Diverse mechanisms of *Athila* retrotransposon epigenetic silencing in *Arabidopsis thaliana*

**DISSERTATION**

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Saivageethi Nuthikattu

Graduate Program in Plant Cellular and Molecular Biology

The Ohio State University

2014

Dissertation Committee:

Professor R. Keith Slotkin, Advisor

Professor Dave Bisaro

Professor Harold Fisk

Professor Juan Alfonzo
Abstract

Transposable elements (TEs) are jumping genes, which when active move from one part of the genome to another and cause mutations and genome instability in their hosts. Transposable element proliferation negatively correlates with the fitness of the organism. Eukaryotes have evolved several defense mechanisms to combat these genetic parasites. In the model organism *Arabidopsis thaliana*, two major defense responses against TE activity are RNA directed DNA methylation (RdDM), which causes transcriptional silencing of TEs and RNA Interference (RNAi), which causes post transcriptional degradation of TE mRNAs. Through my research work, I have shown that when both the transcriptional and post transcriptional regulation of TEs is lost, there is a third host defense response that causes translational inhibition of TEs. I have shown that *UBP1b*, a component of plant stress granules is capable of causing translational inhibition of *Athila* retrotransposons, one of the major retrotransposons in *Arabidopsis thaliana*. Transcriptionally active *Athila* produces 21-22nt siRNAs through RNAi and I have shown that one such 21nt siRNA, siRNA854 causes translational inhibition of *UBP1b*. *Athila* uses the second host defense response, RNAi to target the third defense mechanism, translational regulation by stress granules. Thus, *Athila* targeting host gene *UBP1b* is akin to gene regulation by miRNAs and *trans*-acting siRNAs. I also found that this RNAi pathway which was always known to act post transcriptionally is capable of
acting at transcriptional level and perform similar function as the first level of Athila regulation by RdDM. RNAi pathway plays a role in transcriptional gene silencing through de novo methylation of active TEs via 21-22nt siRNAs. Thus, Athila and its host organism Arabidopsis thaliana are continually evolving new defense strategies to combat each other.
Dedication

This document is dedicated to my family, friends and Slotkin lab.
Acknowledgments

First and foremost I would like to express my sincere gratitude to my advisor Dr. Keith Slotkin for all his valuable guidance, encouragement, motivation and support. He has made my research journey through all these five years extremely exciting and amazing. I would also like to thank my wonderful lab mates in the Slotkin lab for creating such a brilliant atmosphere in the lab and made lab work fun. I would also like to thank all my committee members for their excellent advice and inputs in all the committee meetings. I thank all the Professors and staff of Molecular Genetics department for all their help and support. Last but not the least, I would like to thank my parents, brother, friends and relatives for being there for me at all times.
Vita

2009............................................................ B.Tech. Industrial Biotechnology,
Anna University

2009 to present ...........................................Graduate Teaching Associate, Department
of Molecular Genetics, The Ohio State
University

Publications
Andrea D. McCue, Kaushik Panda, Saivageethi Nuthikattu, Sarah G. Choudury,
EricaN. Thomas and R. Keith Slotkin (2014). ARGONAUTE 6 is recruited to chromatin
by transposable element mRNA degradation products to establish DNA methylation.
Accepted in EMBO journal and published online 11.11.14.

Brittany Cox, Majeda Al Rawahneh, Jacob Anderson, Corey Berta, Evan Boucher, Ryan
Bridenstine, Kimberlee Eakins, Rachel Embly, Candace Flynn, Andrew Francis, Jenna
Gallagher, Brian Jones, Clint Miller, Forrest Payne, Sylvia Phimsouay, Nandita
Rajshekhar, Robert Rosia, Scott Spencer, Kaitlyn Stevelak, David Sun, Valerie Childress,
Saivageethi Nuthikattu, Kaushik Panda, R. Keith Slotkin (2014) The proteins CMT3,
UHRF1 and ESP1 represses select transposable elements in the model plant Arabidopsis
thaliana. Journal of Undergraduate Research at Ohio State (Editor reviewed research publication) for Molecular Genetics Laboratory course MG5601.


A.D. McCue, **Saivageethi Nuthikattu**, S.H. Reeder and R. Keith Slotkin 2012. Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. PLoS Genetics v8: e1002474
Fields of Study

Major Field: Plant Cellular and Molecular Biology
Table of Contents

Abstract .......................................................................................................................... ii

Dedication ..................................................................................................................... iv

Acknowledgments ......................................................................................................... v

Vita ................................................................................................................................... vi

Table of Contents ........................................................................................................... ix

List of Tables ................................................................................................................ xiv

List of Figures ............................................................................................................... xvi

Chapter 1: Introduction to Transposable Elements and different modes of epigenetic regulation by host organisms ................................................................. 1

1.1 Classification of Transposable elements ............................................................... 2

1.2 Transposable elements and their epigenetic regulation ....................................... 3

1.3 Distinction of DNA methylation of TEs from genes ............................................ 6

1.4 Arabidopsis small RNA pathways ........................................................................ 7

1.5 Athila LTR retrotransposons and its life cycle ....................................................... 8

Chapter 2: Mutual translational regulation of Athila, retrotransposons and UBP1b, a stress granule protein .................................................................................... 15
2.1 Abstract ........................................................................................................................................... 16

2.2 Introduction ....................................................................................................................................... 17
  2.2.1 Different levels of *Athila* epigenetic regulation ........................................................................ 17
  2.2.2 TE regulation of genes ................................................................................................................ 18
  2.2.3 *Athila* targets a gene in *trans* .................................................................................................. 19
  2.2.4 UBP1b and stress granules ......................................................................................................... 20

2.3 Materials and Methods .................................................................................................................... 23
  2.3.1 Plant material .............................................................................................................................. 23
  2.3.2 Quantitative Reverse Transcriptase–PCR (qRT-PCR) and qPCR ............................................. 23
  2.3.3 Small RNA Northern blots ......................................................................................................... 24
  2.3.4 Transgene construction and transformation ............................................................................... 24
  2.3.5 GUS staining and MUG assay ................................................................................................. 25
  2.3.6 Fluorescence Microscopy ......................................................................................................... 26
  2.3.7 Preparation of gag particle fractions and extraction of DNA from the particles ................. 26
  2.3.8 Silver staining ............................................................................................................................ 27
  2.3.9 Silique dissection ....................................................................................................................... 28
  2.3.10 GAG protein accumulation by western .................................................................................. 28
  2.3.11 Product Enhanced Reverse Transcriptase Assay (PERT) ...................................................... 28
2.3.12 UBP1b immunoprecipitation-western ................................................................. 29
2.3.13 polyA Northern ................................................................................................. 30
2.4 Results ...................................................................................................................... 30
2.4.1 Transcriptional activation of Athila6 results in production of Athila siRNAs
and siRNA854 ........................................................................................................... 30
2.4.2 siRNA854 accumulation represses reporter gene transcripts with the
UBP1b 3'UTR in seedlings, leaves and inflorescences tissues ......................... 33
2.4.3 siRNA854 accumulation in pollen and immature seeds regulates transgene
transcripts with the UBP1b 3'UTR ........................................................................... 37
2.4.4 Effect of UBP1b on Athila copy number, expression and translation ........ 42
2.4.5 Loss of different levels of Athila regulation leads to sterility and defective
centromeres ............................................................................................................. 45
2.4.6 UBP1b causes translational inhibition of Athila ............................................. 49
2.5 Discussion ................................................................................................................. 57
2.5.1 UBP1b acts as third line of defense against Athila ......................................... 57
2.5.2 Regulation of genes by TEs ................................................................................ 59
2.5.3 Athila via siRNA854 causes translational inhibition of UBP1b ................. 60
2.5.4 Future directions .................................................................................................. 62

Chapter-3: RNAi plays a role in de novo methylation of active transposable elements
through 21-22 nucleotide small RNAs and requires POL V scaffolding transcripts..... 102
3.1 Abstract ......................................................................................................................................... 103

3.2 Introduction ................................................................................................................................... 104

3.2.1 Mechanisms of TE silencing ................................................................................................. 104

3.2.2 Maintenance TE DNA methylation ....................................................................................... 104

3.2.3 Pol IV-RNA directed DNA Methylation (Pol IV-RdDM) .................................................... 105

3.2.4 RDR6-RNA directed DNA Methylation (RDR6-RdDM) .................................................... 107

3.3 Materials and Methods ................................................................................................................ 109

3.3.1 Plant material .......................................................................................................................... 109

3.3.2 Expression Analysis by Quantitative Reverse Transcription-PCR ................................. 109

3.3.3 Bisulfite Conversion and Sequencing ...................................................................................... 110

3.3.4 Northern Blotting .................................................................................................................... 111

3.3.5 Small RNA Library Production, Sequencing, and Analysis ............................................. 111

3.3.6 Chromatin Immunoprecipitation .......................................................................................... 112

3.4 Results .......................................................................................................................................... 113

3.4.1 TE expression-dependent RNA directed DNA methylation (RdDM) activity .................. 113

3.4.2 The RDR6-RdDM pathway plays a role in the establishment of CHH methylation upon TE transcriptional activation ................................................................. 117

3.4.3 Other players involved in RDR6-RdDM ................................................................................. 122
3.4.4 Examination of an additional TE target of RDR6-RdDM demonstrates that 21 and/or 22 nt siRNAs direct methylation ................................................................. 123

3.4.5 The RDR6-RdDM and POL IV-RdDM pathways do not contribute to the maintenance of a trans-generationally silenced TE .............................................. 127

3.4.6 The RDR6-RdDM Pathway Functions to Correctively Reestablish Methylation and Silencing ........................................................................................................ 130

3.4.7 Developmental time point in which re silencing of TEs occurs in Arabidopsis .................................................................................................................. 133

3.4.8 Role of RDR6-RdDM and POL IV-RdDM throughout Arabidopsis development .................................................................................................................. 135

3.4.9 Effect of RDR6 in different gametes .................................................................................................................. 137

3.4.10 Effect of de novo methylation on Athila steady state transcription ............ 139

3.4.11 RDR6-RdDM seen in Athila LTR and IR region but not in ENV .......... 141

3.4.12 Pol V occupancy at transcriptionally active TEs .................................................. 142

3.5 Discussion .......................................................................................................................... 144

3.5.1 RNAi or RDR6-RdDM is responsible for de novo methylation of active TEs through 21-22nt siRNAs .................................................................................. 144

3.5.2 Functional role of RDR6-RdDM and POL IV-RdDM pathways .................... 146

3.5.3 Requirement of POL V for RDR6-RdDM activity ............................................. 147

3.5.4 Future directions ............................................................................................................ 148
Chapter-4 Perspectives........................................................................................................ 177

4.1 Different levels of *Athila* retrotransposon repression by *Arabidopsis thaliana*.... 178

4.2 Regulation of TEs by plant stress granules ................................................................. 179

4.3 Transcriptional regulation of TEs by RDR6-RdDM pathway ...................................... 181

4.4 Role of *Arabidopsis* TE small RNAs in different TE silencing pathways............. 182

4.5 Effect of TE epigenetic regulation on host gene regulation........................................ 183

Appendix A: Data from other small projects................................................................ 185

References..........................................................................................................................192
List of Tables

Table 2. 1 Frequency of 21 nt siRNA in SBS small RNA libraries.......................... 100
Table 2. 2 PCR primers and mutant alleles used. .......................................................... 101

Table 3. 1 PCR primers and mutant alleles used ............................................................. 174
Table 3. 2 Crosses performed to test role of RDR6-RdDM and POLIV-RdDM components in Athila re-silencing ................................................................. 175
Table 3. 3 Crosses performed to test the effect of RDR6 in different gametes. ............ 176

Table A. 1 Primer sequences used in this report............................................................... 192
List of Figures

Figure 1. 1 DNA methylation: Key player in TE epigenetic silencing................................. 11
Figure 1. 2 RNA directed DNA methylation vs RNA Inteference........................................ 12
Figure 1. 3 Maintenance of TE epigenetic silencing............................................................... 12
Figure 1. 4 microRNA and trans-acting siRNA pathways in plants........................................ 13
Figure 1. 5 Life cycle of Retro transposon.............................................................................. 14

Figure 2. 1 Athila experiences three levels of host repression.................................................. 65
Figure 2. 2 The 14kb reconstructed consensus Athila6A element............................................. 66
Figure 2. 3 Expression of the Athila6 retrotransposon leads to accumulation of Athila 21–22 nt siRNAs, including siRNA854................................................................................. 67
Figure 2. 4 Increased levels of siRNA854 leads to UBP1b 3’UTR repression in T1 and T2 Arabidopsis seedlings and T1 inflorescences........................................................................ 69
Figure 2. 5 Accumulation of 21–22 nt siRNA854 negatively regulates transgene transcripts with the UBP1b 3′ UTR in T3 inflorescences......................................................... 72
Figure 2. 6 Accumulation of 21–22 nt siRNA854 in wt pollen regulates transgene transcripts with the UBP1b 3’UTR............................................................... 75
Figure 2. 7 Immature seed GUS staining of Col, 35S: GUS-UBP1b 3’UTR crossed to Col and rdr6........................................................................................................ 79
Figure 2. 8 Number of copies and expression of Athila6 in wt and 12 different mutant background combinations. ................................................................. 80
Figure 2. 9 Isolation of gag particles and silver staining reveals gag encapsidation of Athila................................................................. 82
Figure 2. 10 Loss of epigenetic regulation of Athila affects seed formation in siliques. . 83
Figure 2. 11 siRNA processed from Cen180 and 106B follow pathway distinct from Athila siRNAs. ................................................................. 86
Figure 2. 12 Role of UBP1b in Athila repression. ................................................................. 91
Figure 2. 13 Athila DNA isolated from gag particles from a panel of mutants involved in different layers of Athila repression. ................................................................. 92
Figure 2. 14 UBP1b, a stress granule protein acts to repress Athila protein production. . 95
Figure 2. 15 Product Enhanced Reverse Transcriptase Assay (PERT). ......................... 96
Figure 2. 16 ImmunoPrecipitation-western (IP-Western) of UBP1b-GFP from seedling roots................................................................. 99

Figure 3. 1 POL IV-RdDM is responsible for hyper CHH methylation.................. 150
Figure 3. 2 RDR6-RdDM plays a role in expression dependent de novo methylation of Athila.................................................................................................................. 153
Figure 3. 3 Other players involved in RDR6-RdDM.................................................. 155
Figure 3. 4 RDR6-RdDM plays a role in expression independent de novo methylation of AtENSPM and functions through 21-22nt siRNAs................................. 156
Figure 3. 5 RDR6-RdDM and POL IV-RdDM pathways do not play a role in maintenance of TE silencing. .......................................................... 158
Figure 3. 6 RDR6-RdDM functions in the corrective reestablishment of TE silencing. 159
Figure 3. 7 Time point in development where re-silencing of Athila occurs. ............ 162
Figure 3. 8 Role of RDR6-RdDM and POL IV-RdDM in various Arabidopsis developmental stages. ................................................................. 163
Figure 3. 9 Parent of origin effects of DDM1 and RDR6. ........................................ 165
Figure 3. 10 Effect of de novo methylation on Athila transcription. ......................... 167
Figure 3. 11 IR region of Athila undergoes RDR6-RdDM. .................................... 168
Figure 3. 12 POL V occupancy is seen in transcriptionally active TEs. ................. 169
Figure 3. 13 POL V occupancy in Athila6 LTR and IR but not in ENV. .................... 171
Figure 3. 14 Model of TE epigenetic silencing. .................................................. 173

Figure 4. 1 Regulation of Athila at various stages of its life cycle. ........................ 184

Figure A. 1 Analysis of ddm1 (Col) X ddm1 (Ler) cross and reciprocal crosses. .... 189
Figure A. 2 Athila cartoon .................................................................................... 190
Figure A. 3 Athila transcripts produced by RDR6 ............................................. 191
Chapter 1: Introduction to Transposable Elements and different modes of epigenetic regulation by host organisms
1.1 Classification of Transposable elements

Transposable elements or TEs or Transposons are selfish parasites which when active jump around the genome and cause extensive damage to their host organisms. Major portion of eukaryotic genomes is made up of TEs (For example, humans 48% and maize 85%). There are two major types of transposable elements classified based on the mode of their movement or transposition, namely, type I and type II elements. Type I elements are also called as retrotransposons. These elements move by copying from one location in the genome and pasting in another location of the genome (Slotkin and Martienssen 2007a) and further classified as Long Terminal Repeat, LTR retrotransposons and non-LTR retrotransposons. On the basis of order of protein domains within the elements, LTR retroelements are subdivided into Gypsy and Copia retroelements. Non-LTR retroelements are subdivided into Long Interspersed Elements (LINES) and Short Interspersed Elements (SINES). Type I elements are mainly responsible for huge genome sizes of eukaryotes, for example 49-78% of maize genome and 42% of human genome is made up of retrotransposons. Type II elements are DNA transposons which move by cutting themselves out from one region in the genome and get pasted into different region of the genome. There is class of DNA transposons called Helitrons and these replicate through rolling circle mechanism. Type II elements are highly mutagenic and when active get inserted into genes.
1.2 Transposable elements and their epigenetic regulation

Transposable elements (TEs) are mobile DNA fragments that constitute major portions of eukaryotic genomes, including over 12% of Arabidopsis (Arabidopsis Genome Initiative 2000), 85% of maize (Schnable et al. 2009) and 45% of the human genome (Lander et al. 2001). When transcriptionally active, TEs have the potential to create mutations by insertions, deletions or generating double strand breaks in DNA and hence cause genome instability in their hosts. In order to suppress the mutagenic potential of TEs, the eukaryotic genome has evolved several defense mechanisms to control these jumping TEs (Roudier et al. 2011). Though TE mobilization in animals often results in embryonic lethality or disease (Deininger and Batzer 1999), plants can generally withstand more chromosomal damage, making them ideal system to study both TEs and their host regulation. In plants, TE regulation occurs at different levels of its life cycle.

First level of TE regulation occurs in transcription via chromatin modification.

TEs are targeted for epigenetic repression mediated by the overlapping signals of cytosine DNA methylation, repressive histone tail modifications, and remodeling of chromatin into transcriptionally recalcitrant condensed heterochromatin (reviewed in (Slotkin and Martienssen 2007b)). Gene regulation can be influenced by the epigenetic regulation of TEs; however, this only occurs due to the proximity of a preexisting or newly transposed TE to a gene. This regulation of genes by neighboring TEs in cis can be due to multiple mechanisms, including interruption of a regulatory element, or by local spreading of repressive chromatin modifications such as DNA or histone methylation, resulting in position-effect variegation and potentially the formation of heritable
epialleles (Lippman et al. 2004; Sun et al. 2004). Position-effect variegation is due to atypical proximity of gene with heterochromatin that leads to inactivation of the gene.

DNA methylation is a key player in the epigenetic regulation of TEs and occurs in three steps namely, initiation, establishment and maintenance (Figure 1.1). It is known that establishment of TE methylation occurs through RNA directed DNA methylation (Schoft et al. 2009). TEs are major producers of small RNAs that act to maintain the TE in an epigenetically silenced state. In plants, and in animals, heterochromatin modifications are targeted by the activity of small RNAs. For example, in the mouse TE-derived piwi-interacting RNAs (piRNAs) guide DNA methylation to TEs (Aravin et al. 2008). In the model plant Arabidopsis thaliana, the cycle of RNA-directed DNA methylation (RdDM) or Transcriptional Gene Silencing (TGS) is initiated by the plant-specific RNA Polymerase IV (Pol IV), which produces a non-protein coding transcript that is converted into double stranded RNA (dsRNA) by the activity of RNA-dependent RNA Polymerase 2 (RDR2)(reviewed in (Law and Jacobsen 2010a)). Dicer-like 3 (DCL3) cleaves this TE dsRNA into small interfering RNAs (siRNAs) of 24 nucleotides (nt) in length, which are incorporated into either Argonaute 4(AGO4), AGO6, or potentially AGO9 (Pereira 2004). These siRNA loaded Argonaute proteins act to maintain the heterochromatic or the silent state of TEs by targeting them for DNA and histone tail methylation and repress transcription of TEs (Figure 1.2). This is done by the interaction of Argonaute-siRNA complex with TE scaffolding non coding RNA transcript produced by the activity of another plant specific
RNA polymerase Pol V which then recruits DNA methyltransferases and methylates TE DNA.

Maintenance of DNA methylation occurs in plants by maintenance methyltransferases such as Maintenance of DNA methylation 1 (MET1) that methylates cytosines in CG context, and Chromo methyl transferase 3 and 2 (CMT3 and CMT2) that methylate cytosines in CHG and CHH contexts, where H is A, T or C (Stroud et al. 2014; Teixeira et al. 2009a; Zemach et al. 2013). The CMT3 and CMT2 chromo methyltransferases require Histone 3 Lysine9 di methylation (H3K9me2) to methylate in CHG and CHH context respectively. H3K9me2 is established by KYP (KRYPTONITE) H3 methyl transferases (Stroud et al. 2014)(Law and Jacobsen 2010a). Unlike Animals, where methylation is reset every generation, plants exhibit trans-generational epigenetic silencing from parent to offspring. Methylation is propagated in plants from cell to cell through replication which is then followed by mitosis into daughter cells (Figure 1.3). **Among the three steps of DNA methylation, not much is known about the first step, initiation of DNA methylation.**

Second level of TE regulation occurs at post transcriptional level via RNAi. It was recently discovered that upon TE epigenetic activation and RNA Polymerase II (Pol II) transcription, an endogenous RNAi pathway degrades TE mRNAs in an AGO1-dependent fashion (McCue et al. 2012a). In this endogenous RNAi pathway, RDR6, DCL2, and DCL4 act to degrade Pol II-derived TE transcripts into 21-22 nt small interfering RNAs (siRNAs) that are incorporated into AGO1. Thus, AGO1 does incorporate TE siRNAs; however, this occurs only when TEs are transcriptionally active.
and produce 21‒22 nt siRNAs (Figure 1.2). It is not known if there is any interconnection between the first and second levels of TE regulation. Also, not much is known about the role of TE siRNAs in the gene regulation.

1.3 Distinction of DNA methylation of TEs from genes

DNA methylation is present throughout the mammalian genomes where as in plant genomes it is mainly present only in regions containing centromere, transposable elements, and repeats (Lister et al. 2009; Zhang et al. 2006). However, there are few plant flowering genes which when transcribed show gene body methylation that does not affect expression (Zemach et al. 2010) and this methylation is mainly in the CG context maintained by MET1 (Slotkin et al. 2009). In plants, it is especially important for proper control of RdDM and KYP-CMT3/CMT2 maintenance TE DNA methylation pathways so that there is no misregulation and methylation of endogenous genes. This regulation is done by a histone demethylase named IBM1 and it removes H3K9 methylation to maintain low levels of CHG methylation (Inagaki et al. 2010; Saze and Kakutani 2011). CHG and CHH hypermethylation is observed in methylated gene bodies when IBM1 is mutated (Inagaki et al. 2010; Miura et al. 2009). Thus, IBM1 prevents non CG methylation in methylated gene bodies.

When TEs transpose into genes, epigenetic silencing of TEs may cause harmful effects on the expression of neighbouring genes (Hollister and Gaut 2009). In order to protect genes from these deleterious effects caused by TEs and remove unwanted cytosine methylation, DNA demethylases play a role. In plants, methylated cytosine is removed from DNA by DME (DEMETER)/ROS1 family of DNA glycosylases that work
as DNA demethylases (Zhang and Zhu 2012). IDM1 mediates histone acetylation and aids in the function of DNA demethylases and specifically demethylate TEs with low H3K9me2 and H3K4me2 (Ibarra et al. 2012; Li et al. 2012; Penterman et al. 2007; Qian et al. 2012). In *Arabidopsis*, there is equilibrium between active demethylation and RdDM through gene expression regulation. In *met1* mutant plants, there is down regulation of ROS1, a somatic demethylase (Huettel et al. 2006; Li et al. 2012). However, in plants lacking IDM1, there is increase in ROS1 expression even though its activity is impaired (Qian et al. 2012). DME, DNA demethylase or glycosylase is expressed in the female gametophyte of *Arabidopsis* (Zhang and Zhu 2012). In *dme* mutants, there is loss of DNA methylation in female gametophyte and that leads to seed abortion in *Arabidopsis* (Choi et al. 2002) (Huang and Zhu 2014).

### 1.4 Arabidopsis small RNA pathways

Various kinds of small RNAs are produced in *Arabidopsis thaliana* such as MicroRNAs (miRNAs) and trans-acting siRNAs (tasiRNAs) ([Figure 1.4](#)) (Vaucheret 2006). MicroRNAs act in plants and animals to regulate gene expression on the post-transcriptional or translational level. In *Arabidopsis*, DCL1 produces 21 nt microRNAs from stem-loop precursor transcripts generated by RNA polymerase II (PolII), and these microRNAs are loaded primarily into AGO1. Thus, microRNAs are not amplified by an RNA-dependent RNA polymerase, and only one or two single small RNA species accumulate from the microRNA locus. In contrast, most plant siRNAs are the products of RNA-dependent RNA polymerases, and cleavage of these long dsRNAs produces clusters of siRNAs from each locus. Similar to microRNAs, the trans-acting siRNAs are
also known to regulate genes. This pathway begins with the cleavage of a non-protein coding transcript by the microRNA-loaded AGO1 or AGO7, which initiates the DCL4- and RDR6-dependent phased production of siRNAs (reviewed in (Chapman and Carrington 2007)). These ta-siRNAs are loaded into AGO1 and regulate gene expression similar to microRNAs. Similar to transposons, proteins such as DCL4, RDR6 and AGO1, as well as DCL2, act on viral transcripts in the virus-induced gene silencing (VIGS) pathway to initiate and amplify the 21–22 nt siRNA signal. These signals participate in the post-transcriptional degradation of the viral mRNAs, as well as to transport the siRNAs to unaffected regions of the plant. This helps to mount a systemic resistance to the spread of the virus (Dunoyer, Himber, Voinnet 2005; Dunoyer et al. 2010a).

### 1.5 *Athila* LTR retrotransposons and its life cycle

Retrotransposons (as mentioned in section 1.1 above) are one of two main branches of TEs and are ubiquitous in the genomes of plants, animals and fungi (Flavell et al. 1992; Suoniemi, Tanskanen, Schulman 1998; Voytas et al. 1992a). They are characterized by their use of an RNA intermediate before transposition and comprise major portion of Arabidopsis heterochromatin. They replicate by a cycle of transcription, reverse transcription, and integration of the cDNA copies back into the genome (Kumar and Bennetzen 1999). LTR retrotransposons are a subclass of retrotransposons that have been particularly successful in plants (Sanmiguel and Bennetzen 1998). LTR retrotransposons bear a strong resemblance in sequence and life-cycle to retroviruses. Both contain coding regions for viral particle, integrase, reverse transcriptase, and proteolytic protein domains (Figure 1.5). Many retrotransposons carry a deficient
envelope coding region that may once have allowed them to leave the cell and hence retrotransposons are also known as dead retroviruses. It is unclear whether retrotransposons evolved from retroviruses or vice versa.

The centromeric gypsy family of LTR retrotransposons, *Athila* is the largest family of TEs in *Arabidopsis*, occupying over 2.7% of the genome and located in pericentromeric heterochromatin (Arabidopsis Genome Initiative 2000; Pereira 2004) and preferentially transposes to this location when active. The reconstructed *Athila* element is large (14kb), though there are many smaller, more degenerate copies in the genome (Slotkin 2010). Like other LTR retrotransposons, *Athila* possesses two LTRs, a *gag/pol* ORF (Open Reading Frame) encoding the proteins required for its life cycle such as *Gag*, encoding the structural protein involved in nucleocapsid formation and *Pol*, specifying the activities for the reverse transcription and integration of new copies. *Pol* is a polyprotein and contains domains for an AP (aspartic proteinase), responsible for the post-translational processing of the *Pol* ORF product, RT (reverse transcriptase) and RNaseH, which, as a bifunctional polypeptide, carries out reverse transcription, and IN (integrase), which inserts the new *Athila* copy into the genome (Sabot and Schulman 2006b). Some *Athila* elements may contain degenerate *env* region (Wright and Voytas 1998). Since the *env* region of *Athila* has no known function, it is neither capable of undergoing precise splicing and translation (Wright and Voytas 1998) nor capable of moving from cell to cell unlike *env* gene of retroviruses such as *Drosophila* Gypsy and vertebrate retroviruses (Kim et al. 1994; Song et al. 1994). A gypsy like element in Barley known as Bagy-2 belonging to *Athila* clade is capable of getting transcribed,
particularly env domains and spliced into subgenomic env akin to vertebrate retroviruses (Vicient, Kalendar, Schulman 2001). Because of the similarity in env splicing of Athila relative Bagy-2 to vertebrate retroviruses, there is a possibility that in recent past Athila elements might have had functional env to move from one cell to another and cause infection.

Despite having the highest copy number of any TE family in the Arabidopsis thaliana genome, Athila elements are deeply silenced in wild type ecotype Columbia (wt Col), and this silencing is dependent on symmetrical CG DNA methylation. Athila was first identified by Steimer et al. as a non autonomous element and was called transcriptionally silent information (TSI) (Steimer et al. 2000). This TSI fragment consisted of env region, 3’ non-coding region and 3’ LTR (Steimer et al. 2000). When DNA methylation is removed, either in DNMT1-homolog maintenance of DNA methylation 1 (met1) mutant, or in a swi/snf family chromatin remodeling protein decrease in DNA methylation 1 (ddm1) mutant, transcriptional activation occurs of TSI occurs (Steimer et al. 2000; Wright and Voytas 1998). Athila retrotransposons are also transcriptionally activated in the vegetative nucleus of wt Col pollen grains (Slotkin et al. 2009). In all of these examples, heterochromatin modifications and condensation are lost, and global activation of TEs occurs (Lippman et al. 2004; Lister et al. 2008a; Slotkin et al. 2009). Similar to other TEs, Athila is regulated at transcriptional level by RdDM and at post-transcriptional level by RNAi. It is not known if there are other layers of Athila regulation.
This leads us to series of big questions regarding transposons and their silencing mechanisms acting at different levels of TE lifecycle, which I will answer through my research.

1. **How the initiation of TE methylation occurs?** (Discussed in Chapter-3)

2. **Do only miRNAs and tasiRNAs play a role in regulation of genes?** (Discussed in Chapter-2)

3. **Apart from transcriptional and post transcriptional gene silencing are there any other modes of TE regulation?** (Discussed in Chapter-2)

4. **Are all the small RNA pathways in Arabidopsis independent of each other?** (Discussed in Chapter-3)

![Figure 1. 1 DNA methylation: Key player in TE epigenetic silencing.](image-url)
Figure 1. 2 RNA directed DNA methylation vs RNA Inteference.
(Modified from Panda et al. 2013 Plant signal behavior)

Figure 1. 3 Maintenance of TE epigenetic silencing.
Figure 1. 4 microRNA and *trans*-acting siRNA pathways in plants.

(This figure is from Vaucheret *et al* 2006 Genes and Development (Vaucheret 2006).)
Figure 1. 5 Life cycle of Retro transposon.

This figure is from Sabot and Schulman, 2006, Heridity(Sabot and Schulman 2006a).
Chapter 2: Mutual translational regulation of *Athila*, retrotransposons and *UBP1b*, a stress granule protein
2.1 Abstract

Transposable elements (TEs) are DNA fragments that are capable of moving from one part of the genome to other. During this movement, TEs cause insertional mutagenesis and genome instability in their hosts. Eukaryotic hosts have evolved two known major defense mechanisms to inhibit TE activity: chromatin modification by DNA or histone methylation to repress transcription, and post-transcriptional degradation of TE mRNA by RNA interference. Our lab works on the LTR retrotransposon *Athila* that constitutes 2.7% of *Arabidopsis thaliana* genome. I have shown that upon transcriptional activation of *Athila*, the post-transcriptional process of RNAi generates a small interfering RNA (siRNA854) from *Athila* mRNA that causes translational inhibition of the gene *UBP1b* in a similar fashion to a microRNA or trans-acting siRNA. *UBP1b* is an RNA-binding protein that constitutes a major component of plant stress granules (SGs), sites of translational pausing. It is known in the literature that stress granules have the capacity to sequester and inhibit viral transcripts. This raises the possibility of whether *UBP1b* might act as a third host defense response to TE activity by targeting and sequestering *Athila* retrotransposon transcripts in stress granules. Therefore, this targeting of a stress granule protein by an *Athila* derived small RNA may be analogous to a virus encoding a virulence factor to disrupt the host defense during viral attack.

