The Molecular Characterization of Epigenetic Mutants in *Arabidopsis thaliana*:

Categorization of active *Athila6* elements in *ddm1* and *ddm1/rdr6* mutants

Honors Research Thesis

Presented in partial fulfillment of the requirements for graduation with honors research distinction in Molecular Genetics in the undergraduate colleges of The Ohio State University

By

Jennifer Bosse

The Ohio State University

August 2011

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

Abstract ...................................................................................................................................................pages 3 to 4

Background ..............................................................................................................................................pages 5 to 11

Materials and Methods .......................................................................................................................pages 12 to 14

Data and Analysis.....................................................................................................................................pages 15 to 19

Conclusion ...............................................................................................................................................pages 20 to 21

Future Directions and Current Work ..................................................................................................pages 22 to 24

*Figures are not included in this document. All of the figures are found in the supporting document,

Figures of the Molecular Characterization of Epigenetic Mutants in Arabidopsis thaliana:

Categorization of active Athila6 elements in ddm1 and ddm1/rdr6 mutants.
Abstract:

Transposable elements (TEs) are mobile genetic units that cause mutations, changes in gene regulation, double strand breaks in DNA and non-homologous recombination. For those reasons, TEs are repressed through epigenetic regulation. The TEs studied in this project are the Athila6 family of LTR retrotransposons found in the model plant, Arabidopsis thaliana. One method Arabidopsis employs to repress Athila6 is transcriptional silencing. Decrease in DNA methylation 1 (DDM1) is a chromatin remodeler that represses Athila6 by keeping it in a silent, heterochromatic state. When Athila6 is not silenced transcriptionally, post-transcriptional pathways are utilized. RNA Dependent RNA polymerase 6 (RDR6) is a major protein involved in that process.

Wild type Arabidopsis, which have chromatin-level regulation, do not transcribe Athila6 mRNA as confirmed by Northern analysis. In Arabidopsis mutants where chromatin-level regulation is lost, the numerous Athila6 elements become transcriptionally active and start producing mRNA. Northern analysis of ddm1, rdr6 and ddm1/rdr6 mutant plants reveal that ddm1 has many Athila6 transcripts of various sizes, rdr6 has no transcripts (due to the continued chromatin silencing) and ddm1/rdr6 has additional transcripts in sizes not found in the ddm1 single mutant. The goal of this project is to characterize what Athila6 elements contribute to the transcripts accumulating in ddm1 and ddm1/rdr6 mutants. My research is aimed at deciphering if the same Athila6 elements are transcribed in ddm1 but have undergone different or additional post-transcriptional processing, or if ddm1/rdr6 is transcribing unique Athila6 elements.

Through the characterization of the active Athila6 elements found in ddm1 and ddm1/rdr6, it was demonstrated that the additional transcripts in ddm1/rdr6 are not elements unique to ddm1/rdr6. It is
most likely that the two mutants have the same transcriptionally active \textit{Athila6} elements but some of the elements in \textit{ddm1/rdr6} have undergone different and/or additional post-transcriptional processing.
Background:

Transposable elements:

Transposable elements (TEs) are genomic sequences that are capable of moving from one genomic location to another. It is important to study transposable elements because the transposition of TEs can have a variety of affects on an organism’s genome. The majority of time active transposable elements led to genome instability through mutagenesis and chromosome instability. TEs are mutagenic because their transposition often creates non-functional proteins. TEs cause chromosome instability through double-stranded breaks and non-homologous recombination. Transposable elements also affect gene regulation. TEs often contain promoter and enhancer sequences, and because of this, TEs can influence the host genes near it. TEs can additionally influence host genes by causing nearby host genes to be regulated with the same epigenetic control that the TEs are subject. Transposable elements can positively affect the host’s genome by transposing into protein coding regions and creating novel proteins. There are more than 100 cases where transposable element insertions caused the divergence of orthologous genes from different species. An other positive affect of transposable elements is when TEs are co-opted by the host organism. *Drosophila melanogaster* does not have telomerase to preserve its telomeres. So, instead of telomerase, it uses HET–A and TART LINE like TEs. These elements repetitively transpose and insert at the chromosome ends, causing the telomeres to extend.

