Characterizing the *De Novo* Methylation and Silencing of Exogenous Transposable Elements

Honors Research Thesis

Presented in Partial Fulfillment of the Requirements for Graduation "with Honors Research Distinction" in the undergraduate colleges of The Ohio State University

by
Eric Roose

The Ohio State University
April 2016

Project Advisor: Professor R. Keith Slotkin, Department of Molecular Genetics

Table of Contents

Background

Transposable elements (TEs) are DNA sequences that can move from one location in the genome to another through a process called transposition. These elements are present in all known eukaryotic organisms. Indeed, the human genome itself is comprised of approximately 48% TE sequences [16]. When a TE is active, high levels of duplication and transposition can be seen. This activity can be quite mutagenic, as TEs can cause chromosomal breaks and insert into essential genes, significantly disrupting the genetic sequence. It is well documented that this mutagenic activity extends to TE activity in humans [1]. Furthermore, their activity is implicated in many human diseases, such as breast and colon cancer, Walker-Warburg syndrome, and neurofibromatosis [1]. In summary, one could think of TEs as endogenous, or native, human mutagens that have become a stable, but potentially dangerous, part of the human genome.

Two predominant classes of TEs have been characterized. Class 1 TEs, or retrotransposons, transpose through a copy-and-paste mechanism. This process begins when the TE facilitates transcription of an RNA intermediate of itself. The RNA intermediate then mobilizes to a different location in the genome, where it is copied into DNA through the action of *Reverse Transcriptase*, which helps to give these elements their title, and inserted. Class 1 TEs can be further subdivided into LTR (long terminal repeat) retrotransposons and non-LTR retrotransposons. LTR retrotransposons are similar in structure and life cycle to retroviruses, such as HIV [15]. Indeed the prevailing wisdom is that LTR retrotransposons are retroviruses that lost the ability to exit the host cell, causing them to depend on vertical transmission instead of horizontal infectious transmission. Class 2 TEs, in contrast, do not use an RNA intermediate or Reverse transcriptase. They tend to move through a cut and paste mechanism, although there are exceptions, where the element is excised from its original location and reinserted elsewhere [15].

Since TEs are powerful mutagens in eukaryotic genomes, 1 billion years ago in the common ancestor of plants, animals and fungi, mechanisms began to evolve that identify TEs and reduce TE activity. The reduction, or silencing, of TEs in eukaryotic organisms is accomplished primarily through the action of three ancient pathways, one transcriptional and one post transcriptional. However, in order to understand these important pathways, it is first necessary to understand the concept of DNA methylation. This process involves the addition of a methyl group (CH_3) to a cytosine nucleotide in one of three different contexts (CG, CHG, and CHH), where H represents any base pair except guanine. DNA methylation is a very effective marker of gene silencing, as it prevents the creation of mRNA from the methylated DNA [11]. Thus, DNA methylation acts as a transcriptional silencer.

The first pathway, informally called maintenance methylation, occurs through the action of two different classes of methyltransferases. The first class is known as *de novo* methyltransferases, as they create *de novo* DNA methylation. The protein Domains Rearranged Methylase 2 (DRM2) is the methyltransferase that creates *de novo* DNA methylation in *Arabidopsis thaliana* and is targeted by RdDM (described below) [12]. New DNA (*de novo)* methylation occurs in all contexts (CH, CHG, CHH). The second class is known as maintenance methyltransferases, as they copy methylation patterns on parent strands of DNA to the daughter strands during DNA synthesis. The proteins responsible for this maintenance methylation are DNA methyltransferase 1 (MET1), which produces mainly CG methylation, chromomethylase 3 (CMT3), responsible for CHG methylation, and chromomethylase 2 (CMT2), which is responsible for CHH methylation [11, 12]. Both MET1 and CMT3 are much more active than CMT2, which is why CHH methylation levels are lower in a maintenance context and are associated with *de novo* DNA methylation. Once established, TE silencing is maintained from cell-to-cell and from generation-to-generation by mechanisms that propagate DNA methylation upon cell division, a process of regulation termed *epigenetic inheritance* of TE repression.

If and when the TE is active, the second major pathway, called the RNA interference (RNAi) pathway (Figure 1), acts post transcriptionally by degrading TE mRNAs and thus preventing their translation into functional proteins. This pathway begins with the production of a single stranded RNA molecule (ssRNA) by RNA polymerase 2 (POL II), the conversion of this molecule into a double stranded RNA molecule by RNA-dependent RNA Polymerase 6 (RDR6), and the cleavage of this molecule into 21 or 22 nucleotide small interfering RNA (siRNA) by Dicer-Like 4 protein (DCL4) or Dicer-Like 2 protein (DCL2), respectively. These siRNA's are incorporated into Argonaute 1 (AGO1) and are used to bind to and target TE RNA for degradation [13]. Thus, this pathway is a post-transcriptional backup if the TE is transcribed.

The third pathway, called the *RNA-directed DNA Methylation* (RdDM) pathway (Figure 2), works to establish and maintain DNA methylation of the TEs. It functions as a transcriptional gene silencing pathway as DNA methylation prevents TEs from being transcribed into mRNA, which prevents later translation and production of proteins needed for transposition. The RdDM pathway begins with the transcription of an active TE into a ssRNA by RNA polymerase 4(IV) and the subsequent production of dsRNA by RNA dependent RNA polymerase 2 (RDR2). This dsRNA is then cleaved into 24 nucleotide siRNA by dicer-like 3 (DCL3) and incorporated into Argonaute 4, 6, or 9 (AGO 4/6/9) which uses its siRNA to bind a scaffold transcript produced by another plant RNA polymerase, POL 5(V). This complex recruits cellular machinery that methylates and modifies histones and DNA of the target area [12].

Recently, the RNAi and RdDM pathway have been found to be interconnected. This process is thought to occur through the 21 or 22 nucleotide siRNA's produced by the RNAi pathway. These 21 or 22 nucleotide siRNA's can directly incorporate into AGO6 and direct AGO6 to TE chromatin, aiding the RdDM pathway (Figure 3) [14]. This means that the RNAi pathway, and thus post transcriptional TE silencing, is linked to the RdDM pathway, and thus DNA methylation (Figure 3) [14].

Biological Question

Although a great deal is known about how TE methylation and silencing is maintained, via the above pathways, it is not well understood how TE's arrived at this heterochromatic silenced state [4]. The current belief is that the RdDM pathway is involved mainly in homology-dependent initiation of silencing. The RdDM pathway begins at repressive histone methylation evident at already silenced TEs. Plant genomes contain tens of thousands of silenced TEs. Furthermore, these silenced TEs are known to produce high levels of 24 nucleotide siRNAs through the RdDM pathway [2]. These 24 nucleotide siRNA's can act as a *homology sensor*, identifying TEs with similar sequences, and facilitate *trans*-silencing of active TE copies through the binding of similar POL 5 transcripts and furthering of the RdDM pathway.

However, this pathway is not sufficient to silence new non-homologous TEs as there will be no pre-existing methylation to initiate the production of matching 24 nucleotide siRNAs and none of the existing 24 nucleotide siRNAs will be able to recognize the new element. In these cases, it is dogmatically believed that the RNAi pathway must act in a homology-independent manner to initiate silencing. **Therefore, the over-arching broad biological question of my work has been to determine if and how a non-homologous TE, newly inserted into the genome, is silenced.** If the RNAi pathway is targeted to the new TE, 21/22 nucleotide siRNAs will be produced and will facilitate the degradation of similar TE mRNA. Additionally, these siRNA's may feed into the RdDM pathway, helping to initially methylate the TE to allow for the RdDM pathway to kick in. The issue in the field is that it is not well understood how the RNAi pathway is initiated in response to a new TE mRNA and not genic mRNA.

