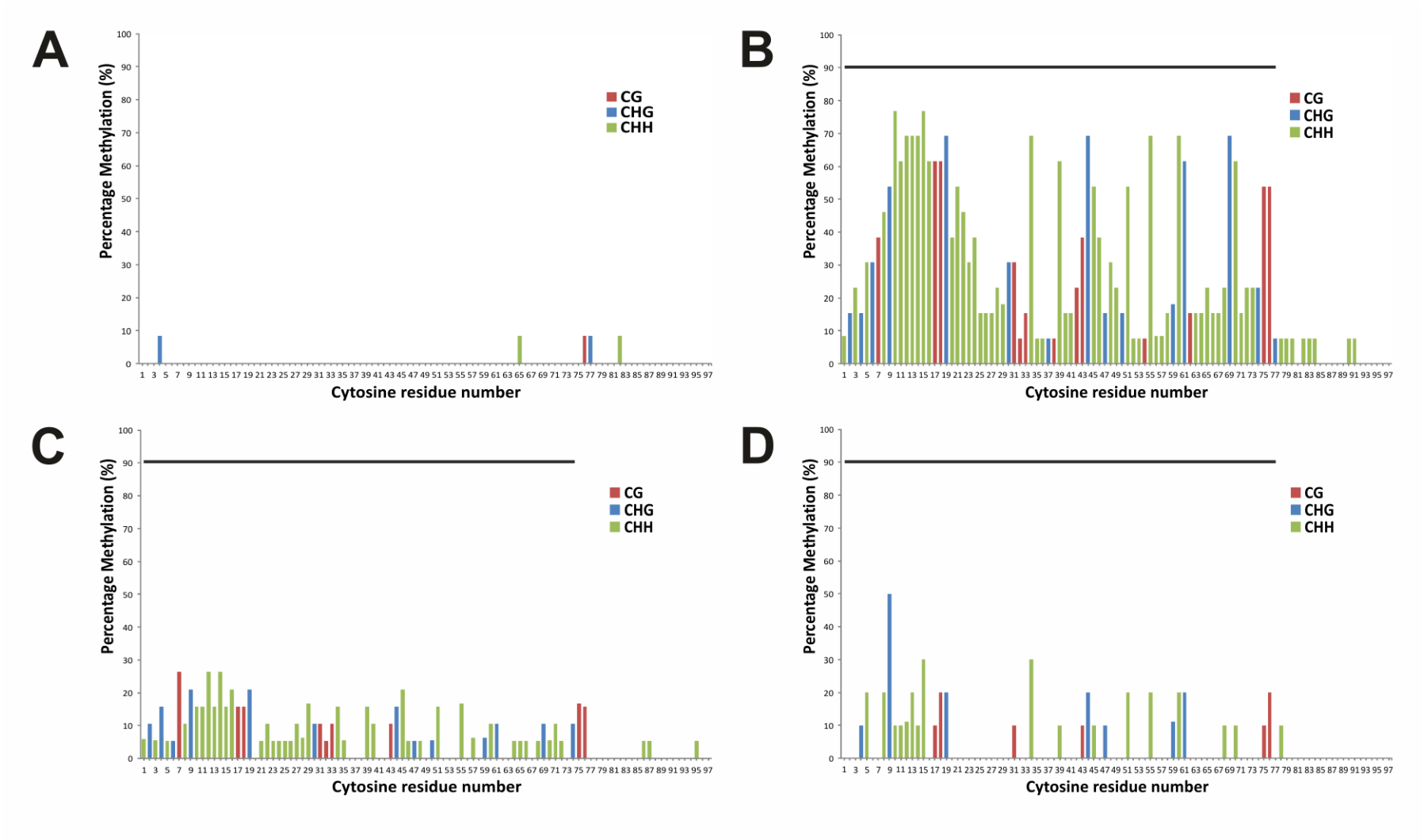


Figure 4.29: Reduction in methylation of the 35S promoter methylation at 28 dpg in M1 and M9

Bar charts of percentage methylation at each cytosine residue in 7 dpg samples from lines 142 (**A**), 142S (**B**), M1 (**C**) and M9 (**D**). The x axis shows the position of the cytosine residues position in the sequence. The y axis shows percentage methylation. The colour of the bar corresponds to the sequence context of the cytosine, red is CG, blue is CHG and green is CHH. The black line above the bars represents the region targeted by the *35S IR* transgene.

The proportion of CG, CHG and CHH residues, which are methylated does not change significantly in M1 (Figure 4.30 B). However, there is a shift in the proportions of CHG and CHH methylation in M9, in that there is an increase in CHG and decrease in CHH. although this change may be caused by the relatively small number of sequences analysed. To confirm that there are two groups within the M1 and M9 samples and that cytosine residue methylation is unique to a single sequence, MDS analysis was again used. MDS shows there are two clear groups for both M1 and M9, with unmethylated sequences clustering together (Figure 4.30 C). The methylation sequences do not group together, but are clearly separate from the unmethylated sequences. This lack of grouping is again due to the unique methylation pattern for each sequence. As there are unmethylated sequences in both the M1 and M9 7 dpg and 28dpg samples and methylated sequences have low levels of methylation, this would support a role for MORC6 in DNA methylation, at least for this particular locus.

4. Silencing phenotype of the 35Sp:GFP transgene system in the MORC6 mutants

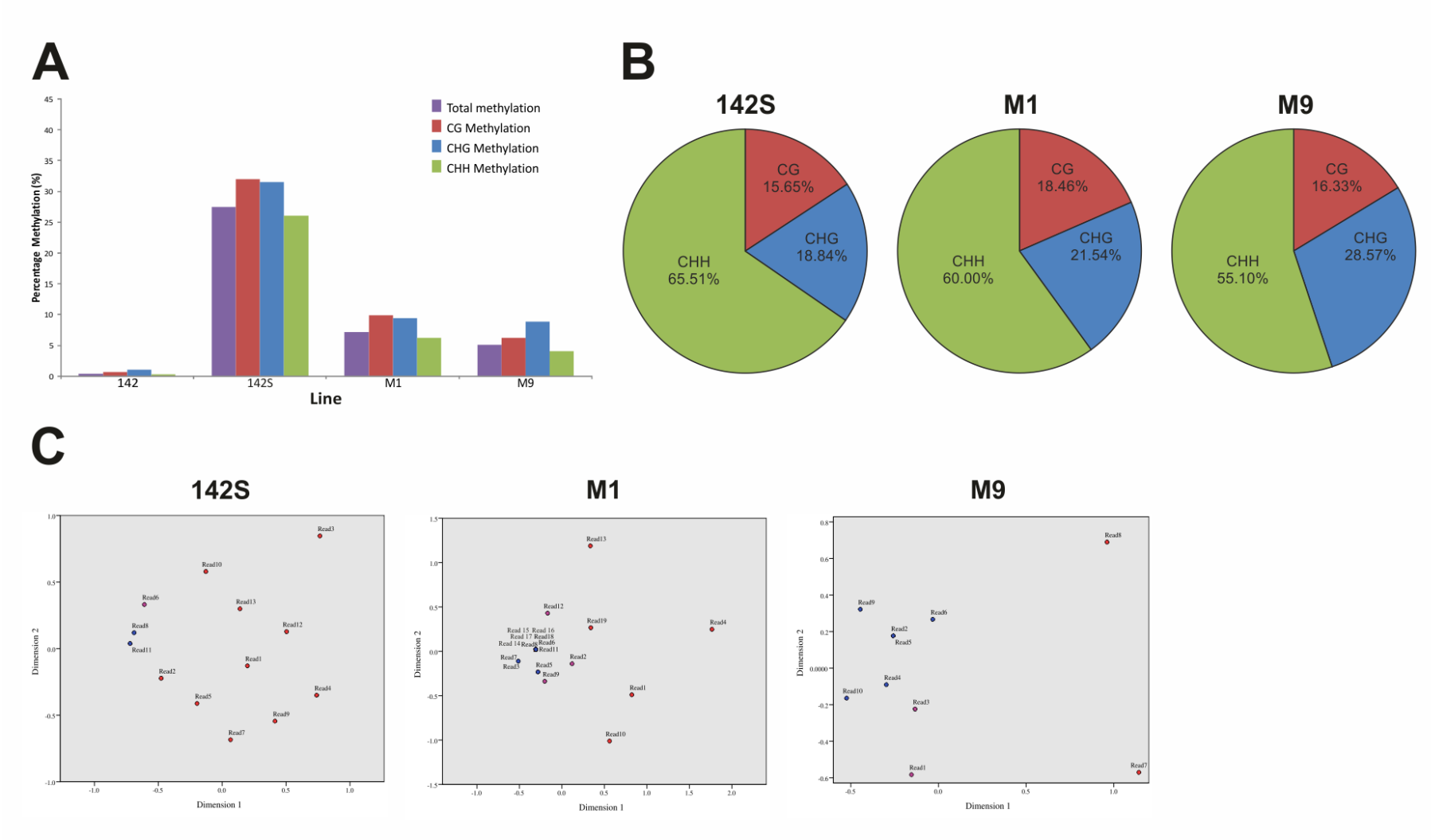


Figure 4.30: Further analysis of the loss of methylation in M1 and M9 at 28 dpg

A: Bar chart giving the proportions of total cytosines in all (purple), CG (red), CHG (blue) and CHH (green) configurations that are methylated. The x axis shows the plant line the sequences come from. The y axis is percentage methylation. **B:** Pie charts for lines 142S, M1 and M9 that show the proportion of the total methylated cytosines that are CG (red), CHG (blue) and CHH (green). The percentage of the total methylated cytosines is also given for each segment of the chart. **C:** Scatter plots of MDS analysis results for lines 142S, M1 and M9. The x axis is dimension 1 and the y axis is dimension 2, both these scales are relative values for difference between sequences. Red dots are methylated sequences, purple dots are sequences with only two or three methylated cytosines and blue dots are those that are assigned as being unmethylated sequences. The read numbers correspond to those in Figure 4.28 for 142S M1 and M9. The positions of several reads on the plots in M1 and M9 overlap, this is indicated by several labels surrounding a dot.

It is expected that the sequences lacking DNA methylation of the *35S* promoter come from cells that express GFP and that sequences that were methylated come from cells in which GFP is silenced. In order to confirm this, cell sorting was used to separate silenced cells from unsilenced cells, the identification of which has been discussed previously (page 241). DNA was then extracted from the separate fractions, but only the M1 unsilenced protoplast sample had a sufficient number of protoplasts to produce DNA of high enough concentration for bisulfite sequencing (Table 4.1). The DNA from this sample was at low concentration, so to prevent loss of DNA, the DNA was not cleaved by restriction enzymes before bisulfite treatment. 142 DNA extracted from whole leaves was used as a control for bisulfite conversion. Sequencing of the *35S* promoter from bisulfite treated DNA showed good conversion of cytosines with most sequences having a maximum number of three unconverted cytosine residues per sequence, however there was one 142 sequence, which had nine unconverted cytosines (Figure 4.31 A). This may be due to lack of bisulfite conversion, possibly due to the different DNA preparation technique used for this bisulfite conversion. It is also possible that a minority of cells in 142 methylate the *35S* promoter and so it cannot be ruled out that the sequence is in fact methylated.

In the M1 unsilenced DNA sample there was little evidence for DNA methylation of the *35S* promoter, eight sequences have three or less unconverted cytosine residues per sequence, which is the same as the 142 sequences (Figure 4.31 B). There are however one sequence with four unconverted cytosine residues and another with six unconverted residues, although this is still less than the 142 sequence with the most unconverted cytosines. Therefore the unconverted cytosines could be caused by a failure of conversion during bisulfite treatment, although methylation cannot be discounted. If these two sequences are methylated, their level of cytosine methylation is low, particularly when compared to levels in methylated M1 sequences from the 7 dpg and 28 dpg time points, which usually have between 15-25 methylated cytosines in each sequence (Figure 4.25 C and Figure 4.28 C). It is therefore likely that this level of

DNA methylation is insufficient for silencing of the *35S* promoter. Although DNA methylation analysis of the silenced M1 protoplasts is lacking, it can be concluded from the lack of DNA methylation in the unsilenced protoplasts that DNA methylation should be restricted to the silenced protoplasts. However, it is unknown whether all silenced protoplasts are methylated and would require bisulfite sequencing of a silence sample in order to determine if this is the case.

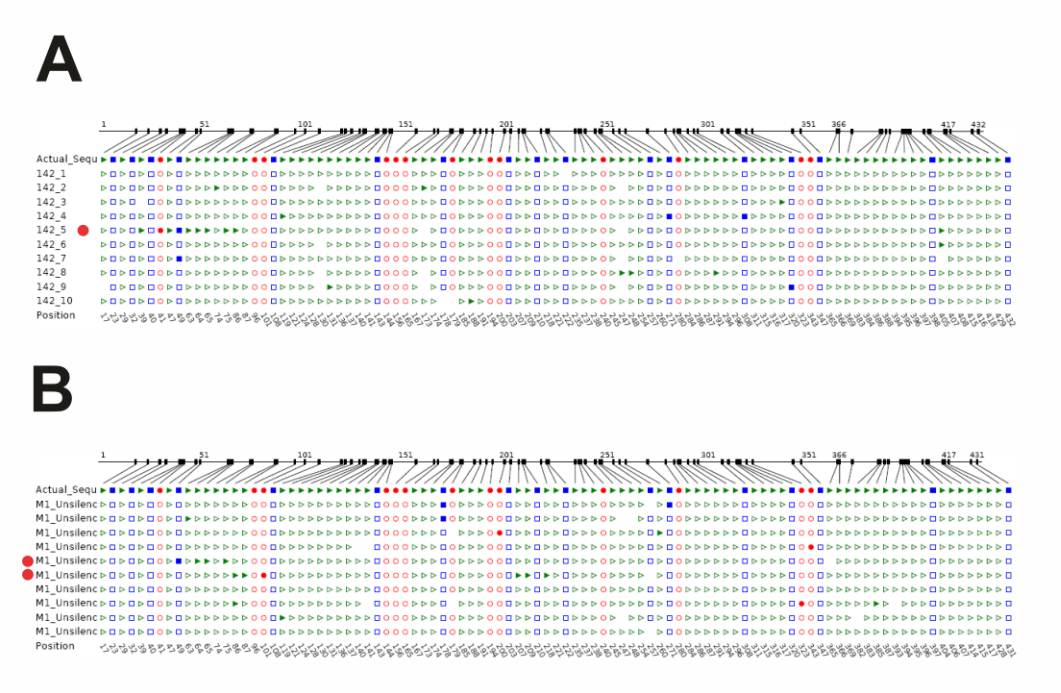


Figure 4.31: Lack of methylation in the unsilenced M1 protoplasts

Alignment of bisulfite sequences to the actual sequnce for DNA samples from 142 (**A**) and unsilenced M1 protoplasts (**B**). The top sequence where all shapes are filled in is the actual sequence. The black lines show where each cytosine residue is found on the sequence. Red circles are CG sites, the blue squares are CHG sites and green triangles are CHH sites. Filled in shapes indicate that cytosine is methylated. Red circles at the beginning of sequences indicate that this sequence is methylated. Sequences are considered methylated if they have 4 or more methylated cytosines.

4.3 Discussion

4.3.1 Role of *MORC6* in RdDM pathway

siRNA levels and DNA methylation in the M1 and M9 *morc6* mutants was assessed. siRNA levels do not differ between the two *morc6* mutants and WT lines indicating that *morc6* does not affect siRNA production. As opposed to siRNA levels DNA methylation is affected in the *morc6* mutants with a total loss of *35S* promoter DNA methylation in the majority of mesophyll cells in M1 and half of the mesophyll cells in M9. There is DNA methylation in a subset of mesophyll cells in M1 and just under half the cells in M9, however even in these cells methylation is lower than seen in WT. This reduction in DNA methylation but lack of change in siRNA levels would suggest that *morc6* is involved in the PolV and AGO4 complex that triggers DNA methylation. This would be contrary to the Moissiard paper which did not identify any significant reduction in DNA methylation and concluded that MORC6's role in RdDM was purely as a higher order chromatin modifier (Moissiard et al. 2012). However, the second paper to identify MORC6 as a RdDM component showed that at a subset of the endogenous RdDM targets tested and the transgene reporter used in the study, DNA methylation levels were reduced in the *morc6* mutant, but no change in methylation was observed at the majority of endogenous loci (Lorković et al. 2012). This would therefore suggest that MORC6's role in DNA methylation itself is loci specific and that the region of the genome where the *GFP* transgene is located is one such locus. To investigate whether the effect of *morc6* on DNA methylation is loci specific, DNA methylation of endogenous targets of RdDM was assessed and will be discussed in the next chapter.

As to MORC6's exact role in DNA methylation, it has previously been shown that MORC6 interacts with DMS3 to form a complex similar to SMC proteins (Lorković et al. 2012). SMC function is varied and consequently the structure of the SMC complexes formed is varied (Hirano 2005). It is suggested that in this case the

MORC6 and DMS3 complex, described as SmcMORC6 henceforth, forms a ring which could hold DNA and/or RNA strands in place during PolV transcription, thus providing a stable scaffold for other proteins to bind and form complexes (Bender 2012). IDN2 and its homologs form complexes that bind to siRNA-AGO4 complexes bound to the PolV transcript and it is thought that this binding and the consequential interaction of the zinc finger motifs of IDN2 and its homologs with DNA may trigger DRM2 to methylate the DNA (Ausin et al. 2012b, Zhang et al. 2012). SmcMORC6 clamping the DNA and RNA strand may therefore help facilitate this interaction between IDN2 and the DNA.

DMS3 has been shown to form part of the DDR complex with DRD1 and RDM1, which is required for PolV transcription (Law et al. 2010). MORC6 is also required for PolV transcription for at least one RdDM endogenous target, therefore indicating that MORC6 is also part of the DDR complex (Lorković et al. 2012). The SmcMORC6 complexes are therefore likely to bind at multiple sites both upstream and downstream of PolV to help provide a platform for PolV transcription and a secure scaffold for initiation of DRM2 methylation. It should be noted that other, as yet undiscovered, proteins may also be involved in SmcMORC6 complexes as this is common amongst SMC proteins (Hirano 2005). Further analysis using yeast 2-hybrid or immuneprecipitation would therefore be required to identify any such proteins.

As noted previously some cells maintain DNA methylation in *morc6* mutants and the reason why methylation could persist in *morc6* mutants is that there are six other *MORC* genes in *Arabidopsis* so there could be redundancy amongst the *MORC* genes. The fact that unlike other SMC proteins, which are encoded by one gene, DMS3 and MORC6 are encoded by separate genes would suggest that modularisation of SMC function has occurred in plants (Lorković et al. 2012). Due to the scale of loss of DNA methylation at the *35S* promoter in this study it is likely that MORC6 is the main MORC protein involved in silencing of this promoter. Other MORC proteins may act as the main MORC for other loci, but

further work is required to determine whether this is indeed the case and if there is redundancy. Testing mutants in other *MORC* genes and producing double mutants should therefore be informative.

There is also a delay in the onset of silencing in both M1 and M9, which may correlate with a change in DNA methylation. There is no obvious change in levels of *35S* promoter DNA methylation between 7 dpg and 21 dpg yet there is a clear change in GFP expression from both visual and molecular data. However, it cannot be discounted that silencing is already occurring at 7 dpg as the exact point of silencing onset has not been established for all tissues. Assuming that there is delay in silencing, it raises the question as to why, despite no detection of a change in methylation, the cells do not silence immediately. An explanation is provided by the other role for MORC6 suggested by Moissiard, that MORC6 is involved in higher order chromatin structure rearrangement (Moissiard et al. 2012). Other SMC proteins are involved in higher order chromatin rearrangement, including chromosome condensation during mitosis and meiosis, and MORC6 has been shown to be involved in the correct organisation of pericentromeric DNA into chromocentres associated with the nuclear envelope (Hirano et al. 1997, Sutani et al. 1999, Moissiard et al. 2012). MORC6 may also reinforce DNA methylation by facilitating the formation of higher order heterochromatin at the *35S* promoter. Other MORCs may therefore be able to compensate for the loss of MORC6 by methylating the DNA in some cells but are unable to form the higher order heterochromatin at earlier stages of development. The onset of silencing would therefore be associated with the establishment of the higher order chromatin by another MORC or other chromatin modifier. The reason for this delay in establishing higher order chromatin could be that the MORC or chromatin modifier in question is not as efficient as MORC6.

4.3.2 Cause of the late onset mosaic silencing pattern and influence of tissue type in *morc6* mutants

The M1 and M9 mutants in *MORC6* display an unusual silencing phenotype. Silencing is delayed in onset, mosaic and influenced by tissue type. The silencing phenotype of the *GFP* transgene also shows that M1 causes a greater loss of silencing than M9 and thus is a stronger defective allele in *MORC6*. Both M1 and M9 are premature stop codon mutations, but in M1 the premature stop codon is in the 41st codon whereas in M9 the premature stop codon is the 267th codon. The MORC6 protein produced in M1 would lack all functional domains but in M9 it is predicted to contain part of the GHKL ATPase domain, including three of the four conserved GHKL sequence motifs. This means that M9 still have some MORC6 functionality whereas M1 is likely to be a full loss-of-function allele. The T-DNA insertion *morc6* mutant, *morc6-3,* used in one of the previous studies to identify *MORC6* is a full knockout*,* but the other four *morc6* mutants may still have some residual MORC6 function (Lorković et al. 2012, Moissiard et al. 2012). These four lines are all premature stop codon mutations in codons 267, 293 and 439, which will produce a MORC6 protein of the same length or longer than that produced by M9. It would therefore be interesting to determine whether there are any differences between the phenotype of M1 and M9 and those observed in the other three *morc6* mutants for endogenous targets of RdDM. This will be discussed in the next chapter.

The mosaic silencing pattern of *morc6* has not been previously reported in any RdDM mutant. This is not necessarily because other mutants do not silence mosaically, as no previous study has presented data on silencing patterns using confocal microscopy. Also many studies have used transgene systems where this type of phenotype cannot be detected due to the nature of the reporter gene. The mosaic silencing pattern is unusual as it occurs in cells from the same cell type, therefore it is unlikely to be due to differences in expression of *MORC6* or other genes that may affect silencing. It is known that 24nt siRNAs involved in

RdDM can act as a mobile silencing signal that can transmit silencing from cell to cell and across the entire plant, via phloem transport (Smith et al. 2007, Molnar et al. 2010). It is not known whether *35S* siRNAs can be transported from cell to cell, however as levels in M1 and M9 are not significantly different from WT it is not the lack of siRNAs causing the mosaicism. Indeed the mosaic phenotype of M1 and M9 is similar to the positional-effect variegation phenomenon, which was first observed in the eye pigmentation of *Drosophila melongaster* and has also been observed in other species, including mice (Muller 1930, Robertson et al. 1995, Wallrath and Elgin 1995). The *White* gene is required for red eye pigmentation in *Drosophila* and a mutant in this gene was identified which displayed a mosaic pattern of cells exhibiting varying levels of pigmentation (Muller 1930). Investigation of this mutant revealed that a chromosome rearrangement had moved the *White* gene so that it was now adjacent to a pericentromeric region and that it was this region that was responsible for suppressing the gene and causing the loss in pigmentation, but that this silencing did not occur in all cells (Wallrath and Elgin 1995). This mechanism of silencing is unlikely to be the cause of the loss of silencing of the *35S* promoter as in WT plants silencing is triggered by the *35S IR* transgene rather than adjacent heterochromatic regions, but does show that stochastic silencing events in single cell types is possible.

The likely cause of the mosaicism in the M1 and M9 mutants may instead be due to a reduction in the efficiency of RdDM. MORC6 forms part of the DDR complex, required for PolV transcription and may also be involved in clamping the DNA and PolV transcript together, thus enabling IDN2 and its homologs to successfully bind to the PolV transcript-siRNA duplex and interact with the DNA through their zinc finger motifs (Ausin et al. 2012b, Lorković et al. 2012, Zhang et al. 2012). Loss of MORC6 would therefore perturb both PolV transcription and IDN2 binding resulting in a failure of RdDM to methylate the *35S* promoter. However, PolV transcription and correct IDN2 binding may still occur but less frequently than in WT, hence why some cells methylate and others do not. As these

processes are downstream of siRNA production they would be independent of cell to cell siRNA transmission so that silencing would be cell autonomous. This inefficiency in RdDM is highlighted by the difference between M1 and M9, as in M1, which is probably a full knockout of *morc6,* only 17.36% of mesophyll cells silence the *GFP* transgene whereas in M9, which may have some residual MORC6 function, 46.97% of mesophyll cells silence the transgene. This shows that even a partially functional MORC6 protein improves the effectiveness of RdDM.

Mosaicism has previously been described in a *morc6* mutant regarding chromocentre formation. When cells were assessed in *morc6* mutants for presence of chromocentres it was found that cells either lacked defined chromocentres, had ill-defined chromocentres or had chromocentres that were unaffected (Moissiard et al. 2012). *morc1* has a similar phenotype to *morc6,* but has a greater reduction in the percentage of unaffected chromocentres, suggesting that although there is redundancy between the two MORC proteins, MORC1 has the greater role in chromocentre formation. It should however be noted that the *morc6* mutant that was used to assess the chromocentre phenotype may not have been a full knockout so the effect of a full knock out of *morc6* may be greater. Unexpectedly the double mutant of *morc1* and *morc6* has a weaker phenotype than the two single mutant lines and the reason for this change in phenotype is unknown, but may be due to the up regulation of another chromatin modifier in response to the double but not the single mutants. Further investigation to determine if any other chromatin modifiers are indeed up regulated would be required to determine if this is the cause of this phenotype. It is possible that other MORC proteins are also involved in chromocentre formation and this would require further investigation by screening mutants in the other *MORC* genes for chromocentre definition.

Both M1 and M9 show clear influence of tissue type in their respective silencing patterns. As mentioned previously the degree of silencing varies between tissues, shown by the lower levels of silencing in the upper epidermis of leaves

compared to the lower epidermis. Redundancy between MORC6 and other MORCs or chromatin modifiers could be an explanation for these differences, but this would require further investigation to prove, as although published microarray data suggests differences in expression between *MORC* proteins, the activity of the MORC proteins has not been assessed directly in different tissue types. In the case of the upper and lower epidermal layers, if there is redundancy between the MORC proteins it would suggest that the other MORCs or chromatin modifiers are more able to replace MORC6 in the lower epidermis than the upper epidermis and hence why the delay in silencing is longer and levels of silencing are lower in the upper epidermis than the lower epidermis. There is an adaxial to abaxial gradient to silencing in both M1 and M9, suggesting that there also may be a gradient in *MORC6* expression, although there is no obvious explanation why such a gradient would occur. Silencing is stronger in leaves than roots, although in this case the difference in silencing may be related to the transgenes as this reduction in silencing in the roots is also observed in 142S. Analysing the expression levels of the transgenes would determine if this root specific silencing pattern is related to reduced expression of the *35S IR* or a reduction in *GFP* silencing. The reason for this difference is unclear as both transgenes should be ubiquitously expressed throughout the plant. A possible explanation is that expression of one or both of the transgenes is affected by nearby endogenous regulatory regions that affect expression levels and hence could impact silencing levels. Another possibility is that either the *GFP* or *35S IR* transgene may reside in a region of the genome that has different chromatin state between tissue types, which could affect silencing.

A delay in silencing has been seen in other RdDM mutants, notably in *nrpd1* (Herr et al. 2005). In that study, silencing was delayed and initially occurred in localised areas before spreading outwards. However, for M1 and M9 the onset of silencing depends on the tissue type, with the upper epidermis silencing between 14 dpg and 20 dpg and the mesophyll layers silencing before 10 dpg. This would suggest that tissue type influences time of onset as well as the strength of silencing in the

morc6 mutants. There is also molecular evidence for a delay in silencing of the *GFP* transgene as levels of *GFP* mRNA decrease between 7 dpg and 21 dpg in M1 and M9. However, as the 7 dpg samples will have a higher proportion of RNA from cotyledon leaves than the 21 dpg samples this decrease may be down to the difference in tissue composition. Silencing of *GFP* in 142S is not as strong in cotyledon leaves as rosette leaves, hence why *GFP* mRNA can be detected at 7 dpg in 142S but not 21 dpg and so this could also occur in M1 and M9. But visualisation of GFP between 7 and 21 dpg shows that levels of GFP decrease between these two time points and therefore I would argue that the decrease does provide molecular evidence for a delay in silencing. The lower epidermis' silencing pattern has been most thoroughly characterised and unlike in the *nrpd1* delay, silencing is sudden with cells going from unsilenced to a mosaic phenotype within six hours. The time of onset for epidermal silencing is also specific, being 13 dpg and 14 dpg for M9 and M1 respectively and this seems to be linked to time after germination rather than a specific developmental stage. The delay itself could be caused by redundancy of MORC6 with other MORC proteins, expressed at a later stage; although the fact silencing is not linked to development and the difference seen between M1 and M9 would discredit this. Another possible explanation is that due to the loss of MORC6, RdDM silencing is less efficient and therefore may take longer to silence the *GFP* transgene. This would not be dependent on developmental stage, although it would be expected that if RdDM was less efficient the time of silencing would be more random.

4.4 Acknowledgements

The production of the 142S transgenic line and EMS mutagenesis was carried out by A. Eamens. The cell sorting was carried out with the assistance of Karen Hogg. Peter O'Toole, Karen Hodgkinson and Graeme Park also provided technical assistance and help with experimental design. Jo Marrison provided guidance in imaging plant leaves under the confocal microscope and with propidium iodide staining. Simon Ramsbottom helped design the protocol for gelatin sectioning of

leaves. The siRNA northern blot detection of 35S siRNAs was carried out by Louise Jones. Tom Smith provided assistance with the MDS analysis of the bisulfite sequencing data.

5. Effect of *MORC6* on silencing of endogenous targets of RdDM

5.1 Introduction

5.1.1 Loci specific effects of RdDM mutants

The effect of different RdDM mutants on DNA methylation varies. This can be seen at a global level as well as at specific targets. Global methylation in *morc6* and *morc1* mutants is reported to be unaffected (Moissiard et al. 2012). Whereas the *drm1*, *drm2* and *cmt3* triple mutant causes near total loss of global CHG and CHH methylation (Lister et al. 2008). These two examples are extremes; most RdDM mutants have a more nuanced phenotype, whereby RdDM mutants' effect on DNA methylation varies between different target loci. This target specific nature of methylation has been shown in several studies. A recent study, identified the *IDN2* homologs *IDP1* and *IDP2* as RdDM components that bind to siRNAs bound to the PolV transcript and may mediate DRM2 methylation of the target DNA (Zhang et al. 2012). The study compared DNA methylation levels of the RdDM targets Solo LTR, MEA-ISR and AtSN1 in the mutant lines *idn2, idp1, idp2* and *nrpd1. idn2* and *idp1* mutants only affected CHG and CHH methylation but not CG methylation, whereas *nrpd1* led to reduction in DNA methylation in all cytosine sequence contexts for all the loci tested and *idp2* did not affect DNA methylation in any sequence context. The effect of *idn2* and *idp1* on CHG and CHH methylation varied between the loci, with a reduction in methylation of Solo LTR in *idn2* but not *idp1,* and reduced methylation of MEA-ISR for both. This shows a clear target specific effect of the *idn2* and *idp1* mutants.

As mentioned above, mutation of *MORC6* does not lead to large scale DNA methylation changes across the genome (Moissiard et al. 2012). This suggested that MORC6 is involved in establishing and maintaining higher order heterochromatin structure in response to DNA methylation at RdDM targets, rather than DNA methylation itself. However, a second study, which identified *MORC6* as a RdDM component showed that *morc6* does cause a loss of DNA methylation at specific loci and this loci-specificity is also seen in the first study

(Lorković et al. 2012, Moissiard et al. 2012). MORC 6 also forms a complex with the protein DMS3, but *dms3* by comparison to *morc6* exhibits a greater reduction in DNA methylation (Kanno et al. 2008, Ausin et al. 2009). It was therefore suggested either that the *morc6* mutant was not a full knockout, resulting in a smaller reduction in DNA methylation, or that there is redundancy with other MORC proteins since there are six *MORC* homologs in *Arabidopsis*. The two *morc6* mutant alleles identified in this study (M1 and M9) both cause a reduction in methylation of the transgene system used to identify mutants (Figure 4.25 and Figure 4.28). This suggests that *morc6* may effect DNA methylation. Therefore this chapter will focus on the effect of M1 and M9 on endogenous RdDM targets.

5.1.2 Selection of RdDM target loci to be tested

Assessment of the DNA methylation status of specific RdDM targets will be presented in this chapter. However as there are a large number of endogenous target loci only a subset of these loci was tested. These loci broadly subdivide into three groups: endogenous genes, repetitive elements and transposable elements (TEs) (Huettel et al. 2006, Lister et al. 2008, Baev et al. 2010). At least one target from each group was chosen. In total five targets were chosen and these were: 5S rDNA, MEA-ISR, AtMu1, Solo LTR and AtSN1 (Table 5.1). The 5S rDNA repeats are an example of an endogenous gene target as well as a repetitive element; MEA-ISR is an example of a repetitive element target and AtMu1, Solo LTR and AtSN1 are examples of TEs.

Abbreviated name	Full name	Target description
5S rDNA	5S rDNA gene repeats	Repeats of the 5S rDNA gene
AtMu1	*Arabidopsis thaliana* Mutator-like 1	Class II transposable element
AtSN1	*Arabidopsis thaliana* SINE 1	Class I (reterotransposon) transposable element
MEA-ISR	*MEDEA*-Intergenic Subtelomeric Repeats	Tandem repeats downstream of the *MEDEA* gene

Abbreviated name	Full name	Target description
SoloLTR	Solo Long Terminal Repeat	LTR of LTR/Copia transposon

Table 5.1: Endogenous targets of RdDM to be tested

Endogenous targets of RdDM that were tested for changes in DNA methylation in this chapter. The abbreviated name, full name and what the target is are given.

The 5S rDNA repeats are tandem repeats of the 5S rDNA gene and the resulting protein forms part of the 60S large ribosome subunit (Campell et al. 1992). There are estimated to be around one thousand copies of the 5S rDNA gene within these repeats. These thousand copies are arranged into four repeat cluster on chromosomes 3, 4 and 5, with two clusters on chromosome 5 (Murata et al. 1997). The exact make up of these clusters varies between ecotypes, with some ecotypes lacking the repeat cluster of chromosome 3 (Fransz et al. 1998). There are two classes of 5S rDNA gene, a major and minor gene, which have one or two SNP differences (Cloix et al. 2002). Major versions of the 5S rDNA genes are constitutively expressed while minor genes are expressed in roots and the cotelydon stage of plant development, but are mostly silenced in adult aerial tissue (Mathieu et al. 2003). The shift to silencing is associated with formation of heterochromatin at the pericentromeric regions. This suggests that as well as decreasing the level of 5S rDNA expression in response to a reduced need for the ribosomal subunit that silencing of the 5S rDNA is also associated with larger alterations in heterochromatin during leaf development. Silencing of the minor 5S rDNA genes involves multiple pathways including RdDM (Vaillant et al. 2007). Several RdDM mutants, including *ago4* and *dms3*, show reduction in methylation of the 5S rDNA repeats while other mutants, such as *drd1* and *hda6,* show no change in methylation (Aufsatz et al. 2002b, Kanno et al. 2004, Vaillant et al. 2007, Ausin et al. 2009).

The *MEDEA*-Intergenic Subtelomeric Repeats (MEA-ISR) is an example of a repetitive element target of RdDM (Deleris et al. 2010). MEA-ISR is a sequence of seven tandem repeats found downstream of the Polycomb group gene *MEDEA,*

part of a *Polycomb* Repressive Complex 2 type complex found in *Arabidopsis* (Kiyosue et al. 1999, Kohler et al. 2003). Each repeat is 183 bp long and shows homology with twelve other subtelomeric repeat elements (Cao and Jacobsen 2002a). MEA-ISR has been used to characterise the effect of several RdDM mutants' effect on endogenous DNA methylation, including *drm2*, *idn2, clsy1, nrpe1, dcl3* and *ktf1,* with some mutants, such as *drm2*, losing all methylation of MEA-ISR while others, such as *dcl3,* result in a partial loss in methylation (Ausin et al. 2009, Greenberg et al. 2011).

TEs are the last group of endogenous RdDM targets, of which *Arabidopsis thaliana Mutator*-like (AtMu1), *Arabidopsis thaliana* SINE 1(AtSN1) and Solo Long terminal repeat (LTR) have been chosen for DNA methylation analysis. AtMu1 is a class two TE that belongs to the Mutator-like element (MULE) family in *Arabidopsis* and is a homolog of the *Mutator* TEs found in maize (Yu et al. 2000). There are two AtMu1 TEs in the Columbia genome and both are transcriptionally inactive (Singer et al. 2001). They are silenced by DNA methylation and become transcriptionally active in DNA methylation defective mutants, such as *ddm1.* AtMu1 has been used to characterise several RdDM mutants including *rdm4*, *ktf1* and *rdm1* (He et al. 2009b, He et al. 2009c, Gao et al. 2010). AtSN1 is another class two TE that belongs to the Short Interspersed Elements (SINE) family, of which there are two sub-families in *Arabidopsis*, AtSN1 and AtSN2 (Myouga et al. 2001). There are seventy one AtSN1 elements, which vary in sequence and structure and are targets of RdDM silencing. These have been used to characterise a large number of RdDM mutants including *nrdp1,nrpd4* and *shh1* (Hamilton et al. 2002, Herr et al. 2005, He et al. 2009a, Law et al. 2011). The final TE, Solo LTR, is a lone LTR so can also be considered a repeat element as well as a TE (Huettel et al. 2007). The solo LTR is the LTR of a LTR/*Copia* reterotransposon, which has been orphaned by either unequal crossing over or intrachromosomal recombination between LTRs (Shirasu et al. 2000, Huettel et al. 2007). The solo LTR is located on the left arm of chromosome five and is methylated in WT, but

loses methylation in RdDM mutants, including *drd1* and *ktf1* (Huettel et al. 2007, Bies-Etheve et al. 2009).

5.1.3 RdDM effect on flowering time

As well as RdDMs role in chromatin regulation, it is also involved in flowering time. The control of flowering time is complex and involves the action of four main regulatory pathways: the autonomous pathway, the vernalization pathway, the photoperiod pathway and gibberellin-dependent pathway (Wilson et al. 1992, Koornneef et al. 1995, Alonso-Blanco et al. 1998, Blázquez et al. 1998, Koornneef et al. 1998a, Koornneef et al. 1998b, Amasino 2004). These pathways all regulate *Flowering time Locus C* (*FLC*), which is a repressor of flowering. It has been shown that the RdDM pathway is also involved in *FLC* regulation and represses expression through methylation of sites downstream of the poly-adenylation site of *FLC* (Swiezewski et al. 2007). Deletion of this target site causes a late flowering phenotype, indicating that the action of RdDM is required for correct regulation of *FLC*.

The *FWA* gene is another floral repressor that is controlled by DNA methylation, however in this case methylation is maintained by MET1 maintenance rather than RdDM (Kinoshita et al. 2004). However, in cases where methylation of *FWA* has been lost, such as in the *decrease in DNA methylation 1* (*ddm1*) mutant, methylation can be re-established by RdDM (Kinoshita et al. 2004). The Jacobsen group also use a transgenic *FWA* gene to screen for RdDM mutants (Greenberg et al. 2011). In WT plants introduction of the *FWA* transgene results in DNA methylation of the transgene by RdDM and flowering time is unaffected; but in RdDM mutants the transgene is not methylated, resulting in a delay in flowering time.

These findings would therefore suggest that RdDM mutants should exhibit a delay in flowering, which has been shown to be the case for *dcl3, nrpd1* and

nrpe1 mutants (Pontier et al. 2005, Greenberg et al. 2011). However, the effect of RdDM mutants on flowering time depends on day length, with only a minor delay in flowering time in long day (16 hour) conditions compared to a greater change in short day (8 hour) conditions (Greenberg et al. 2011). This could therefore suggest an involvement in the photoperiod regulatory pathway that controls flowering time.

5.1.4 Aims of chapter

The aim of this chapter is to characterise the effect of the *morc6* mutants on endogenous targets of RdDM. This will be done by studying:

A) Effect of M1 and M9 mutants on DNA methylation of endogenous targets of DNA methylation

B) Effect of M1 and M9 on siRNA production from endogenous targets

C) The effect of M1 and M9 on flowering time in short day growth conditions

5.2 Chapter-specific methods

5.2.1 Methylation sensitive Southern blotting

To assess whether endogenous RdDM targets have lost DNA methylation, methylation sensitive Southern blotting was used. Initially DNA samples were digested with either *Msp*I or *Hae*III and then run on a gel before being transferred to a membrane. This membrane is then probed with a radiolabelled probe for a specific locus, producing a specific banding pattern for that locus. The pattern is affected by methylation as both restriction enzymes are methylation sensitive. *Msp*I is sensitive to CG and CHG while *Hae*III is sensitive to CHG and CHH methylation (Figure 5.1). If methylation is present at the restriction site the restriction enzyme will not cleave the DNA, but will if methylation is not present.

Cleavage of restriction sites will create lower molecular weight (LMW) bands whereas as failure to cleave will produce higher molecular weight (HMW) bands. If methylation has been lost in the mutants this will increase the number of restriction sites that are cleaved, manifesting in a shift from HMW bands to LMW bands, compared to the WT lines. Bisulfite sequencing can also be used to assess DNA methylation and is in fact more informative as it can test all cytosine residues at a locus rather than a select few that are in restriction sites. The reason for the use of Southern blotting despite its limitations is that it is both more rapid and less expensive than bisulfite sequencing.

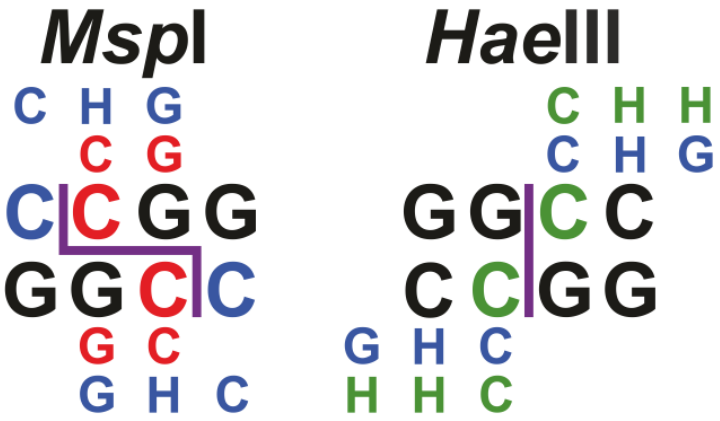


Figure 5.1: How methylation of *Msp*I and *Hae*III restriction sites affects cleavage

Diagram of the sequences of the restriction sites of *Msp*I and *Hae*III. The purple line shows where the respective restriction enzyme cleaves. *Msp*I cleavage creates a two nucleotide overhang while *Hae*III forms blunt ends. Red, blue or green Cs represent the cytosine residues in the restriction sites that if methylated will inhibit cleavage of that sequence. Red Cs are cytosine residues in a CG sequence context; blue Cs are cytosines in a CHG sequence context; and greens Cs are cytosines in a CHH sequence context. For the cytosines in the *Hae*III restriction site they can either be in the CHG or CHH sequence configuration as the third base lies outside the restriction site. The letters above and below the restriction site sequence give the sequence context of the methylation

sensitive cytosines, with the colours of the letters again representing the sequence context of CG, CHG or CHH.

5.2.2 Methylation sensitive PCR

As well as methylation sensitive Southern blotting, methylation sensitive PCR was also used to asses AtSN1 DNA methylation changes. Like Southern blotting the DNA is first digested with the methylation sensitive restriction enzyme *Hae*III, which detects CHG or CHH methylation (Figure 5.1). Two PCRs were then performed on 5 ng of the digested DNA, whereby one amplifies AtSN1, and the other amplifies part of *SHH1*. The AtSN1 PCR product spans a *Hae*III cut site in AtSN1, while the *SHH1* PCR product, described as *SHH1 4* from hence forth, lacks a *Hae*III cleavage site. The *SHH1 4* product will always amplify in the *Hae*III digested DNA so provides a positive control for presence of DNA and its concentration. The AtSN1 PCR product will only be produced if *Hae*III has not cleaved the DNA at the AtSN1 restriction site due to the presence of methylation. The level of amplification of the AtSN1 PCR can therefore be used to provide a relative value for methylation levels at the locus. In order to provide a value for the amount of PCR product produced the intensity of the PCR product on a SYBR gel is measured using ImageJ and normalised against the *SHH1 4* PCR product's intensity. A PCR for AtSN1 was also performed on undigested DNA as another positive control for each sample to show that the AtSN1 PCR works in all instances.

5.2.3 Flowering time assay

A flowering time assay was carried out to assess whether the mutants affected flowering time. As mentioned in the introduction RdDM mutants can affect flowering time and this is most apparent in short day growth conditions (Pontier et al. 2005). For this reason the flowering time assay was carried out in short day (8 hour) conditions. Flowering time was considered to be the point when the first flower opened and the leaf number (LN) of that plant was then recorded

(Koornneef et al. 1998a). LN is the total number of adult rosette and cauline leaves and has been previously shown to correlate well with the number of days until flowering (Koornneef et al. 1991). To allow for statistical analysis of the results, the LN of at least one hundred plants from each line were recorded. In previous flowering time assays an analysis of variance (ANOVA) was used to test for statistical significance (Koornneef et al. 1998a). An ANOVA assumes that each data point is independent; that all data are normally distributed within groups; and variance is homogeneous between groups. For this study it is assumed that data points are independent as this is an assumption made in previous flowering time assays. To check if the data was normally distributed, the Kolmogorov-Smirnova and Shapiro-Wilk tests of normal distribution were used before running the ANOVA. A Levene's test of homogeneity of variance was used to check the variance between groups. If there was no significant difference an ANOVA could be used, but if there was a significant difference in variance between groups the Welch robust test of equality of means, which is similar to an ANOVA but take into account difference in variance, was used instead.

As there is only one form of grouping being tested in this experiment, that being plant line, a one way ANOVA/Welch test was used. If the ANOVA/Welch test showed that there was a significant difference between the means for each line the Least Significant Distance (LSD) post hoc test was then used to identify which groups were different from each other, but if there was not a significant distance the Bonferroni post hoc test was used (Dytham 2005). This is because the LSD test is less powerful if the ANOVA/Welch test is not significant.

5.3 Results

5.3.1 *morc6* has no effect on DNA methylation of endogenous genes

Methylation of the *35S* promoter does not change between 7 dpg and 21 dpg and so methylation of endogenous loci was also tested at 7 dpg and 21 dpg to determine if this is also true for endogenous targets of RdDM and if M1 and M9 affect methylation at these loci. To achieve this methylation sensitive Southern blotting was used and how this technique works is explained in in the methods section of this chapter (page 294). DNA samples from lines 142, 142S, M1, M9, *rdr2-1* and *rmd3* or *nrpd1-2* were taken from 7 dpg and 28 dpg plants. Both 142 and 142S give the WT methylation pattern for each locus tested, while *rdr2-1, nrpd1-2* and *rmd3* are previously characterised mutants in the RdDM pathway that lose DNA methylation at endogenous loci (Xie et al. 2004, Eamens et al. 2008).

The first locus tested was the 5S rDNA repeats which were chosen as an example of an endogenous gene targeted by RdDM. The 5S rDNA repeats had previously been tested during the initial characterisation of the mutant lines (Figure 3.3) but was of poor quality so the one shown in this chapter is a repeat. All three sequence contexts of methylation were assessed. DNA from the *nrpd1-2* mutant line was used instead of *rmd3* for the 5S rDNA Southern blots. Both *nrpd1-2* and *rmd3* are both full knockouts of the *NRPD1* gene so will produce similar DNA methylation patterns, but are in different ecotypes with *nrpd1-2* in the Columbia ecotype and *rmd3* in the C24 ecotype (Eamens et al. 2008). *rmd3* contains the transgene system used in this study and has a nonsense mutation in the *nrpd1* gene, whereas *nrpd1-2* lacks the transgene system and has a T-DNA insertion which affects the *NRPD1* gene. *rmd3* is therefore preferentially used, however at the time when the 5S rDNA Southern blots were completed, *rmd3* seed stocks were being replenished so were not used.

There is no change in *Msp*I CG or CHG methylation of the 5S rDNA repeats in either M1 and M9 or *rdr2-1* and *nrpd1-2* (Figure 5.2 A). The banding pattern of the four mutant lines at 7 dpg is identical to that of 142 and 142S at 7 dpg, with no loss of HMW bands. Unfortunately, both 142 and 142S 21 dpg DNA samples were not properly digested by *Msp*I, which can be seen from the presence of high molecular weight genomic DNA in both the Southern blot and ethidium bromide gel. Although the 21 dpg WT methylation pattern is unknown, since there is no change in digestion pattern between 7 dpg and 21 dpg for M1, M9 and *nrpd1-2,* DNA methylation at 21 dpg in the mutants is likely to be the same as WT. There is also a problem with the *rdr2-1* 21 dpg banding pattern as the signal from DNA bands is weak due to the lower amount of DNA run on the gel for this sample, compared to other DNA samples on the gel. The low amount of DNA is caused by poor recovery of DNA after the precipitation of the restriction digestion reaction, as equal amounts of DNA were digested for all samples.