There are two classes of transposable elements. Class I elements, called retroelements, transpose using an RNA intermediate created by reverse transcription of the template RNA. Class II elements, called DNA elements, transpose using DNA directly, without an RNA intermediate, in a cut and paste mechanism.
The two classes of TEs can further be divided. Class I elements contain the subclasses of LTR and non-LTR elements. Long Terminal Repeats (LTRs) have a structure similar to retroviruses and are closely related evolutionarily. LTR elements encode long terminal repeats and the structural and enzymatic proteins used in their retrotransposion. The element studied in this project, *Athila6*, is a LTR retrotransposon.

Non-LTR elements include SINE and LINE like elements. LINE like elements are the most abundant non-LTR elements and compose over a quarter of the human genome themselves. All TEs families have non-autonomous members. Non-autonomous TEs cannot transpose by themselves and rely on the transposon machinery encoded by autonomous TEs. The SINE subclass is an example of a completely non-autonomous TE subclass. The SINE elements do not encode transposition machinery and are transposed in *trans* by enzymatic machinery encoded by the autonomous LINE elements. The Alu elements, which compose over 10% of the human genome, are members of the SINE subclass of TEs (Bowen and Jordan).

*Athila* is a large family of LTR retrotransposons, with hundreds of intact and fragmented elements, representing 3% of the total *Arabidopsis thaliana* genome. It has many subfamilies including *Athila6*, whose epigenetic regulation is focused on in this project. The structure of *Athila6* LTR retrotransposon is shown in Figure 1.

LTR retrotransposons transpose through a RNA intermediate. More specifically, LTR retrotransposons are first copied by the host’s RNA polymerase II, which is signaled by a promoter found in the 5’ LTR. The translation of the RNA occurs within the cytoplasm and yields the proteins encoded by the *gag* gene, which form the *gag* capsid particle, and the proteins encoded by *pol*, the reverse transcriptase and the integrase. Each *gag* capsid particle contains two RNA copies of the
element and reverse transcription takes place within the gag capsid particle. After the reverse transcriptase creates a double-stranded DNA copy of the retrotransposon, the integrase inserts it into the genome, creating a second copy of the element. This replication of cycle differs from LTR retroviruses in that LTR retrotransposons do not encode for the envelope protein and therefore never leave the cell during their replication cycle.

All LTR retrotransposons have long terminal repeats and autonomous elements have gag and pol genes as well. Some elements also have retained an ancient env-like domain which is contains no currently known function. As mentioned previously retrotransposons are structurally similar to retroviruses. Retroviruses have env or envelope genes, which allow the viruses to leave the host cell and infect other cells. Retrotransposons with an ancient env-like domain contain what was an open reading frame, ORF, in the same position as the env gene in retroviruses. The Athila elements retained these degraded env-like genes from their evolutionary history as retroviruses. Over evolutionary time, Athila’s env gene lost its protein coding capacity and Athila could no longer leave the cell. This switched its replication cycle from that of a LTR retrovirus to that of a LTR retrotransposon. There is also a long non-coding region between the env like domain and the 3’ LTR found in several Athila elements, as well (Havecker, Gao and Voytas).

**Epigenetic Regulation:**

Transposable elements were previously mentioned as having gene regulating and species diverging properites but are better known for their dangerous mutagenic properties. Due to these mutagenic properities, transposable elements like Athila6 are transcriptionally silenced through epigenetic regulation. Certain epigenetic mutants display active Athila6 expression. The two epigenetic
mutants used in this project are decrease in DNA methylation 1 (ddm1), and RNA dependent RNA polymerase 6 (rdr6).

/DDM1/is a SWI2/SNF2- like protein. SWI2/SNF2-like proteins have a variety of functions including transcriptional control, DNA repair, chromatin regulation, and chromosome folding. /DDM1/, as a chromatin remodeler, hydrolyzes ATP to move and slide nucleosomes, resulting in the remodeling of chromatin into compact, transcriptionally silent heterochromatin. DDM1’s homolog in mouse is LSH1, and their function is conserved. DDM1 acts at TEs to compact the DNA into heterochromatin, creating a chromatin environment inaccessible for transcription. TE heterochromatin formation by DDM1 leads to changes in histone tail modifications and cytosine DNA methylation (Kim, Samaranayake and Pradhan). In Arabidopsis, there is a 70% decrease in DNA methylation in ddm1 mutants, and TEs become transcriptionally active (Jeddeloh, Stokes and Richards).