Significance

These findings could have significant future implications. Transposable element activity is implicated in diseases such as Walker-Warburg Syndrome and breast cancer [1]. Furthermore, TE activity is the causal agent behind human diseases such as hemophilia A and B, severe combined

immunodeficiency, and Duchenne muscular dystrophy [9]. An increase in our knowledge of how *de novo* DNA methylation and TE silencing occurs in *Arabidopsis* is particularly applicable to the treatment of these diseases, as the mechanisms of TE silencing in *Arabidopsis* and mammals is highly conserved. Thus, an increase in our understanding of how *de novo* DNA methylation occurs could allow for the prevention or therapeutic treatment of many of these devastating, costly diseases

Aims of Project

This thesis has two primary aims:

1. Investigate the potential role of several gene products in the establishment of non-homology based *de novo* DNA methylation, through the use of the exogenous LTR TE construct Tto and mutant Arabidopsis lines.

In order to study the process of non-homology based *de novo* DNA methylation it is important to have a system in which non-homology based *de novo* DNA methylation is the only initially active pathway towards TE silencing. It is necessary to ensure that no homology between the inserted element and previously silenced Arabidopsis is present, as that will allow for canonical homology-based RdDM silencing and methylation. Furthermore, it must be shown that DNA methylation and silencing of the inserted TE is occurring, as that highlights the presence of *de novo* DNA methylation. In this vein, it has been demonstrated that when the LTR retrotransposons, originally from tobacco, *Tto1* and Tnt1 exogenous TEs are inserted into the Arabidopsis genome these elements have become methylated and silenced [8, 17]. Furthermore, it has been evidenced that these elements are initially actively transposing, helping to show there is no homology dependent pathway at work as these pathways would allow for rapid silencing of an introduced TE [4]. In addition, the Slotkin lab did a number of assays to test for

homology, and found none between Tto and Arabidopsis. This allows for the use of the Tto TE construct as an effective tool to study the process of non-homology based *de novo* DNA methylation. In this thesis, I inserted this TE construct (as a transgene) into several mutant Arabidopsis lines in an effort to identify gene products necessary to achieve non-homology based *de novo* DNA methylation and subsequent silencing of the exogenous TE. Consequently, in this thesis I identified:

a) The methylation levels achieved by the Tto LTR TE construct in wild-type and mutant contexts.

b) The transcription levels of the Tto LTR TE construct present in wild type and mutant contexts.

2. Determine the impact of the insertion site of the Tto LTR TE construct on the methylation levels achieved

In the experimental design utilized in this thesis the transformation of the exogenous LTR construct into the genome is completely random and uncontrollable. As such, it is expected that within each mutant line, each plant should show different insertion sites of the construct. This difference could cause discrepancies in DNA methylation levels achieved between plants, especially between plants with constructs inserted into genic and intergenic regions. It has long been known that silencing and methylation of intergenic and genic regions differs [5]. As such, it may well be that insertion into an intergenic region or non-coding region makes it more difficult for Arabidopsis's gene silencing pathways to act on the exogenous element. The goal of the approach in Aim 1 is to have 10 or so randomly inserted TEs in each wild-type or mutant line to compare. Consequently in this thesis I attempted to:

a) Troubleshoot a procedure for determining the constructs insertion site (inverse PCR)

b) Compare achieved methylation levels with the location of the construct insertion site

Core Techniques

1) **Bisulfite Sequencing:** This technique is necessary to obtain DNA methylation levels. First, DNA must be isolated from all tissues of interest and then subsequently treated with Ribonuclease (RNase A) to degrade any present RNA still present in the DNA extraction. Next, this purified DNA is treated with bisulfite conversion reagent to convert all un-methylated cytosines into uracil, while leaving methylated cytosine residues unchanged. This bisulfite DNA is then amplified using target specific primers in a process called bisulfite PCR. These PCR products are then purified, cloned into E-coli, and plasmid DNA corresponding to the target of interest is obtained, confirmed and sent for sequencing. This data is then compared to a reference (not bisulfite modified) sequence using the software program Kismeth [6]. The program works by comparing the number of converted and unconverted cytosine's against the reference sequence. The output is the varying types of methylation found on the target sample, in terms of CG, CHG, and CHH methylation levels.

2) **Quantitative Reverse-Transcriptase Polymerase Chain Reaction (qRT-PCR):** This technique allows for the quantification of the amount of cDNA, or complement DNA, in a sample. First, RNA must be extracted from the samples of interest and subsequently DNased to remove any present DNA. This added DNAse is then inactivated. The enzyme reverse transcriptase is then utilized to convert the remaining RNA into cDNA using an oligo-dT primer to only make cDNA from the mRNA with a polyA tail. Primers specific to the target sequence are then added, along with SYBER Green fluorescent dye, and Taq polymerase. The cDNA is then denatured and the primers anneal to the single stranded cDNA. The

SYBER green molecules only fluoresce in the presence of double-stranded DNA, and the amount of fluorescence is linear to the amount of double-stranded DNA PCR product. As such, the amount of fluorescence can be used as a measure of the relative amount of cDNA. This fluorescence is measured by the q-PCR machine each PCR cycle and used to subsequently determine the amount of cDNA originally present in the samples and thus the amount of transcription of the target sequence into mRNA (expression level).

3) **Inverse PCR:** This protocol is largely adapted from that presented by Kim et al. [10]. It is used to ultimately, in addition with the BLAST program (detailed below), determine the site of transgene insertions. The first step is to obtain genic DNA and utilize RNase to degrade any present RNA. This is followed by a restriction digest using appropriate enzymes that cut the genome such that the intact, known insertion sequence is recovered along with some flanking region. The flanking region could be of varying size depending on the location of the next restriction site in the flanking DNA, but the important point is that a fragment of some length containing the insertion sequence is obtained. However, many other fragments are recovered in addition to the fragment of interest, as the enzyme will undoubtedly act on similar sequences throughout the genome. The recovered DNA fragments are circularized through the use of T4 DNA ligase after dilution, as this increases the odds of self-ligation and not ligation between fragments. These circularized DNA fragments are then used as PCR templates with primers specific to both ends of the known insertion sequence. The PCR primers do not face each other, but face out, to amplify across the ligation point. This primary PCR is followed by a nested PCR that utilizes shorter primers containing some, but not all, of the original sequence as a way to increase the PCR specificity for the insert and its flanking region. Next, this nested PCR is run out on a gel, to confirm the presence of DNA and purity of the sample, gel extracted, and purified. Lastly, this sequence is sent in for

sequencing and the returned sequence is run through BLAST, which determines what gene the flanking region matches and thus the location in the genome that it inserted into.

4) **Basic Local Alignment Search Tool (BLAST):** This is a GUI program that compares an input of genome nucleotides with a large stockpile of known genes, eventually telling the user what genic regions of an organism the sequence matches. In our case, we are only comparing our sequences to the Arabidopsis genome (TAIR10 annotation). It offers the ability to help match any genic sequence to an organism and location in that organism's genome. As such, it was of paramount importance in this thesis, specifically aim 2.

Characterization of the Establishment of non-homology based *De Novo* DNA Methylation

Materials and Methods

The focus of this thesis is an exogenous, or new, TE retrotransposon (Class 1 LTR) construct native to *Nicotiana tabacum* (tobacco) that was inserted into multiple *Arabidopsis thaliana* mutant lines. Arabidopsis is a powerful model organism for the study of TE's as the mechanisms of TE silencing are highly conserved in plants and mammals [7]. This construct is a truncated version of the whole element, called Tto, and was essentially reduced to a single LTR prior to being obtained by the Slotkin Lab [8].