Neither M1 nor M9 show a change in *Hae*III CHG or CHH methylation of the 5S rDNA repeats (Figure 5.2 B). Both M1 and M9 have a banding pattern that is identical to the WT 142 and 142S patterns. There is also no change between the 7 dpg methylation pattern and 21 dpg methylation patterns. As neither M1 nor M9 cause a change in CG, CHG or CHH methylation of the 5S rDNA repeats, this suggests that *MORC6* does not affect 5S rDNA methylation. In contrast both *rdr2-1* and *nrpd1-2* lose CHH methylation (Figure 5.2 B). This can be seen by the loss of HMW bands in both mutants. There is no change in methylation pattern between 7 dpg and 21 dpg in *rdr2-1*, however DNA methylation appears to change between 7 dpg and 21 dpg for *nrpd1-2*, as there is an increase in the HMW bands between 7 dpg and 21 dpg. This increase is small so may be due to differences in the amount of DNA loaded for each sample rather than a delay in methylation, although if genuine it would correlate with the previously observed delay in *nrpd1* mutants (Herr et al. 2005). *nrpd1-2* also has a greater effect on *Hae*III CHG and CHH methylation of the 5S rDNA repeats than *rdr2-1,* seen by the presence of more HMW bands in *rdr2-1* than *nrpd1-2*, suggesting that the effect

of *nrpd1-2* on DNA methylation is greater than *rdr2-1,* at least for the 5S rDNA repeats.

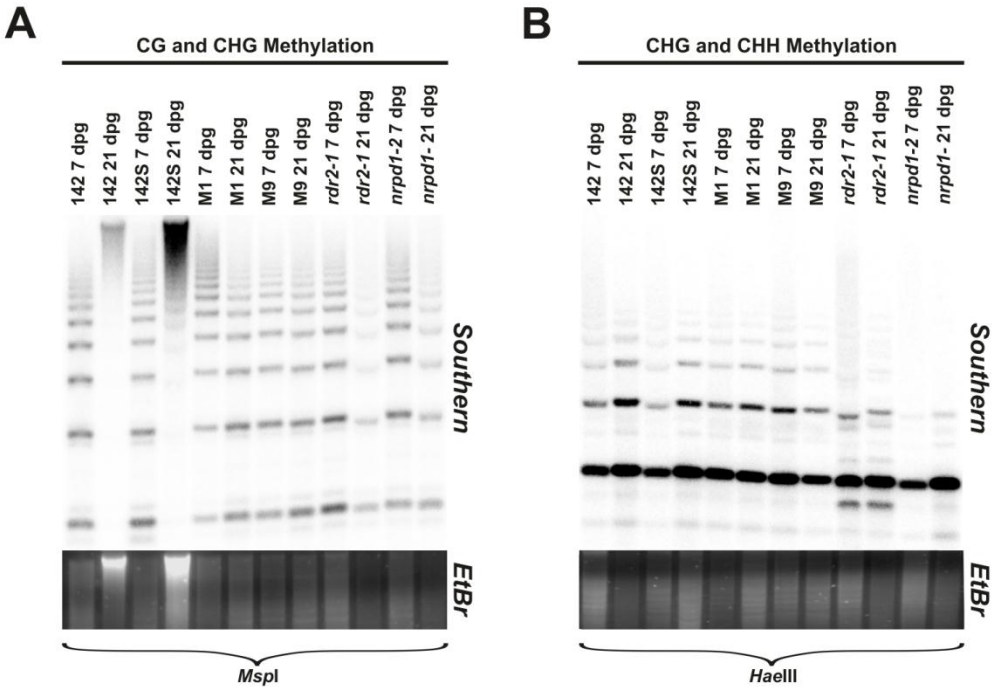


Figure 5.2: M1 and M9 have no effect on 5S RNA repeat CG, CHG and CHH methylation

Southern blots of the 5S rDNA repeats. DNA from 7 dpg and 21 dpg plants of lines 142, 142S, M1, M9, *rdr2-1* and *nrpd1-2* were used. The order of loading is given at the top of each blot. For each blot an image of the probed Southern blot and part of the ethidium bromide stained gel used for blotting are shown. The stained gel provides an indication of the amount of DNA loaded for each sample. **A:** Southern blot of the 5S rDNA assessing CG and CHG methylation. DNA was digested with *Msp*I. **B:** Southern blot of the 5S rDNA assessing CHG and CHH methylation. DNA was digested with *Hae*III.

5.3.2 *morc6* has no effect on DNA methylation of repetitive elements

Repetitive elements in the *Arabidopsis* genome are another major target of RdDM of which the MEA-ISR was assessed for changes to CG and CHG DNA

methylation in the *morc6* mutants. Analysis of MEA-ISR CG and CHG methylation shows no loss of methylation in M1 and M9 compared to WT (Figure 5.3). In WT plants a HMW and a LMW band are detected, which are similar in intensity. This banding pattern is seen in both M1 and M9 at both 7 dpg and 21 dpg, indicating that there is no change in methylation between these time points and no reduction compared to WT. This would suggest that MORC6 therefore does not affect DNA methylation at the MEA-ISR locus. *rdr2-1* and *rmd3* both lose CG and CHG methylation at the MEA-ISR loci, seen by the loss of the HMW band in *rdr2-1* and near total loss in *rmd3*. There is no difference in methylation between 7 dpg and 21 dpg in either mutant. Whereas for the 5S rDNA repeats *rmd3* has a greater loss of DNA methylation, DNA methylation loss is greatest in *rdr2-1* for the MEA-ISR locus, therefore indicating that these two known mutants differentially affect different targets.

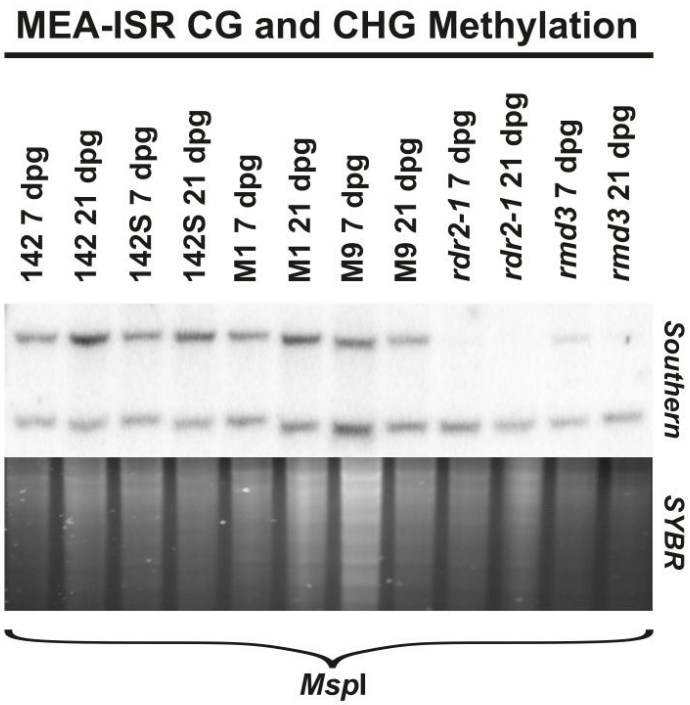


Figure 5.3: M1 and M9 have no effect on MEA-ISR CG and CHG methylation

Southern blot looking at CG and CHG methylation of the MEA-ISR. DNA from 7 dpg and 21 dpg plants of lines 142, 142S, M1, M9, *rdr2-1* and *rmd3* were used. The order of

loading is given at the top of the blot. DNA was digested with *Msp*I, which is sensitive to CG and CHG methylation. An image of the probed Southern blot and part of the SYBR safe stained gel used for blotting are shown. The stained gel indicates the amount of DNA for each sample used for the Southern blot.

5.3.3 Locus specific effect of *morc6* on DNA methylation of transposable elements

The final group of endogenous targets of RdDM are TEs. Three TEs were tested: AtMu1, Solo LTR and AtSN1. There is no change in DNA methylation for either M1 or M9 for the AtMu1 transposon as the banding pattern in both mutants is identical to the WT lines 142 and 142S (Figure 5.4). There is again also no change between 7 dpg and 21 dpg samples for either M1 or M9. AtMu1 is therefore another endogenous RdDM target where *morc6* does not affect DNA methylation. The two previously characterised RdDM mutants, *rdr2-1* and *rmd3,* both cause a loss in DNA methylation. This loss of DNA methylation is greater in *rdr2-1* than *rmd3,* seen by the near total loss of HMW bands in *rdr2-1* compared to the more modest increase in LMW bands in *rmd3*. Again this shows a locus specific effect for the *rdr2-1* and *rmd3* mutants due to the difference in methylation loss.

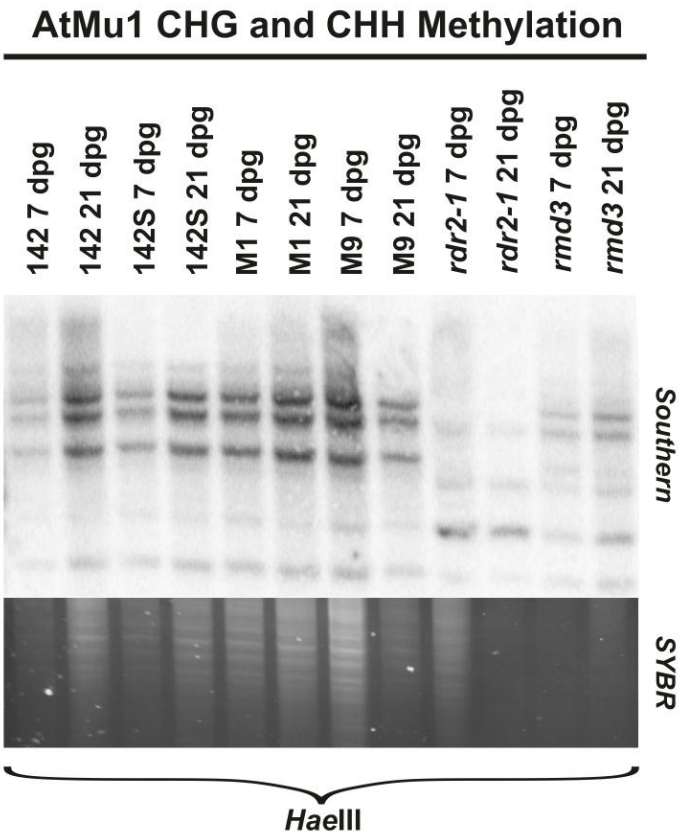


Figure 5.4: M1 and M9 have no effect on AtMu1 CHG and CHH methylation

Southern blot looking at CHG and CHH methylation of AtMu1. DNA from 7 dpg and 21 dpg plants of lines 142, 142S, M1, M9, *rdr2-1* and *rmd3* were used. The order of loading is given at the top of the blot. DNA was digested with *Hae*III, which is sensitive to CHG and CHH methylation. An image of the probed Southern blot and part of the SYBR safe stained gel used for blotting are shown. The stained gel indicates the amount of DNA for each sample used for the Southern blot.

There is a change in CG and CHG methylation of the Solo LTR locus in the M1 and M9 mutants (Figure 5.5). In WT samples two bands are detected, of which the HMW band is more intense than the LMW band. However, in the M1 and M9 samples taken at 21 dpg the intensity of the HMW and LMW bands are equal, indicating a loss in DNA methylation. For the 7 dpg samples, both M1 and M9 have a banding pattern that is more similar to WT in that the HMW band is more intense than the LMW band. Repetition of this blot using different DNA samples produced the same difference in the banding intensity for Solo LTR in lines M1

and M9 between 7 dpg and 21 dpg (Appendix 3.1 page 402). The reason as to why DNA methylation of the Solo LTR appears to decrease between 7 dpg and 21 dpg could be that MORC6 is more active in DNA methylation at this locus at later developmental stages, so would have a greater effect on methylation at the later time point; but this would require further investigation to show this to be the case. However, from the 21 dpg samples there is a clear reduction of DNA methylation in M1 and M9, indicating that *morc6* mutants affect DNA methylation of endogenous targets of RdDM as well as the transgene system used in this study. As the three other RdDM targets so far tested showed no change in DNA methylation, this would suggest that the effect of *morc6* mutants is loci specific in terms of its effect on DNA methylation.

The Solo LTR banding pattern could not be detected in the *rdr2-1* so was not shown in Figure 5.5. The most likely cause for the lack of detection is ecotype differences. The two parental lines, M1, M9 and *rmd3* are all C24 ecotype, hence the Solo LTR probe was produced from WT C24 DNA, however *rdr2-1* is of the Col ecotype (Xie et al. 2004). The failure to detect the Solo LTR in *rdr2-1* could therefore be due to sequence differences between the C24 based probe and the *rdr2-1* genome. The Solo LTR could be detected in the *rmd3* mutant and shows a modest reduction in DNA methylation at 7 dpg but not at 21 dpg. At 7 dpg the ratio between the HMW and LMW bands are of similar intensity, whereas at 21 dpg the HMW band is stronger than the LMW, indicating an increase in DNA methylation. This again matches previously published results that *nrpd1* mutants show a delay in silencing of a *GFP* transgene system (Herr et al. 2005).

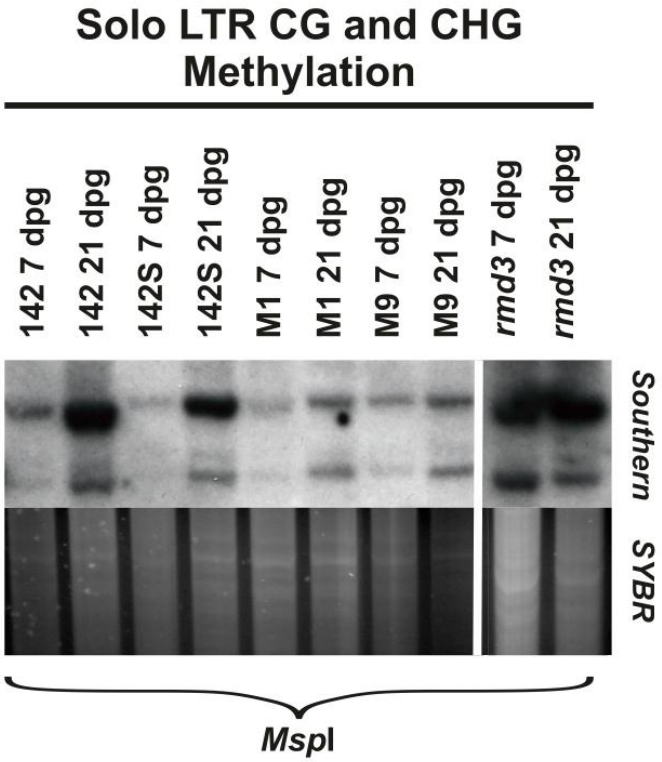


Figure 5.5: M1 and M9 cause a reduction in CHG methylation of Solo LTR

Southern blot looking at CG and CHG methylation of Solo LTR. DNA from 7 dpg and 21 dpg plants of lines 142, 142S, M1, M9 and *rmd3* were used. The order of loading is given at the top of the blot. DNA was digested with *Msp*I, which is sensitive to CG and CHG methylation. An image of the probed Southern blot and part of the SYBR safe stained gel used for blotting are shown. The stained gel indicates the amount of DNA for each sample used for the Southern blot. Two lanes have been removed from the images of the blot and gel which is indicated by the white vertical line between the M9 21 dpg and rmd3 7 dpg samples.

The final TE assessed in this study was AtSN1. Unlike the previous loci, methylation sensitive PCR was used instead of methylation sensitive Southern blotting. Although visually there appears to be little difference in levels of the AtSN1 PCR product between 142 and 142S there is a difference in the normalised AtSN1 values (Figure 5.6). Whereas for both 7 dpg and 21 dpg 142 has a normalised intensity of 0.68 142S has a intensity of 0.50 at 7 dpg and 0. 87 at 21

dpg (Figure 5.6 B). As there is little difference between the samples visually it is likely this difference is due to errors in measurement of intensity. This means that the difference seen in 142S is unlikely to be biologically relevant. By comparison to 142 and 142S both visually and through the normalised intensity values, there appears to be no change in DNA methylation in either M1 or M9. Visually the level of AtSN1 product in M1 and M9 are similar to 142 and 142S and although the relative intensity of M1 and M9 varies, it is within the variation seen in 142S (Figure 5.6). This would indicate that like three of the previous four endogenous RdDM targets, AtSN1 methylation is unaffected in *morc6* mutants. By comparison DNA methylation is reduced in both *rdr2-1* and *rmd3.* In *rdr2-1* there are reduced levels of AtSN1 PCR product, although the reduction is greatest at 21 dpg than 7 dpg, which is unexpected as it suggests methylation decreases from 7 dpg to 21 dpg. This possibly could be caused by tissue specific effects on loss of DNA methylation as a far greater proportion of the 7 dpg DNA sample will come from cotyledon leaves than the 21 dpg sample. AtSN1 PCR amplification is reduced in *rmd3* compared to 142 and 142S levels, but there is a difference between 7 dpg and 21 dpg. Amplification is higher in the 21 dpg sample than the 7 dpg, however similar differences are seen in 142S so this could also be due to errors during measuring intensity. It should be noted that this method is non-quantitative so, although it provides an indication of differences in methylation levels between the mutants, it cannot be used to quantify the level of methylation.

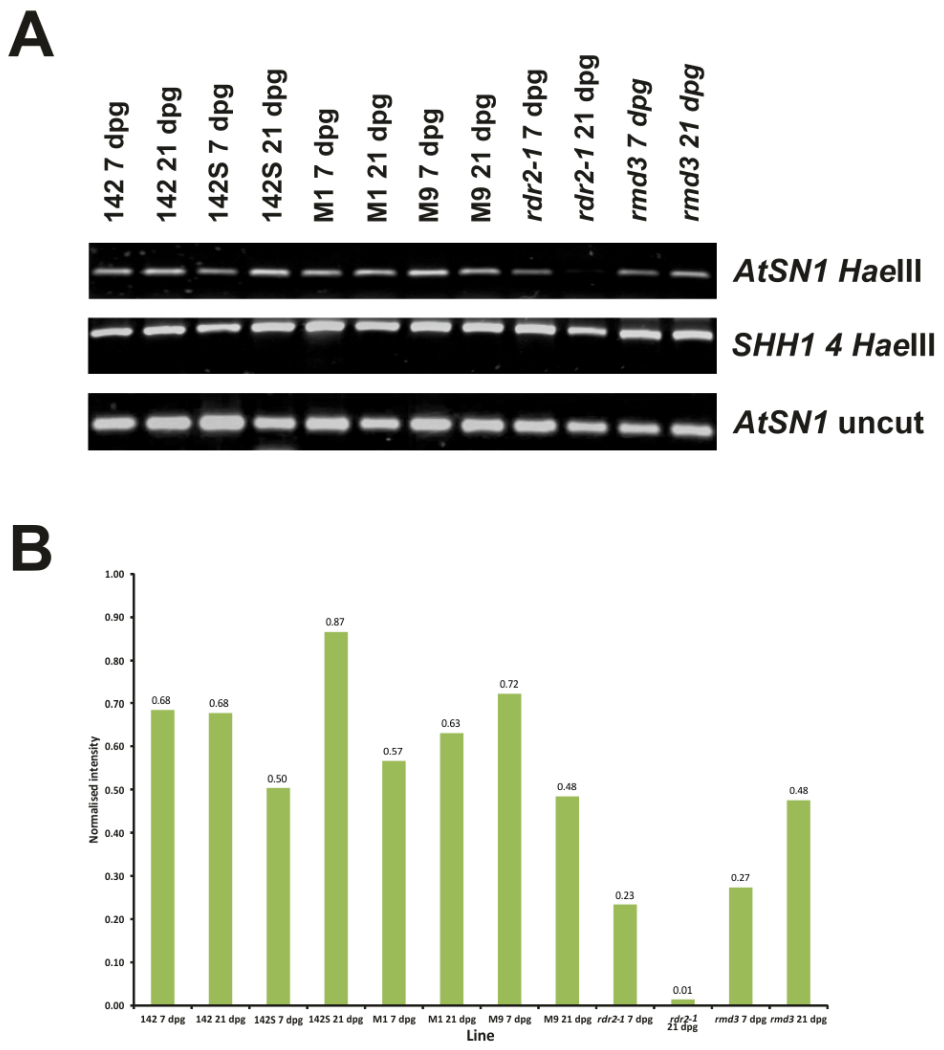


Figure 5.6: M1 and M9 do not affect AtSN1 CHG and CHH methylation

Methylation sensitive PCR analysis of CHG and CHH methylation for the AtSN1 TE. **A:** Gel images of PCR products produced from 7 dpg and 21 dpg DNA sample from lines: 142, 142S, M1, M9, *rdr2-1* and *rmd3.* The order is given above the gel photos. Three PCR products for each DNA sample are shown: AtSN1 from DNA digested with *Hae*III (Top); *SHH1 4* from DNA digested with *Hae*III (Middle); and AtSN1 from undigested DNA (Bottom). The PCR products from *Hae*III digested DNA were run on the same gel whereas the undigested DNA sample PCRs were run on a different gel. **B:** Graph showing the normalised intensity of AtSN1 PCR product in the *Hae*III digested DNA samples AtSN1 is normalised to the intensity of the respective *SHH1* 4 PCR product.

5.3.4 Increase in global DNA methylation levels in *morc6* mutants

Analysis of endogenous RdDM targets reveal localised effects of the *morc6* mutants but cannot be used to show global changes in DNA methylation. In order to determine if there are global effects on DNA methylation in M1 and M9 a Epigentek Methyl Flash kit was used to measure genomic levels of methyl cytosines relative to WT. The kit measures methyl cytosine levels using an anti-methyl cytosine antibody and then quantifying the amount of antibody bound to the DNA colorimetrically through a procedure similar to ELISA. This method provides a relative measurement of the number of genomic methyl cytosines. DNA from 7 dpg 142, 142S, M1 and M9 plants were tested using this kit. The relative levels of DNA methylation for 142S, M1 and M9 were normalised against the 142 value and this was repeated once using different sets of plants grown at different times so are considered to be biological replicates. 142 was used to normalise the other three samples as it should have WT levels of DNA methylation so will show the change in DNA methylation compared to WT. 142S should also have WT DNA methylation, however in both replicates DNA methylation is higher in 142S than 142, which is unexpected (Figure 5.7). The only difference between the lines is the *35S IR* transgene, therefore this increase may be due the presence of the transgene. The transgene should only silence the *GFP* transgene, so its effect on DNA methylation is therefore unlikely due to the function of the transgene itself but instead its location in the genome. Another possibility is that a large scale change in the epigenome has occurred in the 142S line that is maintained in the M1 and M9 mutants that is unrelated to the *35S IR* transgene. It should be noted that there is a difference in the change in DNA methylation levels for 142S, M1 and M9 between the two biological replicates, however as the technical replicates for each biological replicate were not significantly different from each other the pattern in each biological replicate is likely to be genuine. The difference between the two biological replicates could be due to natural variation in DNA methylation or an error in measuring in DNA

methylation between plates as the biological replicates were not done at the same time.

Both M1 and M9 show an increase in levels of DNA methylation in both biological replicates compared to 142 and 142S (Figure 5.7). Again, as with 142S, the relative change in DNA methylation is different between the two biological replicates, which is either due to natural variation in DNA methylation levels or an error in measuring DNA concentration. There is also a difference in M9 between the two biological replicates, in the 1st replicate M9 has a higher level of methylation than M1 but in the 2nd replicate methylation in M9 is lower than in M1. This again could reflect natural variation in DNA methylation levels or be an error in measurement as it is subjective when to stop the colour developing so there can be a difference in the intensity of the absorbance between plates. The relative DNA methylation levels would however suggest that methylation is higher in the two mutant lines than in 142S, which is unexpected as analysis of RdDM targets shows a reduction of DNA methylation at some target loci. This would therefore mean that at some specific loci there is loss of DNA methylation but overall there is an increase in DNA methylation.

A

Relative Level of DNA Methylation		
Line	**1st Biological Repeat**	**2nd Biological Repeat**
142 7 dpg	1.00	1.00
142S 7 dpg	1.32	1.87
M1 7 dpg	1.49	2.72
M9 7 dpg	1.56	2.26

B

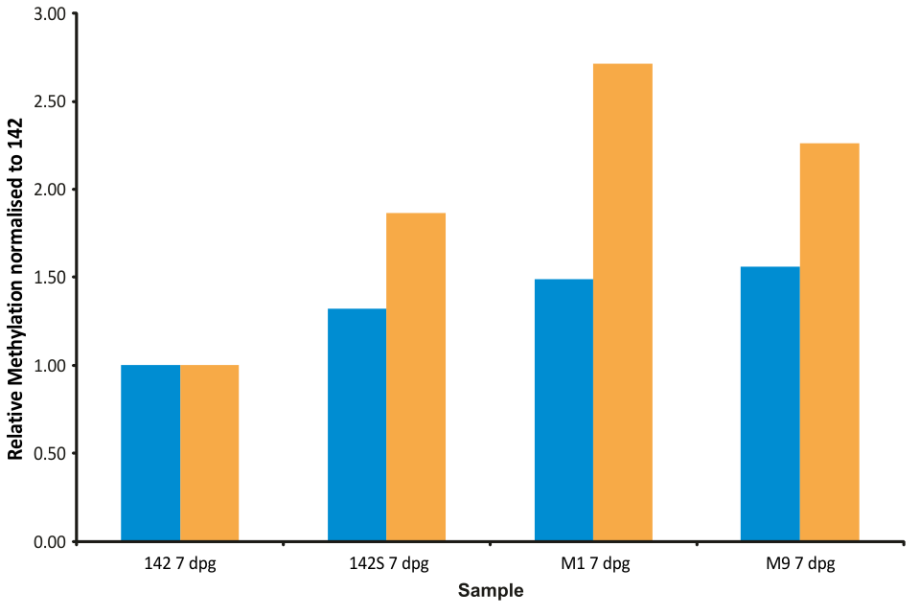


Figure 5.7: Increase in global DNA methylation in *morc6* mutants

Analysis of global DNA methylation levels using an Epigentek Methyl Flash kit. **A:** table of relative levels of DNA methylation in 7 dpg 142, 142S, M1 and M9 plants. There are two biological replicates for each line and the values shown are the mean of two technical replicates, except the 1st biological replicate of 142S 7 dpg, which only has one technical replicate. For each replicate levels of DNA methylation are normalised to 142. **B:** graph of the relative levels of DNA methylation in 142, 142S, M1 and M9. Both biological replicates are shown, the 1st replicate is blue while the 2nd replicate is orange.

5.3.5 Solo LTR expressed in *morc6* mutants

From screening endogenous RdDM targets for loss of DNA methylation, only the Solo LTR showed a reduction in CG and CHG DNA methylation in M1 and M9. Although there is a reduction in methylation it was not known if this resulted in expression of the Solo LTR. To determine whether expression occurred in M1 and M9 a reverse transcriptase (RT) PCR for the Solo LTR was carried out on 28 dpg RNA samples. RNA samples were taken from the WT lines 142 and 142S; the *morc6* mutants M1, M9 and rmd6; and *rmd1*, which is a *nrpe1* mutant (Eamens et al. 2008). There is no Solo LTR expression in either WT lines but all three *morc6* mutants and *rmd1* show expression of Solo LTR and would therefore suggest that the reduction in DNA methylation in *morc6* mutants does release expression of the Solo LTR (Figure 5.8). The expression of the Solo LTR is similar between M1, M9 and *rmd1* but lower in rmd6 and would correlate with rmd6 being the weaker of the three *morc6* alleles. However, this PCR is only non-quantitative so conclusions on exact differences in expression level of the Solo LTR between mutants cannot be made.

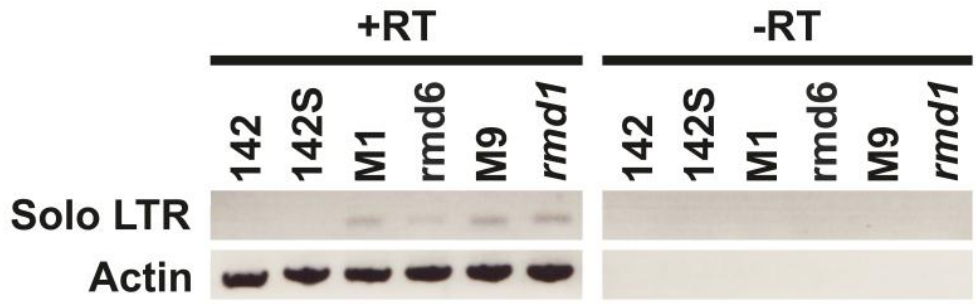


Figure 5.8: Solo LTR expressed in *morc6* mutants

Gel images of RT PCR for Solo LTR from 28 dpg RNA samples from lines 142, 142S, M1, *rmd6,* M9 and *rmd1.* The order of the RNA samples in shown along the top with the RT samples shown on the left and no RT samples shown on the right. Both the Solo LTR and Actin PCR products are the expected size. The top row is the Solo LTR PCR for which

samples were taken at thirty five cycles and the bottom row is the Actin control PCR for which samples were taken at thirty cycles.

5.3.6 *morc6* mutants do not affect Solo LTR siRNA levels

There is a decrease in DNA methylation at the Solo LTR locus, which therefore suggests that MORC6 has an effect on methylation but does not indicate at which point in the RdDM pathway MORC6 acts. To determine if MORC6 is involved in siRNA production, a northern blot of RNA samples that had been enriched for small RNAs was used to assess levels of Solo LTR 24 nt siRNAs. Samples were taken from leaves at 7 dpg and 21 dpg and floral tissue from lines 142, 142S, M1 and M9. The northern blots was initially probed with a radio-labelled 35S probe (described previously in chapter 4 page 250) before being re-probed for Solo LTR. Solo LTR siRNAs are detectable in 7 dpg and 21 dpg samples but are at lower levels than in the floral samples, which is expected as 24 nt siRNAs are up regulated in floral tissue (Mosher et al. 2009). When M1 and M9 Solo LTR levels in 7 dpg, 21 dpg and floral samples were compared against the respective 142 and 142S WT samples there was no difference in levels of Solo LTR 24 nt siRNAs, suggesting that MORC6 does not affect siRNA levels and is therefore not involved in siRNA production in the RdDM pathway (Figure 5.9). However, the miR167 probe, which is used as a loading control, showed a difference in loading between the floral samples with the 142S, M1 and M9 samples being significantly lower than 142. This may indicate that there is a lower amount of RNA for these samples and consequently that Solo LTR 24 nt siRNAs levels are in fact higher in 142S, M1 and M9 compared to 142. The fact that on the acrylamide gel the level of M1 RNA was lower than all other samples would support this idea for M1, but the RNA levels for 142S and M9 were similar to that of 142. The reason for the reduction in 142S and M9 could be either poor RNA transfer during blotting or that as the blot must be first stripped of the Solo LTR probe before being probed with the miR167 probe, some of the RNA from these floral samples were removed during this process. Of these two possibilities

it is likely that the latter is true as 142 and 142S would not be expected to have differing Solo LTR 24 nt siRNA levels, but to determine which of these was the case a second blot of just the floral samples was also assessed (Appendix 3.2 page 403). Again this blot showed no significant difference between 142, 142S, M1 or M9 Solo LTR 24 nt siRNAs. However, although 142 and 142S had similar levels of miR167, the levels in M1 and M9 were again lower. This again raises the question whether this difference is due to loss of RNA during the blot stripping process or if it is a genuine difference in RNA levels. However, from the 7 dpg and 21 dpg samples that have no problems with loading and the fact that there is no difference in the signal for Solo LTR siRNAs between floral samples, I would argue that the blots show no change in Solo LTR siRNA levels.

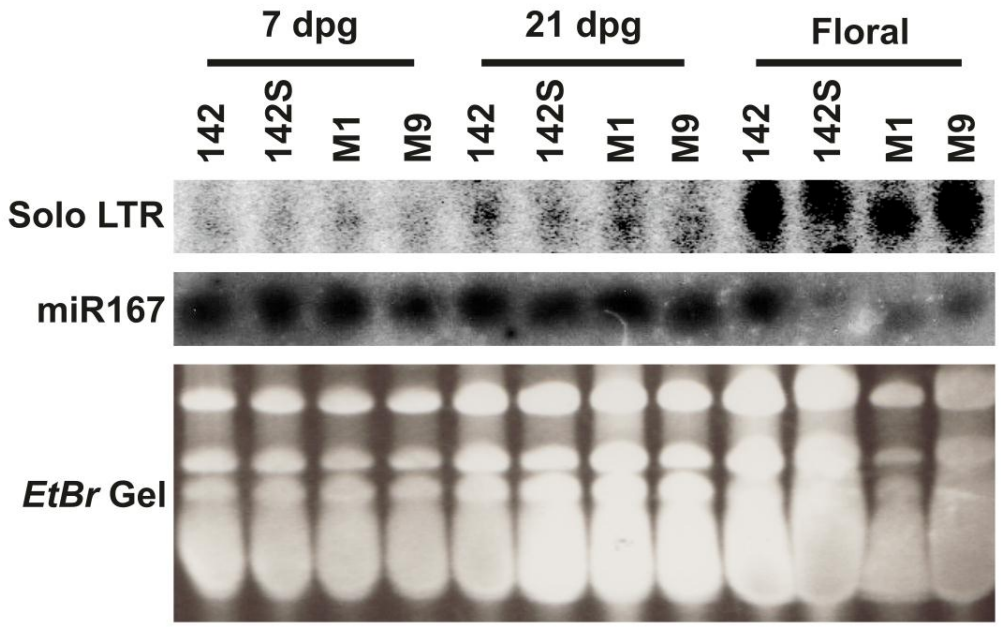


Figure 5.9: Solo LTR siRNAs can be detected in M1 and M9

Image of a northern blot probed with a ^{32}P-labelled probes for Solo LTR (top) and miR167 (middle). 7 dpg, 21 dpg and floral RNA samples of lines 142, 142S, M1 and M9 were used, as indicated. The bottom image is of the polyacrylamide gel stained with ethidium bromide before being blotted and shows RNA loading onto the gel. The miR167 probe acts as a loading control.

Levels of 24 nt siRNAs were also tested at another locus, AtREP2, which is a helitron transposon. For this blot RNA samples were extracted from 7 dpg and 21 dpg leaves and floral tissues of lines 142, 142S, M1, M9, *rdr2-1* and *nrpd1-1.* The two known RdDM mutants, *rdr2-1* and *nrpd1-1* are both involved in siRNA production so should have reduced siRNA levels and indeed no 24 nt siRNA was detected for AtREP2 in these samples (Figure 5.10). Comparison of M1 and M9 to the WT lines 142 and 142S showed no difference between 142, 142S and M1 but a decrease in M9. For this blot there was no difference in the loading control, which used a probe specific for $tRNA^{met}$. This would indicate that the decrease in M9 AtREP2 siRNAs is genuine and not due to reduced levels of RNA. However it would be expected that M1 should also show a decrease in At REP2 siRNAs if this was caused by the *morc6* mutation. Since this is not the case the reason for this apparent decrease in AtREP2 siRNA levels could be due to differences in the stage of floral tissue collected for RNA extraction as the stage of the flower can affect the levels of 24 nt siRNAs (Mosher et al. 2009).

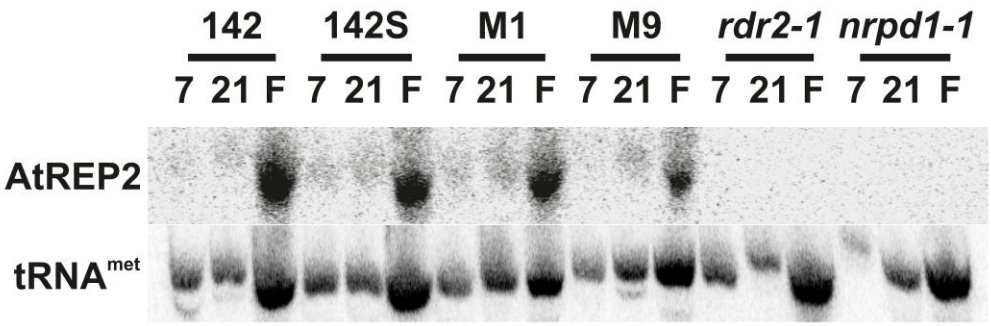


Figure 5.10: M1 and M9 do not affect AtREP2 siRNA levels

Image of a northern blot probed with radio-labelled probes for AtREP2 (top) and $tRNA^{met}$ (bottom). 7 dpg, 21 dpg and floral RNA samples of lines 142, 142S, M1. M9, *rdr2-1* and *nrpd1-1* were used, the order of these is indicated. The $tRNA^{met}$ probe acts as a loading control.

5.3.7 Differing effect of M1 and M9 on flowering time

In order to test whether *morc6* mutants have an effect on flowering time, a flowering time assay was performed, which is described in detail in the methods section of this chapter (page 296). As well as M1 and M9, wild-type (WT) C24, 142, 142S and *rmd1* were also used in the flowering time experiment. 142S is the parental line of M1 and M9 so flowering time in these mutants must be compared against 142S in order to determine if a change has occurred. WT C24 and 142 were also tested in order to assess if the transgenes used in this experiment have any effect on flowering time, 142 lacks the *35S IR* transgene so can be compared against 142S to determine if there is an effect on flowering time for that transgene and WT C24 can be compared against 142 to determine if there is an effect of the *GFP* transgene. Both M1 and M9 have been backcrossed twice to the parental line to remove EMS mutations unrelated to the RdDM phenotype. However, some EMS mutations that are unrelated to the RdDM phenotype but linked to the mutant RdDM gene may not have been removed by this process and therefore may affect the flowering time. The fact that there are two alleles should lessen the effect of linked mutations as they should not have the same set of mutations as each other. *rmd1* is a mutant in the largest PolV subunit, *npre1,* previously identified from EMS mutagenesis of 142S (Eamens et al. 2008). It would be expected that *rmd1* should show a delay in flowering as a *nrpe1* mutant showed a delay in silencing in short day conditions in a study by Pontier, although this was in the Columbia ecotype background whereas *rmd1* is in the C24 background, so ecotype specific effects may alter flowering time (Pontier et al. 2005).

Before statistical analysis of the leaf number (LN) was carried out the LN data for each line was assessed using the Kolmogorov-Smirnova and Shapiro-Wilk tests and a box whisker plot. The two statistical tests are used to see if the data was normally distributed and showed that the LN data for five of the lines was normally distributed but not for 142S (Appendix 3.3 page 404). As five of the

lines show normal distribution a test assuming normal distribution of the data could still be used but meant that the power of the test was reduced due to 142S not meeting one of the assumptions for the test. The box whisker plot showed that the range and distribution of WT C24, 142 and M9 were similar while 142S has a smaller range than these lines but its median is similar (Figure 5.11 A). M1 and *rmd1* have a different range to the other lines with a range shifted above and below that of the other lines respectively. This is also reflected in the means of each line with the M1 mean (66.18 leaves) being higher than that of the other lines and the *rmd1* mean (53.90) being lower than the means of the other lines (Figure 5.11 B). The means for 142 (63.56 leaves) and 142S (62.96 leaves) are similar to each other with a lower mean for WT C24 (61.02 leaves) and M9 (59.17). SPSS also defined several outliers in lines WT C24, 142, 142S, M1 and M9 and an extreme value for 142S (Figure 5.11 A). Outliers were included in the statistical analysis while extreme values were not.

The Levene's test of variance showed a highly significant difference in variance between the lines, with a significance value of <0.001, therefore the Welch test was used instead of an ANOVA as the Welch test does not assume homogeneity in variance. The Welch test showed a highly significant difference (<0.001) between the means, so a least significant difference (LSD) post hoc test was used to identify which lines were significantly different. The LSD test showed that M1 and *rmd1* means are statistically different from all other lines (<0.001 significance). This means that there is a significant delay in flowering time in the M1 mutant, however the mean of M9 was lower than both the M1 and parental means, and this difference is significant. M9 therefore flowers earlier than the parental line, which is opposite to that of M1, suggesting that the difference in flowering time is not due to the *morc6* mutation as both lines would be expected to behave similarly. Other mutations carried in either M1 and M9 are therefore likely to account for this difference, one mutation carried by the M1 line gives rise to a necrotic lesion phenotype which is likely to impact on flowering time. This will be discussed in more detail in the next section (page 322). The earlier

flowering observed in the *rmd1* mutant is unexpected as a previous study that measured flowering time in a *nrpe1* mutant showed a delay in flowering (Pontier et al. 2005). This difference again could be due to other mutations carried in the *rmd1* lines or difference in flowering time between ecotypes, as *rmd1* is in a C24 background whereas the *nrpe1* mutant used in the Pontier study was in the Columbia background. Statistical analysis shows that there is no significant difference between the means of 142 and 142S but that there is a significant difference (<0.05) between WT C24 and 142, which may suggest that the *GFP* transgene but not the *35S IR* transgene has an effect on flowering time. However, as the difference between the means is small it is possible that other biological factors, such as seed quality could account for the difference.

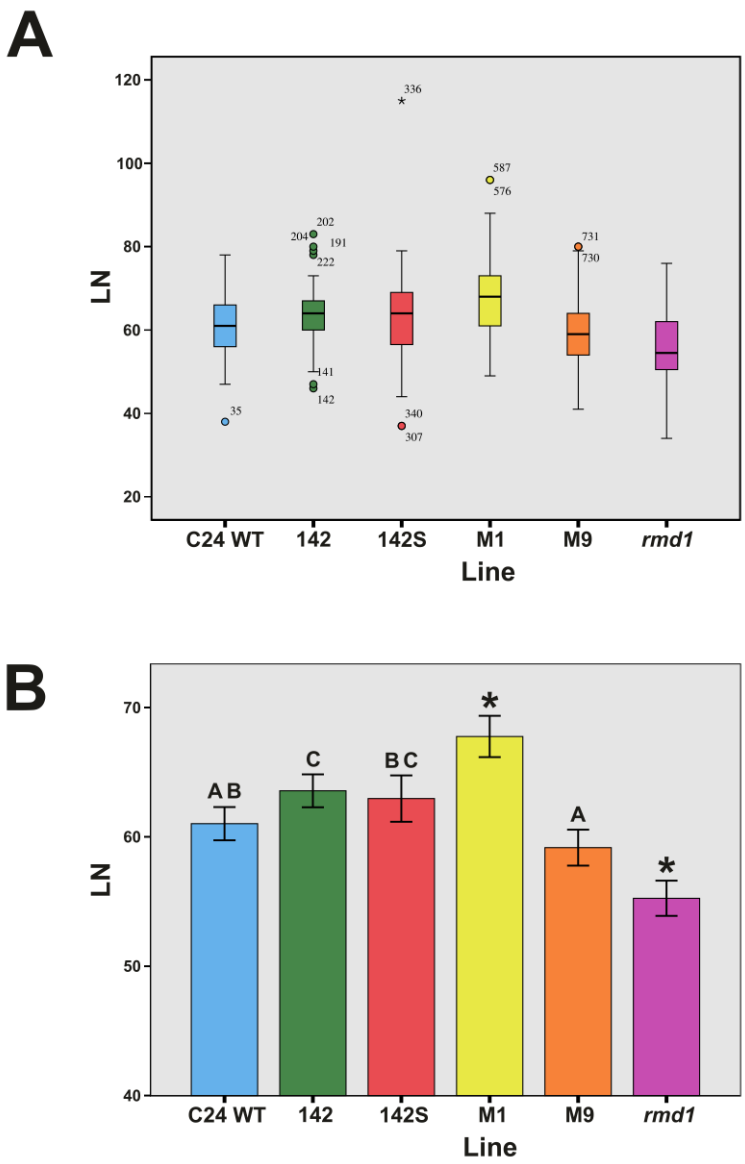


Figure 5.11: Different effect of M1 and M9 on flowering time

Graphs showing LN of lines: WT C24 (blue), 142 (green), 142S (red), M1 (yellow), M9 (orange) and *rmd1* (purple). **A:** Box whisker plot of LN distribution for each line. For each line the box shows the inter quartile range for each line, with the black line within each box being the median. The whiskers show the highest and lowest LN for each line, but do not include values that are outliers (circles) and extreme values (asterisk). **B:** Bar graph of mean LN for each line. The error bars represent two standard errors above and below the mean. Lines with a ***** above the bar (M1 and *rmd1*) have a significance of <0.001 to all other lines. For the other lines they are significantly different from each

other, bar those denoted by letters. The letters above each bar denote which lines they do not have a significant difference between, so any line with an A is not significantly different from any other line with an A and the same is true for B and C.

During the flowering time assay it was observed that there was a difference in plant size between the lines. This difference was not apparent in long day (16 hour) conditions, but only in short day (8 hour) conditions. In order to quantify this difference the rosette of each plant was weighed at the point when the first flower opened. The data showed a clear difference between 142 and WT C24, and 142S and the three RdDM mutants, the mean for WT C24 is 1.52 g and 1.87 g for 142 whereas for 142S and the RdDM mutants the means range from 0.88 g for 142S to 0.65 g for *rmd1,* clearly demonstrating a difference in size between the two groups (Figure 5.12 B). There is also far greater variation in rosette weight in lines WT C24 and 142 than the other four lines, which can be seen in the difference between the interquartile ranges. WT C24 has a interquartile range (IQR) of 0.54 g and 142 has a IQR of 0.57 g, whereas for 142S and the RdDM mutants the IQRs range from 0.29 g for M1 to 0.22 g for *rmd1* (Figure 5.12 A).

The same statistical tests used for the leaf number analysis were used to test the rosette weight data. The data was normally distributed, but the Levene's test showed a <0.001 significant difference in variance between the lines so the Welch test was again used instead of an ANOVA (Appendix 3.4 page 405). The Welch test was significant so a LSD post hoc test was used to identify which lines were significantly different from each other. The LSD test showed that WT C24, 142 and 142S were significantly different from each other (<0.001), with 142 having increased rosette weight compared to WT C24 and 142S having reduced rosette weight compared to WT C24. This could suggest that both the *GFP* and *35S IR* transgenes influence plant growth, possibly due to where each transgene is inserted into the genome (Figure 5.12). The decrease in 142S is far greater than the increase in 142 compared to WT, which may suggest that the dual

transgene background has a greater effect on plant growth than the single *GFP* transgene. However, it cannot be discounted that other experimental variables may be responsible for these differences, such as seed quality or light intensity, so requires repetition. M1 and M9 do not have a significant difference in rosette weight between each other, but both show a statistically significant decrease in weight compared to 142S, although this change is minor, with a decrease of less than 0.2 g compared to a decrease of 1.0 g between 142 and 142S. This could suggest that mutation of *morc6* has a minor effect on growth, but not to the same extent as whatever is causing the decrease in growth in 142S. *rmd1* also has a statistically significant decrease in rosette weight compared to 142S, although again the decrease is minor compared to the one between 142 and 142S. The decrease does however suggest that *nrpe1* affects plant size. This reduced rosette weight is only observed in short day conditions, with visual assessment of plant in long day conditions suggesting no significant difference in rosette weight between 142, 142, M1 and M9. However, this needs to be tested experimentally as smaller differences in rosette weight, not apparent in visual assessment, may exist.