RNA dependent RNA polymerases (RdRPs) work in endogenous RNA interference (RNAi) pathways to amplify small RNA production. RdRPs convert single stranded RNA into double stranded RNA by creating a complementary strand through primer independent mechanisms. These dsRNA molecules are processed into small interfering RNAs (siRNA). These siRNAs can work transcriptionally, post-transcriptionally or on the translational level to silence protein production from mRNAs complementary to the small RNAs. RdRPs are found in fungi, plants, and some invertebrate animals, while in other animals evolutionarily distinct mechanisms are used to amplify endogenous siRNAs (Zong, Yao and Yin). In Arabidopsis thaliana, which has 6 RdRP genes, RDR6 is the major RNA dependent RNA polymerase that acts in post-transcriptional silencing (Curaba and Chen).

Eukaryotic RNA silencing targets TEs, viruses, transgenes and certain genes. The core mechanism for RNA silencing involves a trigger, which is generated through double-stranded RNA
(dsRNA). There are several mechanisms to generate dsRNA including bidirectional transcription of DNA, self-complementary RNA foldbacks or RNA-dependent RNA transcription. In *Arabidopsis*, there are two biogenesis classes in canonical RNA silencing. One class is short interfering RNA (siRNA). SiRNAs are assembled to be perfectly complementary dsRNA by RDRPs or other mechanisms, such as long foldback-inverted repeats. The other class is microRNA (miRNA), which is assembled from imperfect short foldbacks. SiRNAs and miRNAs become active when the RNase III-type activity of a DICER or DICER-LIKE (DCL) enzyme cleaves the dsRNAs to form 21-25 nucleotide (nt) small RNA duplexes. One strand of the RNA duplex is then incorporated into RNase H-like ARGONAUTE (AGO)-containing effector complex. The RNA serves a sequence-specific guide for the AGO protein, and results in target transcript degradation.

In Arabidopsis, there are two pathways to produce siRNAs. The first pathway works on the transcriptional level to silence TEs. This pathway uses RNA-DEPENDANT RNA POLYMERASE2 (RDR2) and DCL3 to make 24 nt-sized siRNAs which when incorporated with AGO4 initiates or maintains DNA methylation and histone modification. This pathway is used to silence transposons and other repeat classes and is illustrated in Figure 2. DDM1 recruitment to TEs is thought to occur through this pathway, resulting in the formation of heterochromatin and transcriptional silencing.

The second siRNA pathway in Arabidopsis uses RDR6 and DLC4 to create siRNAs are 21 nt in size and work post-transcriptionally. This pathway requires a POLII-derived transcript, and works in the cytoplasm, performing RNAi on transcripts that are recognized by the siRNAs (Howell, Fahlgren and Chapman). A diagram of this pathway is shown in Figure 3.
Regulation of \textit{Athila6}:

First, DDM1, on the transcriptional level, and then RDR6, on the post-transcriptional level, work together to silence transposable elements. Figure 4 is a Northern blot performed by the Slotkin lab (unpublished) using polyA+ mRNA and various probes for \textit{Athila6}. In Figure 4, \textit{ddm1} mutants have many different sized \textit{Athila6} gag/pol and \textit{Athila6} 3′CDS sense transcripts, where as \textit{rdr6} mutants have no \textit{Athila6} gag/pol or \textit{Athila6} 3′CDS sense transcripts, which is due to the hierarchical repression of \textit{Athila6} silencing with DDM1 above RDR6. The double mutant, \textit{ddm1/rdr6} has additional, different sized transcripts than found in \textit{ddm1}, as well as increased overall levels of \textit{Athila6} gag/pol transcripts, which has been verified as statistically significant by qRT-PCR (data not shown). These additional transcripts in \textit{ddm1/rdr6} could represent a synergistic effect in the double mutant, meaning that the silencing of the \textit{Athila6} elements represented in those additional transcripts require both DDM1 and RDR6, transcriptional and post-transcriptional control. On the other hand, these additional transcripts could represent a change in the post-transcriptional processing of those \textit{Athila6} elements already expressed in \textit{ddm1} mutants.

The biological question that is raised by the Northern blot analysis in Figure 4 is what accounts for the differences in \textit{Athila6} transcripts between \textit{ddm1} and \textit{ddm1/rdr6}? The goal of this project was to determine the nature of this difference.

To test if the same \textit{Athila6} elements in the genome produce the different transcripts between \textit{ddm1} and \textit{ddm1/rdr6} seen in Figure 4, I have generated two hypotheses. My null hypothesis states that there is no significant difference between which \textit{Athila6} elements in the genome produce transcripts among the two mutants and that any differences are due to random chance. This means that the additional transcripts appearing in the \textit{ddm1/rdr6} double mutant are produced by the same active
Athila6 elements found in the ddm1 single mutant. As such, these transcripts must have been processed differently in the double mutant, resulting in the additional transcripts that appear on the Northern blot.