In order to test if *UBP1b* is inhibiting *Athila* activity, I checked the copy number, gene expression and protein production from *Athila* in wild type and 12 different mutant background combinations. This panel of 12 includes plants mutant for *ddm1* (loss of
chromatin modification and transcriptional regulation), a \textit{ddm1/rdr6} double mutant (loss of chromatin modification and disrupted post-transcriptional RNAi), the triple mutant \textit{ddm1/rdr6/ubp1b} (loss of chromatin modification, disrupted RNAi, and unable to form stress granules), as well as the necessary control genotypes. Our most convincing data to date is from my test of the accumulation of \textit{Athila} GAG protein using our newly produced GAG antibody in all the above mutants by Western. In the mutant where all the three defense responses are lost (\textit{ddm1/rdr6/ubp1b} triple mutant), there was more accumulation of GAG particles and \textit{Athila} DNA intermediates associated with gag particles than WT or any of the other double mutant combinations. This result has been independently replicated three times. I’ve also shown that the triple mutant has highest reverse transcriptase activity. These data indicate that \textit{UBP1b} and stress granule-based translational repression act as a third negative regulatory layer inhibiting \textit{Athila} retrotransposon activity, likely through sequestration of \textit{Athila} transcripts.

2.2 Introduction

2.2.1 Different levels of \textit{Athila} epigenetic regulation

Upon global activation of TEs, there are widespread shifts in the accumulation of small RNAs derived from TE transcripts. Transcriptional activation of most silenced TEs results in the loss of their corresponding siRNAs (Lippman et al. 2003)(Lister et al. 2008b) (Slotkin et al. 2009). However, some retrotransposon families, including \textit{Athila}, are unusual in the fact that they produce more siRNAs when epigenetically active than when epigenetically silenced (Mirouze et al. 2009)(Slotkin et al. 2009)(Tanurdzic et al. 2008).
The *Athila* siRNAs that increase in abundance are primarily 21–22 nt in length and are produced from the 3’ non-protein coding region downstream of the *gag* and *pol* ORFs of the consensus *Athila* element (Figure-1.2). The abundance and specific location of these siRNAs generates islands of 21–22 nt siRNAs in the genome when epigenetic silencing of *Athila* is lost (Slotkin et al. 2009). Unlike 24nt siRNAs, which direct *de novo* methylation, a **first level of Athila regulation which occurs at transcription**, 21 and 22nt siRNAs generally act post transcriptionally to direct target RNA cleavage or translational inhibition (reviewed in (Ghildiyal and Zamore 2009a)). This suggests that *Athila* is targeted for **post transcriptional repression in the absence of transcriptional regulation, a second level of Athila repression**. In addition to these two levels of *Athila* regulation mentioned above, we have discovered a third, subordinate level of *Athila* regulation that acts on *Athila* transcripts in the absence of chromatin condensation and post transcriptional gene silencing.

### 2.2.2 TE regulation of genes

TEs alter the regulation of genes in *cis* by structural polymorphisms, which are produced upon insertion or excision of a TE and this may alter a neighboring gene’s regulation (Xiao et al. 2008). In *Arabidopsis*, as well as in animals, the production of small RNAs and subsequent targeting of TEs is distinct from the production of post-transcriptionally acting gene-regulating small RNAs (reviewed in (Czech and Hannon 2011)). Most plant siRNAs are the products of RNA-dependent RNA polymerases, and cleavage of these long dsRNAs produces clusters of siRNAs from each locus. However,
the notion that only microRNAs regulate genes and endogenous siRNAs do not, is incorrect, since some inverted repeat-derived siRNAs act to regulate genes, (Dunoyer et al. 2010b) and other siRNAs act to regulate genes through the trans-acting siRNA (tasiRNA) pathway in *Arabidopsis*. This latter pathway begins with the cleavage of a non-protein coding transcript by the microRNA-loaded AGO1 or AGO7, which initiates the DCL4- and RDR6-dependent phased production of siRNAs (reviewed in (Chapman and Carrington 2007)). These siRNAs are loaded into AGO1 and regulate gene expression similar to microRNAs.

The first examples of a TE piRNA or siRNA regulating a genic mRNA in *trans* were only recently discovered in *Drosophila* and mouse (Rouget et al. 2010; Watanabe et al. 2011). In addition, recently an example of a plant viral siRNA was shown to regulate a gene (Smith, Eamens, Wang 2011). However, these examples represent exceptions to the general rule of separation between TE/viral and gene-regulating small RNAs (Xie et al. 2004). For example, there is a clear distinction between the biogenesis mechanism and target of TE siRNAs and microRNAs.

### 2.2.3 *Athila* targets a gene in *trans*

Arteaga-Vázquez *et al* demonstrated that 12 elements of the *Athila6* subfamily, each encode a small RNA, for which they predicted and provided indirect evidence that a small RNA targets a genic transcript for translational repression (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006b). They predicted that this small RNA was generated as microRNA from a stem-loop precursor transcript and determined that it was
processed by the microRNA machinery DCL1, HEN1 and HYL1. Additionally, they predicted that this microRNA, which they named microRNA854, targets the 3’ untranslated region (UTR) of the UPB1b gene, a homolog of the mammalian TIA-1 that encodes an RNA binding protein involved in the formation of stress granules (Kedersha et al. 1999)(Weber, Nover, Fauth 2008). They observed that 21 nt microRNA854 accumulates in wt vegetative tissues and found that the microRNA854-targeted UPB1b 3’UTR inhibits translation in wt plants when the 3’UTR is added to a reporter transcript. Lastly, Arteaga-Vázquez et al provided evidence that microRNA854 is highly conserved from plants to mammals. My research began by examining if Athila made microRNAs, and if the accumulation of these microRNAs is influenced by the epigenetic regulation of Athila TEs. This would be significant because if the microRNAs accumulation is dependent on epigenetic regulation of Athila then it would indicate that loss of epigenetic regulation leads to production of these microRNAs and these in turn might act to regulate host defense response.

2.2.4 UBP1b and stress granules

UPB1b, a member of the hnRNP (heterogenous nuclear RiboNucleoProtein) family, has been characterized as a major component of plant stress granule (Weber, Nover, Fauth 2008). Its mammalian orthologs, TIA-1(T-cell Intracellular Antigen 1) and TIAR (TIA-1 Related) are also defining features of stress granules in mammals (Anderson and Kedersha 2002) (Kedersha et al. 2000) (Kedersha et al. 1999).
Both UBP1b and TIA-1 proteins are nuclear-localized during normal physiological conditions (Kedersha et al. 2000) (Lambermon et al. 2000) and play a role in nuclear mRNA stability and in splicing of inefficiently spliced introns (Del Gatto-Konczak et al. 2000) (Lambermon et al. 2000). Additionally, TIA-1 and TIAR have been shown to selectively bind AU-rich elements in mRNA (Zhang et al. 2002), while both UBP1b and TIA-1 show selectivity for binding to oligouridylate-rich sequences in vitro (Dember et al. 1996) (Lambermon et al. 2000). TIA-1 and UBPIb share similar domains, as well, with 3 sequential RNA Recognition Motifs (RRMs) and 1 Prion-like Aggregation Domain (PRD) (Lorkovic and Barta 2002) (Tian et al. 1991). With induction of cellular stress, TIA-1 shuttles to the cytoplasm, mediating the formation of distinct cytoplasmic foci, known as Stress Granules (SGs) (Kedersha et al. 2000) (Kedersha et al. 1999). During cellular stress, TIA-1 proteins prevent the assembly of translation initiation complexes (Kedersha and Anderson 2002), and SG formation occurs via auto-aggregation of the C-terminal PRD of TIA-1 (Gilks et al. 2004). SGs retain some of the components of the 48S small ribosomal preinitiation complex, though they are not sites of translation during stress. Rather, they prevent the translation of mRNA bound within the SG (Kedersha and Anderson 2002). Research in animals within the last decade suggests that SGs may serve as sorting sites, where mRNAs are targeted for storage, reinitiation, or degradation by transfer to processing bodies (Beckham and Parker 2008) (Kedersha et al. 2005). Prolonged cellular stress makes these TIA-1 rich cytoplasmic foci more visible than under normal physiological conditions.
Several studies suggest that TIA-1 and SG formation function to limit a range of viral activities. These include observations in animals that some viruses block SG formation in order to translate their viral proteins (Khong and Jan 2011)(Montero et al. 2008)(White et al. 2007), and mouse embryonic fibroblasts deficient in TIA-1 display increased rates of viral replication from a wide range of RNA and DNA viruses. Some SGs have been shown to contain LINE1 retrotransposon mRNA consisting of ORF1 and ORF2, both of which are necessary for transposition. Though it is unclear if the presence of LINE1 RNA in SGs represents a host anti-TE response, it demonstrates that SGs do not only contain genic mRNA (Goodier et al. 2007a). Additionally, viral infections can induce the formation of SG like foci that contain some, but not all, components of canonical SGs and play a role in limiting the efficiency of viral replication (Piotrowska et al. 2010)(Simpson-Holley et al. 2011).

Do stress granules and the gene UBP1b act to regulate TEss?

I hypothesized that UBP1b acts to sequester Athila for translational inhibition, based on the RRM5 of UBP1b, the defined role of UBP1b in SG formation, and the role of TIA-1 in antiviral defense. This would be significant because if UBP1b acts to sequester Athila mRNAs and hence prevent their translation, then it would indicate that there is an additional third layer of Athila repression. UBP1b might cause translational repression of Athila.
2.3 Materials and Methods

2.3.1 Plant material

The mutant alleles used in this study are listed in Table 2.2. All mutants are in the Col background except ago1, Landsberg (Ler) and ubp1b, Wassilewskija (Ws) background respectively. Plants were grown under standard long day conditions at 22 °C. Inflorescence tissue was used in each experiment unless otherwise noted.

2.3.2 Quantitative Reverse Transcriptase–PCR (qRT-PCR) and qPCR

Total RNA was extracted using TRIzol reagent (Invitrogen) or the RNeasy Plant Kit (Qiagen). Total RNA was DNaseI treated and reverse transcribed using an oligo-dT primer and SuperScript III Reverse Transcriptase (Invitrogen). qRT-PCR was performed with iQ SYBR Green Supermix (BioRad Laboratories) using 3 technical replicates each of 3 or more biological replicates. qRT-PCR primers are shown in Table 2.2. qPCR reactions were annealed at 60°C unless otherwise noted. Since most standard qRT-PCR control genes are not highly expressed in pollen, the relative expression values for all experiments were calculated based on the expression of the experimentally validated control gene At1g08200. qPCR was performed on a CFX96 thermocycler and the results analyzed on the CFX Manager Software package (BioRad Laboratories). Relative expression was calculated using the ‘delta-delta method’ formula $2^{-\text{[ACP sample-ACP control]}}$, where 2 represents perfect PCR efficiency. Statistical significance was calculated using unpaired T-tests. qPCR was performed using DNA isolated using standard phenol-chloroform extraction and rest of the procedure was performed as described above.
2.3.3 Small RNA Northern blots

Total RNA was extracted using TRIzol reagent (Invitrogen), and small RNA was enriched by polyethylene glycol precipitation. The quantity of small RNA loaded in each lane ranged from 16–60 µg between blots, though the same amount of RNA was loaded per lane on each blot for comparison between samples. We accounted for the equal loading and sizes of small RNAs by re-probing our Northern blots with a known 21 nt microRNA (miR161) and/or a known 24 nt siRNA (siRNA02). In addition, our small RNA Northern blot analysis is supported by independent small RNA deep sequencing data (Nakano et al. 2006). Gel electrophoresis, blotting and cross-linking were performed as in Pall et al. (Pall et al. 2007a). Probes for siRNA854, miR161, Cen 180, 106 B and siRNA02 were generated by 5’ labeling DNA oligonucleotides with P$^{32}$-ATP, whereas the probe for Athila 3’ was generated by randomly degrading a P$^{32}$-UTP labeled in vitro transcribed RNA as in (Slotkin, Freeling, Lisch 2005). Sequences of DNA oligonucleotides and primers for generating the in vitro transcription template are listed in Table 2.2.

2.3.4 Transgene construction and transformation

The 35S:amiRNA-siRNA854 transgene was generated by cloning the sequence 5’GATGAGGATAGGGAGGAGGAGGAG into the microRNA319a stem-loop transcript as in (Schwab et al. 2006). This transcript was sub-cloned into the 35S promoter binary plasmid pB2GW7. The wt version of the UBP1b 3’UTR was amplified from the wt Col genome, and the 35S:GUS-UBP1b 3’UTR transgene was produced as in (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006b).
The modified and deleted versions of the \textit{UBP1b} 3'UTR were synthesized by IDT. The Lat52 promoter-driven GFP-\textit{UBP1b} 3'UTR transgene was constructed by cloning the either the wt\textit{UBP1b} 3'UTR, Modified \textit{UBP1b} 3'UTR or Deleted 3'UTR version into the SacI site at the end of the mGFP coding sequence of the binary plasmid pMDC107, and then by adding the Lat52 promoter to the KpnI site upstream of mGFP in these clones.

The 35S:\textit{UBP1b}-GFP transgene was generated by cloning the \textit{UBP1b} coding region into the binary plasmid pK7FWG2.

All \textit{Arabidopsis} transformations were performed using Agrobacterium strain GV3101 and standard laboratory techniques.

\subsection*{2.3.5 GUS staining and MUG assay}

T1 whole seedlings, inflorescences, leaves, siliques and immature seeds obtained from silique dissections were GUS stained in microcentrifuge tubes. T2 Whole seedlings from plates were transferred to 96 well plate for GUS staining. GUS staining was performed as in (Sundaresan et al. 1995), but for immature seed the protocol was modified. Immature seeds were stained for 1 day at 37C in GUS stain and de staining was performed for 4h in 50% Ethanol 50% acetic acid in room temp and cleared O/N in Hoyers at room temperature.

For GUS protein activity quantification, protein was quantified using the \textit{DC} assay (BioRad Laboratories), and 1 mg was used to assay the cleavage of MUG into fluorescent 4-MU as in (Jefferson 1987; Twell, Yamaguchi, McCormick 1990).
Fluorescence was measured in 96-well format with a Tecan-SpectraFluor Plus microplate reader, and the specific activity of GUS in each sample was calculated as nmol of 4-MU formed per hour per mg of protein (nmol 4-MU/h/mg). RT-PCR of the lines used in GUS staining was performed with oligo-dT primed cDNA for 28 cycles of PCR using primers listed in Table 2.2. Statistical significance was calculated using unpaired T-tests.

2.3.6 Fluorescence Microscopy

GFP fluorescence quantification was performed on a Nikon C2 confocal microscope with the NIS-Elements software suite (Nikon Corporation). GFP quantification was performed with the same microscope settings (exposure time, laser power) on the same day. Subtraction of the fluorescence of pollen grains that did not inherit the GFP transgene from the same hemizygous plant negated the background pollen auto-fluorescence. Statistical significance was calculated using unpaired T-tests.

2.3.7 Preparation of gag particle fractions and extraction of DNA from the particles

Preparation of particle fractions was carried out as described by (Gardner and Shepherd 1980; Takeda et al. 2001). *Arabidopsis* inflorescence 0.3 g was homogenized with sea sand in 0.5 ml extracted buffer [0.2 m Tris–HCl pH 7.0, 20 mm EDTA, 1.5 m urea] on ice, then in an additional 0.5 ml of the same buffer supplemented with Triton-X100 (2% final). The extracts were transferred into a 1.5 ml microcentrifuge and centrifuged at 18 000 g for 5 min at 4°C in a conventional microcentrifuge. The supernatant was centrifuged for a further 10 min, and the second supernatant was placed onto 0.25 ml of 15% sucrose, 10 mm potassium phosphate buffer pH 7.2 in a 1.5 ml microtube. Particles are purified from the supernatant by ultracentrifugation at
45 000 r.p.m. (average 109 000 g) for 90 min at 4°C in a rotor (SW 50.1; Beckmann, Fullerton, CA, USA). The pellet was rinsed with the 15% sucrose, 10 mm K-phosphate buffer pH 7.2, and centrifuged again for 20 min.

For DNA experiments, the pellet after ultracentrifugation was resuspended in 0.5 ml of 0.1 m Tris–HCl pH 7.4, 2.5 mm MgCl₂, and treated with DNaseI (10 µg ml⁻¹ final concentration) for 10 min at 37°C. EDTA (10 mm final) was added to stop the reaction, and the mixture was subsequently treated with 0.5 mg ml⁻¹ proteinase K for 10 min at 65°C in the presence of 1% SDS. DNA was extracted three times with phenol/chloroform and once with chloroform, precipitated in ethanol, and resuspended in 12 µl H₂O and used for PCR with primers indicated in Table 2.2.

2.3.8 Silver staining

For protein analysis, the pellet containing gag particles was dissolved in a mixture of 25 µl 2 × SDS dye-loading solution [125 mm Tris–HCl pH 6.8, 4% SDS, 20% glycerol, 0.01% BPB, 1.4 mβ-mercaptoethanol] and 25 µl 100 mm sodium phosphate pH 7.5, 100 mm NaCl. After boiling for 3 min, 15 µl of the protein sample was used for electrophoresis. Protein was run on 12% SDS-PAGE gel. After the separation of proteins, the gel was fixed with 30% ethanol/10% acetic acid for 1 h and rinsed twice with 30% ethanol for 10 min followed by 3 times water rinse for 10 min each time. The gel was then incubated with 100 ml of 0.1% AgNO₃ (freshly prepared). After brief rinse with water, the gel was developed for 10 min (until stain reveals) with 100 ml of 2.5% Na₂CO₃-0.02% formaldehyde (freshly prepared). Exchange the solution if it turns brownish. The reaction
was stopped with 1ml of 1% glacial acetic acid. The gel was placed on a white light box and picture was taken.

2.3.9 Silique dissection

Silique dissections were performed using dissection needle under Zeiss Stemi 2000-C stereo microscope and pictures were taken.

2.3.10 GAG protein accumulation by western

Protein was extracted from 0.2 g of fresh inflorescence tissue using 1 ml of extraction buffer (0.2 M Tris-HCl pH 7.0, 20 mM EDTA, 1.5 M Urea, and 2% Triton-X100). Quantification of soluble protein was performed by DC protein assay (BioRad). One hundred and twenty-five µg/ul of protein was mixed with 1:1 suspension buffer (100 mM sodium phosphate pH 7.5, 100 mM NaCl) and 2X SDS dye-loading solution (200 mM Tris-HCl pH 6.8, 4% SDS, 20% Glycerol, 400 mM DTT, 0.2% bromophenol blue) and boiled for 10 min. Protein was run on a 12% SDS-PAGE gel and transferred to PVDF membrane (Millipore). We raised a mouse polyclonal antibody against the Athila GAG epitope GDKAHQWEKS (Abmart), and used this at a 1:500 dilution. The ACT-11 antibody was used as a loading control and is described above.

2.3.11 Product Enhanced Reverse Transcriptase Assay (PERT)

PERT assay used for retroviruses in animals (Sastry et al. 2005) was modified for use in Arabidopsis thaliana. Fresh tissue was harvested on ice and ground in 300ul of cold protein isolation buffer (50mM KCL, 40mM Tris pH8.1, 20mM dithiothreitol DTT, 0.2% Nonidet-P40 NP-40). Protein quantification was done as described above using biorad DC assay. To 10ul of protein sample (65ug/ul), 10ul of RT cocktail (100mM
KCL, 20mM Tris pH 8.3, 11mM MgCl₂, 1mM dNTPs, 0.4uM pert cDNA primer Reverse, 0.2% NP-40, 20mM DTT, 0.8U/ul RNAse OUT, 0.314 mg/ml Salmon Sperm DNA, 0.15ng/ul MS2 RNA) was added and mixed well. The samples were placed at 48°C for 30 minutes and then placed on ice. 1ul of this reaction was used in PCR with pert RT Forward and pert cDNA Reverse primers (Table 2.2).

2.3.12 UBP1b immunoprecipitation-western

Seedlings were grown on 1/2X MS media for 11 days before their roots were harvested. The UBP1b-GFP protein was immunoprecipitated from the root tissue as follows using a commercially available polyclonal GFP antibody (AbCam). Root tissue that was crosslinked and uncrosslinked was used. Crosslinking was performed by gently immersing 0.15g of root tissue in 5ml of buffer A (0.4M Sucrose, 10mM Tris pH 8, 1mM EDTA, 1mM PMSF, 1% Formaldehyde) and vacuum infiltration for 20 minutes. The crosslinking was stopped by quenching with 0.1M glycine for 10 minutes under vacuum infiltration, and then rinse five times in sterile water. This tissue was then ground with liquid nitrogen and homogenized in 2 ml extraction buffer (100 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, and 5 mM DTT) containing 1 tablet/10 mL protease inhibitor cocktail (Roche) per gram of tissue. In a standard immunoprecipitation reaction, *Arabidopsis* protein extract was pre-cleared by incubation with 10 µl of goat anti-rabbit magnetic beads (NEB). Pre-cleared extracts were then incubated overnight with goat anti-rabbit magnetic beads pre-incubated with 1 µg α-GFP. All washes were performed with extraction buffer. Immunoprecipitated, mock-immunoprecipitated and input sample protein was isolated and run on an SDS-PAGE gel, transferred by Trans
Blot SD Semi-Dry transfer system (BioRad), and immunoblotted with α-GFP antibody (AbCam).

2.3.13 polyA Northern

polyA RNA was extracted from total RNA samples using Oligotex mRNA mini kit (Qiagen, Valentia, CA). A total of 1ug of polyA RNA was used in the northern blot. Northern blot was performed using Northern Max Gly kit (Ambion, Austin, Tx). RNA was then transferred to positively charged nylon membrane (Bright star-plus membrane) and fixed by UV crosslinking. The blot was then hybridized with ultrahyb buffer and probed using Maxiscript Invitro transcription kit (Ambion, Austin, Tx). Sense and antisense transcripts were produced using T7 and T3 RNA polymerases. P32 labelled probes for TyrAT (control gene), Athila6 gag/pol and Athila6 env were generated and used. Sequences of DNA oligonucleotides and primers for generating the in vitro transcription template are listed in Table 2.2.

2.4 Results

2.4.1 Transcriptional activation of Athila6 results in production of Athila siRNAs and siRNA854

We were unable to detect the accumulation of 21 nt microRNA854 in Arabidopsis thaliana wild-type Columbia ecotype plants (wt Col) leaf and inflorescence tissues. Due to the failure to meet multiple criteria in order to validate this small RNA as a microRNA (Meyers et al. 2008), including the biogenesis pathway of this small RNA, we have renamed this small RNA siRNA854 to avoid confusion (see below). To determine when siRNA854 accumulates, we interrogated publicly available deep sequencing small
RNA libraries (Nakano et al. 2006) and found that in wt Col, 21 nt siRNA854 does not accumulate. Only one read of 21 nt siRNA854 was detected in over 6.6 million genome-matched reads of wt Col inflorescence small RNAs combined (Table 2.1). However, when epigenetic repression of Athila6 is lost, 21 nt siRNA854 levels increase. Table 2.1 shows that, compared to the extremely low levels in wt Col inflorescence tissue, siRNA854 accumulates at high levels in met1 and ddm1 mutant inflorescences. Increased siRNA854 levels were also detected in pollen of wt Col plants, albeit to a lower level of 21 nt siRNA854 reads per million than in met1 or ddm1 mutants.

In the plant body, retrotransposons such as Athila6 are epigenetically suppressed by heritable symmetric DNA methylation and RdDM (Slotkin and Martienssen 2007a). In each case of 21 nt siRNA854 accumulation (met1 and ddm1) loss of TE epigenetic silencing is known to occur (Hirochika, Okamoto, Kakutani 2000)(Johnson, Cao, Jacobsen 2002)(Slotkin et al. 2009) (Steimer et al. 2000). To determine if the Athila6 retrotransponson is specifically activated in these mutants, I performed real-time quantitative RT-PCR (qRT-PCR) and found that in met1 and ddm1 mutants inflorescence tissue, Athila6 transcript accumulation is significantly increased compared to wt Col (Figure 2.3 A). To further examine siRNA854 accumulation, I performed small RNA Northern blots and found in wt Col, ddm1 and met1 inflorescences, a 24 nt version containing the siRNA854 sequence accumulates, while 21–22 nt versions of this sequence accumulate only in ddm1 and met1 (Figure 2.3 B), as well as in pollen (Table 2.1). 24nt siRNAs are produced in both transcriptionally silent TEs and transcriptionally active TEs epigenomes through RNA directed DNA methylation (RdDM) pathway
components such as POL4, RDR2 and DCL3. However, 21-22nt siRNAs come only from actively transcribing TEs and are produced via RNA interference (RNAi) pathway components such as POL2, RDR6, DCL2 and DCL4. I then probed this Northern blot with a 300 bp siRNA854-flanking region of *Athila6* (*Athila6* 3’ probe) and found that this region also produces other 24 and 21–22 nt siRNAs at levels comparable to those of siRNA854. These results demonstrate that the production of 21–22 nt siRNAs from this entire region is under the same epigenetic regulation as siRNA854, and combined with the results of deep sequencing of small RNAs from *ddm1*, *met1* and pollen (Lister et al. 2008a; Slotkin et al. 2009), demonstrates that siRNA854 is one member of a larger region of siRNA production. Our data refutes previous data that characterized a specific microRNA produced from this region of the *Athila6* retrotransposon (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006b).

The phenotype of *ddm1* mutant plants becomes more severe in progressive generations. Second generation homozygotes for the recessive *ddm1*-2 allele (*ddm1* F2) display little to no morphological phenotype, while after propagation as a homozygote for four additional generations (*ddm1* F6), leaf size and infertility phenotypes are severe (Kakutani et al. 1996). Figure 2.3 C shows that increasing transcript accumulation of the *Athila6* retrotransposon is associated with the progression of *ddm1* from the F2 to F6 generation. To determine if the different transcript levels of *Athila6* directly correlate with the accumulation of 21–22 nt siRNA854 and flanking 21–22 nt *Athila6* 3’ siRNAs, I examined siRNA854 accumulation by Northern blot in *ddm1* F2 and F6 individuals. F6 *ddm1* individuals produce increased levels of siRNA854 and other *Athila6* 3’ siRNAs.
compared to F2 generation *ddml* individuals (Figure 2.3 D). These data, together with the transcript accumulation and siRNA accumulation in *metl* and pollen (Figure 2.3 A and B, Table 2.1), suggests that the epigenetic activation and level of *Athila6* steady-state transcripts directly and positively correlates with the accumulation level of *Athila6* 21–22 nt siRNAs, including siRNA854.

### 2.4.2 siRNA854 accumulation represses reporter gene transcripts with the *UBP1b* 3'UTR in seedlings, leaves and inflorescences tissues

Arteaga-Vázquez et al predicted that microRNA854 (now known as siRNA854, see section 2.4.1 above), targets the 3' untranslated region (UTR) of the *UBP1b* gene, a homolog of the mammalian TIA-1 that encodes an RNA binding protein involved in the formation of stress granules (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006b)(Kedersha et al. 1999)(Weber, Nover, Fauth 2008). They observed that siRNA854 accumulates in wt vegetative tissues and found that the siRNA854-targeted *UBP1b* 3'UTR inhibits translation in wt plants when the 3'UTR is added to a reporter transcript. I wanted to determine whether the increased levels of endogenous 21–22 nt siRNA854 observed when *Athila6* is epigenetically activated in *ddml* mutants (Figure 2.3 B and D) can alter the regulation of UBP1b. In order to determine this, I transformed wt Col and *ddml* plants with the β-glucuronidase (*GUS*) reporter gene fused to the 3'UTR of *UBP1b* under the control of the constitutive cauliflower mosaic virus 35S promoter. I examined GUS expression in transgenic individuals harboring either the 35S: *GUS-UBP1b* 3'UTR construct or a 35S: GUS construct (with no 3'UTR) used as a control. GUS activity was monitored qualitatively by GUS staining of T1 (first
generation transgenic plant) seedlings, leaves, inflorescences and siliques and also T2 (progeny of T1) seedlings shown in Figure 2.4 A and B. Wt Col T1 seedlings with UBP1b 3′ UTR transgene show less GUS activity (mostly seen in roots) than the control 35S: GUS transgene which may be due to low levels of siRNA854. However, Wt Col T1 inflorescences, siliques and leaves do not display a difference in GUS activity between control and the UBP1b 3′ UTR transgene. This is because of negligible Athila 3’ expression (Figure 2.3 A and B) and siRNA854 production (Table 2.1 and Figure 2.3 B, D) in inflorescences and leaves. In ddm1 mutants, due to increased production of siRNA854 upon transcriptional activation of Athila6 the GUS activity of the UBP1b 3′ UTR transgene is reduced compared to the control transgene in seedlings, leaves and inflorescences tissues. Alternatively, in siliques, GUS activity of UBP1b 3′ UTR transgene when compared to control transgene seems to be only slightly reduced.

Progeny of T1 plants were grown to seedling stage and GUS staining was performed in T2 seedlings in 96 well plate. Nearly 85% of T2 seedlings showed significantly less GUS activity of the UBP1b 3′UTR transgene in ddm1 than control Col T2 seedlings with the same transgene. From this combined analysis of GUS staining pattern, I conclude that for inflorescence, leaf and slique tissue in a wt epigenetically silenced background; Athila does not produce 21 and 22nt siRNAs and therefore UBP1b is neither affected nor regulated by Athila. In seedling tissue, UBP1b regulation occurs in wt background due to low levels of Athila6 3’ expression and siRNA854 production. When epigenetic silencing is removed from Athila in a ddm1 mutant, the TE is transcribed and produces 21and 22nt siRNAs. One of these Athila siRNAs, siRNA854, regulates the UBP1b in trans.
To determine if the increased levels of endogenous 21–22 nt siRNA854 observed when Athila6 is transcriptionally active can regulate the UBP1b 3’UTR in T3 (progeny of T2 plants from Figure 2.4 B), I assayed GUS activity in plants homozygous (T3 generation) for the transgenes used in Figure 2.4 and found that in wt Col, the presence of the UBP1b 3’UTR did not affect GUS activity (Figure 2.5 B). In contrast, when this same analysis was previously published, the same GUS-UBP1b3’UTR transgene in a wt Col plant resulted in little to no GUS protein production in leaves and inflorescences (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006b). My data, which demonstrates no inhibition of the UBP1b 3’UTR in wt Col siliques, leaves and inflorescences (Figure 2.4 A and Figure 2.5 B), is supported by the fact that there is no 21/22 nt siRNA854 detected in leaves or inflorescence by either Northern blot (Figure 2.3 B and D) or small RNA deep sequencing (1 read in a combined 6.6 million) (Table 2.1). In ddm1 mutants, the GUS-UBP1b 3’UTR and GUS control (no 3’UTR) transgenes both display reduced GUS activity (Figure 2.5 B). It remains enigmatic why the constitutively expressed GUS transgene without a 3’UTR has reduced expression in ddm1 compared to wt Col. However, the presence of the UBP1b 3’UTR resulted in a significant reduction of GUS activity compared to the no 3’UTR control in ddm1 (Figure 2.5 B). To make sure that position effects of these transgene insertions were not the cause of this differential regulation, we crossed a wt Col plant with the UBP1b 3’UTR transgene that displayed high GUS activity to a ddm1 homozygote, and the resulting heterozygote has Athila6 3’siRNAs and hence siRNA854 accumulation as shown in the small RNA northern blot in Figure 2.5 E. The GUS activity in this ddm1 heterozygote is significantly reduced
compared to both the wt Col homozygous GUS transgene parent and to wt Col plants heterozygous for the same GUS transgene (Figure 2.5 C).