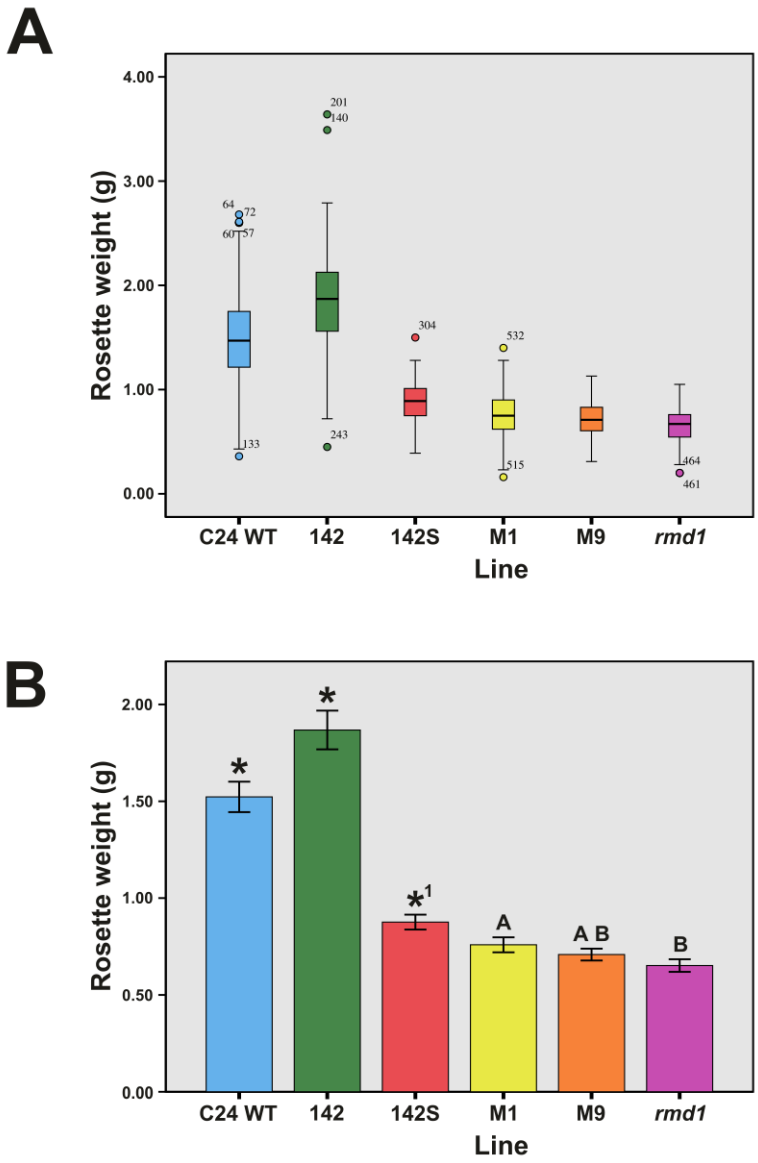


Figure 5.12: Reduction in plant size of 142S, M1, M9 and *rmd1*

Graphs showing rosette weight of lines: WT C24 (blue), 142 (green), 142S (red), M1 (yellow), M9 (orange) and *rmd1* (purple). **A:** Box whisker plot of rosette weight distribution for each line. For each line the box shows the inter quartile range for each line, with the black line within each box being the median. The whiskers show the highest and lowest rosette weight for each line, but does not include values that are outliers (circles). The y axis scale is in grams. **B:** Bar graph of mean rosette weight in grams for each line. The error bars represent two standard errors above and below the mean. Lines with a * above the bar (WT C24 and 142) have a significance of <0.001 to all

other lines. 142S (*[1]) has a significance of <0.001 to all lines except M1 to which it has a significance of <0.05. For the other lines they are significantly different from each other, bar those denoted by letters. The letters above each bar denote which lines they do not have a significant difference between, so any line with an A is not significantly different from any other line with an A and the same is true for B.

5.3.8 M1 necrotic lesion phenotype

During the flowering time assay it was observed that M1 developed necrotic lesions on its leaves approximately four weeks post germination. These lesions manifested as small areas of necrosis with multiple lesions found on a single leaf and this occurred on nearly all leaves on affected M1 plants (Figure 5.13). This phenotype was only observed in short day (8 hour) growth conditions and not in long day (16 hour) growth conditions. No other line, including M9, showed this necrotic lesion phenotype (Figure 5.13 and Table 5.2). The necrotic lesion phenotype occurred in 91.6% (131 of 143) M1 plants and this coupled with the lack of a necrotic lesion phenotype in M9 suggests that this phenotype, although interesting, is not related to the *morc6* mutation but instead is caused by another mutation closely linked to the *morc6* mutation in M1.

Line	Plants with necrotic lesions	Plants without necrotic lesions
142	0	121
142S	0	128
M1	131	12
M9	0	136
rmd1	2	131
WT C24	0	148

Table 5.2: Number of plants displaying necrotic lesion phenotype

Table showing number of plant that have necrotic lesion or lack necrotic lesion for lines 142, 142S, M1, M9, *rmd1* and WT C24. Plants are considered to have necrotic lesions if they have at least five rosette leaves with at least one 1 mm diameter lesion.

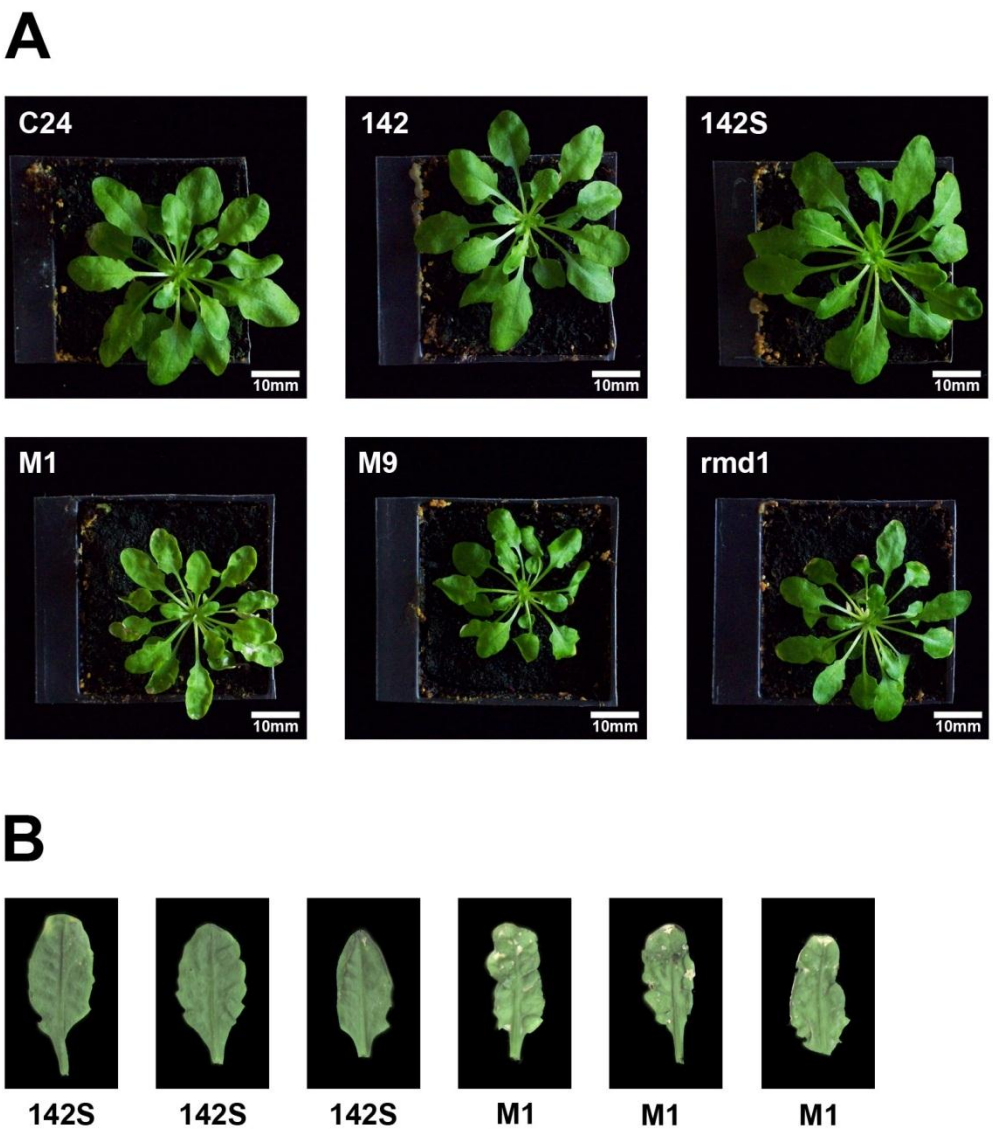


Figure 5.13: Necrotic lesion phenotype of M1

Images of leaves and whole plants showing the M1 necrotic lesion phenotype in short day growth conditions. **A:** Whole plant images of lines WT C24, 142, 142S, M1, M9 and *rmd1* at 37 dpg. A scale bar is shown in the bottom of each image. **B:** Images of adaxial side of adult rosette leaves of lines 142S and M1 at 100 dpg.

5.4 Discussion

5.4.1 *MORC6* has a locus-specific effect on DNA methylation

Analysis of DNA methylation of the *35S* promoter revealed that in both M1 and M9 *morc6* methylation is reduced. In order to test whether the phenotype was specific to the transgene only, a series of endogenous RdDM targets was analysed. Of the five loci tested only one, Solo LTR, showed a reduction in DNA methylation suggesting that MORC6's role in DNA methylation is locus-specific. The loss of DNA methylation at the Solo LTR results in it being expressed but does not correspond to a reduction in siRNA levels suggesting that MORC6 is not involved in siRNA production at this locus and is therefore more likely to be involved in the DNA methylation process itself. The locus specificity matches the results seen in the two previous studies of *MORC6* where global methylation was not affected but methylation was reducted at specific target loci (Lorković et al. 2012, Moissiard et al. 2012). The studies showed no change in methylation for AtMu1 and AtSN1, which I have also tested and found no change in DNA methylation levels. However, neither study showed a significant reduction in Solo LTR methylation and the Moissiard study showed an increase in CHG methylation levels, whereas in this study Solo LTR has reduced CHG methylation. The study by Lorković assessed CHG and CHH methylation with the *Alu*I restriction enzyme whereas I tested for CG and CHG methylation using *Msp*I, so it is possible that methylation at *Alu*I sites are unaffected at this locus. However, this is not the case for the Moissiard study which used bisulfite sequencing rather than a restriction enzyme. One of the limitations of methylation sensitive Southern blotting in that it can only detect methylation changes at a specific site whereas bisulfite sequencing can assess methylation across the entire locus. It is therefore unknown whether the apparent decrease in methylation at the *Msp*I site in the Solo LTR in this study occurs for other cytosines at the locus. For this reason bisulfite sequencing of the Solo LTR should be carried out to confirm the

decrease in methylation seen in this study. Assuming that the decrease in Solo LTR CHG methylation is accurate one possible explanation for the difference in methylation of the Solo LTR between this study and the previous two is that there may be an ecotype specific effect on the methylation of this locus, as this study used plants in the C24 background whereas the other studies used plants in the Columbia background. Another possible explanation is that there appears to be a loss of methylation between 7 dpg and 21 dpg in *morc6* mutants. The study by Moissiard analysed methylation at 14 dpg therefore it is possible that Moissiard assessed methylation before the decrease occurred. However, this would require bisulfite sequencing analysis of Solo LTR methylation at 7 dpg and 21 dpg to demonstrate that such an effect occurs. If there is a decrease in DNA methylation from 7 dpg to 21 dpg this would suggest a complex role for MORC6 in Solo LTR methylation.

The other endogenous RdDM targets where a reduction in DNA methylation has been reported are Intergenic loci (IGN) 23, IGN25 and AtCopia28, none of these loci have been tested in this study so this change in DNA methylation cannot be confirmed (Lorković et al. 2012, Moissiard et al. 2012). As the *GFP* transgene also shows a reduction in methylation, this would suggest that the region in which it is inserted is also targeted by MORC6. However, analysis of the Illumina sequencing data failed to identify this region or whether there were multiple copies of the transgene at this locus. This is likely due the transgene location residing within a region of the C24 genome that has not been assembled. It would be interesting to determine the identity of the region and if MORC6 affects DNA methylation when the insertion is not present. Of the fifteen RdDM targets tested in both this and the two previous studies, four show a reduction in DNA methylation. This is a significant subset of the RdDM loci tested and could therefore suggest that MORC6 is involved in DNA methylation at a large number of loci. However, the study by Moissiard found no large scale changes in DNA methylation so it is possible that, due to the low number of loci tested, this is an overestimation of the effect of MORC6 in DNA methylation (Moissiard et al.

2012). Another possibility for this disparity between locus specific loss of methylation and a lack of change in global methylation levels could be that a perturbation of RdDM by MORC6 could alter which regions of the genome are methylated, which is what has recently been shown to occur in *nrpe1* mutants (Wierzbicki et al. 2012). To determine the extent of the locus-specific decrease in DNA methylation in *morc6* mutants a further endogenous targets must be assessed using bisulfite sequencing.

The reason for the locus-specific nature of MORC6's role in DNA methylation is unclear but may be due to redundancy with other MORC proteins. There are six other *MORC* genes in the *Arabidopsis* genome and each could be specific to a subset of RdDM targets. Unlike Metazoan MORC proteins, Plantae MORC proteins lack a zinc finger domain so are unlikely to be able recognise specific DNA sequences and target specific loci themselves. However, *Arabidopsis* MORC proteins have a coiled coil domain that allows for protein-protein interactions and five MORCs have low complexity regions LCR that are thought to increase the number of interactions a protein can make (Coletta et al. 2010). These domains should allow the MORC proteins to interact with other proteins and it may be these that may infer the locus specificity. Differences in protein sequence amongst the *Arabidopsis* MORCs may result in each being able to bind to a different subset of proteins and so to be targeted to a different set of loci. This hypothesis would need to be tested by looking at *Arabidopsis* MORC protein-protein interactions and locus interaction using ChIP.

5.4.2 *morc6* mutants show an increase in global DNA methylation levels

Global DNA methylation levels in both M1 and M9 was higher than in the parental line. The parental line itself was also unusual as DNA methylation was higher in the 142S parental line than 142. The only known difference between these two lines is the presence of the *35S IR* transgene, thus the transgene may

be the cause of the change in global DNA methylation levels. However, unless the transgene's location affected the regulation of a methylation or demethylation gene, it would seem unlikely that the presence of the *35S IR* transgene would cause such drastic changes to global methylation. Another explanation of this disparity between 142 and 142S is that there may have been a large scale change in the epigenome in 142S that is unrelated to the additional transgene, but the cause of this change is unknown. 142S also shows a reduction in size in short day conditions and this may be related to the increase in DNA methylation. Further work would be required to elucidate what is causing both the reduced size and increased methylation levels in 142S. There is also the question of whether this increase in methylation occurs across the entire genome or is localised, as it should be noted that 142S did not show an increase in methylation, compared to 142, in any of the five endogenous RdDM loci tested.

Returning to the increase in global methylation in M1 and M9, the fact that global methylation is not reduced matches the findings of Moissiard, however they did not find an increase in methylation either (Moissiard et al. 2012). If the increase in DNA methylation in M1 and M9 is caused by the mutation of *MORC6* the increase could be a compensatory mechanism. MORC6 is predicted to have a role in higher order heterochromatin formation and the movement of this chromatin to the nuclear membrane (Moissiard et al. 2012). When perturbed in *morc6* mutants the plant could compensate for this loss in high order heterochromatin by increasing DNA methylation. Indeed a similar compensatory event has been seen previously in DNA methylation pathways. In *met1 drm1 drm2* triple mutants, gene body CG methylation is eliminated, however CMT3 acts to methylate CHG residues in the gene bodies to compensate (Cokus et al. 2008). However both the differences in the mutants and WT lines could be due to technical issues with the kit used. In order to accurately quantify the DNA methylation levels the concentration of the DNA used must be known and this can be inaccurate particularly in low concentration samples. Steps were taken to

minimise this inaccuracy, including using a highly concentrated DNA sample and taking repeated measurements of the DNA concentration and using an average of these values. This should limit inaccuracies in regard to concentration and indeed the technical replicates for each sample were similar. However, another problem with the assay is that when comparing different plates the end point of the assay is subjective in regard to the intensity of the colour at which the measurement is taken for each plate. This could affect the relative DNA methylation levels and so explain the significant difference between the 1st and 2nd biological replicates. It is possible that the difference is due to natural variation in DNA methylation but in this case I would favour a technical error. For this reason in order to get a more accurate assessment of global DNA methylation the analysis using the kit should be repeated with both biological and technical replicates on the same plate. This would need to be completed before making firm conclusions on changes in global DNA methylation in the mutants.

5.4.3 *morc6* mutant's effect on flowering time

The flowering time assay produced contrasting results for M1 and M9, with M1 causing a delay in flowering time and M9 bringing forward flowering time. This difference is therefore not due to *morc6* and indicates that other EMS mutations carried by the M1 and M9 mutants are affecting the assay. Both M1 and M9 have been backcrossed twice to remove other EMS mutations carried in the two lines, but as sequencing of M1 revealed this line still carries fifteen EMS mutations that lie in exons and that cause a change in amino acid sequence and there is likely to be more still segregating in the M1 population (Table 3.3). Mutations in one or several of these genes may therefore skew the results and make it hard to discern the effect of *morc6* on flowering time. This is exemplified by the necrotic lesion phenotype in M1, where 91.6% of the plants used in the flowering time assay had the necrotic lesion phenotype. The necrotic lesion phenotype may also be the cause of the delay in flowering time in M1. To determine the actual effect

of *morc6* mutants on flowering time the assay could be repeated on the F1 progeny of a M1 M9 cross as this should be homozygous for *morc6* but heterozygous for all other EMS mutants. As well as the reduction in length of flowering time in M9, *rmd1* also has a reduced flowering time, seen in the lower leaf number (LN). This is contrary to previous findings that *nrpe1* mutants cause a delay in flowering time (Pontier et al. 2005). This contradictory result may be due to ecotype differences in flowering as my study has used lines in a C24 background whereas the *nrpe1* mutant in the Pontier study was in the Columbia background. To determine if there if this is an ecotype difference *morc6* and *nrpe1* mutants in several difference ecotypes could also be tested using this assay.

The necrotic phenotype would also be an interesting candidate for future study. The phenotype is observed in M1 but not M9 suggesting that it is not a result of the *morc6* mutations but another mutation carried by M1 and the high percentage of M1 plants displaying the phenotype suggests that it is closely linked to the *morc6* locus. It is also possible that as M1 is a null allele whereas M9 still produces a truncated MORC6 protein that this phenotype is in fact caused by MORC6 but is only apparent in a full kncokout. However, I would argue that the phenotype is not associated with the M1 *morc6* mutation because there are three genes with an amino acid substitution and a known function that could cause the necrotic phenotype. The three candidate genes are *NTT2, FMO1* and *LIG1,* of which *NTT2* is the best candidate as a previous study showed that *ntt2* has a necrotic lesion phenotype in short day growth conditions (Reinhold et al. 2007). NTT2 is a ATP/ADP plastid antiporter and it is thought that a lack of ATP during the night in *ntt2* mutants results in necrotic cell death through photooxidative damage (Reiser et al. 2004, Reinhold et al. 2007). However, the amino acid change in *NTT2* in M1 is alanine to valine, which only results in a change in side chain size so may not necessarily affect function (Table 3.3). *FMO1* is involved in pathogen defence and mutants display a reduced ability to acquire immunity to the pathogen (Mishina and Zeier 2006). However, mutations of

FMO1 could only cause necrotic lesions in response to an infection, which is unlikely to have occurred. *LIG1* is a DNA ligase involved in DNA replication and base excision repair, so in this case as base excision is compromised, cell death could occur if mutations are not repaired (Taylor et al. 1998). The amino acid substitution in *LIG1* is glutamic acid to lysine so results in a change from an acid to a base and this could conceivably have an effect on LIG1 function. Another possibility is a gene with an unknown function, At1G11200, which has a premature stop codon mutation in M1.

The flowering time assay also established that line 142S, and any line derived from it, has a reduced size compared with 142 and WT C24. This was quantified by measuring the rosette weight at the point of flowering. As mentioned during the discussion of the effect of the 142S on global DNA methylation, this may be due to a change in the 142S epigenome. Further work is required to determine whether this is the case and what the cause of the epigenetic change is. In short day conditions both M1 and M9 have a statistically significant decrease in plant size compared to 142S, although this is not seen in long day conditions. Again, as with the flowering time itself, there is a possibility of other mutations carried in the two mutant lines causing this result. However, as there is no statistically significant difference in plant size between M1 and M9 this would suggest that *morc6* is involved in the reduction in plant size compared to 142S. *rmd1* also causes a reduction in plant size so would also support the idea that RdDM mutants cause a reduction in plant size, possibly through failure to silence a negative regulator of growth or leaf expansion. However, this finding would need to be retested in lines that do not carry other mutations and are from different ecotypes, to determine if this is an ecotype-specific effect.

5.5 Acknowledgements

The siRNA northern blot detection of Solo LTR siRNAs was carried out by Louise Jones. Tom Smith provided assistance with the Epigentek Methyl Flash DNA methylation kit.

6. General discussion

6.1 MORC6 function in RdDM

The aim of this study was to identify novel RdDM genes and characterise both their phenotype and function in the pathway. This study has identified three mutant alleles in the *MORC6* gene and determined that this gene has a role in RdDM. The MORC6 protein is a GHKL ATPase domain containing protein with a coiled coil domain at the C terminal end of the protein. However, two other studies have also identified MORC6 as an RdDM component, thus *MORC6* is no longer a novel RdDM gene (Lorković et al. 2012, Moissiard et al. 2012). The phenotype from this study can however be used, in conjunction with the previous two studies to provide a better understanding of MORC6's role in RdDM. The evidence from both this and the previous two studies into MORC6 function in RdDM point to multiple roles in the pathway. In all three studies siRNA levels are not affected and so would preclude a role for MORC6 in siRNA production (Lorković et al. 2012, Moissiard et al. 2012). There is however a reduction in DNA methylation that occurs in a locus specific manner. This would therefore suggest that MORC6 has a role in facilitating the methylation of DNA by DRM2 and raises the question of what MORC6's exact role during this process is.

Findings from the study by Lorković suggest a possible function for MORC6 in DNA methylation. The study showed that MORC6 interacts with DMS3, a hinge domain containing protein, through coiled coil domains found on both proteins (Lorković et al. 2012). DMS3 has two coiled coil domains so can in theory interact with two MORC6 proteins and it was noted that the resulting complex bears a striking similarity to structural maintenance of chromosomes (SMC) protein, which are characterised by two ATPase domains, in this case provided by the two MORC6 proteins, and a hinge domain, provided by the DMS3 protein (Hirano 2005, Lorković et al. 2012). The complex of DMS3 and two MORC6 proteins is therefore described as SmcMORC6 in accordance with nomenclature of other SMC proteins. SMC proteins function by oscillating between a closed and open

conformations, which is controlled by the binding and hydrolysis of ATP respectively (Figure 6.1 A) (Hirano 2005). Upon ATP binding to the two GHKL ATPase domains SmcMORC6 may undergo a conformational change resulting in the two MORC6 proteins binding to each other forming a closed circle that could clamp onto a DNA or RNA strand or strands (Figure 6.1 B). SMC proteins can also bind to other SMC proteins upon ATP binding and so form hooks that bring DNA or RNA strands into close proximity (Figure 6.1 B). It is not known whether SmcMORC6 can do either or both of these interaction types, so further analysis would be required to both confirm the existence of the SmcMORC6 complex and determine its conformational changes. The human MORC3 protein forms a homodimers that is thought to behave similarly to a SMC protein and so could indicate that a MORC6 homodimer may also be able to carry out a similar function (Mimura et al. 2010). It is known that MORC6 can form a homodimer as well as a binding to DMS3 but it is not known if either or both are involved in RdDM and if they are both involved whether they have differing functions.

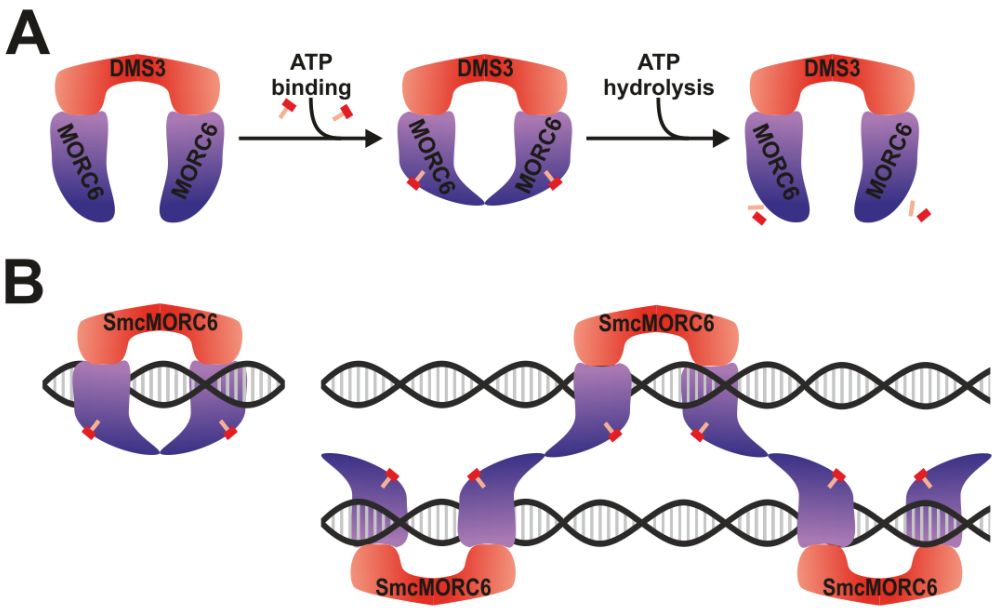


Figure 6.1: SmcMORC6 open and closed conformations and DNA binding possibilities

A: Model proposing how a SmcMORC6 complex undergoing ATP binding and hydrolysis may result in a change between the closed and open conformations of the complex. **B:** Model proposing the formation of circle (left) and hook (right) conformations of SmcMORC6. The hook formation requires the interaction between multiple SmcMORC6 complexes. DMS3 is shown in red and the two MORC6 proteins are shown in purple. ATP is shown as a red and orange T shape, which is broken apart upon ATP hydrolysis. The DNA strands are shown as black double helices with grey bars representing the base-base pairing. This diagram has been adapted from Hirano et al, 2005.

In terms of the SmcMORC6 complex's function in RdDM, DMS3 has previously been shown to form the DDR complex with RDM1 and DRD1 and this complex is required for PolV transcription of the target DNA (Kanno et al. 2008, Ausin et al. 2009, Law et al. 2010, Lorković et al. 2012). Mutation of *MORC6* results in a decrease in PolV transcript accumulation and so would suggest that MORC6 is also part of the DDR complex (Figure 6.2 B). In this case the SmcMORC6 complex is likely to form a circular conformation around the target DNA strand, rather than the hook conformation, and could, along with RDM1, provide a stable platform for the nucleosome modifier, DRD1 to modify the chromatin structure of the target DNA strand, thus allowing PolV transcription to occur (Kanno et al. 2004, Gao et al. 2010, Lorković et al. 2012). Since DMS3 is required for the DDR complex this would favour a SmcMORC6 complex rather than a MORC6 homodimer being involved.

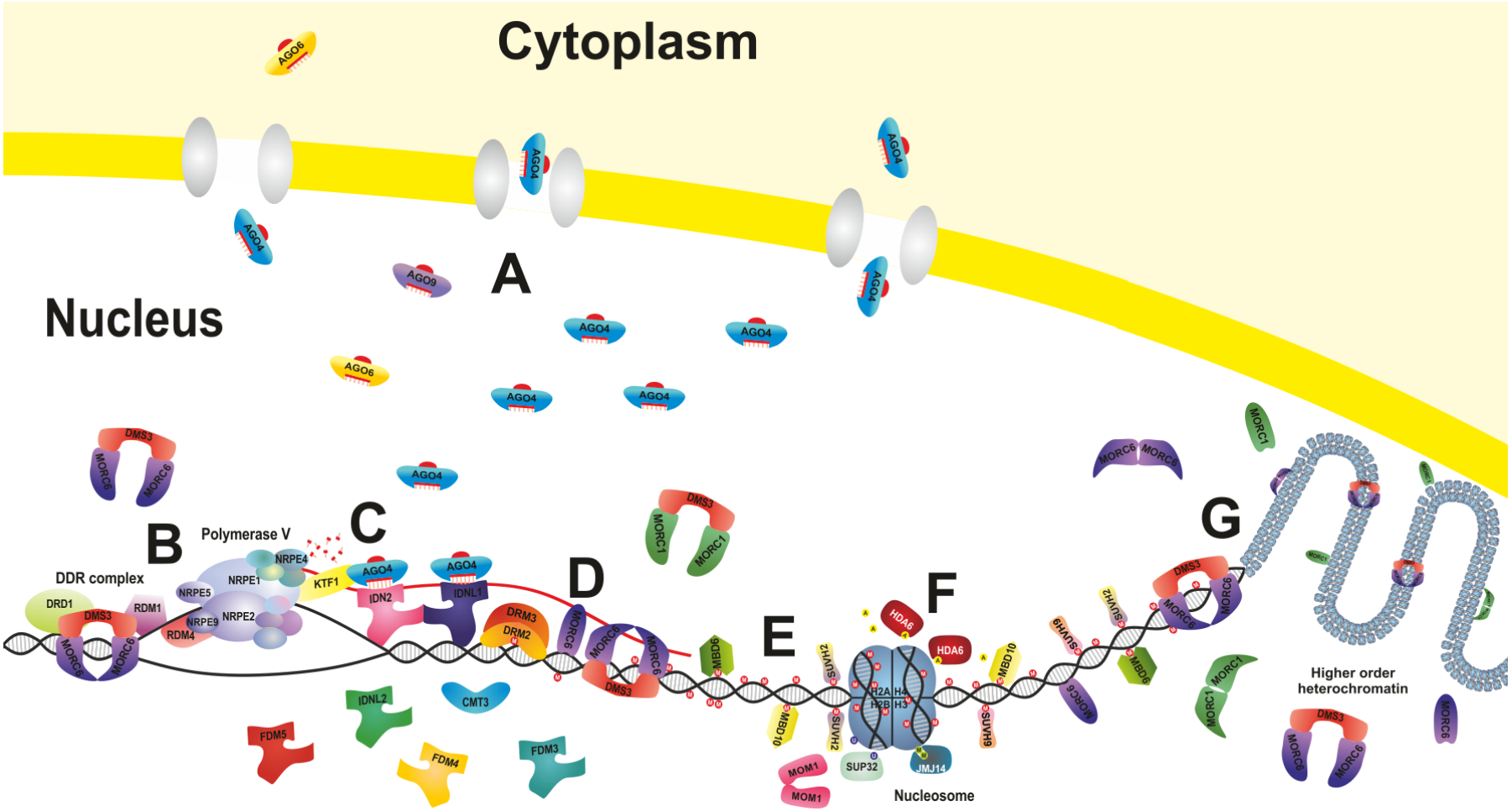

Cytoplasm
Nucleus
A
B
Polymerase V
DDR complex
C
D
E
F
G
Nucleosome
Higher order heterochromatin
AGO6
AGO4
AGO9
DMS3
MORC6
MORC1
NRPE4
NRPE1
NRPE5
NRPE2
NRPE9
KTF1
RDM1
RDM4
DRD1
IDN2
IDNL1
IDNL2
DRM3
DRM2
CMT3
FDM5
FDM4
FDM3
SDMBD5
MBD10
MBD6
SUVH2
SUVH9
HDA6
MOM1
SUP32
JMJ14
H2A
H2B
H3
H4

Figure 6.2: Role of MORC6 in the RdDM pathway

A: siRNA loaded AGO reentre the nucleus. **B:** PolV transcribes the target locus and this activity is facilitated by the DDR complex including MORC6. **C:** The siRNA loaded AGOs bind to the PolV transcript and this aided by interaction with KTF1. The siRNA-PolV transcript duplex is recognised by a IDN2 heterodimer that then binds to the duplex. **D:** A SmcMORC6 complex, MORC6 homodimer (not shown) or MORC6 protein binds to the PolV transcript and target DNA locus which brings the two into close proximity. This allows the IDN2 heterodimer then interacts with the DNA through its zinc finger motifs, thus triggering DNA methylation by DRM2 and its co-factor DRM3. **E** Methylated cytosine binding proteins such as MBD6, MBD10, SUVH2 and SUVH9 bind to the DNA locus at methylation sites. **F:** The methylated cytosine binding proteins interact with histone modifiers, such as HDA6, SUP32 and JMJ14 leading to modification of the nucleosomes into a heterochromatic state. **G:** DNA methylation and histone modification at the target locus is recognised and triggers higher order heterochromatin formation which then localises to the nuclear membrane. MORC6 and MORC1 proteins are involved in this process as either part of SMC like complex with DMS3, lone proteins or as homodimers. The nuclear envelope is depicted as a yellow line with protein channels within this membrane shown as pairs of silver ovals. Areas with a white background are within the nucleus and areas in pale yellow are in the cytoplasm. RNA strands are shown as in red while DNA strands are black. RNA nucleotides are shown as red and orange T shapes. A single nucleosome has been depicted in this diagram in blue and four of the eight histone proteins that make up the nucleosome are shown. DNA methylation marks are shown as white Ms in red circles while histone methylation marks are shown as black Ms in green circles. Histone acetylation marks are shown as black As in yellow circles and histone ubiquitination marks are shown as white Us in purple circles

Another possibility for SmcMORC6 function, suggested by Bender, is that the complex clamps both the target DNA and PolV transcript so that the two are in close proximity to each other (Bender 2012). This would provide part of the solution to how the binding of siRNA-loaded AGO4 complexes to the PolV transcript triggers DRM2 to methylate the corresponding DNA sequence. The current view of how this binding results in methylation is that initially siRNA-AGO4 complexes bind to the PolV transcript soon after the complementary sequence to the siRNA has been transcribed. This step may be facilitated by KTF1 which can interact with both AGO4 and PolV (Bies-Etheve et al. 2009, He et al. 2009c, Wang and Dennis 2009, Rowley et al. 2011). The RNA duplex between the siRNA and PolV transcript is then recognised and bound by a dimer comprising of IDN2 or one of its homologs (IDP1, IDP2 or IDNL2) (Ausin et al. 2009, Ausin et al. 2012b, Xie et al. 2012, Zhang et al. 2012). The zinc finger motifs of IDN2 and its homologs are then thought to interact with the DNA and it is proposed that this is the trigger for DRM2 to methylate the DNA. However, this mechanism only works if the DNA and RNA strands are in close enough proximity in order for IDN2 to bind to both the DNA and RNA. Thus SmcMORC6 may provide a mechanism by which the DNA and RNA strands are clamped together and hence are in close enough proximity for IDN2 binding (Figure 6.2 D). This function could occur in either the hook or circle formation and further work would be required to firstly demonstrate the clamping mechanism and secondly to determine which conformation the SmcMORC6 complex is in. Unlike the role in the DDR complex this role could also be carried out by a MORC6 homodimer.

The study by Moissiard has also suggested another function for MORC6 in the pathway. The study found that in *morc6* mutants, transcripts from RdDM targets were upregulated but that this does not always correspond to a reduction in DNA methylation (Moissiard et al. 2012). This suggests that although the loci are methylated they are not fully silenced and could therefore lack chromatin modification in response to the DNA methylation. Analysis of histone modification showed no change in *morc6* mutants, however mutants have a

reduction in the proportion of pericentromeric chromocentres that is a result of a loss in higher order heterochromatin structure. Higher order chromatin involves the formation of tertiary and quaternary DNA structures such as solenoid-like coils of nucleosomes and the bunching together of large loops of these coils, both of which result in the DNA becoming more condensed and inaccessible to proteins that bind to DNA (Woodcock and Ghosh 2010, Li and Reinberg 2011). Although disruption of higher order chromatin formation has not been demonstrated specifically for RdDM targets upregulated in *morc6* mutants, the changes in pericentromeric heterochromatin organisation would suggest MORC6 could have a role in higher order chromatin modification in RdDM. In this scenario MORC6 would facilitate the formation of higher order chromatin structures in response to DNA methylation at RdDM targets and the resulting heterochromatin would then associate with the nuclear envelope, thus becoming fully silenced (Figure 6.2 G) (Ye et al. 1997, Solovei et al. 2004, Fang and Spector 2005, Moissiard et al. 2012). However, it is not known what the exact function of MORC6 is in the formation of higher order chromatin. It is not also known if this activity requires DMS3 as no study has implicated DMS3 in higher order heterochromatin modification. This may suggest that a MORC6 homodimer rather than SmcMORC6 is involved in higher order heterochromatin modification, but since SMC proteins are involved in such modifications it is possible that the SmcMORC6 complex is also involved in this process (Hirano 2005).

The three possible functions for MORC6 in RdDM so far described may also involve proteins other than DMS3. SMC proteins are known to bind to other proteins in order to provide specificity to the resulting complex, for example the condensin complex that have a role in chromosome condensation and separation during eukaryotic mitosis have several none SMC proteins associated with the complex (Hirano et al. 1997, Sutani et al. 1999, Fujimoto et al. 2005). Other proteins, such as zinc finger motif containing proteins, could therefore target the SmcMORC6 complex to specific loci which could explain why MORC6

appears to only have a role in DNA methylation at specific loci. Other proteins could facilitate catalytic activity, particularly with regard to higher order chromatin modificaiton. It is possible that some of the proteins that bind specifically to methylated cytosines in the RdDM pathway, SUVH2, SUVH9, MBD6 and MBD10, could be recognised by MORC6, prompting its binding to the DNA and initiation of higher order chromatin modification (Johnson et al. 2008, Preuss et al. 2008). Further analysis using co-immuno precipitation and yeast 2-hybrid studies would be required to determine whether other proteins, in addition to DMS3, also bind to MORC6.

The previous two studies suggest that MORC6 functions either as a dimer or as part of a complex with DMS3 (Lorković et al. 2012, Moissiard et al. 2012). However, evidence from this study would suggest MORC6 has a RdDM function that does not require either DMS3 or dimerisation. This is because there is a detectable phenotypic difference between M1, a full knockout of *morc6,* and M9, which is predicted to produce a truncated form of the MORC6 protein that includes part of the GHKL ATPase domain but not the coiled coil domain. In M1 only around 20% of mesophyll cells silence the *GFP* transgene system used in this study, whereas in M9 around 50% of mesophyll cells silence the transgene. The fact that even a truncated form of the GHKL ATPase domain is capable of maintaining some level of silencing would suggest that the coiled coil domain and hence DMS3 and MORC6 dimerisation, is not required for some forms of MORC6 activity. The exact nature of this lone function of MORC6 is unknown as it is not known whether all cells in the silenced mesophyll cell population are methylated and so the solo function could be in DNA methylation or higher order chromatin modification or in both. It should be noted that although this function does not require the coiled coil domain it cannot be discounted that proteins could bind to the N-terminal section of MORC6 and so still be able to confer some level of normal function to the truncated protein.

6.2 Redundancy with other MORC proteins

There are a total of seven *MORC* genes within the *Arabidopsis* genome, of which, only *MORC1* and *MORC6* have been shown to be involved in RdDM (Moissiard et al. 2012). It is not known if any of the other five *MORC* genes are involved in RdDM and further analysis is required in order to determine this. Of the other five *MORC2* is the best candidate due to its close homology to *MORC1* and the fact that both proteins have been shown to have a redundant role in pathogen response (Kang et al. 2008, Kang et al. 2010). *MORC3* is perhaps the least likely of the five remaining *MORC* genes as homozygous mutants in this gene are embryonically lethal, suggesting a highly specialised role for *MORC3* during early plant development. This role would be interesting to identify, even if not associated with RdDM, although the lethality of the homozygous mutant would make it difficult to test for MORC3's function.

Analysis of the two *MORC* genes with a known role in RdDM by Moissiard showed double mutants of *morc1* and *morc6* do not have a strong additive effect on the number of loci where there is an increase in expression levels compared to the single mutants (Moissiard et al. 2012). This means that they regulate the expression of a similar subset of loci and therefore show redundancy in regard to their targets. However, as both single mutants result in activation at these loci they are not redundant in terms of function. This is demonstrated by their effect on DNA methylation of the AtMu1 TE. In both this study and the one by Moissiard it was shown that *morc6* mutants do not cause a reduction in methylation of AtMu1, however Moissiard showed that there is a reduction in AtMu1 methylation in a *morc1* mutant. As both MORCs cause a release in AtMu1 expression this would suggest that MORC1 is required for DNA methylation while MORC6 may be required for higher order chromatin modification. This would mean that for different loci the two MORC proteins have different functions in the RdDM pathway, either in DNA methylation or in higher order heterochromatin formation and so would explain the locus specific loss of DNA

methylation in *morc6* mutants. Currently the methylation status of five RdDM targets has been assessed in *morc1* mutants and fifteen for *morc6* mutants, so more targets need to be tested to determine if there is indeed locus specificity between the two MORC proteins. If other MORC proteins are involved in RdDM these may also have a locus specific role in DNA methylation.

In both *morc1* and *morc6* mutants the number of RdDM targets with increased expression levels is larger than the number of loci that have a reduction in methylation and so raises the question of why DNA methylation seems to have target specificity but not higher order chromatin modification. One explanation for this could be MORC1's and MORC6's respective binding partners. Sequence differences between the two MORCs could mean that they are able to bind a different set of proteins to each other. For example if these included different zinc finger proteins they would recognise different DNA sequences and hence target the MORCs to specific loci. The reason why higher order heterochromatin modification would not be affected in this way could be that as modification is in response to DNA methylation and repressive histone modification, the proteins that target MORC1 and MORC6 to the correct targets recognise DNA methylation and histone marks rather than specific sequences. Therefore there may be two subsets of proteins that bind to MORC1 and MORC6, ones that target the MORCs for DNA methylation activity and ones that target MORCs for higher order chromatin modification. This would require further investigation of protein interactions for MORC1 and MORC6 in order to prove.

6.3 MORC activity in Plantae and Metazoa

RdDM is a plant specific silencing pathway as the key components PolIV and PolV have only found in higher plants, but other eukaryotes do have similar transcriptional gene silencing pathways (Pal-Bhadra et al. 2004, Verdel et al. 2004, Wang et al. 2006, Luo and Hall 2007, Ream et al. 2009). This raises the question of whether MORC proteins found in other eukaryotes are associated

with RNA silencing or RNA interference. Amongst higher plants, other species have orthologs of both the *Arabidopsis MORC1* and *MORC6* genes and so it would seem likely that these MORCs are associated with RdDM. However, there is no experimental evidence to support this and so would require further investigation. There has been large scale divergence in MORCs between the higher and lower plant species suggesting a significant expansion or change in function. It is also interesting to note that RdDM has evolved between lower and higher plants as lower plants have PolIV but lack PolV and so this rapid divergence in MORCs could be related to the changes in RdDM (Luo and Hall 2007). The fact that MORC6 activity in DNA methylation centres around PolV would support this idea.

There is some evidence that Metazoan MORC proteins may have a similar role to the *Arabidopsis* MORC6, however much of this evidence is circumstantial. Mutation of the *C. elegans* MORC1 protein has been shown to release silencing of a *GFP* transgene in seam cells (Moissiard et al. 2012). A similar release of silencing also occurs in mutants defective in the RNA interference component *rde4*, however no direct link has been demonstrated between the *C. elegans* MORC1 and the RNA interference pathway. *C. elegans* also lacks DNA methylation so any interaction between MORC1 and an RNA interference pathway cannot be connected to methylation as it is with RdDM in *Arabidopsis* (Simpson et al. 1986, Moissiard et al. 2012). The *H. sapiens* and *M. musculus* MORC1 proteins are both localised to the male germ cells and in mice mutation of the gene results in infertility (Watson et al. 1998, Inoue et al. 1999). *M. musculus* MORC1 is localised to the nucleus and is thought to be involved in higher order chromatin modification during meiosis. The PIWI RNA (piRNA) silencing pathway which silences transposable elements in Metazoan germ cells also results in male sterility when components of the pathway are mutated in mice (Carmell et al. 2007, Kuramochi-Miyagawa et al. 2008). Both RdDM and piRNA silencing target transposable elements and since MORC1 and MORC6 are associated with RdDM in *Arabidopsis,* it is possible that MORC1 may therefore be

associated with the PIWI pathway, however no direct link between the two has been observed (Carmell et al. 2007, Slotkin et al. 2009, Mosher et al. 2009, Lorković et al. 2012, Moissiard et al. 2012). When considering the domain structure of Metazoan MORC proteins the majority of MORCs have multiple coiled coil domains and a zinc finger motif as well as the GHKL ATPase domain. This means that they could theoretically be able to perform a similar function to that of the *Arabidopsis* MORC1 and MORC6, since their coiled coil domains could allow for dimerization and interaction with a DMS3 homolog, but they may also have other functions due to the presence of further coiled coil domains and the zinc finger motif. It is possible that *Arabidopsis* MORC1 and MORC6 bind to a zinc finger motif containing protein that provides similar functionality to the domain in Metazoan MORCs, but it would requires further investigation to determine if there is a divergence in function. This means that there is no conclusive evidence of an association between RNA silencing or RNA interference pathways and MORC proteins outside *Arabidopsis*.

6.4 Future directions

Having identified MORC6 as an RdDM component and shown that it has a locus specific effect on DNA methylation there are several avenues of research that I would wish to pursue if this project was continued. In terms of short term aims the first would be to use a cell sorter to separate the unsilenced and silenced protoplasts for the M9 line and then assess DNA methylation in the silenced population by bisulfite sequencing. This would demonstrate whether or not all silenced cells have methylation and therefore if loss of methylation results in loss of silencing of the *GFP* transgene. Another aim would be to re-assess the DNA methylation status of Solo LTR using bisulfite sequencing for 7 dpg and 21 dpg DNA samples to get a better idea of the changes to the DNA methylation pattern at this locus. This would then be expanded to include other loci that I have not tested in this study to try and identify further loci specific effects on DNA methylation in *morc6* mutants. Ideally the methylome would be obtained for the

mutants and this could be analysed to identify regions where there are changes in DNA methylation. The final short term aim would be to test the DNA methylation status of DNA loci in other *morc* mutants to determine if these other MORCs are involved in RdDM and if they result in loss of methylation at different loci to MORC6. Double and triple *morc* mutants would also be created and their effect on DNA methylation compared against single mutants. this would show if there is redundancy between MORC proteins are certain loci. Long term aims for the project this would include carrying out protein interaction analysis to identify whether MORC6 functions in a complex. This should help to determine its function in higher order chromatin formation. Further biochemical analysis may also be required to determine this function in higher order heterochromatin formation. Another long term aim would be to assess the function of MORC proteins in other Plantae species, particularly the *M. truncatula* MORC that lacks a coiled coil domain but has a zinc finger motif. This protein would especially be of interest as the difference in domain structure would suggest an effect on function.

6.5 Concluding remarks

Since its discovery, our understanding of the RdDM pathway has increased rapidly so that now most of the mechanism is well understood. The identification of MORC6 in this study and the two previous studies adds to our understanding of the pathway. This protein has roles in higher order chromatin modification and DNA methylation in the RdDM pathway, of which its role in DNA methylation may help solve one of the unresolved issues in RdDM mechanism, that of how siRNAs trigger DRM2 to methylate the DNA target. There are six paralogs of *MORC6* in *Arabidopsis,* some of which may also have roles in RdDM and are also part of a wider family of *MORC*s found in both Plantae and Metazoa. This family could have an ancient role in forming higher order chromatin structure.