The construct described above was utilized to allow for the efficient study of the role of several gene products in the establishment of non-homology based *de novo* DNA methylation in the model organism *Arabidopsis thaliana*. This construct, along with an endogenous gene promoter (pAXS2), was obtained and subsequently inserted into several mutant *Arabidopsis thaliana* lines (described below) or wild type Arabidopsis, respectively, through the use of transformative bacteria, Agrobacterium. The endogenous promotor was used as a control to insure that the insertion of genic material itself would not cause DNA methylation to occur. The Agrobacterium-infected transformed plants (T0) were then

selfed and their progeny planted out (T1). A selective spray, Basta, was used to select for only plants that had taken up the transgene construct, as these plants would have also obtained a gene encoding Basta resistance. Consult Figure 4 for a simplified overview of this process.

Next, T1 plants with the germinal-inserted transgene had their inflorescence (flower bud) tissue and leaf tissue collected from each mutant line in a pooled manner (3-12 T1 plants per pool). This allowed a comparison between methylation levels achieved by the construct at different developmental levels, as a way to see if developmental stage impacted the DNA methylation achieved in each mutant. DNA was then extracted using Damon's all Natural Genomic Prep (DANGP). Bisulfite sequencing was then used to determine methylation levels for both the inflorescence and leaf pooled mutant tissue in the Tto LTR construct background (Figure 5 and 6 respectively). Additionally, bisulfite sequencing was used to determine the methylation levels achieved by the endogenous promoter construct in the wild-type background (Figure 5). This data was then analyzed for statistical significance and conclusions.

In addition, inflorescence tissue was separately obtained from four individual wild type (Columbia) *Arabidopsis thaliana* plants. This tissue was treated the same as the pooled mutant tissue, described above. The purpose of isolating this tissue was to see how methylation levels of the TE LTR construct varied among individuals and whether pooled tissue was truly representative of the individual methylation levels found in a given mutant background. Wild type Arabidopsis was used as it was known that it would attain some level of methylation, based on background information. This would ensure we had levels of methylation to compare for this control experiment. The data for this experiment is displayed in Figure 7.

Lastly, RNA was obtained from *pol4, rdr6,* and *dcl2/3/4* mutants pooled inflorescence tissue and used to perform qRT-PCR (Figure 8). This experiment provided data on the level of transcript abundance of TE LTR construct, allowing us to compare how methylation levels and silencing of the TE construct in a *de novo* setting compared. It was quite possible, with the enigmatic nature of *de novo* TE silencing, that

there could be a discrepancy between methylation levels achieved and the actual effective silencing of the TE LTR construct.

Mutant Lines

The mutant Arabidopsis lines used in this experiment are *dcl1*, *ago1*, *rdr6, ddm1*, *suvh2/9*, *sde3* and *rrp6/1*. Mutants were selected, on an overarching level, for their ability to impact the homology-independent RNAi pathway, as this is where *de novo* DNA methylation initiation is thought to occur. Mutations in gene products necessary for the RNAi pathway to function should limit the production of 21/21 nucleotide siRNA and thus limit their ability to facilitate DNA methylation through the association with Ago6. Mutants deficient in the RNAi pathway include *rdr6*, *ago1*, *dcl1,* and *sde3*. Dicer-like protein 1 (DCL1) has a role in the processing of miRNA and the incorporation of these into the RNA-induced silencing complex (RISC) containing AGO1 [3]. The gene RRP6/1 is implicated in RNA degradation pathways, and was thus of interest for possible impacts on the RNAi pathway. The gene DDM1 is implicated in chromatin condensation. When this gene is mutated, the cell can no longer silence TEs. Lastly, SUVH2/9 potentially aids POL5 recruitment to DNA as its products bind to methylated DNA.

Data and Analysis

As expected from background research, the Tto LTR construct achieved significant levels of methylation in all three contexts in a wild type background (Figures 5 and 6). Furthermore, the expression levels of the Tto LTR construct was much lower than would be seen in a plant with an active construct. However, it is worth noting that the expression levels of the Tto LTR construct were around ten percent relative expression (Figure 8), which is not insignificant. This could be due to the nature of the element, as just a LTR, or perhaps the nature of recently acquired *de novo* DNA methylation.

It is also worth noting that the levels of methylation achieved in this context varied considerably between individual wild type plants (Figure 7). This variance in methylation could be due to simple variance in conditions (light level, moisture etc), or biological variance. However, it is not unlikely that this difference could be a reflection of differential TE insertion sites into the Arabidopsis genome and subsequent changes in the level of *de novo* methylation achieved (see Aim 2). When the individual wild type plants methylation data for the exogenous Tto LTR construct was averaged it was essentially the same as the pooled value. This, along with the range of individual methylation values, from low to medium, helps to show that the experimental design utilizing pooled tissue was representative of the population of mutant plants with the exogenous Tto LTR construct.

The endogenous promotor construct did not achieve any significant methylation in the wild type background (Figure 7). Indeed the levels of methylation did not rise above five percent in any context, or even all three contexts combined. This helps show that it was not the act of insertion of some genic construct that facilitated methylation but rather the nature of the **TE LTR construct** and *de novo* DNA methylation pathway.

The levels of methylation achieved in the rrp6L1, suvh2/9 and ddm1 mutant backgrounds were all statistically comparable to the wild type levels, showing that these products do not have a significant role in *de novo* DNA methylation. Furthermore, the Tto LTR construct methylation levels achieved in the leaf and inflorescence tissue was not statistically different, showing that developmental stage of the plant has no significant impact on methylation level (Figures 5 and 6). However, three of the mutants (*ago1, rdr6, and dcl1*) showed a statistically significant increase in methylation of the exogenous Tto LTR construct. Though some contexts were not statistically different from the methylation levels achieved in wild type plants, the CHH levels of methylation in these plants were. This is particularly applicable for this thesis as CHH levels of methylation are primarily associated with *de novo* DNA methylation, as the level of maintenance CHH methylation (via CMT2) is very low. When the level of

transcription/expression of the Tto LTR construct in the *rdr6* background was measured, high levels of Tto LTR construct expression were observed (Figure 8).

Conclusions

It appears that the act of inserting a genic sequence into the genome is not sufficient to cause DNA methylation and that the developmental stage of the plant has little impact on the methylation levels the Tto LTR construct achieves. This demonstrates that our assay is correctly determining *de novo* methylation at TEs, not just methylation upon transformation. According to the bisulfite sequencing data most of the mutants surveyed (*ddm1, suvh2/9, rrp6/*1) have no real impact on *de novo* DNA methylation. The wild-type gene products corresponding to these genes are involved in TE silencing, Pol 5 recruitment, and RNA degradation, respectively. This shows that these well-studied pathways are not involved in non-homology based *de novo* DNA methylation. Interestingly *rdr6*, *ago1*, and *dcl1* mutants show unexpected high levels of methylation when compared to the levels achieved in wild type plants, which is considered the benchmark for "silenced." This increased level of methylation is unexpected as one would expect mutations in components known to facilitate the RNAi pathway would hamper the production of 21/22 nucleotide siRNAs. By extension, it would seem that DNA methylation should decrease in these mutants without a functional RNAi pathway to feed 21/22 nucleotide siRNAs into the RdDM methylation pathway, via incorporation into AGO6. However, the exact opposite is observed; the RNAi mutants listed above have significantly higher methylation levels than the wild.

However, higher methylation levels do not necessarily correspond to effectively silenced Tto LTR constructs. According to the qRT-PCR data, it appears that although *rdr6, ago1,* and *dcl1* can achieve high levels of methylation, this is not sufficient to silence the construct. This lack of silencing is, in opposition to the methylation data, what one would expect for an RNAi mutant. One possible explanation is that when one of the above RNAi proteins is nonfunctional, the RNAi silencing of the TE is

pulled away from RNAi and pushed towards the RdDM. This suggests a new level of interconnectedness or competitive sharing of gene silencing pathway components, such as 21-22 and 24 nucleotide small RNAs. Effectively, **it seems that RNAi is competing with the RdDM pathway**. Such an undogmatic connection between these pathways could form the basis of how *de novo* DNA methylation of TEs occurs. Further study of this connection could shed a great deal of light on the process of *de novo* DNA methylation.