The 21 nt version of siRNA854 was previously predicted to target the 3′UTR of the *UBP1b* gene in four locations using modified criteria that allows for non-canonical or ‘wobble’ G:U base pairing (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006b). G:U base pairing has been demonstrated to be tolerated in microRNA target sites, even within the critical first 7 base pairs (bp) or ‘seed’ pairing region which is critical for microRNA function in animals (Didiano and Hobert 2006). However, the targeting of the 3′UTR by small RNAs was not previously shown directly, and complementarity of siRNA854 to the *UBP1b* 3′UTR relies heavily on non-canonical base pairing and lacks a strong 7 bp seed-pairing region (shown in Figure 2.6 A). To directly test if the 21 nt siRNA854 sequence has the ability to target the *UBP1b* 3′UTR, I took advantage of the fact that wt Col inflorescences do not accumulate 21–22 nt siRNA854 (Table 2.1, Figure 2.3 B and D). To directly examine the role of the siRNA854 sequence, we constructed plants constitutively expressing a GUS reporter gene with the GUS mRNA fused to the *UBP1b* 3′UTR driven by 35S promoter from Figure 2.4 and transformed these plants with artificial microRNA (amiRNA) constructs expressing the siRNA854 sequence, or an unrelated sequence as a control, from the *Arabidopsis* microRNA319a stem-loop transcript (Schwab et al. 2006). I performed quantitative assays to detect GUS activity, as well as qualitative plant staining, to demonstrate that the plants with the control amiRNA have high levels of GUS activity, while plants expressing the siRNA854 sequence from a microRNA stem-loop display significantly lower levels of GUS activity (Figure 2.5A).
For GUS protein activity quantification, I quantified protein and used 1 mg to assay the cleavage of MUG into fluorescent 4-MU by GUS protein (Jefferson 1987; Twell, Yamaguchi, McCormick 1990). Next, I measured 4-MU fluorescence and calculated the specific activity of GUS in each sample as nmol of 4-MU formed per hour per mg of protein (nmol 4-MU/h/mg). These data demonstrate that although the alignment of siRNA854 to the \textit{UBP1b} 3'UTR lacks a strong seed pairing region and relies on G:U base pairing, the 21 and/or 22 nt siRNA854 sequence can target the \textit{UBP1b} 3'UTR resulting in decreased reporter protein accumulation. Therefore, the regulation of the \textit{UBP1b} 3'UTR is sensitive to the levels of siRNA854, with either the production of this sequence as an amiRNA, or accumulation of this sequence as a siRNA in \textit{ddm1} resulting in repression of GUS activity. I determined that all of the transgenes in either wt Col or \textit{ddm1} from Figure 2.5 B have GUS transcripts that accumulate to similar levels (Figure 2.5 D), indicating that the regulation of these transgene transcripts is not due to post-transcriptional degradation of the GUS RNA, but is likely rather due to the inhibition of translation of these mRNAs.

2.4.3 siRNA854 accumulation in pollen and immature seeds regulates transgene transcripts with the \textit{UBP1b} 3'UTR

In addition to the increased levels of siRNA854 in \textit{ddm1} mutants, siRNA854 also accumulates in wt Col pollen (Table 2.1) due to natural expression of \textit{Athila}. In the wt pollen vegetative cell, there is a global loss of heterochromatin, which may be due to the lack of DDM1 in the pollen vegetative nucleus, and this results in the transcriptional activation of \textit{Athila} and other TEs (Slotkin et al. 2009). To determine if endogenous
siRNA854 accumulation in pollen is able to regulate the *UBP1b* 3'UTR, I performed similar transgene reporter assays as above in wt Col pollen. We used a pollen vegetative cell promoter to drive GFP and added one of three different 3'UTRs to these reporter transgenes (Figure 2.6 A). I quantitatively measured GFP fluorescence by subtracting the fluorescence of segregating pollen grains that did not inherit the transgene from the fluorescence of pollen grains that did inherit the transgene. With no 3'UTR, transgene protein accumulates, and a moderate level of fluorescence is observed (Figure 2.6 B). When the wt *UBP1b* 3'UTR is added to this transgene, significantly less fluorescence is observed, likely due to the accumulation of siRNA854 in wt Col pollen. To test if the binding of siRNA854 was specifically responsible for this regulation, we generated a version of the *UBP1b* 3'UTR that lacks all four of the siRNA854 predicted target sites, resulting in a shorter 3'UTR (shown in Figure 2.6 A). This deleted 3'UTR transgene (DEL transgene) resulted in significantly increased fluorescence compared to the wt *UBP1b* 3'UTR (Figure 2.6 B). We also produced a *UBP1b* 3'UTR variation of the same length as the wt *UBP1b* 3'UTR, in which each of the perfectly complementary base pairs in all four of the siRNA854 predicted target sites have been switched to bases that do not show complementarity (or G:U pairing) with siRNA854 (Figure 2.6 A). Pollen grains with the base-modified 3'UTR (MOD transgene) on the GFP transcript display significantly increased fluorescence compared to the wt *UBP1b* 3'UTR (Figure 2.6 B), demonstrating that these bases are necessary for the repressive regulation of the *UBP1b* 3'UTR. Pollen from both the MOD and DEL 3'UTR transgenes display fluorescence levels even higher than the control lacking a 3'UTR, likely due to the ability
of the *UBP1b* 3′UTR, when not targeted by small RNAs, to stabilize transcripts or promote their translation. Lastly, we transformed the GFP transgene with and without the wt *UBP1b* 3′UTR into *rdr6* mutants. We had previously found that the *Athila* 21–22 nt siRNA biogenesis is dependent on AGO1, DCL2, DCL4 and RDR6, the same pathway as the tasiRNA gene regulating pathway (A.D. McCue, data not shown)(McCue et al. 2012a). I observed that the expression of Lat52: GFP (no 3′UTR) in *rdr6* mutant pollen is higher than that of the same transgene in wt Col pollen (*Figure 2.6 B*). This difference is likely due to the role of *RDR6* in post-transcriptional silencing of transgenes (Luo and Chen 2007). I speculate that the wt Col Lat52: GFP transgene is subject to a certain low amount of post-transcriptional regulation mediated by *RDR6*. When this transgene is present in *rdr6* mutant plants, this post-transcriptional regulation does not occur, resulting in higher expression of the transgene compared to wt Col. Most importantly, I did not observe a reduction in pollen fluorescence for the Lat52: GFP-*UBP1b* 3′UTR transgene in *rdr6* compared to the no-3′UTR control transgene in *rdr6* (*Figure 2.6 B*), demonstrating that *RDR6* is necessary for the targeting of the *UBP1b* 3′UTR in pollen. The combined data from *Figure 2.5* and *Figure 2.6* demonstrate that the *RDR6*-dependent accumulation of siRNA854 and base pairing with the *UBP1b* 3′UTR target sites are required for the inhibitory regulation of *UBP1b* 3′UTR reporter genes.

To determine which Argonaute protein plays a role in repression of UBP1b in pollen, I transformed the GFP transgenes with pollen-specific expression with the *UBP1b* 3′UTR into *ago* mutants *ago1*, *ago6*, *ago8* and *ago9*. Argonaute proteins are involved in binding small interfering RNA (siRNA) fragments and have
endonuclease activity directed against messenger RNA (mRNA) strands that are complementary to their bound siRNA fragment (Mallory and Vaucheret 2010), or cause the translational inhibition of target transcripts. Arabidopsis AGO1 is involved in the regulation of gene expression through the microRNA pathway and the plant-specific trans-acting small interfering RNA (tasiRNA) pathway (Vazquez et al. 2004). In contrast to the roles of AGO1 in gene regulation, the other Argonaute proteins (Arabidopsis has ten Argonaute family proteins) AGO6, and possibly AGO9, function in RNA-directed DNA methylation (RdDM) to establish and maintain transposable element (TE) silencing (Havecker et al. 2010b; Tran et al. 2005; Zheng et al. 2007) (Zilberman, Cao, Jacobsen 2003). AGO8 is a pseudogene of no known function in Arabidopsis (Takeda et al. 2008). These proteins, AGO1, AGO6, AGO9 and AGO8 were chosen to determine the role of different Argonaute proteins in the repression of UBP1b in pollen. ago1 mutant is in Landsberg (Ler, ecotype of Arabidopsis thaliana) background and ago6, ago8 and ago9 mutants are in Col background. Statistically significant increase in GFP fluorescence (Figure 2.6 C) is seen in ago1 (p<0.0001), ago9, ago6 and ago8 mutants with the GFP: UBP1b 3’UTR transgene when compared to Col, shown in Figure 2.6 C. Due to the accumulation of siRNA854 in wt Col pollen significantly less (p<0.0001) fluorescence is observed in Col pollen with GFP: UBP1b 3’UTR than in Col pollen with only GFP. The ideal control of Ler, GFP: UBP1b 3’UTR was not generated to compare the fluorescence levels of ago1, GFP: UBP1b 3’UTR and represents a way this experiment could be improved in the future. While each of these four ago mutants show statistically significant increases compared to control, the increase of ago6 was
significantly higher (p<0.0001) than \textit{ago1}, \textit{ago8} and \textit{ago9} mutants. This suggests that AGO6 has the largest role in the repression of the \textit{UBP1b} 3’ UTR. It is surprising to observe AGO6 role in UBP1b repression in pollen because it is known in literature that AGO6 binds 24nt siRNAs and functions to keep repetitive regions of the Arabidopsis genome in a transcriptionally silenced state (Eun et al. 2011; Olmedo-Monfil et al. 2010a). We have shown in the lab that AGO6 binds to 21-22nt siRNAs as well (Unpublished data, see Chapter-3). This suggests that AGO6 plays a role in \textit{UBP1b} repression through binding to 21-22nt siRNA854.

Lastly, I wanted to examine the repression of \textit{UBP1b} 3’UTR by siRNA854 in immature seeds obtained from crossing Col, 35S: GUS-\textit{UBP1b} 3’UTR as a female parent to wt Col and \textit{rdr6} as male parent respectively. Through these crosses, I wanted to determine if \textit{Athila} also plays a role in \textit{UBP1b} repression in immature seeds through siRNA854 produced via RDR6. I isolated immature seeds from the resulting siliques of the crosses and monitored GUS activity by qualitative GUS staining. If RDR6 plays a role in UBP1b repression then in Col, 35S: GUS-\textit{UBP1b} 3’UTR X Col cross due to the presence of 21-22nt siRNA854 in wt Col pollen, I expect that immature seeds will be stained very light blue to no color whereas in Col, 35S: GUS-\textit{UBP1b} 3’UTR X \textit{rdr6} cross I expect that that immature seeds will be stained dark blue due to lack of siRNA854 production in \textit{rdr6} pollen. Surprisingly, immature seeds from Col, 35S: GUS-\textit{UBP1b} 3’UTR X Col cross were mostly stained dark blue whereas immature seeds from Col, 35S: GUS-\textit{UBP1b} 3’UTR X \textit{rdr6} cross were mostly stained light blue to no color Figure \textbf{2.7}. Out of 50 immature seeds from Col, 35S: GUS-\textit{UBP1b} 3’UTR X Col cross 27 are
stained dark blue and 23 are stained light blue. Out of 73 immature seeds from Col, 35S: GUS-UBP1b 3’UTR X rdr6 cross 30 are stained very light blue and 40 showed no color. The graph in the left of Figure 2.7 shows the number of immature seeds stained as dark blue, light blue and no color for both the Col, 35S: GUS-UBP1b 3’UTR X Col and Col, 35S: GUS-UBP1b 3’UTR X rdr6 crosses. From this data I conclude that RDR6 does not regulate the UBP1b 3’UTR in seeds, and in seeds the UBP1b repression is following a different mechanism than what is observed with other tissues such as inflorescences and pollen.

2.4.4 Effect of UBP1b on Athila copy number, expression and translation

To determine why TE 21 nt siRNAs regulate some genes, we focused on the critical unresolved question of whether the regulation of genes by TE siRNAs is simply due to random partial complementarity between siRNAs and genes, or if this regulation imparts some advantage to the host cell or TE that may be evolutionarily selected for over time (McCue and Slotkin 2012). To address this question, we used the Athila6-derived siRNA854 and its target UBP1b genic mRNA. In plants, UBP1b protein is located in the nuclei of unstressed cells, and translocates to punctate cytoplasmic bodies in stress conditions (such as dark grown or osmotic stress conditions) or during the loss of heterochromatic control and TE activation in ddm1 mutants (McCue et al. 2012a). Therefore, we hypothesized that UBP1b, and potentially stress granules, may form due to TE activation and act to translationally inhibit TE activity, and that Athila6 has retained the siRNA854 sequence as a mechanism to suppress the host UBP1b-based translational silencing of TEs (McCue and Slotkin 2012).
In order to test UBP1b’s role in Athila repression, first I checked the effect of UBP1b on the copy number and expression of Athila. I produced a panel of single, double and triple mutants with various combinations of the ddm1, rdr6, and ubp1b mutations. These three proteins DDM1, RDR6 and UBP1b, are representatives of different levels of Athila regulation and are responsible for Athila silencing in a WT background. In the ddm1 mutant, epigenetic activation of Athila occurs that leads to loss of chromatin modification (first level of Athila regulation), in rdr6 mutant RNAi is disrupted (second level of Athila regulation) and in ubp1b mutant there is loss of stress granule protein. To test if UBP1b plays a role at the third level of Athila regulation, I made double mutants removing both the transcriptional regulation (ddm1) and RNAi (rdr6) and a triple mutant ddm1rdr6/ubp1b that has loss of chromatin modification, disrupted RNAi, and inability to form stress granules. I performed qPCR to test the Athila6 gag/pol protein coding region copy number in wt and 12 different mutant background combinations with ddm1, rdr6, and ubp1b mutations. ubp1b mutation is in Wassilewskija (Ws, ecotype of Arabidopsis thaliana). ubp1b insertion mutant allele results in a lack of polyadenylated transcripts, although un-spliced and non-polyadenylated transcripts are still produced(McCue et al. 2012a). Interestingly, the triple mutant has more Athila copies than wt (p<0.05) as shown in Figure 2.8 A. Both chromosomal DNA copies as well as extra-chromosomal cDNA copies of transposing Athila elements are detected by this qPCR. This shows that when all the three levels of Athila regulation are lost, there is increase in Athila DNA.
To determine the effect of *UBP1b* on *Athila* expression, I performed qRT-PCR of the *Athila6* *gag/pol* protein coding region. In the *ddm1/rdr6* double mutant (loss of chromatin modification and disrupted post-transcriptional RNAi), plants exhibit higher *Athila6* transcript accumulation than the single *ddm1* mutant (Figure 2.8 B). The transcript level is highest in *ddm1/rdr6* because these transcripts aren’t getting degraded by RNAi, and hence they are likely being sequestered by UPB1b. In *ubp1b* mutants the steady-state transcript level likely goes down as more of these transcripts are translated and turned over.

Since UBP1b seems to have an effect on *Athila* copy number, next I wanted to determine the role of UBP1b in *Athila* translation. In the retrotransposon replication cycle, the process subsequent to RNA transcription is the translation of Gag and Pol products and the formation of Gag particles (Takeda et al. 2001). It is known in literature that Gag particles of retrotransposons of yeast and animals are purified by ultracentrifugation through a sucrose density gradient (Eichinger and Boeke 1988; Shiba and Saigo 1983). I prepared a putative Gag-particle fraction of *Athila* by a similar method, which was originally developed for isolating DNA from particles of cauliflower mosaic virus(Gardner and Shepherd 1980). I ran this Gag-particle fraction in SDS-PAGE gel and then performed silver staining to detect the protein bands in the panel of single, double and triple mutants with various combinations of the *ddm1, rdr6*, and *upb1b* mutations which represent loss of chromatin modification, disrupted RNAi, and inability to form stress granules respectively (Figure 2.9). Since *Athila* is silenced in *wt Col and Ws*, there is no gag protein accumulation. The same is true for single mutants
such as \textit{ddm1} 2\textsuperscript{nd} generation, \textit{rdr6} and \textit{ubp1b} because there is only loss of one level of \textit{Athila} regulation and other two levels act to keep \textit{Athila} in its silent state in these single mutants. However, in \textit{ddm1} 6\textsuperscript{th} generation there’s higher expression of \textit{Athila} than \textit{ddm1} 2\textsuperscript{nd} generation as shown in Figure 2.3 C. This increase in expression leads to accumulation of \textit{Athila} gag particles which correspond to protein bands in the range of 28-32kDa. In contrast to \textit{ddm1} 6\textsuperscript{th} generation mutant, \textit{ddm1/rdr6} double mutant (loss of chromatin modification and disrupted RNAi) shows less gag protein accumulation in spite of high \textit{Athila6} gag/pol transcript expression (Figures 2.8 B and 2.9). This suggests that UBP1b may be playing a role in sequestering \textit{Athila} transcripts and inhibiting their translation. Both \textit{ddm1/ubp1b} double mutant (loss of chromatin modification and unable to form stress granules) and the triple mutant \textit{ddm1/rdr6/ubp1b} (loss of chromatin modification, disrupted RNAi, and unable to form stress granules) 28-32kDa protein bands are detected, which are consistent with gag polypeptide region. From this data, I conclude that UBP1b has effect on copy number and gag encapsidation of \textit{Athila}. This suggests that UBP1b may act at the third level of \textit{Athila} regulation in translation.

2.4.5 Loss of different levels of \textit{Athila} regulation leads to sterility and defective centromeres

Next I wanted to determine the effect of loss of the three layers of defenses against \textit{Athila} in the function of host organism, \textit{Arabidopsis thaliana}. I looked at the phenotype of single, double and triple mutants with various combinations of the \textit{ddm1}, \textit{rdr6}, and \textit{upb1b} mutations. I wanted to see if the loss of \textit{Athila} regulation has any effect on fertility of the plants. Hence I looked at whole siliques from the panel of
mutants described above. Mutant siliques varied in length when compared to wt as shown in Figure 2.10 A. Siliques from double and triple mutants with epigenetic activation of *Athila* have stunted growth. I then performed dissection of the whole siliques (Figure 2.10 B). Dissected siliques from *ddm1* and *rdr6* are organized and fertile as in wt whereas *upb1b* has disorganized siliques but produce fertile seeds. In *ddm1/rdr6* and *ddm1/upb1b* double mutants, where there is loss of two levels of *Athila* regulation, very few seeds are produced. In *ddm1het/upb1b* double mutant, the seed production is increased when the function of transcriptional regulation of *Athila* is partially restored. Dissected siliques of the triple mutant have severe sterility with no seed production. Therefore loss of different levels of epigenetic regulation of *Athila* affects seed formation in siliques and leads to sterility. How this genotype translates to the seed sterility phenotype is currently unknown, however *Athila* and other TE mobilization and regulation of genes likely play a role.

It has been shown in the literature that when there is loss of small RNA-induced heterochromatin, the centromere function is perturbed, resulting in lagging chromosomes and potential sterility (Volpe et al. 2003). Since silique dissections of mutants that reactivate *Athila* revealed sterility, another interesting question to study would be to check if these defects are caused by TE transposition or due to improper centromere function. We were not able to detect any successful transposition of *Athila* even in the triple mutant *ddm1/rdr6/upb1b*, which loses all the three levels of *Athila* regulation when we performed transposable element display (A.D. McCue, unpublished data). This may be because there may be some other level of regulation or mechanism which is
preventing the movement of *Athila* even though the regulation at transcriptional, post-transcriptional and translational level is lost.

Next, I wanted to test if the improper centromere function in the different mutant combinations affecting the different levels of *Athila* regulation is responsible for the sterility defects seen in these mutants. In plant genomes retrotransposons are the building blocks of heterochromatin. Figure 2.11 A displays the accumulation of *gypsy* family LTR retrotransposons and DNA transposons on a stereotypical plant chromosome. It is interesting to note the massive accumulation of both retrotransposons and DNA transposons at the centromere, although the exact location of their accumulation differs between the centromere core and the peripheral regions of the centromere (the pericentromere)(Arabidopsis Genome Initiative 2000). As mentioned above, plant genomes often use *gypsy* family LTR retrotransposons and alpha satellite repeats as the building blocks of the centromere core (Figure 2.11 B). The centromeric core regions are large assemblies of thousands of short (approximately 151 to 340 base pairs [bp]) satellite repeats in head-to-tail orientation with interspersed retroelements(Hall, Kettler, Preuss 2003; Thompson, Schmidt, Dean 1996). In *Arabidopsis thaliana*, these comprise 177- to 179-bp satellite repeats known as Cen180. 106B repeats are 398-bp internal portions of *Athila2* LTRs that are interspersed among the Cen180 arrays and act as promoters for Cen180 transcription (Slotkin 2010)(May et al. 2005).

Improper centromere function leads to expression of Cen180 alpha satellite repeats which are usually kept silent in a normal centromere(May et al. 2005). In order to test whether centromere is affected in single and double mutants of different levels of
Athila silencing, I performed RT-PCR to check the expression of 106B or Athila2, Athila6 and Cen180 (Figure 2.11 C). 106B and Athila6 are not expressed in wt controls Col, Ws and single mutants such as rdr6 and ubp1b. I found that 106B or Athila2 and Athila6 expression is high in ddm1, ddm1/rdr6 and ddm1/ubp1b mutants than met1. Cen180 expression is higher in ddm1 than in Col, Ws, met1, rdr6 and ubp1b. Cen180 expression is increased in ddm1/rdr6 and ddm1/ubp1b double mutants than in ddm1 single mutants. Improper centromere function in ddm1 gets worse with ddm1/rdr6 and ddm1/ubp1b double mutants. This misregulation of centromere and Athila expression that occurs when two levels of Athila silencing either transcriptional and post transcriptional or transcriptional and translational are lost may be responsible for the sterility defects observed in ddm1/rdr6 and ddm1/ubp1b double mutants as shown in Figure 2.10 B. Figure 2.8 B shows that Athila is expressed in the triple mutant ddm1/ubp1b/rdr6 where all the three levels of Athila regulation are lost but I have not yet looked at the Cen180 expression in this triple mutant. Hence the reason for severe sterility seen in the triple mutant (Figure 2.10 B) is still unknown.

Next, I wanted to determine if the pathway that is essential for the processing of Athila siRNAs and the pathway for Cen180 and 106B siRNAs production is the same. I performed small RNA northern blot and detected 24 nt Cen180 and 106B siRNAs (Figure 2.11 D). 24nt siRNAs from Cen180 and 106B are produced in both TEs transcriptionally silent and transcriptionally active epigenomes through RNA directed DNA methylation (RdDM) pathway components such as POL4, RDR2 and DCL3. This is distinct from the biogenesis of Athila siRNAs which involves RDR6, DCL4 and AGO1. siRNAs
processed from *Cen180* and *106B* follow heterochromatin silencing pathway. From this data, I conclude that loss of two different layers of *Athila* regulation (*ddm1/rdr6* and *ddm1/ubp1b*) leads to sterility defects which may be due to improper centromeric function that is characterized by the expression and siRNAs production of Cen180 repeats and 106B.

### 2.4.6 UBP1b causes translational inhibition of *Athila*

To test the role of UBP1b in Athila transcript accumulation, I performed polyA northern blot with polyA RNA extracted from Col, *rdr6, ddm1, ubp1b* in Ws, *ddm1/rdr6* and *ddm1/ubp1b* (Col/Ws). I probed polyA northern with *Athila6 gag/pol* and *Env T7* probes which detects sense transcripts. I probed with TyrAT T7 control probe. I created *ddm1/ubp1b* double mutant by crossing *ddm1* in Col background to *ubp1b* in Ws background. Hence *ddm1/ubp1b* (Col/Ws) is a hybrid made up of two mixed backgrounds. In Figure 2.12, *ddm1* mutants have many different sized *Athila6 gag/pol* and *Athila6 env* sense transcripts, whereas *rdr6* mutants have no *Athila6 gag/pol* or *Athila6 env* sense transcripts, which is due to the hierarchical repression of *Athila6* silencing with DDM1 above RDR6. The double mutants, *ddm1/rdr6* (loss of chromatin modification and disrupted post-transcriptional RNAi) and *ddm1/ubp1b* (loss of chromatin modification and unable to form stress granules) have additional different sized *Athila gag/pol* and *env* transcripts than found in *ddm1*. These additional transcripts in *ddm1/rdr6* and *ddm1/ubp1b* could represent a synergistic effect of the double mutant. In *ddm1/ubp1b*, these additional transcripts might also be due to the effect of two mixed backgrounds. On the other hand, these additional transcripts could represent a change in
the post-transcriptional processing and possibly the stress granule sequestration of *Athila6* elements already expressed in *ddm1* mutants due to loss of RDR6 and UBP1b respectively. We have shown in the lab that *Athila* gag/pol and env transcripts seen in *ddm1* and *ddm1/rdr6* are coming from the same *Athila* elements. We are yet to determine if there is difference in the generation of *Athila* transcripts between *ddm1* and *ddm1/ubp1b* mutants. Thus, inability to either perform RNAi or form stress granules in epigenetically active background leads to new *Athila* transcripts suggesting that UBP1b might play a role in processing or sequestration of *Athila* mRNA. To further determine if UPB1b regulates *Athila* transcripts, I generated a more controlled full mutant panel of all combinations of *ddm1*, *rdr6* and *ubp1b* mutants and used direct assays to investigate *Athila* RNA location, translation and reverse transcription.

Gag particles of retrotransposons, in which reverse transcription and second-strand DNA synthesis take place, contain translation products, RNA intermediates and linear DNA intermediate (Takeda et al. 2001). I previously preliminarily showed in Figure 2.9 that gag particles accumulate when all the three levels of *Athila* regulation are lost. Next, I wanted to find out whether these gag particles contain *Athila6* DNA intermediates. I isolated leaf DNA from gag particles from the triple and single mutants that are defective in different levels of *Athila* regulation and performed PCR to amplify the *Athila6* CDS (Figure 2.13 A). I used the Tyrosine amino transferase (TyrAT), a single copy gene and Cen180, a multi copy gene as negative control since gag particles will not contain TyrAT and Cen180 DNA. I used genomic DNA isolated from wt Col as positive control for the PCR reaction. Positive control of the PCR reaction worked for all the three
PCR targets. TyrAT negative control worked and no bands were seen in any of the genotyped tested. However, Cen180 didn’t amplify for Col, ddm1 F6 but amplified in rdr6 and ddm1/rdr6/ubp1b. Since the multi copy negative control gene didn’t work for few genotypes it is difficult to conclude about the presence or absence of Athila6 DNA in gag particles isolated from the triple and single mutants that are defective in different levels of Athila silencing. Then, I performed gag particle purification using ultra centrifugation on inflorescence tissue from the panel of single, double and triple mutants that have lost different layers of Athila regulation (Figure 2.13 B). I isolated DNA from the gag particles and performed PCR to amplify Athila6, Cen180 and TyrAT genes. TyrAT and Cen180 genes showed non-specific amplification in all the genotypes. In order to eliminate this non-specific TyrAT and Cen180 negative control genes amplification, I resuspended the pellet containing gag particles with DNAse-I buffer and treated the gag particles with DNAse I. Theoretically, DNAse I treatment will not remove Athila6 DNA because they are enclosed within gag particles. From these DNAse I treated gag particle containing pellets, I isolated DNA and then performed PCR to amplify the target genes.

Even after DNAse I treatment, I still observed non specific amplification of TyrAT and Cen180 negative control genes. Athila6 DNA amplified in all of the genotypes tested. Since both the negative control genes showed amplification, it is still unclear whether Athila6 DNA intermediates are present or absent in gag particles isolated from various mutants that are defective in different levels of Athila repression.
To investigate the role of UBP1b in silencing *Athila* at translational level, I wanted to determine if *Athila6* gag protein levels increase in *ubp1b* mutants or not. This would indicate whether the wt UBP1b protein acts to repress *Athila* gag protein levels or not. We raised a mouse polyclonal antibody against the *Athila* GAG epitope GDKAHQWEKS. I screened 11 ascite lines of the GAG antibody and found the ascite line that worked the best. Then I sent this ascite line to Abmart for purification. I used this antibody for Westerns using protein extracted from inflorescence tissue in a panel of single, double and triple mutants with various combinations of the *ddm1*, *rdr6*, and *ubp1b* mutations which represent loss of chromatin modification, disrupted RNAi, and inability to form stress granules respectively that I produced. A representative Western of three independent experimental and biological replicates is shown in Figure 2.14 A, with the levels and variation of GAG accumulation normalized to Actin across these biological replicates shown in Figure 2.14 B. In wt Col or *rdr6* single mutant plants, *Athila6* is transcriptionally silenced (McCue et al. 2012a) and the level of GAG protein is virtually undetectable. In *ddm1* F2 or F6 single mutants, *Athila6* is transcriptionally activated, and the levels of GAG protein increase 4-fold. I next investigated *ubp1b* mutants, which are not in the reference Col background as are the *ddm1* and *rdr6* mutations, but rather in a WS background(McCue et al. 2012a). I tested this *ubp1b* mutant in WS, as well after introgressing the *ubp1b* mutation into Col over six generations. Some individual bioreps of any genotype with a WS background show low levels of GAG protein accumulation (Figure 2.14 A), but over the course of the three biological replicates the level of GAG protein does not show significant changes in
wt WS or upb1b single mutants (in Col or WS backgrounds) compared to wt Col (Figure 2.14 B). In rdr6/ubp1b double mutants (which have a mixed Col/WS background), Athila6 TEs are still transcriptionally silenced, and GAG protein does not accumulate. In ddm1/ubp1b mutants, Athila6 elements are transcriptionally active (due to the ddm1 mutation), but GAG protein levels decrease compared to ddm1 single mutants, and we are unable to conclude if this decrease is due to the upb1b mutation or the mixed Col/WS background. As a control for this mixed background, I generated two types of ddm1/rdr6 double mutants in which TEs are transcriptionally active but these transcripts are not degraded by RNAi via RDR6. One is in a pure Col background, and one is in a Col/WS mixed background. I found that the background does play a role in the accumulation of GAG protein, as the mixed background plant accumulates higher GAG protein levels compared to the pure Col background. Most importantly, in each of the three biological replicates, the ddm1/rdr6/upb1b triple mutant displayed the highest level of GAG protein accumulation (Figure 2.14 A and B). I observed a 2-fold increase in GAG protein accumulation compared to the mixed-background ddm1/rdr6 double mutant (p < 0.05). These data demonstrate that the UBP1b stress granule protein plays a role in repressing Athila6 GAG protein production when the TE is transcriptionally activated (in a ddm1 mutant) and RNAi is abolished (in the rdr6 mutant). Therefore, UBP1b serves as a third line of defense in the cell against the activity of TEs, and is only required once transcriptional and post-transcriptional silencing have failed (or at least I can only observe UBP1b’s effect once transcriptional and post-transcriptional silencing are abolished). I thus conclude that the inhibition of UPB1b protein production by the Athila-
derived 21–22 nt siRNA854 acts as a suppressor of a host TE silencing mechanism, demonstrating a likely derived function for an individual epigenetically regulated TE siRNA.

*Athila* possesses a gag/pol ORF (Open Reading Frame) encoding the proteins requires for its life cycle such as the Gag, encoding the structural protein involved in nucleocapsid formation; the Pol, specifying the activities for the reverse transcription and integration of new copies (Sabot and Schulman 2006b). Since I found that UBP1b causes translational repression of *Athila* by inhibiting the formation of gag particles, I wanted to know if UBP1b also has an effect on reverse transcriptase, one of the translational products of *Athila*. This would provide a second line of evidence that UPB1b represses *Athila* activity. Hence I performed Product Enhanced Reverse Transcriptase Assay (PERT) modified from the method used in retroviruses to assay reverse transcriptase activity (Sastry et al. 2005). Figure 2.15 A shows schematic of PERT assay, which is performed to determine the endogenous reverse transcription activity in a cell or tissue sample. I performed PERT by first isolating protein from different tissues from the panel of mutants defective in three modes of *Athila* repression and then adding the normalized protein to a mixture containing buffer, dNTPs, primers and MS2 bacteriophage RNA. The samples are then placed in thermocycler under conditions for RT to produce cDNA from MS2 RNA and then are amplified by PCR. The PCR products are run in gel electrophoresis and are shown in Figure 2.15 B and C. If a particular genotype tested has endogenous reverse transcriptase activity then the protein isolated from the genotype will be able to convert MS2 RNA to MS2 cDNA, which will amplify in PCR. I used protein
isolated from Col spiked with ssIII Reverse Transcriptase (RT) and no protein control spiked with ssIII RT as positive controls and both these controls showed amplification of MS2 cDNA Figure 2.15 B and C. I used no cDNA template and no protein as negative controls and both these controls didn’t show any amplification for MS2 cDNA. Inflorescence, leaf and silique tissues of the triple ddm1/rdr6/ubp1b mutant (all three modes of Athila repression are absent), exhibit endogenous reverse transcriptase activity (Figure 2.15 B). In leaf tissue, I observed slightly higher endogenous reverse transcriptase activity in ddm1/rdr6/ubp1b triple mutant than in rdr6 and ddm1/ubp1b mutants. I didn’t detect any reverse transcriptase activity in Col, Ws, ddm1, ubp1b and ddm1/rdr6 leaf tissues. In inflorescence tissue, I observed slightly higher endogenous reverse transcriptase activity in ddm1/rdr6/ubp1b triple mutant than in WS, ubp1b, ddm1/rdr6 and ddm1/ubp1b mutants. I didn’t detect any reverse transcriptase activity in Col, ddm1 and rdr6 inflorescence tissues. In order to verify the results obtained with inflorescence tissue shown in Figure 2.15 B, I performed biological replicate of PERT assay (Figure 2.15 C) from the panel of mutants defective in three modes of Athila repression. I observed endogenous reverse transcriptase activity in ddm1 F6, ubp1b, ubp1b in Col, rdr6/ubp1b, ddm1/rdr6/UBP1b in Ws and ddm1/rdr6/ubp1b mutants. I didn’t detect any reverse transcriptase activity in Col, ddm1 F2, rdr6, Ws, ddm1/rdr6 and ddm1/ubp1b mutants. It is difficult to infer anything about the effect of different levels of Athila silencing on endogenous reverse transcriptase activity of Athila because the results of both the biological replicates of PERT experiment on inflorescence tissue differ
from each other. Thus, I conclude that UBP1b affects the translation of *Athila* gag particles but its effect on reverse transcriptase activity of *Athila* is still unclear.