My alternative hypothesis states that there is a significant difference between which Athila6 elements produce transcripts between the two mutants, and this difference can not attributed to random chance. The biological source of this difference could come from either ddm1 or ddm1/rdr6. Assuming the double mutant has all of the transcripts that the single mutant has in the Northern blot and the predicted hierarchical nature of Athila6 repression, I predict that ddm1/rdr6 is causing the significant difference. This is my alternative prediction and if the alternative hypothesis is right then it is because the additional transcripts found in the ddm1/rdr6 double mutant represent the transcription of additional Athila6 elements not transcribed in the ddm1 single mutant. If this prediction is correct then these additional Athila6 elements could represent an area of overlap in transcriptional and post-transcriptional regulation meaning both pathways could be needed to act on certain elements.
Materials and Methods:

RNA was extracted from three biological replicates from each genotype; wild type *Arabidopsis thaliana* in the Columbia ecotype (Col), *ddm1* mutants and *ddm1/rdr6* mutants. RNA was extracted in order to study the activity of the Athila6 elements rather than extracting DNA, which would only show the elements themselves. The mRNA in the total RNA collected from the plants was reverse transcribed using an oligo-dT primer and reverse transcriptase. The resulting cDNA was PCR amplified using four different sets of primers. The four primer sets corresponded to four different regions of the Athila6 element as shown in Figure 5.

After PCR amplification, the products were run on a 2% agarose gel as shown in Figure 6. The PCR products for each replicate of the same genotype were extracted from the agarose gel and then pooled and cloned into TOPO vector pCR4 for sequencing. The purpose of this pooling was to reduce the variation due to individuals. The sequences were BLASTed to the *Arabidopsis thaliana* genome using The *Arabidopsis* Information Resource (TAIR) website.

The Athila6 gag/pol primers were tested using this method before the other three primer sets. One problem encountered using the gag/pol primers was that the gag/pol region is a highly conserved, and some of the clone sequences matched more that one Athila6 element, making the determination of which element was transcribed impossible. To get a more complete idea of the Athila6 transcripts present, the other 3 primer sets were used. The Athila6 3’CDS and Athila6 3’845 were chosen particularly for specificity. The Athila6 3’ CDS region is more variable than gag/pol, as it does not encode a protein. Also not all Athila6 elements have the 3’ CDS, as some are fragmented non-autonomous elements.
Shown in Figure 6, when the 3 other primer sets were run out on the agarose gel the *Athila6 LTR* PCR product ran as two products. These products were cut out and cloned separately. The top band is called *Athila6 LTR* set 1, and the bottom band is called *Athila6 LTR* set 2.

A total of 100 clones, 50 with plasmids containing *ddm1* transcripts and 50 with plasmids containing *ddm1/rdr6* transcripts were collected for each primer set. The *Athila6 3’CDS* set is an exception, as the TOPO reaction yielded fewer clones, as well as the *gag/pol* set where 72 or 73 clones from each mutant were sequenced. As mentioned previously, the clones were sequenced and BLASTed against the TAIR website. The top hit (lowest $e$ value) from either the TAIR10 all transcripts annotated sequence, or the TAIR10 annotated intergenic sequences was determined to be the source of the transcript, while in some cases several different elements in the genome matched the sequence with the same $e$ value, and in those cases each sequence was divided as not to over-represent nearly identical *Athila6* elements. Both transcript and intergenic annotated sequences were queried because the annotation of TEs is often poor, and the best sequence match could be in either category.

The gene numbers for best matches to sequenced transcripts were placed in an Excel spreadsheet in the column following the clone number. Once all the clones had been sequenced, then all genes present for that primer set were ordered numerically. Each time a gene appeared in *ddm1* or *ddm1/rdr6*, it was scored as such. The clones, which matched to more than one gene, were scored by dividing 1 by the total of genes that the clone matched. For example if a gene appeared in *ddm1* in one clone but that clone also matched two other genes, then that gene would be scored as appearing in *ddm1* $1/3$ times, not 1 time.