7. Appendices

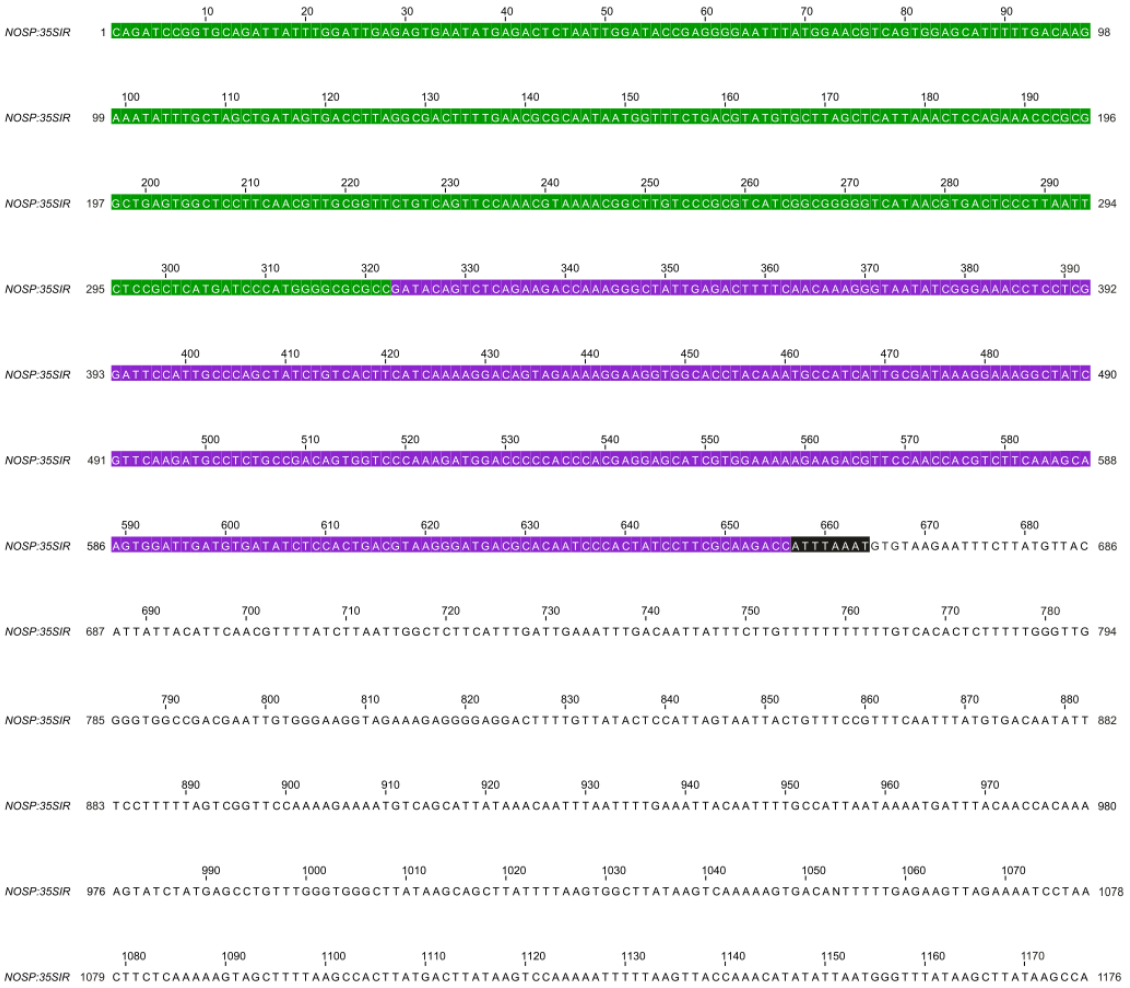
Appendix 1

Appendix 1.1 *35Sp:GFP* and *NOSp:35S IR* transgene sequences


35Sp_GFP 1 AGATTAGCCTTTTCAATTTCAGAAAGAATGCTAACCCACAGATGGTTAGAGAGGCTTACGCAGCAGGTCTCATCAAGACGATCTACCCGAGCAATAATCTCCAGGAAATCAAATACCTTCCCAAG 125
35Sp_GFP 126 AAGGTTAAAGATGCAGTCAAAAGATTCAGGACTAACTGCATCAAGAACACAGAGAAAGATATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAAGGCTTGCTTCACAAACCAAG 250
35Sp_GFP 251 GCAAGTAATAGAGATTGGAGTCTCTAAAAAGGTAGTTCCCACTGAATCAAAGGCCATGGAGTCAAAGATTCAAATAGAGGACCTAACAGAACTCGCCGTAAAGACTGGCGAACAGTTCATACAGA 375
35Sp_GFP 376 GTCTCTTACGACTCAATGACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAA 500
35Sp_GFP 501 CAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCAT 625
35Sp_GFP 626 CGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTG 750
35Sp_GFP 751 ACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGAACACGGGGGACTCTAGAGGATCCAAGGAGATATAACAATGAAGAC 875
35Sp_GFP 876 TAATCTTTTTCTCTTTCTCATCTTTTCACTTCTCCTATCATTATCCTCGGCCGAATTCAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATG 1000
35Sp_GFP 1001 GGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCT 1125
35Sp_GFP 1126 TATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTCAAGGACGACGGGAACTACAA 1250
35Sp_GFP 1251 GACACGTGCTGAAGTCAAGTTTGAGGGAGACACCCTCGTCAACAGGATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGGAAACATCCTCGGCCACAAGTTGGAATACAACTACAACTCCCACA 1375
35Sp_GFP 1376 ACGTATACATCATGGCCGACAAGCAGAAGAACGGCATCAAAGCCAACTTCAAGACCCGCCACAACATCGAAGACGGCGGCGTGCAACTCGCTGATCATTATCAACAAAATACTCCAATTGGCGAT 1500
35Sp_GFP 1501 GGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGG 1625
35Sp_GFP 1626 CATGGATGAACTATACAAACATGACGAACTCTAAGAGCTA 1665

Figure 7.1A: DNA sequence of the *35Sp:GFP* transgene

The sequence shown is only of the transgene and does not show the flanking sequence of the transgene. The sequence highlighted in purple is the *35Sp* and the sequence highlighted in green is the *GFP* gene. The start and stop codons for the *GFP* gene are highlighted in blue and red respectively.


NOSP:35SIR 1 CAGATCCGGTGCAGATTATTTGGATTGAGAGTGAATATGAGACTCTAATTGGATACCGAGGGGAATTTATGGAACGTCAGTGGAGCATTTTTGACAAG 98
NOSP:35SIR 99 AAATATTTGCTAGCTGATAGTGACCTTAGGCGACTTTTGAACGCGCAATAATGGTTTCTGACGTATGTGCTTAGCTCATTAAACTCCAGAAACCCGCG 196
NOSP:35SIR 197 GCTGAGTGGCTCCTTCAACGTTGCGGTTCTGTCAGTTCCAAACGTAAAACGGCTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCTTAATT 294
NOSP:35SIR 295 CTCCGCTCATGATCCCATGGGGCGCGCCGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCG 392
NOSP:35SIR 393 GATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATC 490
NOSP:35SIR 491 GTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCA 588
NOSP:35SIR 586 AGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCATTTAAATGTGTAAGAATTTCTTATGTTAC 686
NOSP:35SIR 687 ATTATTACATTCAACGTTTTATCTTAATTGGCTCTTCATTTGATTGAAATTTGACAATTATTTCTTGTTTTTTTTTTTGTCACACTCTTTTTGGGTTG 794
NOSP:35SIR 785 GGGTGGCCGACGAATTGTGGGAAGGTAGAAAGAGGGGAGGACTTTTGTTATACTCCATTAGTAATTACTGTTTCCGTTTCAATTTATGTGACAATATT 882
NOSP:35SIR 883 TCCTTTTTAGTCGGTTCCAAAAGAAAATGTCAGCATTATAAACAATTTAATTTTGAAATTACAATTTTGCCATTAATAAAATGATTTACAACCACAAA 980
NOSP:35SIR 976 AGTATCTATGAGCCTGTTTGGGTGGGCTTATAAGCAGCTTATTTTAAGTGGCTTATAAGTCAAAAAGTGACANTTTTTGAGAAGTTAGAAAATCCTAA 1078
NOSP:35SIR 1079 CTTCTCAAAAAGTAGCTTTTAAGCCACTTATGACTTATAAGTCCAAAAATTTTTAAGTTACCAAACATATATTAATGGGTTTATAAGCTTATAAGCCA 1176

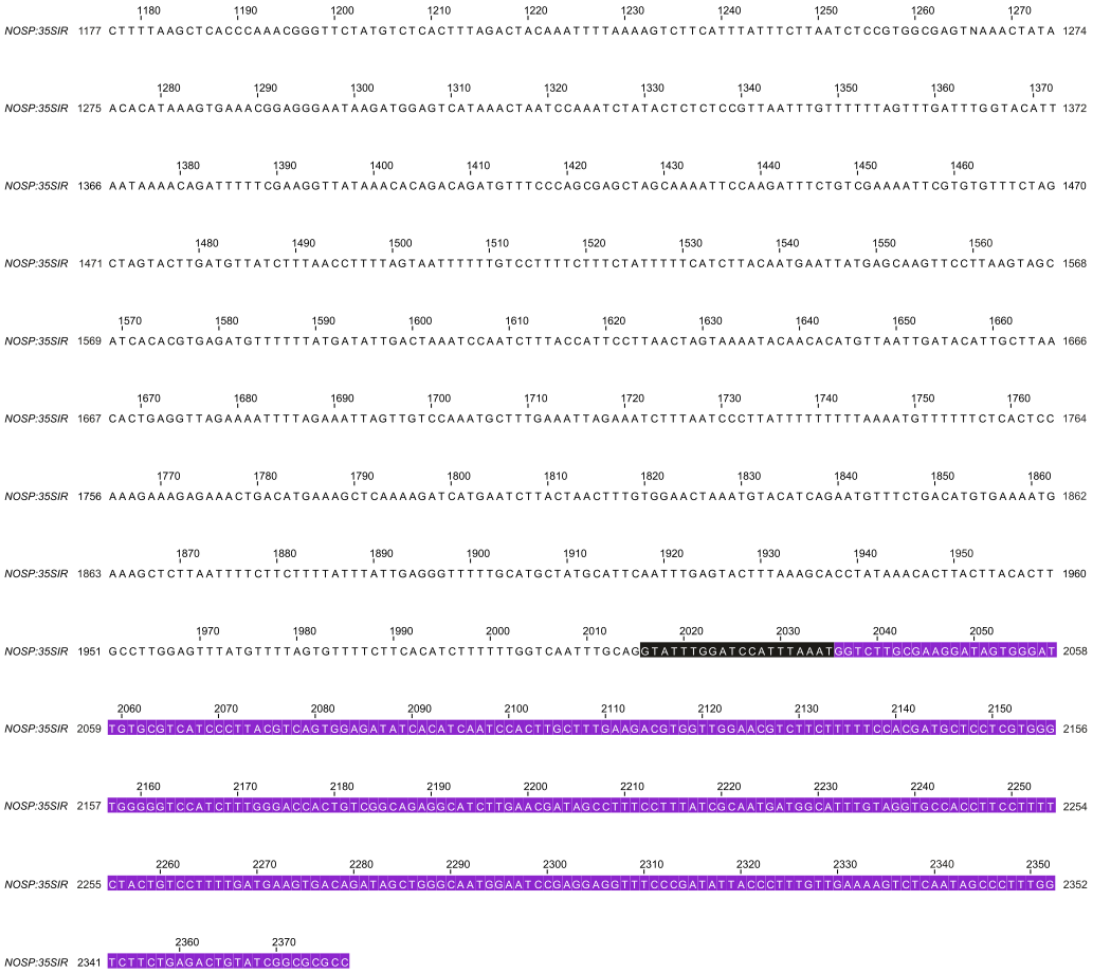

NOSP:35SIR 1177 CTTTTAAGCTCACCCAAACGGGTTCTATGTCTCACTTTAGACTACAAATTTTAAAAGTCTTCATTTATTTCTTAATCTCCGTGGCGAGTNAAACTATA 1274
NOSP:35SIR 1275 ACACATAAAGTGAAACGGAGGGAATAAGATGGAGTCATAAACTAATCCAAATCTATACTCTCTCCGTTAATTTGTTTTTTAGTTTGATTTGGTACATT 1372
NOSP:35SIR 1366 AATAAAACAGATTTTTCGAAGGTTATAAACACAGACAGATGTTTCCCAGCGAGCTAGCAAAATTCCAAGATTTCTGTCGAAAATTCGTGTGTTTCTAG 1470
NOSP:35SIR 1471 CTAGTACTTGATGTTATCTTTAACCTTTTAGTAATTTTTTGTCCTTTTCTTTCTATTTTTCATCTTACAATGAATTATGAGCAAGTTCCTTAAGTAGC 1568
NOSP:35SIR 1569 ATCACACGTGAGATGTTTTTTATGATATTGACTAAATCCAATCTTTACCATTCCTTAACTAGTAAAATACAACACATGTTAATTGATACATTGCTTAA 1666
NOSP:35SIR 1667 CACTGAGGTTAGAAAATTTTAGAAATTAGTTGTCCAAATGCTTTGAAATTAGAAATCTTTAATCCCTTATTTTTTTTTAAAATGTTTTTTCTCACTCC 1764
NOSP:35SIR 1756 AAAGAAAGAGAAACTGACATGAAAGCTCAAAAGATCATGAATCTTACTAACTTTGTGGAACTAAATGTACATCAGAATGTTTCTGACATGTGAAAATG 1862
NOSP:35SIR 1863 AAAGCTCTTAATTTTCTTCTTTTATTTATTGAGGGTTTTTGCATGCTATGCATTCAATTTGAGTACTTTAAAGCACCTATAAACACTTACTTACACTT 1960
NOSP:35SIR 1951 GCCTTGGAGTTTATGTTTTAGTGTTTTCTTCACATCTTTTTTGGTCAATTTGCAGGTATTTGGATCCATTTAAATGGTCTTGCGAAGGATAGTGGGAT 2058
NOSP:35SIR 2059 TGTGCGTCATCCCTTACGTCAGTGGAGATATCACATCAATCCACTTGCTTTGAAGACGTGGTTGGAACGTCTTCTTTTTCCACGATGCTCCTCGTGGG 2156
NOSP:35SIR 2157 TGGGGGTCCATCTTTGGGACCACTGTCGGCAGAGGCATCTTGAACGATAGCCTTTCCTTTATCGCAATGATGGCATTTGTAGGTGCCACCTTCCTTTT 2254
NOSP:35SIR 2255 CTACTGTCCTTTTGATGAAGTGACAGATAGCTGGGCAATGGAATCCGAGGAGGTTTCCCGATATTACCCTTTGTTGAAAAGTCTCAATAGCCCTTTGG 2352
NOSP:35SIR 2341 TCTTCTGAGACTGTATCGGCGCGCC

Figure 7.2A: DNA sequence of the *NOSp:35S IR* transgene

The sequence shown is only of the transgene and does not show the flanking sequence of the transgene. The sequence highlighted in purple are the inverted repeat (IR) of the part of the *35Sp* and the sequence highlighted in green is the *NOS* promoter. The section not highlighted is the intron of the *CHS A* gene, which allows for the correct formation of dsRNA from the IR. The section in black is linker sequence that forms a loop after the *CHS* intron is spliced out.

Appendix 1.2 Allelism tests of M1 and M9 with *hda6* and *mom1*

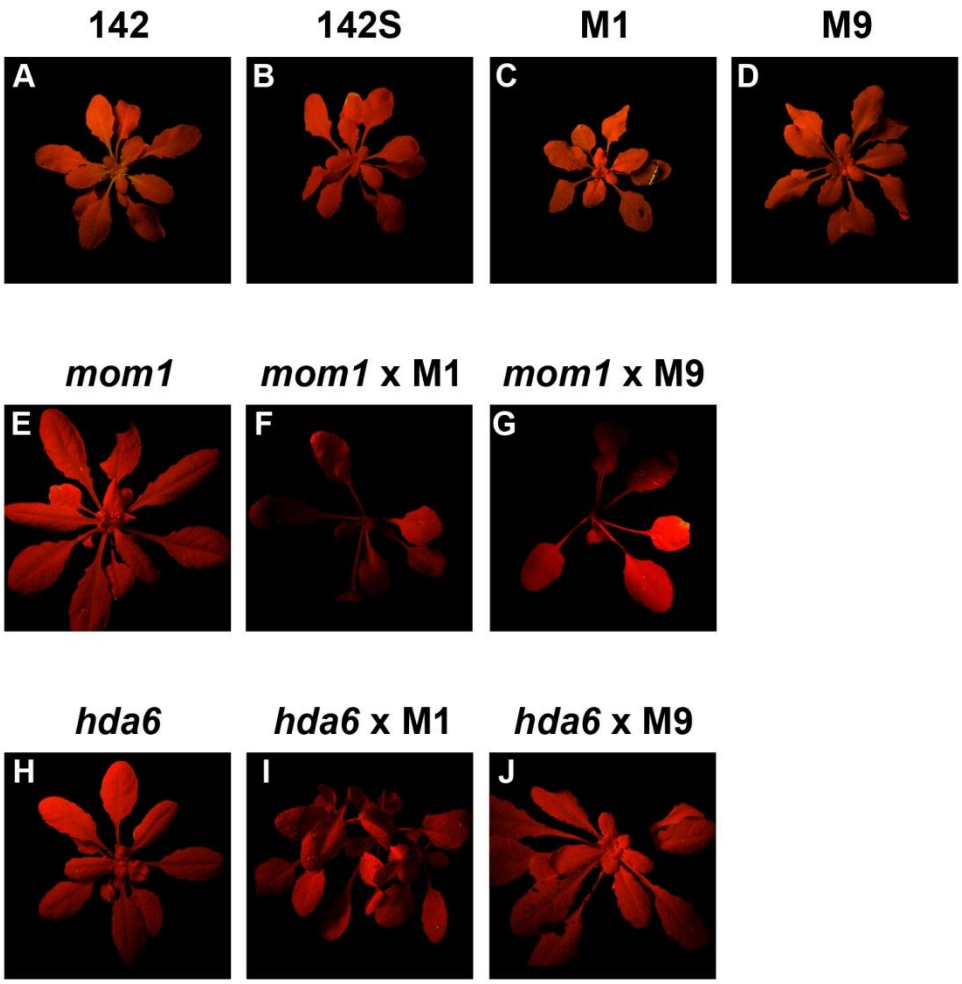


Figure 7.3A: Allelism tests showing M1 and M9 are not *mom1* and *hda6* mutants

Images taken under UV light of parents and crosses. The parental lines shown are 142, 142S, M1, M9, *mom1* and *hda6*. For all crosses the names of the parents are given, the first name is the male and the second is the female. Images are representative for each line. The bright fluorescence seen on leaves for the 142S and *mom1* x M9 images is due to necrosis and has no relevance to the silencing phenotype. The bright fluorescence seen on the right hand leaf of the M1 image is soil on the leaf.

Appendix 1.3 All homozygous SNPs in M1

Chromosome	Scaffold Number	Position within scaffold (bp)	Base in 142S	Base in M1	Proportion of concordance	Number of concordant reads
Chr1_left_arm	4	199539	C	T	0.86	127
Chr1_left_arm	4	220454	C	T	0.85	97
Chr1_left_arm	4	22854	G	A	0.87	111
Chr1_left_arm	4	38773	C	T	0.93	99
Chr1_left_arm	14	129088	C	T	0.86	113
Chr1_left_arm	14	301830	C	T	0.93	90
Chr1_left_arm	20	197351	C	T	0.89	89
Chr1_left_arm	20	310670	C	T	0.89	76
Chr1_left_arm	20	388068	C	T	0.91	105
Chr1_left_arm	20	405956	C	T	0.89	78
Chr1_left_arm	20	610420	C	T	0.86	96
Chr1_left_arm	20	66903	C	T	0.86	96
Chr1_left_arm	24	210128	C	T	0.83	90
Chr1_left_arm	25	110148	C	T	0.87	109
Chr1_left_arm	25	204931	C	T	0.87	127
Chr1_left_arm	25	291069	C	T	0.88	122
Chr1_left_arm	25	328638	C	T	0.85	112
Chr1_left_arm	25	81944	C	T	0.83	80
Chr1_left_arm	26	129417	C	T	0.87	103
Chr1_left_arm	26	278549	C	T	0.93	126
Chr1_left_arm	26	318083	C	T	0.90	96
Chr1_left_arm	26	359473	C	T	0.89	96
Chr1_left_arm	32	166444	C	T	0.90	107
Chr1_left_arm	32	187188	C	T	0.92	119
Chr1_left_arm	36	705641	C	T	0.91	125
Chr1_left_arm	36	89588	C	T	0.86	95
Chr1_left_arm	48	31037	C	T	0.93	132
Chr1_left_arm	49	78579	G	A	0.93	82
Chr1_left_arm	57	2458	C	T	0.94	65
Chr1_left_arm	58	7103	C	T	0.94	113
Chr1_left_arm	66	21383	C	T	0.95	120
Chr1_left_arm	66	93482	C	T	0.88	82
Chr1_left_arm	69	147996	C	T	0.96	127
Chr1_left_arm	69	40455	C	T	0.96	121
Chr1_left_arm	69	95888	C	T	0.96	126
Chr1_left_arm	73	105251	C	T	0.99	78
Chr1_left_arm	73	200949	C	T	0.98	105

Chromosome	Scaffold Number	Position within scaffold (bp)	Base in 142S	Base in M1	Proportion of concordance	Number of concordant reads
Chr1_left_arm	73	2832	C	T	0.97	91
Chr1_left_arm	73	56157	C	T	0.98	106
Chr1_left_arm	74	10327	C	T	1.00	138
Chr1_left_arm	74	481	C	T	0.98	120
Chr1_left_arm	79	103938	A	C	0.83	15
Chr1_left_arm	79	214406	C	T	0.97	99
Chr1_left_arm	79	24264	T	C	0.93	14
Chr1_left_arm	79	24290	T	C	0.93	14
Chr1_left_arm	79	24316	G	T	0.93	14
Chr1_left_arm	79	4095	C	T	0.96	103
Chr1_left_arm	84	118930	C	T	0.97	86
Chr1_left_arm	84	186948	C	T	0.99	92
Chr1_left_arm	88	173559	C	T	0.97	109
Chr1_left_arm	88	229237	C	T	0.96	126
Chr1_left_arm	88	296804	C	T	0.99	134
Chr1_left_arm	88	303908	T	C	0.80	133
Chr1_left_arm	98	20578	C	T	0.98	112
Chr1_left_arm	110	183196	C	T	0.98	106
Chr1_left_arm	110	333140	C	T	0.93	124
Chr1_left_arm	110	840513	C	T	0.93	76
Chr1_left_arm	110	959682	C	T	0.90	112
Chr1_left_arm	110	977488	C	T	0.95	115
Chr1_left_arm	115	44531	C	T	0.96	111
Chr1_left_arm	115	82933	C	T	0.96	87
Chr1_left_arm	115	88160	C	T	0.93	81
Chr1_left_arm	124	520	C	T	0.89	104
Chr1_left_arm	150	120473	C	T	1.00	13
Chr1_left_arm	177	225540	A	G	1.00	13
Chr1_left_arm	194	973	G	A	1.00	12
Chr1_left_arm	201	31931	A	T	1.00	14
Chr1_left_arm	334	22547	T	C	0.81	189
Chr1_left_arm	360	57313	C	T	1.00	8
Chr1_right_arm	569	32269	C	T	0.83	202
Chr1_right_arm	574	28015	G	A	1.00	10
Chr1_right_arm	597	6794	A	G	0.82	36
Chr1_right_arm	599	372	T	G	1.00	14
Chr1_right_arm	613	28105	T	C	0.87	220
Chr1_right_arm	622	10024	T	G	1.00	12
Chr1_right_arm	638	46124	A	G	1.00	15

Chromosome	Scaffold Number	Position within scaffold (bp)	Base in 142S	Base in M1	Proportion of concordance	Number of concordant reads
Chr1_right_arm	693	113281	A	C	0.86	12
Chr1_right_arm	693	35337	A	C	1.00	13
Chr1_right_arm	693	35339	G	T	0.93	13
Chr1_right_arm	693	64633	T	C	0.86	61
Chr1_right_arm	717	31	C	T	0.83	107
Chr1_right_arm	772	50	T	G	1.00	6
Chr1_right_arm	806	27261	A	T	1.00	12
Chr1_right_arm	899	44011	T	A	1.00	11
Chr1_right_arm	916	50346	C	A	0.80	4
Chr1_right_arm	922	7942	A	G	0.80	8
Chr2_left_arm	1141	83041	G	A	1.00	15
Chr2_left_arm	1176	59497	T	C	0.81	54
Chr2_left_arm	1208	38725	T	C	0.83	174
Chr2_right_arm	1514	29	G	A	0.80	12
Chr2_right_arm	1544	2779	A	C	1.00	16
Chr2_right_arm	1544	2780	G	A	1.00	16
Chr2_right_arm	1544	2784	C	A	1.00	20
Chr2_right_arm	1549	54	A	G	0.82	180
Chr2_right_arm	1550	96974	T	C	0.80	140
Chr2_right_arm	1558	394259	T	A	0.80	8
Chr2_right_arm	1575	202590	T	C	1.00	11
Chr2_right_arm	1575	202591	C	T	1.00	12
Chr2_right_arm	1575	202593	T	C	1.00	15
Chr2_right_arm	1584	20730	T	G	1.00	11
Chr2_right_arm	1584	20731	G	A	1.00	11
Chr2_right_arm	1584	662620	G	T	1.00	20
Chr2_right_arm	1636	78709	C	A	1.00	13
Chr2_right_arm	1639	359665	T	C	0.93	13
Chr2_right_arm	1673	588538	T	G	1.00	9
Chr2_right_arm	1686	31716	A	G	0.81	42
Chr2_right_arm	1691	53015	T	G	0.81	17
Chr2_right_arm	1724	83633	A	T	0.91	10
Chr2_right_arm	1724	83634	T	C	0.90	9
Chr3_left_arm	1937	325912	A	T	0.82	125
Chr3_left_arm	1937	326605	T	C	0.80	176
Chr3_left_arm	1937	611249	A	G	1.00	12
Chr3_left_arm	1937	611250	G	T	1.00	12
Chr3_left_arm	1937	62824	T	A	1.00	12
Chr3_left_arm	1951	1001405	C	T	0.86	42

Chromosome	Scaffold Number	Position within scaffold (bp)	Base in 142S	Base in M1	Proportion of concordance	Number of concordant reads
Chr3_left_arm	1976	109607	A	G	0.82	9
Chr3_left_arm	2010	18327	T	C	1.00	22
Chr3_left_arm	2010	18328	C	T	1.00	22
Chr3_left_arm	2020	13122	A	T	1.00	11
Chr3_left_arm	2038	845	T	A	1.00	4
Chr3_right_arm	2312	29492	C	T	1.00	15
Chr3_right_arm	2339	18007	C	T	0.85	72
Chr3_right_arm	2347	1073	C	A	1.00	22
Chr3_right_arm	2428	1813	A	C	1.00	26
Chr3_right_arm	2428	1814	T	A	1.00	19
Chr3_right_arm	2428	3385	C	T	1.00	17
Chr3_right_arm	2571	32431	C	G	1.00	19
Chr3_right_arm	2619	127658	T	A	1.00	12
Chr3_right_arm	2659	53774	T	G	1.00	20
Chr3_right_arm	2659	56942	G	A	1.00	8
Chr4_left_arm	2855	176204	A	G	1.00	15
Chr4_left_arm	2855	56718	T	A	0.89	71
Chr4_right_arm	2956	68240	T	C	0.83	126
Chr4_right_arm	2956	68264	T	C	0.83	129
Chr4_right_arm	2980	147	G	C	0.80	224
Chr4_right_arm	2991	3546	G	T	1.00	13
Chr4_right_arm	3020	1811	G	T	1.00	9
Chr4_right_arm	3022	84053	G	A	0.83	10
Chr4_right_arm	3054	15612	A	T	0.80	12
Chr4_right_arm	3054	15613	A	T	0.82	14
Chr4_right_arm	3056	21488	A	C	1.00	9
Chr4_right_arm	3073	15163	G	A	0.81	184
Chr4_right_arm	3093	18090	C	T	0.92	11
Chr4_right_arm	3101	31011	C	A	1.00	14
Chr4_right_arm	3101	74492	T	A	0.91	10
Chr4_right_arm	3118	61253	C	T	1.00	21
Chr4_right_arm	3256	388360	T	C	0.81	52
Chr4_right_arm	3256	91244	T	C	1.00	14
Chr4_right_arm	3277	213456	C	A	1.00	12
Chr4_right_arm	3278	31224	G	A	1.00	14
Chr4_right_arm	3285	39	G	T	0.93	25
Chr5_left_arm	3550	16104	T	C	0.82	41
Chr5_left_arm	3558	149509	A	G	0.83	99
Chr5_left_arm	3597	37559	A	C	0.81	13

Chromosome	Scaffold Number	Position within scaffold (bp)	Base in 142S	Base in M1	Proportion of concordance	Number of concordant reads
Chr5_left_arm	3605	14863	A	C	1.00	14
Chr5_left_arm	3650	138706	T	A	0.83	5
Chr5_left_arm	3650	138707	A	T	1.00	5
Chr5_left_arm	3650	57614	A	G	0.84	70
Chr5_left_arm	3722	14362	T	A	1.00	14
Chr5_right_arm	3849	3397	A	G	1.00	12
Chr5_right_arm	3862	8703	T	G	1.00	7
Chr5_right_arm	4039	35985	A	G	0.82	84
Chr5_right_arm	4041	5618	T	G	0.91	30
Chr5_right_arm	4058	13411	C	T	1.00	10
Chr5_right_arm	4059	184522	A	G	1.00	11
Chr5_right_arm	4059	189832	C	A	1.00	24
Chr5_right_arm	4147	36733	A	T	1.00	10
Chr5_right_arm	4209	81150	A	C	0.93	315
Chr5_right_arm	4209	81161	A	G	0.94	317
Chr5_right_arm	4216	385116	A	C	0.92	11
Chr5_right_arm	4226	578362	A	G	0.82	40
Chr5_right_arm	4228	123775	C	T	1.00	14
Chr5_right_arm	4229	357992	T	A	0.94	15

Table 7.1A: All SNPs identified in the M1 line

Table showing the location of each SNP in M1 in terms of which chromosome and which arm of the chromosome they are located on. The table also gives which scaffold of the C24 genome they are located on and the SNPs exact position within that scaffold. The base in 142S and M1 is given and the proportion of concordance and number of concordant reads supporting the M1 mutation are given for each SNP.

Appendix 1.4 All left arm of chromosome one SNPs in M1 and exact position in annotated genome

SNP position (bp)	Mutation location	TAIR ID	Gene name	Base change	Proportion concordance in M1	Mutation location and amino acid change
272314	Exon	AT1G01740	AT1G01740	G to A	0.87 (111)	Exon 10 Codon 417: CAA (Glutamine) to UAA (Stop codon)
288235	Intronic	AT1G01790	KEA1	C to T	0.93 (99)	Intron 10
448847	Exon	AT1G02280	TOC33	C to T	0.86 (127)	Exon 6 Codon 264: GGA (Glycine) to AGA (Arginine)
469888	Exon	AT1G02350	AT1G02350	C to T	0.85 (97)	Exon 1 Codon 25: UCC (Serine) to UCU (Serine)
613895	Exon	AT1G02810	AT1G02810	C to T	0.86 (113)	Exon 1 Codon 38: UCC (Serine) to UUC (Phenylalanine)
788675	Intergenic			C to T	0.93 (90)	Intergenic, 1415bp upstream AT1G03230
859201	Intron	AT1G03445	BSU1	C to T	0.86 (96)	Intron 2
989786	Exon	AT1G03890	AT1G03890	C to T	0.89 (89)	Exon 2 Codon 145: UUC (Phenylalanine) to UUA (Leucine)
1101701	Intergenic			C to T	0.89 (76)	Intergenic, 2921bp upstream of AT1G04180 (YUC9)
1180293	Intron	AT1G04390	AT1G04390	C to T	0.91 (105)	Intron 8
1197642	Intergenic			C to T	0.89 (78)	Intergenic, 475bp upstream AT1G04430
1403045	3'UTR	AT1G04945	AT1G04945	C to T	0.86 (96)	3' UTR
1732989	Intergenic			C to T	0.83 (90)	Intergenic 49bp upstream AT1G05790
1864858	Exon	AT1G06140	AT1G06140	C to T	0.83 (80)	Exon 1 Codon 21: AAC (Asparagine) to AAU (Asparagine)
1893527	Exon	AT1G06190	AT1G06190	C to T	0.87 (109)	Exon 1 Codon 49: GGG (Glycine) to GAG (Glutamic acid)
1988596	Exon	AT1G06490	GSLO7	C to T	0.87 (127)	Exon 40 Codon 1884: CUC (Leucine) to CUU (Leucine)
2076617	3'UTR	AT1G06760	AT1G06760	C to T	0.88 (122)	3' UTR of AT1G06760 and AT1G06750
2114176	5'UTR	AT1G06890	AT1G06890	C to T	0.85 (112)	5' UTR
2351992	Intergenic			C to T	0.87 (103)	Intergenic 300bp downstream AT1G07630 (PLL5)
2503774	Exon	AT1G08060	MOM1	C to T	0.93 (126)	Exon 11 Codon 1561: GCG (Alanine) to GCA (Alanine)

SNP position (bp)	Mutation location	TAIR ID	Gene name	Base change	Proportion concordance in M1	Mutation location and amino acid change
2543135	Exon	AT1G08130	LIG1	C to T	0.90 (96)	Exon 16 Codon 756: GAA(Glutamic acid) to AAA(Lysine)
2584446	Intron	AT1G08230	AT1G08230	C to T	0.89 (96)	Intron 4
2908764	Exon	AT1G09030	NF-YB4	C to T	0.90 (107)	Exon 1 Codon 89: AAG (Lysine) to AAA (Lysine)
2929529	Exon	AT1G09080	BIP3	C to T	0.92 (119)	Exon 6 Codon 589: ACG (Threonine) to ACA (Threonine)
3130907	5' UTR	AT1G09660	AT1G09660	C to T	0.86 (95)	5' UTR
3753966	Exon	AT1G11200	AT1G11200	C to T	0.91 (125)	Exon 2 Codon 24: UCG (Serine) to UAG (Stop codon)
3816744	Exon	AT1G11340	AT1G11340	C to T	0.93 (132)	Exon 1 Codon 226: GGA (Glycine) to GAA (Glutamic acid)
4280595	Exon	AT1G12610	DDF1	G to A	0.93 (82)	Exon 1 Codon 80: UUG (Leucine) to UUA (Leucine)
4316338	Intergenic			C to T	0.94 (65)	Intergenic between two pseudogenes, 1970bp upstream AT1G12672
4564003	Exon	AT1G13320	PP2AA3	C to T	0.94 (113)	Exon 12 Codon 527: GUU (Valine) to AUU (Isoleucine)
4798406	Exon	AT1G14000	VIK	C to T	0.95 (120)	Exon 3 Codon 146: CCA (Proline) to UCA (Serine)
4866043	Intron	AT1G14240	AT1G14240	C to T	0.88 (82)	Intron 3
5218865	Exon	AT1G15165	AT1G15165	C to T	0.96 (121)	Exon 4 Codon 119: AGG (Arginine) to AAG (Lysine)
5274409	Exon	AT1G15330	AT1G15330	C to T	0.96 (126)	Exon 1 Codon 14: AUC (Isoleucine) to AUU (Isoleucine)
5326847	Exon	AT1G15500	ATNTT2	C to T	0.96 (127)	Exon 1 Codon 142: GCU (Alanine) to GUU (Valine)
5360698	Intergenic			C to T	0.97 (91)	Intergenic 1310bp upstream AT1G15570 (CYCA2;3)
5419895	Exon	AT1G15750	TPL	C to T	0.98 (106)	Exon 4 Codon 98: GUG (Valine) to AUG (Methionine)
5469631	5'UTR	AT1G15920	AT1G15920	C to T	0.99 (78)	5' UTR
5575831	Intron	AT1G16300	GAPCP-2	C to T	0.98 (105)	Intron 6
5613754	Exon	AT1G16420	MC8	C to T	0.98 (120)	Exon 1 Codon 26: GUC (Valine) to AUC (Isoleucine)
5623557	Intron	AT1G16470	PAB1	C to T	1.00 (138)	Intron 4

SNP position (bp)	Mutation location	TAIR ID	Gene name	Base change	Proportion concordance in M1	Mutation location and amino acid change
5677535	Exon	AT1G16610	SR45	C to T	0.96 (103)	Exon 6 Codon 195: GAU (Aspartic acid) to AAU (Asparagine)
5701373	Intergenic			T to C	0.93 (14)	Intergenic 1379bp upstream AT1G16680
5701399	Intergenic			T to C	0.93 (14)	Intergenic 1353bp upstream AT1G16680
5701425	Intergenic			G to T	0.93 (14)	Intergenic 1327bp upstream AT1G16680
5781761	Intron	AT1G16900	AT1G16900	A to C	0.83 (15)	Intron 1
5892010	Exon	AT1G17230	AT1G17230	C to T	0.97 (99)	Exon 1 Codon 212: AGC (Serine) to AGU (Serine)
6324487	Intergenic			C to T	0.97 (86)	Intergenic 667 bp upstream AT1G18370 (HIK)
6393196	Intergenic			C to T	0.99 (92)	Intergenic 2667bp upstream AT1G18580 (GAUT11)
6595680	Exon	AT1G19100	AT1G19100	C to T	0.97 (109)	Exon 2 Codon 41: CAA (Glutamine) to UAA (Stop codon)
6651377	Exon	AT1G19250	FMO1	C to T	0.96 (126)	Exon 4 Codon 360: GGG (Glycine) to GAG (Glutamic acid)
6729457	Exon	AT1G19440	KCS4	C to T	0.99 (134)	Exon 1 Codon 113: UAC (Tyrosine) to UAU (Tyrosine)
6736478	Intergenic			T to C	0.80 (133)	Intergenic 1595bp upstream of AT1G19450
6952070	Intron	AT1G20060	AT1G20060	C to T	0.98 (112)	Intron 16
7280563	Exon	AT1G20910	AT1G20910	C to T	0.98 (106)	Exon 2 Codon 25: GAA (Glutamic acid) to AAA (Lysine)
7445662	Exon	AT1G21270	WAK2	C to T	0.93 (124)	Exon 1 Codon 222: CCU (proline) to UCU (Serine)
7963776	Intergenic			C to T	0.93 (76)	Intergenic 265bp upstream AT1G22540
8090097	Exon	AT1G22870	AT1G22870	C to T	0.90 (112)	Exon 2 Codon 176: CUC (Leucine) to CUU (Leucine)
8107946	Exon	AT1G22910	AT1G22910	C to T	0.95 (115)	Exon 6 Codon 347: UGC (Cysteine) to UGU (Cysteine)
8352560	3'UTR	AT1G23550	SRO2	C to T	0.96 (111)	3'UTR
8397784	Intron	AT1G23730	BCA3	C to T	0.96 (87)	Intron 8
8403007	Exon	AT1G23760	JP630	C to T	0.93 (81)	Exon 2 Codon 243: GGC (Glycine) to GGU (Glycine)

SNP position (bp)	Mutation location	TAIR ID	Gene name	Base change	Proportion concordance in M1	Mutation location and amino acid change
8450341	Intergenic			C to T	0.89 (104)	Intergenic 725bp downstram AT1G23910
9304860	Intergenic			C to T	1.00 (13)	Intergenic 720gp upstream AT1G26850
10047444	3'UTR	AT1G28590	AT1G28590	A to G	1.00 (13)	3' UTR
10922907	Intergenic			G to A	1.00 (13)	Intergenic 1011bp downstream AT1G30780
11050895	Intergenic			A to T	1.00 (14)	Intergenic 165bp upstream AT1G30980
13085698	Exon	AT1G3552	ARF15	T to C	0.81 (189)	Exon 2 Codon 19: AGA (Arginine) to GGA (Glycine)
13477372	Intergenic			C to T	1.00 (8)	Intergenic 4292bp downstream AT1G36078

Table 7.2A: All homozygous SNPs on the left arm of chromosome one in M1 and their location in the Col genome

Table showing the exact position of each SNP in bp on chromosome one and also shows whether this position is intergenic or genic. If the SNP is within a gene it also gives the gene name and where is the gene the mutation is (UTR, exon or intron). The base change from WT to mutant is also given as well as the proportion of reads that show concordance in this change, with the actual number with this change in brackets.

SNP position (bp)	Mutation location	TAIR ID	Function
272314	Exon	AT1G01740	Protein kinase containing tetratricopeptide repeat domain
288235	Intronic	AT1G01790	Potassium efflux antiporter, ion transporter
448847	Exon	AT1G02280	GTPase involved in chloroplast import machinary
469888	Exon	AT1G02350	Protoporphyrinogen oxidase

SNP position (bp)	Mutation location	TAIR ID	Function
613895	Exon	AT1G02810	Plant inverstase/pectin methyltransferase inhibitor, inhibits cell wall modification
788675	Intergenic		Eukaryotic aspartyl protease
859201	Intron	AT1G03445	Serine–threonine protein phosphatase, involved in response to brassinosteroid
989786	Exon	AT1G03890	RmlC-like cupins superfamily protein, involved in metabolism
1101701	Intergenic		Reduction of NADPH to NADP
1180293	Intron	AT1G04390	BTB/POZ domain containing protein
1197642	Intergenic		S-adenosyl-L-methionine-dependent methyltransferase, involved in protein trafficking
1403045	3'UTR	AT1G04945	HIT-type zince finger protein
1732989	Intergenic		Lipase, lipid cleavage
1864858	Exon	AT1G06140	Pentatricopeptide repeat protein
1893527	Exon	AT1G06190	Rho termination factor, involved in transciption termination
1988596	Exon	AT1G06490	Callose Synthase, involved in creation of sieve elements in phloem
2076617	3'UTR	AT1G06760	Winged helix transciption factor
2114176	5'UTR	AT1G06890	Nodulin MtN21 /EamA-like transporter
2351992	Intergenic		Protein phosphatase
2503774	Exon	AT1G08060	Transciptional silencing
2543135	Exon	AT1G08130	DNA ligase involved in DNA replication and base excision repair
2584446	Intron	AT1G08230	Transmembrane amino acid transporter
2908764	Exon	AT1G09030	Transcription factor
2929529	Exon	AT1G09080	ATP binding protein involved in protein in the ER in response to heat stress
3130907	5' UTR	AT1G09660	RNA binding KH domain containing protein
3753966	Exon	AT1G11200	Unknown function
3816744	Exon	AT1G11340	S-locus protein kinase, involved in pollen recognition

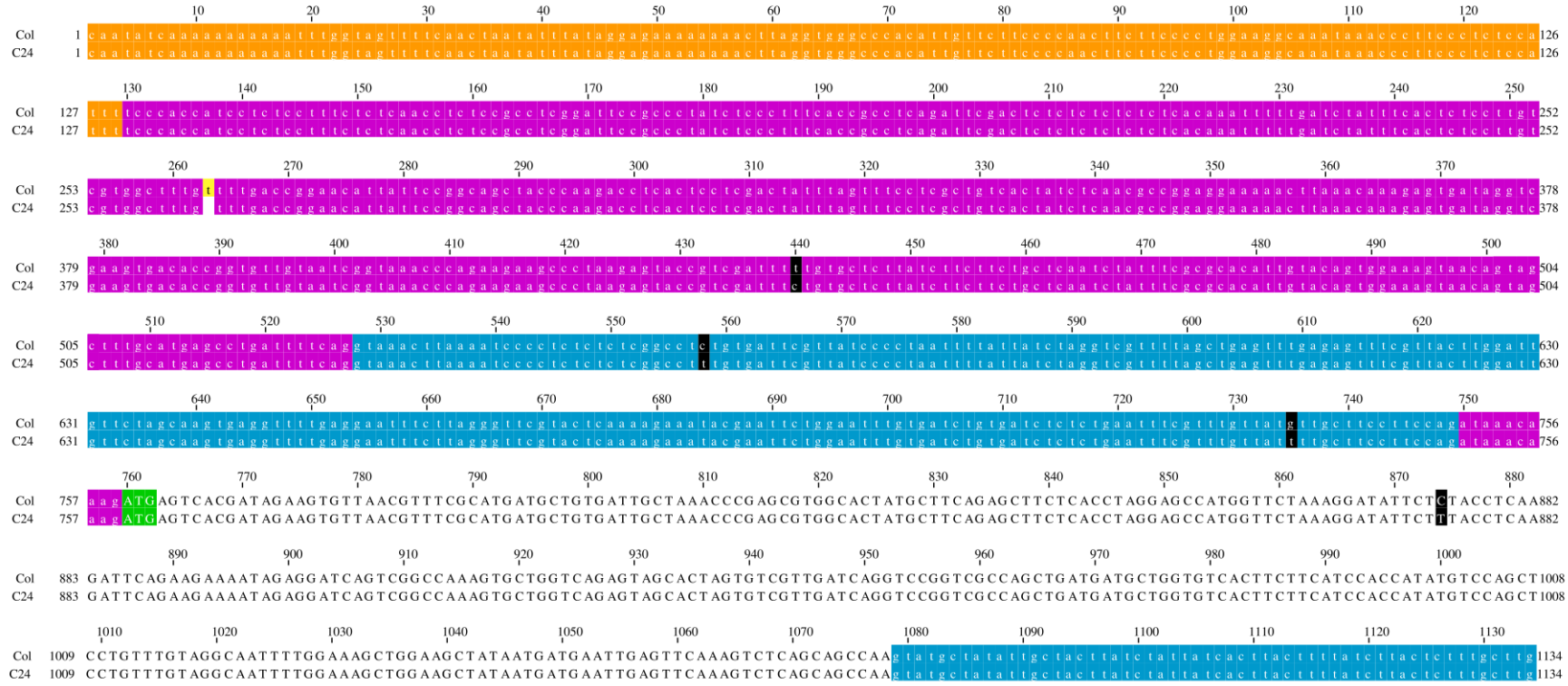
SNP position (bp)	Mutation location	TAIR ID	Function
4280595	Exon	AT1G12610	DREB family transciption factor involved in flowering and growth
4316338	Intergenic		Unknown function
4564003	Exon	AT1G13320	Part of regulatory subunit of protein phosphatase 2
4798406	Exon	AT1G14000	MAP Kinase involved in auxin signalling
4866043	Intron	AT1G14240	GDA1/CD39 nucleotide phosphatase
5218865	Exon	AT1G15165	Zinc finger protein associated with a ubiquitin ligase complex which ubiquitinates proteins
5274409	Exon	AT1G15330	Cystathionine beta-synthase protein
5326847	Exon	AT1G15500	ADP antiporter transport protein found in chloroplasts
5360698	Intergenic		A2-type cyclin involved in cell division control
5419895	Exon	AT1G15750	Transcription factor involved in repressing root development in aeriel tissues
5469631	5'UTR	AT1G15920	Polynucleotidyl transferace, related to ribonuclease H
5575831	Intron	AT1G16300	Encodes a chloroplast GAPDH protein
5613754	Exon	AT1G16420	Metacaspase which promotes cell death in response to UV light, hydrogen peroxide and methyl violegen stress
5623557	Intron	AT1G16470	Proteasome subunit
5677535	Exon	AT1G16610	Involved in RdDM and spliceosome
5701373	Intergenic		Protein chaperone
5701399	Intergenic		Protein chaperone
5701425	Intergenic		Protein chaperone
5781761	Intron	AT1G16900	Protein kinase in curculin-like lectin protein family
5892010	Exon	AT1G17230	Leucine rich protein kinase found in membrane
6324487	Intergenic		Kinesin involved in cytokenesis in pollen

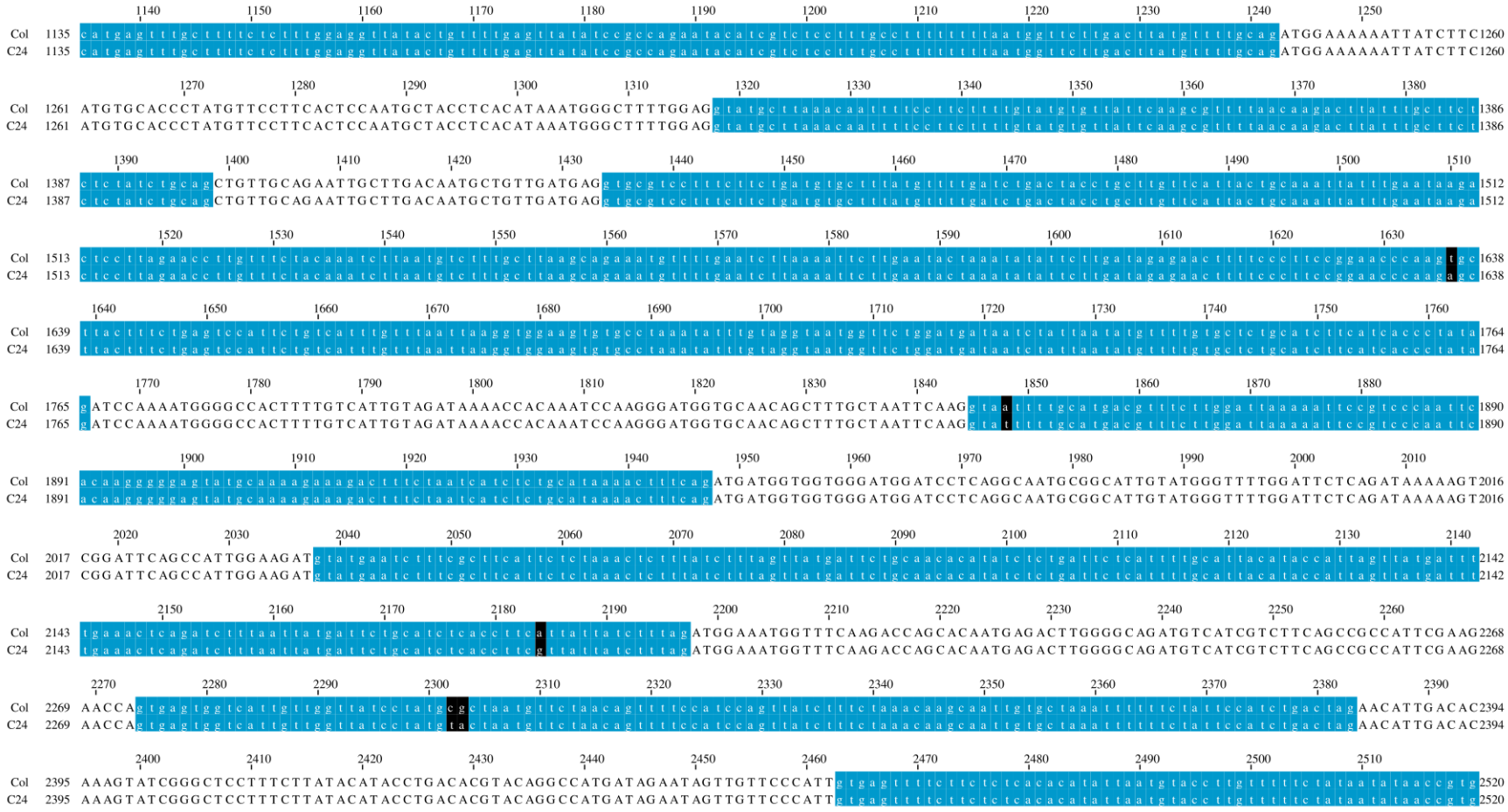
SNP position (bp)	Mutation location	TAIR ID	Function
6393196	Intergenic		Protein with galacturonosyltransferase activity.
6595680	Exon	AT1G19100	Chloroplast envelope protein similar to histidine kinases, DNA gyrases and heat shock proteins
6651377	Exon	AT1G19250	Response to viral infection, promotes cell death
6729457	Exon	AT1G19440	3-ketoacyl-CoA synthase involved in fatty acid biosytnhesis
6736478	Intergenic		Major facilitator protein involved in carbohydrate membrane transport
6952070	Intron	AT1G20060	ATP binding microtubule motor protein
7280563	Exon	AT1G20910	DNA binding proteins which contains ARID/BRIGHT domains
7445662	Exon	AT1G21270	Serine/threonine protein kinase involved in cell expansion
7963776	Intergenic		Oligopeptide transport protein
8090097	Exon	AT1G22870	Serine/threonine protein kinase
8107946	Exon	AT1G22910	RNA binding protein
8352560	3'UTR	AT1G23550	ADP ribosylation of proteins
8397784	Intron	AT1G23730	Carbonate dehydratase
8403007	Exon	AT1G23760	Aromatic rich glycoprotein
8450341	Intergenic		Polyketide cyclase/dehydrase and lipid transport protein
9304860	Intergenic		S-adenosyl-L-methionine-dependent methyltransferase
10047444	3'UTR	AT1G28590	GDSL-like Lipase/Acylhydrolase
10922907	Intergenic		F-box associated ubiquitination effector protein
11050895	Intergenic		Transposable element from mutator like transposon family
13085698	Exon	AT1G3552	Auxin response transcription factor
13477372	Intergenic		Unknown function