In animals, stress granules have been shown to contain LINE1 retrotransposon mRNA consisting of ORF1 and ORF2, both of which are necessary for transposition (Goodier et al. 2007a). I wanted to determine if UBP1b sequesters *Athila* in stress granules similar to mammalian stress granules that sequester LINE1 retrotransposons. I first transformed Col and *ddm1* plants with *UBP1b* driven by a 35S constitutive promoter translationally fused to GFP (35S:*UBP1b*-GFP). I harvested seedling roots and performed Immuno Precipitation-Western (IP-Western) with crosslinked and uncrosslinked tissues. I used seedlings roots for UBP1b immunoprecipitation because we have visualized UBP1b-GFP protein stress granule fluorescence in seedling roots tissue through fluorescence microscopy (McCue et al. 2012a). I found that it was better to use crosslinked tissues than uncrosslinked tissues for IP-Western because non-specific bands were seen with uncrosslinked tissues (Figure 2.16). I observed increase in immunoprecipitated UBP1b-GFP protein levels in *ddm1* than in Col when I performed IP-Western on the crosslinked tissues. Therefore in *ddm1* mutant plants where Athila is epigenetically activated there is increase in UBP1b protein than in epigenetically silent background Col plants. In order to determine UBP1b’s role in sequestration of Athila transcripts in stress granules, I performed immuno precipitation again with crosslinked seedling roots from Col and *ddm1* containing 35S:*UBP1b*-GFP transgene and isolated RNA after immune precipitation from Input (no IP control), mock (no antibody control) and IP samples. I then made cDNA from this RNA using random
hexamers and performed RT-PCR to amplify for Athila6 CDS and TyrAT. This experiment was unsuccessful. I didn’t see any amplification of Athila in IP samples. From this data I conclude that more stress granules accumulate when Athila is in transcriptionally active state than in transcriptionally silent state. It is yet unknown whether these stress granules are capable of sequestering Athila and other transposable elements or not.

2.5 Discussion

2.5.1 UBPIb acts as third line of defense against Athila

The plant TIA-1 and TIAR homolog, UBPIb, a member of the hnRNP (heterogeneous nuclear RibonucleoProtein) family, has been characterized as a major component of plant stress granule (Weber, Nover, Fauth 2008) (Aparicio et al. 2010; Kedersha et al. 2000). In animals, TIA-1, acts to repress viral translation, and hence prevents viral replication (Aparicio et al. 2010; Kedersha et al. 2000). Animal stress granules have been shown to contain LINE-1 retrotransposon mRNA(Goodier et al. 2007a; Goodier et al. 2010). Plant stress granule’s role in translational repression of viruses and transposons is poorly studied. I examined different levels of repression of Athila, LTR retrotransposon in Arabidopsis, each operating at a different stage of Athila life cycle. First, transcriptional regulation dependent on DNA methylation epigenetically silences Athila. Second, when transcriptionally active, Athila mRNA accumulation is inhibited by the post-transcriptional regulation mediated by the tasiRNA/VIGS siRNA pathway components DCL2, DCL4, RDR6 and AGO1. It is known in literature that when TEs are activated they elicit stress response(Arnault and Dufournel 1994; Capy et al.
I observed that when first layer of repression at transcriptional level is removed (in ddm1 mutants), Athila is epigenetically activated and there is increase in UBP1b stress granule protein than in epigenetically silent Col plants (Figure 2.16). I made a triple mutant ddm1/rdr6/ubp1b that has loss of chromatin modification, disrupted RNAi, and inability to form stress granules to test the role of UBP1b in Athila regulation. I found that the triple mutant has more Athila copies than wt (p<0.05) as shown in Figure 2.8 A. Hence, UBP1b seems to have an effect on Athila copy number. Next, I tested the role of UBP1b in regulation of Athila at translational level. In contrast to loss of only transcriptional regulation (ddm1 mutant), ddm1/rdr6 double mutant that has loss of chromatin modification and disrupted RNAi shows less gag protein accumulation in spite of high Athila6 gag/pol transcript expression (Figures 2.8 B and 2.9). This suggests that UBP1b may be playing a role in sequestering Athila transcripts and inhibiting their translation. When there is loss of chromatin modification, disrupted RNAi and inability to form stress granules in the triple mutant (ddm1/rdr6/ubp1b), highest level of Athila GAG protein accumulation was observed. Thus, through my research, I have shown that akin to mammalian TIA-1, plant ortholog UBP1b causes translational repression of Athila and serves as an additional, third negative regulatory layer inhibiting Athila activity beyond chromatin modification and endogenous RNAi. It is yet unknown whether these stress granules are capable of sequestering Athila and other transposable elements or not similar to LINE1 retrotransposons sequestration by mammalian stress granules (Goodier et al. 2007b).
2.5.2 Regulation of genes by TEs

Transposable elements (TEs) or the jumping genes which when active get inserted into existing host genes and may disrupt genomic functions by structural polymorphisms. Nevertheless, TEs are also capable of regulating genes through other mechanisms apart from insertional mutagenesis. TEs such as LTR retrotransposons and LINEs when inserted in *cis* of neighboring genes, may act as their promoters and drive transcription (Xiao et al. 2008)(Slotkin and Martienssen 2007a) (Dunn, Medstrand, Mager 2003; Dunn et al. 2006) (Wolff et al. 2010)(Mourier et al. 2014). TEs are also known to affect genes in *cis* by influencing promoter and enhancer regions present at huge distances from protein coding regions of genes (Nakanishi et al. 2012; Sasaki et al. 2008). On the other hand, during various environmental conditions such as stress, epigenetic regulation of TEs is altered and TEs can also regulate non-neighboring genes in *trans* (Grandbastien 2004; Tittel-Elmer et al. 2010). Only recently for the first time, it was discovered that piRNA or siRNA derived from TE, regulates genic mRNA in *Drosophila* and mouse (Rouget et al. 2010; Watanabe et al. 2011). Also, in tobacco the reason for yellowing symptom caused by Cucumber mosaic virus (CMV) was not known for a long time, until recently it was found to be caused by siRNA produced from CMV targeting chlorophyll biosynthetic gene in *trans*. (Smith, Eamens, Wang 2011). In both plants and animals, epigenetic activation of TEs and viral attacks cause unexplained physical and genetic damage to the host organisms. Given that host translation machinery is used for the survival of viruses, translational inhibition by host stress granules is very essential for fighting these viruses (Katoh et al. 2013). To combat this host defense, most viruses such
as Japanese encephalitis virus (JEV), Polio Virus (PV), Influenza A virus, human immunodeficiency virus 1 (HIV-1) have evolved to get around SG formation during infection (Cristea et al. 2010; Emara and Brinton 2007; Smith and Gray 2010; White et al. 2007)(Lindquist et al. 2010)(Khaperskyy, Hatchette, McCormick 2012)(Abrahamyan et al. 2010).

2.5.3 *Athila* via siRNA854 causes translational inhibition of UBP1b

Adenovirus associated RNAs, an LTR retrovirus in animals, produces miRNAs such as miR854 and miR855 that target silencing of TIA-1 (Aparicio et al. 2010). Arteaga-Vázquez *et al* predicted that these two miRNAs are present in *Athila* retrotransposons in plants (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006a). Since miR854 and miR855 are very much conserved between both animals and plants, these miRNAs may play a role in silencing plant stress granule protein, UBP1b which in turn might cause translational inhibition of TE mRNAs in stress granules (Slotkin 2010). Arteaga-Vázquez *et al* predicted that microRNA854 (now known as siRNA854), targets the 3′ untranslated region (UTR) of the *UPB1b* gene (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006b)(Weber, Nover, Fauth 2008). They observed that siRNA854 accumulates in wt vegetative tissues and found that the siRNA854-targeted *UPB1b* 3′UTR inhibits translation in wt plants when the 3′UTR is added to a reporter transcript. Despite its non microRNA status, I sought to answer the question: Can siRNA854 produced from *Athila* retrotransposon target a non-TE gene, UBP1b to suppress third level of TE silencing mechanism? When epigenetic silencing is removed from *Athila* in a *ddm1* mutant, the *Athila* is transcribed (Figure 2.3 A and B) and
produces 21 and 22 nt siRNAs (Figure 2.3 B and D). I used the UBPlb 3'UTR in reporter assays to demonstrate that whenever I observe the accumulation of the 21–22 nt siRNA854 sequence in ddm1 mutants (Figure 2.3 B and D), I observe decreased reporter protein accumulation in Arabidopsis seedlings, leaves, inflorescences and siliques (Figure 2.4 and 2.5 B). Thus one of Athila siRNAs, siRNA854, regulates the UBPlb in trans. In addition to the increased levels of siRNA854 in ddm1 mutants, siRNA854 also accumulates in wt Col pollen (Table 2.1) due to natural expression of Athila. I used similar transgene reporter assays as above in wt Col pollen and observed that the RDR6-dependent accumulation of siRNA854 and base pairing with the UBPlb 3'UTR target sites are required for the inhibitory regulation of UBPlb 3’UTR reporter genes (Figure 2.6 A and B). Amongst different Argonaute proteins such as AGO1, AGO6, AGO8 and AGO9, AGO6 has the largest role in the repression of the UBPlb 3’ UTR (Figure 2.6 C). It is surprising to observe AGO6 role in UBPlb repression in pollen because it is known in literature that AGO6 binds 24nt siRNAs and functions to keep repetitive regions of the Arabidopsis genome in a transcriptionally silenced state (Eun et al. 2011; Olmedo-Monfil et al. 2010a). We have shown in the lab that AGO6 binds to 21-22nt siRNAs as well (Unpublished data, see Chapter-3). This suggests that AGO6 plays a role in UBPlb repression through binding to 21-22nt siRNA854. Hence, a TE-derived siRNA854 is a TE defense strategy that uses the second level of Athila control (RNAi) to act as a suppressor against the third level of regulation, UBPlb repression of Athila. This data provides evidence that the interaction between TE small RNAs and their mRNA
targets might be an advantage for TEs and aid in their life cycle. Such interactions can be selected evolutionally.

**2.5.4 Future directions**

To date only one example is known for translational inhibition of TEs namely LINE-1 or L1 in animals (Goodier et al. 2007a) by Stress granules. Through my work in this chapter, I have shown a second example for translational inhibition of TEs by SGs and it occurs in plants, where *Athila* is translationally repressed by *UBP1b*, a component of plant stress granules. For both these examples it is yet to be determined whether translational repression by stress granules represents accumulation depots for L1/ *Athila* or if they play an important role in L1/ *Athila* retrotransposition. In addition to translational inhibition by SGs, even P bodies play a role in translational inhibition of TEs. For example, in yeasts, P body components cause translational repression and boost the retrotransposition of Ty3 and Ty1 retrotransposons Virus Like Particles (VLPs) (Beliakova-Bethell et al. 2006; Checkley et al. 2010; Dutko et al. 2010). Moreover, important question to address would be what other TEs may be regulated by stress granules or P bodies. It is not yet known how and why stress granules specifically take in TE mRNAs during stress. Much work remains to be done to fully understand the interactions between TEs and SGs, a field that is in its infancy. Future experiments are needed to determine if viral or TE cleavage of the SG components such as TIA-1, UBP1b is due to the need to inhibit SGs or a consequence of requirement for TE life cycle in either case it would be beneficial for TEs (Goodier et al. 2010; McCue et al. 2012b). It would be interesting to study if there are any other transposons apart from *Athila* that
could manipulate SG responses and the precise mechanisms by which transposons inhibit SG formation. Recently, it was shown in *Drosophila* that gene regulating small RNAs are produced due to TE reactivation occurring because of epigenetic and developmental regulation. Exact mechanism of this gene regulation by TE piRNAs is unknown. As different environmental stresses may elicit epigenetic activation of TEs, it would be compelling to study the effect of transient gene regulation by TE siRNAs and their potential contribution to an epigenetic effect on an organism’s stress granules. Details of how genomes with naturally active transposons deal with continuous gene regulation that occurs due to the constant presence of gene regulating TE small RNAs are not yet known.

To answer the above mentioned future perspectives on SG regulation of TEs and TEs regulation of host genes, I propose to perform mRNP (mRNA bound RiboNucleoProteins) Immuno Precipitation (IP) with N-terminal FLAG tagged UBP1b in *ddm1* and *ddm1/rdr6* plants (loss of chromatin modification and disrupted RNAi) as described by Sorenson *et al* (Sorenson and Bailey-Serres 2014). From this mRNP-IP, I would isolate the mRNA transcripts to determine if *Athila* is sequestered and also if any other TEs are also present in SGs. By looking at different types of mRNA transcripts found in stress granules when there is epigenetic activation of TEs, the reason for recruitment of specific types of transcripts during stress can be determined. The same experiment can be performed in genomes with active TEs such as maize to examine the translational inhibition of TEs by stress granules. Multidimensional protein identification technology can be applied on isolated mRNPs to identify UBP1b interactions with other proteins in stress granules. Lack of UBP1b along with key proteins involved in first and
second levels *Athila* silencing led to increased GAG protein accumulation. I was not able to determine if other protein coding region of *Athila* POL (RT) translation is also affected in the triple mutant (ddm1/ubp1b/rdr6). Also, we couldn’t detect any *Athila* transposition in triple mutant. Even after the loss of three levels of *Athila* repression, there’s some other undiscovered mechanism that’s preventing *Athila* transposition and affects the fertility of the plant. Since transposition assays didn’t work, it would be valuable to assay for *Athila* cRNAA intermediated by isolating GAG particles or by performing immune precipitations with GAG antibody and then isolating RNA or cDNA intermediates.

UBP1b regulation by siRNA854 produced from *Athila* retrotransposon was at translational level in inflorescence tissue and post transcriptional level in pollen. In different tissues the mechanism of regulation of genes by TE siRNAs seems to be different. It would be interesting to test the mechanism of regulation in different tissues of Arabidopsis such as seedlings, leaves, siliques etc. Also, UBP1b repression by *Athila* in pollen seems to be mainly dependent on AGO6. Due to AGO6’s recent role in RNA directed DNA methylation (see chapter-3), it would be interesting to test the promoter of UBP1b-GFP transgene for methylation and hence determine if siRNA854 causes transcriptional silencing of UBP1b or not.

Rouget *et al.* and McCue *et al.* have bioinformatically predicted different genic targets of TE derived small RNAs (McCue, Nuthikattu, Slotkin 2013; Rouget et al. 2010). Some of the targets seemed to be playing a role in mRNA processing and stability while some other targets were involved in metabolism. It would be interesting to use the
similar bioinformatics approaches to determine TE regulation of genes in organisms with active TEs such as humans, maize etc.

**Figure 2.1 Athila experiences three levels of host repression.**

The first level of repression occurs at transcriptional level via chromatin modification. The second level of repression, RNAi, which occurs at post transcriptional level, produces siRNA854 that targets the third level of *Athila* repression at translational level by inhibiting UBP1b, a protein involved in stress granule formation.
Figure 2. The 14kb reconstructed consensus Athila6A element.

Athila6A contains 2 1.8kb LTRs, 2 ORFS (gag and pol), a degenerate ORF (env), and a 3’ non-coding region, which is the region of increased 21-22nt siRNAs production. miRNA854 has 4 predicted binding sites in the 3’ UTR of UBP1b (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006a).
Figure 2. Expression of the Athila6 retrotransposon leads to accumulation of Athila 21–22 nt siRNAs, including siRNA854.

(A) qRT-PCR of steady-state mRNAs coming from the Athila6 env region. Expression in wt Col inflorescences (infl) is activated in ddm1 and met1 mutants. Athila6 transcript accumulation increases >24,000-fold in ddm1 mutants compared to wt Col expression. (B) Northern blot detecting siRNA854 and Athila6 3′ region in ddm1 and met1 mutant infl. 21–22 nt Athila6 siRNAs and siRNA854 only accumulate when the retrotransposon is transcriptionally activated. The DNA oligonucleotide probe used to detect siRNA854 is 21 nt in length, and is shown in Table 2.2. (C) qRT-PCR of the Athila6 3′ region of gag/pol, demonstrating that higher retrotransposon expression levels accumulate when ddm1 is maintained as a homozygote over several generations. F2 = second generation of homozygosity, F6 = sixth generation. (D) Small RNA Northern blot detecting siRNA854 and the Athila6 3′ region. Increased levels of Athila6 21–22 nt siRNAs and siRNA854 correspond to samples with higher transcript levels. For Northern blots in parts B and D, microRNA161 (miR161) and a heterochromatic-region 24 nt siRNA (siRNA02) are shown as loading controls.

Continued
Figure 2.3: Continued

A

![Graph showing Athila 3' expression](image)

**Athila 3' expression (arbitrary units)**

- Col
- ddm1
- mer1

**p<0.001**

B

![Image of Western blots](image)

- Athila 3'
- siRNA854
- miR161
- siRNA02

C

![Graph showing Athila 6 3' expression](image)

**Athila 6 3' expression (arbitrary units)**

- Col
- ddm1 F2
- ddm1 F6

**p<0.001**

**p<0.01**
Figure 2. 4 Increased levels of siRNA854 leads to UBP1b 3'UTR repression in T1 and T2 Arabidopsis seedlings and T1 inflorescences

(A) Col and ddm1 plants were transformed with GUS reporter transgene with 35S constitutive promoter (35S:GUS) or the same reporter transgene with the UBP1b 3' UTR (35S:GUS-UBP1b 3'UTR). GUS activity was monitored by GUS staining of T1 seedlings, leaves, inflorescences and siliques. Wt Col seedlings show less GUS activity of the UBP1b 3' UTR transgene than the control 35S:GUS transgene due to low levels of siRNA854. Wt Col siliques, leaves and inflorescences didn’t show much difference in GUS activity between control and the UBP1b 3' UTR transgene, because wt Col leaves and inflorescences have negligible Athila 3’ expression and siRNA854 production. In ddm1, due to increased production of siRNA854 the GUS activity of the UBP1b 3' UTR transgene is very less than the control 35S:GUS transgene in seedlings, leaves and inflorescences but slightly less in siliques. (B) GUS staining was performed in T2 seedlings in 96 well plate. Nearly 85% of T2 seedlings showed less GUS activity of the UBP1b 3'UTR transgene in ddm1 produced from 5 different T1 plants (7-11) than control Col T2 seedlings produced from 6 different T1 plants (1-6) with the same transgene.
Figure 2. 4: Continued

A  Col, 35S:GUS

<table>
<thead>
<tr>
<th>Seedlings</th>
<th>Leaves</th>
<th>Inflorescences</th>
<th>Siliques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col, 35S:GUS-UBP1b3’UTR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ddm1, 35S:GUS</th>
</tr>
</thead>
</table>

| ddm1, 35S:GUS-UBP1b 3’UTR |

Continued
Figure 2.4: Continued

B

[Image of a bar graph showing the percentage of T2 seedlings for different genotypes, including Col, 35S:GUS-UBP1B 3'UTR and ddm1, 35S:GUS-UBP1B 3'UTR.]
Figure 2. 5 Accumulation of 21–22 nt siRNA854 negatively regulates transgene transcripts with the UBP1b 3’ UTR in T3 inflorescences.

(A) Plants homozygous for a transgene constitutively expressing the transgene35S:GUS-UBP1b 3'UTR were transformed with an artificial microRNA (amiRNA) with the siRNA854 sequence (35S:amiRNA-854), or a second control sequence that does not target UBP1b (35S-amiRNA-control). GUS activity was monitored using a quantitative assay (left) and inflorescence staining (right). Plants expressing the siRNA854 sequence as an artificial microRNA show significantly reduced GUS levels. (B) Col and ddm1 plants homozygous for a constitutively expressed GUS transgene (35S:GUS) or the same reporter transgene with the UBP1b 3’ UTR from part A. As in A, GUS activity was monitored using a quantitative assay (left) and plant staining (right). Wt Col plants show no differential regulation between the two transgene variations, while in the ddm1 mutant background the GUS activity of the UBP1b 3’ UTR transgene is significantly less than the control 35S:GUS transgene. (C) A 35S:GUS-UBP1b 3'UTR transgene in the wt Col background was crossed to a ddm1 homozygote, and the GUS activity was measured in the F1 plant. GUS activity of the same hemizygous transgene in the wt Col background is shown as a control. (D) RT-PCR of the same transgenic individuals from part B. The GUS activity differences observed in part B are not reflected in transgene transcript levels, demonstrating that this regulation is not due to post-transcriptional mRNA degradation. Primers are shown in Table 2.2. In A, B and C, the box plot whiskers represent the minimum and maximum of the dataset, the top and
bottom of the box are the 75\textsuperscript{th} and 25\textsuperscript{th} percentile (respectively), the middle line is the median, and + is the mean. The number of individuals assayed (n) is shown in or near the box. (E) 21–22 nt *Athila6* 3′ siRNAs and siRNA854 accumulate in *ddm1* heterozygotes produced by crossing wt to an individual homozygous for the recessive *ddm1-2* allele (*Col x ddm1*).
Figure 2. 5: Continued
Figure 2. 6 Accumulation of 21–22 nt siRNA854 in wt pollen regulates transgene transcripts with the UBP1b 3’UTR.

(A) Base pairing of siRNA854 to the wt, MOD and DEL versions of the UBP1b 3’ UTR. The UBP1b3’ UTR is shown from 5’ to 3’. The four sites of the UBP1b 3’UTR targeted by siRNA854 predicted by Arteaga-Vázquez et al (Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006a; Arteaga-Vazquez, Caballero-Perez, Vielle-Calzada 2006a) are shown. In the MODified 3’ UTR variant, all perfect complementary base pairing was replaced. In the DELeted 3’ UTR variant, the entire target sites have been removed, resulting in a shorter 3’ UTR. (B) GFP transgenes with pollen-specific expression were used to assay the activity of endogenous siRNA854 in pollen. Background corrected GFP fluorescence levels are shown on the left, while representative pollen images are on the right. On each image, an asterisk marks non-fluorescent pollen grains that did not inherit the transgene from the hemizygous parent that were used for background correction. Addition of the UBP1b 3’UTR to a pollen-expressed GFP reporter transgene results in decreased GFP fluorescence in wt Col transgenic pollen grains, as siRNA854 accumulates in the wt pollen vegetative nucleus. Abrogation of the perfectly complementary base pairing in the predicted siRNA854 target sites of the UBP1b 3’UTR (MOD), or removal of these target sites altogether (DEL) alleviates this repression. The repression of the UBP1b 3’ UTR in pollen is lost in rdr6 mutant plants. Box plot whiskers represent the 90th and 10th percentile of the dataset, the top and bottom of the box are the 75th and 25th percentile (respectively), the middle line is the median, and + is the mean. Number of pollen grains measured and the number of transgenic individuals they came from (in parentheses) is shown in or near the box. Scale bars = 10 µm. (C) GFP

Continued
transgenes with pollen-specific expression were used to assay the effect of different Argonaute proteins in the expression of \textit{UBP1b}, \textit{ago1}, \textit{ago6}, \textit{ago8} and \textit{ago9} mutants with the GFP: \textit{UBP1b} 3'UTR transgene show statistically significant increases in GFP fluorescence (p<0.0001) when compared to Col control with the GFP: \textit{UBP1b} 3'UTR transgene. The increase in fluorescence of \textit{ago6} was significantly higher (p<0.0001) than \textit{ago1}, \textit{ago8} and \textit{ago9} mutants with the GFP \textit{UBP1b} 3'UTR transgene and hence AGO6 plays a larger role in the repression of \textit{UBP1b} 3’ UTR. Background corrected GFP fluorescence levels are shown. Box plot whiskers represent the 90\textsuperscript{th} and 10\textsuperscript{th} percentile of the dataset, the top and bottom of the box are the 75\textsuperscript{th} and 25\textsuperscript{th} percentile (respectively), the middle line is the median, and + is the mean. Number of pollen grains measured and the number of transgenic individuals they came from (in parentheses) is shown in or near the box.

\textbf{Figure 2. 6: Continued}
Figure 2.6: Continued

A  **UBP1b 3'UTR**

![Diagram of UBP1b 3'UTR with complementary base pairs and siRNA854](image)

B  **Saturation Fluorescence**

![Box plot showing fluorescence levels](image)

Continued
Figure 2.6: Continued

C

Fluorescence (grey units)

Col Lat52: GFP
Col Lat52: GFP-UBP1b 3' UTR
ago6, Lat52:GFP-UBP1b 3' UTR
ago8, Lat52:GFP-UBP1b 3' UTR
ago9, Lat52:GFP-UBP1b 3' UTR
ago1-11, Lat52:GFP-UBP1b 3' UTR
Ler Lat52: GFP

p < 0.0001

59 (2)
56 (5)
130 (9)
28 (1)
115 (4)

7 (41)
11 (73)
Figure 2. 7 Immature seed GUS staining of Col, 35S: GUS-UBP1b 3’UTR crossed to Col and rdr6.

Col, 35S: GUS-UBP1b 3’UTR (female) was crossed to Col (male) and rdr6 (male) respectively. Immature seeds were isolated from the resulting siliques of the crosses and GUS activity was monitored by qualitative GUS staining. Immature seeds from Col, 35S: GUS-UBP1b 3’UTR X Col cross were mostly stained dark blue to light blue whereas immature seeds from Col, 35S: GUS-UBP1b 3’UTR X rdr6 cross were mostly stained very light blue to no color. Representative pictures of GUS stained immature seeds from both the Col, 35S: GUS-UBP1b 3’UTR X Col and Col, 35S: GUS-UBP1b 3’UTR X rdr6 crosses are shown in the left side of this figure. The graph in the right side of this figure shows the number of immature seeds stained as dark blue, light blue, very light blue and no color for both the Col, 35S: GUS-UBP1b 3’UTR X Col and Col, 35S: GUS-UBP1b 3’UTR X rdr6 crosses.
Figure 2. 8 Number of copies and expression of Athila6 in wt and 12 different mutant background combinations.

(A) qPCR of the Athila6 gag/pol protein coding region. This panel of 12 includes plants mutant for ddm1 (loss of chromatin modification and transcriptional regulation), a ddm1/rdr6 double mutant (loss of chromatin modification and disrupted post-transcriptional RNAi), the triple mutant ddm1/rdr6/ubp1b (loss of chromatin modification, disrupted RNAi, and unable to form stress granules), as well as control genotypes. Interestingly, the triple mutant has more Athila copies than wt. This assay will detect both chromosomal DNA copies as well as extra-chromosomal cDNA copies of transposing Athila elements. Y axis is relative to number of copies of control gene At1g08200. (B) qRT-PCR of the Athila6 gag/pol protein coding region. The ddm1/rdr6 double mutant (loss of chromatin modification and disrupted post-transcriptional RNAi) exhibits higher Athila6 transcript accumulation than the single ddm1 mutant. Primers used are shown in Table 2.2.
Figure 2. 8: Continued
Figure 2. 9 Isolation of gag particles and silver staining reveals gag encapsidation of Athila.

Silver staining of SDS-PAGE gel containing proteins in the pellet with gag particles from ultracentrifugation was performed for the panel of mutants described in Figure 2.8. Triple mutant, ddm1/rdr6/ubp1b (loss of chromatin modification, disrupted RNAi, and unable to form stress granules), ddm1F6 and ddm1/ubp1b mutants show more protein bands in the range of 28-32kDa consistent with gag polypeptide region than ddm1/rdr6 mutant.
Figure 2. Loss of epigenetic regulation of Athila affects seed formation in siliques. (A) Whole siliques from the panel of mutants described in Figure 2. Siliques from double and triple mutants with epigenetic activation of Athila have stunted growth. (B) Silique dissection of the whole siliques in A was performed. Dissected siliques of double and triple mutants had severe sterility with less to no seed production respectively.
Figure 2. 10: Continued

A

<table>
<thead>
<tr>
<th>Col back ground</th>
<th>Col/Ws mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col</td>
<td>ubp1b in WS</td>
</tr>
<tr>
<td>ddm1 F2</td>
<td>ddm1/het/ubp1b</td>
</tr>
<tr>
<td>G6</td>
<td>ddm1/ubp1b</td>
</tr>
<tr>
<td>ddm1/rd6/ubp1b</td>
<td>ddm1/ubp1b</td>
</tr>
</tbody>
</table>

Continued
Figure 2.10: Continued

B
Figure 2. 11 siRNA processed from Cen180 and 106B follow pathway distinct from Athila siRNAs.

(A) The distribution of gypsy family retrotransposons, DNA transposons and genes on a stereotypical plant chromosome.

(B) The stereotypical layout of the internal region of a plant centromere. Alpha satellite repeats, CEN180 are interspersed with retrotransposons in centromeres. Fragments of gypsy family LTR retrotransposons drive the expression the satellite repeats and are referred as 106B (May et al. 2005). A and B are from our book chapter on “Impact of transposable elements on gene and genome evolution” in Molecular Biology and Evolution of the Plant Genome.

(C) RT-PCR with single and double mutants of different levels of Athila silencing. 106B or Athila2 and Athila6 expression is high in ddm1, ddm1/rdr6 and ddm1/ubp1b mutants than met1. Cen180 expression is high in ddm1 than in Col, Ws, met1, rdr6 and ubp1b. Cen180 expression is increased in ddm1/rdr6 and ddm1/ubp1b double mutants than in ddm1 single mutants and this indicates that improper centromere function in ddm1 gets worse with ddm1/rdr6 (loss of chromatin modification and disrupted post-transcriptional RNAi) and ddm1/ubp1b (loss of chromatin modification and unable to form stress granules) double mutants. Primers used for RT-PCR are shown in Table 2.2.

(D) Small RNA northern blot detecting 24 nt Cen180 and 106B siRNAs. 24nt siRNAs from Cen180 and 106B are produced in both TEs transcriptionally silent and transcriptionally active epigenomes through RNA directed DNA methylation (RdDM) Continued
pathway components such as POL4, RDR2 and DCL3 (red boxes). This is distinct from the biogenesis of 21-22nt *Athila* siRNAs. The DNA oligonucleotide probes used to detect siRNAs from *Cen180* and *106B* are 24 nt in length, and are shown in Table 2.2. microRNA161 (miR161) is shown as loading control.
Figure 2.11: Continued

A

B

alpha satellite tandem repeats
fragments of young gypsy LTR centromeric retrotransposons
complete young gypsy LTR centromeric retrotransposons
fragments of old gypsy LTR centromeric retrotransposons

CEN180

106B

Continued

Continued
Figure 2.11: Continued

C

[Image: Gel electrophoresis images for different samples labeled A, 16S, Cen180, and Tyr2T. Each lane contains samples marked as 'Parent', 'Col', 'rest', 'delt', 'rdg', 'delt/rdg', 'delt/rdg/le', 'delt/rdg/te', 'Vig', 'Col/DNA', and 'H2O'.]
Figure 2.11: Continued

D
Figure 2. 12 Role of UBP1b in Athila repression.

PolyA northern shows new Athila bands in ddm1/rdr6 and ddm1/ubp1b double mutants which are not present in ddm1 single mutant. PolyA northern was probed with Athila6 gag/pol and Env T7 probes shown in Table 2.2 detecting sense transcripts. TyrAT T7 probe shown in Table 2.2 serves as control.
Figure 2. 13 Athila DNA isolated from gag particles from a panel of mutants involved in different layers of Athila repression.