Characterization of the Impact of Tto LTR Genic Insertion Site on DNA Methylation

Materials and Methods

In this experiment I focused on wild type *Arabidopsis thaliana* plants with the exogenous Tto LTR construct. This was because I already had individual methylation data for this line, as part of my validation of the pooled inflorescence tissue technique (Figure 7). The only step left to complete this experimental aim was to determine where in the genome the exogenous construct had inserted for the four individual plants and to compare this to the methylation levels achieved. Towards this end, I first obtained genomic DNA from each individual plant's inflorescence tissue. I then followed the inverse PCR protocol detailed above and subsequently used BLAST to determine the site of insertion (Table 1). However, it is worth noting here that it took a significant amount of work to troubleshoot this protocol and fine tune it into something that experimentally worked in the lab. As such I consider the production of a working protocol as one of the major gains of this thesis. The protocol is now operational and can serve as a guideline for all future inverse PCR endeavors in the Slotkin Lab (Supplement 1).

Data and Analysis

In three of the four individuals I tested the inverse PCR failed (individual three) or came back as vector sequence (individual 1 and 4). This vector sequence traces back to the Agrobacterium insertion

of the transgene into the parental *Arabidopsis thaliana*. This simply means that the primers and other minor parameters must be rearranged to obtain the insertion site of the TE, as we know that the plants have the transgene or they could not survive the selective Basta spray. However, in the case of the second individual I successfully obtained the insertion site of the exogenous Tto LTR construct (Table 1). The exogenous construct has inserted into the *Arabidopsis thaliana* gene ATG31005, internal to the coding sequence, on which very little research has been done. When this information is paired with the methylation levels achieved in this individual (Figure 7), one can see that this context has significant DNA methylation in the CHH, or *de novo*, context.

Conclusions

The Tto LTR construct that inserted into the *Arabidopsis thaliana* gene ATG31005 can now be correlated with a high level of CHH methylation, the highest seen in the four individual plants tested. This preliminary data seems to suggest that perhaps the genic insertion site of the Tto construct, and thus TEs in general, has a good deal to do with the methylation levels achieved. However, until methylation data corresponding to Tto elements that inserted into intergenic or non-gene regions can be obtained, this association will remain a simple correlation. Nonetheless, this is an interesting piece of data that may explain the variation in individual plants Tto methylation levels. Furthermore, this difference in insertion site could have more impact on the methylation of newly inserted TEs, compared to maintenance methylation, as perhaps it impacts the ability of the RNAi pathway to selectively degrade TE mRNA into 21/22 nucleotide siRNAs. As such more research into this correlation and hypothesized connection will need to be performed to facilitate a better understanding of future *de novo* DNA methylation studies.

Final Conclusions

When I began my research at the Slotkin Lab, precious little was known about the initiation of non-homology based *de novo* DNA methylation. We had recently discovered that the RNAi pathway can feed into the RdDM pathway and thus DNA methylation, through AGO6 [14], but did not adequately understand how homology independent (RNAi) gene silencing of TEs was initiated. **My research has identified a previously unknown competitive connection between the RNAi pathway and the RdDM pathway, and thus DNA methylation itself.** This knowledge will provide a solid groundwork for future experiments into the nature and mechanism of this competitiveness, which can only increase our knowledge of *de novo* DNA methylation. Furthermore, **my work has suggested that the insertion site of TEs may be particularly important for the ability of *de novo* DNA methylation to act upon them**. Future experiments based off this initial correlation may allow for a greater understanding of how homology independent pathways target TE mRNA over genic mRNA.

Acknowledgements

I would like to thank the Slotkin Lab and all its members for their tutelage, expertise, and time. In particular, I would like to thank the graduate students Dalen Fultz, who served as a mentor during this project, and Kaushik Panda and Saivageethi Nuthikattu, who were both instrumental in my development as a competent experimental researcher. I would also like to thank my funding sources, including the Ohio State University Office of Undergraduate Education for its funding of my Undergraduate Research Fellowship, the Arts and Sciences Honors Committee for its funding of my Undergraduate Research Scholarship, and the Slotkin Lab NSF funding MCB-1252370. Furthermore, I would like to thank the Ohio State Plant-Microbe Genomics Facility for its help in sequencing DNA methylation. Lastly I would like to specifically thank my PI Keith Slotkin, for his expertise and help in the formation and revision of this thesis and the grants that made it possible.

Works Cited

1. Belancio V, Deininger P, Roy-Engel A. 2009 LINE dancing in the human genome: transposable elements and disease. *Genome Med.* **1**, 97.

2. Cho SH, Addo-Quaye C, Coruh C, Arif MA, Ma Z, Frank W, Axtell MJ. 2008 *Physcomitrella patens* DCL3 is required for 22–24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. *PLoS Genet*. **4**, 12.

3. Eamons AL, Smith NA, Curtin SJ, Wang MB, Waterhouse PM. 2009 The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA*. **15**, 2219-2235.

4. Fultz D, Choudury SG, Slotkin RK. 2015 Silencing of active transposable elements in plants. *Curr. Opin. Plant Biol*. **27**, 67-76.

5. Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES. 2012 Trans chromosomal methylation in *Arabidopsis* hybrids. *Proc. Natl. Sci. U.S.A*. **109**, 3570-3575.

6. Gruntman, Eyal, Yijun QI, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R. 2008 Kismeth: Analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics*. **9**, 371.

7. He XJ, Chen T, Zhu JK. 2011 Regulation and function of DNA methylation in plants and animals. *Cell Res.* **21**, 442-465.

8. Hirochika H, Okamoto H, Kakutani T. 2000 Silencing of retrotransposons in Arabidopsis and

reactivation by the *ddm1* mutation. *The Plant Cell*. **12**, 357-368.

9. Kapitonov VV, Pavlicek A, Jurka J. 2006 Anthology of human repetitive DNA. Encyclopedia of Molecular Cell Biology and Molecular Medicine.

10. Kim SR, Jeon JS, An G. 2011 Development of an efficient inverse PCR method for isolating gene tags from T-DNA insertional mutants in rice. *Methods in Mol. Biol.* **678**, 139-146.

11. Lisch D. 2009 Epigenetic regulation of transposable elements in plants. *Rev. Plant Biol*. **60**, 43-66.

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

13. McCue AD, Nuthikattu S, Reeder SH, Slotkin RK. 2012 Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLOS. Genet*. **8**, 1.

14. McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK. 2014 ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO. J*. **34**, 20-35.

15. Munoz-Lopez M and Garcia-Perez JL. 2010 DNA transposons: Nature and applications in genomics. *Curr. Genome*. **11**, 115-128.

16. Pace J, Feschotte C. 2007 The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res.* **17**, 422-432.

17. Perez-Hormaeche J, Frederique P, Beauclair L, Masson IL, Courtial B, Bouche N, Lucas H. 2008 Invasion of the Arabidopsis genome by the tobacco retrotransposon Tnt1 is controlled by reversible transcriptional gene silencing. *Plant Physiol*. **147**, 1264-1278.