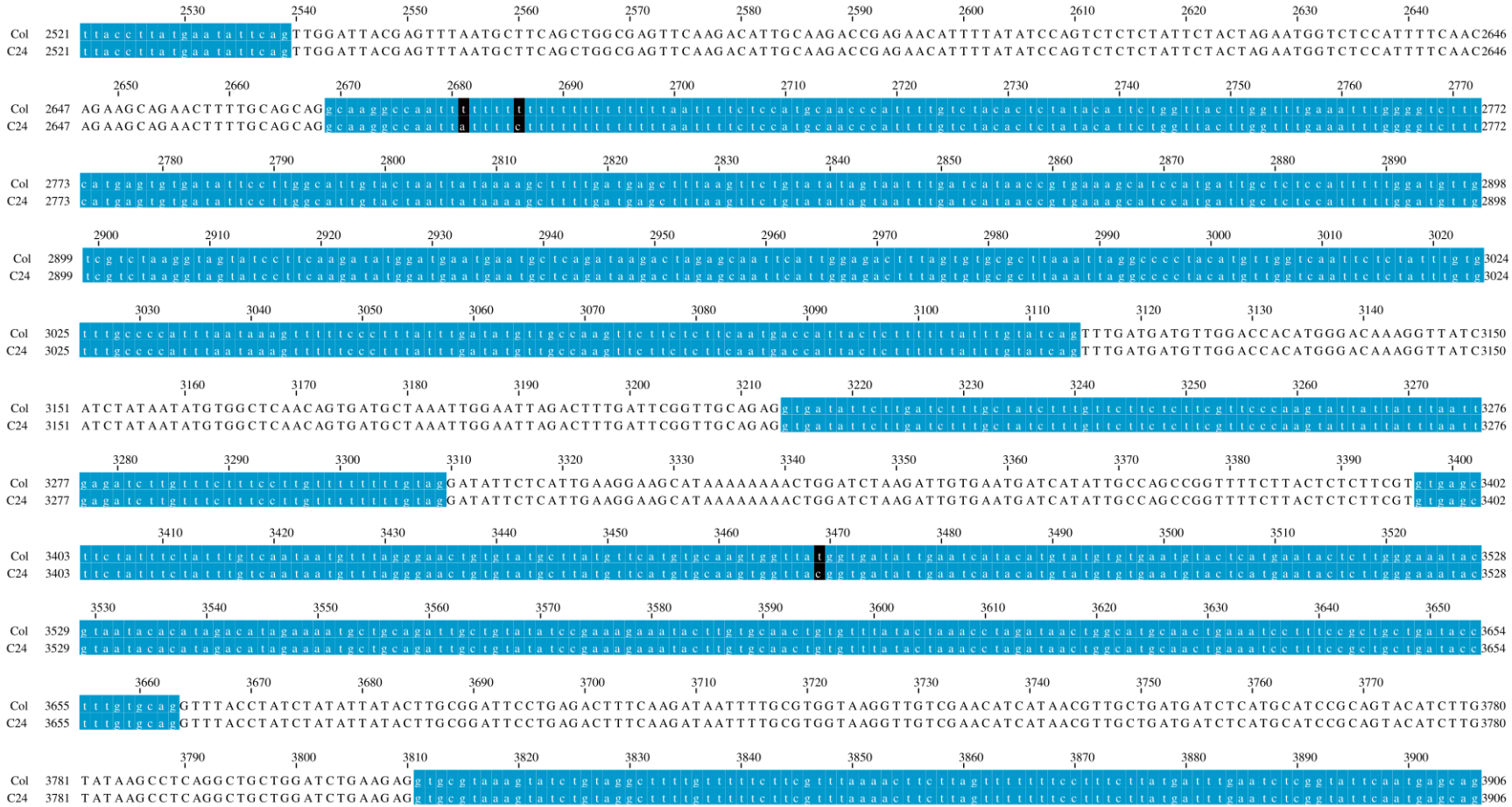
Table 7.3A: Functions of genes that the SNPs on the left arm of chromosome one are within or closest to

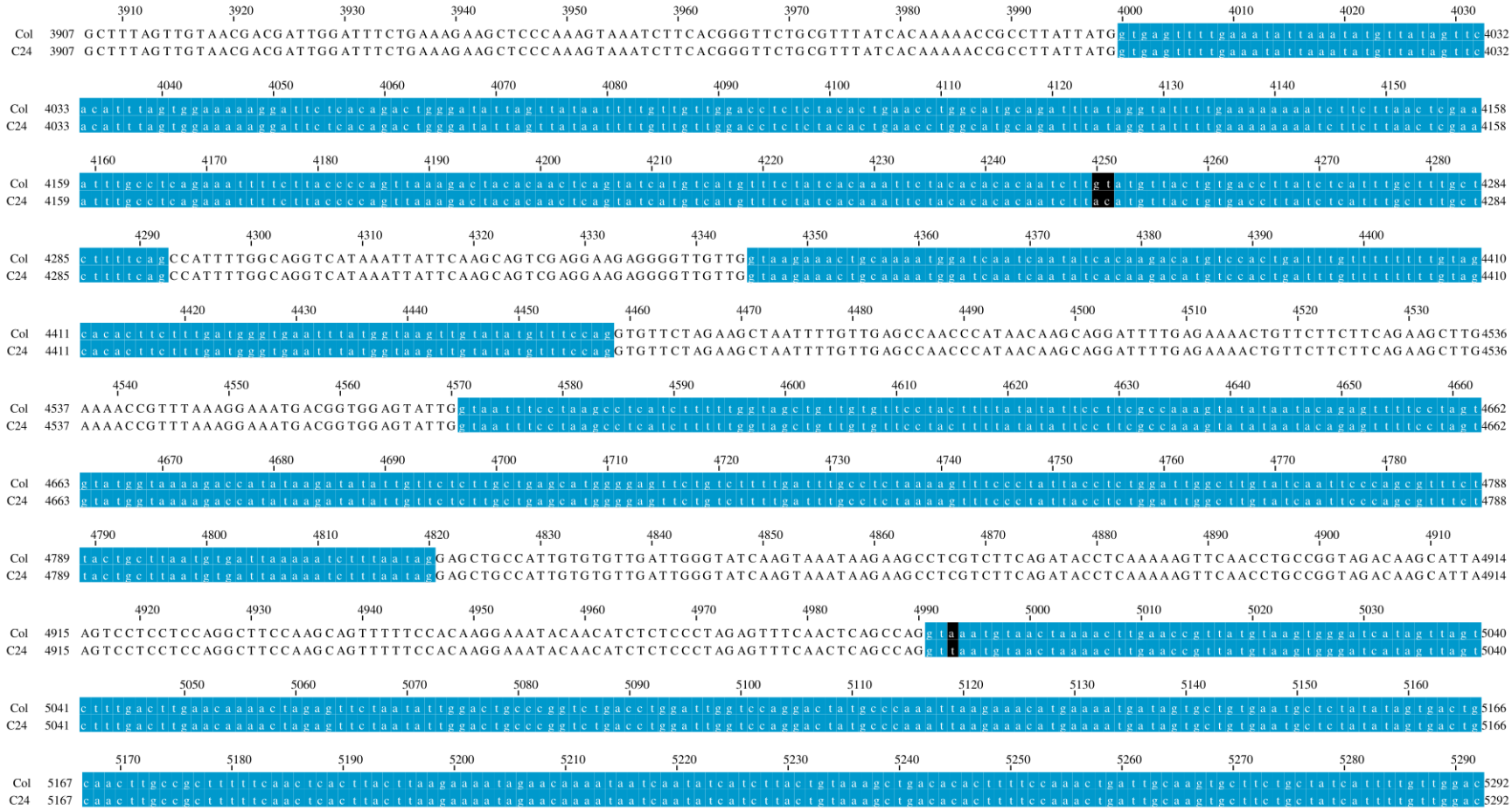
Table showing the position of each SNP of chromosome one in bp and whether the SNP is genic or intergenic. For cases where the SNP is within a gene it also gives the name of that gene. The function of the gene the SNP is within or closest to is given for all SNPs.

Appendix 1.5 At1G19100 (*MORC6*) sequence differences between WT Col and WT C24









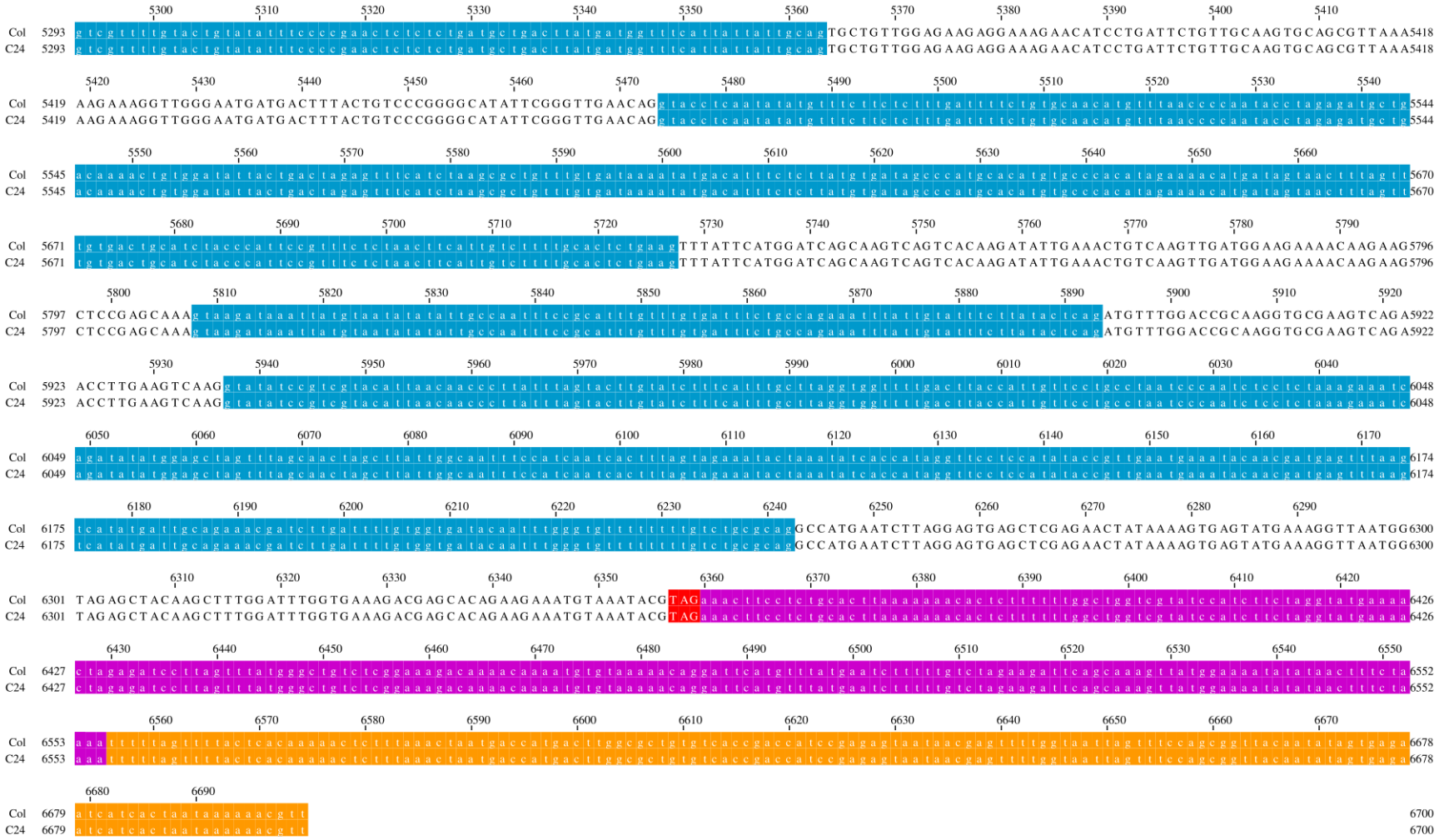

Col 5293 gtcgttttgtactgtatatttccccgaactctctctgatgctgacttatgatggttttcattattattgcagTGCTGTTGGAGAAGAGGAAAGAACATCCTGATTCTGTTGCAAGTGCAGCGTTAAA5418
C24 5293 gtcgttttgtactgtatatttccccgaactctctctgatgctgacttatgatggttttcattattattgcagTGCTGTTGGAGAAGAGGAAAGAACATCCTGATTCTGTTGCAAGTGCAGCGTTAAA5418
Col 5419 AAGAAAGGTTGGGAATGATGACTTTACTGTCCCGGGGCATATTCGGGTTGAACAGgtacctcaatatatgttttcttctctttgattttctgtgcaacatgtttaacccaatacctagagatgctg5544
C24 5419 AAGAAAGGTTGGGAATGATGACTTTACTGTCCCGGGGCATATTCGGGTTGAACAGgtacctcaatatatgttttcttctctttgattttctgtgcaacatgtttaacccaatacctagagatgctg5544
Col 5545 acaaaactgtggatattactgactagagtttcatctaagcgctgtttgtgataaaatatgacatttctcttatgtgatagcccatgcacatgtgcccacatagaaaacatgatagtaacttttagtt5670
C24 5545 acaaaactgtggatattactgactagagtttcatctaagcgctgtttgtgataaaatatgacatttctcttatgtgatagcccatgcacatgtgcccacatagaaaacatgatagtaacttttagtt5670
Col 5671 tgtgactgcatctacccattccgtttctctaacttcattgtctttttgcactctgaagTTTATTCATGGATCAGCAAGTCAGTCACAAGATATTGAAACTGTCAAGTTGATGGAAGAAAACAAGAAG5796
C24 5671 tgtgactgcatctacccattccgtttctctaacttcattgtctttttgcactctgaagTTTATTCATGGATCAGCAAGTCAGTCACAAGATATTGAAACTGTCAAGTTGATGGAAGAAAACAAGAAG5796
Col 5797 CTCCGAGCAAAgtaagataaattatgtaatatatattgccaatttccgcatttgtttgtgatttctgccagaaatttattgtatttcttatactcagATGTTTGGACCGCAAGGTGCGAAGTCAGA5922
C24 5797 CTCCGAGCAAAgtaagataaattatgtaatatatattgccaatttccgcatttgtttgtgatttctgccagaaatttattgtatttcttatactcagATGTTTGGACCGCAAGGTGCGAAGTCAGA5922
Col 5923 ACCTTGAAGTCAAGgtatatccgtcgtacattaacaaccctatttagtacttgtatctttcatttgcttaggtggttttgacttaccattgttcctgcctaatcccaatctcctctaaagaaatc6048
C24 5923 ACCTTGAAGTCAAGgtatatccgtcgtacattaacaaccctatttagtacttgtatctttcatttgcttaggtggttttgacttaccattgttcctgcctaatcccaatctcctctaaagaaatc6048
Col 6049 agatatatggagctagtttagcaactagcttattggcaaatttccatcaatcactttagtagaaatactaaatatcaccataggttcctccatataccgttgaatgaaatacaacgatgagtttaag6174
C24 6049 agatatatggagctagtttagcaactagcttattggcaaatttccatcaatcactttagtagaaatactaaatatcaccataggttcctccatataccgttgaatgaaatacaacgatgagtttaag6174
Col 6175 tcatatgattgcagaaacgatcttgatttttgtggtgatacaatttgggtgttttttttttgtctgcgcagGCCATGAATCTTAGGAGTGAGCTCGAGAACTATAAAAGTGAGTATGAAAGGTTAATGG6300
C24 6175 tcatatgattgcagaaacgatcttgatttttgtggtgatacaatttgggtgttttttttttgtctgcgcagGCCATGAATCTTAGGAGTGAGCTCGAGAACTATAAAAGTGAGTATGAAAGGTTAATGG6300
Col 6301 TAGAGCTACAAGCTTTGGATTTGGTGAAAGACGAGCACAGAAGAAATGTAAATACGTTAGaaaacttcctctgcacttaaaaaaaacactctttttttggctggtctgtatccatcttctaggtatgaaaat6426
C24 6301 TAGAGCTACAAGCTTTGGATTTGGTGAAAGACGAGCACAGAAGAAATGTAAATACGTTAGaaacttcctctgcacttaaaaaaaacactctttttttggctggtctgtatccatcttctaggtatgaaaat6426
Col 6427 ctagaggatccttagtttatggggctgtctcggaaagacaaaacaaaatgtgtaaaaacaggatttcatgtttatgaatctttttgtctagaagattcagcaaagttatggaaaatatataactttcta6552
C24 6427 ctagaggatccttagtttatggggctgtctcggaaagacaaaacaaaatgtgtaaaaacaggatttcatgtttatgaatctttttgtctagaagattcagcaaagttatggaaaatatataactttcta6552
Col 6553 aaattttttagtttttactcacaaaaaactctttaaactaatgaccatgacttggcgctgtgtcaccgaccatccgagagtaataacgagttttggtaatttagtttccagcggttacaatatagtgaga6678
C24 6553 aaattttttagtttttactcacaaaaaactctttaaactaatgaccatgacttggcgctgtgtcaccgaccatccgagagtaataacgagttttggtaatttagtttccagcggttacaatatagtgaga6678
Col 6679 atcatcactaataaaaaacgtt 6700
C24 6679 atcatcactaataaaaaacgtt 6700

Figure 7.4A: SNP differences between WT Col and WT C24 for the At1G19100 (*MORC6*) gene

DNA sequence alignment of WT Col and WT C24 for the At1G19100 (*MORC6*) gene. Intergenic DNA is highlighted in gold, the UTRs are highlighted in purple and the introns are highlighted in blue. The areas of black uppercase text are exons with the start codon highlighted in green and the stop codon highlighted in red. Base substitutions have been highlighted in black and the 5' UTR single base deletion highlighted in yellow. There are fifteen base substitutions of which most are intronic (thirteen), one is in the 5'UTR and one is in the 2nd exon, but does not cause an amino acid change.

Appendix 1.6 *MORC* gene family and genome location

Name	Locus	Chromsome number	Chromsome arm	Gene size (Kb)	Number of exons	Protein size (amino acids)
MORC1	At4G36290	4	Right	3.75	18	635
MORC2	At4G36280	4	Right	4.05	18	626
MORC3	At4G36270	4	Right	4.22	18	486
MORC4	At5G50780	5	Right	4.94	20	819
MORC5	At5G13130	5	Left	3.9	18	706
MORC6	At1G19100	1	Left	6.43	20	663
MORC7	At4G24970	4	Right	4.44	19	707

Table 7.4A: *MORC* gene family in Arabidopsis

Table showing all *MORC* genes found in *Arabidopsis* and their location within the genome. The size of each gene, exon number and protein size is also given.

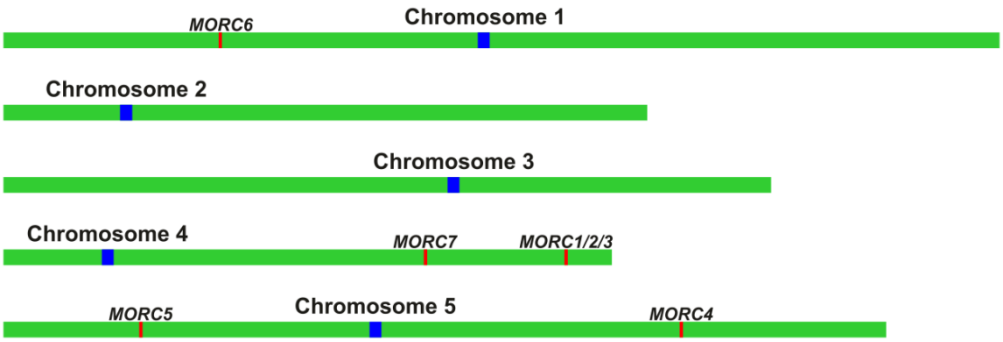


Figure 7.5A: Location of *MORC* genes in the *Arabidopsis* genome

Diagram showing the location of *MORC* genes within the *Arabidopsis* genome. Each chromosome's centromere is shown in blue and the *MORC* gene location is shown in red.

Appendix 1.7 Gene expression of *Arabidopsis MORC* genes in C24 and Columbia ecotypes

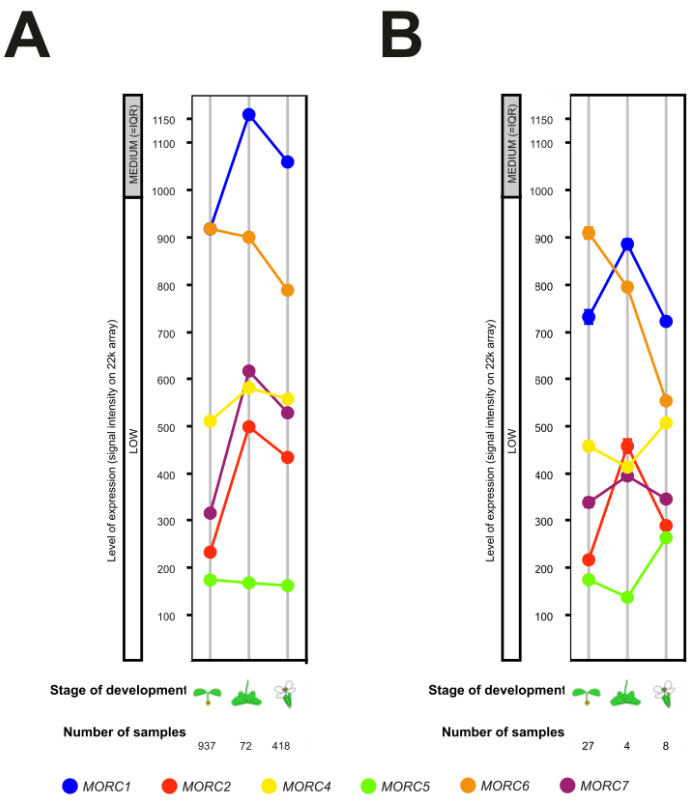
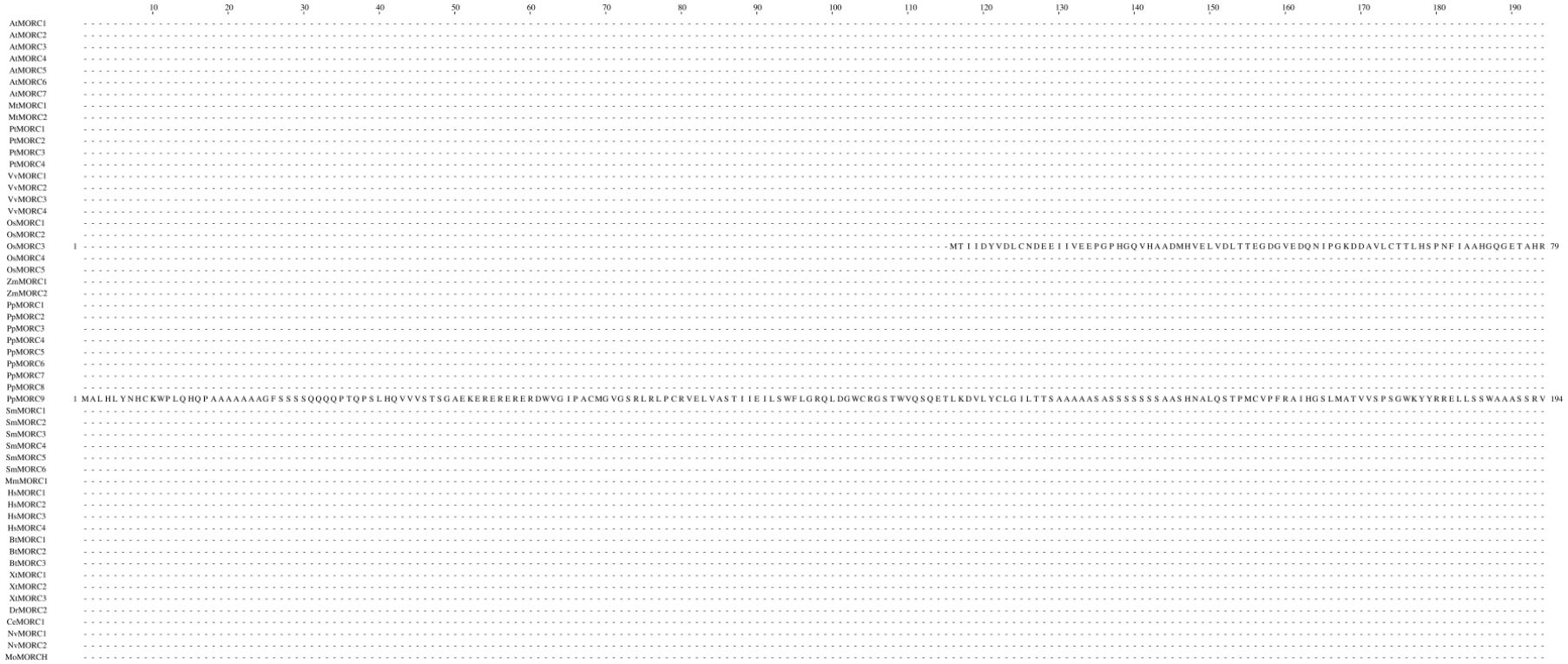
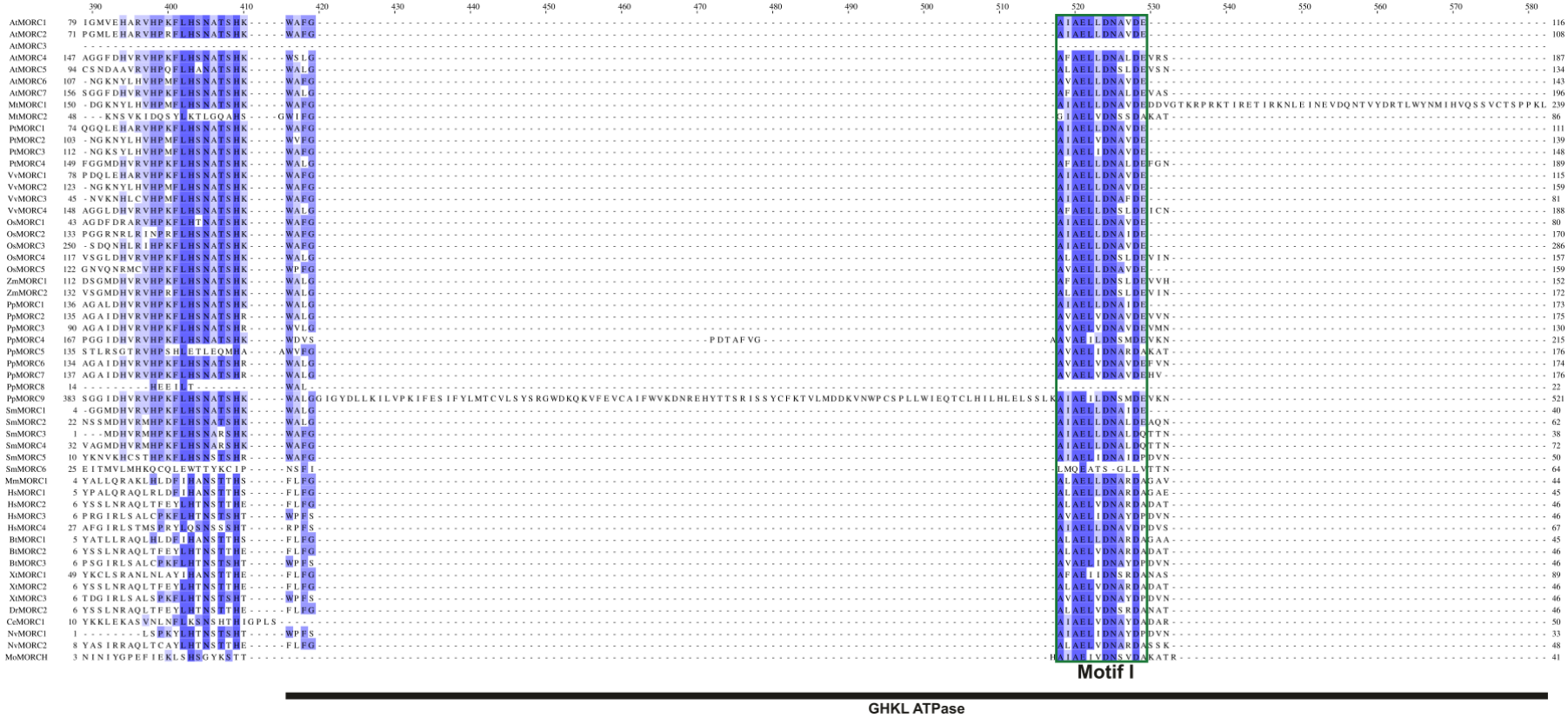


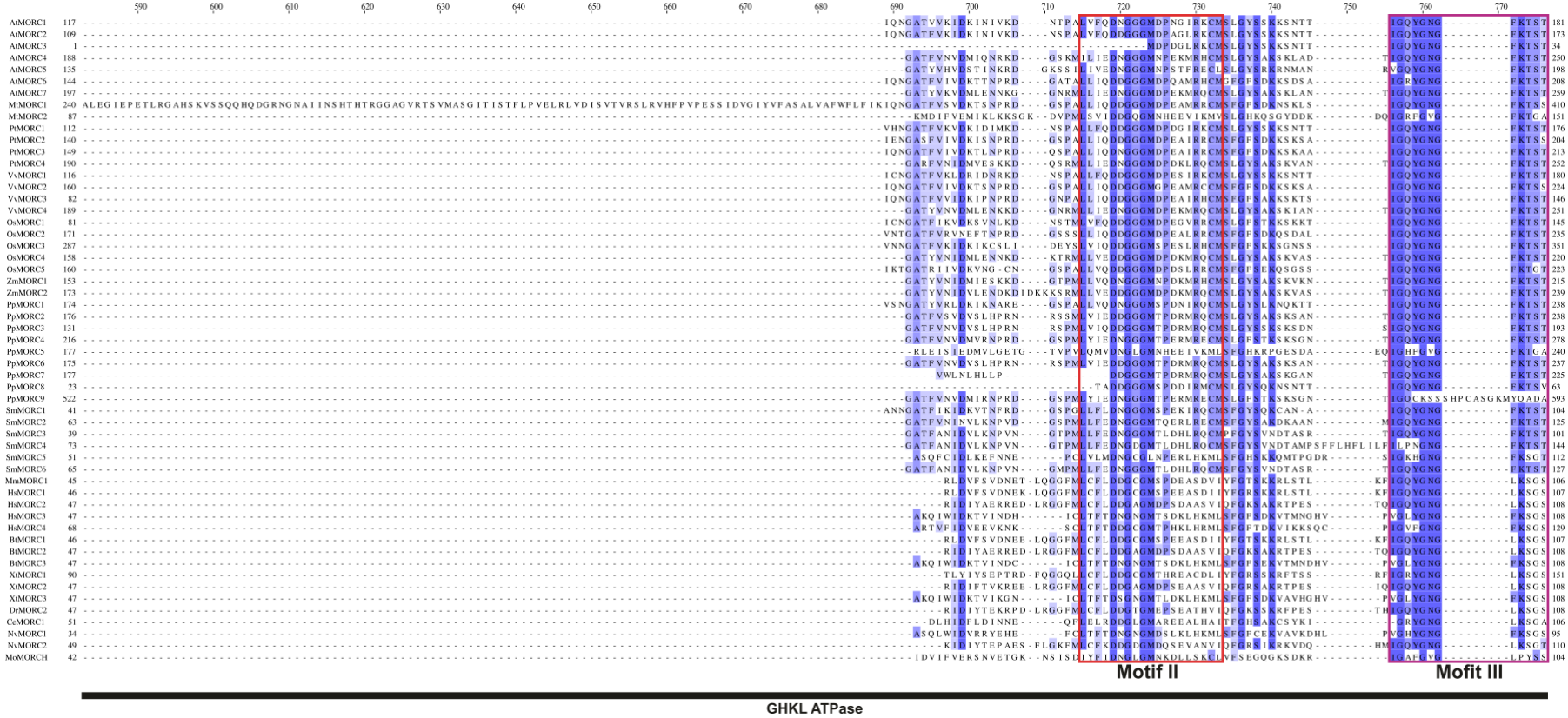
Figure 7.6A: Developmental expression patterns of *MORC*s in Columbia and C24 ecotypes

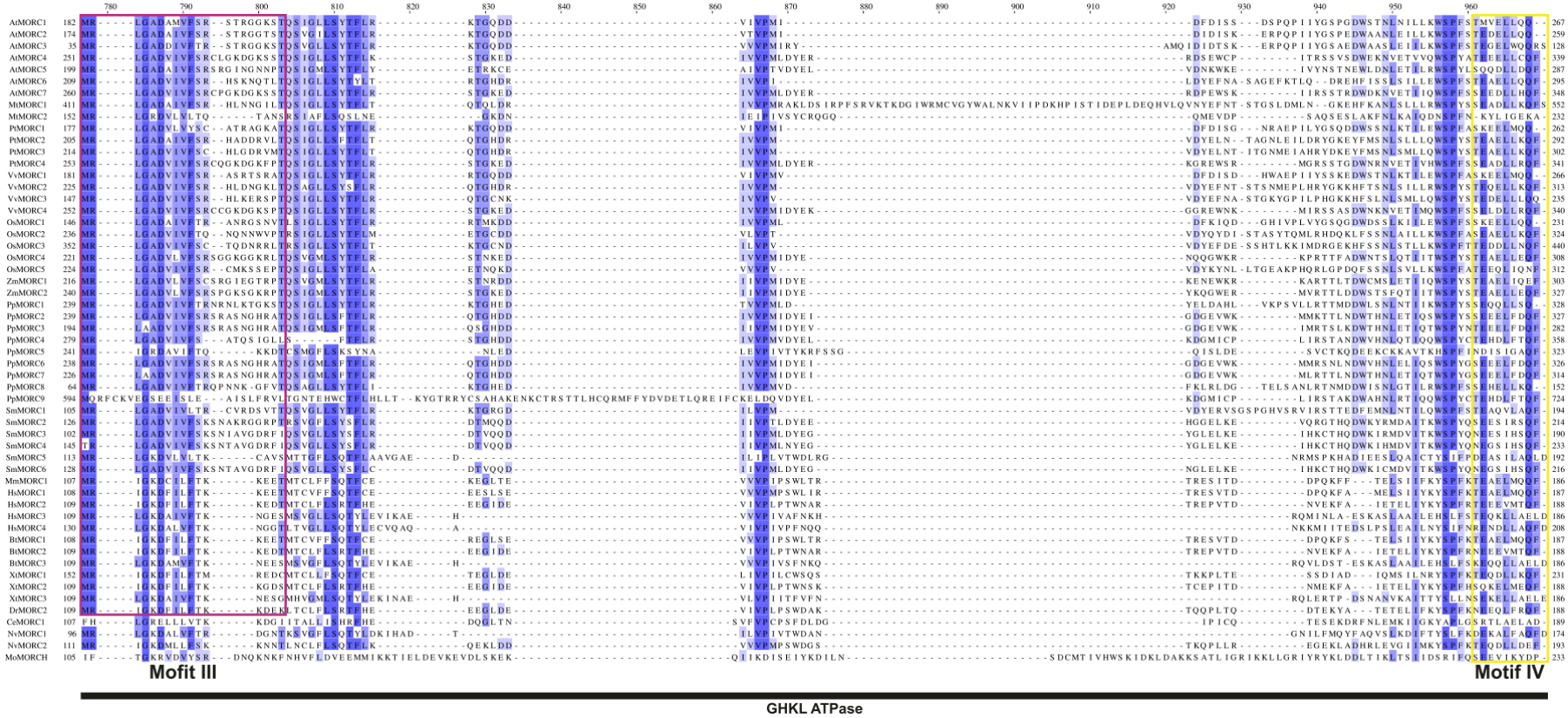
Graphs showing mean expression levels of expression patterns of *MORC* genes for different developmental stages of Columbia (**A**) and C24 (**B**) taken from microarray data from Genevestigator. There is a key at the bottom of the figure showing which *MORC* corresponds to each coloured circle. Each circle on the graph represents the mean relative signal intensity. Low levels of gene expression are considered to be genes with a relative intensity among the 25% lowest intensities of all genes and medium levels of gene expression are considered to be those within the interquartile range (IQR) of relative intensity. The different developmental stages are represented as pictures and are from left to right, are: seedling, bolting plant and developed flower. The number of microarrays with probes for *MORC* genes at that developmental stage are shown below each developmental stage.

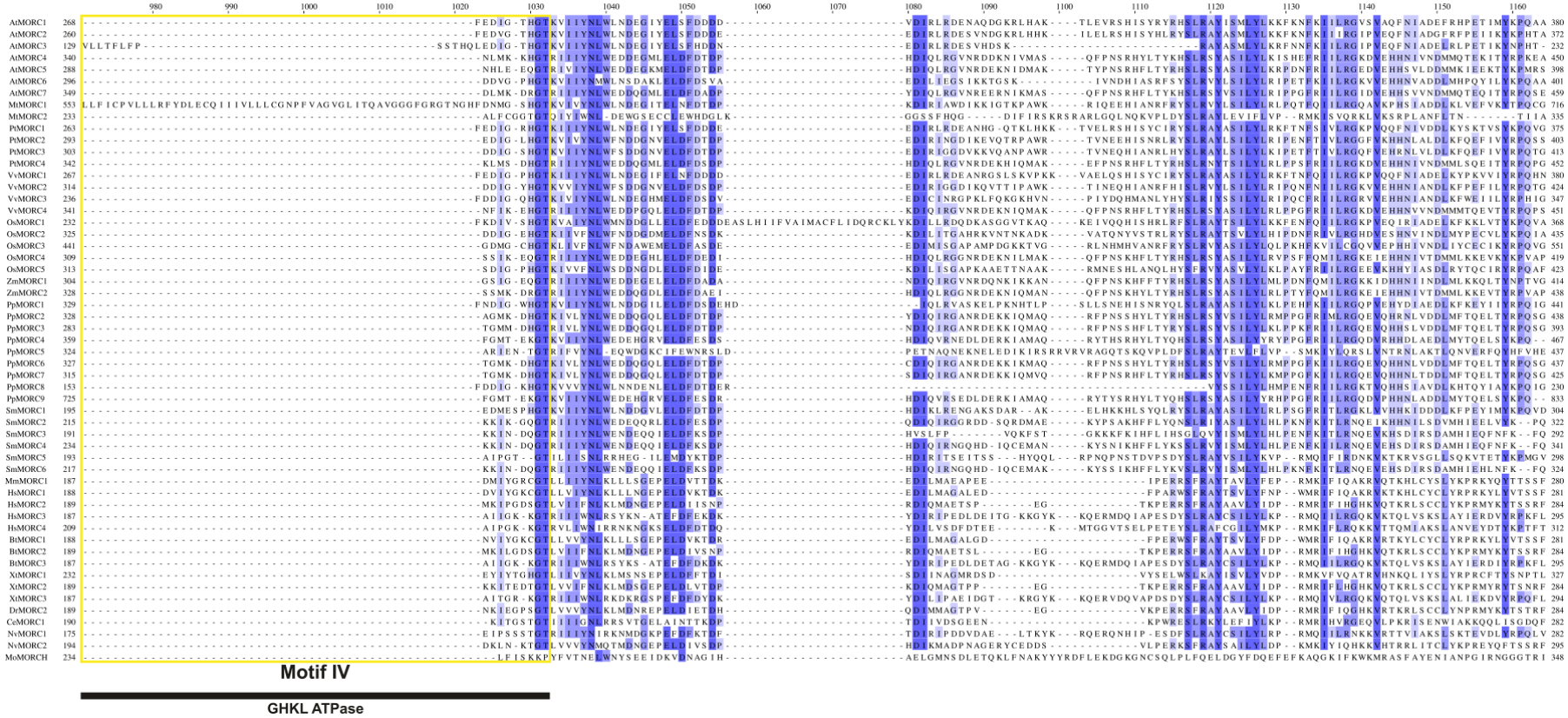
Appendix 1.8 Alignment of MORC proteins from *Arabidopsis* other Plantae and Metazoa

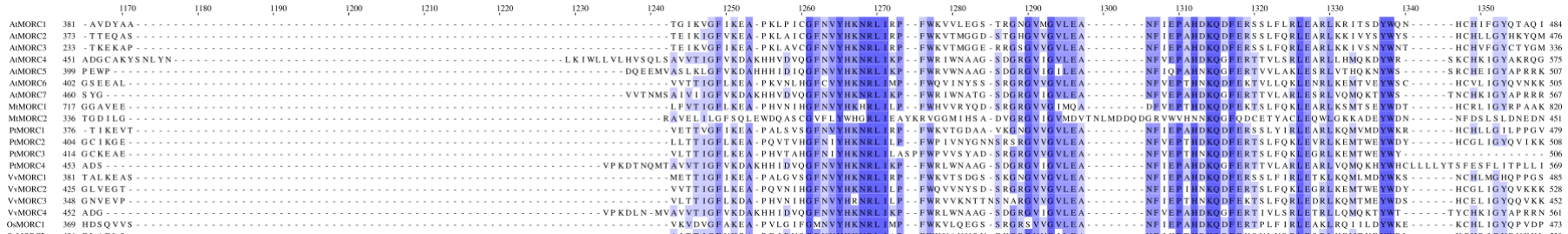


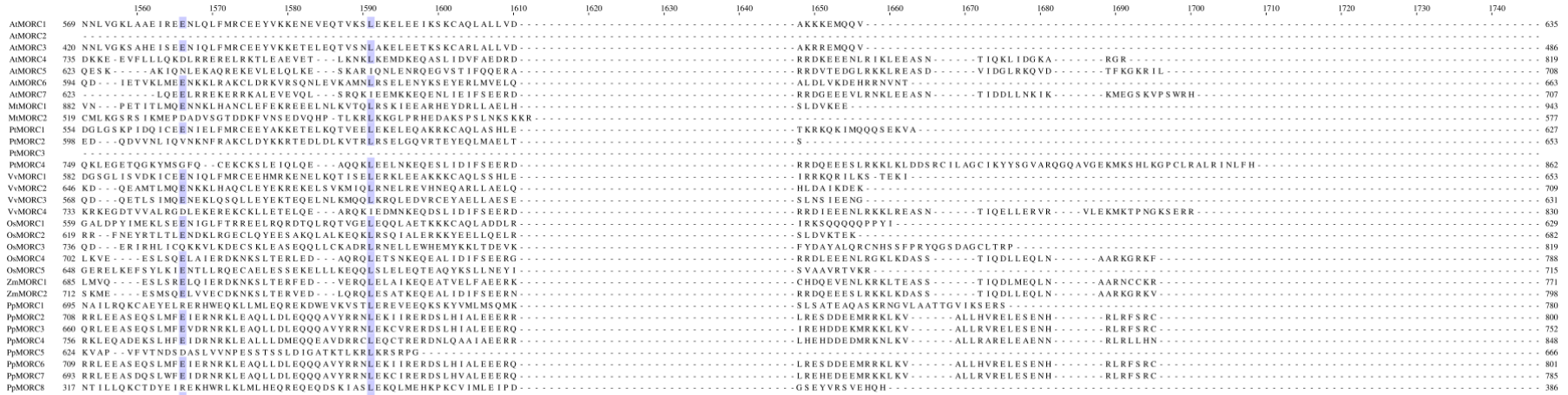

Motif I
GHKL ATPase


Motif II
Mofit III
GHKL ATPase


Mofit III
Motif IV
GHKL ATPase


Motif IV
GHKL ATPase




1560
1570
1580
1590
1600
1610
1620
1630
1640
1650
1660
1670
1680
1690
1700
1710
1720
1730
1740
AtMORC1
AtMORC2
AtMORC3
AtMORC4
AtMORC5
AtMORC6
AtMORC7
MtMORC1
MtMORC2
PtMORC1
PtMORC2
PtMORC3
PtMORC4
VvMORC1
VvMORC2
VvMORC3
VvMORC4
OsMORC1
OsMORC2
OsMORC3
OsMORC4
OsMORC5
ZmMORC1
ZmMORC2
PpMORC1
PpMORC2
PpMORC3
PpMORC4
PpMORC5
PpMORC6
PpMORC7
PpMORC8
PpMORC9
SmMORC1
SmMORC2
SmMORC3
SmMORC4
SmMORC5
SmMORC6
MmMORC1
HsMORC1
HsMORC2
HsMORC3
HsMORC4
BtMORC1
BtMORC2
BtMORC3
XtMORC1
XtMORC2
XtMORC3
DrMORC2
CeMORC1
NvMORC1
NvMORC2
MoMORCH


1750 1760 1770 1780 1790 1800 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930
AtMORC1
AtMORC2
AtMORC3
AtMORC4
AtMORC5
AtMORC6
AtMORC7
MtMORC1
MtMORC2
PtMORC1
PtMORC2
PtMORC3
PtMORC4
VvMORC1
VvMORC2
VvMORC3
VvMORC4
OsMORC1
OsMORC2
OsMORC3
OsMORC4
OsMORC5
ZmMORC1
ZmMORC2
PpMORC1
PpMORC2
PpMORC3
PpMORC4
PpMORC5
PpMORC6
PpMORC7
PpMORC8
PpMORC9
SmMORC1
SmMORC2
SmMORC3
SmMORC4
SmMORC5
SmMORC6
MmMORC1 773 VSGDCQLPASP-MPSQSMSVEETARKLLSNLREILLYFVPEFQLSSEFECTSVEELITNPELER-----------CPENINEKLKTCFNQIQNIYMAQYEK-RLKRKMQSIVYEANRRGLLNQVFLGQCELKRKRTEEKLSDLRAKLALLLQKLQLGGPAGDPQQIDAYLEDLLKED-------- 936
HsMORC1 796 VSGSCKVASSP-ASSQSTPVKETVRKLKSKLREILLYFFPEHQLPSEL-----EEPALSCELEQ-----------CPEQMNKKLKMCFNQIQNTYMVQYEK-KIKRKLQSIIYDSNTRGIHNEISLGQCENKRKISEDKLKNLRIKLALLLQKLQLGGPEGDLEQTDTYLEALLKEDNLLFQNNL 962
HsMORC2 805 REWYTGRVTAVEVGKHVVRWKVKFDYVPTDTTPRDRWVEKGSEDVRLMKPPSPEHQSLDTQQEGGEEEVGP-------VAQQAIAVAEPSTSECLRIEPDTTALSTNHETIDLLVQILRNCLRYFLPPSFPISKKQLSAMNSDELISFPLKEYFKQYEVGLQNLCNSYQSRADSRAKASEESLRTSERKLRETEE 992
HsMORC3 699 L-RNQLLLVTEEKENYKRQCHMFTDQIKVLQQRILEMNDKYVKKETCHQSTETDAVFLLESINGKSESPDHMVSQYQQALEEIERLKKQCSALQHVKAECSQCSNN--ESKSEMDEMAVQLDDVFRQLDKCSIERDQYKSEVELLEMEKSQIRSQCEELKTEVEQLKSTNQQTATDVSTSSNIEESVNHMDGES 889
HsMORC4 711 IPFNREELAERRKAVESWNPVPYSVASAAIPAAAIGEKARGYEESEGHNTPKLKNQRELEELKRTTEKLERVLAERNLFQQKVEELEQERNHWQSEFKKVQHELVIYSTQEAEG-------LYWSKKHMGYRQAEFQILKAELERTKEEKQELKEKLKETETHLEMLQKAQVSYRTP----------EGDDLERAL 889
BtMORC1 790 VSGDCKVASALTESSQSTSVKETVRKLTSNLREILLYFFPEYQLPSEFGQASLEELTACSELEQ-----------CPEQTNKKLKMCFNQICDVYMVQYEK-KLKRRIQSVICDANKRAVVNEVSLGQCEKKRKLTEDKLNKLRVKLAVLLQKLQLGGPAGDLEQIDAYLEALLKEDNLLFQN-- 960
BtMORC2 805 REWYTGRVTAVEVGKNAMRWKVKFDYVPTDTTPRDRWVEKGSDDVRLMKPPSPEYESPDTRQGREEVAVAAAPAGDLVAQQAIATVEPSTSDCIPIEPDTTAPSTNHETIDLLVQILRNCLRYFLPPSFPISKKELSAMNSDELISFPLKEYFKQYEVGLQNLCHSYQSRADSRAKASEESLRTSERKLRETEE 998
BtMORC3 700 L-RNELLLVTQKKKK----------713
XtMORC1 749 IAEAGRTLEVP-----SCSRKETEEKLKLTLGEVISYFVPDCQLP----------ILETIYKGDQAQLGSLVNDYFKEYEENVLAVYLHAANDHFEAKAELNELMGEACEEQIKVVEQKISALSKKAARIFKKLHINVPE----------ESVLDTE-------- 880
XtMORC2 804 KEWFTGRVTAVEVGKQGVRWKVKFDYVPTDTTPRDRWVEKGCEDVRLMRPPSPEHQAPDTQQ----------AEPSTSDNTRIEPDTTEPNASPAPLEQLVQTLRNCLRYFWPPNFTISKNELSTMSAEDLIGFPLSAGDPHWQVSGPIY---------- 943
XtMORC3 666 Q-EKSVGARDREKENLKKHCEELLVRVKGLEGEVSELRNNQVKKEVSHQVTQTDAS----PPNITELPSKQLVAQYEQALKEINRLKAQCEVLQDLKAEPGTGPQKPAENSSDMDDMALQLDSLFRQLDACSTERDRYKSEAEKVKQQSAQLESQTRQLGTQIEQLKHMVQDVQKKESASTSVASASQTN 850
DrMORC2 825 GEWFTGRVTAVETGKDNVRWKVKFDYVPMN-TPRDRWVFKGSDDVRLMRPASPESRSPDTNQG----------PARSVTPPTAAEPDTTQSGPSRETSEGLVTMMRTLLRYFFPPDFRIPKDSVNTMSAEDLVAFPIKEYFQQYEAGLQSLCNSYQTRAESRARAVEEKSSGAETKLRESEE 995
CeMORC1 795 IVDAQQAIMRQIGSVLDHLKRNPTHNFKIPASGTVEEKLKNIEHQIKGKKK----------845
NvMORC1
NvMORC2
MoMORCH

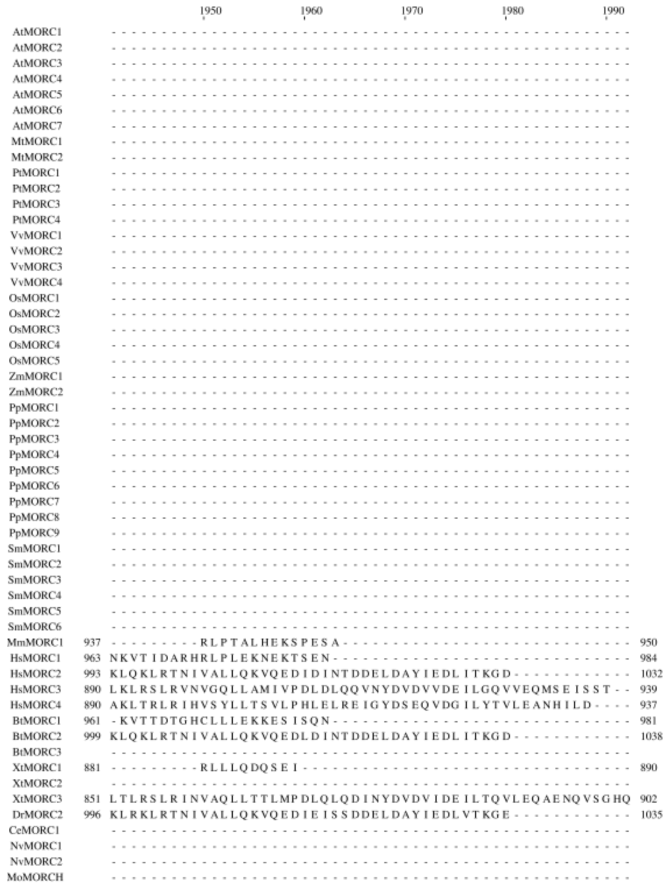


Figure 7.7A: Alignment of Plantae and Metazoa MORC proteins

Alignment of protein sequences from Plantae and Metazoa species for MORC proteins. There is also a bacterial MORC homolog from *Myroides odoratimimus* (Mo). The Plantae species are: *Arabidopsis thaliana* (At), *Medicago truncatula* (Mt), *Populus trichocarpa* (Pt), *Vitis vinifera* (Vv), *Oryza sativa* (Os), *Zea mays* (Zm), *Physcomitrella patens* (Pp) and *Selaginella moellendorffii* (Sm); and the Metazoa species are: *Mus musculus* (Mm), *Homo Sapiens* (Hs), *Bos taurus* (Bt), *Xenopus tropicalis* (Xt), *Danio rerio* (Dr), *Caenorhabditis elegans* (Ce) and *Nematostella vectensis* (Nv). For most species there are MORC paralogs in each species, bar *M. Musculus, D. rerio* and *C. elegans* which have only one MORC. Amino acid residues that are conserved are highlighted in blue, with the darkness of the blue indicating how conserved each residue is, dark blue are the most highly conserved while light blue are only conserved in a few species; amino acid residues that are not highlighted are not conserved between species. The amino acid sequence which codes for the GHKL ATPase domain is represented by a labelled black line underneath the relevant sections of the alignment. The GHKL domain begins at the start of the region identified by Pfam and SMART as the GHKL ATPase but ends beyond the region identified by these systems and instead ends after the fourth conserved motif of the GHKL ATPase domain. The four conserved motifs of GHKL ATPases are highlighted with boxes on the alignment: motif I is a green box, motif II is a red box, motif III is a purple box; and motif IV is a yellow box.