(A) Leaf DNA was isolated from gag particles from the triple and single mutants described in Figure 2.8. PCR was performed with primers for Athila6 CDS and primers for TyrAT (a single copy gene) and Cen180 (a multi copy gene) used as negative controls. Col genomic DNA was used as positive control for the PCR reaction and it worked for all the three PCR targets. TyrAT negative control worked and no bands were seen in any of the genotyped tested. However, Cen180 didn’t amplify for Col, ddm1 F6 but amplified in rdr6 and ddm1/rdr6/ubp1b. Athila6 DNA amplified only in rdr6 and ddm1/rdr6/ubp1b. It is difficult to conclude about the presence or absence of Athila6 DNA in gag particles isolated from rdr6 and ddm1/rdr6/ubp1b since negative control Cen180 showed amplification in these two genotypes. (B) Gag particle purification using ultra centrifugation was performed on inflorescence tissue from the panel of mutants described in Figure 2.8. PCR was performed to amplify Athila6, Cen180 and TyrAT genes from DNA isolated from the gag particles. TyrAT and Cen180 negative control genes showed non-specific amplification in all the genotypes. In order to eliminate this non-specific amplification of negative control genes, the pellet containing gag particles was treated with DNAse I. DNA was isolated from these DNAse I treated gag particle containing pellets, and then PCR was used to amplify the target genes. In spite of DNAse I treatment, non specific amplification of TyrAT and Cen180 negative control genes was still observed. Athila6 DNA amplified in all of the genotypes tested. Since both the negative control genes showed amplification, it is still unclear whether Athila6 DNA

Continued
Figure 2.13: Continued

intermediates are present or absent in gag particles isolated from various mutants that are
defective in different levels of *Athila* repression. Primers used are shown in Table 2.2.
Figure 2.13: Continued

A

B

Mock (without DNase1 treatment) (Inflorescence)  After DNase1 treatment

94
Figure 2. 14 UBP1b, a stress granule protein acts to repress Athila protein production.

(A) Western analysis of *Athila* GAG protein levels in double and triple mutant plants. Actin is shown as a protein loading control. The highest level of GAG protein is detected in *ddm1/rdr6/ubp1b* triple mutant plants. (B) Statistical analysis of GAG protein accumulation in three independent westerns as in (A). For each of the three analyses, GAG protein is highest in the *ddm1/rdr6/ubp1b* triple mutant plants. GAG protein levels in each western and for each sample were normalized to the level of Actin protein accumulation.
Figure 2.15 Product Enhanced Reverse Transcriptase Assay (PERT).

(A) Schematic of Product Enhanced Reverse Transcriptase Assay methodology. (B) PERT assay was performed from the panel of mutants described in Figure 2.8. Protein isolated from Col spiked with ssIII Reverse Transcriptase (RT) and no protein control spiked with ssIII RT were used as positive controls and both these controls amplified MS2 cDNA. No cDNA template and no protein were used as negative controls and both these controls didn’t show any MS2 cDNA amplification. Inflorescence, leaf and silique tissues of the triple ddm1/rdr6/ubp1b mutant (all three modes of Athila repression are absent), exhibits endogenous reverse transcriptase activity. Slightly higher endogenous reverse transcriptase activity was observed in ddm1/rdr6/ubp1b triple mutant than in rdr6 and ddm1/ubp1b mutants leaf tissues. Col, Ws, ddm1, ubp1b and ddm1/rdr6 leaf tissues didn’t show any reverse transcriptase activity. In inflorescence tissue, slightly higher endogenous reverse transcriptase activity was observed in ddm1/rdr6/ubp1b triple mutant than in WS, ubp1b, ddm1/rdr6 and ddm1/ubp1b mutants. No reverse transcriptase activity was detected in Col, ddm1 and rdr6 inflorescence tissues. (C) In order to verify the results obtained with inflorescence tissue shown in Figure 2.14 B, biological replicate of PERT assay was performed from the panel of mutants defective in three modes of Athila repression. Endogenous reverse transcriptase activity was seen in ddm1 F6, ubp1b, ubp1b in Col, rdr6/ubp1b, ddm1/rdr6/UBP1b in Ws and ddm1/rdr6/ubp1b mutants but not in Col, ddm1 F2, rdr6, Ws, ddm1/rdr6 and ddm1/ubp1b mutants. It is difficult to infer anything about the effect of different levels of Athila silencing on endogenous reverse transcriptase activity of Athila because the results of both the

Continued
Figure 2.15: Continued

biological replicates of PERT experiment on inflorescence tissue differ from each other. Primers used are shown in Table 2.2.
Figure 2.15: Continued

A

B

C
Figure 2. 16 ImmunoPrecipitation-western (IP-Western) of UBP1b-GFP from seedling roots.

Col and ddm1 plants were transformed with UBP1b driven by 35S constitutive promoter along with GFP reporter transgene (35S:UBP1b-GFP). Seedling roots were harvested and IP-Western was performed with crosslinked (C) and non crosslinked (N) tissues. Increase in immunoprecipitated UBP1b-GFP protein levels is observed in ddm1 than in Col.
<table>
<thead>
<tr>
<th></th>
<th>siRNA854 raw counts</th>
<th>genome matched reads&lt;sup&gt;A&lt;/sup&gt;</th>
<th>siRNA854 N&lt;sup&gt;B&lt;/sup&gt;</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col inflorescence</td>
<td>0</td>
<td>2,516,337</td>
<td>0</td>
<td>Lister &lt;i&gt;et al.&lt;/i&gt;, 2008 Cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lister &lt;i&gt;et al.&lt;/i&gt;, 2008 Cell</td>
</tr>
<tr>
<td>&lt;i&gt;met1&lt;/i&gt; inflorescence</td>
<td>290</td>
<td>1,506,711</td>
<td>192</td>
<td>Slotkin &lt;i&gt;et al.&lt;/i&gt;, 2009 Cell</td>
</tr>
<tr>
<td>Col inflorescence</td>
<td>1</td>
<td>3,200,398</td>
<td>1</td>
<td>Slotkin &lt;i&gt;et al.&lt;/i&gt;, 2009 Cell</td>
</tr>
<tr>
<td>&lt;i&gt;ddm1&lt;/i&gt; inflorescence</td>
<td>268</td>
<td>3,568,226</td>
<td>75</td>
<td>Tanurdzic &lt;i&gt;et al.&lt;/i&gt;, PLoS Biology</td>
</tr>
<tr>
<td>Col leaf</td>
<td>0</td>
<td>926,951</td>
<td>0</td>
<td>Slotkin &lt;i&gt;et al.&lt;/i&gt;, 2009 Cell</td>
</tr>
<tr>
<td>Col pollen</td>
<td>14</td>
<td>551,394</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

<sup>A</sup> Does not include tRNA, rRNA, snRNA, snoRNA reads

<sup>B</sup> Number of siRNA854 reads normalized per 1 million reads

Table 2. Frequency of 21 nt siRNA854 in SBS small RNA libraries.
Table 2. 2 PCR primers and mutant alleles used.
Chapter-3: RNAi plays a role in de novo methylation of active transposable elements through 21-22 nucleotide small RNAs and requires POL V scaffolding transcripts.
3.1 Abstract

Transposable elements (TEs) are DNA fragments that, when active, move from one part of the genome to another. During this movement, TEs cause genome instability in their hosts. Eukaryotes have evolved epigenetic silencing mechanisms involving small interfering RNAs (siRNAs) to inhibit TE activity. The epigenetic silencing of TEs in plants at transcriptional level is mediated by the epigenetic inheritance of cytosine DNA methylation patterns. These methylation patterns are established via a process of RNA-directed DNA methylation (RdDM) through 24 nucleotide(nt) siRNAs. When the epigenetic repression of TEs is lost, TEs become transcriptionally active, and the host cell acts to repress mutagenic TEs by post transcriptionally degrading TE mRNAs into 21-22nt siRNAs. How the repression of transcriptionally active TEs changes from post-transcriptional control to transcriptional level epigenetic silencing has remained unknown. We have recently identified a pathway that acts independent of 24nt siRNAs. Our discovery began when I observed that upon mutation of one of the major components of RdDM, there is complete loss of 24nt siRNAs in our model organism Arabidopsis thaliana. However, this loss of 24nt siRNAs does not lead to complete loss of RdDM and TE de novo methylation. Hence I hypothesized that the RNA interference (RNAi) that causes post-transcriptional degradation of TE mRNAs might play a non-traditional role to methylate TE chromatin via RdDM. I identified that RDR6, a major component of the RNAi, functions in RdDM and hence we named this pathway as RDR6-RdDM. I determined that only transcriptionally active TEs produce RDR6-dependent 21-22nt siRNAs and experimentally identified Athila retrotransposon as one of the targets of
RDR6-RdDM. Expression dependent function of RDR6-RdDM is dependent on 21-22nt Athila siRNAs and Pol V scaffolding transcripts. My data highlights an expression-dependent pathway to bridge the gap between post-transcriptional degradation of TEs and the establishment of TE DNA methylation for epigenetic silencing.

3.2 Introduction

3.2.1 Mechanisms of TE silencing

Transposable elements (TEs) constitute large percentages of both animal and plant genomes. The mobilization of TEs results in chromosome instability and mutation. In order to maintain genome integrity, fungi, plants and animals modify TE chromatin through multiple endogenous gene-silencing pathways that act to limit their expression to repress the production of TE mRNAs to inhibit TE mobilization (Girard and Hannon 2008). To study TE silencing, the process has been divided into three distinct mechanisms: the de novo initiation/triggering of silencing, the establishment of silencing, and the epigenetic maintenance of TE silencing.

3.2.2 Maintenance TE DNA methylation

In mammals and flowering plants, cytosine DNA methylation is critical for inhibiting TE activity (Zemach and Zilberman 2010). Once established at TEs, DNA methylation in the symmetrical CG context is propagated by the DNMT1 family of CG methyltransferases (MET1 in plants) (Law and Jacobsen 2010b) through mitotic cell divisions. In plants, non-CG context DNA methylation (CHG or CHH, where H = A, T or C) is also epigenetically maintained via recruitment of the CMT3 and CMT2 DNA methyltransferases through their interaction with repressive histone modifications (Stroud
et al. 2014). In contrast to animals, plants do not erase the DNA methylation patterns of their gametes; therefore, CG and CHG (where H = A, T, or C) symmetrical DNA methylation patterns established in one generation are inherited and act to maintain TE silencing in the next generation through a process termed transgenerational epigenetic inheritance (Mathieu et al. 2007)(Becker et al. 2011). Thus, once established, robust mechanisms exist to propagate TE DNA methylation, resulting in the epigenetic transcriptional silencing of TEs. However, the mechanism of how DNA methylation and epigenetic silencing is originally targeted to TEs is not known.

3.2.3 Pol IV-RNA directed DNA Methylation (Pol IV-RdDM)

In contrast to the maintenance of methylation, the methylation of previously unmethylated cytosines (de novo methylation) occurs through the DNMT3 family of DNA methyltransferases (DRM2 in plants), which methylate cytosines in any sequence context (Law and Jacobsen 2010b). In plants and mammals, de novo TE DNA methylation is targeted through an RNA-directed DNA methylation (RdDM) pathway, which utilizes TE small interfering RNAs (siRNAs) in plants and piRNAs in animals to guide ARGONAUTE (AGO) family proteins to TE chromatin (Castel and Martienssen 2013). Similar pathways of small RNA-mediated chromatin modification exist in fission yeast, Drosophila and C. elegans, which have evolutionarily lost cytosine DNA methylation (Castel and Martienssen 2013; Grewal 2010; Zemach and Zilberman 2010). In each of these models, a small RNA-targeted AGO family protein is recruited to the TE locus via a scaffolding RNA that functionally tethers the AGO protein to the target chromatin (Castel and Martienssen 2013). The recruitment of the AGO protein to the
TE locus then initiates a poorly understood cascade of chromatin modifications, which includes de novo DNA methylation in plants and mammals.

In the reference plant Arabidopsis thaliana, the well-studied RdDM pathway begins with transcription of a noncoding TE RNA by RNA Polymerase IV (Pol IV). Pol IV is a plant-specific specialized derivative of RNA Polymerase II (Pol II) that utilizes multiple shared Pol II subunits (Haag and Pikaard 2011). Although the recruitment of Pol IV is not fully understood, transcriptionally-repressive histone modifications have been shown to guide Pol IV recruitment to previously-silenced TEs (Law et al. 2013). Once produced, the Pol IV-derived transcript is converted into double-stranded RNA by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and cleaved into distinctly sized 24 nucleotide (nt) siRNAs by DICER-LIKE 3 (DCL3). These 24nt siRNAs have been a long-standing hallmark of plant RdDM activity. The 24nt siRNAs are incorporated into the AGO4 and AGO6 proteins and guide these proteins to TE loci through their interaction with a TE scaffolding transcript produced from RNA Polymerase V (Pol V) (another plant-specific specialized derivative of Pol II) (Haag and Pikaard 2011). This Pol IV-dependent RdDM pathway (herein referred to as the Pol IV-RdDM pathway) utilizes 24nt siRNAs to constantly retarget methylation to previously silenced TEs, particularly small euchromatic TEs located near genes in order to maintain chromatin boundaries (Zemach et al. 2013). Hence, the Pol IV-RdDM pathway acts as a loop reinforcing methylation states at regions of heterochromatin and silenced TEs.
3.2.4 RDR6-RNA directed DNA Methylation (RDR6-RdDM)

The Pol IV-RdDM pathway has been previously shown to be necessary for the initiation of transgene silencing (Aufsatz et al. 2002; Chan et al. 2004; Greenberg et al. 2011), the corrective reestablishment of TE silencing (Ito et al. 2011; Teixeira et al. 2009b), and the maintenance of some TE silencing (Herr et al. 2005; Huettel et al. 2006). Pol IV is thought to transcribe regions of the genome that are already DNA methylated (Wierzbicki et al. 2012; Zheng et al. 2009). Therefore, the initiation of TE silencing solely by the Pol IV-RdDM pathway presents a chicken-and-egg dilemma: if Pol IV transcription is guided to previously heterochromatic and methylated TEs, how can TE silencing be initiated by Pol IV transcription? Hence I hypothesized that the RNA interference (RNAi) that causes post-transcriptional degradation of TE mRNAs might play a non-traditional role to methylate TE chromatin via RdDM. When transcriptionally active, Pol II-derived TE mRNAs can be post transcriptionally degraded into siRNAs that retarget complementary transcripts for further degradation in the cyclic RNA interference (RNAi) pathway (Chung et al. 2008; Sijen and Plasterk 2003). In Arabidopsis, RDR6, DCL2, DCL4, and AGO1 degrade some TE mRNAs to produce siRNAs of 21 to 22 nts (McCue et al. 2012c).

Recently, three publications by independent groups have co-discovered a mechanism of RdDM that acts independently of Pol IV, DCL3 and 24nt siRNAs. In the first two reports, this Pol IV-independent de novo DNA methylation utilizes SDE3 (an RNA helicase), SGS3 and SDE5 (both proteins of unknown function), AGO2, RDR1, RDR6, and NERD (a PHD finger protein) to initiate CHH DNA methylation (Garcia et
This pathway was shown to be necessary for the methylation of a *Copia* LTR retrotransposon on chromosome 5, a *Helitron* family TE on chromosome 1, and several intergenic regions. The methylation produced by this novel pathway was responsible for maintaining the repressed expression of these TEs. Additionally, the third report identified a role for RDR6, SDE3, AGO4, AGO6, PolIV and all four DCL proteins in targeting *de novo* CHH methylation to TAS (tasiRNA-producing) loci (Wu, Mao, Qi 2012). However, the removal of this methylation had no apparent consequence on the expression of TAS loci. Surprisingly, these DNA methylation pathways are composed of the following protein components that were previously characterized to only have roles in PTGS processing of TAS, virus, and/or TE transcripts: SDE3, SGS3, SDE5, AGO2, RDR1, RDR6, DCL1, DCL2 and DCL4. In addition, these pathways were able to function in the absence of PolIV, suggesting that they operate on siRNAs derived from RNA Polymerase II (PolII) transcripts.

This novel Pol IV-independent methylation pathway is not well understood, as factors such as SDE3 are important for methylation at individual *Copia* and *Helitron* elements, but dispensable for TAS methylation (Garcia et al. 2012; Pontier et al. 2012; Wu, Mao, Qi 2012). Therefore, I have particularly focused on RDR6, a protein identified as a necessary component of the Pol IV-independent DNA methylation of TAS loci as well as the single *Copia* and *Helitron* elements. I refer to this methylation pathway as RDR6-dependent *RNA directed DNA Methylation* (RDR6-RdDM). Recent publications have shown that RDR6-RdDM pathway plays a role in the maintenance of TE silencing. My research began by testing if this pathway also functions in the initiation,
reestabishment of TE silencing in transcriptionally active TEs. I focused specific experiments mainly on *Athila* LTR retrotransposons, the largest family of Arabidopsis TEs. If RDR6-RdDM pathway is involved in these processes of TE silencing, I aimed to determine the relative contributions of the RDR6-RdDM and Pol IV-RdDM pathways. I also wanted to determine the molecular mechanism of RDR6-RdDM pathway to find out what small RNA size class and other downstream proteins are utilized to direct *de novo* DNA methylation using this newly identified RdDM pathway. Knowing the function and mechanism of this newly identified RDR6-RdDM pathway would be significant because this will be a possible resolution to the Pol IV-RdDM initiation conundrum and would also shed light on common players between both the RdDM pathways. **Therefore, the RDR6-RdDM pathway could provide a link between posttranscriptional gene silencing/RNAi mediated by RDR6-dependent 21-22 nt siRNAs and the DNA methylation responsible for the initiation of Pol IV-RdDM and TE transgenerational silencing.**

3.3 Materials and Methods

3.3.1 Plant material

The mutant alleles of Arabidopsis (*Arabidopsis thaliana*) used in this study are described in Table 3.1. Plants were grown under standard long-day conditions at 23°C. Inflorescence tissue was used in each experiment unless otherwise noted.

3.3.2 Expression Analysis by Quantitative Reverse Transcription-PCR

Three biological replicates were performed for each genotype. Each replicate consisted of a nonoverlapping pool of individuals. Quantitative reverse transcription
(qRT)-PCR was performed and analyzed as described (McCue et al. 2012c), with the exception that complementary DNA was generated using an oligo(dT) primer and Tetro Reverse Transcriptase (Bioline). Quantitative real-time PCR was performed using SensiMix (Bioline) on a Mastercycler ep realplex thermocycler (Eppendorf). The At1g08200 gene was used as a reference in all qRT-PCRs. qRT-PCR primers are listed in Table 3.1.

3.3.3 Bisulfite Conversion and Sequencing

DNA was extracted via fractional precipitation, treated with RNaseA, and purified using phenol-chloroform extraction and precipitation. DNA was bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo Research). Converted DNA was amplified using EpiTaq DNA polymerase (Takara) using the primers listed in Table 3.1. PCR products were TOPO-TA cloned into pCR4 (Invitrogen) and sequenced. Analysis of individual DNA sequences was performed using the Kismeth analysis tool (Gruntman et al. 2008). For each bisulfite-converted DNA sample, an unmethylated exon of the gene At2g20610 was amplified, and at least three clones were sequenced to determine the C→T conversion efficiency, which is listed on each figure. Error bars represent Wilson score interval 95% confidence limits (Henderson et al. 2010). Differences between genotype methylation status were analyzed with a two-tailed Student’s t test. Each cloned sequence was given a methylated percentage based on the number of unconverted cytosines that were present out of total cytosines in the sequence. Each methylation context (CG, GHG, CHH, or total) was analyzed independently in this manner. Individual methylation percentages for each clone were grouped into populations by genotype, and
these populations were compared with Student’s t test. This statistical approach is more stringent compared with the \( \chi^2 \) tests previously used for statistical comparisons of bisulfite data (Henderson et al. 2010), as it takes into account clone-to-clone methylation variation. All of our bisulfite methylation comparisons that were statistically significant (\( P < 0.05 \)) using the Student’s t test were also statistically significant at \( P < 0.0001 \) using the \( \chi^2 \) approach.

### 3.3.4 Northern Blotting

Total RNA was isolated using Trizol reagent (Invitrogen), and small RNAs were concentrated with polyethylene glycol. Fourteen micrograms of small RNA-enriched RNA were loaded in each lane. Gel electrophoresis, blotting, and cross linking were performed as described (Pall et al. 2007b). *Athila* probes were generated by randomly degrading a \(^{32}\)P-labeled in vitro RNA transcript. PCR primers used to generate the probes are listed in Table 3.1. The miR161 control probe was generated by 5′ end labeling the DNA oligonucleotide shown in Table 3.1.

### 3.3.5 Small RNA Library Production, Sequencing, and Analysis

Inflorescence small RNA was isolated with Trizol reagent (Invitrogen) and concentrated using the mirVana microRNA Isolation Kit (Ambion). Libraries were produced using the TruSeq Small RNA Sample Preparation Kit (Illumina) as recommended by the manufacturer. Each library was barcoded and sequenced in the same lane of the Illumina Genome Analyzer IIx. The resulting sequences were demultiplexed, adapter trimmed, and filtered on length and quality, and transfer RNA/ribosomal RNA and low complexity reads were removed. Small RNAs were
matched to the Arabidopsis genome, and sequences that did not perfectly align were
discarded. Library size is normalized by calculating reads per million of 18 to 28 nt
genome-matched small RNAs. Small RNAs were also matched to The Arabidopsis
Information Resource TAIR10 (www.arabidopsis.org) and Repbase (www.girinst.org)
annotations of the TE portion of the Arabidopsis genome using bowtie. To best handle
multimapping sequences generated from repetitive regions of the genome, the bowtie
modifiers “−−best−M1−−strata” were employed (Salzberg et al. 2012). If more than one
genome perfect match for a TE siRNA exists, only one random match is assigned per
small RNA read. Counts of multimapping siRNAs were not amplified using our
approach, and we have not overestimated read number for high-copy TEs. The raw
sequencing and genome-matched small RNAs analyzed are available from the National
Center for Biotechnology Information Gene Expression Omnibus repository under
number GSE41755.

3.3.6 Chromatin Immunoprecipitation

ChIP experiments were performed as described in (Huettel et al. 2006), except for
the crosslinking. Crosslinking was performed as follows: Inflorescence tissue was soaked
in 1X PBS for 2 hours on ice, rinsed twice in sterile water, infiltrated under vacuum twice
for 10 minutes in 0.4M sucrose, 10mM Tris pH 8, 1mM EDTA, 1mM PMSF, 1%
formaldehyde, then quenched with infiltration of 0.1M glycine for 5 minutes under
vacuum, and then rinsed five times in sterile water. The qPCR was carried out in a 20µL
total reaction volume with an annealing temperature of 60°C. Chromatin was
immunoprecipitated with antibodies to RNA polymerase II CTD repeat YSPTSPS
(phospho S) (Abcam), FLAG (Sigma) and H3Ac (Millipore) at 5μg per IP, and immune complexes were collected with salmon sperm DNA-blocked protein A agarose beads (Millipore). The qPCR was carried out with primers shown in Table 3.1. The results shown in Figure 3.13 represent three independent biological replicates for each genotype.

3.4 Results

3.4.1 TE expression-dependent RNA directed DNA methylation (RdDM) activity

In the wild type Arabidopsis thaliana ecotype Columbia (Col), TEs are in silent state due to very tight epigenetic regulation. DNA methylation is a key player in epigenetic regulation. In order to study DNA methylation of TEs, I focused on the largest TE family in the Arabidopsis genome, the Athila family of LTR retrotransposons (Figure 3.1D). To characterize the methylation status of the Athila6 transcriptional regulatory region, we first mapped the transcriptional start site (TSS) of the gag/pol protein-coding transcript of the Athila6 subfamily and subsequently used the surrounding 350 bp of this 5′ LTR region for analysis of DNA methylation as shown in Figure 3.1D. In wt Col, the Athila6 LTR TSS has high levels of CG (94%) and intermediate levels of CHG symmetrical methylation (55%) and lower levels of CHH asymmetrical methylation (18%) (Figure 3.1E). CG and CHG methylation are likely higher than CHH methylation in each genotype tested due to siRNA-directed targeting of new methylation to cytosines in any sequence context, followed by S-phase replication of DNA methylation patterns only in the symmetrical CG and CHG contexts. Transcriptional activation of Athila6 is associated with and potentially caused by a reduction in specifically CG methylation (Figure 3.1E). I utilized the global TE transcriptional reactivation found in plants that
lack a functional DDM1 protein to determine the effect of DNA methylation upon transcriptional activation of Athila. DDM1 is a swi/snf family ATPase that coordinates linker histone chromatin compaction (Zemach et al. 2013)(Lippman et al. 2004). In ddm1 mutants CG methylation is reduced 2.4-fold to 39%, and CHG methylation is unaffected (Figure 3.1E). However, the high level of transcriptional activation in ddm1 mutants is accompanied by a 2.6-fold increase in CHH methylation of the Athila6 LTR (from 18% in wt Col to 47% in ddm1). I previously demonstrated that the expression of a second region, the degenerate protein-coding envelope (env) open reading frame, increases in successive inbreeding generations of ddm1 plants(McCue et al. 2012a)(see Chapter-2, Figure 2.3 C). I find that the same increase is observed with gag/pol transcripts, as ddm1 F2 plants (the second generation homozygous for ddm1) have lower expression than ddm1 F6 plants (sixth generation homozygous for ddm1) (Figure 3.1A). I find that this increase in Athila6 gag/pol expression is directly proportional to the level of LTR CHH methylation, as LTR CHH methylation increases from 18% in wt Col to 47% in ddm1 F2 and 68% in ddm1 F6 (Figure 3.1E). This CHH hypermethylation of Athila has been previously reported to occur in the wt pollen vegetative nucleus (Schoft et al. 2009), where Athila is known to be transcriptionally active (Slotkin et al. 2009). In addition, global TE activation in the pollen vegetative nucleus is associated with an increase in the CHH methylation levels of many different TE families (Calarco et al. 2012; Ibarra et al. 2012).

It is known in literature that CHH hypermethylation of TEs is dependent on the components of RNA directed DNA methylation pathway (I hereby refer it as POL IV-
RdDM pathway) (Figure 3.1E) ((Schoft et al. 2009); (Teixeira et al. 2009b)). I generated various ddm1 double mutants to determine the role of POL IV- RdDM pathway in transcriptional silencing of Athila. ddm1 double mutants with Pol IV-RdDM components, such as dcl3, rdr2, ago4 and pol IV (or nrpD1), displayed significant reduction in CHH methylation (28%, 13.8%, 19% and 24%, respectively). However, CHH methylation was not completely abolished in the above ddm1 double mutants, suggesting that an alternate route for the establishment of CHH methylation exists. The least amount of methylation in any sequence context was detected in the ddm1 ago6 (6.2%), ddm1drm2 (1.9%) and ddm1 pol V (or nrpE), (2.7%) double mutants (Figure 3.1E). The CHH methylation detected in the ddm1drm2 and ddm1 pol V double mutant is the minimum baseline that can be detected in these samples, as this is very close to my experimentally determined conversion efficiency of 98.3% (Figure 3.1E). The methylation status of the Athila6 TSS has a direct effect on the steady-state transcript accumulation of the gag/pol protein-coding transcript. Activation of expression is detected in ddm1 single mutants and increases (p < 0.001) when hyper methylation is reduced in ddm1 rdr2, ddm1 pol IV, and ddm1 pol V double mutants (Figure 3.1A). Athila6 expression was particularly high in ddm1 pol V double mutants, 10-fold higher compared with ddm1 single mutants. This expression level likely represents close to the maximum transcript accumulation potential of Athila6 when virtually uninhibited by repressive DNA methylation.

Transcriptional activation of Athila leads to production of small RNAs. To characterize the small RNAs contributing to RdDM of transcriptionally active TEs, I performed small RNA deep sequencing and small RNA northern blot. This analysis
focused on the Athila6A TSS in a series of single mutants in TE-transcriptionally silent epigenomes and double mutants in ddm1 TE-transcriptionally active epigenomes (Figure 3.1B and C) (Small RNA deep sequencing analysis of all the mutants for Athila6 LTR region was performed by Kaushik Panda). When TEs are transcriptionally silenced in the wt Col epigenome, I find the CHH methylation of the Athila6A TSS to be independent of the RdDM-targeted DNA methyltransferase DRM2, even though this region produces abundant 24nt siRNAs (Figure 3.3.B and C). drm2 single mutant has similar kind of CG, CHG and CHH methylation patterns as wt Col (Figure 3.1E). Thus, in spite of 24nt siRNA production by POL IV-RdDM pathway (Xie et al. 2004), POL IV-RdDM does not continually target methylation of this TE region. In the TE-activated epigenome of ddm1 plants, the methylation of the Athila6A TSS requires the POL IV-RdDM-dependent activity of DRM2 since in ddm1drm2 mutant all the CHH methylation (and nearly all total methylation) is nearly completely lost (Figure 3.1E). Thus, all DNA methylation when TEs are transcriptionally active is de novo methylation. In ddm1 plants, when Athila6 is transcriptionally active, 21 and 22 nt siRNAs accumulate (Figure 3.1B and C). In addition, 24 nt siRNAs accumulate to higher levels in ddm1 compared to wt Col. In ddm1 double mutants with the POL IV-RdDM pathway (ddm1 rdr2, ddm1 dcl3, and ddm1 pol IV), 24 nt siRNAs are lost when compared to the single ddm1 mutant (Figure 3.1C). In contrast to small RNA northern blot (Figure 3.1C), small RNA deep sequencing is even more precise technique to detect and produce quantitative small RNA accumulation data (Figure 3.1B). My quantitative analysis of siRNA accumulation through small RNA deep sequencing demonstrates that when TEs are activated, ddm1 pol
IV mutants continue to produce low-levels of 24nt siRNAs (Figure 3.1B), suggesting that some 24nt siRNA production is dependent upon POL II-derived TE mRNAs, as has been recently theorized (Mari-Ordonez et al. 2013). These 24mers are DCL3-dependent, because ddm1 dcl3 double mutants lose all Athila6 TSS 24 nt siRNAs (Figure 3.1E), confirming that there is an alternative pathway besides the production of 24 nt siRNAs for the establishment of the 28% CHH methylation in this double mutant. The ddm1 ago4, ddm1 ago6 and ddm1 pol V double mutants do not lose siRNAs compared to ddm1 single mutants, suggesting that AGO4, AGO6 and POL V function downstream of siRNA production (Figure 3.1B and C). This demonstrates that these RdDM factors are important for DNA methylation but not for siRNA production, and that in ddm1 double mutants with Pol IV-RdDM components (including pol IV and dcl3) the level of TE siRNAs increases, displaying a positive relationship between TE transcript accumulation shown in Figure 3.1A and TE 21-22nt siRNA production shown in Figure 3.1B and C.

3.4.2 The RDR6-RdDM pathway plays a role in the establishment of CHH methylation upon TE transcriptional activation

Since CHH methylation was not completely abolished in the ddm1 double mutants with Pol IV-RdDM components, an alternate pathway must exist which has the ability to circumvent Pol IV dependent RdDM pathway in the establishment of CHH DNA methylation. I hypothesize that RNAi pathway which plays a role in post transcriptional gene silencing might play a role in transcriptional gene silencing via RdDM. I aimed to understand the alternate pathway of CHH hypermethylation upon Athila transcriptional activation in ddm1 mutant plants, and to specifically determine if
the RNAi pathway is responsible for this methylation. Therefore, I generated a series of
*ddm1* double mutants with RNAi components to investigate DNA methylation patterns. I
directly compared *ddm1* double mutants to *ddm1* F2 generation plants to minimize the
impact of the generational effect of *ddm1* inbreeding on our results, as the double mutants
vary between the F2 and F4 generation. Interestingly, I found that CHH hypermethylation
was also reduced in plants that were double mutants for *ddm1* and various components of
the RNAi pathway. In particular, *ddm1 rdr6* has a 3.9-fold reduction in CHH DNA
methylation compared to *ddm1* single mutants (from 47% in *ddm1* F2 single mutants to
12% in *ddm1 rdr6*) *(Figure 3.2 C).* As a control, I assayed sibling plants from a
segregating family, both of which were *ddm1* homozygous mutants but differed in the
presence or absence of a functional RDR6 gene, and found a statistically significant (*P <
0.001*) dependence on RDR6 for establishing CHH hypermethylation *(Figure 3.2 C).* To
determine if CHH hypermethylation of the *Athila6* LTR plays a role in repressing *gag/pol*
expression; I performed qRT-PCR on *ddm1 rdr6* double mutants. My analysis of *ddm1
rdr6* demonstrates that when the CHH hypermethylation is removed *(Figure 3.2 C),
expression increases *(Figure 3.2 A).* This suggests that upon transcriptional activation,
the RNAi pathway, which is henceforth referred as RDR6- RdDM pathway is acting to
inhibit expression levels of active TEs. I then set out to determine if other previously
identified components of the RDR6- RdDM pathway function to increase the CHH
methylation of the *Athila6* LTR in the mutant *ddm1* background. I found that CHH
methylation is significantly decreased in both *ddm1 dcl2* (33%) and *ddm1 dcl4* (19%)
double mutants *(Figure 3.2 C), confirming that these two genes have a role in RDR6-
RdDM ((Pontier et al. 2012);(Wu, Mao, Qi 2012)). In contrast to *ddm1 rdr6*, in *ddm1 dcl2* there’s no significant change in *Athila* gag/pol expression when compared to *ddm1* (Figure 3.2 A). Also, expression of *Athila* gag/pol decreases in *ddm1 dcl4* when compared to *ddm1* (Figure 3.2 A). It’s not known why there is no correlation between expression and methylation patterns for *ddm1 dicer* double mutants. Lastly, three *ddm1* double mutants, *ddm1 ago6, ddm1drm2* and *ddm1 pol V*, displayed almost no CHH methylation (2%) (Figure 3.2 C). In addition to Pol V, AGO6 and DRM2 have been identified as common components and convergence point of both the RDR6- RdDM and POL IV- RdDM pathways ((Wu, Mao, Qi 2012)).

To further illuminate which small RNAs are driving CHH hypermethylation of the *Athila6* LTR TSS, I analyzed small RNA accumulation in the *ddm1* double mutants with RDR6-RdDM pathway components. When *Athila6* is transcriptionally activated in the *ddm1*, 21 and 22 nt siRNAs accumulate (Figure 3.1C and 3.4 B). *ddm1* RDR6-RdDM pathway double mutants have either absent (as in *ddm1 rdr6*) or reduced (as in *ddm1 dcl2* and *ddm1 dcl4*) levels of 21 and 22 nt siRNAs. In *ddm1 dcl2*, the same siRNA size classes are found compared to the *ddm1* single mutant; however, the quantities are reduced, which supports the role of DCL2-dependent siRNAs in the re-targeting of RDR6 activity and the amplification of secondary siRNAs (Figure 3.1C) ((Chen 2010)). In *ddm1 rdr6, ddm1 dcl2* and *ddm1 dcl4* the levels of 24 nt siRNAs also decrease, suggesting that 21/22 nt siRNA production is upstream of 24 nt siRNA production and that the 21/22 nt siRNAs are potentially required to direct Pol IV or to amplify the POL IV- RdDM pathway’s production of 24 nt siRNAs (Figure 3.1C). Lastly, *ddm1 pol V* has
reduced levels of 24 nt siRNAs and increased 21 nt siRNAs (Figure 3.1C). This suggests that the PolV polymerase is not required to generate any particular siRNA size class, but its absence causes a shift towards RDR6- RdDM pathway-dependent siRNAs, and away from POL IV- RdDM and PolIV-dependent siRNAs. Therefore, from my methylation and small RNA analysis in Figure 3.2, I conclude that the RDR6- RdDM pathway is responsible for de novo methylation at an active TE, which is associated with the production of 21 and 22 nt siRNAs, as well as the amplification of 24 nt siRNAs.

My results from Athila6, indicate that RDR6-RdDM and Pol IV-RdDM function independently to establish CHH hypermethylation on transcriptionally active TEs. To determine if these are the only two pathways responsible for establishing TE methylation, I constructed a ddm1 pol IV rdr6 triple mutant, inactivating critical components of both the RDR6-RdDM and Pol IV-RdDM pathways at the same time and pol IV rdr6 double mutant. As expected in pol IV rdr6 double mutant, methylation in all contexts were similar to wt Col (84% CG, 50.6% CHG and 14.6% CHH) (Figure 3.2 C) indicating that without transcriptional activation there’s no establishment of de novo methylation by both the RdDM pathways. Interestingly, the ddm1 pol IV rdr6 triple mutant has very low levels of Athila6 LTR TSS methylation in all sequence contexts (10.6% CG, 2.2% CHG, and 1.6% CHH), lower than both the ddm1 rdr6 and ddm1 pol IV double mutants that inactivate RDR6-RdDM and Pol IV-RdDM individually (Figure 3.2 C). I have shown plant picture of ddm1 pol IV rdr6 triple mutant in Figure 3.2 D to indicate that when TEs are transcriptionally active (ddm1) and lose major components of the two RdDM pathways (pol IV and rdr6) the plants are severely affected with no seed production due
to loss of three different transcriptional gene silencing pathways. I have shown $+/ddm1\ pol\ IV\ rdr6$ plant picture as control to point out that with partial restoration of DDM1 the plants are healthier than $ddm1\ pol\ IV\ rdr6$ triple mutant (Figure 3.2 D). Therefore, I conclude that RDR6 and Pol IV are the only two entry points into RdDM responsible for the expression-dependent CHH hypermethylation of Athila6, and these pathways work independently and additively on their target(s) to establish TE methylation levels. When the 21-22nt siRNAs are specifically lost in the $ddm1\ pol\ IV\ rdr6$ triple mutant, all CHH methylation targeted by RdDM is also lost (Figure 3.2 B and C). I also generated $ddm1\ dcl2\ dcl4$ triple mutant, to analyze the effect of loss of dicers that produce 21-22nt siRNAs in transcriptionally active background (Figure 3.2 D). I found that the $ddm1\ dcl2\ dcl4$ triple mutant has 2-fold increase in CG methylation, 6-fold increase in CHG methylation and 1.5-fold increase in CHH methylation in comparison to $ddm1\ pol\ IV\ rdr6$ triple mutant. This increase in methylation in all sequence contexts when compared to $ddm1\ pol\ IV\ rdr6$ triple mutant may be because of the activity of either DICER3 or DICER1. In addition, we know that the 21-22nt siRNAs that drive RDR6-RdDM are not processed from Pol IV-dependent transcripts. Previous work demonstrated that Pol IV-derived transcripts are not processed into siRNAs in the absence of RDR2 (Kasschau et al. 2007). I have previously shown that $ddm1\ rdr2$ double mutants retain the same level of Athila6A TSS methylation as $ddm1\ dcl3$ double mutants (Figure 3.1E), demonstrating that Pol IV transcripts are not required for RDR6-RdDM. From this combined data I conclude that RDR6-RdDM is driven by Pol IV-independent 21-22nt siRNAs.
3.4.3 Other players involved in RDR6-RdDM

Next, I investigated the role of other proteins such as RDR1, DCL1, NERD, and SDE3 in RDR6-RdDM. Traditionally DCL1 is known to play a role in the biogenesis of tasiRNAs (Ghildiyal and Zamore 2009b). However, Wu et al. have now shown that DNA methylation of TAS3 loci requires DCL1 protein (Wu, Mao, Qi 2012). Silencing Defective 3 or SDE3 (RNA helicase) has always been known to play a role in defense against viruses through amplification of post transcriptional gene silencing pathway (Dalmay et al. 2001). Recently Wu et al. and Garcia et al. have shown that SDE3 plays a role in TAS3 loci, Copia retrotransposon and several intergenic regions DNA methylation (Garcia et al. 2012; Wu, Mao, Qi 2012). A novel plant protein Needed for RDR2-independent DNA methylation or NERD was recently identified to play a role in POL IV independent methylation of repeats, pseudogenes and intergenic regions (Pontier et al. 2012). RDR1 is known to act in anti viral defense of some RNA viruses by producing small RNAs (Rakhshandehroo et al. 2009) (Wang et al. 2014). Along with NERD, Pontier et al. have also shown that RDR1 plays a role in non canonical 21nt siRNAs mediated RdDM pathway (Pontier et al. 2012). To test the role of RDR1, DCL1, NERD, and SDE3 in RDR6-RdDM, I generated ddm1 double mutants with these proteins. The SDE3 line was generated by using an amiRNA to knock down expression because knocking out the SDE3 gene results in embryonic lethal plants. First I looked at the expression of Athila6 gag/pol protein coding region in the above generated ddm1 double mutants. I’ve only tested DCL1, NERD and SDE3 in this expression analysis. Upon transcriptional activation and loss of either DCL1 or NERD seems to have very little
effect on *Athila* expression when compared to *ddm1* single mutant (*Figure 3.3 A*). *ddm1/sde3amiRNA* has higher expression of *Athila* than *ddm1/dcl1* and *ddm1/nerd* double mutants but not as high as *ddm1*. This makes sense because SDE3 is a RNA helicase and cofactor of RDR6 in PTGS and hence similar to RDR6’s role in repression of *Athila* through RNAi (*Figure 3.2 A*), SDE3 is also playing a role in post transcriptional silencing of *Athila* (*Figure 3.3 A*). Next, I studied *Athila6* LTR TSS methylation in *ddm1/rdr1*, *ddm1/dcl1*, *ddm1/nerd* and *ddm1/sde3amiRNA* double mutants. *De novo* methylation observed in *ddm1* is lost in *ddm1/rdr1*, *ddm1/dcl1*, *ddm1/nerd* and *ddm1/sde3amiRNA* double mutants (*Figure 3.3 B*). Thus, through this analysis, I conclude that RDR1, DCL1, SDE3 and NERD are playing role in the *de novo* methylation of *Athila* through RDR6-RdDM pathway. Further investigation is required to determine the exact mechanism of action of these above proteins in RDR6 dependent *de novo* methylation.

3.4.4 Examination of an additional TE target of RDR6-RdDM demonstrates that 21 and/or 22 nt siRNAs direct methylation

Previous reports have suggested that the RDR6-RdDM is responsible for methylating target DNA through the production of 21nt siRNAs ((Garcia et al. 2012);(Pontier et al. 2012);(Wu, Mao, Qi 2012)). My data agrees with this suggestion, as CHH hypermethylation can be established in a *ddm1 dcl3* mutant without detectable levels of 24 nt siRNAs by sRNA deep sequencing (*Figure 3.1 B*). However, my deep sequencing and northern hybridization data of *ddm1/rdr6* show that there is decrease in 24 nt siRNAs and loss of 21-22nt siRNAs (*Figure 3.1 C* and *3.2 B*), and therefore it is
still a possibility that these 24 nt siRNAs are directing methylation, and not the 21/22 nt siRNAs through RDR6-RdDM pathway. Therefore, direct methylation through a small RNA size other than 24 nt has not been formally demonstrated. We wanted to know whether the 21/22 or 24 nt siRNAs are directly responsible for RDR6-RdDM methylation. Analysis of the siRNAs responsible for RDR6-RdDM of *Athila* is confounded by the fact that *Athila* LTR TSS that displays RDR6-dependent CHH hypermethylation has increases in 21, 22 and 24 nt siRNAs (Figure 3.1 B and C). Since I find that all *Athila6* TSS siRNA size classes are dependent on RDR6, I cannot conclude whether increases in 21/22 or 24 nt siRNAs are required for RDR6-RdDM. To address this problem, we searched our small RNA sequencing datasets for a TE sub-family that has RDR6-dependent 21 and 22 nt siRNAs in *ddm1* in single mutants, and 24 nt siRNAs that are not RDR6-dependent. We identified the TE sub-family *AtENSPM6* a DNA transposon in Arabidopsis genome, which has increased 21 and 22 nt siRNAs in *ddm1* compared to wt Col (Figure 3.4 A) (Small RNA deep sequencing analysis was performed by Kaushik Panda). These 21 and 22 nt siRNAs are dependent on RDR6, as they are lost in *ddm1 rdr6* double mutants (Figure 3.4 A). Importantly, the 24 nt siRNAs are not RDR6-dependent, as they slightly increase in *ddm1 rdr6* compared to *ddm1* single mutants (Figure 3.4 A). We further analyzed *AtENSPM6* to determine the specific region of the element that shows increased levels of 21 and 22 nt siRNAs from wt Col to *ddm1*, and consistent levels of 24 nt siRNAs in *ddm1* and *ddm1 rdr6*. We identified the first and second exon of *AtENSPM6* for further analysis (red box, Figure 3.4 D). Examination of this region shows equal amounts of 24 nt siRNAs in *ddm1* and *ddm1 rdr6* (Figure 3.4 A).
Bisulfite DNA methylation analysis of this region identified strong CHH hypermethylation in \textit{ddm1} single mutants (from 9\% CHH in wt Col to 63\% in \textit{ddm1}) (Figure 3.4 B) (Andrea McCue performed bisulfite analysis of some of these mutants). In \textit{ddm1 rdr6} double mutants, this CHH hypermethylation is significantly reduced to 38\% (p<0.05) (Figure 3.4 B), demonstrating that the specific loss of the 21 and/or 22 nt siRNAs are responsible for a portion of the CHH hypermethylation. In addition, the CG and CHG methylation levels of \textit{AtENSPM6} also decrease in \textit{ddm1 rdr6} compared to \textit{ddm1} single mutants (Figure 3.4 B).

In addition to being targeted for RDR6-RdDM methylation, this same region of \textit{AtENSPM6} is also targeted by 24 nt siRNAs and CHH methylation via Pol IV-RdDM. Mutations in \textit{pol IV} in double mutant combination with \textit{ddm1} also show reduced methylation compared with \textit{ddm1} single mutants (Figure 3.4 B). Mutations in either the Pol IV-RdDM or RDR6-RdDM pathway each produce roughly one-half of the level of \textit{AtENSPM6} CHH methylation (both 32\%) compared with when the TE is active in \textit{ddm1} mutant plants (63\%; Figure 3.4 B). Akin to \textit{Athila} LTR TSS, the \textit{ddm1 pol IV rdr6} triple mutant has very low levels of \textit{Athila6} LTR TSS methylation in all sequence contexts (25\% CG, 9.36\% CHG, and 3.89\% CHH), lower than both the \textit{ddm1 rdr6} and \textit{ddm1 pol IV} double mutants that inactivate RDR6-RdDM and Pol IV-RdDM individually (Figure 3.4 B). This demonstrates that RDR6-RdDM and Pol IV-RdDM function on the same TE to establish CHH hypermethylation upon TE transcriptional activation. Similar to \textit{Athila} LTR TSS, mutations in \textit{pol V} and \textit{ago6} in combination with \textit{ddm1} show a more severe reduction compared with \textit{ddm1 rdr6} and \textit{ddm1 pol IV} (Figure 3.4 B), corroborating
previous findings that suggest that Pol V is a shared component of both the RDR6-RdDM and Pol IV-RdDM pathways (Wu, Mao, Qi 2012) and also showing that AGO6 also acts as common factor to both these RdDM pathways. Examination of the AtENSPM6 steady-state polyadenylated expression levels demonstrates that the methylation of this coding region of the AtENSPM6 element examined in Figure 3.4 B does not correlate with or control the element’s transcriptional activity. AtENSPM6 activation in a ddm1 single mutant displays the same expression level as ddm1 pol IV and ddm1 pol V double mutants (Figure 3.4 C), which have significantly less AtENSPM6 methylation (Figure 3.4 B). ddm1 rdr6 double mutants have increased expression levels compared with the ddm1 single or other ddm1 double mutants (Figure 3.4 C); however, this change is not statistically significant. Rather than acting through methylation and on the transcriptional level, in this case, RDR6 may be acting posttranscriptionally to degrade the activated AtENSPM6 mRNA into 21 and 22 nt siRNAs. Therefore, in contrast to Athila, RDR6-RdDM and Pol IV-RdDM play no role in the repression of AtENSPM6 expression, once activated in ddm1, although AtENSPM6 is targeted for methylation by both Pol IV-RdDM and RDR6-RdDM. This is similar to the tasiRNA-generating loci, where methylation induced by RDR6-RdDM plays no role in altering gene expression levels (Wu, Mao, Qi 2012), as this methylation is likely too distant from the promoter of the tasiRNA-generating loci or AtENSPM6 to exert an influence on transcriptional rates. This data suggests that by establishing CHH hypermethylation, the RDR6-RdDM pathway is likely attempting to re-methylate and re-silence these active TEs. This analysis of AtENSPM6 also demonstrates that locating sites of RDR6-dependent 21 and 22 nt siRNA
production can be used to identify TE or other candidate genomic locations of RDR6-RdDM activity.

### 3.4.5 The RDR6-RdDM and POL IV- RdDM pathways do not contribute to the maintenance of a trans-generationally silenced TE

To determine the functional role of RDR6-RdDM and POL IV- RdDM pathways in maintaining long-term trans-generational TE silencing, I tested the contribution of both pathways on maintaining silenced TE expression and DNA methylation levels. It is known in the literature that DDM1 and MET1 (Methyl Transferase 1) are responsible for maintaining CG methylation (Law and Jacobsen 2010b) while Chromo methyltransferase 3 (CMT3) is responsible for maintaining CHG methylation and CHH methylation is maintained by CMT2 (Stroud et al. 2014). I performed bisulfite sequencing on cmt3 and cmt2 single mutants to see the effect of these maintenance methyltransferases on Athila LTR TSS methylation. I found that in cmt3, CHG methylation was decreased to 8% from 54% in wt Col and there was not much change in CG and CHH methylation levels whereas in cmt2, CHH methylation was decreased to 2.5% from 15.9% in wt Col. Thus CHG and CHH methylation of Athila6 LTR TSS is maintained in transcriptionally silent background by CMT3 and CMT2 respectively. I confirmed that expression of an Athila6 protein-coding region required for retrotransposition, gag/pol, is activated when it loses CG DNA methylation in ddm1 (CG 52%, CHG 71%, CHH 68%) and met1 mutants compared to wild-type plants (wt Col) (Figure 3.5 A and B) (Lippman et al. 2004; McCue et al. 2012d). In met1 mutants, CG methylation is reduced by half (from 94% in wt Col to 35% in met1), while CHG and
CHH methylation (44% and 19% respectively) remains unaltered (Figure 3.5 B). I generated $ddm1$ double mutants with either $met1$ or $cmt2$ to see the effect of maintenance methyl transferases when TEs are active. Both $ddm1/met1$ and $ddm1/cmt2$ double mutants showed decrease in CG methylation (29.6% and 29.91% respectively) when compared to $ddm1$, $met1$ and $cmt2$ single mutants respectively. CHG methylation (43.54%) was decreased in $ddm1/met1$ when compared to $ddm1$, but in comparison to $met1$ CHG methylation remains unaltered. CHH methylation was also decreased in $ddm1/met1$ (8.76%) when compared to $ddm1$ and $met1$ single mutants (Figure 3.5 B). $ddm1/cmt2$ showed decrease in CHG methylation (32.23%) when compared to $ddm1$ and $cmt2$. Also, $ddm1/cmt2$ (10.35%) showed decrease in CHH methylation when compared to $ddm1$ and increase when compared to $cmt2$ (Figure 3.5 B). Therefore this data preliminarily suggests that when there is transcriptional activation of TEs ($ddm1$) and there is simultaneous loss of a maintenance methyl transferase ($met1$ or $cmt2$) there is more reduction in symmetrical CG methylation than that observed in either of the single maintenance methylation mutants ($ddm1$, $met1$ and $cmt2$). The combination of either DDM1 and MET1 or DDM1 and CMT2 might be causing this drastic reduction in CG methylation. It is known in literature that hyper CHH methylation in $ddm1$ mutants is due to RdDM activity ((Schoft et al. 2009);(Teixeira et al. 2009b)) and hence I would expect that in both $ddm1/met1$ and $ddm1/cmt2$ double mutants there will be not much effect on CHH methylation. Surprisingly, hyper asymmetrical (CHH) methylation seen in $ddm1$ mutants is reduced in $ddm1/met1$ and $ddm1/cmt2$ double mutants when compared to $ddm1$ single mutant. The reason for this reduction in CHH methylation in these double
mutants is not yet known. There is 5-fold increase in CHH methylation in *ddm1/cmt2* when compared to *cmt2* single mutant and this increase may be due to the activity of RdDM and DRM2 when TEs are transcriptionally active as shown in Figure 3.1 E.

*Athila6* *gag/pol* expression is not detected by qRT-PCR in plants bearing mutations in components of the RDR6-RdDM pathway (*rdr6* and *dcl2*) or mutants of the POL IV- RdDM pathway (*rdr2, pol IV, and pol V*) as shown in Figure 3.8 A. These data demonstrate that the RDR6-RdDM and POL IV- RdDM pathways are not responsible for maintaining the transcriptional repression of a long-term *trans*-generationally repressed TE. We used bisulfite sequencing of the transcriptional start site (TSS) region of the *gag/pol* transcript to determine if the RDR6-RdDM or POL IV- RdDM pathways play a role in maintaining CG, CHG or CHH methylation of the *trans*-generationally silenced *Athila6* element. We mapped the TSS of *gag/pol* to the *Athila6* 5’ LTR (Nuthikattu et al. 2013) (Dalen Fultz performed this methylation analysis). In wt Col, the *Athila6* LTR TSS has high levels of CG methylation (94%), intermediate levels of CHG methylation (55%) and lower levels of CHH methylation (18%) (Figure 3.5 B). DNA methylation levels were not significantly different (p>0.05) in any sequence context between wt Col and mutants of the POL IV- RdDM pathway (*rdr2, dcl3, pol IV and pol V*) or mutants of the RDR6-RdDM pathway (*rdr6 and dcl2*) (Figure 1B). Therefore, I conclude that the *trans*-generational transcriptional repression of TEs, such as *Athila6*, is controlled independently of either the POL IV- RdDM or RDR6-RdDM pathways. Interestingly, we were reproducibly able to detect >10% CHH methylation of the *Athila6* LTR TSS in *pol*
and rdr6 single mutants (Figure 3.5 B), suggesting that an additional PolIV- and RDR6-independent mechanism exists to maintain CHH methylation levels.

3.4.6 The RDR6-RdDM Pathway Functions to Correctively Reestablish Methylation and Silencing

The POL IV-RdDM pathway is necessary to re-silence TEs that were previously active in a ddm1 mutant ((Teixeira et al. 2009b)). I set out to determine if the RDR6-RdDM pathway plays a role in the re-silencing of active TEs, and, if so, to identify the relative contributions of POL IV-RdDM and RDR6-RdDM to this process. I conducted a series of genetic crosses between mutant ddm1 and WT DDM1 plants (Table 3.2). For the majority of TEs, the transcriptionally reactivated state in ddm1 mutants is inherited, and these TEs remain active upon crossing to a wt plant and the subsequent formation of ddm1/+ heterozygous progeny ((Kakutani et al. 1999)). I find that, independent of the gender of the ddm1 mutant parent, Athila6 re-silencing occurs very efficiently and quickly, as Athila6 expression is reduced 52-fold back to the levels observed in wt Col (Figure 3.6 A). I used ddm1 double mutants to perform an Athila6 re-silencing assay with parents that were homozygous for a mutation in a second gene that produces a component either involved in the RDR6-RdDM or POL IV-RdDM pathway (see Table 3.2). I found that Athila expression is not fully re-silenced in mutant backgrounds of either the POL IV-RdDM pathway or the RDR6-RdDM pathway (Figure 3.6 A). The POL IV-RdDM pathway mutants (rdr2, dcl3, pol IV) play a larger role in this re-silencing, for on average, their Athila6 expression levels are 9-fold higher than the RDR6-RdDM pathway mutants. However, the RDR6-RdDM pathway mutants (rdr6, and dcl2) do show statistically
significant (p<0.01) expression that is greater than the control cross of wt Col x ddm1, demonstrating that both the POL IV-RdDM and RDR6-RdDM pathways play a role in the re-silencing of Athila TE expression.

I next determined if the lack of full Athila re-silencing in a +/-ddm1 rdr6 mutant background is due to incomplete reestablishment of DNA methylation. I performed bisulfite DNA sequencing on the progeny of the crosses in Table 3.2 and found that the expression levels in Figure 3.6 A correlate with the methylation levels of the Athila6 LTR TSS. The high level of CG and low level of CHH methylation found in wt Col is restored in the progeny of wt Col x ddm1 (86% CG and 9% CHH)(Figure 3.6 D). However, when this cross is performed in an rdr6 background, CG and CHH methylation levels are only partially restored (57% CG and 27% CHH) compared to the +/-ddm1 control cross. This data demonstrates that RDR6 is necessary for the complete re-methylation and re-silencing of Athila6 TEs. In the analysis of ddm1 double mutants, the polV mutant behaves differently than the rest of the POL IV-RdDM components. In the polV mutant background, ddm1 heterozygous progeny from Table 3.2 display very low CG and CHH methylation levels (33% CG and 8% CHH) (Figure 3.6 D), resulting in the highest levels of Athila6 expression (Figure 3.6 A). Athila6A is almost completely resilienced in the absence of DCL3 or RDR6 (Figure 3.6 A), demonstrating that neither 24nt nor 21-22nt siRNAs are solely required for resiliencing (Figure 3.6 E). However, the lack of resiliencing of Athila6A in the ago6 mutant background is similar to that of pol V mutants. This demonstrates that, unlike DCL3 or RDR6 that only function in one RdDM pathway, Pol V and AGO6 function in both RdDM pathways to fully reestablish Athila6A
silencing. Therefore, these combined experiments genetically demonstrate that AGO6 functions not only in the Pol IV-RdDM pathway, but also in the RDR6-RdDM pathway. The combined data from my crosses performed in Table 3.2 demonstrate that the RDR6-RdDM pathway plays a role, albeit a minor role compared to POL IV-RdDM, in the efficient corrective reestablishment of TE DNA methylation levels and TE re-silencing. I then selfed the plants used in Figure 3.6 A to generate F2 progeny and performed expression analysis of Athila gag/pol protein coding region (Figure 3.6 B). As the generation progresses from F1 to F2, POL IV-RdDM seems to be playing a huge role in re-silencing Athila when compared to RDR6-RdDM pathway in re-silencing. This evident from increased expression levels in the F2 progeny from crosses with (rdr2, pol IV and pol V) POL IV-RdDM pathway mutants and decrease in expression to nearly no expression in RDR6-RdDM pathway mutants (rdr6, dcl2 and dcl4) (Figure 3.6 B).

The biological relevance of the resiliencing of TEs by RDR6-RdDM would be increased if it were shown that this pathway works to regulate TEs other than Athila6. Another TE that I was interested in investigating is ENSPM5. In order to investigate the expression levels of ENSPM5, I performed qPCR. The backgrounds analyzed in this experiment include the following: Col, ddm1, ddm1 het, ddm1 het; rdr6 homo, ddm1 het; dcl2 homo, ddm1 het; dcl4 homo, ddm1 het; rdr2 homo, ddm1 het; dcl3 homo, ddm1 het; pol4 homo, ddm1 het; pol5 homo. In each one of the mutant backgrounds, ENSPM5 shows similar levels of expression to Athila. ENSPM5 levels of expression increase as expected when Pol4-RdDM components are lost, but also when RDR6-RdDM components are lost as well (Figure 3.6 C). Thus, RDR6-RdDM functions more
ubiquitously than exclusively working to silence *Athila6*. The data from this experiment shows that when its components are lost, the expression of ENSPM5 significantly increases when compared to *ddm1* background levels. Since *ENSPM5* is a DNA element, this shows that RDR6-RdDM can function to regulate a wide array of TEs in the genome. Future work will be needed to identify other targets of this pathway in the genome to understand its importance in the maintenance of genomic integrity.

### 3.4.7 Developmental time point in which re silencing of TEs occurs in Arabidopsis

After I identified that one of the biological functions of RDR6-RdDM is transgenerational resilencing and corrective reestablishment of TE methylation, I wanted to determine the developmental time point of Arabidopsis during which this resilencing occurs. For the majority of TEs, the transcriptionally reactivated state in *ddm1* mutants is inherited, and TEs retain transcriptional activity during the subsequent formation of +/-*ddm1* heterozygous progeny (Kakutani et al. 1999). In contrast, I find that *Athila6* resilencing of *gag/pol* expression occurs very efficiently and quickly in an F1 +/-*ddm1* heterozygote ([Figure 3.6 A](#)), which is likely due to the abundant siRNAs these TEs produce when active ([Figure 3.1 B](#)). The developmental stage where *Athila* expression between +/-*ddm1 rdr6* (*ddm1* heterozygous; *rdr6* homozygous) plants and +/-*ddm1* (*ddm1* heterozygous) plants remains the same is the time where RDR6-RdDM initially starts to repress to TE expression. I had already investigated the role of RDR6-RdDM in re silencing TEs in inflorescence tissue which is one of the mature tissues in Arabidopsis ([Figure 3.6](#)). Next, I began the re silencing analysis in tissues involved in earlier stages of development such as leaf, seedling and seed ([Figure 3.7](#)).
I performed expression analysis of *Athila*6 transcripts in the following F1 progeny *ddm1* het, *rdr6* homo;*ddm1* het, *dcl2* homo;*ddm1* het, and *pol IV* homo;*ddm1* het from crosses mentioned in Table 3.2. In leaf, seedling and seed tissues, *Athila* expression in *ddm1* het is less than the *ddm1* het; *rdr6* mutants (Figure 3.7 A, C, D). However, the difference in expression levels between +/∗ddm1 and +/∗ddm1 rdr6 decreases as I move from leaf to seed tissue. Though I could not find out the exact developmental time point in which resilingencing of TEs occurs by RDR6-RdDM, I did show that resilingencing of *Athila* occurs as early as in younger tissues, such as seedling and seed (Figure 3.7 A and D). Future work would be to test tissues from earlier stages of embryogenesis to identify the exact developmental stage before which or during which resilingencing starts.

Next, I analyzed RDR6-RdDM’s role in corrective re establishment of methylation in seedling tissue. Through methylation analysis, I found that unlike hyper CHH methylation (68%) seen in inflorescence tissue of *ddm1* plants (Figure 3.1 E), only 17% CHH methylation is observed in seedling tissue of *ddm1* plants (Figure 3.7 B). Then I wanted to determine if the lack of full *Athila* re-silencing in a +/∗ddm1 rdr6 mutant background is due to incomplete reestablisment of DNA methylation. The high level of CG and low level of CHH methylation found in wt Col is restored in the progeny of wt Col x *ddm1* (89% CG and 12% CHH)(Figure 3.7 B). However, when this cross is performed in an rdr6 background, CG and CHH methylation levels are only partially restored (53% CG and 26% CHH) compared to the +/∗ddm1 control cross. This data demonstrates that RDR6 is necessary for the complete re-methylation and re-silencing of *Athila*6 TEs in inflorescence as well as seedling tissues of Arabidopsis.
3.4.8 Role of RDR6-RdDM and POL IV-RdDM throughout Arabidopsis development

Both the RdDM pathways play a role in *de novo* methylation in inflorescence tissue of Arabidopsis (Figure 3.1 and 3.2). Next, I wanted to test the contribution of both the RDR6-RdDM and POL IV-RdDM pathways in different developmental stages of Arabidopsis such as pollen, seedling, meristem, leaf, inflorescence and silique tissues (Figure 3.8 A). I also tested the role played by AGO6, a common factor between both the pathways in different developmental stages of Arabidopsis. I performed bisulfite analysis on Col, *ddm1, ddm1/rdr6, ddm1/pol IV* and *ddm1/ago6* in the above mentioned tissue types. I did not test few genotypes for particular type of tissues. Through methylation analysis, I found that unlike hyper CHH methylation in the *Athila6 LTR TSS* (68%) seen in inflorescence tissue of *ddm1* plants (Figure 3.1 E), CHH methylation observed in pollen, seedling, meristem, leaf and silique tissues of *ddm1* plants is only 18%, 17%, 14%, 23% and 21% respectively (Figure 3.8 B-D,G). *Athila6 LTR TSS* methylation is almost similar in all the different types of tissues from wt Col with high levels of CG and CHG symmetrical methylation and low levels of CHH methylation (Figure 3.8 B-D, G).

Next, I calculated the percentage CHH methylation dependent on POL IV, RDR6 and AGO6 (Figure 3.8 F). In pollen and leaf tissues, POL IV (63% and 61% respectively) contributes more to CHH methylation than RDR6 (54% and 42% respectively). Whereas in inflorescence and silique tissues, RDR6 (82% and 93% respectively) contributes more to CHH methylation than POL IV (69% and 79% respectively). Both POL IV and RDR6 contribute to 37% of CHH methylation in seedling tissue. In meristem tissue, RDR6
contributes to 41% of CHH methylation and AGO6 contributes to 26% of CHH methylation. It is interesting to note that in leaf, inflorescence and silique tissues AGO6 dependent CHH methylation 79%, 90% and 84% respectively and its contribution to CHH methylation is also in between or higher than the contribution of both the RdDM pathways. This is because AGO6 is a common player between RDR6-RdDM and POL IV-RdDM pathways. Thus, my analysis of RdDM pathways through development shows that RDR6-RdDM plays higher role than POL IV-RdDM in later stages of development in more mature tissues while POL IV-RdDM plays higher role than RDR6-RdDM in earlier stages of development in younger tissues.