Figures for Characterizing *De Novo* DNA Methylation

By
Eric Roose
Ohio State University
April 2016
Project Advisor: Dr. R. Keith Slotkin, Department of Molecular Genetics

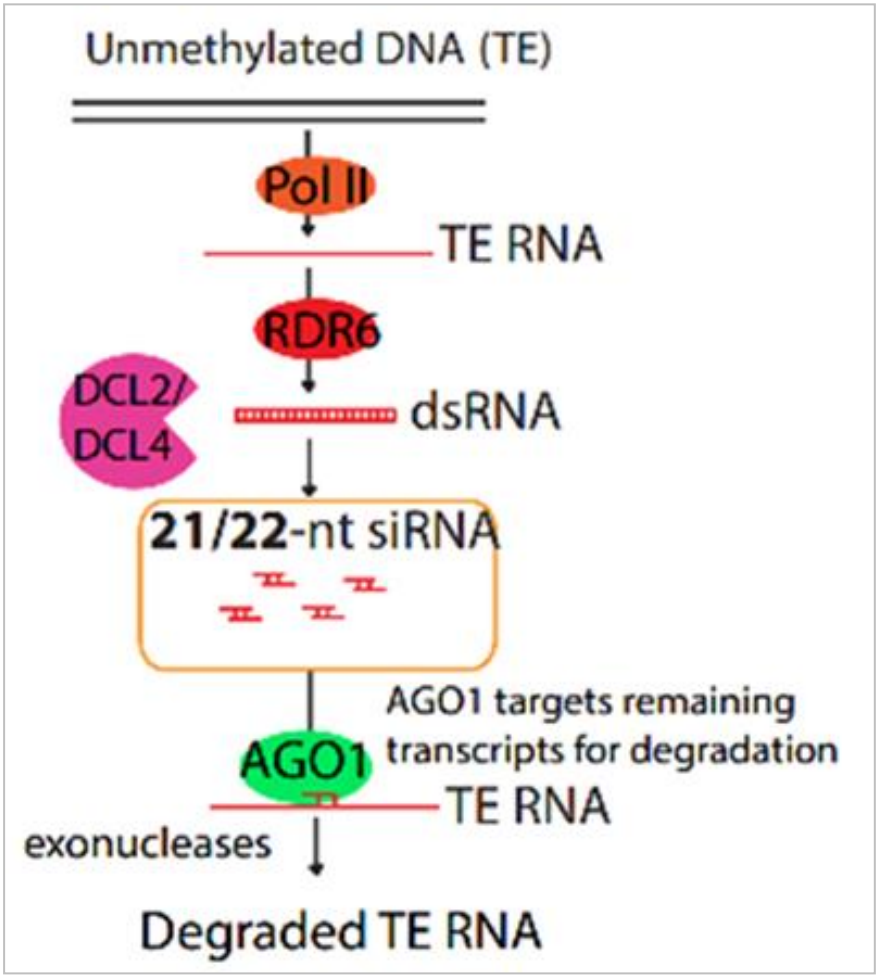

Figure 1: The RNAi pathway

An active TE, originating from un-methylated TE DNA, is first transcribed into mRNA by POL II, before being converted into dsRNA by RDR 6. This dsRNA is then degraded into 21/22 nt small RNA fragment, which then act to facilitate destruction of mRNA with similar sequences through interaction with AGO 1.

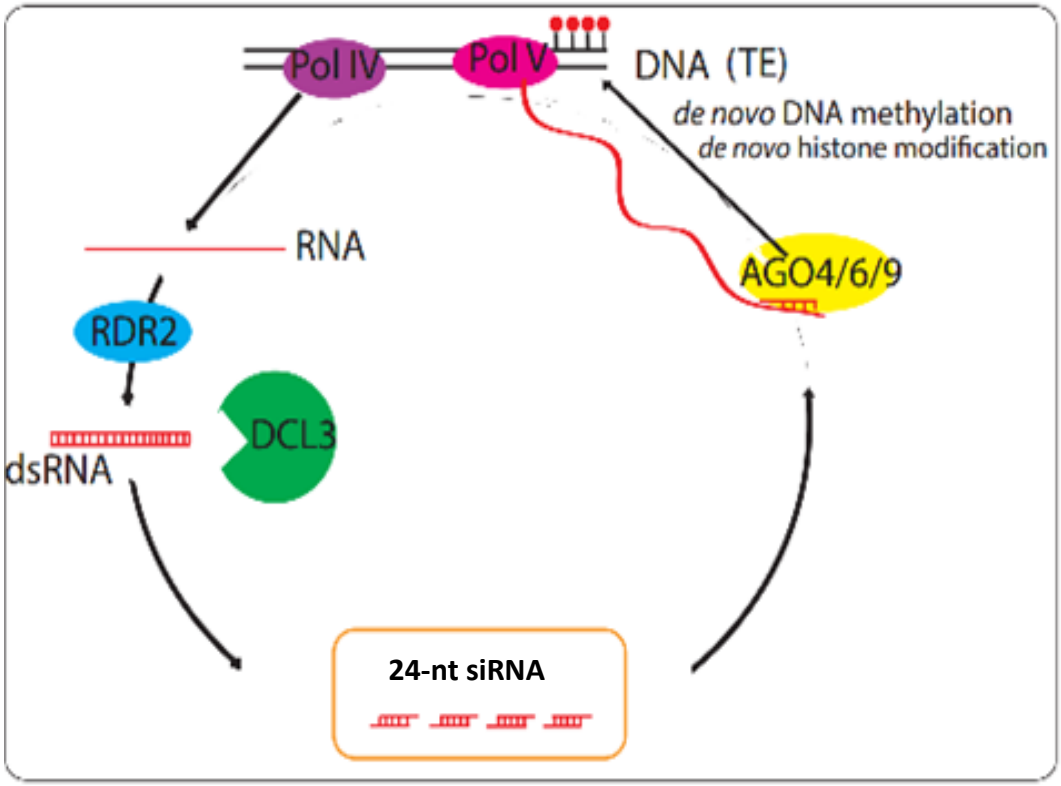

Figure 2: The RdDM pathway

Transcription into long non-coding RNA and cleavage into 24 nucleotide (nt) siRNAs drives the RNA-directed DNA methylation of TEs. First POL IV transcribes an ssRNA molecule, which is subsequently converted into into a dsRNA molecule by RDR2. This molecule is then cleaved into 24 nucleotide siRNA, by DCL3, before incorporating into AGO4/6/9. This complex then recognizes a scaffold transcript of the active TE produced by POL5, which recruits other proteins and initiates DNA methylation.

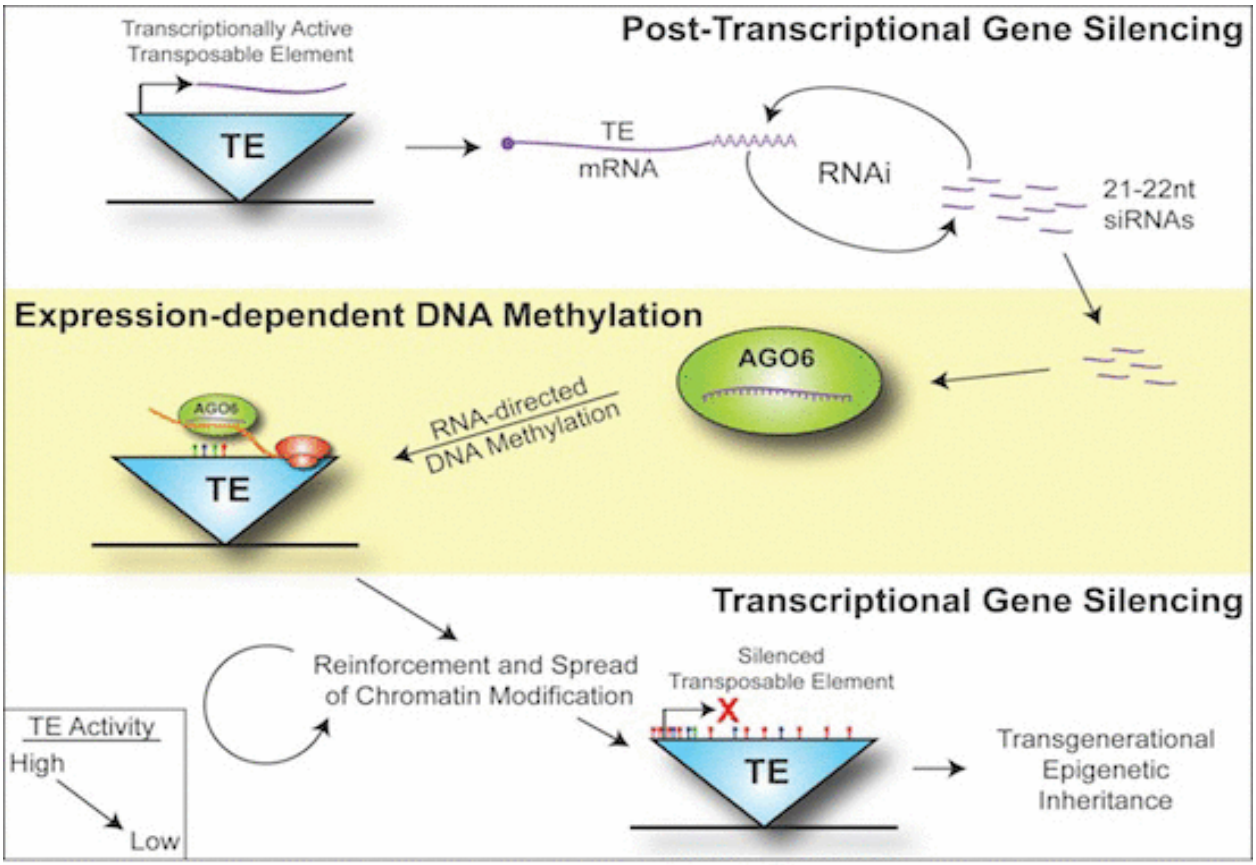

Figure 3 [13]: Overview of connection between RNAi (first line) pathway and RdDM (second line) pathway, leading to transcriptional silencing of the TE (bottom line**).** The RNAi pathway dogmatically leads to 21-22 nucleotide siRNAs which can incorporated into AGO 6. This complex can then help facilitate DNA methylation through the RdDM pathway. Once methylation is initiated, its reinforcement and spread is facilitated by maintenance methylation, leading to effective silencing of the element. This silencing is maintained through generations via mechanisms of epigenetic inheritance. This figure originates from a recent publication of the Slotkin lab in the EMBO Journal [13].

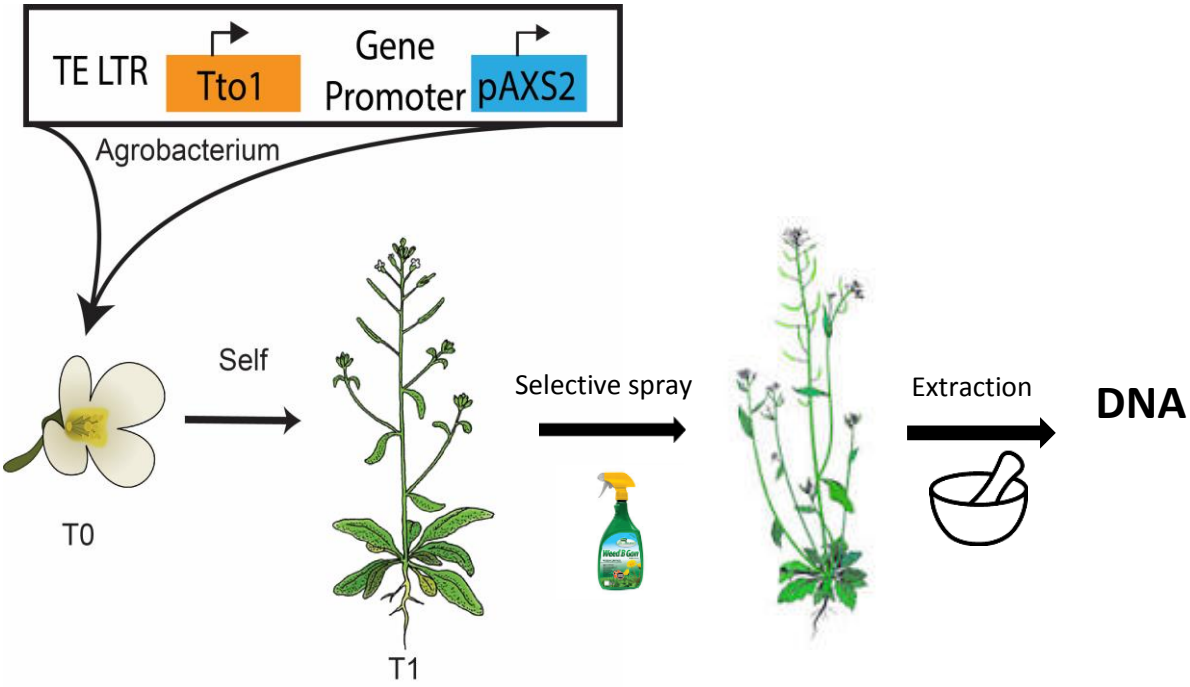

Figure 4: Overview of construct insertion into *Arabidopsis thaliana*. Agrobacterium is used to insert the construct of interest in into mature *Arabidopsis* plants. These plants are then selfed and their progeny are planted out before being sprayed with a selective agent (Basta). This agent selects for only plants that have taken up the transgene. The surviving plants are then used for DNA extraction using DANGP. This figure was adapted from a presentation made by a member of the Slotkin lab, Dalen Fultz.