The scores for *ddm1* were compared to the scores for *ddm1/rdr6* by F-test analysis. The F-test analyzes the variances between the two mutants and gives the two-tailed probability that the two are not
significantly different (FTEST). The two mutants were interpreted as significantly different if the F-test yielded a probability lower than 0.05.
Data and Analysis:

**gag/pol primer set:**

The *Athila6* copies sequenced for the *gag/pol* primers are shown in Figure 7. There were a total of 145 elements represented in all of the sequencing combined (72 in *ddm1*, and 73 in *ddm1/rdr6*) The F-test analysis of the *gag/pol* primers returned a probability of 0.51. This means that the variances between *ddm1* and *ddm1/rdr6* are not statistically significant, and, therefore, *ddm1* and *ddm1/rdr6* are expressing the same *Athila6* elements. This supports the null hypothesis and that the additional transcripts seen on Figure 4 represent the same *Athila6* transcripts produced by *ddm1*, but these same transcripts are processed differently in the *ddm1/rdr6* double mutant. There was one element, AT1G37060, where *ddm1/rdr6* expression was much higher than *ddm1*. This was invested further by creating a PCR primer unique to the AT1G37060. The *ddm1* and *ddm1/rdr6* PCR products were about equal in size and intensity, therefore the greater number of AT1G37060 transcripts found in *ddm1/rdr6* clones was due to random chance and not biologically real.

**Athila6 LTR primers:**

The *Athila6* elements identified by amplifying and sequencing RT-PCR products from the *Athila6* LTR are shown in Figure 8. In all cases, these LTR regions matched intergenic regions of the genome, and the LTR is named for the two transcripts that surround it. The F-test from the *Athila6* LTR primer set 1 gave a statically significant probability of 0.00168. This means there is a significant difference in variances between *ddm1* and *ddm1/rdr6*, which is not attributed to random chance. The intergenic region between elements AT1G40310-AT1G40076, (Figure 8), is sequenced 15 times in *ddm1* and never in *ddm1/rdr6*. While the F-test for the *Athila6* LTR primer set 1 does not support the null hypothesis, the region between elements AT1G40310-AT1G40076 does not support the alternative
prediction either, since a large portion of the difference in variances is coming from a region found only in *ddm1*, and not *ddm1/rdr6*. When that region is removed the probability is increased over 18 fold to 0.03139. Since, *Athila6* LTR primer set 1 does not support the alternative prediction, it must support the other possible cause for the alternative hypothesis which is that *ddm1* is the source of the difference in variants. Region AT1G40310-AT1G40076 also supports *ddm1* being the source of the variance for this primer set but this region was sequenced one time in *ddm1/rdr6* in the *Athila6* 3’845 primer set. So, it is not a transcript unique to *ddm1* and is not the source of the additional transcripts in Figure 4 because it was sequenced in both mutants. Because this region did not fit the alternative prediction, element-specific RT-PCR as in the bottom panel of Figure 7 was not performed.

The data from the *Athila6* LTR primer set 2 conflicts with the data from the *Athila6* LTR primer set 1, because the F-test for *Athila6* LTR primer set 2 gave probability of 0.69. This means that the variances between *ddm1* and *ddm1/rdr6* are not statistically significant and are caused by random chance. This supports the null hypothesis and that *ddm1* and *ddm1/rdr6* are transcribing the same *Athila6* elements. Figure 9 shows the distribution of *Athila6* elements in the *ddm1* and *ddm1/rdr6* clones.

To analyze all of the LTR data together, I pooled both sets of data and reanalyzed. Figure 10 depicts the combined data from both *Athila6* LTR primer sets. The F-test gave a not statistically significant probability of 0.38 meaning that the difference in variances between *ddm1* and *ddm1/rdr6* are due to random chance for the *Athila6* LTR primer set overall. So, the combined *Athila6* LTR data supports the null hypothesis and that the two mutants are expressing the same *Athila6* elements. The combined data also disputes the alternative prediction, because two of the three most highly expressed intergenic regions, AT1G40310-AT1G40076 and AT5G34985-AT5G34990, are only expressed in *ddm1* single mutants.
**Athila6 3’ 845 primer set:**

The *Athila6* elements identified by amplifying and sequencing RT-PCR products from the *Athila6* 3’845 are shown in Figure 11. In all cases, these 3’ to miRNA845 regions matched the intergenic regions of the genome, and the region is named for the two transcripts that surround it. The F-test for the *Athila6* 3’ 845 primer set gave a probability of 0.93. This means that the difference in variances between *ddm1* and *ddm1/rdr6* are not statistically significant and are caused by random chance. This supports the null hypothesis and that *ddm1* and *ddm1/rdr6* are expressing the same *Athila6* elements.