Next, I analyzed methylation levels of *Athila6 LTR TSS* in silique tissue of *ddm1/ago* double mutants for each of the five Argonaute proteins (AGOs) in order to determine which AGO protein plays the most prominent role in this tissue apart from AGO6 as shown above in Figure 3.8 D in RDR6 dependent establishment of *de novo* methylation. The mutant lines included five Arabidopsis AGOs such as AGO1, AGO4, AGO6, AGO7 and AGO9. AGO1 is involved in gene regulation through miRNA cleavage and translation inhibition (Meister 2013). AGO4 and AGO6 are involved in RdDM through heterochromatic siRNAs (Zilberman, Cao, Jacobsen 2003). AGO7 participates in ta-siRNA generation (Jouannet et al. 2012; Montgomery et al. 2008). AGO9 is known to play a role in TE silencing in female gametes (Olmedo-Monfil et al. 2010b). I included a strong and a weak allele for AGO1 and AGO9, because the strong allele for each of these lines generated very unhealthy plants. *ddm1/ago1* SA plants were very unhealthy and I couldn’t get any silique tissues off of these plants. My methylation
analysis shows that loss of function of AGO1, AGO4, AGO6 and AGO7 proteins in
_ddm1_ background results in decrease in methylation levels in all sequence contexts when
compared to _ddm1_ single mutants (Figure 3.8 E). The most significant decrease in
methylation levels is observed in _ddm1/ago6_ double mutant. Interestingly, _ddm1/ago9_ SA
and WA showed increase in CG and CHG methylation and no change in CHH
methylation when compared to _ddm1_. From this analysis of silique tissue of different
AGOs, I conclude that AGO6 is the main AGO in the targeting of _Athila6 LTR TSS_ for _de
novo_ methylation in the silique tissue, while AGO1, AGO4 and AGO7 also seem to play
a role in _Athila6 LTR TSS_ for _de novo_ methylation.

3.4.9 Effect of RDR6 in different gametes

After analyzing the role of RDR6-RdDM through different developmental time
points, I wanted to determine the effect of _ddm1_ and _rdr6_ mutants as either male or
gendered gametes. I performed several crosses and reciprocal crosses as mentioned in Table
3.3. First I looked at the effect of the progeny of these crosses on expression of _Athila_ and
then I performed methylation analysis to determine the effect of _ddm1_ and _rdr6_ gametes
on _Athila6 LTR TSS_ methylation (Figure 3.9). Expression analysis shows that
independent of the gender of the _ddm1_ mutant parent, _Athila6_ re-silencing occurs very
efficiently and quickly, as _Athila6_ expression is reduced 52-fold back to the levels
observed in wt Col (Figure 3.9 A). Next I crossed _rdr6_ as female parent to _ddm1_ as male
parent and reciprocal cross. I found that _ddm1 X rdr6_ has significantly increased
expression than _rdr6 X ddm1_ (p<0.01). Then I crossed Col as female parent to _ddm1/rdr6_
as male parent and vice versa. I found no change in the _Athila_ expression levels (Figure
3.9 A). Then, I switched around some of the re-silencing crosses from Figure 3.6 A to determine if there is a parent of origin effect. $ddm1/rdr6 \times rdr6$ shows statistically significant ($p<0.005$) $Athila$ expression than $rdr6 \times ddm1/rdr6$. The parent of origin effect of $ddm1/rdr6 \times rdr6$ cross is similar to $ddm1 \times rdr6$ cross. However, when I looked at the expression levels of $Athila$ between $ago6 \times ddm1/ago6$ and $ddm1/ago6 \times ago6$, I didn’t see much change between the two crosses. This indicates that unlike DDM1 and RDR6, AGO6 does not exhibit parent of origin effects. I next determined if parent of origin effects also cause changes to DNA methylation patterns depending on the direction of crosses. I performed bisulfite DNA sequencing on the progeny of the crosses in Table 3.3 and found that the expression levels in Figure 3.9 A correlate with the methylation levels of the $Athila6$ LTR TSS (Figure 3.9 B), but changing the direction of parents in the crosses does not cause difference in DNA methylation. The high level of CG and low level of CHH methylation found in wt Col is restored in the progeny of Col x $ddm1$ and its reciprocal cross (86% CG and 9% CHH) (Figure 3.9 B). However, when this cross is performed in an $rdr6$ background, either as $rdr6 \times ddm1/rdr6$ or as $ddm1/rdr6 \times rdr6$, CG and CHH methylation levels are only partially restored (57% CG and 27% CHH) compared to the $+/ddm1$ control cross. In Col X $ddm1/rdr6$, there is reduction in CG and CHG methylation levels and increase in CHH methylation when compared to $+/ddm1$ control cross. This methylation pattern correlates with increase in $Athila$ expression in the progeny of Col X $ddm1/rdr6$ cross (Figure 3.9 A). To conclude the above analysis shows that $ddm1$ as a female parent and $rdr6$ as a male parent play
more prominent role in \textit{Athila} resiling and change in the direction of \textit{ddm1} and \textit{rdr6} parents in the crosses does not affect the DNA methylation patterns.

I wanted to further test the role of \textit{RDR6} in \textit{Athila} silencing and I performed three different crosses to produce progeny with different mutant combinations of \textit{ddm1} and \textit{rdr6}. First I crossed \textit{ddm1het/rdr6homo} to \textit{ddm1het/RDR6 WT} and vice versa. I analyzed the effect of three different progeny namely \textit{ddm1het/rdr6het}, \textit{ddm1homo/rdr6het} and \textit{DDM1 WT/rdr6het} on \textit{Athila gag/pol} expression. Out of three different types of progeny, only \textit{ddm1homo/rdr6het} progeny from both the crosses shows similar \textit{Athila} expression which not as high as either \textit{ddm1} or \textit{ddm1/rdr6} (\textit{Figure 3.9 C}). \textit{DDM1 WT/rdr6het} progeny from \textit{ddm1het/RDR6 WT X ddm1het/rdr6homo} cross , shows very low levels of \textit{Athila} expression (\textit{Figure 3.9 C}). I then performed a cross between Col as female and \textit{ddm1het/rdr6homo} as male and analyzed the two progeny \textit{ddm1het/rdr6het} and \textit{DDM1 WT/rdr6het} for \textit{Athila} expression. I didn’t detect any \textit{Athila} expression in either of these progeny (\textit{Figure 3.9 C}). This analysis shows that out of different progeny from crosses between different mutant combinations of \textit{ddm1} and \textit{rdr6} only the one progeny which is homozygous for \textit{ddm1} and heterozygous for \textit{rdr6} shows \textit{Athila} expression in all the other progeny tested \textit{Athila} is efficiently silenced.

\textbf{3.4.10 Effect of de novo methylation on \textit{Athila} steady state transcription}

I have shown in \textit{Figure 3.1} and \textit{3.2} that decrease in CG methylation and increase in CHH methylation in \textit{ddm1} led to transcriptional activation of TEs. Both POL IV-RdDM and RDR6-RdDM pathway components were found to be needed for this hyper CHH methylation. Loss of \textit{de novo} CHH methylation in \textit{ddm1} double mutants with
components of POL IV-RdDM and RDR6-RdDM pathways led to increase in *Athila* expression when compared to *ddm1* single mutant. What is the effect of this *de novo* methylation on steady state transcription of TEs is not yet known. In order to determine this, I performed ChIP analysis using POL2 antibody in Col, *ddm1 F6*, *ddm1/rdr6*, *ddm1/pol4*, *ddm1/dcl3* and *ddm1/ago6*. RNA polymerase II is essential for transcription of DNA to synthesize mRNA. I analyzed POL2 occupancy in Actin and two *Athila6* protein coding regions *gag/pol* and *env* (Figure 3.10 A). I found POL2 occupancy to be more in *env* region than in *gag/pol* region. Expression analysis of *Athila6 env* region shows increased expression in *ddm1* than in *ddm1/rdr6* (Figure 3.10 B), while *Athila6* *gag/pol* expression is more in *ddm1/rdr6* than compared to *ddm1* (Figure 3.2 A). POL2 occupancy was found to be decreased in *ddm1/rdr6* than *ddm1* in *Athila6 env* region in confirmation with the expression data (Figure 3.10 B and C). I have found through methylation analysis that *Athila6 env* region does not undergo RDR6 dependent *de novo* methylation and hence this is not a good region of *Athila* to study the effect of *de novo* methylation on steady state transcription (Figure 3.11 C). However there is no change in POL2 occupancy between *ddm1* and *ddm1/rdr6* for *gag/pol* region and this does not correlate with the expression data for this region (Figure 3.2 A). Since the positive control for POL2 chip did not show equal levels of POL2 occupancy in all of the genotypes tested, it is hard to conclude anything from this experiment. More over this POL2 ChIP data is from only one bio replicate. I need to perform three more bio replicates and find a good positive control to be able to determine the effect of *de novo* methylation on *Athila* steady state transcription.
3.4.11 RDR6-RdDM seen in *Athila* LTR and IR region but not in **ENV**

I have shown that RDR6-RdDM plays a role in *de novo* methylation of *Athila*6 LTR region which is a promoter for *Athila*6 *gag/pol* protein coding region (Figure 3.2 C). Next, I wanted to determine if *Athila*6 IR region, which is promoter for *Athia*6 *env* protein coding region also exhibits RDR6 dependent *de novo* methylation similar to *Athila*6 LTR (Figure 3.11 A). We mapped the TSS of the endogenous *Athila*6 *env* gene and found it begins in the intergenic region between *gag/pol* and *env* (Figure 3.11 A). I assayed for methylation and small RNAs in this region. In wt Col, the native IR region has high CG methylation (80%) and low CHH methylation (14%) (Figure 3.11 C). When transcriptionally activated in *ddm1* mutants, this region has a reduction of CG methylation (to 35%) (Figure 3.11 C). The IR region undergoes CHH hypermethylation in *ddm1* mutants (from 13.7% in wt Col to 38% in *ddm1*). This CHH hypermethylation is mediated through RDR6-RdDM pathway, as *ddm1 rdr6* mutants display only 3% CHH methylation (Figure 3.11 C). Small RNA deep sequencing data in Figure 3.11 B shows that IR region produces RDR6 dependent 21-22nt sRNAs upon transcriptional activation. Thus, the RDR6-RdDM pathway is also acting to induce methylation of this IR promoter region (second region of *Athila*) through 21-22nt siRNAs (Figure 3.11 B and C). Next, I wanted to determine if RDR6-RdDM plays a role in *Athila*6 *env* methylation. When active, *Athila* makes abundant 21-22nt siRNAs from ENV coding region, which are dependent on RDR6 (Figure 3.11 B). Methylation analysis of *Athila* env shows that it does not gain CHH methylation in *ddm1* mutants and there is no significant difference between methylation levels of *ddm1* and *ddm1 rdr6* (Figure 3.11 C) (Andrea McCue
performed this methylation analysis). RDR6-RdDM is absent in *Athila ENV* even though it produces high levels of 21-22nt siRNAs. Why only some targets of RDR6-dependent siRNAs become methylated while others do not is not known. I hypothesize that difference in POLV occupancy might be the reason for RDR6 dependent de novo methylation in *Athila LTR* and *IR* but not in *ENV*, even though all the three regions produce RDR6 dependent methylation.

3.4.12 Pol V occupancy at transcriptionally active TEs

Pol V is present at silenced TEs including the *Athila6A* TSS, even though this region is not undergoing significant RdDM (*Figure 3.5 B*) (Johnson et al. 2014). Pol V occupancy has never been determined for transcriptionally active TEs. Therefore, I set out to determine if Pol V is also present at the Pol II transcriptionally active *Athila6 LTR* TSS. I crossed a FLAG-epitope tagged POL V protein into a *ddm1 pol V* TE-transcriptionally active background (Wierzbicki et al. 2012) and TE-transcriptionally silent background, *pol V* respectively, where FLAG-POL V complements the *pol V* mutation. I performed ChIP for FLAG-POL V both in plants with the FLAG-POL V transgene and, as a no-antigen negative control, in wt Col plants without the transgene. At the same time, I performed ChIP for H3Ac as a control (*Figure 3.12 B*). Along with *Athila6 LTR* (promoter region) and *Athila6 ENV* (coding region), I tested POL V occupancy in other TEs such as *Simple Hat2* (DNA transposon) (Zheng et al. 2007) and *At Copia66* or *Ta3* (Retrotransposon) (Voytas et al. 1992b). I found that FLAG-POL V is not enriched at the constitutively expressed Actin2 gene, while it is present at the *Athila6 LTR*, *Ta3* and *Simple Hat2*, both when the TE is transcriptionally silenced and
transcriptionally active (Figure 3.12 B). Increase in POL V Occupancy is associated with increase in acetylated histone3 marks (Figure 3.12 B). I could not detect any difference in POL V occupancy in Athila6 ENV region between transcriptionally silent and transcriptionally active backgrounds. The data in Figure 3.12 B is from only one bio replicate and therefore I repeated the ChIP for FLAG-POL V and H3Ac in plants with the FLAG-POL V transgene in pol V and ddm1 pol V plants and, in wt Col and ddm1 plants without the transgene (no antigen controls) for three more bio replicates to validate the above results. I find that POL V presence is enriched at both transcriptionally inactive and transcriptionally active Athila6 LTR when compared to respective no transgene controls wt Col and ddm1. In addition, I also detect FLAG-POL V enrichment at the Athila6A IR region when the TE is both transcriptionally silenced and transcriptionally active, and at the ancient env protein-coding region only when the TE is transcriptionally silenced (Figure 3.13 A).

Transcriptionally active state has more acetylated histone3 marks as seen in Actin, Athila6 LTR, IR and ENV in ddm1 and ddm1 pol V with FLAG-POL V transgene (Figure 3.13 B). Hence, even though Athila6 LTR, IR and ENV regions produce RDR6-dependent 21-22nt siRNAs (Figures 3.2 B and 3.11 B), RDR6 dependent de novo methylation occurs only in Athila LTR and IR regions (Figure 3.2 C and 3.11 C) but not in Athila ENV due to difference in POLV occupancy (Figure 3.12 A). Thus, as in Pol IV-RdDM, I suggest that RDR6-RdDM requires a Pol V scaffolding transcript at the target locus.
3.5 Discussion

3.5.1 RNAi or RDR6-RdDM is responsible for *de novo* methylation of active TEs through 21-22nt siRNAs

RNAi and RdDM are two mechanisms by which TEs get silenced at post transcriptional and transcriptional levels. It was always known in the literature that these two are independent pathways and act at different levels of TE regulation. RNAi was always known to act in the post transcriptional degradation of TE mRNAs (Chung et al. 2008; Sijen and Plasterk 2003). Through my research in this chapter, I have shown that surprisingly, the components of traditional RNAi pathway such as RDR6, DCL2 and DCL4 also play a role in *de novo* methylation of *Athila6 LTR TSS* when *Athila* is transcriptionally active (Figure 3.2). Since RDR6 plays a major role in this *de novo* methylation, this pathway is referred as RDR6-RdDM. It was always known in literature that components of RdDM or POL IV-RdDM pathway such as POL IV, RDR2, DCL3, AGO4, AGO6, POL V and DRM2 are responsible for establishment of silencing of TEs (Ito et al. 2011; Teixeira et al. 2009b). I have also shown that transcriptionally active *Athila* is silenced by POL IV-RdDM pathway (Figure 3.1). Loss of Components such as AGO6, POL V and DRM2 in transcriptionally active background led to huge decrease in *de novo* methylation when compared to all the other mutants of both RDR6 and POL IV RdDM pathways. Thus these three proteins are connecting bridge between both the pathways and hence establish a common point where RNAi transitions to RDR6-RdDM (Figure 3.14). I also observed that when key components of both the pathways are lost in a transcriptionally active background, there is nearly complete loss of *de novo*
methylation (Figure 3.2). Thus, both RDR6 and POL IV RdDM pathways together cause expression dependent *de novo* methylation in *Athila*.

It was not clear if this expression dependent *de novo* methylation in *Athila* was because of 21-22nt siRNAs or 24nt siRNAs because loss of RDR6 in *Athila* active background leads to increase in 21, 22 and 24nt siRNAs. This loss in 24nt siRNAs may be because RDR6-RdDM is playing a role in initiation of *Athila* methylation and so POL IV-RdDM which requires a methylated template to get started may not be that active in *ddm1/rdr6* background and therefore produces less 24nt siRNAs. Hence we analyzed *AtENSPM* which is a DNA transposon and has only RDR6 dependent increase in 21-22nt siRNAs in transcriptionally active back ground. Methylation analysis showed that *AtENSPM* also undergoes RDR6-RdDM but in an expression independent manner (Figure 3.4). This shows that RDR6-RdDM acts through 21-22nt siRNAs and not via 24nt siRNAs (Figure 3.14). The reason for expression independent role of RDR6-RdDM in *AtENSPM* in contrast to *Athila* may be because of the difference in the region analyzed for methylation, LTR TSS promoter region for *Athila* and protein coding region for *AtENSPM*. Also, the fact that *AtENSPM* is a DNA transposon and *Athila* is a LTR retro transposon might also be a reason for this difference. LTR retro transposons are characterized by 2 LTRs at 5’ and 3’ end respectively and it is quite possible that the 3’ LTR gets transcribed and converted to siRNAs which then target 5’ LTR for methylation. Nevertheless, further work is needed to investigate RDR6-RdDM in other TEs, both DNA and retro transposons, to conclude anything about expression dependent or independent role of this pathway. The role played by maintenance methyl transferases
when TEs are active is unknown. I found that hyper asymmetrical (CHH) methylation seen in \textit{ddm1} mutants is reduced in \textit{ddm1/met1} and \textit{ddm1/cmt2} double mutants when compared to \textit{ddm1} single mutant (Figure 3.5). This was very surprising because in \textit{ddm1/drm2} all the TE \textit{de novo} methylation was abolished indicating that RdDM pathway is responsible for this. May there is some unknown mechanism of interplay between maintenance pathway components and RdDM components.

3.5.2 Functional role of RDR6-RdDM and POL IV-RdDM pathways

Next, I explored the functional role of RDR6-RdDM and POL IV-RdDM pathways in TE maintenance and establishment of silencing. Previous work has shown that POL IV-RdDM plays a role in maintenance of \textit{AtSN1} and solo LTRs silencing (Herr et al. 2005; Huettel et al. 2006). However, I have shown that both RDR6 and POL IV RdDM pathways do not play a role in maintenance of TE silencing (Figure 3.5). Only MET1/CMT3/CMT2 plays a role in the maintenance of TE silencing. I also found that in \textit{wt Col}, where TEs are in silent state, all the CHH methylation is maintained by CMT2 and DRM2, a key component of POL IV-RdDM only acts when TEs are transcriptionally active (Figure 3.1, 3.5 and 3.14). Therefore, when there is loss of RDR6 in a transcriptionally silent background, both \textit{Athila} and \textit{AtENSPM} are silenced by maintenance methylation pathway and not through POL IV-RdDM (Figure 3.2 and 3.4).

Then, I looked at the role of RDR6 and POL IV RdDM pathways in corrective re-establishment of methylation and re-silencing of TEs. I found that both these pathways contribute to resilencing of \textit{Athila} and \textit{AtENSPM} but POL IV-RdDM plays a major role than RDR6-RdDM (Figure 3.6). Havecker \textit{et al.} have shown that levels of AGO6 protein
decreases in *pol IV* and *rdr2* mutants (Havecker et al. 2010a) and this may be causing
POL IV-RdDM’s major role in resilencing of TEs. Since both the RdDM pathways share
AGO6, when AGO6 is low in *pol IV* and *rdr2* mutants, even RDR6-RdDM can get
affected in directly. Also, through my analysis of role of RDR6 and POL IV RdDM in
development, I found that POL IV-RdDM plays more prominent role in earlier stages of
development and RDR6-RdDM plays more prominent role in the later stage of
development (Figure 3.8). AGO6 dependent *de novo* methylation is either in between the
*de novo* methylation contributed by both the pathways or similar to the pathway that
plays more prominent role in that particular tissue time. It is not known why different
tissue types exhibit differences in preference to a particular type of RdDM pathway.

### 3.5.3 Requirement of POL V for RDR6-RdDM activity

In contrast to *AtENSPM* and *Athila*, we saw that *Athila env* methylation region did
not undergo RDR6-RdDM in spite of large amounts of 21-22nt siRNAs produced from
this region (Figure 3.11). I also showed that *Athila6 IR* TSS, which is a promoter for
*Athila env* region, undergoes RDR6 dependent *de novo* methylation. It is not known why
different targets of RDR6 dependent small RNAs show difference in methylation. I found
that difference in POL V occupancy is the reason as POL V occupancy was only seen in
*Athila6 LTR* and *IR* promoter regions but not in *Athila env* protein coding region (Figure
3.12 and 3.13). In agreement with literature (Johnson et al. 2014), I also observed POL V
occupancy in transcriptionally silent TEs. It is yet to be determined how POL V is first of
all recruited to TEs with no methylation. There are two possible explanations for the
recruitment issue of POL V. First, may be POL V is able to get recruited without the need
for methylation. Secondly, as seen in fission yeast and animals (Zheng et al. 2009); (Castel and Martienssen 2013), another polymerase such as Pol II can act in place of POL V.

### 3.5.4 Future directions

RDR6-RdDM has been experimentally verified for only two TEs. We have shown in lab that through genome analysis there are 13 other TE targets both DNA and retro transposons which also produce RDR6 dependent 21-22nt siRNAs. It is not yet known whether there is a more preference for RDR6-RdDM on a particular type of transposon or not. This can be found out by performing genome wide methylation seq to determine RDR6-RdDM target TEs. Also, I found that RDR1, DCL1, NERD and SDE3 are other players in RDR6-RdDM ([Figure 3.3](#)). An exact mechanism of action of these proteins is not yet known. In *ddm1 dcl2 dcl4* triple mutant, even in the absence of 21-22nt siRNAs, I didn’t see complete loss of de novo methylation. This may be due the activity of DCL1 which is known to be involved in the biogenesis of 21nt siRNAs or DCL3 can also act through 24nt siRNAs and POL IV-RdDM. To determine how these proteins are acting in RDR6-RdDM pathway, I would perform protein-protein interaction studies, co immune precipitation assays and small RNA northern analysis. I found that resilencing of TEs occurs even later than seed stage of Arabidopsis ([Figure 3.7](#)). It is not yet known which stage of Arabidopsis life cycle TE resilencing occurs. This can be found out by look at early stages of embryogenesis by performing laser capture micro dissection to isolate the tissue types and perform expression analysis to analyze the time point in development where TEs are getting re silenced. It is also not known why POL IV-RdDM and RDR6-
RdDM differ in their preference for activity in a particular tissue type. Tissues such as Leaf, inflorescence and siliques were tested for CHH dependency on POL IV, RDR6 and AGO6. It would be interesting to explore the analysis in meristem tissue for POL IV as 41% and 26% CHH dependency is contributed by RDR6 and AGO6 (Figure 3.8). Also, it would be interesting to find the role of AGO6 in pollen CHH methylation as POL IV and RDR6 contribute 63% and 54% CHH methylation (Figure 3.8). Parent of origin effect seen when DDM1 is female and RDR6 is male (Figure 3.9), suggests that it would be interesting to study this RDR6-RdDM pathway in pollen, which is natural ddm1 mutant, that has global TE activation. It is not yet known the effect of this de novo methylation caused by RDR6-RdDM pathway on expression. In order to determine this, I would perform nuclear run on assay on the same mutants that I tested for POL2 ChIP in Figure 3.10 and determine the steady state transcript levels for various regions of Athila and AtENSPM. Another puzzling observation from my results was that though Athila6 IR TSS region undergoes RDR6-RdDM (Figure 3.11), in contrast to Athila6 LTR TSS where loss of RDR6-RdDM results in Athila6 gag/pol expression (Figure 3.2), there is no increase in Athila6 env protein coding region expression in ddm1/rdr6 when compared to ddm1 (Figure 3.10). Why RDR6 is playing different roles in regulating the expression of two different regions of Athila is not known.
Figure 3. 1 POL IV-RdDM is responsible for hyper CHH methylation.
A) qRT-PCR analysis of the steady-state polyadenylated gag/pol transcript levels of Athila6 demonstrate that the expression level increases in ddm1 double mutants with components of POL IV-RdDM that lose hypermethylation compared with ddm1 single mutant plants. B) Illumina deep sequencing showing the accumulation of siRNAs from the Athila6 LTR TSS region assayed for DNA methylation by bisulfite sequencing in E (Kaushik Panda performed sRNA deep sequencing analysis in all the mutants analyzed for Athila6 LTR throughout this chapter). In Col Athila is silenced and low levels of Athila6 derived sRNAs are produced. In ddm1, transcriptionally active Athila produces high levels of 21, 22, 24nt sRNAs. When DCL3 the major component of RdDM, is mutated in ddm1 mutant background there’s complete loss of 24nt sRNAs. C) Small RNA northern blot of the Athila6 LTR TSS region demonstrates that all 24 nt siRNA production is dependent on RDR2, DCL3 and Pol IV. All 21 and 22 nt siRNA production is dependent on RDR6. miR161 is used as a loading control. D) The reconstructed consensus Athila6A element. Athila6A contains two 1.8kb LTRs and two Open Reading Frame (ORF) gag and pol, a degenerate ORF (env). We mapped Transcriptional Start Site (TSS) in the 5’ LTR and used the surrounding 350 bp of this 5’ LTR TSS region for analysis of DNA methylation. E) In the wild type Arabidopsis thaliana ecotype Columbia (Col), TEs are silent and bisulfite analysis demonstrates that Athila6 LTR TSS has high symmetrical methylation (CG,CHG) and low asymmetrical methylation (CHH). In ddm1 (SWI/SNF chromatin remodeler) mutant, Athila6 LTR TSS becomes hyper CHH methylated (de novo methylation) when transcribed. Loss of DRM2, DCL3, POL IV, POL V, AGO4 and AGO6 in ddm1, leads to complete loss of de novo methylation

Continued
Figure 3.1: Continued
present in ddm1 single mutants. For all the bisulfite sequencing performed in this chapter C→T conversion efficiency is judged by sequencing a non-methylated genic exon and usually the conversion efficiency varies between 97%-100% and the number of sequenced individual LTR clones varies between 7-30. H= A, T or C. Error bars represent Wilson score interval 95% confidence limits.
Figure 3.1: Continued

A

B

Athila6 5'LTR TSS sRNAs

C

D

E

Athila6 TSS

miR161

LTR

gagpol

prot

rt

int

Bisulfite region

LTR

env

0 1 2 3 4 5 6 7 8 9 10 11 11.6 kb

1

2

3

4

5

6

7

8

9

10

11

11.6

152
Figure 3. 2 RDR6-RdDM plays a role in expression dependent de novo methylation of Athila.

A) Expression analysis of the Athila6 gag/pol coding region demonstrates that RDR6-dependent CHH hypermethylation repress further element expression when Athila6 is transcribed by Pol II. B) The accumulation of siRNAs from the region of the Athila6 LTR TSS region assayed for DNA methylation by bisulfite sequencing in C. RDR6-dependent 21, 22 and 24 nt siRNAs increase in ddm1. Nearly, small RNAs of all size classes are lost in ddm1 pol IV rdr6 triple mutant C) Bisulfite analysis demonstrates that the Athila6 TSS becomes CHH hypermethylated when transcribed (de novo methylation). This methylation is dependent upon components of RDR6 methylation pathway. Two biological replicates of ddm1 rdr6 were performed. A further decrease of overall methylation is apparent in pol V and ago6 mutants, a convergence point for both pathways, because they phenocopy ddm1 pol IV rdr6 mutants. D) Representative picture of triple mutant plant, ddm1/pol IV/rdr6 showing severe phenotypic effects whereas control plant +/-ddm1/pol IV/rdr6, where DDM1 is partially restored looks healthy.
Figure 3.2: Continued
Figure 3. Other players involved in RDR6-RdDM

A) Upon transcriptional activation and loss of either DCL1 or NERD seems to have very little effect on *Athila* expression when compared to *ddm1* single mutant. *ddm1/sde3amiRNA* has higher expression of *Athila* than *ddm1/dcl1* and *ddm1/nerd* double mutants but not as high as *ddm1*. B) Bisulfite analysis indicates that *de novo* methylation decreases in all of the double mutants when compared to *ddm1*, suggesting that DRM2, NERD, and SDE3 could all potentially play a role in RDR6-RdDM.
Figure 3. 4 RDR6-RdDM plays a role in expression independent de novo methylation of AtENSPM and functions through 21-22nt siRNAs.

A) The DNA TE subfamily, *AtENSPM6*, has RDR6-dependent 21-22nt sRNAs in the region used for further analysis. These RDR6-dependent sRNAs are only produced when the TE is active (*ddm1* mutant) (Kaushik Panda performed sRNA deep sequencing analysis). B) Bisulfite sequencing of AtENSPM6 in various mutants was used to independently assess the role of the RDR6-dependent sRNAs and Pol IV on DNA methylation (Andrea McCue performed bisulfite analysis of some of these mutants). I find that AtENSPM is CHH hypermethylated in *ddm1*, in an attempt at corrective reestablishment of DNA methylation. This CHH hypermethylation in *ddm1* is dependent upon both RDR6 and POLIV, as both PolIV- and RDR6- dependent DNA methylation can target the same transcriptionally active TE. Further loss of methylation occurs in *ddm1 pol V* mutants, like with *Athila6*. Pol V is a downstream component of both pathways. C) Though both Pol IV and RDR6 act to methylate the TE when it is active, this methylation does not affect its expression. D) The region assessed is over 2kb downstream of the transcriptional start site, possibly explaining the lack of expression correlation with methylation.
Figure 3.4: Continued
Figure 3. RDR6-RdDM and POL IV-RdDM pathways do not play a role in maintenance of TE silencing.

A) qRT-PCR analysis of expression of the gag/pol region from the Athila6 LTR retrotransposon. The expression of this element remains silenced in the RDR6-RdDM pathway mutants rdr6 and dcl2. In addition, mutations in the POL IV-RdDM pathway (rdr2, polIV and polV) also do not activate Athila6 expression. Only mutants that abolish CG methylation (ddm1 and met1) result in Athila6 gag/pol mRNA accumulation. Loss of CMT3 led to slight increase in Athila6 expression. B) Bisulfite sequencing of the Athila6 LTR around the gag/pol transcriptional start site (TSS). Methylation in the CG, CHG and CHH sequence contexts are shown. Mutations in the RDR6-RdDM and POL IV-RdDM pathways do not significantly affect the high CG, medium CHG and low CHH methylation levels (Dalen Fultz performed bisulfite analysis of some of these single mutants). The removal of CG methylation by ddm1 and expression of the gag/pol transcript is associated with increased levels of Athila6 LTR CHH methylation. Mutations in MET1, CMT3 and CMT2 affect corresponding methylation such as CG, CHG and CHH repectively.CMT2 is responsible for most of the CHH methylation in the TE-silenced wt Col epigenome.
Figure 3. 6 RDR6-RdDM functions in the corrective reestablishment of TE silencing.

A) qRT-PCR analysis of Athila6 gag/pol transcript accumulation in the progeny of the crosses detailed in Table 3.2. Active Athila6 TEs in a ddm1 background are efficiently resienced upon crossing to wild-type (wt) plants. Mutations in the Pol IV-RdDM and RDR6-RdDM pathways result in only partial resilencing of the Athila6 gag/pol transcript. Corrective reestablishment of silencing in the pol V and ago6 mutant backgrounds is impaired compared to the more efficient resilencing that occurs without Athila6 24nt siRNAs in the dcl3 mutant background. B) qRT-PCR analysis of F2 progeny produced from self of the F1 plants used in A. As the generation progresses from F1 to F2, POL IV-RdDM seems to be playing a huge role in re silencing Athila when compared to RDR6-RdDM pathway that seems to have even lesser role in re silencing. This evident from increased expression levels in the F2 progeny from crosses with (rdr2, pol IV and pol V) POL IV-RdDM pathway mutants and decrease in expression to nearly no expression in RDR6-RdDM pathway mutants (rdr6, dcl2 and dcl4). C) qRT-PCR analysis of ENSPM5. The expression levels in progeny from crosses of the RDR6-RdDM and Pol4-RdDM backgrounds are similar to those observed in Athila. D) Bisulfite analysis of DNA methylation indicates that the efficient re-silencing of Athila in the F1 +/-ddm1 is due to re-methylation, which is not fully reestablished when either the RDR6-RdDM or Pol IV-RdDM pathways are absent. E) Small RNA northern blot showing small RNAs from progeny of crosses performed in Table 3.2, 24nt siRNAs produced in +/-ddm1 are responsible for efficient re silencing of Athila. RDR6-RdDM mutant re

Continued
Figure 3.6: Continued

silencing crosses lose 21/22nt or 21&22nt siRNAs while POL IV-RdDM mutant re
silencing crosses lose 24nt siRNAs.
Figure 3.6: Continued
Figure 3. 7 Time point in development where re-silencing of Athila occurs.

A) qRT-PCR analysis of seedling tissue. The expression of Athila in the ddm1 het and ddm1 het/ rdr6 is not significantly similar. Resilencing of TEs is not occurring in this tissue type. B) Bisulfite analysis of DNA methylation indicates that the efficient re-silencing of Athila in the F1 +/-ddm1 seedlings is due to re-methylation, which is not fully reestablished when either the RDR6-RdDM pathway component RDR6 is absent. Hyper CHH methylation is not observed in ddm1 seedlings. C) qRT-PCR analysis of juvenile leaf tissue. The expression of Athila in the ddm1 het and ddm1 het/ rdr6 or pol IV mutants is not significantly similar. Resilencing of TEs is not occurring in this tissue type. D) qRT-PCR analysis of seed tissue. Bioreps could not be performed in this assay due to a shortage of tissue. The expression of Athila in the ddm1 het and ddm1 het/ homo mutants of interest is not significantly similar. Resilencing of TEs is not occurring in this tissue type.
Figure 3. 8 Role of RDR6-RdDM and POL IV-RdDM in various Arabidopsis developmental stages.

A) Plant pictures of various developmental stages of Arabidopsis such as Pollen, Seedling, Meristem, Leaf, Inflorescence and Silique tissues. B) Bisulfite analysis of RDR6-RdDM in various tissue types mentioned in A. Hyper CHH methylation in ddm1 mutants is only observed in inflorescence tissue but not in any other tissues. RDR6 plays more prominent role in later stages of development such as inflorescence and silique tissues. C) Bisulfite analysis of POL IV-RdDM in various tissue types mentioned in A. POL IV plays more prominent role in earlier stages of development such as leaf and pollen tissues. Both POL IV and RDR6 play equal role in seedling tissue. D) Bisulfite analysis of contribution of AGO6 in various tissue types mentioned in A. AGO6 dependent CHH methylation is in between or higher than the contribution of both the RdDM pathways. E) Bisulfite sequencing from various ddm1/ago mutants. Col exhibits high levels of CG (heritable) methylation, while ddm1 shows a decrease in CG methylation and an increase in de novo methylation (CHG and CHH). There is a significant decrease in all levels of Athila6 LTR TSS methylation, when AGO1, AGO4, AGO6 and AGO7 are not functioning in a ddm1 background. F) Percentage CHH methylation dependent on POL IV, RDR6 and AGO6. G) Bisulfite analysis of RDR6-RdDM and POL IV-RdDM in pollen tissue.

Continued
Figure 3.8: Continued
Figure 3. 9 Parent of origin effects of DDM1 and RDR6.

A) qRT-PCR analysis shows that *ddm1* X *rdr6* has significantly increased expression than *rdr6* X *ddm1* (p<0.01). Switching around re-silencing crosses shows that *ddm1/rdr6* X *rdr6* shows statistically significant (p<0.005) *Athila* expression than *rdr6* X *ddm1/rdr6*. DDM1 as a female parent and RDR6 as a male parent play more prominent role in *Athila* resilencing. B) Bisulfite analysis shows that high level of CG and low level of CHH methylation found in wt Col is restored in the progeny of Col x *ddm1* and its reciprocal cross (86% CG and 9% CHH). In *rdr6* X *ddm1/rdr6* or as *ddm1/rdr6* X *rdr6*, CG and CHH methylation levels are only partially restored (57% CG and 27% CHH) compared to the +/∗ddm1 control cross. Col X *ddm1/rdr6*, shows reduction in CG and CHG methylation levels and increase in CHH methylation when compared to +/∗ddm1 control cross. This methylation pattern correlates with increase in *Athila* expression in the progeny of Col X *ddm1/rdr6* cross. Change in the direction of DDM1 and RDR6 parents in the crosses does not affect the DNA methylation patterns. C) qRT-PCR analysis of progeny from *ddm1het/rdr6homo* X *ddm1het/RDR6 WT* cross and vice versa. Analysis of the effect of three different progeny *ddm1het/rdr6het*, *ddm1homo/rdr6het* and DDM1 WT/rdr6het on *Athila* gag/pol expression shows that only *ddm1homo/rdr6het* progeny from both the crosses shows similar level of *Athila* expression which not as high as either *ddm1* or *ddm1/rdr6*.

Continued
Figure 3.9: Continued
Figure 3. 10 Effect of de novo methylation on Athila transcription.

A) Athila6A contains two LTRs, two Open Reading Frame (ORF) gag and pol, a degenerate ORF (env) and IR region acts as promoter for env. B) qRT-PCR of analysis of Athila6 ENV shows increased expression in ddm1 than in ddm1/rdr6. C) ChIP analysis using POL2 antibody in Col, ddm1 F6, ddm1/rdr6, ddm1/pol4, ddm1/dcl3 and ddm1/ago6. POL2 occupancy was tested in Athila6 gag/pol, env and Actin. Since positive control Actin did not work, it is difficult to draw conclusion from this experiment. D) Re analysis of B, POL2 enrichment is calculated as fold enrichment over mock normalized to At1g08200 control gene. Only in ddm1/dcl3 double mutant, there is significant enrichment of POL2 over mock normalized to control gene for both Athila6 gag/pol and Athila6 env regions.
Figure 3. 11 IR region of Athila undergoes RDR6-RdDM.

A) Athila6A contains two LTRs, two Open Reading Frame (ORF) gag and pol, a degenerate ORF (env) and IR region acts as promoter for env. B) Small RNA deep sequencing analysis shows RDR6 dependent 21-22nt siRNAs in Athila6 IR TSS and Athila6 env. C) Bisulfite analysis shows that IR region undergoes CHH hypermethylation in ddm1 mutants (from 13.7% in wt Col to 38% in ddm1). This CHH hypermethylation is mediated through RDR6-RdDM pathway, as ddm1 rdr6 mutants display only 3% CHH methylation. Methylation analysis of Athila env shows that it does not gain CHH methylation in ddm1 mutants and there is no significant difference between methylation levels of ddm1 and ddm1 rdr6. ns= not statistically significant. (Andrea McCue performed Athila env methylation analysis)
Figure 3. 12 POL V occupancy is seen in transcriptionally active TEs.

A) *Athila6A* contains two LTRs, two Open Reading Frame (ORF) *gag* and *pol*, a degenerate ORF (*env*) and IR region acts as promoter for *env*. This figure shows the hypothesis that the reason for why only some targets of RDR6 derived siRNAs play a role in methylation is because of difference in POL V occupancy.

B) ChIP qPCR analysis of the FLAG-tagged POL V protein in a TE-silent epigenome and the *ddm1* TE transcriptionally-active epigenome from one bio replicate. POL V occupancy is present in transcriptionally active *Ta3*, *Simple Hat2* and *Athila6 LTR*. H3Ac ChIP-qPCR with the same chromatin at the same
time as the FLAG-POL V ChIP performed as an experimental verification of ChIP.
Figure 3. 13 POL V occupancy in Athila6 LTR and IR but not in ENV.

A) ChIP qPCR analysis of the FLAG-tagged POL V protein in a TE-silent epigenome and the ddm1 TE transcriptionally-active epigenome from three bio replicates. POL V occupancy is present in transcriptionally active Athila6 LTR and IR promoter regions undergoing RDR6-RdDM but not in Athila6 ENV which does not undergo RDR6-RdDM. ChIP signals are normalized to the wt Col no-antigen control. B) As POL V-ChIP controls, I performed ChIP of H3Ac on the chromatin of each biorep at the same time as the FLAG-IP. ChIP-qPCR experiments of H3Ac show that the enrichment of a constitutively expressed gene, At1g08200, exhibits no difference from wt Col to ddm1 mutant plants. However, as previously reported the enrichment of H3Ac at the Athila6A TSS is significantly increased (6.4-fold) in ddm1 mutant plants compared to wt Col, reflecting the increased transcription of Athila6A observed in the ddm1 epigenome (Gendrel et al. 2002). This IP was performed from the same chromatin at the same time as the FLAG-IPs from A as an experimental verification of ChIP.
Figure 3.13: Continued

A

B
Figure 3. 14 Model of TE epigenetic silencing.

Expression dependent RDR6-RdDM requires 21-22nt siRNAs bound to AGO6 and POLV scaffolding transcript. AGO6, POL V and DRM2 are common players between RDR6-RdDM and POL IV-RdDM. (This figure is modified from Panda et al. 2013 Plant and Signal Behavior (Panda and Slotkin 2013))
Table 3. 1 PCR primers and mutant alleles used

<table>
<thead>
<tr>
<th>Target</th>
<th>Experiment</th>
<th>Figure</th>
<th>Forward primer sequence</th>
<th>Reverse primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ath56d FSS</td>
<td>Baseline Sequencing</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Mq29019 exon</td>
<td>Baseline Sequencing control</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d LTR TSS region</td>
<td>Northern Probe</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>m1851</td>
<td>Northern Probe</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d GAPDH3</td>
<td>qRT-PCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d region 7’</td>
<td>qRT-PCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>A1glqR0360</td>
<td>qRT-PCR, qPCR control</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>SimpleHat2</td>
<td>qRT-PCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d FSS</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Aa3</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Actin2</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d IR</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d ccw</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron8</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron9</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron10</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron11</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron12</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron13</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron14</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron15</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron16</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron17</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron18</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron19</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron20</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron21</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron22</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron23</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron24</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron25</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron26</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron27</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron28</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron29</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron30</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron31</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron32</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron33</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron34</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron35</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron36</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron37</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron38</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron39</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron40</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron41</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron42</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron43</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron44</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron45</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron46</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron47</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron48</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron49</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
<tr>
<td>Ath6d Intron50</td>
<td>qPCR</td>
<td>Figure 3.18.2</td>
<td>AGGAAAGAAGAAGAAGA</td>
<td>TCAACTCAACACAGATGTCACAA</td>
</tr>
</tbody>
</table>

Key for degenerate bases: M = A, C, G, T; W = A, or T; S = G, C; Y = C, T; K = G, T
<table>
<thead>
<tr>
<th>Female Parent</th>
<th>Male Parent</th>
<th>Progeny Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt Col</td>
<td>ddm1 F3</td>
<td>+/-ddm1</td>
</tr>
<tr>
<td>ddm1 F3</td>
<td>wt Col</td>
<td>ddm1/+</td>
</tr>
<tr>
<td>rdr6</td>
<td>ddm1 rdr6</td>
<td>+/-ddm1 rdr6</td>
</tr>
<tr>
<td>dcl2</td>
<td>ddm1 dcl2</td>
<td>+/-ddm1 dcl2</td>
</tr>
<tr>
<td>dcl4</td>
<td>ddm1 dcl4</td>
<td>+/-ddm1 dcl4</td>
</tr>
<tr>
<td>rdr2</td>
<td>ddm1 rdr2</td>
<td>+/-ddm1 rdr2</td>
</tr>
<tr>
<td>dcl3</td>
<td>ddm1 dcl3</td>
<td>+/-ddm1 dcl3</td>
</tr>
<tr>
<td>pol IV</td>
<td>ddm1 pol IV</td>
<td>+/-ddm1 pol IV</td>
</tr>
<tr>
<td>ago6</td>
<td>ddm1 ago6</td>
<td>+/-ddm1 ago6</td>
</tr>
<tr>
<td>pol V</td>
<td>ddm1 pol V</td>
<td>+/-ddm1 pol V</td>
</tr>
</tbody>
</table>

Table 3. 2 Crosses performed to test to role of RDR6-RdDM and POLIV-RdDM components in Athila re-silencing
<table>
<thead>
<tr>
<th>Female Parent</th>
<th>Male Parent</th>
<th>Progeny Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt Col</td>
<td>ddm1</td>
<td>+/-ddm1</td>
</tr>
<tr>
<td>ddm1</td>
<td>wt Col</td>
<td>ddm1/+</td>
</tr>
<tr>
<td>rdr6</td>
<td>ddm1</td>
<td>+/-rdr6 +/-ddm1</td>
</tr>
<tr>
<td>ddm1</td>
<td>rdr6</td>
<td>ddm1/+ rdr6/+</td>
</tr>
<tr>
<td>wt Col</td>
<td>ddm1/rdr6</td>
<td>+/-ddm1/rdr6</td>
</tr>
<tr>
<td>ddm1/rdr6</td>
<td>wt Col</td>
<td>ddm1/+ rdr6/+</td>
</tr>
<tr>
<td>rdr6</td>
<td>ddm1/rdr6</td>
<td>+/-ddm1 rdr6</td>
</tr>
<tr>
<td>ddm1/rdr6</td>
<td>rdr6</td>
<td>ddm1/+ rdr6</td>
</tr>
<tr>
<td>ago6</td>
<td>ddm1/ago6</td>
<td>+/-ddm1 ago6</td>
</tr>
<tr>
<td>ddm1/ago6</td>
<td>ago6</td>
<td>ddm1/+ ago6</td>
</tr>
<tr>
<td>ddm1/+ rdr6</td>
<td>ddm1/+ RDR6</td>
<td>ddm1/+ rdr6/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ddm1homo rdr6/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDM1 rdr6/+</td>
</tr>
<tr>
<td>ddm1/+ RDR6</td>
<td>ddm1/+ rdr6</td>
<td>ddm1/+ rdr6/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ddm1homo rdr6/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDM1 rdr6/+</td>
</tr>
<tr>
<td>wt Col</td>
<td>ddm1/+ rdr6</td>
<td>ddm1/+ rdr6/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DDM1 rdr6/+</td>
</tr>
</tbody>
</table>

Table 3. 3 Crosses performed to test the effect of RDR6 in different gametes.
Chapter-4 Perspectives
4.1 Different levels of *Athila* retrotransposon repression by *Arabidopsis thaliana*

*Athila* retrotransposons constitute more than 3% of *Arabidopsis* genome and are completely silenced in wild type Columbia. I have shown through my thesis work that *Athila* is under tight epigenetic regulation and *Arabidopsis* uses different defense mechanisms to keep *Athila* under control. In wt Col, *Athila* is transcriptionally inactive due to dense CG and CHG symmetrical methylation and low levels of CHH asymmetric methylation. This transcriptional repression of *Athila* in Col occurs through DDM1 and maintenance methylation pathway. I have shown in Chapter-3 that when *Athila* is active, its transcriptional regulation occurs mainly through RdDM pathway and post transcriptional regulation occurs via RNAi. In Chapter-2, I have shown that active *Athila* uses siRNA854 produced from RNAi mechanism to target non TE neighboring gene, *UBP1b*, a component of plant stress granules in *trans*. Thus, *Athila* targeting host gene *UBP1b* for translational repression is similar to miRNAs and tasiRNAs regulating genes (Chapman and Carrington 2007; Vaucheret 2006). Hence I have shown that gene regulation not only occurs via miRNAs and tasiRNAs but also by TE siRNAs. I found that similar to mammalian *UBP1b* homolog, TIA-1 which sequesters LINE-1 retrotransposons (Aparicio et al. 2010; Goodier et al. 2007a; Kedersha et al. 2000), *UBP1b* causes translational repression of *Athila* Gag protein, thus forming as a third negative regulatory layer against *Athila* activity (Chapter-2).

RNAi, the second host defense response, produces *Athila* derived 21-22nt siRNAs when *Athila* is transcriptionally active. In Chapter-3, I have shown that these *Athila* derived 21-22nt siRNAs and RDR6, a key component of post transcriptional RNAi
pathway plays a role in de novo methylation of active TEs. This new pathway is referred as RDR6-RdDM. I also found that both RDR6 and POL IV RdDM (the traditional RdDM) pathways function to repress active TEs at transcriptional level and also play a role in corrective reestablishment of TE methylation. Through my research, I have shown that Arabidopsis employs three defense mechanisms against Athila at transcriptional, post-transcriptional and post transcriptional levels through RDR6 and POL IV RdDM, RNAi, UBP1b and stress granule formation respectively (Figure 4.1). Athila manipulates host defense at post-transcriptional level (RNAi) to fight with host defense at translational level (formation of stress granules). Thus transposable elements are in constant battle with their host organisms and vice versa. Though there has been much progress in the field of transposable element regulation in the past few years, there are a lot more questions that are unanswered and remain to be explored.

4.2 Regulation of TEs by plant stress granules

My research on the effect of UBP1b, a component of plant stress granules on Athila retrotransposon showed that UBP1b causes translational repression of Athila GAG protein when key components of Athila regulation at transcriptional (ddm1) and post transcriptional level (rdr6) are lost. Till date, UBP1b regulation of Athila is the first example of a plant stress granule component regulating a TE. We were not able to detect any retrotransposition of Athila in the triple mutant with loss of all three levels of Athila regulation (ddm1/rdr6/ubp1b). I have also shown that when Athila is transcriptionally active, there is increase in accumulation of stress granules (chapter-2). These results raise the question whether similar to sequestration of LINE-1 retro transposon by mammalian
stress granule component (TIA-1) (Goodier et al. 2007b); do plant stress granules through UBP1b also sequester Athila and other TE transcripts and thereby cause translational repression? Also, it is currently unknown why and how TE mRNAs are specifically sequestered by stress granules during various conditions of stress. It is not known if there exists any other mechanism of TE regulation which acts on different stages of TE life cycle apart from transcription, post transcription and translation.

In order to address the above mentioned questions, the key experiment would be creation of ddm1 and ddm1/rdr6 plants (loss of transcriptional and post transcriptional regulation) with N-terminal FLAG tagged UBP1b transgene and performing Immuno Precipitation (IP) of mRNPs (mRNA bound RiboNucleoProteins) using the technique described by Sorenson et al (Sorenson and Bailey-Serres 2014). Analysis of transcripts from this mRNP-IP would determine if plant stress granules sequester Athila and other TE mRNA transcripts and would also answer why specifically TE mRNAs are recruited into SGs during stress due to epigenetic reactivation of TEs. To determine other mechanisms of TE regulation that affect TE life cycle, first I would analyze the next step after translation into GAG protein, which is formation of virus like particles (VLPs), in the panel of single, double and triple mutants defective in transcriptional, post transcriptional and translational regulation of Athila. Based on the technique described by Bohmdorfer et al (Bohmdorfer et al. 2008), I would perform VLP isolation and detection using 50 mesh gold grids coated with anti-gag anti serum. This experiment would determine if Athila is capable of forming VLPs in the triple mutant. If Athila VLPs are formed and can be visualized under electron microscope, the next step is to analyze
Athila cDNA intermediates by isolating GAG particles to perform immuno precipitations with GAG antibody in order to isolate Athila RNA or cDNA intermediates. For thorough knowledge on the regulation of TEs by stress granules and vice versa, in both plants and animals, a lot more work needs to be done as this is relatively new area of research.

4.3 Transcriptional regulation of TEs by RDR6-RdDM pathway

Though through my work, I have determined a novel role for major component of RNAi, RDR6 in de novo methylation, there is lot more to be known about the function, other players of this new RdDM pathway, RDR6-RdDM in different tissue types of Arabidopsis thaliana. I have shown that proteins such as NERD, SDE3 and RDR1 play a role in TE de novo methylation (chapter-3). Mode of action of these proteins in RDR6-RdDM is not clearly understood. To determine how these proteins are acting in RDR6-RdDM pathway, I would perform protein-protein interaction studies and co immune precipitation assays of these proteins with other well known components of RDR6-RdDM pathway in TE epigenetically active background. I would also find out the effect of NERD and SDE3 in small RNA biogenesis in TE active background by small RNA northern blots.

I found that resilencing of TEs in Arabidopsis occurs even later than seed stage of development. The developmental time point of Arabidopsis life cycle where TE resilencing occurs is still unknown. I would perform laser capture micro dissection to isolate the tissue types at early stages of embryogenesis and perform expression analysis to analyze the time point in development where TEs are getting re silenced. Parent of origin effect occurs when ddm1 is female and rdr6 is male, suggests that it would be
interesting to study this RDR6-RdDM pathway in pollen. Since pollen is natural ddm1 mutant and has global TE activation.

4.4 Role of Arabidopsis TE small RNAs in different TE silencing pathways

I’ve shown through my thesis work that when transcriptionally active Athila produces 21-22nt small RNAs through RNAi pathway (RDR6, DCL4, DCL2 and AGO1). These Athila 21-22nt small RNAs not only play a role in post transcriptional degradation of Athila mRNAs (PTGS) but also in gene regulation (UBP1b, a component of stress granules) through translational inhibition and Athila DNA de novo methylation (RDR6-RdDM). Thus there seems to be interplay between different TE silencing pathways in Arabidopsis. It is not yet known if there also exists similar kind of interplay between TE defense mechanisms in animals. Interplay between several other small RNA and TE silencing pathways in Arabidopsis remains to be studied.

What role maintenance methylation pathway plays in TE active background is not known. Through my work in chapter-3, I have shown that CHH hyper methylation in ddm1 mutants is reduced in ddm1/met1 and ddm1/cmt2 double mutants when compared to ddm1 single mutant. I have also shown that RdDM pathway is responsible for all of the TE de novo methylation when TEs are active. There may be some unknown mechanism of interaction between RdDM and maintenance methylation pathway. It would be interesting to test the role of CMT3 in methylating active TEs by creating ddm1/cmt3 double mutant and looking for Athila promoter methylation. I also found that DCL1, a component of plant miRNA pathway plays a role in TE de novo methylation and hence
RDR6-RdDM. Thus there is interplay of miRNA and RDR6-RdDM and this may be due to the involvement of DCL1 in 21nt siRNAs biogenesis.

4.5 Effect of TE epigenetic regulation on host gene regulation

In chapter-2, I have shown that loss of all three layers of Athila regulation leads to sterility defects in the triple mutant (ddm1/rdr6/ubp1b). It is not known what causes this sterility in these mutant plants since no retrotransposition of Athila was detected. Sterility defects may be caused by improper centromere function or even due to loss of demethylase activity. In order to test for centromere defects, I would perform quantitative RT-PCR, in the panel of single, double and triple mutants defective in RdDM, RNAi and stress granule formation for Athila2 and CEN180. Increase in expression of CEN180 indicated defective centromere. Since Athila2 is promoter for CEN180 expression, loss of epigenetic activity of Athila leads to centromere dysfunction and hence sterility defects in plants. Sterility might also be caused due to lack of demethylase activity, therefore in order to determine if this is true, I would perform western in the above mutant panel and probe the western with antibody for DME.

Another important question that remains to be further studied is how the TE regulation mechanisms do not affect the host gene regulation. It is not known how the proteins responsible for maintaining the gene body methylation in plants would act when TEs are transcriptionally active. It would be interesting to create ddm1 double mutants with proteins such as IDM1 and ROS1 and test the DNA demethylation of Athila retro transposon promoter region as well as some active genes that undergo CG methylation to see the effect of transcriptional activation of TEs on activity of proteins responsible for
gene body methylation. The test can be performed using Thymine DNA glycosylase activity kit.

Transposable elements epigenetic regulation in plants and animals is widely researched and the more it is studied the more complex it gets making way to lot more questions. In recent years there has been tremendous progress in TE epigenetic regulation studies and through my research work I was able to answer few of the important questions in this vast amazing TE epigenetics.

Figure 4. 1 Regulation of *Athila* at various stages of its life cycle.
Appendix A: Data from other small projects

In Slotkin lab, I worked on two main projects (Chapter-2 and 3) and I also did two small projects. Data from these projects was not sufficient to write into entire chapter. Therefore, I’ve described the results from both the projects in this addendum.

**ddml hybrids project**

It is known in literature that hybrids cause activation of TEs (Josefsson, Dilkes, Comai 2006; Rakhshandehroo et al. 2009). It is also known that global activation of TEs occurs in ddm1 mutants (DDM1, SWI/SNF chromatin remodeler) (Steimer et al. 2000)(Wright and Voytas 1998). I wanted to find out if TE activity increases if I cross ddm1 mutants in different ecotype backgrounds to generate hybrids. In order to determine this, I crossed ddm1 mutant in Columbia (Col) background to ddm1 mutant in Landsberg (Ler) background and also the reciprocal cross. Then I performed qRT-PCR on the resulting hybrid seedlings to check the effect of these hybrids on Athila6 gag/pol and env coding regions. I found that when I crossed ddm1 in Col to ddm1 in Ler, I saw statistically significant increase in Athila6 env expression when compared to parents, ddm1 in Col or ddm1 in Ler (Figure A.1), while the reciprocal cross did not show any effect on Athila6 env expression. However, I did not find any difference in Athila6 gag/pol expression between parents and hybrids. Since I performed qRT-PCR with only two biological replicates, it is hard to draw any solid conclusion from this experiment.
But, the effect of hybrids on *Athila6 env* expression is hinting that *ddm1* ecotype hybrids have increased TE activation. It would be interesting to repeat this experiment with more biological replicates in and more TE candidates. Changing the type of tissue analyzed will also be useful, because we have seen that *ddm1* inflorescences have more *Athila* expression than *ddm1* seedlings.

**Analysis of *Athila* transcripts project**

I have shown in Chapter-2, Figure 2.3 that upon transcriptional activation of *Athila6* in *ddm1* mutants, there is huge accumulation of *Athila6* 3’ region transcripts. These transcripts do not accumulate to high levels in *ddm1/rdr6*, indicating that the amplification of *Athila6* 3’ is dependent on RDR6. (Chapter-3, Figure 3.10). Also, in Chapter-2, Figure 2.12 I have shown that additional *Athila* transcripts are produced when RDR6 is lost in transcriptionally active background. It is not known whether these RDR6 generated transcripts are polyadenylated, or do they have an internal stretch of As, as *Athila* retrotransposon has internal stretch of As between 7164 to 7201 positions. Therefore, I wanted to analyze these transcripts further by Reverse Transcriptase PCR (RT-PCR). I hypothesize that if these RDR6- generated transcripts are not polyadenylated, then only one strand of cDNA will accumulate to high levels as the other strand will have Ts.

First, I prepared cDNA from both total RNA and poly A RNA using gene specific primers such as *Athila6* 3’ of CDS, *Athila6* 3’ of 845, *Athila6* 5’ of intron, *Athila6* 3’ of intron and *Athila6* Intron (Table A.1). I used both forward (F) and reverse (R) gene specific primers to detect AntiSense (AS) and Sense (S) transcripts of *Athila* respectively.
I also prepared cDNA from total RNA using oligo dT. Then, I performed various RT-PCR reactions with different sets of primers as described in the Table A.1 below, spanning inside and outside of intron, around the 3’ coding region and 3’ of miRNA 845 binding regions (Figure A.2 and A.3). F and R Primers for both the Athila6 3’ CDS and Athila6 5’ and 3’ of intron regions detects both spliced and un-spliced Athila transcripts, while F and R Primers for Athila6 3’845 detects only un-spliced Athila transcripts. Also, the primer combinations Athila6 3’ CDS F and 3’ 845 R as well as Athila6 5’ of intron F and intron R detects only un-spliced Athila transcripts. All the RT-PCR analysis for various regions of Athila was done in transcriptionally silent background Col, transcriptionally active ddm1 and ddm1/rdr6 (Figure A.3). I used Col DNA as a control to test for any genomic DNA contamination in the cDNA and also the amplification of Athila transcript in Col would indicate that the transcript is unspliced.

From this overall RT-PCR analysis, I found that there are two polyAAAAA Pol II transcripts are produced from Athila6. First one is the sense Athila6 transcript from the 3’ region which gets immediately and efficiently spliced. Using Athila6 5’ of Intron F and 3’ of intron R primers, I detected this 500bp sense Athila6 transcript (Figure A.3). Second one is full length antisense and un-spliced Athila6 transcript approximately 2.5 Kb of this transcript is detected using Athila6 5’ of Intron F and 3’ of intron R primers and is not dependent on RDR6 (Figure A.3). RDR6 plays a role on both transcripts. RDR6 copies the spliced sense polyA Athila6 transcript into another antisense polyAAA transcript (Figure A.3). This antisense transcript itself is spliced. I was able to detect 300bp Athila6 transcript using Athila6 5’ of Intron F and 3’ of intron R primers. Hence spliced Athila6
transcript gets copied and does not get degraded. The unspliced full length antisense transcript can continue to copy back to sense, which is dependent on RDR6 (Figure A.3). The RDR6 copy of this unspliced transcript produces dsRNA of the unspliced element. This dsRNA is very quickly and efficiently turned over into siRNAs. Therefore the unspliced Athila6 transcript gets copied and degraded.

**Materials and Methods**

**Plant material**

The mutant alleles used in this study are in the Columbia (Col) background. For ddm1 hybrids experiment, I used ddm1 in either Col background or ddm1 in Landsberg (Ler) background. Plants were grown under standard long day conditions at 22 °C. Inflorescence tissue was used in each experiment unless otherwise noted.

**Quantitative Reverse Transcriptase–PCR (qRT-PCR)**

Total RNA was extracted using TRIzol reagent (Invitrogen) or the RNeasy Plant Kit (Qiagen). Total RNA was DNAsel treated and reverse transcribed using an oligo-dT primer and Tetro Reverse Transcriptase (Bioline). qRT-PCR was performed using 3 technical replicates each of 2 biological replicates. qRT-PCR primers are shown in Table below. qRT-PCR reactions were annealed at 60°C. The relative expression values for all experiments were calculated based on the expression of the experimentally validated control gene At1g08200. Quantitative real-time PCR was performed using SensiMix (Bioline) on a Mastercycler ep realplex thermocycler (Eppendorf). Relative expression was calculated using the ‘delta-delta method’ formula $2^{\Delta\Delta CP_{\text{sample}}}$, where 2
represents perfect PCR efficiency. Statistical significance was calculated using unpaired T-tests.

**Reverse transcriptase PCR (RT-PCR)**

Total RNA was extracted from seedlings using mini hot phenol extraction as described by Box et al (Box et al. 2011). TRIzol reagent (Invitrogen). polyA RNA was extracted from total RNA samples using Oligotex mRNA mini kit (Qiagen, Valentia, CA). RT-PCR was performed DNAseI treated total RNA or polyA RNA and reverse transcribed using either oligo-dT primed cDNA or Gene specific primed cDNA and SuperScript III Reverse Transcriptase (Invitrogen). PCR was performed for 28 cycles using primers listed in Table below.

![Graph A](image1.png)

**Figure A. 1 Analysis of ddm1 (Col) X ddm1 (Ler) cross and reciprocal crosses.**
Figure A. 2 *Athila* cartoon

*Athila* has two LTRs, two protein coding regions, *gag/pol* and *env* and Intron in the 3’ region of *Athila*. Arrows around the env and intron region show the primers used for the analysis of RDR6 generated transcripts in Figure A.3.
Figure A. 3 *Athila* transcripts produced by RDR6.
<table>
<thead>
<tr>
<th>Target</th>
<th>Experiment</th>
<th>Figure</th>
<th>Forward primer sequence</th>
<th>Reverse primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athabasca 1 pgp1</td>
<td>qRT-PCR</td>
<td>Figure 4.1</td>
<td>CTAATCCGCCAGATCCCAAG</td>
<td>AGACGGAATAGCATCACATTGC</td>
</tr>
<tr>
<td>Athabasca F region</td>
<td>qRT-PCR</td>
<td>Figure 4.1</td>
<td>ACAGGAAGTGCGGAGACACC</td>
<td>CTUCAACGGAGGCAAGGACTTCT</td>
</tr>
<tr>
<td>N/a</td>
<td>qRT-PCR, qPCR control</td>
<td>Figure 4.1</td>
<td>AGCCCAAGGAAGACATGAGTGA</td>
<td>GATGGTGACACCTCAGTAGG</td>
</tr>
<tr>
<td>Athabasca F CDS</td>
<td>RT-PCR</td>
<td>Figure 4.3</td>
<td>TCACTTTCGCTCTGTGAGGTT</td>
<td>TGAGGAAACGTGAAAGCTTGGAA</td>
</tr>
<tr>
<td>Athabasca F c1455</td>
<td>RT-PCR</td>
<td>Figure 4.3</td>
<td>TCAAGGATACGATACGGAAGG</td>
<td>TCTGGAAGCTGAAATCAAGGAT</td>
</tr>
<tr>
<td>Athabasca F 5cl25</td>
<td>RT-PCR</td>
<td>Figure 4.3</td>
<td>GCAACTCAGAGGAGCTTTGAC</td>
<td>TCCTGACTGCTCGCCTTCCCTC</td>
</tr>
<tr>
<td>Athabasca F VpVp</td>
<td>RT-PCR</td>
<td>Figure 4.3</td>
<td>GCACTCAGAGGAGCTTTGAC</td>
<td>TCCTGACTGCTCGCCTTCCCTC</td>
</tr>
<tr>
<td>TpA/T</td>
<td>RT-PCR, control</td>
<td>Figure 4.3</td>
<td>GCTCTGCTCCTCCCCTCACTAG</td>
<td>CCACCCGCGGAATCATTCCAG</td>
</tr>
</tbody>
</table>

**Table A.1** Primer sequences used in this report
References


epigenomic mapping defines four main chromatin states in arabidopsis. EMBO J 30(10):1928-38.


Sanmiguel P and Bennetzen JL. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Annals of Botany 82:37-44.