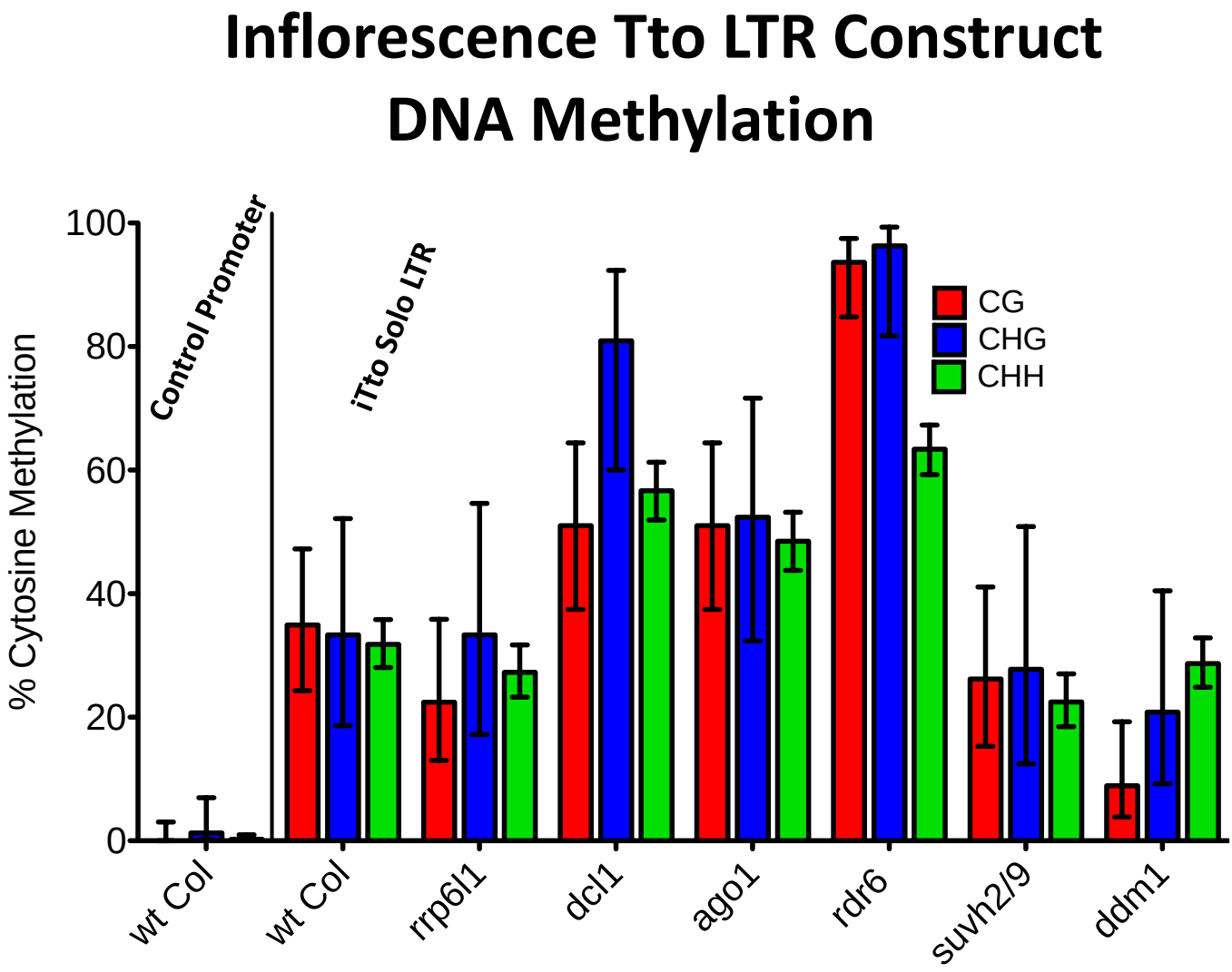

Figure 5: Methylation of the exogenous TE LTR construct and control promoter in three different cytosine contexts. The Tto LTR construct achieves significant methylation in the wild type background. The control construct (AT2G3005) does not become significantly methylated. Only the *rdr6, ago1,* and *dcl1* mutants achieve significantly higher methylation levels than the wild type control. *Sde3* methylation levels are not shown in this graph as the bisulfite sequencing for this mutant failed in the inflorescent context

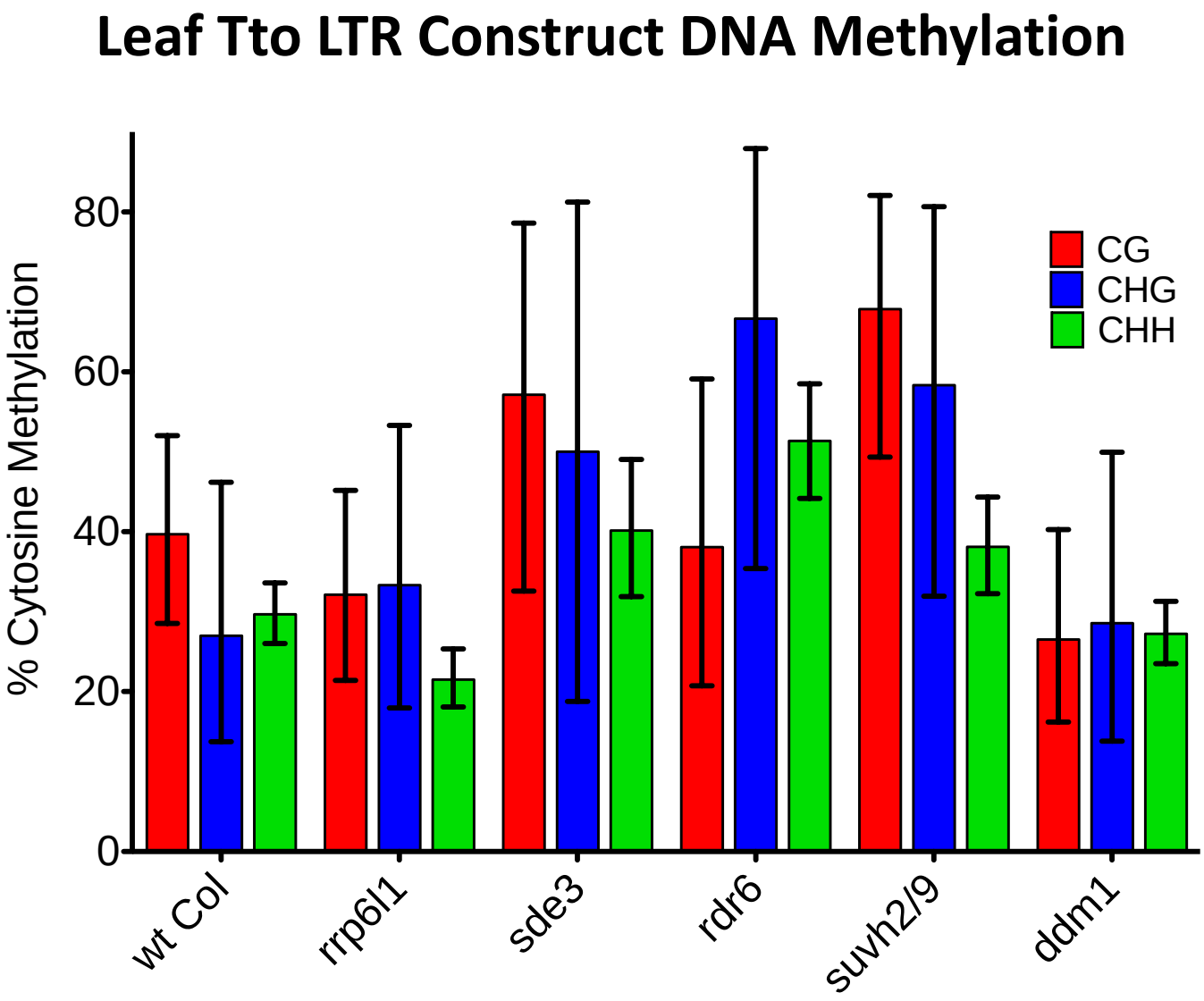

Figure 6: Methylation of iTto solo LTR cytosines in three different cytosine contexts. The methylation data corresponding to the *dcl1* and *ago1* mutants is not shown as the bisulfite sequencing failed in the leaf context. Methylation of *rrp6/1*, *sde3, suvh2/9*, and *ddm1* is similar to that seen in the inflorescent context and the wild type control. Again, *rdr6* is seen to have higher levels of CHH methylation than that achieved in the wild type context, reinforcing the inflorescence findings for this mutant.

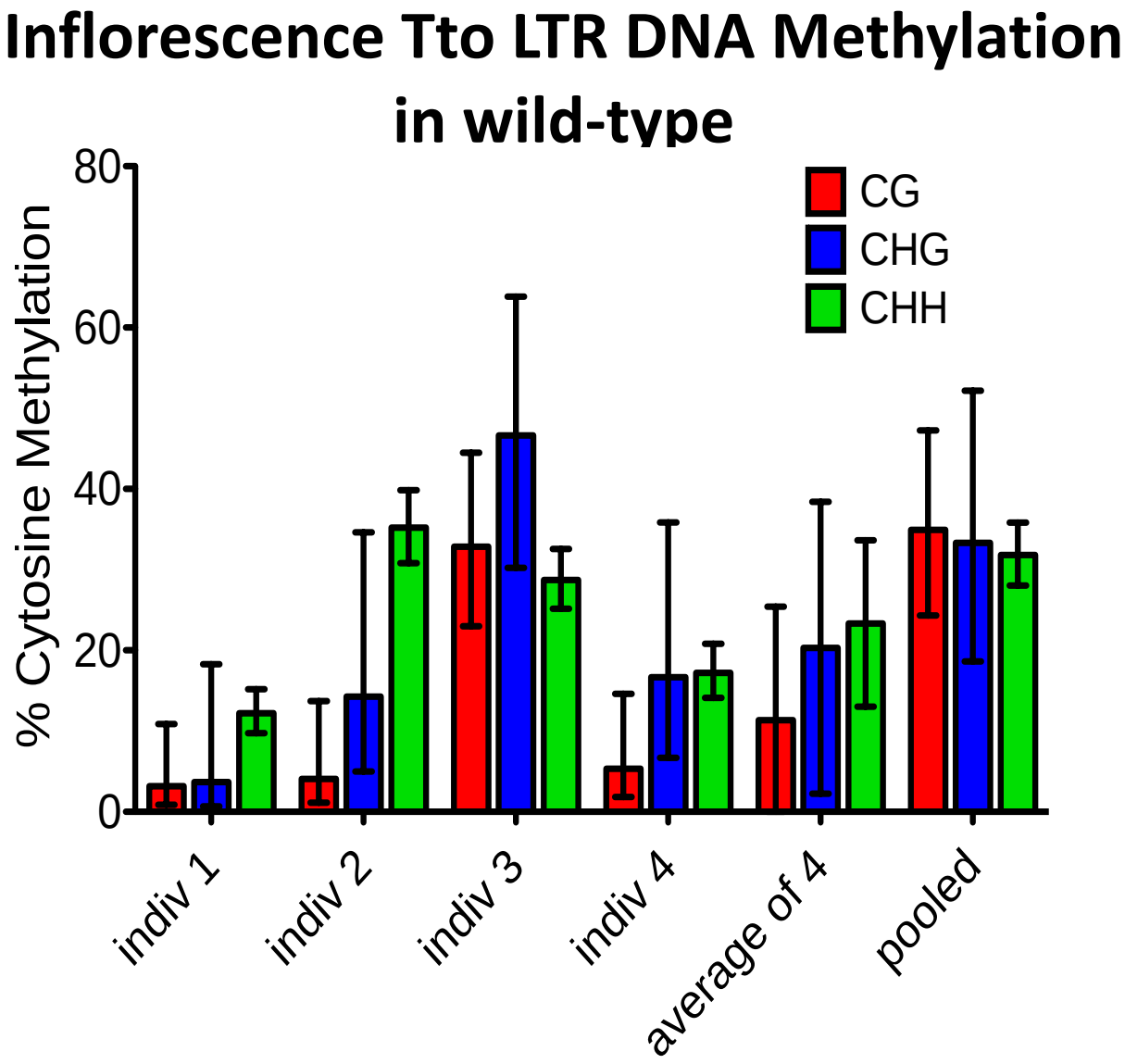

Figure 7: Methylation of exogenous TE LTR construct in wild type *Arabidopsis thaliana* for three different cytosine contexts. Methylation of the exogenous Tto LTR construct can be seen to vary significantly between individual plants. However, it appears that the average of the mutant tissue and majority of the CHH methylation appears to be similar to the pooled methylation levels, validating the pooled experimental procedure utilized in this thesis.

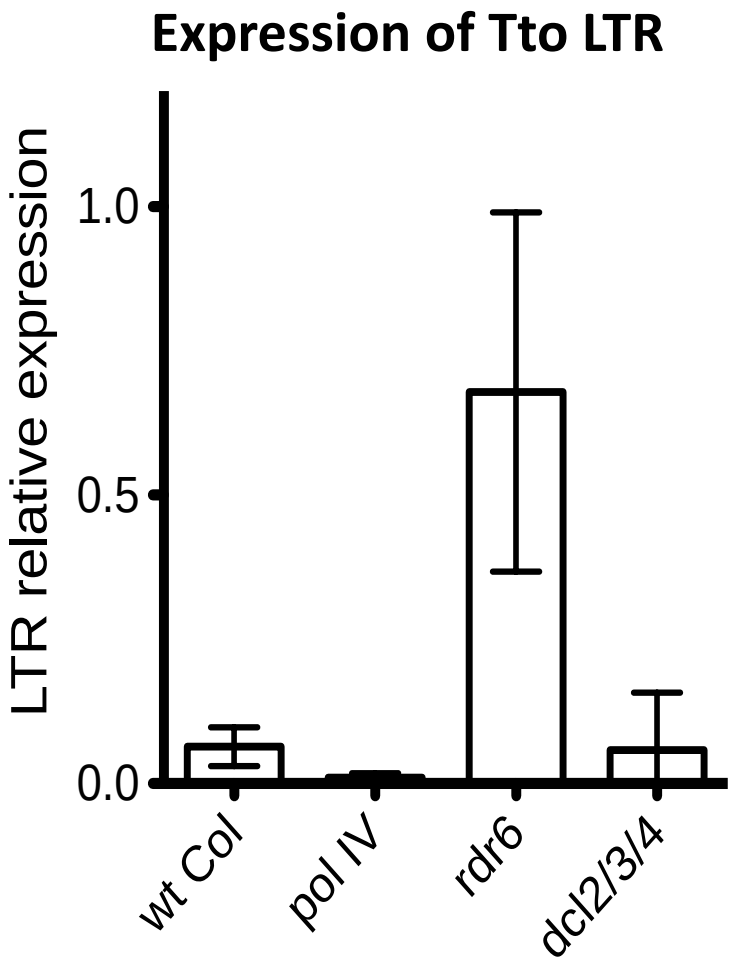

Figure 8: Relative mRNA level of exogenous Tto solo LTR. The expression level of the Tto LTR construct is low, suggesting that the level of methylation achieved in this context is sufficient to silence the element. However, it must be noted that the relative expression level remains around 10%, which is not insignificant. This could be due to the nature of the Tto construct used or the *de novo* DNA methylation pathway. Methylation levels in *rdr6* are significantly higher than that achieved in the wild type background, despite the high methylation in *rdr6*. This suggests that this methylation is not sufficient to silence the construct. The *pol4* and *dcl2/3/4* mutant lines are present for comparison. It appears that mutants in these pathways decrease expression, which is an unexpected finding. This could be due to the nature of the *de novo* DNA methylation pathway and demands future investigation.

Individual	Position
1	Only vector sequence
2	Internal to AT2G31005
3	Inverse PCR failure
4	Only vector sequence

Table 1: The genomic location occupied by the Tto LTR element in four wild type *Arabidopsis thaliana* individual plants. Individual 1 and 4 only mapped to vector sequence, likely corresponding to the insertion of the Tto element into *Arabidopsis* through the use of an Agrobacterium vector. The Inverse PCR procedure failed in individual 3. The Tto LTR construct in individual 2 maps internal to the little studied *Arabidopsis thaliana* gene AT2G31005.

Inverse PCR Protocol (Supplement 1)

1. Obtain tissue of interest (Dry ice or liquid nitrogen for inflorescence and regular ice for leaf).

2. Extract DNA using a standard DANGP protocol

3. RNAse the sample following the standard RNAse protocol and subsequent cleanup

4. Nanodrop (will need approximately 1 ug of RNAsed DNA)

5. Digest (using appropriate restriction enzymes) with 1 Ug of DNA in 10 units of enzyme (total of 50 ul reaction) for 10 hours. No need to heat inactivate enzymes.

6. Add the **entire** ligation mixture to a mixture containing 15 units of T4 DNA ligase, 10 ul of 10X ligation buffer, and H_20. Mix gently and incubate at 14^0C for 10 hours.

7. Prepare two sequential, nested PCR.

7a) Utilize standard 20 ul PCR MyTaq recipe using 5X MyTaq buffer and Taq polymerase. Using 3 ul of ligated DNA improves output in tricky experiments. Use appropriate primers for the target of interest and ensure PCR conditions (standard MyTaq) are adjusted accordingly. Use 32 cycles instead of 30.

7b) Utilize the same PCR recipe and cycler conditions for the secondary, nested PCR. Ensure the correct, **Different (from the primary PCR)** primers are used and cycler conditions are adjusted accordingly (still use 32 cycles). Use only 1 ul of PCR product. **Do NOT DILUTE the primary PCR. This lowers yields significantly.**

8. Run this out on a gel to confirm presence and purity of unique bands. Utilize an appropriate ladder. I have found a 2%, medium gel allows for easier subsequent extraction.

9. Gel extract the bands of interest (perhaps the brightest from each sample) and follow standard gel extraction cleanup procedure.

10. This gel extracted product can be directly submitted to sequencing at the PMGF, utilizing the correct format for PCR products. **DO NOT SUBMIT AS YOU WOULD A PLASMID PRODUCT!**