**Athila6 3’CDS primer set:**

As mentioned in the materials and methods, the *Athila6* 3’CDS primer set yielded few clones. Figure 12 shows the data from the first attempt at TOPO cloning. There were 13 *ddm1* clones and 15 *ddm1/rdr6* clones. The F-test gave a probability of 0.069, which is not statistically significant and supports the null hypothesis. Nevertheless there were a low number of clones and the probability was close to 0.05. For those reasons, the *Athila6* 3’CDS primer set was TOPO cloned again. Figure 13 shows the combined data from the original TOPO cloning and the repeated TOPO cloning.

The repeated TOPO cloning did not yield many clones either, and the Slotkin lab now has confirmed that *E.coli* are transformed this portion of the *Athila6* element with low efficiency, potentially due to the predicted strong secondary structure in this region. The total number of clones between the original TOPO cloning and the repeated was 20 *ddm1* clones and 16 *ddm1/rdr6* clones. The F-test for the combined clones gave a statistically significant probability of 0.0116 meaning that the difference in variances between the mutants is not due to random chance (Figure 13). So, while the original TOPO cloning results were not statistically significant and supported the null hypothesis, the combined
clonings are statistically significant and do not support the null hypothesis. Despite not supporting the null hypothesis, the combined data does not support the alternative prediction either. In Figure 13, there are no elements unique to \textit{ddm1/rdr6} and 5 elements found only in \textit{ddm1} single mutants. Figure 13 also shows there are 4 elements that have higher expression in \textit{ddm1/rdr6}: AT5G29975, AT5G31719, AT5G32107 and AT5G32306. These 4 elements would be the best candidates for the cause of the increased expression in \textit{ddm1/rdr6} seen with \textit{gag/pol} probe in Figure 4. However, looking at the mRNA Northern blot in Figure 4 probed with the same \textit{Athila6} 3’CDS region as amplified and cloned in Figure 13, \textit{ddm1} expression for this region seems to be much higher than \textit{ddm1/rdr6}, and this is likely due to the elements At3g33067, At4g06517, AT5G32197, AT5G32228, AT5G32511, AT5G32624 and AT5G34836-AT5G34837. The source of the difference in variance is coming from both \textit{ddm1} and \textit{ddm1/rdr6} but it may be that the small size what is making those differences statistically significant.

Combined data from the primer sets:

The four different primer sets show which \textit{Athila6} elements are active and their levels of expressions in those individual parts of the \textit{Athila6} element, but Figure 14 gives a more global picture of \textit{Athila6} expression. Figure 14 combines the data from all the primer sets. The F-test for the combined data gives a probability of 0.67, which is not statistically significant. This supports the null hypothesis and that \textit{ddm1} and \textit{ddm1/rdr6} are transcribing the same \textit{Athila6} elements. Also, 2 out the 4 most highly expressed elements are only found in the \textit{ddm1} single mutant, which refutes the alternative prediction.

The combined primer set data can be broken down into genic and intergenic elements to see if there is difference between the two mutants in the genic and intergenic part of the \textit{Athila6} element. Figure 15 is the combined genic data from the \textit{gag/pol} and \textit{Athila6} 3’CDS primer sets. The F-test gives a probability of 0.33, which means that the difference in variances between the \textit{ddm1} and \textit{ddm1/rdr6} in
the genic region are not statistically significant and due to random chance. This supports the null hypothesis and that the two mutants are transcribing the same *Athila6* elements.

Figure 16 is the combined intergenic data from the *Athila6* LTR primers sets 1 and 2 and *Athila6* 3′845. The F-test gives a probability of 0.62, meaning that the difference in variances between the two mutants is not statistically significant and caused by random chance. This means in the intergenic region supports the null hypothesis. The intergenic region also disputes the alternative prediction because two of the three most highly expressed regions are found only in the *ddm1* single mutant.