Appendix 2

Appendix 2.1 Confocal images of the M9 3rd leaf at 7 hpd and 13 hpd 13 dpg

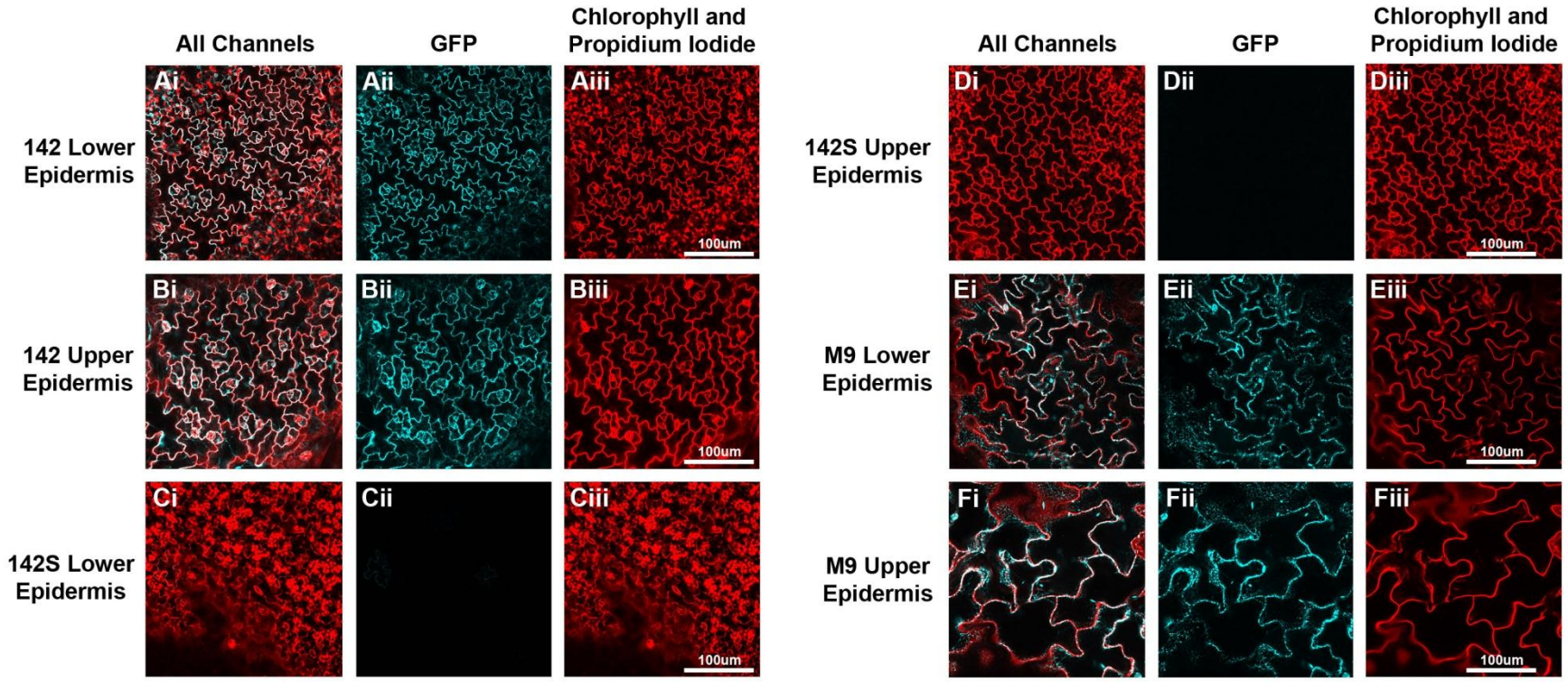


Figure 7.8A: GFP expression in epidermal layers of 3rd leaves of 13 dpg plants at 7 hpd

The 3rd leaves of 13 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 7hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were image: 142 (**A** and **B**), 142S (**C** and **D**) and M9 (**E** and **F**). A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

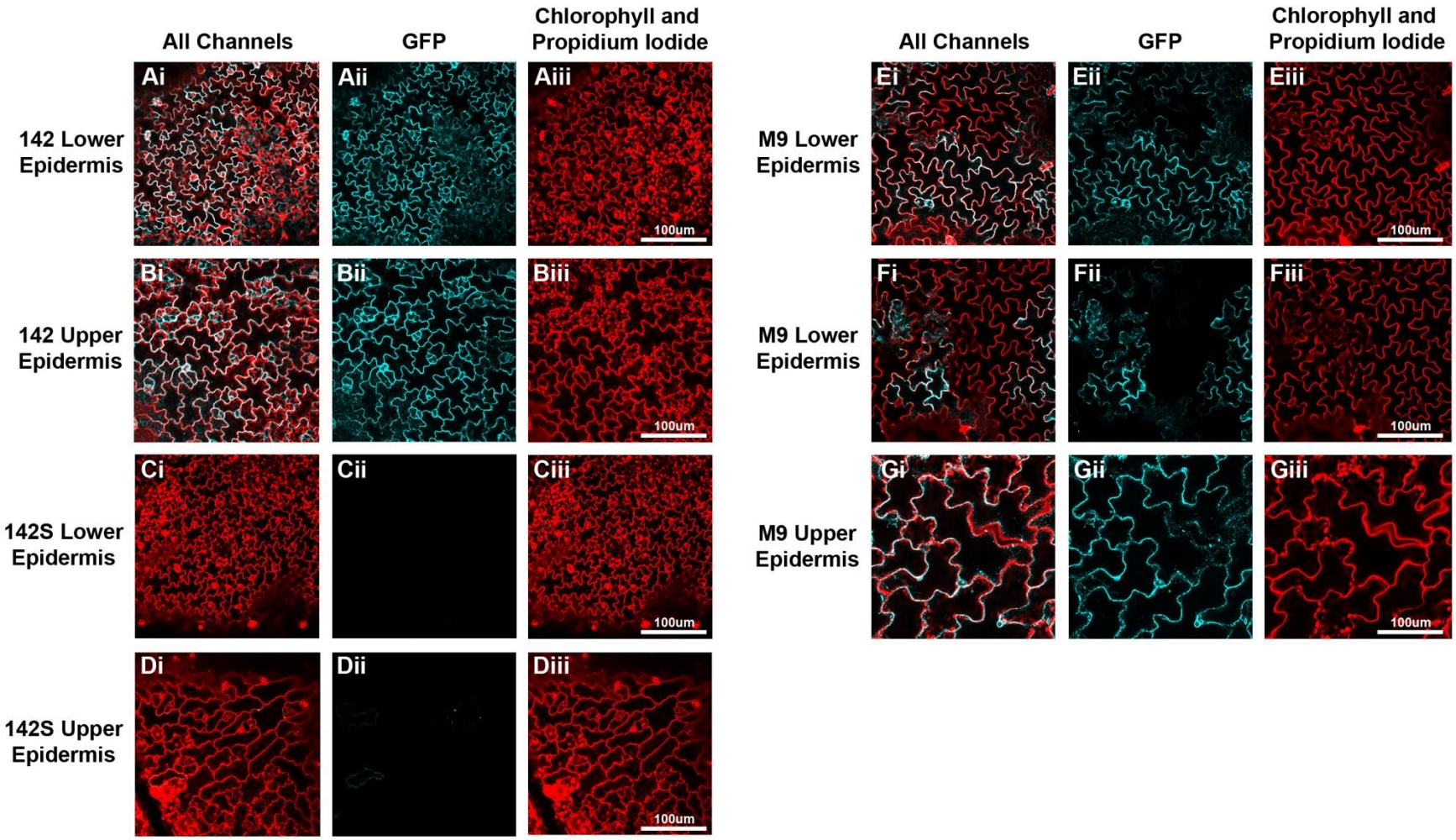

All Channels
GFP
Chlorophyll and Propidium Iodide
142 Lower Epidermis
142 Upper Epidermis
142S Lower Epidermis
142S Upper Epidermis
M9 Lower Epidermis
M9 Lower Epidermis
M9 Upper Epidermis
Ai
Aii
Aiii
Bi
Bii
Biii
Ci
Cii
Ciii
Di
Dii
Diii
Ei
Eii
Eiii
Fi
Fii
Fiii
Gi
Gii
Giii
100um

Figure 7.9A: GFP expression in epidermal layers of 3rd leaves of 13 dpg plants at 13 hpd

The 3rd leaves of 13 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 13 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were image: 142 (**A** and **B**), 142S (**C** and **D**) and M9 (**E**, **F** and **G**). **E** and **F** show areas of the leaf where silencing is less (**E**) or more (**F**) prevalent. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

Appendix 2.2 Confocal images of the M1 3rd and 5th leaves at 7 hpd and 13 hpd 14 dpg

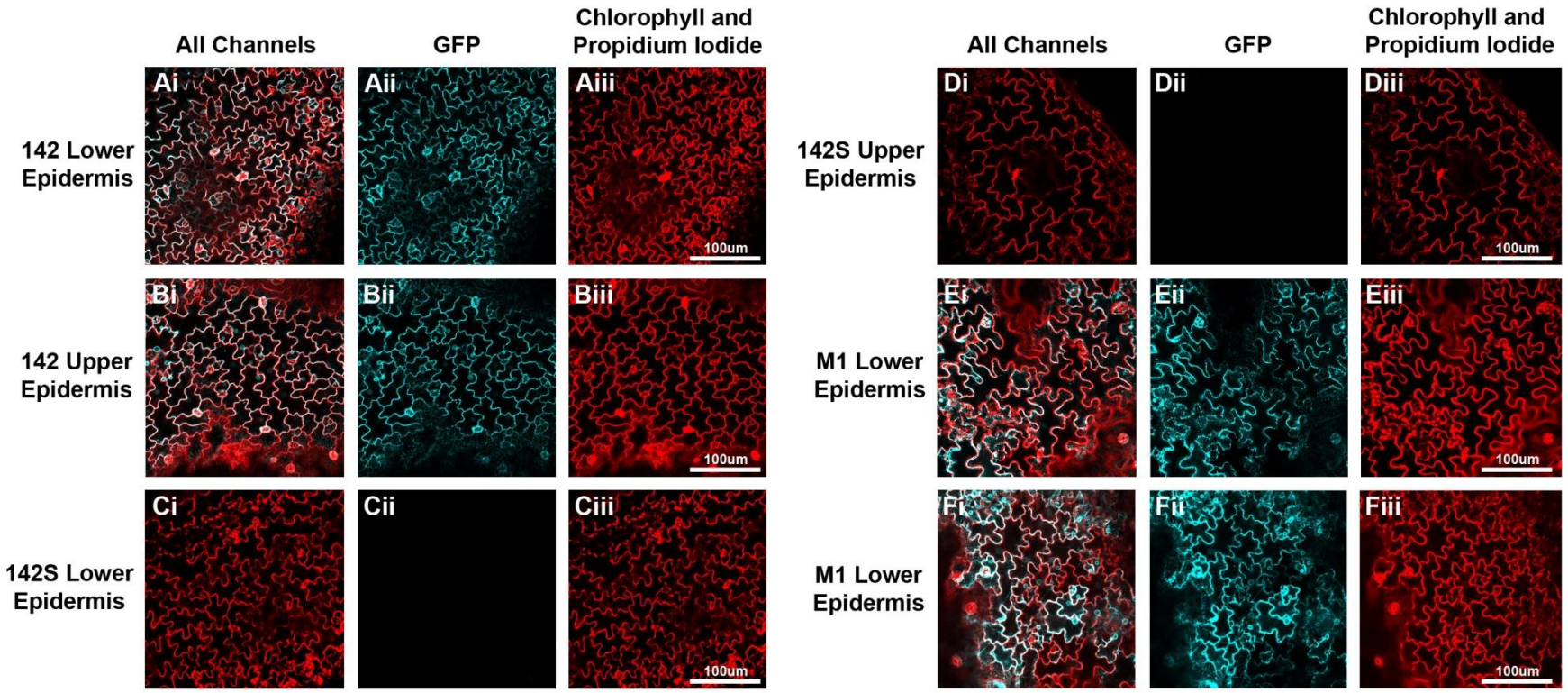


Figure 7.10A: GFP expression in epidermal layers of 3rd leaves of 14 dpg plants at 7 hpd

The 3rd leaves of 14 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 7 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were image: 142 (**A** and **B**), 142S (**C** and **D**) and M1(**E** and **F**). A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

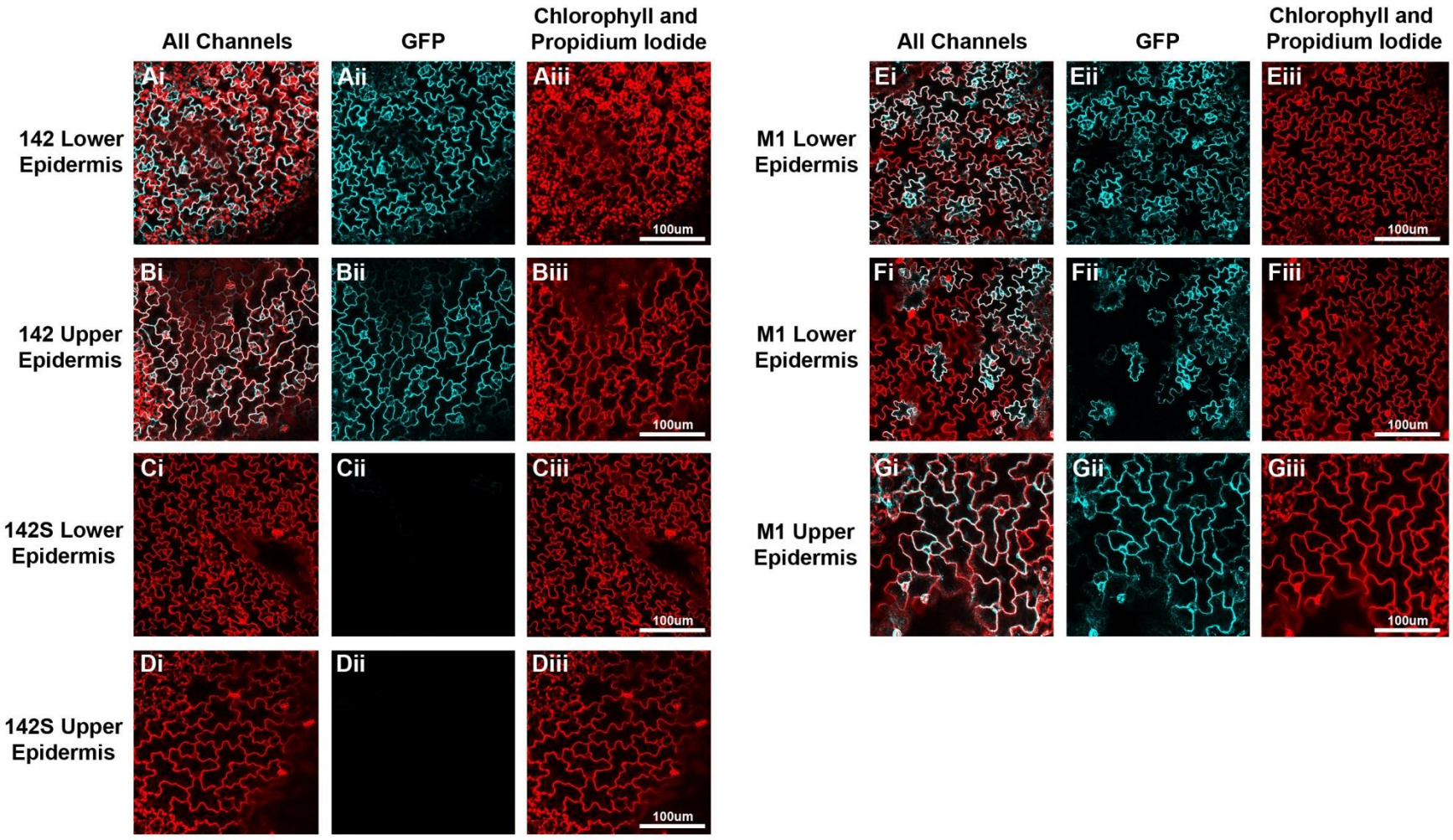

All Channels
GFP
Chlorophyll and Propidium Iodide
142 Lower Epidermis
142 Upper Epidermis
142S Lower Epidermis
142S Upper Epidermis
M1 Lower Epidermis
M1 Lower Epidermis
M1 Upper Epidermis
Ai
Aii
Aiii
Bi
Bii
Biii
Ci
Cii
Ciii
Di
Dii
Diii
Ei
Eii
Eiii
Fi
Fii
Fiii
Gi
Gii
Giii
100um

Figure 7.11A: GFP expression in epidermal layers of 3rd leaves of 14 dpg plants at 13 hpd

The 3rd leaves of 14 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 13 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were image: 142 (**A** and **B**), 142S (**C** and **D**) and M1 (**E**, **F** and **G**). **E** and **F** show areas of the leaf where silencing is less (**E**) or more (**F**) prevalent. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

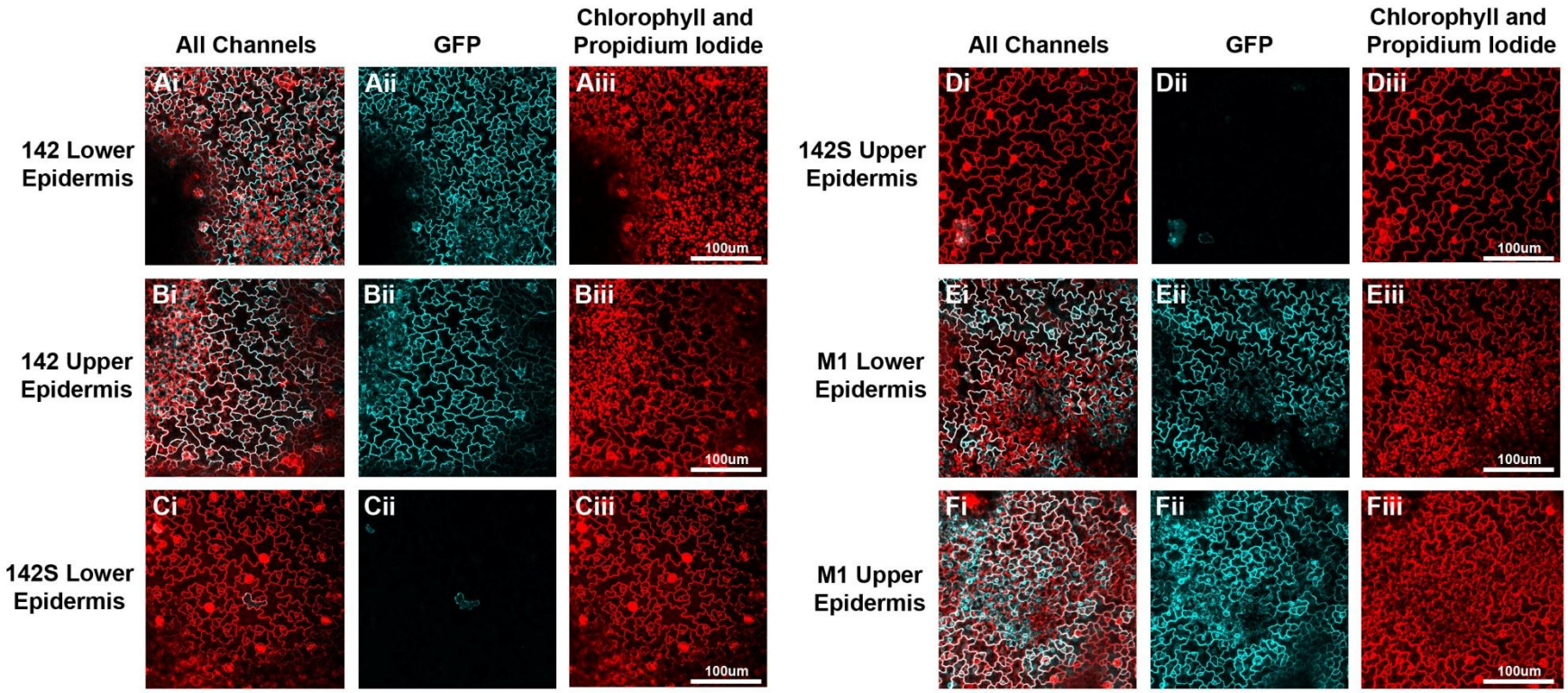


Figure 7.12A: GFP expression in epidermal layers of 5th leaves of 14 dpg plants at 7 hpd

The 5th leaves of 14 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 7 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan)

and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were image: 142 (**A** and **B**), 142S (**C** and **D**) and M1 (**E** and **F**). A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

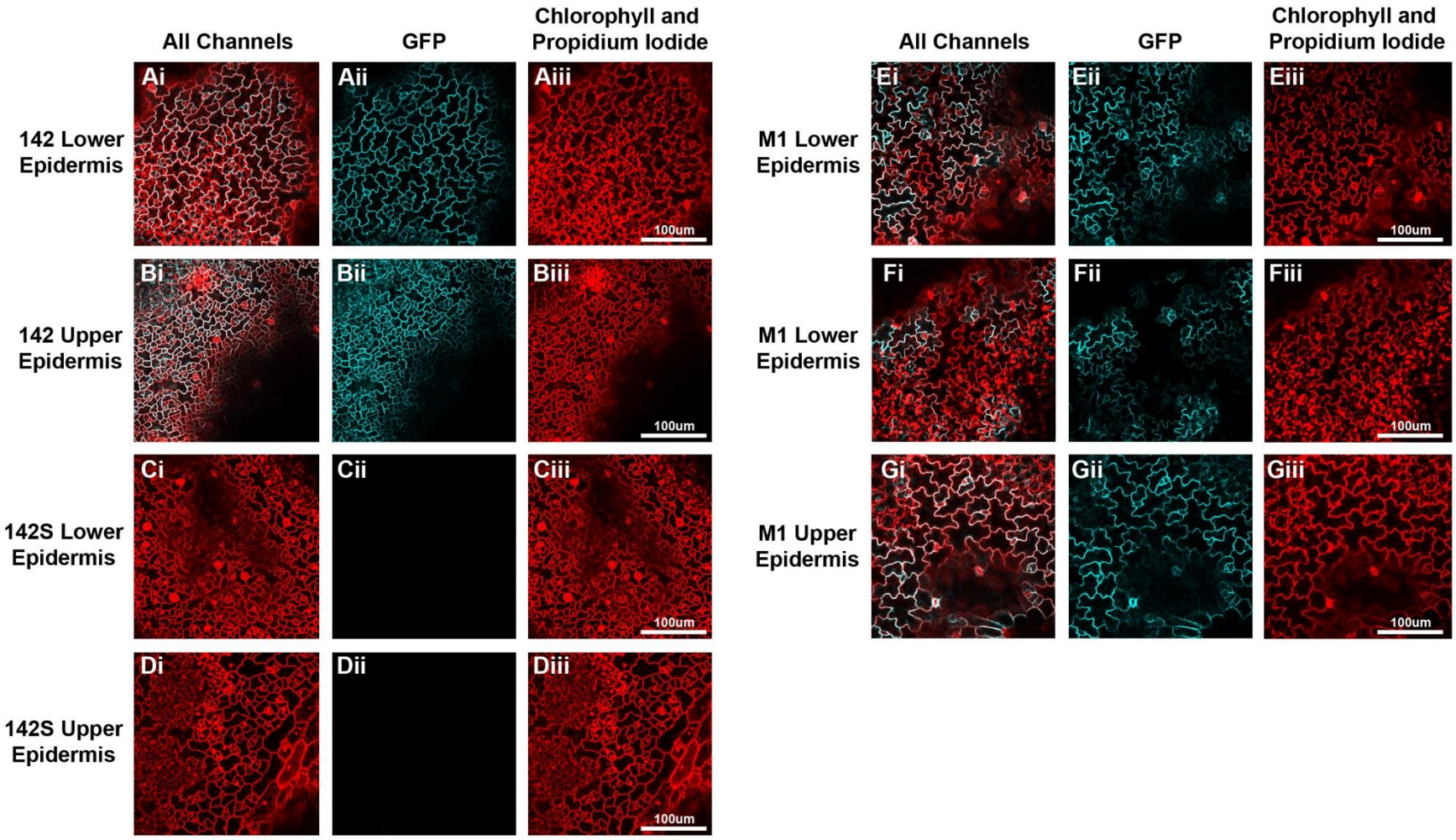

All Channels
GFP
Chlorophyll and Propidium Iodide
142 Lower Epidermis
142 Upper Epidermis
142S Lower Epidermis
142S Upper Epidermis
M1 Lower Epidermis
M1 Lower Epidermis
M1 Upper Epidermis
Ai
Aii
Aiii
Bi
Bii
Biii
Ci
Cii
Ciii
Di
Dii
Diii
Ei
Eii
Eiii
Fi
Fii
Fiii
Gi
Gii
Giii
100um

Figure 7.13A: GFP expression in epidermal layers of 5th leaves of 14 dpg plants at 13 hp

The 5th leaves of 14 dpg seedlings were imaged under a confocal microscope to observe GFP expression. These images were taken at 13 hpd. The upper and lower epidermis of each leaf was imaged. All images were taken at 20 time magnification. Two separate channels were used, one for GFP (cyan) and one for propidium iodide and chlorophyll (red). In all cases a combined image of both channels (**i**) and separate images of just the GFP channel (**ii**) and propidium iodide and chlorophyll channel (**iii**) are shown. Leaves from the following plant lines were image: 142 (**A** and **B**), 142S (**C** and **D**) and M1 (**E**, **F** and **G**). **E** and **F** show areas of the leaf where silencing is less (**E**) or more (**F**) prevalent. A scale bar is shown for each image in the propidium channel only image. Images are representative of the three leaves visualised.

Appendix 2.3 Sorting GFP negative and positive protoplasts in M1 and M9 samples

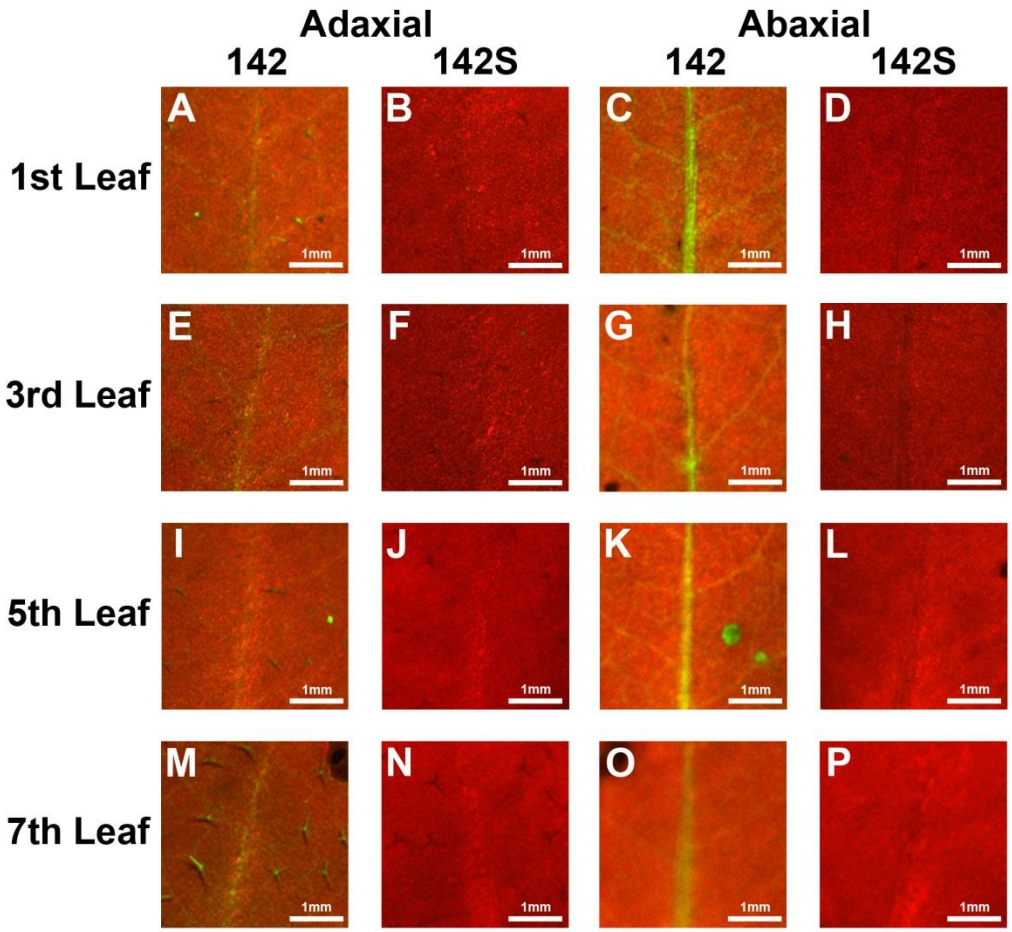


Figure 7.14A: GFP expression of the adaxial and abaxial side of M1 and M9 leaves at 21 dpg

Images of the adaxial and abaxial side of odd numbered leaves from lines 142 and 142S, taken under a fluorescence microscope at 21 dpg. Images **A**, **B**, **C** and **D** are of the 1st leaf, images **E**, **F**, **G** and **H** are of the 3rd leaf, images **I**, **J**, **K** and **L** are of the 5th leaf and images of **M**, **N**, **O** and **P** are of the 7th leaf. The first column (**A**, **E**, **I** and **M**) are image of the adaxial side of 142 leaves. The second column (**B**, **F**, **J** and **N**) are images of the adaxial side of 142S leaves. The third column (**C**, **G**, **K** and **O**) are image of the abaxial side of 142 leaves. The fourth column (**D**, **H**, **L** and **P**) are images of the abaxial side of 142S leaves. A scale bar is shown at the bottom right of each image.

Appendix 2.4 Sorting GFP negative and positive protoplasts in M1 and M9 samples

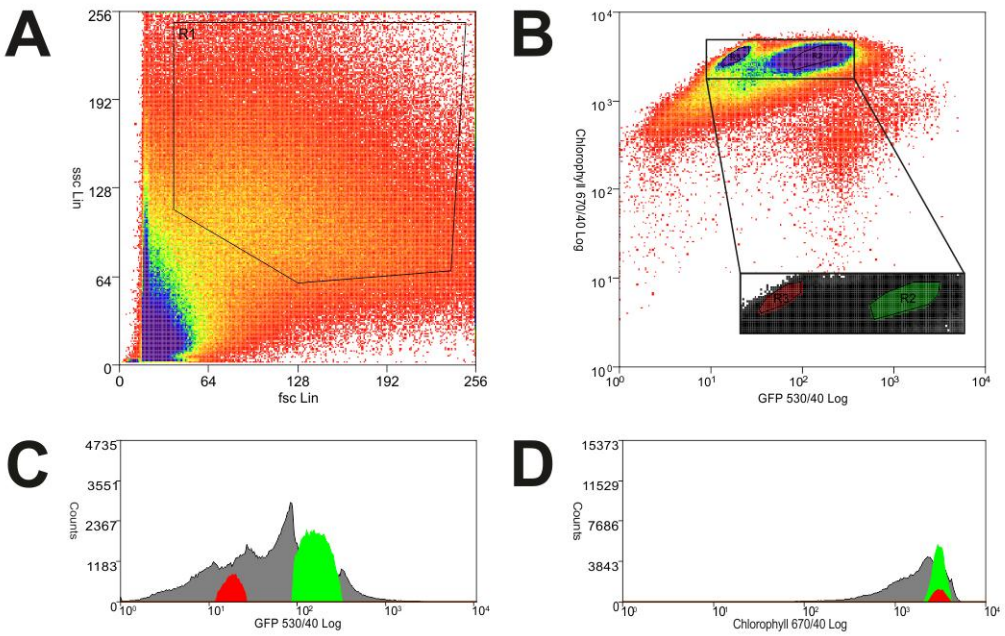


Figure 7.15A: Sorting GFP positive and negative protoplasts in M1 sample

Graphs showing results of flow cytomtery analysis of the M1 protoplast sample. **A**: Dot plot of ssc against fsc. The values for the y axis and x axis are relative values of ssc and fsc respectively. The dots colour indicates how many events are found at this point. It ranges from red (low number events) to purple (high number events).The R1 box highlights the area of the graph where protoplasts are found. Anything below or to the left of the R1 box is likely to be cell debris, while anything above or to the right of the R1 box is likely to be multiple protoplasts still attached to each other. The R1 box acts as a gate to the other 3 graphs, so only points found within the R1 box are shown on the other graphs. **B**: Dot plot of chlorophyll fluorescence intensity against GFP fluorescence intensity. Both axes give relative values of intensity for the two flurophores. The R2 box highlights the GFP positive protoplast population. The R3 box highlights the GFP negative protoplast population. This area of the graph has been blown up. **C**: Histogram of intensity of GFP fluorescence. **D**: Histogram of intensity of chlorophyll fluorescence. For both **C** and **D** the areas of the histograms corresponding to the GFP positive protoplasts is highlighted in green while the areas if the histograms corresponding to the

GFP negative protoplasts is highlighted in red. The x axes for both **C** and **D** are relative measurements of intensity. In D the GFP positive and negative protoplast populations overlap each other.

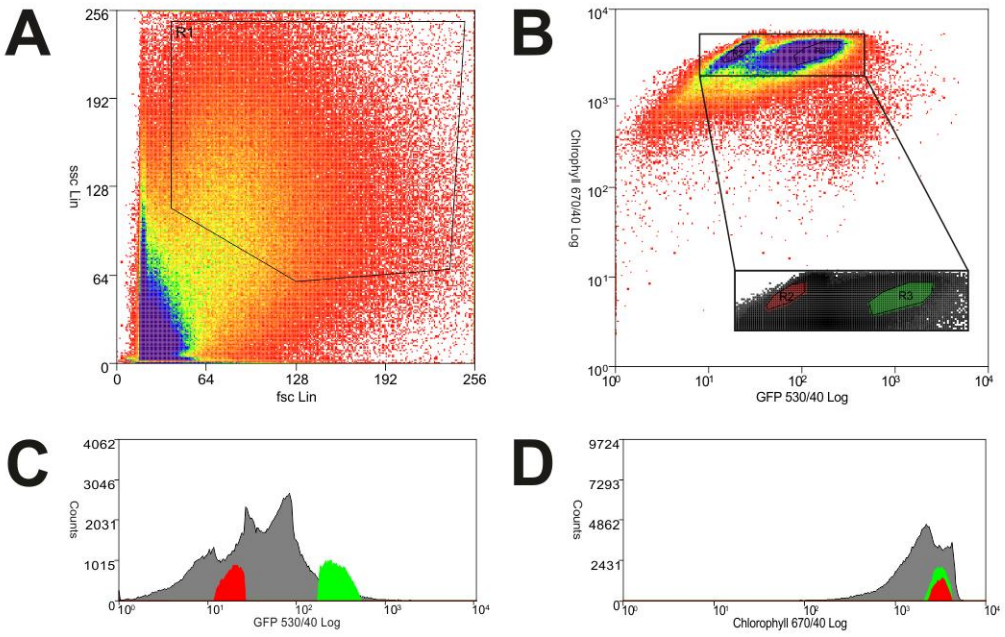


Figure 7.16A: Sorting GFP positive and negative protoplasts in M9 sample

Graphs showing results of flow cytomtery analysis of the M9 protoplast sample. **A**: Dot plot of ssc against fsc. The values for the y axis and x axis are relative values of ssc and fsc respectively. The dots colour indicates how many events are found at this point. It ranges from red (low number events) to purple (high number events).The R1 box highlights the area of the graph where protoplasts are found. Anything below or to the left of the R1 box is likely to be cell debris, while anything above or to the right of the R1 box is likely to be multiple protoplasts still attached to each other. The R1 box acts as a gate to the other 3 graphs, so only points found within the R1 box are shown on the other graphs. **B**: Dot plot of chlorophyll fluorescence intensity against GFP fluorescence intensity. Both axes give relative values of intensity for the two flurophores. The R2 box highlights the GFP positive protoplast population. The R3 box highlights the GFP negative protoplast population. This area of the graph has been blown up. **C**: Histogram of intensity of GFP fluorescence. **D**: Histogram of intensity of chlorophyll fluorescence. For both **C** and **D** the areas of the histograms corresponding to the GFP positive

protoplasts is highlighted in green while the areas if the histograms corresponding to the GFP negative protoplasts is highlighted in red. The x axes for both **C** and **D** are relative measurements of intensity. In D the GFP positive and negative protoplast populations overlap each other.

Appendix 3

Appendix 3.1. Repeat of Solo LTR Southern blot

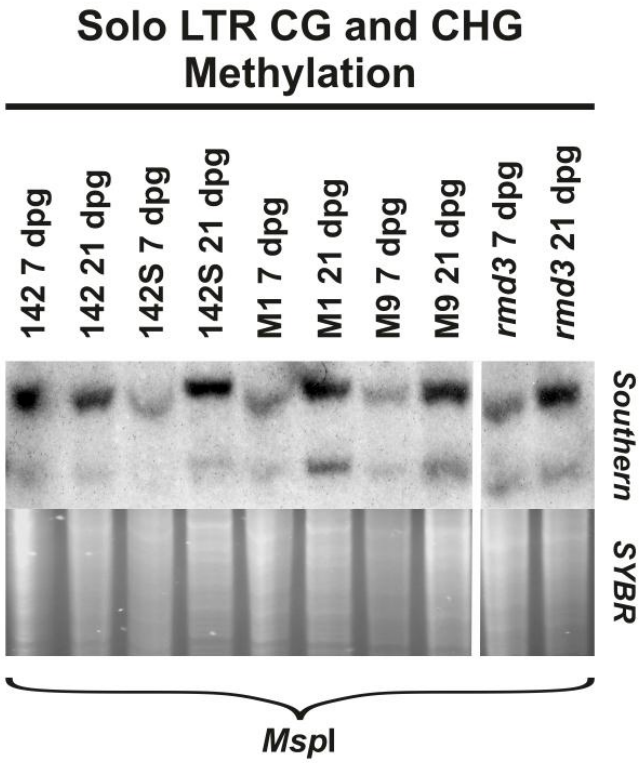


Figure 7.17A: CHG methylation pattern of Solo LTR

Southern blot looking at CG and CHG methylation of Solo LTR. DNA from 7 dpg and 21 dpg plants of lines 142, 142S, M1, M9, *rdr2-1* and *rmd3* were used. The order of loading is given at the top of the blot. DNA was digested with *Msp*I*,* which is sensitive to CG and CHG methylation. An image of the probed Southern blot and part of the SYBR safe stained gel used for blotting are shown. The stained gel indicates the amount of DNA for each sample used for the Southern blot. Two lanes have been removed from the images

of the blot and gel which is indicated by the white vertical line between the M9 21 dpg and rmd3 7 dpg samples.

Appendix 3.2 Detection of Solo LTR siRNAs in floral tissue

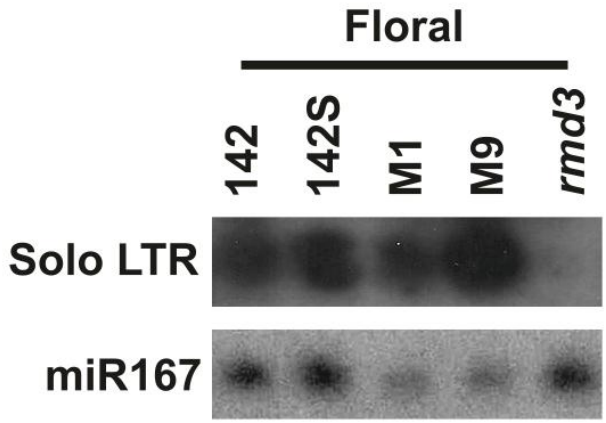


Figure 7.18A: Northern blot for detection of Solo LTR siRNAs

Images of a northern blot probed with ^{32}P-labelled probes for Solo LTR (top) and miR167 (bottom). Floral RNA samples of lines 142, 142S, M1, M9 and *rmd3* were used and the order of loading is given above the blots. For this blot the miR167 probe acts as a loading control. The *rmd3* sample is a negative control as it is a mutant in *nrpd1* and so should lack 24 nt siRNAs.

Appendix 3.3 Confirmation of normal distribution of LN and descriptive statistics for each line

	Line					
Statistic	**C24**	**142**	**142S**	**M1**	**M9**	***rmd1***
Kolmogorov-Smirnova	0.200	0.075	0.020	0.088	0.200	0.193
Shapiro-Wilk	0.488	0.326	<0.001	0.067	0.303	0.675

Table 7.5A: Table showing statistics testing normal distribution of LN data for each line tested

Two statistical tests that both look at the normal distribution of the LN data for each line. The significance value for each line is given for both tests. For all lines, except 142S, both tests show there is no significant difference and so the data is normally distributed. For 142S both tests shows that 142S is not normally distributed.

	Line					
Statistic	**C24**	**142**	**142S**	**M1**	**M9**	***rmd1***
Mean	61.02 (0.642)	62.30 (0.636)	62.96 (0.895)	67.76 (0.799)	59.17 (0.696)	55.25 (0.681)
Median	61.00	64.00	64.00	68.00	59.00	54.50
Variance	56.85	42.52	96.11	84.22	60.99	57.46
SD	7.54	6.52	9.80	9.18	7.81	7.58
Min. Value	38.00	46.00	37.00	49.00	41.00	34.00
Max. Value	78.00	83.00	115.00	96.00	80.00	76.00
Range	40.00	37.00	78.00	47.00	39.00	42.00
IQR	10.00	8.00	13.00	12.00	10.00	12.00

Table 7.6A: Table of descriptive statistic for LN data from each line tested

Table showing the following descriptive statistics for leaf number (LN) data for each line tested: mean, median, variance, standard deviation (SD), minimum value, maximum value, range and interquartile range (IQR). The lines tested were C24 WT, 142, 142S, M1, M9 and *rmd1.* For the mean the number in brackets is the standard error of each mean. There are no units for these statistics as the data is count data.

Appendix 3.4 Confirmation of normal distribution of rosette weight and descriptive statistics for each line

	Line					
Statistic	**C24**	**142**	**142S**	**M1**	**M9**	***rmd1***
Kolmogorov-Smirnova	0.200	0.200	0.200	0.200	0.200	0.200
Shapiro-Wilk	0.079	0.060	0.78	0.982	0.561	0.735

Table 7.7A: Table showing statistics testing normal distribution of rosette weight data for each line tested

Two statistical tests that both look at the normal distribution of the rosette weight data for each line. The significance value for each line is given for both tests. For all lines both tests show there is no significant difference and so the data is normally distributed.

	Line					
Statistic	**C24**	**142**	**142S**	**M1**	**M9**	***rmd1***
Mean	1.523 (0.039)	1.868 (0.050)	0.876 (0.019)	0.759 (0.019)	0.709 (0.015)	0.652 (0.016)
Median	1.470	1.870	0.890	0.750	0.710	0.670
Variance	0.210	0.260	0.043	0.050	0.029	0.032
SD	0.459	0.510	0.208	0.223	0.170	0.179
Min. Value	0.360	0.450	0.390	0.160	0.310	0.200
Max. Value	2.680	3.640	1.500	1.400	1.130	1.050
Range	2.320	3.190	1.110	1.240	0.820	0.850
IQR	0.540	0.570	0.280	0.290	0.230	0.220

Table 7.8A: Table of descriptive statistic for rosette weight data from each line tested

Table showing the following descriptive statistics for rosette weight data for each line tested: mean, median, variance, standard deviation (SD), minimum value, maximum value, range and interquartile range (IQR). The lines tested were C24 WT, 142, 142S, M1, M9 and *rmd1.* For the mean the number in brackets is the standard error of each mean. The units for all the statistics, bar the variance, are in grams (g) and for the variance the units is g^2.