Figure 17 combines the data from all the primer sets and shows that there is 22 *Athila6* elements found only in *ddm1* single mutants as compared to 15 *Athila6* elements found only in *ddm1/rdr6* double mutants. Not only are there more elements unique to *ddm1*, but also those elements are more often expressed. The elements expressed only in *ddm1* were sequenced 48.083 times, where as the elements expressed only in *ddm1/rdr6* were sequenced 19 times. This refutes the alternative prediction that *ddm1/rdr6* is transcribing additional *Athila6* elements since there are more elements unique to *ddm1* than *ddm1/rdr6*. If these unique elements are significant and not a product of the limited sample size, then Figure 17 supports the alternative hypothesis and supports *ddm1* as the source of the difference in transcripts.
Conclusion:

The data from the \textit{gag/pol} primer set, \textit{Athila6} LTR primer set 2, \textit{Athila6} LTR primer sets 1 and 2 combined, \textit{Athila6} 3’845 primer set, the original \textit{Athila6} 3’CDS cloning, the genic primer sets combined, the intergenic primer sets combined and all of the primer sets combined support the null hypothesis. The F-tests gave probabilities above 0.05 meaning the different in variances between \textit{ddm1} and \textit{ddm1/rdr6} is not statistically significant and are caused by random chance events. This indicates that the two mutants are transcribing the same \textit{Athila6} elements. Since the same elements are being transcribed, the additional transcripts on Figure 4 must represent RNAi degradation of some \textit{Athila6} transcripts by a RDR6 pathway in \textit{ddm1} single mutants, or some different or additional post-transcriptional processing that the \textit{Athila6} elements are undergoing in the double mutant related to the loss of RDR6. Also RNAi degradation would explain the increase expression of \textit{gag/pol} transcripts seen in Figure 4, although we currently do not understand why some transcripts are degraded, and others are not.

Although most of the data supports the null hypothesis, the data from the \textit{Athila6} LTR primer set 1 and the data from the combined clonings of the \textit{Athila6} 3’CDS do not support the null hypothesis. The F-tests gave probabilities of below 0.05 meaning that the difference in variances between the two mutants is significant and that something besides random chance is causing the difference. Even though those data sets do support the alternative hypothesis, they still do not support the alternative prediction that \textit{ddm1/rdr6} is transcribing additional different \textit{Athila6} elements.

The \textit{Athila6} LTR primer set 1 does not support the alternative prediction because a major source of the difference in variation is the intergenic region of AT1G40310-AT1640076 which is expressed in that primer set 15 times in \textit{ddm1} and 0 times in \textit{ddm1/rdr6}. The combined clonings of the \textit{Athila6}
3’CDS does not support the alternative prediction, in the combined data set there are 5 element unique to 
*ddm1* and no elements unique to *ddm1/rdr6*. In fact, as shown Figure 17, there are more elements unique to 
*ddm1* than *ddm1/rdr6*, which does not support the alternative prediction either.

In addition, while Figure 17 does show different *Athila6* elements in *ddm1/rdr6* which is a part of the alternative prediction, because the number of unique elements found in double mutant is less than number of unique element found the single, and also those elements found in the double mutant are not as frequently sequenced as the unique elements found in the single mutant, it’s more likely that those elements unique to *ddm1/rdr6* are found in both mutants but due to the size of the sample it was random chance that those elements weren’t found in *ddm1*. Random chance is also probably the reason the unique elements found in *ddm1* were not found in *ddm1/rdr6*, with the exception of AT5G34985-AT5G34990, which was sequenced 16.5 times in *ddm1*. On the other hand, the more unique elements found in *ddm1* could support the other possible prediction from the alternative hypothesis which states that *ddm1* is the source of the difference in variances between the mutants.

Since all of the primer sets, combined primer sets and Figure 17 do not support the alternative prediction, it must be rejected. Even though the alternative prediction is rejected, two of the data sets support the alternative hypothesis and that *ddm1* is the source of the difference in variances.

It is mostly likely that the null hypothesis is correct and that the additional transcripts seen on Figure 4 were transcribed from the same *Athila6* elements that are actively expressed in *ddm1*. Of the two sets that disagree with the null hypothesis, the *Athila6* 3’CDS primer set had such a low number of clones, it cannot be trusted until more clones are grown and sequenced. This leaves the *Athila6* LTR primer set 1 as the only evidence against the null hypothesis and Figure 4 was not probed for the LTR, so the LTR could not contribute to the additional transcripts on Figure 4.
Current Work and Future Directions:

My current work at this time is a PCR based approach. Although it was attempted, it was not possible to clone the gel-excised transcripts of each size from a repeated Northern analysis of ddm1 and ddm1/rdr6. Instead I am trying to use PCR to capture the different transcripts. Starting with oligo-dT primed cDNA, I am using different combinations of primers to scan across the Athila6 element and search for differences in the ddm1/rdr6 PCR products when compared to ddm1 PCR products. Since this approach is centered on differences in the PCR products, it is able to analyze the transcript differences in ddm1/rdr6 as compared to ddm1 based off more than size differences.