Abbreviations

A	Adenine
aa	Amino acid
AGO	Argonaute
ANOVA	Analysis of variance
APS	Ammonium persulfate
At	*Arabidopsis thaliana* (Thale Cress)
AtMu1	*Arabidopsis thaliana Mutator*-like 1
ATP	Adenosine triphosphate
AtSN1	*Arabidopsis thaliana* SINE 1
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
Bt	*Bos Taurus* (Cattle)
C	Cytosine
$CaCl_2$	Calcium di-chloride
CaMV	Cauliflower mosaic virus
CAPS	Cleaved amplified polymorphic sequences
CDS	Coding sequence
Ce	*Caenorhabditis elegans* (Nematode worm)
Chr	Chromosome
CHS	Chalcone synthase
CLSY	Classy
CMM	Conserved MOM1 motif
CMT	Chromomethylase
CNV	Copy number variant
Col	Columbia
CRH	Compromised recognition of TCV homolog
CRT	Compromised recognition of TCV

CTAB	Cetyltrimethylammonium bromide
CTP	Cytidine triphosphate
dATP	Deoxyadenosine triphosphate
DCL	Dicer like
dCTP	Deoxygcytidine triphosphate
DDM	Decrease in DNA methylation
dGTP	Deoxyguanidine triphosphate
dH_2O	Deionised water
DME	Demeter
DML	Demeter-like
DMS	Defective in meristem silencing
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
dpg	Days post germination
Dr	*Danio rerio* (Zebrafish)
DRB	Double stranded RNA binding protein
DRD	Defective in RNA-directed DNA methylation
DRM	Domains rearranged methyltransferase
ds	Double stranded
DTF	DNA binding transcription factor
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EMS	Ethyl ester methane sulfonic acid
ER	Endoplasmic reticulum
EtBr	Ethidium bromide
FACT	Facilitates chromatin transcription/transaction)

FDM	Factor of DNA methylation
FLC	Flowering time locus C
fsc	Forward scatter
FWA	Flowering wageningen
FUBAR	Rude word for properly screwed
G	Guanine
g	Gram
GFP	Green fluorescent protein
GHKL	Gyrase, Hsp90, Histidine Kinase, MutL
GTP	Guanidine triphosphate
GW/WG	Glycine tryptophan/tryptophan glycine
H	Cytosine, adenine or thymine
H1	Histone one
H2	Histone two
H3	Histone three
H4	Histone four
H_2O	Water
HCl	Hydrochloric acid
HDA	Histone deacetylase
HEN	HUA enhancer
HMW	High molecular weight
hpd	Hours post dawn
Hs	*Homo sapiens* (Humans)
HSP	Heat shock protein
Hz	Hertz
IDM	Increased DNA methylation
IDN	Involved in *de novo*
IDNL	Involved in *de novo*-like
IDP	IDN paralog
IGN	Intergenic loci

IGR	Intergenic region
Indel	Insertion or deletion
IQR	Interquartile range
IR	Inverted repeat
JMJ	Jumonji
kb	Kilobase
K	Lysine
KCl	Potassium chloride
KOH	Potassium hydroxide
KTF	KOW domain-containing transcription factor
l	Litre
LB	Lysogeny broth
LCR	Low-complexity region
Ler	Landsberg *erecta*
LHP	Like heterochromatin protein
LMW	Low molecular weight
LN	Leaf number
LSD	Lysine specific demethylase (Chapter one)
LSD	Least significant difference (Chapter five)
LTR	Long terminal repeat
M	Molar
mA	Miliamps
Max.	Maximum
MBD	Methyl- CpG- binding domain
MBq	Mega becquerels
MDS	Multi-dimensional scaling
MEA-ISR	*Medea*-intergenic subtelomeric repeats
MES	2-(N-Morpholino)ethanesulfonic acid
MET	Methyltransferase
mg	Milligram

Min	Minimum
miRNA	Micro RNA
ml	Millilitre
Mm	*Mus musculus*
mM	Millimolar
mm	Millimetre
Mo	*Myroides odoratimimus* (Bacteria species)
MOM	Morpheus molecule
MOPS	3-(N-morpholino)propanesulfonic acid
MORC	Microrchidia
mRNA	Messenger RNA
MS	Murashige and Skoog
Mt	*Medicago truncatula* (Barrel clover)
MULE	*Mutator*-like element
MZ	Metazoan
N	Cytosine, guanine, adenine or thymine
Na_2HPO_4	Di-sodium phosphate
Na_3CIT	Tri-sodium citrate
NaAc	Sodium acetate
NaCl	Sodium chloride
NaH_2PO_4	Sodium di-hydrogen phosphate
NaOH	Sodium hydroxide
nat-siRNA	Natural-antisense small interfering RNA
NB	Nuclear bodies
NERD	Needed for RDR2-independent DNA methylation
ng	Nanogram
NLS	Nuclear localisation signal
nm	Nano metres
NOS	Nopaline synthase
NRPB	Nuclear RNA polymerase B (II)

NRPD	Nuclear RNA polymerase D (IV)
NRPE	Nuclear RNA polymerase E (V)
nt	Nucleotide
Nv	*Nematostella vectensis* (Starlet sea anemone)
Oz	*Oryza sativa* (Rice)
p	Promoter
^{32}P	^{32}Phosphorous
PAZ	Piwi Argonaute Zwille
PCR	Polymerase chain reaction
PDS	Phytoene desaturase
PEG	Polyethylene glycol
PHB	Phabulosa
PHV	Phavoluta
Phyre	Protein homology/analogy recognition engine
piRNA	PIWI RNA
PIWI	P-element induced wimpy testis
PL	Plantae
PML	Promyelocytic leukemia
PolIV	Nuclear RNA polymerase IV (D)
PolV	Nuclear RNA polymerase V (E)
Pp	*Physcomitrella patens* (Moss)
Pt	*Populus trichocarpa* (Black Cottonwood)
PTGS	Post-transcriptional gene silencing
PVP	Polyvinylpyrrolidone
rcf	Relative centrifugal force
rDNA	Ribosomal RNA gene
RdDM	RNA-directed DNA methylation
RDM	RNA-directed DNA methylation
RDR	Plant RNA-dependent RNA polymerase
RING	Real interesting new gene

RISC	RNA-induced silencing complex
RMD	RNA-directed DNA methylation defective
RMR	Required to maintain repression
RNA	Ribonucleic acid
RNAi	RNA interference
ROI	Region of interest
ROS	Repressor of silencing
rpm	Revolutions per mintue
rRNA	Ribosomal RNA
RT	Reverse transcriptase
s	Seconds
SAM	S-adenosyl methionine
SD	Standard deviation
SDC	Suppressor of drm2 and cmt3
SDS	Sodium dodecyl sulfate
SET	Suppressor of variegation 3-9, Enhancer of zeste and Trithorax
SHH	Sawadee homeodomain homolog
SINE	Short interspersed elements
siRNA	Small interfering RNA
Sm	*Selaginella moellendorffii* (Lycophyte)
SMART	Simple Modular Architecture Research Tool
SMC	Structural maintenance of chromosomes
SmcMORC	Structural maintenance of chromosomes microrchidia
SNP	Single nucleotide polymorphism
SR	Serine/arginine rich
SRA	SET and RING associated
sRNA	Small RNA
ss	Single stranded
SSC	Saline-sodium citrate
ssc	Side scatter

SSLP	Simple sequence length polymorphisms
SSRP	Structure specific recognition protein
SUP	Suppressor of ros1
SUVH	SU(Var)3-9 homolog
T	Thymine
TAIR	The Arabidopsis information resource
tasiRNA	Trans-acting small interfering RNAs
TBE	Tris borate EDTA
TCV	Turnip crinkle virus
TE	Transposable element
te	Tris-EDTA
TEMED	Tris(hydroxymethyl)aminomethane
TGS	Transcriptional gene silencing
Tm	Primer melting temperature
tRNA	Transfer RNA
T-DNA	Transfer-DNA
U	Uracil
UBP	Ubiquitin protease
UTP	Uridine triphosphate
UTR	Untranslated region
UV	Ultra violet
V	Volts
Vv	*Vitis vinifera* (Grape vine)
WT	Wild type
x g	Times gravity
Xt	*Xenopus tropicalis* (West African Clawed Frog)
Zm	*Zea mays* (Maize)
α	Alpha
β	Beta
γ	Gamma

µg	Microgram
µl	Microlitre
°C	Degrees Celcius
%	Percentage

References

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouché, N., Gasciolli, V. and Vaucheret, H. (2006) 'DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7', *Current Biology,* 16(9), 927-932.

Agius, F., Kapoor, A. and Zhu, J.-K. (2006) 'Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation', *Proceedings of the National Academy of Sciences of the United States of America,* 103(31), 11796-11801.

Agorio, A. and Vera, P. (2007) 'ARGONAUTE4 is required for resistance to pseudomonas syringae in *Arabidopsis*', *The Plant Cell,* 19(11), 3778-3790.

Ahmed, I., Sarazin, A., Bowler, C., Colot, V. and Quesneville, H. (2011) 'Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in *Arabidopsis*', *Nucleic Acids Research,* 39(16), 6919-6931.

Allen, E. and Howell, M. D. (2010) 'miRNAs in the biogenesis of trans-acting siRNAs in higher plants', *Seminars in Cell & Developmental Biology,* 21(8), 798-804.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C. C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D. E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W. L., Berry, C. C. and Ecker, J. R. (2003) 'Genome-wide insertional mutagenesis of *Arabidopsis thaliana*', *Science,* 301(5633), 653-657.

Alonso-Blanco, C., El-Assal, S. E.-D., Coupland, G. and Koornneef, M. (1998) 'Analysis of natural allelic variation at flowering time loci in the Landsberg

erecta and Cape Verde Islands Ecotypes of *Arabidopsis thaliana*', *Genetics,* 149(2), 749-764.

Amasino, R. (2004) 'Vernalization, competence, and the epigenetic memory of winter', *The Plant Cell,* 16(10), 2553-2559.

Amedeo, P., Habu, Y., Afsar, K., Scheid, O. M. and Paszkowski, J. (2000) 'Disruption of the plant gene *MOM* releases transcriptional silencing of methylated genes', *Nature,* 405(6783), 203-206.

An, G., Costa, M. A., Mitra, A., Ha, S. B. and Marton, L. (1988) 'Organ-specific and developmental regulation of the *Nopaline Synthase* promoter in transgenic tobacco plants', *Plant Physiology,* 88(3), 547-552.

Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., Mühlenbock, P., Skirycz, A., Gonzalez, N., Beemster, G. T. S. and Inzé, D. (2012) 'Exit from proliferation during leaf development in *Arabidopsis thaliana*: A not-so-gradual process', *Developmental Cell,* 22(1), 64-78.

Aravin, A. A. and Bourc'his, D. (2008) 'Small RNA guides for *de novo* DNA methylation in mammalian germ cells', *Genes & Development,* 22(8), 970-975.

Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M. and Gvozdev, V. A. (2001) 'Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline', *Current Biology,* 11(13), 1017-1027.

Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T. and Hannon, G. J. (2008) 'A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice', *Molecular Cell,* 31(6), 785-799.

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. and Hannon, G. J. (2007) 'Developmentally regulated piRNA clusters implicate MILI in transposon control', *Science,* 316(5825), 744-747.

Ashapkin, V. V., Kutueva, L. I. and Vanyushin, B. F. (2002) 'The gene for domains rearranged methyltransferase (DRM2) in *Arabidopsis thaliana* plants is methylated at both cytosine and adenine residues', *Federation of European Biochemical Societies Letters,* 532(3), 367-372.

Aufsatz, W., Mette, M., Matzke, A. and Matzke, M. (2004) 'The role of MET1 in RNA-directed *de novo* and maintenance methylation of CG dinucleotides', *Plant Molecular Biology,* 54(6), 793-804.

Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, A. J. M. and Matzke, M. (2002a) 'RNA-directed DNA methylation in *Arabidopsis*', *Proceedings of the National Academy of Sciences of the United States of America,* 99(Supplement 4), 16499-16506.

Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, M. and Matzke, A. J. (2002b) 'HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA', *Europen Molecular Biology Organization Journal,* 21(24), 6832-6841.

Aukerman, M. J. and Sakai, H. (2003) 'Regulation offlowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes', *Plant Cell,* 15(11), 2730-2741.

Ausin, I., Greenberg, M. V., Li, C. F. and Jacobsen, S. E. (2012a) 'The splicing factor SR45 affects the RNA-directed DNA methylation pathway in *Arabidopsis*', *Epigenetics,* 7(1), 29-33.

Ausin, I., Greenberg, M. V. C., Simanshu, D. K., Hale, C. J., Vashisht, A. A., Simon, S. A., Lee, T.-f., Feng, S., Española, S. D., Meyers, B. C., Wohlschlegel, J. A., Patel, D. J. and Jacobsen, S. E. (2012b) 'INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in *Arabidopsis*', *Proceedings of the National Academy of Sciences of the United States of America,* 109(22), 8374-8381.

Ausin, I., Mockler, T. C., Chory, J. and Jacobsen, S. E. (2009) 'IDN1 and IDN2: two proteins required for de novo DNA methylation in *Arabidopsis thaliana*', *Nature Structural Molecular Biology,* 16(12), 1325-1327.

Axtell, M. J., Jan, C., Rajagopalan, R. and Bartel, D. P. (2006) 'A two-hit trigger for siRNA biogenesis in plants', *Cell,* 127(3), 565-577.

Baek, D., Jiang, J., Chung, J.-S., Wang, B., Chen, J., Xin, Z. and Shi, H. (2011) 'Regulated *AtHKT1* gene expression by a distal enhancer element and DNA Methylation in the promoter plays an important role in salt tolerance', *Plant and Cell Physiology,* 52(1), 149-161.

Baev, V., Naydenov, M., Apostolova, E., Ivanova, D., Doncheva, S., Minkov, I. and Yahubyan, G. (2010) 'Identification of RNA-dependent DNA-methylation regulated promoters in *Arabidopsis*', *Plant Physiology and Biochemistry,* 48(6), 393-400.

Ballestar, E. and Wolffe, A. P. (2001) 'Methyl-CpG-binding proteins', *European Journal of Biochemistry,* 268(1), 1-6.

Ban, C., Junop, M. and Yang, W. (1999) 'Transformation of MutL by ATP binding and hydrolysis: A switch in DNA mismatch repair', *Cell,* 97(1), 85-97.

Ban, C. and Yang, W. (1998) 'Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis', *Cell,* 95(4), 541-552.

Bao, N., Lye, K.-W. and Barton, M. K. (2004) 'microRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome', *Developmental Cell,* 7(5), 653-662.

Barber, W. T., Zhang, W., Win, H., Varala, K. K., Dorweiler, J. E., Hudson, M. E. and Moose, S. P. (2012) 'Repeat associated small RNAs vary among parents and following hybridization in maize', *Proceedings of the National Academy of Sciences,* 109(26), 10444-10449.

Bartee, L., Malagnac, F. and Bender, J. (2001) '*Arabidopsis CMT3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene', *Genes & Development,* 15(14), 1753-1758.

Bateman, A. (2002) 'The SGS3 protein involved in PTGS finds a family', *BioMed Central Bioinformatics,* 5(3), 21.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P. and Izaurralde, E. (2006) 'mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes', *Genes & Development,* 20(14), 1885-1898.

Bell, C. J. and Ecker, J. R. (1994) 'Assignment of 30 Microsatellite loci to the linkage map of *Arabidopsis*', *Genomics,* 19(1), 137-144.

Bender, J. (2012) 'RNA-Directed DNA Methylation: Getting a grip on mechanism', *Current Biology,* 22(10), R400-R401.

Berg, A., Meza, T. J., Mahić, M., Thorstensen, T., Kristiansen, K. and Aalen, R. B. (2003) 'Ten members of the *Arabidopsis* gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, AtMBD11, is crucial for normal development', *Nucleic Acids Research,* 31(18), 5291-5304.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.-C., Nicolas, A. and Forterre, P. (1997) 'An atypical topoisomerase II from archaea with implications for meiotic recombination', *Nature,* 386(6623), 414-417.

Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J. (2001) 'Role for a bidentate ribonuclease in the initiation step of RNA interference', *Nature,* 409(6818), 363-366.

Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R. and Lagrange, T. (2009) 'RNA-directed DNA methylation requires an AGO4-

interacting member of the SPT5 elongation factor family', *European Molecular Biology Organization Reports,* 10(6), 649-654.

Bilwes, A. M., Alex, A., Crane, B. R. and Simon, M. I. (1999) 'Structure of CheA, a signal-transducing histidine kinase', *Cell,* 96(1), 131-141.

Blewitt, M. E., Gendrel, A.-V., Pang, Z., Sparrow, D. B., Whitelaw, N., Craig, J. M., Apedaile, A., Hilton, D. J., Dunwoodie, S. L., Brockdorff, N., Kay, G. F. and Whitelaw, E. (2008) 'SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation', *Nature Genetics,* 40(5), 663-669.

Blázquez, M. A., Green, R., Nilsson, O., Sussman, M. R. and Weigel, D. (1998) 'Gibberellins promote flowering of *Arabidopsis* by activating the *LEAFY* promoter', *The Plant Cell,* 10(5), 791-800.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M. and Benning, C. (1998) 'AGO1 defines a novel locus of *Arabidopsis* controlling leaf development', *European Molecular Biology Organization Journal,* 17(1), 170-80.

Bolden, A., Ward, C., Siedlecki, J. A. and Weissbach, A. (1984) 'DNA methylation. Inhibition of *de novo* and maintenance methylation in vitro by RNA and synthetic polynucleotides', *Journal of Biological Chemistry,* 259(20), 12437-12443.

Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R. and Zhu, J.-K. (2005) 'Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*', *Cell,* 123(7), 1279-1291.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G. J. (2007) 'Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*', *Cell,* 128(6), 1089-1103.

Brosnan, C. A., Mitter, N., Christie, M., Smith, N. A., Waterhouse, P. M. and Carroll, B. J. (2007) 'Nuclear gene silencing directs reception of long-

distance mRNA silencing in Arabidopsis', *Proceedings of the National Academy of Sciences of the United State of America,* 104(37), 14741-14746.

Bruno, M., Flaus, A., Stockdale, C., Rencurel, C., Ferreira, H. and Owen-Hughes, T. (2003) 'Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities', *Molecular cell,* 12(6), 1599-1606.

Brzeski, J. and Jerzmanowski, A. (2003) 'Deficient in DNA Methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors', *Journal of Biological Chemistry,* 278(2), 823-828.

Buschhausen, G., Wittig, B., Graessmann, M. and Graessmann, A. (1987) 'Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene', *Proceedings of the National Academy of Sciences of the United State of America,* 84(5), 1177-1181.

Campell, B. R., Song, Y., Posch, T. E., Cullis, C. A. and Town, C. D. (1992) 'Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*', *Gene,* 112(2), 225-228.

Cao, J., Schneeberger, K., Ossowski, S., Gunther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., Wang, X., Ott, F., Muller, J., Alonso-Blanco, C., Borgwardt, K., Schmid, K. J. and Weigel, D. (2011) 'Whole-genome sequencing of multiple *Arabidopsis thaliana* populations', *Nature Genetics,* 43(10), 956-963.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M. F., Huang, M. S., Matzke, M. and Jacobsen, S. E. (2003) 'Role of the DRM and CMT3 methyltransferases in RNA-directed DNA Methylation', *Current Biology,* 13(24), 2212-2217.

Cao, X. and Jacobsen, S. E. (2002a) 'Locus-specific control of asymmetric and CpNpG methylation by the *DRM* and *CMT3* methyltransferase genes', *Proceedings of the National Academy of Sciences of the United States of America,* 99(Suppl 4), 16491-16498.

Cao, X. and Jacobsen, S. E. (2002b) 'Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing', *Current Biology,* 12(13), 1138-1144.

Cao, X., Springer, N. M., Muszynski, M. G., Phillips, R. L., Kaeppler, S. and Jacobsen, S. E. (2000) 'Conserved plant genes with similarity to mammalian de novo DNA methyltransferases', *Proceedings of the National Academy of Sciences of the United States of America,* 97(9), 4979-4984.

Carmell, M. A., Girard, A., van de Kant, H. J. G., Bourc'his, D., Bestor, T. H., de Rooij, D. G. and Hannon, G. J. (2007) 'MIWI2 Is essential for spermatogenesis and repression of transposons in the mouse male germline', *Developmental cell,* 12(4), 503-514.

Chan, S. W. L., Henderson, I. R., Zhang, X., Shah, G., Chien, J. S. C. and Jacobsen, S. E. (2006) 'RNAi, DRD1, and Histone Methylation Actively Target Developmentally Important Non-CG DNA Methylation in *Arabidopsis*', *Public Library of Science Genetics,* 2(6).

Chan, S. W. L., Zilberman, D., Xie, Z., Johansen, L. K., Carrington, J. C. and Jacobsen, S. E. (2004) 'RNA silencing genes control *de Novo* DNA methylation', *Science,* 303(5662), 1336.

Chen, H. M., Chen, L. T., Patel, K., Li, Y. H., Baulcombe, D. C. and Wu, S. H. (2010) '22-nucleotide RNAs trigger secondary siRNA biogenesis in plants', *Proceedings of the National Academy of Sciences of the United States of America,* 107(34), 15269-15274.

Chen, Z. J. (2010) 'Molecular mechanisms of polyploidy and hybrid vigor', *Trends in Plant Science,* 15(2), 57-71.

Chinnusamy, V. and Zhu, J.-K. (2009) 'Epigenetic regulation of stress responses in plants', *Current Opinion in Plant Biology,* 12(2), 133-139.

Chiu, A., Revenkova, E. and Jessberger, R. (2004) 'DNAinteraction and dimerization of eukaryotic SMC hinge domains', *Journal of Biological Chemistry,* 279(25), 26233-26242.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., Jacobsen, S. E. and Fischer, R. L. (2002) 'DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*', *Cell,* 110(1), 33-42.

Cloix, C., Tutois, S., Yukawa, Y., Mathieu, O., Cuvillier, C., Espagnol, M.-C., Picard, G. and Tourmente, S. (2002) 'Analysis of the 5S RNA pool in *Arabidopsis thaliana*: RNAs are heterogeneous and only two of the genomic *5S* loci produce mature 5S RNA', *Genome Research,* 12(1), 132-144.

Cogoni, C., Irelan, J. T., Schumacher, M., Schmidhauser, T. J., Selker, E. U. and Macino, G. (1996) 'Transgene silencing of the *AL-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation', *European Molecular Biology Organization Journal,* 15(12), 3153-3163.

Cogoni, C. and Macino, G. (1999) 'Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase', *Nature,* 399(6732), 166-169.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. y. F., Pellegrini, M. and Jacobsen, S. E. (2008) 'Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning', *Nature,* 452(7184), 215-219.

Coletta, A., Pinney, J., Solis, D., Marsh, J., Pettifer, S. and Attwood, T. (2010) 'Low-complexity regions within protein sequences have position-dependent roles', *BioMed Central Systems Biology,* 4(43).

Collins, S. R., Miller, K. M., Maas, N. L., Roguev, A., Fillingham, J., Chu, C. S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., Ding, H., Xu, H., Han, J., Ingvarsdottir, K., Cheng, B., Andrews, B., Boone, C., Berger, S. L., Hieter, P., Zhang, Z., Brown, G. W., Ingles, C. J., Emili, A., Allis, C. D., Toczyski, D.

P., Weissman, J. S., Greenblatt, J. F. and Krogan, N. J. (2007) 'Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map', *Nature,* 446(7137), 806-810.

Crick, F. H. C. (1952) 'Is α-Keratin a Coiled Coil?', *Nature,* 170(4334), 882-883.

Cuperus, J. T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R. T., Takeda, A., Sullivan, C. M., Gilbert, S. D., Montgomery, T. A. and Carrington, J. C. (2010) 'Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*', *Nature Structural Molecular Biology,* 17(8), 997-1003.

Dalakouras, A., Dadami, E., Zwiebel, M., Krczal, G. and Wassenegger, M. (2012) 'Transgenerational maintenance of transgene body CG but not CHG and CHH methylation', *Epigenetics,* 7(9), 1071-1078.

Dalmay, T., Hamilton, A., Mueller, E. and Baulcombe, D. C. (2000a) 'Potato Virus X amplicons in *Arabidopsis* mediate genetic and epigenetic gene silencing', *Plant Cell,* 12(3), 369-380.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S. and Baulcombe, D. C. (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(5), 543-553.

Davuluri, R., Sun, H., Palaniswamy, S., Matthews, N., Molina, C., Kurtz, M. and Grotewold, E. (2003) 'AGRIS: *Arabidopsis* Gene Regulatory Information Server, an information resource of *Arabidopsis* cis-regulatory elements and transcription factors', *BioMed Central Bioinformatics,* 4(1), 25.

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A. J. M. and Matzke, M. (2009) 'A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation', *European Molecular Biology Organization Journal,* 28(1), 48-57.

Dehio, C. and Schell, J. (1994) 'Identification of plant genetic loci involved in a posttranscriptional mechanism for meiotically reversible transgene silencing', *Proceedings of the National Academy of Sciences of the United States of America,* 91(12), 5538-5542.

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K. D., Carrington, J. C. and Voinnet, O. (2006) 'Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense', *Science,* 313(5783), 68-71.

Deleris, A., Greenberg, M. V. C., Ausin, I., Law, R. W. Y., Moissiard, G., Schubert, D. and Jacobsen, S. E. (2010) 'Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation', *European Molecular Biology Organization Reports,* 11(12), 950-955.

Ding, L. and Han, M. (2007) 'GW182 family proteins are crucial for microRNA-mediated gene silencing', *Trends in Cell Biology,* 17(8), 411-416.

Donnelly, P. M., Bonetta, D., Tsukaya, H., Dengler, R. E. and Dengler, N. G. (1999) 'Cell cycling and cell enlargement in developing leaves of *Arabidopsis*', *Developmental Biology,* 215(2), 407-419.

Doskočil, J. and Šormová, Z. (1965) 'The occurrence of 5-methylcytosine in bacterial deoxyribonucleic acids', *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis,* 95(3), 513-515.

Douet, J., Tutois, S. and Tourmente, S. (2009) 'A Pol V–mediated silencing, independent of RNA–directed DNA Methylation, applies to 5S rDNA', *Public Library of Science Genetics,* 5(10).

Dunn, D. B. and Smith, J. D. (1958) 'The occurrence of 6-methylaminopurine in deoxyribonucleic acids', *Biochemical Journal,* 68(4), 627-36.

Dunoyer, P., Lecellier, C.-H., Parizotto, E. A., Himber, C. and Voinnet, O. (2004) 'Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing', *The Plant Cell,* 16(5), 1235-1250.

Dutta, R. and Inouye, M. (2000) 'GHKL, an emergent ATPase/kinase superfamily', *Trends in Biochemical Sciences,* 25(1), 24-28.

Dytham, C. (2005) 'Post hoc testing: after one-way ANOVA' in *Choosing and Using Statistics: A Biologist's Guide*, 2nd ed., UK: Blackwell Publishing, 116-117.

Eamens, A., Vaistij, F. E. and Jones, L. (2008) 'NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in *Arabidopsis*', *Plant Journal,* 55(4), 596-606.

Ebbs, M. L. and Bender, J. (2006) 'Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase', *The Plant Cell,* 18(5), 1166-1176.

Eden, S. and Cedar, H. (1994) 'Role of DNA methylation in the regulation of transcription', *Current Opinion in Genetics & Development,* 4(2), 255-259.

Ehrlich, M., Gama-Sosa, M. A., Carreira, L. H., Ljungdahl, L. G., Kuo, K. C. and Gehrke, C. W. (1985) 'DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N5methyladenine', *Nucleic Acids Research,* 13(4), 1399-1412.

Ekman, D., Light, S., Bjorklund, A. and Elofsson, A. (2006) 'What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*?', *Genome Biology,* 7(6), R45.

El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.-A., Jacobsen, S. E., Cooke, R. and Lagrange, T. (2007) 'Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components', *Genes & Development,* 21(20), 2539-2544.

Elbashir, S. M., Lendeckel, W. and Tuschl, T. (2001) 'RNA interference is mediated by 21- and 22-nucleotide RNAs', *Genes & Development,* 15(2), 188-200.

Elmayan, T., Proux, F. and Vaucheret, H. (2005) '*Arabidopsis* RPA2: A genetic link among transcriptional gene silencing, DNA repair, and DNA replication', *Current Biology,* 15(21), 1919-1925.

Elmayan, T. and Vaucheret, H. (1996) 'Expression of single copies of a strongly expressed 35S transgene can be silenced post-transcriptionally', *The Plant Journal,* 9(6), 787-797.

Erhard Jr., K. F., Stonaker, J. L., Parkinson, S. E., Lim, J. P., Hale, C. J. and Hollick, J. B. (2009) 'RNA Polymerase IV functions in paramutation in *Zea mays*', *Science,* 323(5918), 1201-1205.

Eulalio, A., Huntzinger, E. and Izaurralde, E. (2008) 'GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay', *Nature Structural Molecular Biology,* 15(4), 346-353.

Eun, C., Lorkovic, Z. J., Naumann, U., Long, Q., Havecker, E. R., Simon, S. A., Meyers, B. C., Matzke, A. J. and Matzke, M. (2011) 'AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in *Arabidopsis thaliana*', *Publci Library of Science One,* 6(10).

Exner, V., Aichinger, E., Shu, H., Wildhaber, T., Alfarano, P., Caflisch, A., Gruissem, W., Köhler, C. and Hennig, L. (2009) 'The chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 is essential for H3K27me3 binding and function during *Arabidopsis* development', *Public Library of Science ONE,* 4(4).

Fagard, M., Boutet, S., Morel, J.-B., Bellini, C. and Vaucheret, H. (2000) 'AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals', *Proceedings of the National Academy of Sciences of the United States of America,* 97(21), 11650-11654.

Fang, R. X., Nagy, F., Sivasubramaniam, S. and Chua, N. H. (1989) 'Multiple cis regulatory elements for maximal expression of the Cauliflower Mosaic Virus *35S* promoter in transgenic plants', *The Plant Cell,* 1(1), 141-150.

Fang, Y. and Spector, D. L. (2005) 'Centromere positioning and dynamics in living *Arabidopsis* plants', *Molecular Biology of the Cell,* 16(12), 5710-5718.

Finnegan, E. J. and Dennis, E. S. (1993) 'Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana', *Nucleic Acids Research,* 21(10), 2383-2388.

Finnegan, E. J., Liang, D. and Wang, M.-B. (2011) 'Self-incompatibility: Smi silences through a novel sRNA pathway', *Trends in Plant Science,* 16(5), 238-241.

Finnegan, E. J., Margis, R. and Waterhouse, P. M. (2003) 'Posttranscriptional gene silencing is not compromised in the *Arabidopsis CARPEL FACTORY* (*DICER-LIKE1*) mutant, a homolog of *Dicer-1* from Drosophila', *Current Biology,* 13(3), 236-240.

Finnegan, E. J., Peacock, W. J. and Dennis, E. S. (1996) 'Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development', *Proceedings of the National Academy of Sciences of the United States of America,* 93(16), 8449-8454.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998) 'Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*', *Nature,* 391(6669), 806-811.

Flaus, A., Martin, D. M. A., Barton, G. J. and Owen-Hughes, T. (2006) 'Identification of multiple distinct Snf2 subfamilies with conserved structural motifs', *Nucleic Acids Research,* 34(10), 2887-2905.

Foss, E., Lande, R., Stahl, F. W. and Steinberg, C. M. (1993) 'Chiasma interference as a function of genetic distance', *Genetics,* 133(3), 681-91.

Fransz, P., Armstrong, S., Alonso-blanco, C., Fischer, T. C., Torres-ruiz, R. A. and Jones, G. (1998) 'Cytogenetics for the model system *Arabidopsis thaliana*', *The Plant Journal,* 13(6), 867-876.

Fujimoto, S., Yonemura, M., Matsunaga, S., Nakagawa, T., Uchiyama, S. and Fukui, K. (2005) 'Characterization and dynamic analysis of *Arabidopsis* condensin subunits, AtCAP-H and AtCAP-H2', *Planta,* 222(2), 293-300.

Gao, Z., Liu, H.-L., Daxinger, L., Pontes, O., He, X., Qian, W., Lin, H., Xie, M., Lorkovic, Z. J., Zhang, S., Miki, D., Zhan, X., Pontier, D., Lagrange, T., Jin, H., Matzke, A. J. M., Matzke, M., Pikaard, C. S. and Zhu, J.-K. (2010) 'An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation', *Nature,* 465(7294), 106-109.

Garcia, D., Garcia, S., Pontier, D., Marchais, A., Renou, J. P., Lagrange, T. and Voinnet, O. (2012) 'Ago hook and RNA helicase motifs underpin dual roles for SDE3 in antiviral defense and silencing of nonconserved intergenic regions', *Molecular Cell,* 48(1), 109-120.

Gehring, M., Bubb, K. L. and Henikoff, S. (2009) 'Extensive demethylation of repetitive elements during seed development underlies gene imprinting', *Science,* 324(5933), 1447-1451.

Gehring, M., Huh, J. H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J. J., Goldberg, R. B. and Fischer, R. L. (2006) 'DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation', *Cell,* 124(3), 495-506.

Genger, R. K., Kovac, K. A., Dennis, E. S., Peacock, W. J. and Finnegan, E. J. (1999) 'Multiple DNA methyltransferase genes in *Arabidopsis thaliana*', *Plant Molecular Biology,* 41(2), 269-278.

Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldán-Arjona, T., David, L. and Zhu, J.-K. (2002) 'ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase', *Cell,* 111(6), 803-814.

Greenberg, M. V. C., Ausin, I., Chan, S. W. L., Cokus, S. J., Cuperus, J. T., Feng, S., Law, J. A., Chu, C., Pellegrini, M., Carrington, J. C. and Jacobsen, S. E. (2011) 'Identification of genes required for de novo DNA methylation in *Arabidopsis*', *Epigenetics,* 6(3), 344-354.

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., Enns, L. C., Burtner, C., Johnson, J. E., Odden, A. R., Comai, L. and Henikoff, S. (2003) 'Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*', *Genetics,* 164(2), 731-740.

Groszmann, M., Greaves, I. K., Albertyn, Z. I., Scofield, G. N., Peacock, W. J. and Dennis, E. S. (2011) 'Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor', *Proceedings of the National Academy of Sciences of the United States of America,* 108(6), 2617-2622.

Gruenbaum, Y., Cedar, H. and Razin, A. (1982) 'Substrate and sequence specificity of a eukaryotic DNA methylase', *Nature,* 295(5850), 620-622.

Gruenbaum, Y., Naveh-Many, T., Cedar, H. and Razin, A. (1981) 'Sequence specificity of methylation in higher plant DNA', *Nature,* 292(5826), 860-862.

Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H. and Siomi, M. C. (2007) 'A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*', *Science,* 315(5818), 1587-1590.

Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L. and Luo, J. (2005) 'DATF: a database of Arabidopsis transcription factors', *Bioinformatics,* 21(10), 2568-2569.

Gyorgy, H. (2005) 'Small RNA asymmetry in RNAi: Function in RISC assembly and gene regulation', *Federation of European Biochemical Societies Letters,* 579(26), 5850-5857.

Habu, Y., Mathieu, O., Tariq, M., Probst, A. V., Smathajitt, C., Zhu, T. and Paszkowski, J. (2006) 'Epigenetic regulation of transcription in intermediate heterochromatin', *European Molecular Biology Organizations Reports,* 7(12), 1279-1284.

Hamilton, A., Voinnet, O., Chappell, L. and Baulcombe, D. (2002) 'Two classes of short interfering RNA in RNA silencing', *European Molecular Biology Organization Journal,* 21(17), 4671-4679.

Hamilton, A. J. and Baulcombe, D. C. (1999) 'A species of small antisense RNA in posttranscriptional gene silencing in plants', *Science,* 286(5441), 950-952.

Hamilton, A. J., Brown, S., Yuanhai, H., Ishizuka, M., Lowe, A., Solis, A.-G. A. and Grierson, D. (1998) 'A transgene with repeated DNA causes high frequency, post-transcriptional suppression of ACC-oxidase gene expression in tomato', *The Plant Journal,* 15(6), 737-746.

Hammond, S. M., Bernstein, E., Beach, D. and Hannon, G. J. (2000) 'An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells', *Nature,* 404(6775), 293-296.

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. and Hannon, G. J. (2001) 'Argonaute2, a link between genetic and biochemical analyses of RNAi', *Science,* 293(5532), 1146-1150.

Haseloff, J., Siemering, K. R., Prasher, D. C. and Hodge, S. (1997) 'Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly', *Proceedings of the National Academy of Sciences of the United States of America,* 94(6), 2122-2127.

Havecker, E. R., Wallbridge, L. M., Hardcastle, T. J., Bush, M. S., Kelly, K. A., Dunn, R. M., Schwach, F., Doonan, J. H. and Baulcombe, D. C. (2010) 'The *Arabidopsis* RNA-directed DNA Methylation argonautes functionally diverge based on their expression and interaction with target loci', *Plant Cell,* 22(2), 321-334.

He, X.-J., Hsu, Y.-F., Pontes, O., Zhu, J., Lu, J., Bressan, R. A., Pikaard, C., Wang, C.-S. and Zhu, J.-K. (2009a) 'NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is

required for RNA-directed DNA methylation', *Genes & Development,* 23(3), 318-330.

He, X.-J., Hsu, Y.-F., Zhu, S., Liu, H.-L., Pontes, O., Zhu, J., Cui, X., Wang, C.-S. and Zhu, J.-K. (2009b) 'A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development', *Genes & Development,* 23(23), 2717-2722.

He, X.-J., Hsu, Y.-F., Zhu, S., Wierzbicki, A. T., Pontes, O., Pikaard, C. S., Liu, H.-L., Wang, C.-S., Jin, H. and Zhu, J.-K. (2009c) 'An effector of RNA-directed DNA Methylation in *Arabidopsis* is an ARGONAUTE 4- and RNA-binding protein', *Cell,* 137(3), 498-508.

Henderson, I. R., Deleris, A., Wong, W., Zhong, X., Chin, H. G., Horwitz, G. A., Kelly, K. A., Pradhan, S. and Jacobsen, S. E. (2010) 'The *de novo* cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA–directed DNA Methylation in *Arabidopsis thaliana*', *Public Library of Science Genetics,* 6(10).

Henderson, I. R. and Jacobsen, S. E. (2008) 'Tandem repeats upstream of the *Arabidopsis* endogene *SDC* recruit non-CG DNA methylation and initiate siRNA spreading', *Genes & Development,* 22(12), 1597-1606.

Hendrich, B. and Bird, A. (1998) 'Identification and characterization of a family of mammalian methyl-CpG binding proteins', *Molecular and Cellular Biology,* 18(11), 6538-6547.

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

Henz, S. R., Cumbie, J. S., Kasschau, K. D., Lohmann, J. U., Carrington, J. C., Weigel, D. and Schmid, M. (2007) 'Distinct expression patterns of natural antisense transcripts in *Arabidopsis*', *Plant Physiology,* 144(3), 1247-1255.

Herr, A. J., Jensen, M. B., Dalmay, T. and Baulcombe, D. C. (2005) 'RNA Polymerase IV directs silencing of endogenous DNA', *Science,* 308(5718), 118-120.

Hetzl, J., Foerster, A. M., Raidl, G. and Mittelsten Scheid, O. (2007) 'CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing', *Plant Journal,* 51(3), 526-36.

Hirano, T. (2002) 'The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair', *Genes & Development,* 16(4), 399-414.

Hirano, T. (2005) 'SMC proteins and chromosome mechanics: from bacteria to humans', *Philosophical Transactions of the Royal Society Biology: Biological Sciences,* 360(1455), 507-514.

Hirano, T., Kobayashi, R. and Hirano, M. (1997) 'Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a *Xenopus* homolog of the *Drosophila* Barren protein', *Cell,* 89(4), 511-521.

Hirschhorn, J. N., Brown, S. A., Clark, C. D. and Winston, F. (1992) 'Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure', *Genes & Development,* 6(12a), 2288-2298.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W. and Zimmermann, P. (2008) 'Genevestigator v3: A reference expression database for the meta-analysis of transcriptomes', *Advances in Bioinformatics,* 2008.

Hsieh, T.-F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L. and Zilberman, D. (2009) 'Genome-wide demethylation of *Arabidopsis* endosperm', *Science,* 324(5933), 1451-1454.

Huang, L., Jones, A. M. E., Searle, I., Patel, K., Vogler, H., Hubner, N. C. and Baulcombe, D. C. (2009) 'An atypical RNA polymerase involved in RNA

silencing shares small subunits with RNA polymerase II', *Nature Strucutal Molecular Biology,* 16(1), 91-93.

Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A. J. M. and Matzke, M. (2006) 'Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*', *European Molecular Biology Organization Journal,* 25(12), 2828-2836.

Huettel, B., Kanno, T., Daxinger, L., Bucher, E., van der Winden, J., Matzke, A. J. M. and Matzke, M. (2007) 'RNA-directed DNA methylation mediated by DRD1 and Pol IVb: A versatile pathway for transcriptional gene silencing in plants', *Biochimica Et Biophysica Acta-Gene Structure and Expression,* 1769(5-6), 358-374.

Huh, J. H., Bauer, M. J., Hsieh, T.-F. and Fischer, R. L. (2008) 'Cellular programming of plant gene imprinting', *Cell,* 132(5), 735-744.

Huson, D., Richter, D., Rausch, C., Dezulian, T., Franz, M. and Rupp, R. (2007) 'Dendroscope: An interactive viewer for large phylogenetic trees', *BioMed Central Bioinformatics,* 8(1), 460.

Hutvagner, G. and Simard, M. J. (2008) 'Argonaute proteins: key players in RNA silencing', *Nature Reviews Molecular Cell Biology,* 9(1), 22-32.

Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., Konagaya, A. and Shinozaki, K. (2005) 'RARTF: Database and tools for complete sets of *Arabidopsis* transcription factors', *DNA Research,* 12(4), 247-256.

Ikeda, Y., Kinoshita, Y., Susaki, D., Ikeda, Y., Iwano, M., Takayama, S., Higashiyama, T., Kakutani, T. and Kinoshita, T. (2011) 'HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in *Arabidopsis*', *Developmental Cell,* 21(3), 589-596.

Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M. C., Matsumoto-Yokoyama, E., Mitsuhara, I., Meshi, T. and Ishikawa, M. (2010) 'In vitro assembly of plant

RNA-induced Silencing Complexes facilitated by molecular chaperone HSP90', *Molecular Cell,* 39(2), 282-291.

Ingelbrecht, I., Van Houdt, H., Van Montagu, M. and Depicker, A. (1994) 'Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation', *Proceedings of the National Academy of Sciences of the United States of America,* 91(22), 10502-10506.

Inoue, N., Hess, K. D., Moreadith, R. W., Richardson, L. L., Handel, M. A., Watson, M. L. and Zinn, A. R. (1999) 'New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis', *Human Molecular Genetics,* 8(7), 1201-1207.

Irvine, D. V., Zaratiegui, M., Tolia, N. H., Goto, D. B., Chitwood, D. H., Vaughn, M. W., Joshua-Tor, L. and Martienssen, R. A. (2006) 'Argonaute slicing is required for heterochromatic silencing and spreading', *Science,* 313(5790), 1134-1137.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. and Paszkowski, J. (2011) 'An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress', *Nature,* 472(7341), 115-119.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T. and Tomari, Y. (2010) 'Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes', *Molecular Cell,* 39(2), 292-299.

Iyer, L., Abhiman, S. and Aravind, L. (2008) 'MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases', *Biology Direct,* 3(8).

Jackson, J. P., Lindroth, A. M., Cao, X. and Jacobsen, S. E. (2002) 'Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase', *Nature,* 416(6880), 556-560.

Jacobsen, S. E. and Meyerowitz, E. M. (1997) 'Hypermethylated SUPERMAN epigenetic alleles in *Arabidopsis*', *Science,* 277(5329), 1100-1103.

Jeddeloh, J. A., Stokes, T. L. and Richards, E. J. (1999) 'Maintenance of genomic methylation requires a SWI2/SNF2-like protein', *Nature Genetics,* 22(1), 94-97.

Jin, H., Vacic, V., Girke, T., Lonardi, S. and Zhu, J. K. (2008) 'Small RNAs and the regulation of cis-natural antisense transcripts in *Arabidopsis*', *BioMed Central Molecular Biology,* 9(6).

Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J. and Jacobsen, S. E. (2007) 'The SRA methyl-cytosine-binding domain links DNA and histone methylation', *Current Biology,* 17(4), 379-384.

Johnson, L. M., Law, J. A., Khattar, A., Henderson, I. R. and Jacobsen, S. E. (2008) 'SRA-domain proteins required for DRM2-mediated *de novo* DNA methylation', *Public Library of Science Genetics,* 4(11).

Johnson, T. B. and Coghill, R. D. (1925) 'Researches on pyrimidines. C111. The Discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the Tubercule *Bascillus*', *Journal of the American Chemical Society,* 47(11), 2838-2844.

Jones, A. L., Thomas, C. L. and Maule, A. J. (1998) 'De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus', *European Molecular Biology Organization Journal,* 17(21), 6385-6393.

Jones, L., Ratcliff, F. and Baulcombe, D. C. (2001) 'RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance', *Current Biology,* 11(10), 747-757.

Kang, H.-G., Kuhl, J. C., Kachroo, P. and Klessig, D. F. (2008) 'CRT1, an *Arabidopsis* ATPase that interacts with diverse resistance proteins and modulates disease resistance to Turnip Crinkle Virus', *Cell Host & Microbe,* 3(1), 48-57.

Kang, H.-G., Oh, C.-S., Sato, M., Katagiri, F., Glazebrook, J., Takahashi, H., Kachroo, P., Martin, G. B. and Klessig, D. F. (2010) 'Endosome-associated CRT1 functions early in resistance gene–mediated defense signaling in *Arabidopsis* and tobacco', *The Plant Cell,* 22(3), 918-936.

Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., Riddle, N. C., Verbsky, M. L. and Richards, E. J. (2003) '*Arabidopsis* MET1 cytosine methyltransferase mutants', *Genetics,* 163(3), 1109-1122.

Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Kreil, D. P., Matzke, M. and Matzke, A. J. M. (2008) 'A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation', *Nature Genetics,* 40(5), 670-675.

Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Kreil, D. P., Breinig, F., Lind, M., Schmitt, M. J., Simon, S. A., Gurazada, S. G. R., Meyers, B. C., Lorkovic, Z. J., Matzke, A. J. M. and Matzke, M. (2009) 'RNA-directed DNA methylation and plant development require an IWR1-type transcription factor', *European Molecular Biology Organization Reports,* 11(1), 65-71.

Kanno, T., Huettel, B., Mette, M. F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D. P., Matzke, M. and Matzke, A. J. M. (2005) 'Atypical RNA polymerase subunits required for RNA-directed DNA methylation', *Nature Genetics,* 37(7), 761-765.

Kanno, T., Mette, M. F., Kreil, D. P., Aufsatz, W., Matzke, M. and Matzke, A. J. M. (2004) 'Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation', *Current Biology,* 14(9), 801-805.

Kass, S. U., Goddard, J. P. and Adams, R. L. (1993) 'Inactive chromatin spreads from a focus of methylation', *Molecular and Cellular Biology,* 13(12), 7372-7379.

Kass, S. U., Landsberger, N. and Wolffe, A. P. (1997) 'DNA methylation directs a time-dependent repression of transcription initiation', *Current biology,* 7(3), 157-165.

Kassabov, S. R., Zhang, B., Persinger, J. and Bartholomew, B. (2003) 'SWI/SNF unwraps, slides, and rewraps the nucleosome', *Molecular cell,* 11(2), 391-403.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J.-K., Staskawicz, B. J. and Jin, H. (2006) 'A pathogen-inducible endogenous siRNA in plant immunity', *Proceedings of the National Academy of Sciences of the United States of America,* 103(47), 18002-18007.

Kato, M., Miura, A., Bender, J., Jacobsen, S. E. and Kakutani, T. (2003) 'Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*', *Current Biology,* 13(5), 421-426.

Kazama, T., Ichihashi, Y., Murata, S. and Tsukaya, H. (2010) 'The mechanism of cell cycle arrest front progression explained by a KLUH/CYP78A5-dependent mobile growth factor in developing leaves of *Arabidopsis thaliana*', *Plant and Cell Physiology,* 51(6), 1046-1054.

Kelley, L. A. and Sternberg, M. J. E. (2009) 'Protein structure prediction on the web: a case study using the Phyre server', *Nature Protocols,* 4(3), 363-371.

Kennerdell, J. R. and Carthew, R. W. (1998) 'Use of dsRNA-mediated genetic interference to demonstrate that Frizzled and Frizzled 2 act in the Wingless pathway', *Cell,* 95(7), 1017-1026.

Kerschen, A., Napoli, C. A., Jorgensen, R. A. and Müller, A. E. (2004) 'Effectiveness of RNA interference in transgenic plants', *Federation of European Biochemical Societies Letters,* 566(1–3), 223-228.

Khraiwesh, B., Arif, M. A., Seumel, G. I., Ossowski, S., Weigel, D., Reski, R. and Frank, W. (2010) 'Transcriptional control of gene Eexpression by microRNAs', *Cell,* 140(1), 111-122.

Khvorova, A., Reynolds, A. and Jayasena, S. D. (2003) 'Functional siRNAs and miRNAs exhibit strand bias', *Cell,* 115(2), 209-216.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Fischer, R. L. and Kakutani, T. (2004) 'One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation', *Science,* 303(5657), 521-523.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J. J., Goldberg, R. B. and Fischer, R. L. (1999) 'Control of fertilization-independent endosperm development by the MEDEA polycomb gene in *Arabidopsis*', *Proceedings of the National Academy of Sciences of the United States of America,* 96(7), 4186-4191.

Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P. and Weisshaar, B. (2012) 'GABI-Kat SimpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database', *Nucleic Acids Res,* 40(D1), D1211-D1215.

Kohler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U. and Gruissem, W. (2003) '*Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development', *European Molecular Biology Organization Journal,* 22(18), 4804-4814.

Konieczny, A. and Ausubel, F. M. (1993) 'A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers', *Plant Journal,* 4(2), 403-10.

Koornneef, M., Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C. J. and Peeters, A. J. M. (1998a) 'Genetic interactions among late-flowering mutants of *Arabidopsis*', *Genetics,* 148(2), 885-892.

Koornneef, M., Alonso-Blanco, C., Peeters, A. J. M. and Soppe, W. (1998b) 'Genetic control of flowering time in *Arabidopsis*', *Annual Review of Plant Physiology and Plant Molecular Biology,* 49, 345-370.

Koornneef, M., Hanhart, C., van Loenen-Martinet, P. and Blankestijn de Vries, H. (1995) 'The effect of daylength on the transition to flowering in phytochrome-deficient, late-flowering and double mutants of *Arabidopsis thaliana*', *Physiologia Plantarum,* 95(2), 260-266.

Koornneef, M., Hanhart, C. J. and Veen, J. H. (1991) 'A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*', *Molecular and General Genetics,* 229(1), 57-66.

Kovalchuk, I., Kovalchuk, O. and Hohn, B. (2000) 'Genome-wide variation of the somatic mutation frequency in transgenic plants', *European Molecular Biology Organization Journal,* 19(17), 4431-4438.

Krieg, D. R. (1963) 'Ethyl methanesulfonate-induced reversion of bacteriophage TRrII mutants', *Genetics,* 48(4), 561-580.

Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrín-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H. Y., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A. and Greenblatt, J. F. (2006) 'Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*', *Nature,* 440(7084), 637-643.

Krysan, P. J., Young, J. C. and Sussman, M. R. (1999) 'T-DNA as an insertional mutagen in *Arabidopsis*', *The Plant Cell,* 11(12), 2283-2290.

Kumagai, M. H., Donson, J., della-Cioppa, G., Harvey, D., Hanley, K. and Grill, L. K. (1995) 'Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA', *Proceedings of the National Academy of Sciences of the United States of America,* 92(5), 1679-1683.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T. W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H. and Nakano, T. (2008) 'DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes', *Genes & Development,* 22(7), 908-917.

Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T. and Lagrange, T. (2009) 'PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit', *Proceedings of the National Academy of Sciences of the United States of America,* 106(3), 941-946.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics,* 23(21), 2947-2948.

Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J. U., Rätsch, G. and Weigel, D. (2008) 'Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*', *Proceedings of the National Academy of Sciences of the United States of America,* 105(25), 8795-8800.

Law, J. A., Ausin, I., Johnson, L. M., Vashisht, A. A., Zhu, J.-K., Wohlschlegel, J. A. and Jacobsen, S. E. (2010) 'A protein complex required for Polymerase V transcripts and RNA-directed DNA Methylation in *Arabidopsis*', *Current Biology,* 20(10), 951-956.

Law, J. A., Vashisht, A. A., Wohlschlegel, J. A. and Jacobsen, S. E. (2011) 'SHH1, a homeodomain proteinrequired for DNA methylation, as well As RDR2, RDM4, and chromatin remodeling factors, associate with RNA Polymerase IV', *Public Library of Science Genetics,* 7(7).

Lee, T.-F., Gurazada, S. G. R., Zhai, J., Li, S., Simon, S. A., Matzke, M. A., Chen, X. and Meyers, B. C. (2012) 'RNA polymerase V-dependent small RNAs in *Arabidopsis* originate from small, intergenic loci including most SINE repeats', *Epigenetics,* 7(7), 781-795.

Letunic, I., Doerks, T. and Bork, P. (2012) 'SMART 7: recent updates to the protein domain annotation resource', *Nucleic Acids Research,* 40(D1), D302-D305.

Li, C. F., Pontes, O., El-Shami, M., Henderson, I. R., Bernatavichute, Y. V., Chan, S. W. L., Lagrange, T., Pikaard, C. S. and Jacobsen, S. E. (2006) 'An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*', *Cell,* 126(1), 93-106.

Li, G. and Reinberg, D. (2011) 'Chromatin higher-order structures and gene regulation', *Current Opinion in Genetics & Development,* 21(2), 175-186.

Liggins, A. P., Cooper, C. D. O., Lawrie, C. H., Brown, P. J., Collins, G. P., Hatton, C. S., Pulford, K. and Banham, A. H. (2007) 'MORC4, a novel member of the MORC family, is highly expressed in a subset of diffuse large B-cell lymphomas', *British Journal of Haematology,* 138(4), 479-486.

Lindbo, J. A., Silva-Rosales, L., Proebsting, W. M. and Dougherty, W. G. (1993) 'Induction of a highly specific antiviral state in transgenic plants: Implications for regulation of gene expression and virus resistance', *The Plant Cell,* 5(12), 1749-1759.

Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S. and Jacobsen, S. E. (2001) 'Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation', *Science,* 292(5524), 2077-2080.

Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., Jenuwein, T., Khorasanizadeh, S. and Jacobsen, S. E. (2004) 'Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3', *European Molecular Biology Organization Journal,* 23(21), 4146-4155.

Lingel, A., Simon, B., Izaurralde, E. and Sattler, M. (2004) 'Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain', *Nature Structural Molecular Biology,* 11(6), 576-577.

Lippman, Z., May, B., Yordan, C., Singer, T. and R., M. (2003) 'Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification', *Public Library of Science Biology,* 1(3).

Lisch, D. (2009) 'Epigenetic regulation of transposable elements in plants', *Annual Review of Plant Biology,* 60, 43-66.

Lisch, D. and Bennetzen, J. L. (2011) 'Transposable element origins of epigenetic gene regulation', *Current Opinion in Plant Biology,* 14(2), 156-161.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H. and Ecker, J. R. (2008) 'Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*', *Cell,* 133(3), 523-536.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B. and Ecker, J. R. (2009) 'Human DNA methylomes at base resolution show widespread epigenomic differences', *Nature,* 462(7271), 315-322.

Liu, J., Bai, G., Zhang, C., Chen, W., Zhou, J., Zhang, S., Chen, Q., Deng, X., He, X. J. and Zhu, J. K. (2011) 'An atypical component of RNA-directed DNA methylation machinery has both DNA methylation-dependent and -independent roles in locus-specific transcriptional gene silencing', *Cell Research,* 21(12), 1691-700.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J.-J., Hammond, S. M., Joshua-Tor, L. and Hannon, G. J. (2004) 'Argonaute 2 is the catalytic engine of mammalian RNAi', *Science,* 305(5689), 1437-1441.

Llave, C., Kasschau, K. D., Rector, M. A. and Carrington, J. C. (2002) 'Endogenous and silencing-associated small RNAs in plants', *The Plant Cell,* 14(7), 1605-1619.

Lohmann, J. U., Endl, I. and Bosch, T. C. G. (1999) 'Silencing of developmental genes in hydra', *Developmental Biology,* 214(1), 211-214.

Lorković, Z. J., Naumann, U., Matzke, A. J. M. and Matzke, M. (2012) 'Involvement of a GHKL ATPase in RNA-directed DNA Methylation in *Arabidopsis thaliana*', *Current Biology,* 22(10), 933-938.

Lu, C., Meyers, B. C. and Green, P. J. (2007) 'Construction of small RNA cDNA libraries for deep sequencing', *Methods,* 43(2), 110-117.

Lu, J., Zhang, C., Baulcombe, D. C. and Chen, Z. J. (2012) 'Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds', *Proceedings of the National Academy of Sciences of the United States of America,* 109(14), 5529-5534.

Luo, J. and Hall, B. D. (2007) 'A multistep process gave rise to RNA polymerase IV of land plants', *Journal of Molecular Evoultion,* 64(1), 101 - 112.

López, A., Ramírez, V., García-Andrade, J., Flors, V. and Vera, P. (2011) 'The RNA silencing enzyme RNA Polymerase V is required for plant immunity', *Public Library of Science Genetics,* 7(12).

Ma, J.-B., Yuan, Y.-R., Meister, G., Pei, Y., Tuschl, T. and Patel, D. J. (2005) 'Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein', *Nature,* 434(7033), 666-670.

Malagnac, F., Bartee, L. and Bender, J. (2002) 'An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation', *European Molecular Biology Organization Journal,* 21(24), 6842-6852.

Marcussen, T., Oxelman, B., Skog, A. and Jakobsen, K. (2010) 'Evolution of plant RNA polymerase IV/V genes: Evidence of subneofunctionalization of duplicated NRPD2/NRPE2-like paralogs in Viola (*Violaceae*)', *BioMed Central Evolutionary Biology,* 10(1), 45.

Mason, J. M. and Arndt, K. M. (2004) 'Coiled coil domains: Stability, specificity, and biological implications', *ChemBioChem,* 5(2), 170-176.

Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.-V., Colot, V., Schubert, I. and Tourmente, S. (2003) 'Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in *Arabidopsis*', *The Plant Cell,* 15(12), 2929-2939.

May, B. P., Lippman, Z. B., Fang, Y., Spector, D. L. and Martienssen, R. A. (2005) 'Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats', *Public Library of Science Genetics,* 1(6).

McCallum, C. M., Comai, L., Greene, E. A. and Henikoff, S. (2000) 'Targeted screening for induced mutations', *Nature Biotechnology,* 18(4), 455-457.

McElver, J., Tzafrir, I., Aux, G., Rogers, R., Ashby, C., Smith, K., Thomas, C., Schetter, A., Zhou, Q., Cushman, M. A., Tossberg, J., Nickle, T., Levin, J. Z., Law, M., Meinke, D. and Patton, D. (2001) 'Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*', *Genetics,* 159(4), 1751-1763.

Melnyk, C. W., Molnar, A., Bassett, A. and Baulcombe, D. C. (2011) 'Mobile 24 nt small RNAs direct transcriptional gene silencing in the root meristems of *Arabidopsis thaliana*', *Current Biology,* 21(19), 1678-1683.

Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. and Matzke, A. J. M. (2000) 'Transcriptional silencing and promoter methylation triggered by double-stranded RNA', *European Molecular Biology Organization Journal,* 19(19), 5194-5201.

Meyer, P., Niedenhof, I. and ten Lohuis, M. (1994) 'Evidence for cytosine methylation of non-symmetrical sequences in transgenic *Petunia hybrida*', *European Molecular Biology Organziation Journal,* 13(9), 2084-8.

Meyer, R. C., Törjék, O., Becher, M. and Altmann, T. (2004) 'Heterosis of biomass production in *Arabidopsis*. Establishment during early development', *Plant Physiology,* 134(4), 1813-1823.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G. J. and Qi, Y. (2008) 'Sorting of small RNAs into *Arabidopsis* Argonaute complexes is directed by the 5' terminal nucleotide', *Cell,* 133(1), 116-127.

Mimura, Y., Takahashi, K., Kawata, K., Akazawa, T. and Inoue, N. (2010) 'Two-step colocalization of MORC3 with PML nuclear bodies', *Journal of Cell Science,* 123(12), 2014-2024.

Mishina, T. E. and Zeier, J. (2006) 'The *Arabidopsis* flavin-dependent monooxygenase FMO1 is an essential component of biologically induced systemic acquired resistance', *Plant Physiology,* 141(4), 1666-1675.

Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H. and Kakutani, T. (2009) 'An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites', *European Molecular Biology Organization Journal,* 28(8), 1078-1086.

Moissiard, G., Cokus, S. J., Cary, J., Feng, S., Billi, A. C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B. R., McCord, R. P., Hale, C. J., Feng, W., Michaels, S. D., Frand, A. R., Pellegrini, M., Dekker, J., Kim, J. K. and Jacobsen, S. (2012) 'MORC family ATPases required for heterochromatin condensation and gene silencing', *Science,* 336(6087), 1448-1451.

Molnar, A., Melnyk, C. W., Bassett, A., Hardcastle, T. J., Dunn, R. and Baulcombe, D. C. (2010) 'Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells', *Science,* 328(5980), 872-875.

Morales-Ruiz, T., Ortega-Galisteo, A. P., Ponferrada-Marín, M. I., Martínez-Macías, M. I., Ariza, R. R. and Roldán-Arjona, T. (2006) 'DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases', *Proceedings of the National Academy of Sciences of the United States of America,* 103(18), 6853-6858.

Moreno, N., Bougourd, S., Haseloff, J. and Feijo, J. (2006) 'Imaging plant cells' in J., P., ed. *Handbook of Biological Confocal Microscopy*, New York: Springer Science and Business Media, 769-87.

Morgan, T. H. (1911a) 'Chromosomes and associative inheritance', *Science,* 34(880), 636-638.

Morgan, T. H. (1911b) 'Random segregation versus coupling in Mendelian inheritance', *Science,* 34(873), 384.

Mosher, R. A., Melnyk, C. W., Kelly, K. A., Dunn, R. M., Studholme, D. J. and Baulcombe, D. C. (2009) 'Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*', *Nature,* 460(7252), 283-286.

Mosher, R. A., Tan, E. H., Shin, J., Fischer, R. L., Pikaard, C. S. and Baulcombe, D. C. (2011) 'An atypical epigenetic mechanism affects uniparental expression of Pol IV-dependent siRNAs', *Public Library of Science One,* 6(10).

Mourrain, P., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., Jouette, D., Lacombe, A.-M., Nikic, S., Picault, N., Rémoué, K., Sanial, M., Vo, T.-A. and Vaucheret, H. (2000) '*Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance', *Cell,* 101(5), 533-542.

Mukherjee, K., Brocchieri, L. and Bürglin, T. R. (2009) 'A comprehensive classification and evolutionary analysis of plant homeobox genes', *Molecular Biology and Evolution,* 26(12), 2775-2794.

Mull, L., Ebbs, M. L. and Bender, J. (2006) 'A histone methylation-dependent DNA methylation pathway is uniquely impaired by deficiency in *Arabidopsis* S-Adenosylhomocysteine Hydrolase', *Genetics,* 174(3), 1161-1171.

Muller, H. (1930) 'Types of visible variations induced by X-rays in *Drosophila*', *Journal of Genetics,* 22(3), 299-334.

Murata, M., Heslop-Harrison, J. S. and Motoyoshi, F. (1997) 'Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multi-color fluorescence in situ hybridization with cosmid clones', *The Plant Journal,* 12(1), 31-37.

Mushegian, A. R., Bassett, D. E., Boguski, M. S., Bork, P. and Koonin, E. V. (1997) 'Positionally cloned human disease genes: Patterns of evolutionary conservation and functional motifs', *Proceedings of the National Academy of Sciences of the United States of America,* 94(11), 5831-5836.

Myouga, F., Tsuchimoto, S., Noma, K., Ohtsubo, H. and Ohtsubo, E. (2001) 'Identification and structural analysis of SINE elements in the *Arabidopsis thaliana* genome', *Genes & Genetic Systems,* 76(3), 169-179.

Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirosawa, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O. (1998) 'Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro', *DNA Research,* 5(6), 355-364.

Napoli, C., Lemieux, C. and Jorgensen, R. (1990) 'Introduction of a chimeric *Chalcone Synthase* gene into *Petunia* results in reversible co-suppression of homologous genes in trans', *The Plant Cell,* 2(4), 279-289.

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A.-C., Dorn, R., Jenuwein, T. and Reuter, G. (2005) 'Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*', *European Molecular Biology Organization Journal,* 24(7), 1418-1429.

Naumann, U., Daxinger, L., Kanno, T., Eun, C., Long, Q., Lorkovic, Z. J., Matzke, M. and Matzke, A. J. M. (2011) 'Genetic evidence that DNA methyltransferase DRM2 has a direct catalytic role in RNA-directed DNA Methylation in *Arabidopsis thaliana*', *Genetics,* 187(3), 977-979.

Ng, H. H., Xu, R.-M., Zhang, Y. and Struhl, K. (2002) 'Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 Lysine 79', *Journal of Biological Chemistry,* 277(38), 34655-34657.

Nishimura, T., Molinard, G., Petty, T. J., Broger, L., Gabus, C., Halazonetis, T. D., Thore, S. and Paszkowski, J. (2012) 'Structural basis of transcriptional gene silencing mediated by *Arabidopsis* MOM1', *Public Library of Science Genetics,* 8(2).

Numa, H., Kim, J.-M., Matsui, A., Kurihara, Y., Morosawa, T., Ishida, J., Mochizuki, Y., Kimura, H., Shinozaki, K., Toyoda, T., Seki, M., Yoshikawa, M. and Habu, Y. (2010) 'Transduction of RNA-directed DNA methylation signals to repressive histone marks in *Arabidopsis thaliana*', *European Molecular Biology Organization Journal,* 29(2), 352-362.

Nykänen, A., Haley, B. and Zamore, P. D. (2001) 'ATP requirements and small interfering RNA structure in the RNA interference pathway', *Cell,* 107(3), 309-321.

Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O. and Pikaard, C. S. (2005) 'Plant nuclear RNA Polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation', *Cell,* 120(5), 613-622.

Ortega-Galisteo, A., Morales-Ruiz, T., Ariza, R. and Roldán-Arjona, T. (2008) '*Arabidopsis* DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks', *Plant Molecular Biology,* 67(6), 671-681.

Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N. and Weigel, D. (2008) 'Sequencing of natural strains of *Arabidopsis thaliana* with short reads', *Genome Research,* 18(12), 2024-2033.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D. and Lynch, M. (2010) 'The rate and molecular

spectrum of spontaneous mutations in *Arabidopsis thaliana*', *Science,* 327(5961), 92-94.

Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Rao, M., Bhadra, U., Birchler, J. A. and Elgin, S. C. R. (2004) 'Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery', *Science,* 303(5658), 669-672.

Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002) 'CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein,act in microRNA metabolism in *Arabidopsis thaliana*', *Current biology,* 12(17), 1484-1495.

Parker, J. (2010) 'How to slice: snapshots of Argonaute in action', *Silence,* 1(1), 3.

Parkinson, S. E., Gross, S. M. and Hollick, J. B. (2007) 'Maize sex determination and abaxial leaf fates are canalized by a factor that maintains repressed epigenetic states', *Developmental Biology,* 308(2), 462-473.

Partridge, J. F., DeBeauchamp, J. L., Kosinski, A. M., Ulrich, D. L., Hadler, M. J. and Noffsinger, V. J. P. (2007) 'Functional separation of the requirements for establishment and maintenance of centromeric heterochromatin', *Molecular Cell,* 26(4), 593-602.

Penterman, J., Uzawa, R. and Fischer, R. L. (2007) 'Genetic interactions between DNA demethylation and methylation in *Arabidopsis*', *Plant Physiology,* 145(4), 1549-1557.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L. and Poethig, R. S. (2004) 'SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*', *Genes & Development,* 18(19), 2368-79.

Perry, J. and Zhao, Y. (2003) 'The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants', *Trends in Biochemical Sciences,* 28(11), 576-580.

Pikaard, C. S., Haag, J. R., Ream, T. and Wierzbicki, A. T. (2008) 'Roles of RNA Polymerase IV in gene silencing', *Trends in Plant Science,* 13(7), 390-397.

Pintor-Toro, J. A. (1987) 'Adenine methylation in zein genes', *Biochemical and Biophysical Research Communications,* 147(3), 1082-1087.

Pontes, O., Costa-Nunes, P., Vithayathil, P. and Pikaard, C. S. (2009) 'RNA Polymerase V functions in Arabidopsis interphase heterochromatin organization independently of the 24-nt siRNA-directed DNA methylation pathway', *Molecular Plant,* 2(4), 700-710.

Pontes, O., Li, C. F., Nunes, P. C., Haag, J., Ream, T., Vitins, A., Jacobsen, S. E. and Pikaard, C. S. (2006) 'The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center', *Cell,* 126(1), 79-92.

Pontier, D., Picart, C., Roudier, F., Garcia, D., Lahmy, S., Azevedo, J., Alart, E., Laudié, M., Karlowski, W. M., Cooke, R., Colot, V., Voinnet, O. and Lagrange, T. (2012) 'NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in *Arabidopsis*', *Molecular Cell,* 48(1), 121-132.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.-A., Lerbs-Mache, S., Colot, V. and Lagrange, T. (2005) 'Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*', *Genes & Development,* 19(17), 2030-2040.

Preuss, S. B., Costa-Nunes, P., Tucker, S., Pontes, O., Lawrence, R. J., Mosher, R., Kasschau, K. D., Carrington, J. C., Baulcombe, D. C., Viegas, W. and Pikaard, C. S. (2008) 'Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins', *Molecular Cell,* 32(5), 673-684.

Probst, A. V., Fagard, M., Proux, F., Mourrain, P., Boutet, S., Earley, K., Lawrence, R. J., Pikaard, C. S., Murfett, J., Furner, I., Vaucheret, H. and Scheid, O. M. (2004) '*Arabidopsis* histonedeacetylase HDA6 is required for maintenance

of transcriptional gene silencing and determines nuclear organization of rDNA repeats', *Plant Cell,* 16(4), 1021-1034.

Promega (2010) 'pGEM®-T Easy Vector Systems', http://www.promega.com/products/pcr/pcr-cloning/pgem_t-easy-vector-systems/ [last accessed: 18th December 2012].

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L. L., Eddy, S. R., Bateman, A. and Finn, R. D. (2012) 'The Pfam protein families database', *Nucleic Acids Research,* 40(D1), D290-D301.

Pélissier, T., Clavel, M., Chaparro, C., Pouch-Pélissier, M.-N., Vaucheret, H. and Deragon, J.-M. (2011) 'Double-stranded RNA binding proteins DRB2 and DRB4 have an antagonistic impact on polymerase IV-dependent siRNA levels in *Arabidopsis*', *RNA,* 17(8), 1502-1510.

Qi, Y., He, X., Wang, X.-J., Kohany, O., Jurka, J. and Hannon, G. J. (2006) 'Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation', *Nature,* 443(7114), 1008-1012.

Qian, W., Miki, D., Zhang, H., Liu, Y., Zhang, X., Tang, K., Kan, Y., La, H., Li, X., Li, S., Zhu, X., Shi, X., Zhang, K., Pontes, O., Chen, X., Liu, R., Gong, Z. and Zhu, J.-K. (2012) 'A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*', *Science,* 336(6087), 1445-1448.

Qu, F., Ye, X. H. and Morris, T. J. (2008) '*Arabidopsis* DRB4, AG01, AG07, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1', *Proceedings of the National Academy of Sciences of the United States of America,* 105(38), 14732-14737.

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P. and Jaenisch, R. (2000) 'Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a', *Proceedings of the National Academy of Sciences United States of America,* 97(10), 5237-5242.

Ratcliff, F. G., MacFarlane, S. A. and Baulcombe, D. C. (1999) 'Gene silencing without DNA: RNA-mediated cross-protection between viruses', *The Plant Cell,* 11(7), 1207-1215.

Ream, T. S., Haag, J. R., Wierzbicki, A. T., Nicora, C. D., Norbeck, A. D., Zhu, J.-K., Hagen, G., Guilfoyle, T. J., Pasa-Tolic, L. and Pikaard, C. S. (2009) 'Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA Polymerase II', *Molecular Cell,* 33(2), 192-203.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. and Bartel, D. P. (2002) 'MicroRNAs in plants', *Genes & Development,* 16(13), 1616-1626.

Reinhold, T., Alawady, A., Grimm, B., Beran, K. C., Jahns, P., Conrath, U., Bauer, J., Reiser, J., Melzer, M., Jeblick, W. and Neuhaus, H. E. (2007) 'Limitation of nocturnal import of ATP into *Arabidopsis* chloroplasts leads to photooxidative damage', *The Plant Journal,* 50(2), 293-304.

Reiser, J., Linka, N., Lemke, L., Jeblick, W. and Neuhaus, H. E. (2004) 'Molecular physiological analysis of the two plastidic ATP/ADP transporters from *Arabidopsis*', *Plant Physiology,* 136(3), 3524-3536.

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B. and Bartel, D. P. (2002) 'Prediction of plant microRNA targets', *Cell,* 110(4), 513-520.

Riano-Pachon, D., Ruzicic, S., Dreyer, I. and Mueller-Roeber, B. (2007) 'PlnTFDB: an integrative plant transcription factor database', *BioMed Central Bioinformatics,* 8(1), 42.

Rieseberg, M., Kasper, C., Reardon, K. F. and Scheper, T. (2001) 'Flow cytometry in biotechnology', *Applied Microbiology and Biotechnology,,* 56(3-4), 350-360.

Rigal, M., Kevei, Z., Pelissier, T. and Mathieu, O. (2012) 'DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin

modification patterns', *European Molecular Biology Organization Journal,* 31(13), 2981-2993.

Rivas, F. V., Tolia, N. H., Song, J.-J., Aragon, J. P., Liu, J., Hannon, G. J. and Joshua-Tor, L. (2005) 'Purified Argonaute 2 and an siRNA form recombinant human RISC', *Nature Structural Molecular Biology,* 12(4), 340-349.

Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D. and Whitelaw, E. (1995) 'Position-dependent variegation of globin transgene expression in mice', *Proceedings of the National Academy of Sciences of the United States of America,* 92(12), 5371-5375.

Rogers, J. C. and Rogers, S. W. (1995) 'Comparison of the effects of N6-methyldeoxyadenosine and N5-methyldeoxycytosine on transcription from nuclear gene promoters in barley', *The Plant Journal,* 7(2), 221-233.

Ronemus, M. J., Galbiati, M., Ticknor, C., Chen, J. and Dellaporta, S. L. (1996) 'Demethylation-induced developmental pleiotropy in *Arabidopsis*', *Science,* 273(5275), 654-657.

Rose, T. M., Schultz, E. R., Henikoff, J. G., Pietrokovski, S., McCallum, C. M. and Henikoff, S. (1998) 'Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences', *Nucleic Acids Research,* 26(7), 1628-1635.

Rowley, M. J., Avrutsky, M. I., Sifuentes, C. J., Pereira, L. and Wierzbicki, A. T. (2011) 'Independent chromatin binding of ARGONAUTE 4 and SPT5L/KTF1 mediates transcriptional gene silencing', *Public Library of Science Genetics,* 7(6), e1002120.

Ruiz, M. T., Voinnet, O. and Baulcombe, D. C. (1998) 'Initiation and maintenance of virus-induced gene silencing', *The Plant Cell Online,* 10(6), 937-946.

Schauer, S. E., Jacobsen, S. E., Meinke, D. W. and Ray, A. (2002) 'DICER-LIKE1: blind men and elephants in *Arabidopsis* development', *Trends in Plant Science,* 7(11), 487-491.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., Jorgensen, J.-E., Weigel, D. and Andersen, S. U. (2009) 'SHOREmap: simultaneous mapping and mutation identification by deep sequencing', *Nature Methods,* 6(8), 550-551.

Schneeberger, K., Ossowski, S., Ott, F., Klein, J. D., Wang, X., Lanz, C., Smith, L. M., Cao, J., Fitz, J., Warthmann, N., Henz, S. R., Huson, D. H. and Weigel, D. (2011) 'Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes', *Proceedings of the National Academy of Sciences of the United States of America,* 108(25), 10249-10254.

Schneeberger, K. and Weigel, D. (2011) 'Fast-forward genetics enabled by new sequencing technologies', *Trends in Plant Science,* 16(5), 282-288.

Schwarz, D. S., Hutvágner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P. D. (2003) 'Asymmetry in the assembly of the RNAi enzyme complex', *Cell,* 115(2), 199-208.

Searle, I. R., Pontes, O., Melnyk, C. W., Smith, L. M. and Baulcombe, D. C. (2010) 'JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in *Arabidopsis*', *Genes & Development,* 24(10), 986-991.

Shao, Y., Li, Y., Zhang, J., Liu, D., Liu, F., Zhao, Y., Shen, T. and Li, F. (2010) 'Involvement of histone deacetylation in MORC2-mediated down-regulation of carbonic anhydrase IX', *Nucleic Acids Research,* 38(9), 2813-2824.

Shapiro, R., Servis, R. E. and Welcher, M. (1970) 'Reactions of uracil and cytosine derivatives with sodium bisulfite', *Journal of the American Chemical Society,* 92(2), 422-424.

Shirasu, K., Schulman, A. H., Lahaye, T. and Schulze-Lefert, P. (2000) 'A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion', *Genome Research,* 10(7), 908-915.

Simpson, V. J., Johnson, T. E. and Hammen, R. F. (1986) '*Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging', *Nucleic Acids Research,* 14(16), 6711-6719.

Sims, R. J., Belotserkovskaya, R. and Reinberg, D. (2004) 'Elongation by RNA polymerase II: The short and long of it', *Genes & Development,* 18(20), 2437-2468.

Singer, T., Yordan, C. and Martienssen, R. A. (2001) 'Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1)', *Genes & Development,* 15(5), 591-602.

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

Slater, A., Scott, N. and Fowler, M. (2003) 'The 35S promoter' in *Plant Biotechnology: The Genetic Manipulation of Plants*, Oxford, Uk: Oxford University Press, 82-83.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J. D., Feijó, J. A. and Martienssen, R. A. (2009) 'Epigenetic reprogramming and small RNA silencing of transposable elements in pollen', *Cell,* 136(3), 461-472.

Smith, L. M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F. K., Herr, A. J., Pikaard, C. S. and Baulcombe, D. C. (2007) 'An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis*', *Plant Cell,* 19(5), 1507-1521.

Smith, S. S., Kaplan, B. E., Sowers, L. C. and Newman, E. M. (1992) 'Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation', *Proceedings National Academy of Sciences of the United States of America,* 89(10), 4744-4748.

Solovei, I., Grandi, N., Knoth, R., Volk, B. and Cremer, T. (2004) 'Positional changes of pericentromeric heterochromatin and nucleoli in postmitotic Purkinje cells during murine cerebellum development', *Cytogenetic and Genome Research,* 105(2-4), 302-310.

Song, J.-J., Smith, S. K., Hannon, G. J. and Joshua-Tor, L. (2004) 'Crystal structure of Argonaute and its implications for RISC slicer activity', *Science,* 305(5689), 1434-1437.

Soppe, W. J. J., Jacobsen, S. E., Alonso-Blanco, C., Jackson, J. P., Kakutani, T., Koornneef, M. and Peeters, A. J. M. (2000) 'The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene', *Molecular Cell,* 6(4), 791-802.

Soppe, W. J. J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M. S., Jacobsen, S. E., Schubert, I. and Fransz, P. F. (2002) 'DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*', *European Molecular Biology Organization Journal,* 21(23), 6549-6559.

Sridhar, V. V., Kapoor, A., Zhang, K., Zhu, J., Zhou, T., Hasegawa, P. M., Bressan, R. A. and Zhu, J.-K. (2007) 'Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination', *Nature,* 447(7145), 735-738.

Sturtevant, A. H. (1913) 'The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association', *Journal of Experimental Zoology,* 14(1), 43-59.

Sugimoto, K., Takeda, S. and Hirochika, H. (2003) 'Transcriptional activation mediated by binding of a plant GATA-type zinc finger protein AGP1 to the AG-motif (AGATCCAA) of the wound-inducible Myb gene NtMyb2', *The Plant Journal,* 36(4), 550-564.

Sugiyama, T., Cam, H., Verdel, A., Moazed, D. and Grewal, S. I. S. (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(1), 152-157.

Sun, Z.-W. and Allis, C. D. (2002) 'Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast', *Nature,* 418(6893), 104-108.

Sutani, T., Yuasa, T., Tomonaga, T., Dohmae, N., Takio, K. and Yanagida, M. (1999) 'Fission yeast condensin complex: essential roles of non-SMC subunits for condensation and Cdc2 phosphorylation of Cut3/SMC4', *Genes & Development,* 13(17), 2271-2283.

Swiezewski, S., Crevillen, P., Liu, F., Ecker, J. R., Jerzmanowski, A. and Dean, C. (2007) 'Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, FLC', *Proceedings of the National Academy of Sciences of the United States of America,* 104(9), 3633-3638.

Sánchez Alvarado, A. and Newmark, P. A. (1999) 'Double-stranded RNA specifically disrupts gene expression during planarian regeneration', *Proceedings of the National Academy of Sciences of the United States of America,* 96(9), 5049-5054.

Tabara, H., Yigit, E., Siomi, H. and Mello, C. C. (2002) 'The dsRNA Binding Protein RDE-4 Interacts with RDE-1, DCR-1, and a DExH-Box Helicase to Direct RNAi in *C. elegans*', *Cell,* 109(7), 861-871.

Takahashi, K., Yoshida, N., Murakami, N., Kawata, K., Ishizaki, H., Tanaka-Okamoto, M., Miyoshi, J., Zinn, A. R., Shime, H. and Inoue, N. (2007) 'Dynamic Regulation of p53 Subnuclear Localization and Senescence by MORC3', *Molecular Biology of the Cell,* 18(5), 1701-1709.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M. and Watanabe, Y. (2008) 'The mechanism selecting the guide strand from small RNA duplexes is different among Argonaute proteins', *Plant and Cell Physiology,* 49(4), 493-500.

Takuno, S. and Gaut, B. S. (2011) 'Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly', *Molecular Biology and Evolution,* 29(1), 219-227.

Tan, E. H., Blevins, T., Ream, T. S. and Pikaard, C. S. (2012) 'Functional consequences of subunit diversity in RNA Polymerases II and V', *Cell Reports,* 1(3), 208-214.

Tanaka, T., Saha, S. K., Tomomori, C., Ishima, R., Liu, D., Tong, K. I., Park, H., Dutta, R., Qin, L., Swindells, M. B., Yamazaki, T., Ono, A. M., Kainosho, M., Inouye, M. and Ikura, M. (1998) 'NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ', *Nature,* 396(6706), 88-92.

Tang, G., Reinhart, B. J., Bartel, D. P. and Zamore, P. D. (2003) 'A biochemical framework for RNA silencing in plants', *Genes & Development,* 17(1), 49-63.

Tariq, M., Saze, H., Probst, A. V., Lichota, J., Habu, Y. and Paszkowski, J. (2003) 'Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin', *Proceedings of the National Academy of Sciences of the United States of America,* 100(15), 8823-8827.

Tarutani, Y., Shiba, H., Iwano, M., Kakizaki, T., Suzuki, G., Watanabe, M., Isogai, A. and Takayama, S. (2010) 'Trans-acting small RNA determines dominance relationships in *Brassica* self-incompatibility', *Nature,* 466(7309), 983-986.

Taylor, R. M., Hamer, M. J., Rosamond, J. and Bray, C. M. (1998) 'Molecular cloning and functional analysis of the *Arabidopsis thaliana* DNA ligase I homologue', *The Plant Journal,* 14(1), 75-81.

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

Tompa, R., McCallum, C. M., Delrow, J., Henikoff, J. G., van Steensel, B. and Henikoff, S. (2002) 'Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3', *Current Biology,* 12(1), 65-68.

Tran, R. K., Zilberman, D., de Bustos, C., Ditt, R. F., Henikoff, J. G., Lindroth, A. M., Delrow, J., Boyle, T., Kwong, S., Bryson, T. D., Jacobsen, S. E. and Henikoff, S. (2005) 'Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*', *Genome Biology,* 6(11), R90.

Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V. and Zamore, P. D. (2006) 'A distinct small RNA pathway silences selfish genetic elements in the germline', *Science,* 313(5785), 320-324.

Vaillant, I., Schubert, I., Tourmente, S. and Mathieu, O. (2006) 'MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in *Arabidopsis*', *European Molecular Biology Organization Reports,* 7(12), 1273-1278.

Vaillant, I., Tutois, S., Cuvillier, C., Schubert, I. and Tourmente, S. (2007) 'Regulation of *Arabidopsis thaliana* 5S rRNA genes', *Plant Cell Physiology,* 48(5), 745-752.

Van Blokland, R., Ross, S., Corrado, G., Scollan, C. and Meyer, P. (1998) 'Developmental abnormalities associated with deoxyadenosine methylation in transgenic tobacco', *The Plant Journal,* 15(4), 543-551.

Van Blokland, R., Van der Geest, N., Mol, J. N. M. and Kooter, J. M. (1994) 'Transgene-mediated suppression of Chalcone Synthase expression in *Petunia hybrida* results from an increase in RNA turnover', *The Plant Journal,* 6(6), 861-877.

van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. and Stuitje, A. R. (1990a) 'Flavonoid Genes in *Petunia*: Addition of a limited number of gene copies may lead to a suppression of gene expression', *Plant Cell,* 2(4), 291-299.

van der Krol, A. R., Mur, L. A., Lange, P., Gerats, A. G. M., Mol, J. N. M. and Stuitje, A. R. (1990b) 'Antisense *Chalcone Synthase* genes in *Petunia*: Visualization of variable transgene expression', *Molecular and General Genetics,* 220(2), 204-212.

Vaucheret, H. (2005) 'RNA polymerase IV and transcriptional silencing', *Nature Genetics,* 37(7), 659-660.

Vaughn, M. W. and Martienssen, R. A. (2005) 'Finding the right template: RNA Pol IV, a plant-specific RNA Polymerase', *Molecular Cell,* 17(6), 754-756.

Vazquez, F., Blevins, T., Ailhas, J., Boller, T. and Meins, F. (2008) 'Evolution of *Arabidopsis MIR* genes generates novel microRNA classes', *Nucleic Acids Research,* 36(20), 6429-38.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., Hilbert, J.-L., Bartel, D. P. and Crété, P. (2004) 'Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs', *Molecular Cell,* 16(1), 69-79.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I. S. and Moazed, D. (2004) 'RNAi-mediated targeting of heterochromatin by the RITS complex', *Science,* 303(5658), 672-6.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S. and Martienssen, R. A. (2002) 'Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi', *Science,* 297(5588), 1833-1837.

Vongs, A., Kakutani, T., Martienssen, R. A. and Richards, E. J. (1993) '*Arabidopsis thaliana* DNA methylation mutants', *Science,* 260(5116), 1926-1928.

Vrbsky, J., Akimcheva, S., Watson, J. M., Turner, T. L., Daxinger, L., Vyskot, B., Aufsatz, W. and Riha, K. (2010) 'siRNA–mediated methylation of *Arabidopsis* telomeres', *Public Library of Science Genetics,* 6(6).

Wada, Y., Ohya, H., Yamaguchi, Y., Koizumi, N. and Sano, H. (2003) 'Preferential *de novo* methylation of cytosine residues in non-CpG sequences by a Domains Rearranged DNA Methyltransferase from tobacco pants', *Journal of Biological Chemistry,* 278(43), 42386-42393.

Wallrath, L. L. and Elgin, S. C. (1995) 'Position effect variegation in *Drosophila* is associated with an altered chromatin structure', *Genes & Development,* 9(10), 1263-1277.

Wang, F., Koyama, N., Nishida, H., Haraguchi, T., Reith, W. and Tsukamoto, T. (2006) 'The assembly and maintenance of heterochromatin initiated by transgene repeats are independent of the RNA interference pathway in mammalian cells', *Molecular Cell Biology,* 26(11), 4028-40.

Wang, G.-L., Wang, C.-Y. u., Cai, X.-Z., Chen, W., Wang, X.-H. and Li, F. (2010) 'Identification and expression analysis of a novel CW-type zinc finger protein MORC2 in cancer cells', *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology,* 293(6), 1002-1009.

Wang, M.-B. and Dennis, E. S. (2009) 'SPT5-like, a new component in plant RdDM', *European Molecular Biology Organization Reports,* 10(6), 573-575.

Wang, R. Y.-H., Gehrke, C. W. and Ehrlich, M. (1980) 'Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues', *Nucleic Acids Research,* 8(20), 4777-4790.

Wang, X.-J., Gaasterland, T. and Chua, N.-H. (2005) 'Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*', *Genome Biology,* 6(4), R30.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T. and Patel, D. J. (2008) 'Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex', *Nature,* 456(7224), 921-926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G. S., Tuschl, T. and Patel, D. J. (2009) 'Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes', *Nature,* 461(7265), 754-761.

Wassenegger, M., Heimes, S., Riedel, L. and Sänger, H. L. (1994) 'RNA-directed de novo methylation of genomic sequences in plants', *Cell,* 76(3), 567-576.

Wassenegger, M. and Pélissier, T. (1998) 'A model for RNA-mediated gene silencing in higher plants', *Plant Molecular Biology,* 37(2), 349-362.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) 'Jalview Version 2—a multiple sequence alignment editor and analysis workbench', *Bioinformatics,* 25(9), 1189-1191.

Waterhouse, P. M., Graham, M. W. and Wang, M.-B. (1998) 'Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA', *Proceedings of the National Academy of Sciences of the United States of America,* 95(23), 13959-13964.

Watson, M. L., Zinn, A. R., Inoue, N., Hess, K. D., Cobb, J., Handel, M. A., Halaban, R., Duchene, C. C., Albright, G. M. and Moreadith, R. W. (1998) 'Identification of morc (microrchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse', *Proceedings of the National Academy of Sciences of the United States of America,* 95(24), 14361-14366.

Whitehouse, I., Flaus, A., Cairns, B. R., White, M. F., Workman, J. L. and Owen-Hughes, T. (1999) 'Nucleosome mobilization catalysed by the yeast SWI/SNF complex', *Nature,* 400(6746), 784-787.

Wianny, F. and Zernicka-Goetz, M. (2000) 'Specific interference with gene function by double-stranded RNA in early mouse development', *Nature Cell Biology,* 2(2), 70-75.

Wierzbicki, A. T., Cocklin, R., Mayampurath, A., Lister, R., Rowley, M. J., Gregory, B. D., Ecker, J. R., Tang, H. and Pikaard, C. S. (2012) 'Spatial and functional

relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the *Arabidopsis* epigenome', *Genes & Development,* 26(16), 1825-1836.

Wierzbicki, A. T., Haag, J. R. and Pikaard, C. S. (2008) 'Noncoding transcription by RNA Polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes', *Cell,* 135(4), 635 -648.

Wierzbicki, A. T., Ream, T. S., Haag, J. R. and Pikaard, C. S. (2009) 'RNA Polymerase V transcription guides ARGONAUTE4 to chromatin', *Nature Genetics,* 41(5), 630-634.

Wilson, R. N., Heckman, J. W. and Somerville, C. R. (1992) 'Gibberellin is required for flowering in *Arabidopsis thaliana* under short days', *Plant Physiology,* 100(1), 403-408.

Wingard, S. A. (1928) 'Hosts and symptoms of ring spot, a virus disease of plants', *Journal of Agricultural Research,* 12(3), 127-153.

Woodcock, C. L. and Ghosh, R. P. (2010) 'Chromatin higher-order structure and dynamics', *Cold Spring Harbor Perspectives in Biology,* 2(5).

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C. and Qi, Y. (2010) 'DNA methylation mediated by a microRNA pathway', *Molecular Cell,* 38(3), 465-475.

Wyatt, G. R. (1951) 'Recognition and estimation of 5-methylcytosine in nucleic acids', *Biochemical Journal,* 48(5), 581-584.

Xie, M., Ren, G., Costa-Nunes, P., Pontes, O. and Yu, B. (2012) 'A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation', *Nucleic Acids Research,* 40(10), 4422-4431.

Xie, Z., Fan, B., Chen, C. and Chen, Z. (2001) 'An important role of an inducible RNA-dependent RNA polymerase in plant antiviral defense', *Proceedings of the National Academy of Sciences of the United States of America,* 98(11), 6516-6521.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Jacobsen, S. E. and Carrington, J. C. (2004) 'Genetic and functional diversification of small RNA pathways in plants', *Public Library of Science Biology,* 2(5).

Yamaguchi-Shinozaki, K. and Shinozaki, K. (1994) 'A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress', *Plant Cell,* 6(2), 251-64.

Yao, Y., Bilichak, A., Golubov, A., Blevins, T. and Kovalchuk, I. (2010) 'Differential sensitivity of *Arabidopsis* siRNA biogenesis mutants to genotoxic stress', *Plant Cell Reports,* 29(12), 1401-1410.

Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J.-C. and Worman, H. J. (1997) 'Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR', *Journal of Biological Chemistry,* 272(23), 14983-14989.

Ye, R., Wang, W., Iki, T., Liu, C., Wu, Y., Ishikawa, M., Zhou, X. and Qi, Y. (2012) 'Cytoplasmic assembly and selective nuclear import of Arabidopsis ARGONAUTE4/siRNA complexes', *Molecular cell,* 46(6), 859-870.

Yokthongwattana, C., Bucher, E., Caikovski, M., Vaillant, I., Nicolet, J., Scheid, O. M. and Paszkowski, J. (2010) 'MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing', *European Molecular Biology Organization Journal,* 29(2), 340-351.

Yoshikawa, M., Peragine, A., Park, M. Y. and Poethig, R. S. (2005) 'A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*', *Genes & Development,* 19(18), 2164-75.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., Steward, R. and Chen, X. (2005) 'Methylation as a crucial step in plant microRNA biogenesis', *Science,* 307(5711), 932-935.

Yu, D., Fan, B., MacFarlane, S. A. and Chen, Z. (2003) 'Analysis of the involvement of an inducible *Arabidopsis* RNA-dependent RNA Polymerase in antiviral defense', *Molecular Plant-Microbe Interactions,* 16(3), 206-216.

Yu, Z., Wright, S. I. and Bureau, T. E. (2000) 'Mutator-like elements in *Arabidopsis thaliana*: Structure, diversity and evolution', *Genetics,* 156(4), 2019-2031.

Yuki, K., Hidetoshi, S., Tetsu, K., Asuka, M., Wim, J. J. S., Maarten, K. and Tetsuji, K. (2007) 'Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats', *The Plant Journal,* 49(1), 38-45.

Zamore, P. D., Tuschl, T., Sharp, P. A. and Bartel, D. P. (2000) 'RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals', *Cell,* 101(1), 25-33.

Zemach, A. and Grafi, G. (2003) 'Characterization of *Arabidopsis thaliana* methyl-CpG-binding domain (MBD) proteins', *Plant Journal,* 34(5), 565-572.

Zemach, A., Li, Y., Wayburn, B., Ben-Meir, H., Kiss, V., Avivi, Y., Kalchenko, V., Jacobsen, S. E. and Grafi, G. (2005) 'DDM1 binds *Arabidopsis* methyl-CpG binding domain proteins and affects their subnuclear localization', *Plant Cell,* 17(5), 1549-1558.

Zhang, C. J., Ning, Y. Q., Zhang, S. W., Chen, Q., Shao, C. R., Guo, Y. W., Zhou, J. X., Li, L., Chen, S. and He, X. J. (2012) 'IDN2 and its paralogs form a complex required for RNA-directed DNA Methylation', *Public Library of Science Genetics,* 8(5).

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W. L., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E. and Ecker, J. R. (2006) 'Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*', *Cell,* 126(6), 1189-1201.

Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.-Y. and Chen, X. (2009) 'Intergenic transcription by RNA Polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in *Arabidopsis*', *Genes & Development,* 23(24), 2850-2860.

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

Zheng, X., Zhu, J., Kapoor, A. and Zhu, J.-K. (2007) 'Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing', *European Molecular Biology Organization Journal,* 26(6), 1691-1701.

Zheng, Z., Xing, Y., He, X.-J., Li, W., Hu, Y., Yadav, S. K., Oh, J. and Zhu, J.-K. (2010) 'An SGS3-like protein functions in RNA-directed DNA methylation and transcriptional gene silencing in *Arabidopsis*', *The Plant Journal,* 62(1), 92-99.

Zilberman, D., Cao, X. and Jacobsen, S. E. (2003) 'ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation', *Science,* 299(5607), 716-719.

Zilberman, D., Cao, X., Johansen, L. K., Xie, Z., Carrington, J. C. and Jacobsen, S. E. (2004) 'Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA Methylation triggered by inverted repeats', *Current Biology,* 14(13), 1214-1220.

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. and Henikoff, S. (2007) 'Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription', *Nature Genetics,* 39(1), 61-69.

Zubko, E., Gentry, M., Kunova, A. and Meyer, P. (2012) 'De novo DNA methylation activity of METHYLTRANSFERASE 1 (MET1) partially restores

body methylation in *Arabidopsis thaliana*', *The Plant Journal,* 71(6), 1029-1037.

Čaikovski, M., Yokthongwattana, C., Habu, Y., Nishimura, T., Mathieu, O. and Paszkowski, J. (2008) 'Divergent evolution of CHD3 proteins resulted in MOM1 refining epigenetic control in vascular plants', *Pulbic Library of Science Genetics,* 4(8).