A program called Amplify 3x allows virtual PCR to be done on the Athila6 consensus sequence using a database of primers. Choosing one forward primer the program can show, for different reverse primers, the length of the PCR product and location of the PCR product. The program also shows how well the primer matches the consensus sequence.

Twenty-six different combinations of primers, summarized in Figure 18, have been tested so far. After the PCR is completed, the product is run on an agarose gel and imaged. Then the images are analyzed for differences between ddm1 and ddm1/rdr6. Any differences found in ddm1/rdr6, such as additional PCR products, size or intensity differences, are excised from the gel and cloned using TOPO cloning. The clones are then sequenced and BLASTed to TAIR10 transcript and TAIR10 intergenic sequences, separately.

Sets 1 through 5:

Sets 3, 11, 13, 15, 17, 20 and 21 each had differences between ddm1 and ddm1/rdr6. Sets 1 through 5 can be seen in Figure 19. All of the PCR product(s) from sets 3 and 5 were gel-excised and TOPO cloned. In set 3, the ddm1/rdr6 PCR product is much brighter than ddm1, meaning that there are
more transcripts that align to those primers in the double mutant. Set 5 was chosen to be cloned, because in both mutants the PCR product was not one 8431 bp sized product but two products, one around 2000 bp and another around 6000 bp. These two products may represent post-transcriptional processing.

Sets 6 through 11:

Out of sets 6 though 11, shown in Figure 20, set 11 was chosen for cloning. The top PCR product is unique to \textit{ddm1/rdr6} and was therefore cloned. Set 11 was the \textit{Athil6} LTR forward primer and \textit{Athila6} LTR reverse outer primer. Since there are two LTRs flanking \textit{Athila6}, there should only be a maximum of four PCR products. The fact that there are more, and one is unique to \textit{ddm1/rdr6} may indicate there was a difference post-transcriptional processing.

Sets 12 through 17:

Figure 21 displays the PCR products from sets 12 through 17. Sets 13, 15 and 17 had differences between the two mutants but only set 13 was cloned. Set 15 had an additional PCR product revealed after the PCR products were run out on the agarose gel longer but this additional PCR product was too faint to obtain clones. In set 17 \textit{ddm1/rdr6} had two PCR products, where as was only one PCR product in \textit{ddm1}. These PCR products were also too faint to clone. Set 13 has no additional PCR products in the double mutant but was substantially brighter that the single mutant. When this brighter \textit{ddm1/rdr6} PCR product was excised from the gel, it was found that it was actually two PCR products that were extremely close together and had to be cloned together.
Sets 18 through 21:

Sets 18 through 21 showed no significant difference until the agarose gel was run longer, as shown in Figure 22. Set 20 has a very faint additional PCR product towards the top of the *ddm1/rdr6* lane. Set 21 also has a very faint additional PCR product under the first PCR product in the *ddm1/rdr6* lane. Neither of these bands were TOPO cloned because of low amount PCR product present.

Sets 22 through 26:

Figure 23 shows the PCR products of sets 22 through 26. None of these sets had noteworthy difference in *ddm1/rdr6* when compared to *ddm1*. The agarose gel was even run longer and still no differences were found.

Restriction digest:

After the bands from sets 3, 5, 11 and 13 were TOPO cloned, the plasmids were digested using EcoRI to confirm that they had the inserted PCR product. Figure 24 shows the results of the restriction digests. Some of clones have multiple bands, because the PCR products they contain had multiple EcoRI restriction sites. The clones that had a PCR product include clones 2, 4 and 5 from set 3; 1, 2, 4 and 5 from set 5, the 2 kb fragment; 1, 2, 3, 4 and 5 from set 5, the 6 kb fragment; and 1, 2, 3, 4 and 5 from set 11. These clones will be sequenced to find out which *Athila6* elements comprise those different PCR products. This will determine which *Athila6* elements in *ddm1/rdr6* are producing transcripts different from those found in *ddm1*. If the *Athila6* elements are found to be unique to *ddm1/rdr6*, then this approach supports the alternative prediction that *ddm1/rdr6* is transcribing additional *Athila6* elements not found in *ddm1*. If they are found to be *Athila6* elements that are present in both mutants, then this approach would support the null hypothesis and that *ddm1* and *ddm1/rdr6* are transcribing the same *Athila6* elements.
References:


