The Thesis Committee for Thomas MacElliott Johnson
Certifies that this is the approved version of the following thesis:

The link between small RNA-directed gene regulation and heterosis in Arabidopsis allotetraploids

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Zengjian Jeffrey Chen

Hong Qiao

The link between small RNA-directed gene regulation and heterosis in Arabidopsis allotetraploids

by

Thomas MacElliott Johnson, B.S.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

December 2016

Abstract

The link between small RNA-directed gene regulation and heterosis in Arabidopsis allotetraploids

Thomas MacElliott Johnson, MA

The University of Texas at Austin, 2016

Supervisor: Zengjian Jeffrey Chen

Hybridization between different species of plants and animals commonly leads to superior levels of biomass, growth rate and stress resistance, because of a phenomenon known as hybrid vigor or heterosis. Despite the evolutionary significance and agricultural importance of this vigor, a true understanding of the genetic and molecular parameters remains unknown. Genome-wide changes in gene expression are well documented in hybrids, and could be under the control of a multitude of epigenetic factors. To examine the relative impact that small RNA (sRNA) directed cleavage of target transcripts has on *Arabidopsis* hybrid vigor, transcriptome, degradome and small RNA-sequence libraries were produced for five *Arabidopsis* lines: *A. thaliana* autotetraploids, *A. arenosa*, F1 resynthesized allotetraploids, F8 allotetraploids, and the natural allotetraploid *A. suecica*. Together these RNA libraries allowed for degradome analysis to be performed. This genome-wide approach allows for direct detection of sRNA-directed cleavage of target mRNAs without the need for prior predictions or sequence knowledge. The analysis output provides evidence of sRNA-target pairs that are biologically functioning *in vivo* at

a particular spatial or temporal capacity. When transcriptome or protein expression data is incorporated with degradome analysis, an overall model of expression and regulation patterns can be devised. During this investigation I detected novel and significantly differentiated sRNA-target interactions between the allotetraploids and their parents. Allelic expression frequencies within the allotetraploids allowed for the identification of homoeolog cleavage silencing modifications, along with the classification of cleavage events as being additive, exhibiting parental level dominance, or manifesting transgressive up- or down-regulation. I found that there is a significant amount of homoeolog expression bias that is being influenced by sRNA-mediated cleavage. Bias cleavage was found to not only preferentially eliminate homoeologous transcripts, but also function a buffering action to control threshold concentrations of target mRNA productions.

Results from this investigation indicates that sRNA-directed cleavage is contributing to the stimulation of heterosis within *Arabidopsis* allotetraploids. Integration of what we learned about this sRNA influence with previous reports on circadian rhythm, methylation and histone modification will better our knowledge of the mechanisms driving heterosis.

Table of Contents

List of Tables

List of Figures

Chapter 1: Background and Introduction:

Introduction

Plants utilize a large array of endogenous small RNAs (sRNA) to regulate the expression patterns of protein coding genes at the post-transcriptional level via targeted mRNA cleavage. These sRNAs play important roles during development, signal transduction, disease and stress resistance. The sRNA regulatory network can be viewed as an intersection between the transcriptome, non-coding RNAs, and the RNA degradome, which is the endogenous pool of RNA molecular decay. There is a plethora of information in this degradome pool, including a snapshot of mRNA transcripts that have undergone sRNA-directed cleavage. The onset of next generation sequencing has allowed for the identification of hundreds of sRNAs, however identifying their associated target genes has proven to be challenging. The identification of these sRNA-target pairs is fundamental to expanding our ability to model the mechanism sRNA-mediated regulation. One area of study that would benefit from a better understanding of sRNA regulation is hybrid vigor. Hybrid plants and animals often exhibit superior levels of biomass, stature, growth rate, and/or stress resistance. This phenomenon is known as heterosis, or hybrid vigor. A true understanding of the genetic and molecular parameters driving this vigor has remains unknown. To this end, genome-wide profiling of the transcriptome, degradome and sRNA network of *Arabidopsis* allotetraploids and their parents will be performed utilizing a technique simply known as degradome analysis. Previous published studies have demonstrated that within Arabidopsis allotetraploids many miRNAs and ta-siRNAs are non-additively expressed when compared to parental values (Ha et al., 2009). Furthermore, there is analysis from Ng et al. (2011), which not only highlighted the non-additive accumulation of miR163 in *Arabidopsis* allotetraploids,

they also demonstrated that the non-additivity is influenced by a dominant A. arenosa trans-acting repressor that is absent in A. thaliana and led to differential homoeolog target degradation (Ng et al., 2012). I hypothesize that homoeolog cleavage bias mediated by small RNAs is stimulating the observed heterosis within *Arabidopsis* allotetraploids. Exploring the extent that sRNA-directed target cleavage contributes to heterosis will lead to a better understanding of sRNA regulatory networks as well as homoeolog regulation bias and parental level dominance.

Allopolyploids and heterosis

Formed by cross-fertilization between different species, ecotypes or strains of plants or animals, hybrids often exhibit superior levels of biomass, stature, growth rate, and/or fertility. This phenomenon is known as heterosis, or hybrid vigor, and the molecular basis behind how hybrid offspring can perform greater than the sum of their parts remains to be fully elucidated. When a hybrid undergoes a chromosome doubling following formation it can lead to an allopolyploid, which can become genetically stable after dealing with the immediate shock from the combination of divergent paternal genomes. Rapid modifications at the genomic, epigenetic, metabolic, and transcript level occur to deal with this shock. Heritable changes that arise can allow for the establishment of novel phenotypes and speciation events. It is estimated that somewhere between 25% of plants and 10% of animals have undergone hybridization with related species, this includes some of more important agriculture crops, such as maize, cotton, and wheat (Wu et al. 2016).

Gene expression in hybrids

A litany of RNA-seq and microarray studies have been applied to characterize genome-wide expression changes within hybrids. A common association with the appearance of hybrid vigor is non-additive or differential gene expression. Here we define additivity as expression that is equal to the expected mid-parent value (MPV), which is the average between the parental values. The level of non-additive gene expression has been found to be correlated heterosis and as genetic distance between the parents increases, so does the amount of vigor. Acknowledging the importance of differential gene expression has done little to tease out the mechanisms that can fully explain heterosis. Gene expression becomes more complicated when the concept of homoeolog expression bias is explored. This concept follows the expression of the gene pairs within an allopolyploid that are derived from the two separate parental genomes. These homoeolog pairs can exhibit equal expression rates or be found to be bias towards one allele over another. Total gene pair expression may line up additively with the expected MPV, but beneath the surface bias expression of homoeolog alleles could be influencing observed vigor. Sometimes, the allopolyploids total expression level of a homoeolog pair will be most similar to that of only one of the parents, and is said to be under expression level dominance (Yoo et al. 2013).

Epigenetic changes in hybrids

Epigenetic changes that have been investigated, in relation to heterosis, include chromatin modification, DNA methylation, circadian rhythms, and small RNAs (sRNAs). These studies highlighted the extensive regulatory element reshuffling that occurs during hybridization. In A. thaliana intraspecific hybrids between ecotypes Ler and C24, the levels of all categories of DNA methylation were found to be increased, and also

enriched in transposable elements (Shen et al., 2012). Circadian rhythms have been found to be altered in *Arabidopsis* allotetraploids and shown to directly be influencing pathways known to be important in the establishment of hybrid vigor, including photosynthesis, metabolism and stress response (Wang et al., 2006; Ni et al., 2009).

Plant small RNAs (sRNAs)

Plants utilize a large array of endogenous sRNAs to regulate the expression patterns of protein coding genes that play important roles during development, signal transduction, disease and stress resistance. microRNAs (miRNAs) are a prominent class of sRNAs that are 20-24 nucleotides long and function through complementary binding of target mRNAs and the induction of mRNA cleavage. In animal species, miRNAs are found to also post-transcriptionally repress targets without cleavage events, but so far this is thought to be a much rarer phenomenon in plants (Brodersen et al., 2008). A second group of sRNAs that can direct mRNA-target cleavage are the small-interfering RNAs (siRNAs), which can be further classified into trans-acting siRNAs (ta-siRNAs), and natural antisense transcript siRNAs (nat-siRNAs).

The different sRNA types are categorized by their distinct characteristics and pathways of biogenesis. miRNAs and siRNAs are both processed by the Dicer RNAseIII family of enzymes, but miRNAs are derived from local stem-loop structures and siRNAs are derived from long double-stranded RNAs. Another similarity between miRNAs and siRNAs is that they are both are packaged into silencing complexes containing Argonaute proteins from where target repression is guided (Jones-Rhoades et al., 2006).

There is a high level of sequence conservation with some classes of sRNAs, especially within miRNAs, but divergent expression patterns are often seen between

species and in allopolyploids. These differences can arise from sequence changes within promoter regions or in trans-acting factors (Chen, 2013).

sRNA-target identification and degradome analysis methods

Traditional methods of target identification have focused on computational techniques that compare sequence similarity between small RNAs and the complementary binding site of their targets. Prediction algorithms are not sensitive enough to differentiate between false positives and real targets. This can be especially confounded when dealing with small RNAs that can regulate targets with limited complementarity (Rhoades et al., 2002).

A second method for target identification tries to take advantage of the evolutionary conserved nature of many of the more prominent small RNAs, such as miRNAs. This method assumes that miRNAs that are conserved between species would also have conserved targets, and thus a previously validated small RNA-target pair in one species may also be present in a related species. Issues arise with this hypothesis, because some targets are not conserved and additional non-conserved targets will be missed (Jones-Rhoades and Bartel, 2004). Moreover, this method completely ignores any species-specific miRNAs.

A third method of target identification involves the overexpression or repression of a particular small RNA. Following the general theory that the overexpression of a regulatory small RNA leads to the downregulation of its target mRNAs, microarray analysis can be used to compare genome wide expression values to WT. Repression of a small RNA can be achieved by expression of target mimics or small RNA 'sponges', which contain the complementary sequence of the assayed small RNA and acts as a

competitive inhibitor (Ebert et al., 2007). In the case of small RNA repression, microarray analysis would be used to identify possible targets with increased expression. However, these methods limit target identification to a single small RNA at a time and increase the chance that prospective targets are only differentially expressed because of indirect effects.

Each of the above traditional methods requires previous sequence knowledge and substantial validation of the small RNA-mediated cleavage. Validation can be performed with quantitative-PCR, but is most commonly achieved by 5' rapid amplification of cDNA ends (5'-RACE). 5'RACE can show the 5' position of the target mRNA, which represents the site of cleavage and generally aligns between the 10th and 11th nucleotide of the corresponding small RNA that is inducing cleavage (Addo-Quaye et al., 2008).

Continued advancement of high-throughput sequencing technology has recently allowed for genome-wide approaches to directly detect small RNA-mediated cleavage products without the need for prior predictions. Three independently developed procedures combining modified 5'-RACE with high-throughput deep sequencing and bioinformatics tools have delivered this capability. The three approaches are known respectively as degradome sequencing, parallel analysis of RNA ends (PARE) and genome-wide mapping of uncapped transcripts (GMUCT) (Addo-Quaye et al., 2008; German et al., 2008; Gregory et al., 2008).

These new methods are based on the fact that cleaved mRNA fragments and full-length mRNA have different structures on their 5' ends. Mature full-length mRNAs have modified 5' ends comprised of a 7-methylguanosine cap, whereas cleaved fragments that are the products of endonucleolytic cuts directed by small RNAs, possess a 5' monophosphate. The 5' monophosphate is ligation competent and allows for the selective cloning of these cleavage products.

Significance

Degradome sequencing methods are valuable tools for the identification and validation of sRNA-mediated target cleavage. In addition to providing evidence for which sRNA-target pairs are functional in vivo within a specific tissue or at a particular developmental stage, the relative abundance of cleaved target fragments can be directly compared bioinformatically with transcriptome, sRNA and protein expression data to formulate an overall model of expression patterns. Furthermore, sRNA-directed cleavage is only a fraction of the biological information found within a degradome library. Many phases of RNA metabolism can be isolated from a degradome library, including the maturation, quality control, and turnover. Almost all intermediates that constitute these RNA metabolism phases are found to have 5' monophosphates, and thus are free for selective cloning during degradome library construction. For example the intermediates found during the maturation of transfer RNAs (tRNAs) have been reported to have a ability to modify RNA-silencing pathways (Haussecker et al., 2010). In addition, degradome analysis allows for the identification of sequence motifs that are enriched upstream or downstream of small RNA-independent cleavage fragments. A recent study from Hou et al., (2014) began such an investigation and discovered 11 short motifs within the 3'UTR of enriched in Arabidopsis and rice PARE libraries. The biological relevance of these motifs is still to be elucidated, but the enrichment of sites for the PUF family of RNA-binding proteins is interesting because of their role in RNA localization and repression.

Additionally, the integrated nature of degradome analysis adds power to the significance of the experiment because the resulting output is more biologically relevant as if illustrates a snapshot of functioning regulation within samples at any time. This should allow for a better understanding of the layered molecular mechanisms occurring in

heterosis exhibiting allopolyploids. This study could influence a movement towards being able to modify gene networks in wildtype plant species for the purpose of reproducing the vigor phenotype. Furthermore, generating more productive agricultural crops or disease resistant crops could spur increasing food production to keep up with the growing world population.

Chapter 2: Degradome library construction and sRNA-target pair prediction

Rational

With the recent advances in high-throughput sequencing technology large sets of RNA sequence can be identified quickly and with more accuracy than ever before. In order to elucidate how sRNAs and their cleaved target mRNAs affect the heterosis phenotype observed in *Arabidopsis* allopolyploids, RNA-sequencings ability to identify levels of homoeologous RNA molecules is extremely important. Through the use of transcriptome, degradome and small RNA-sequence libraries produced from five Arabidopsis lines, *A. thaliana* autotetraploids, *A. arenosa*, F1 resynthesized allotetraploids, F8 allotetraploids, and the natural allotetraploid *A. suecica*, degradome analysis of heterosis was be performed. Degradome analysis is becoming more popular in plant sRNA studies and is now also being used in animal model organisms (Bracken et al., 2013; Park et al., 2013; Liu et al., 2014; Thatcher et al., 2014). However, this is the first study to look genome-wide at the direct contribution that sRNA-directed posttranscriptional cleavage has on growth vigor in interspecific hybrids.

Production of biological material and library prep

Arabidopsis allotetraploids (AlloF1) were resynthesized by using tetraploid *A. arenosa* (A.a, 2n=4x=32, Arabidopsis Biological Resource Center, CS3901) to pollinate autotetraploid *A. thaliana* Landsberg *erecta* (L*er*) (At4, 2n=4x=20, CS3900). Seven generations of self-pollination followed to obtain F8 allotetraploids (AlloF8, CS3895). The natural allotetraploid A. suecica (A.s, 2n=4x=26, CS22508) was also used for sampling. All plants were grown under 16 h of illumination per day with temperature conditions of 22°/20°C day/night. Rosette leaves were collected just prior to bolting (3-4

weeks old A. thaliana; 6-7 weeks old A. arenosa or allotetraploids plants) for small RNA, degradome and transcriptome analysis. Transcriptome and small RNA raw sequence data was derived from previous studies Shi et al., (2012) and Ha et al., (2009) respectively.

The degradome libraries were constructed following a slightly modified method described previously by German et al., (2009). Degradome products still contain a 3' polyA tail, so library construction begins with the enrichment of polyA RNA with oligo d(T)25 magnetic beads (New England BioLabs, Inc., Beverly, MA, USA), followed by the ligation of a 5' RNA adaptor. The ligated RNA was then subjected to reverse transcription using an oligo d(T)18 primer (New England BioLabs, Inc.) and second-strand synthesis using primers specific for the ligated 5' adaptor. The German et al., (2009) approach utilizes a MmeI restriction site within its 5' RNA adaptor to digest the resulting cDNA to produce equal-sized fragments comprised of a 20-21 nucleotide degradome tag attached to the 5' adaptor. This is because MmeI cleaves 20 nucleotides 3' of the recognition site located within the 5' adaptor. Contrastingly to this we sonicated the cDNA to produce fragments of 50-500 nucleotides in length. Addition of a 3'-double-strand DNA adaptor, gel purification and a final PCR amplification step leaves the degradome RNA libraries ready for high-throughput pyrosequencing,.

It is important to note that following the MmeI digestion method produces templates that at approximately 20 nucleotides of sequence may be a limiting factor for some organism systems. In the majority of cases reported, this has provided enough sequence to map uniquely to the particular genomic reference of study (German et al., 2008). For investigations needing additional sequence read length, such as those studying hybrids or organisms with highly repetitive genomes, the sonication method is preferred.

Sequencing analysis of *Arabidopsis* degradome and sRNA libraries

Transcriptome and degradome raw sequence data were first processed with cutadapt software (Martin, 2011) to remove the sequencing adaptor sequences. Customized perl scripts were then used to remove low quality reads, including those less than 20 nucleotides in length, those containing ambiguous bases and those with > 5% of sequence with Q values < 20. Additionally, within the degradome library, sequences containing the preliminarily ligated 5' RNA adaptor were selected for. The high quality sequences were then mapped to the *Arabidopsis thaliana* cDNA reference genome (The Arabidopsis Information Resource (TAIR) version 10; http://www.arabidopsis.org) using Bowtie2 (Langmead and Salzberg 2012) to define the coverage rate. Because of the lack of an A. arenosa reference genome, a higher amount of mismatches were allowed, in order to account for the sequence divergence between the *A. arenosa* and *A. thaliana* species.

The small RNA data from Ha et al., (2009) underwent adaptor trimming and quality control, followed by alignment to the full genome sequence of *A. thaliana* (TAIR 10). Sequences that matched the annotated areas of the mitochondria, chloroplast or genomic mRNAs, pretRNAs, rRNAs, snoRNAs, and snRNAs were excluded from analysis. The remaining 20-25 nucleotide sequences were considered to be small RNAs. The sRNAs that originated in *A. arenosa* and *A. suecica* but could not perfectly match the A. thaliana reference genome were excluded from most analyses because the complete *A. arenosa* genome sequence is unavailable. Thus, the analysis is limited to A. thaliana sRNAs.

The identified sRNAs were then selected and aligned against the reference cDNA of A. thaliana. sRNAs that aligned to the reverse strand of the reference cDNA were considered to have the potential to bind complementarily to the respective mRNA derived

from that alignment locus. Alignment settings were modified to allow for short insertions/deletions (indels) and up to six mismatches between the sRNA and the reference. This was necessary because plant sRNAs do not have to bind with perfect complementarity to elicit cleavage and will even bind with nucleotide bulges (Brousse et al., 2014).

Selection for degradome reads that contain the 5' RNA adaptor is important because the pertinent degradome information is found at the junction between the small RNA-directed cleaved fragment and the 5' adaptor. The exact junction site indicates the 5' terminus of the original cleavage fragment. If the fragment is derived from small RNA-directed cleavage then the 5' terminus commonly corresponds to the 10th nucleotide of a small RNA complementary sequence. To account for occasional positional heterogeneity seen between some mature miRNAs and their cleavage sites, miRNA isoform (isomiRs) production, as well as the more random cleavage of siRNAs, potential sRNA-mediated cleavage products were expanded to degradome fragments found to align to the 8th, 9th, 10th, 11th or 12th corresponding nucleotide of the proposed sRNA (Ahmed et al., 2014; Li et al., 2010).

sRNA-target cleavage prediction

Currently there are a variety of bioinformatics tools available that follow sets of pre-defined rules to predict targets of sRNAs in plants and included among these tools are a handful that specifically analyze degradome data (Addo-Quaye et al., 2008; German et al., 2008; Gregory et al., 2008). The stringency of these tools in dealing with how to score mismatches between the sRNA and its proposed target led me to develop my own workflow of custom perl, python and linux scripts. Together these gave me better control

over setting thresholds, parsing through data and dealing with homoeolog identity assignment. I also have the advantage of being able to re-analyze data from a multitude of intermediate steps along my computational workflow.

The first step in developing the library of potential small RNA-mediated cleavage products involved determining the intersect positions between the 5'-end of the degradome reads and the sRNAs using BEDtools (Quinlan and Hall, 2010). Intersections between the 8th-12th sRNA nucleotide were pooled together to represent a prospective cleavage event site. Pooling helped to account for the heterogeneity in the sequence of sRNAs, in particular, mature miRNAs. Typically miRNAs are annotated as having a single mature sequence, however, there are a growing number of recent studies observing isomiRs with several length and/or sequence variants. These variants were initially thought to be experimental artifacts that may have been generated by PCR errors, but are prevalent in sRNA sequencing libraries. Heterogeneity in miRNA length may not affect target selection, but the cleavage site on the target mRNA can shift by a nucleotide or two. An example of this cleavage site shift was reported by Jeong et al. (2013), they found that Arabidopsis miR858 is generated as two equally abundant isomiRs, miR858.1 and miR858.2. Utilizing PARE data, miR858.2, which is generated 1 nucleotide downstream of the annotated miR858.1 start site, mediated cleavage of the same set of MYB transcription factors, but the cleavage site was shifted 1 nucleotide downstream.

To minimize false negative sRNA-target interactions the threshold for alignment score was set at ≤ 4 between the respective sRNA and its target. There are sRNAs that can cause the cleavage of targets through poor binding complementarity with 5 or more mismatches. These include miR319, which has been shown function in leaf serration by targeting transcription factor TCP4 for degradation through 6 mismatch complementarity (Palatnik et al., 2003). Therefore, it is known that our data sets are missing some

functionally validated sRNA-target interactions. Yet, increasing the allowed number of mismatches when processing the data truly confounds the resulted output because of the short sequence reads being compared.

All degradome and transcriptome reads are normalized to reads per 10 million (RP10M) via a two-step normalization process that factors in the number of times an individual read aligns to the reference cDNA and the total number of mapped reads within its particular library. Firstly, each sequencing read is weighted by the number of times it aligns to the reference cDNA. For example, if a read aligns to five different positions across the entire reference cDNA then each separate alignment is scored as 0.2 (1 read / 5 total alignments). Secondly the weighted alignments are then normalized by mapped reads per 10 million. For example, if a sequencing library has 24,000,000 mapped reads in total than each read is normalized by dividing by 2.4 (24 million / 10 million = 2.4).

Biological functionality of cleavage events were further estimated by categorizing cleavage sites based on their abundance compared to the total degradome reads across the whole respective transcript. This is done by creating a degradome read distribution/density plot, which summarizes the alignment position of the 5' end of every degradome read across each locus of the reference sequence. These plots are also called target plots (t-plots) and they visually illustrate the abundance of potential small RNA-mediated cleavage products against random mRNA degradation products for each locus (German et al., 2008).

SNP database creation

In order to estimate allelic expression/cleavage in the allotetraploids, an *A. arenosa* single nucleotide polymorphism (SNP) database was built so that each transcriptome and degradome read could be given a SNP identity, which designated the homoeolog, A. thaliana or *A. arenosa*, form which the read derived. The SNP database was created using the sequence reads from the *A. arenosa* transcriptome and degradome libraries. A reconstituted A. thaliana reference was also built to account for sequence differences that arise from the A. thaliana ecotype used in this study, Landsberg *erecta*, being different from the Columbia ecotype that makes up the TAIR 10 reference sequences.

SNP database building began by labeling every nucleotide from the *A. arenosa* mapped reads by its alignment position, including locus and numbered position on that locus. Customized perl scripts were written to make sure that any insertions and deletions, identified during the bowtie2 alignment, were correctly labeled also. Nucleotide frequency tables were created for each nucleotide position where *A. arenosa* had mapped coverage. Then the *A. arenosa* and *A. thaliana* nucleotide distributions were compared and differential consensus nucleotides that were found to be statistically significantly (P-value 0.05) and had at least 4x coverage were registered as SNPs.

There are still concerns that allotetraploid SNP calling could be biased because of increased sequence divergence between homoeologous alleles and the degree of heterozygosity within these alleles. When assigning an allelic identity to the allotetraploid reads, two or more nucleotides of SNP coverage was needed. Reads with no coverage from the SNP database or with ambiguous SNP information were excluded.

Chapter 3: sRNA-directed cleavage event differences between *Arabidopsis* allotetraploids and their parents

Differences in total number of cleavage events

The degradome analysis found a total of 4795 cleavage events, directed by sRNAs, across all of the Arabidopsis samples. Pairwise comparisons between the allopolyploids and their parents were performed in order to identify the degree of differentially cleaved events. When cleavage events that are shared between at least two replicates per sample are pooled together and compared between hybrids and parents we find a decrease in the total number of cleavage during the selfing process from F1 to F8 allotetraploid (1104 to 741) (Table 3.1). This could be a product of stabilizing selection as the heterosis phenotype becomes fixed. There is also an observed increase in the pool of cleavage events between F8 allotetraploids and the natural *A. suecica* (741 to 934). It is possible that this is a result of the divergent evolutionary distance between them, because *A. suecica* was formed 12,000-300,000 years ago via a crossing of *A. thaliana* and *A. arenosa*. This observation runs parallel with the fact that there is a higher level of gene-expression divergence between *A. suecica* and the F1 and F8 allotetraploids, which again suggests many genes have diverged their expression overtime.

Table 3.1: Events of sRNA-directed target cleavage

	A. thaliana	***A. arenosa***	**F1 Allotetraploid**	**F8 Allotetraploid**	***A. suecica***
Cleavage Events Shared Between At Least Two Replicates	1262	1333	1104	741	934

Novel cleavage events in allotetraploids compared to parents

The data set was also explored for novel cleavage events that arose in the hybrids. Novel cleavage was inferred when both parental species had zero degradome reads for the particular sRNAs-target pair, yet the hybrids had an average greater than 5 reads per 10 million (RP10M) across its biological replicates. Each allopolyploid was found to have novel cleavage events (Table 3.2).

In the F1, F8 and *A. suecica* hybrids, only 5 (0.10%), 7 (0.15%) and 2 (0.04%) sRNA-directed cleavage events exhibited novel cleavage, respectively. Within F1 and F8 allotetraploids, these novel events may function as a buffering control on target transcript levels because the level of target gene expression is not significantly different from the MPV. However, the novel cleavage events established within *A. suecica* are both examples of a rare sRNA acting upon a gene exhibiting greater than 10-fold expression increase compared to *A. arenosa* and *A. thaliana*. The two interactions are ath-miR5654-5p and the siRNA ATsRR235816 targeting a transposable element gene (AT2G16140) and an uncharacterized gene (AT4G38932) respectively. These results indicate that novel cleavage events are most likely not driving heterosis and are either acting to buffer the genomic shock of early hybridization or are products of long periods of divergent evolution.

Table 3.2. Novel and homoeolog cleavage silencing modifications

Taxa	**Hybrid Novel**	***A. thaliana* Cleavage Silencing**	***A. arenosa* Cleavage Silencing**	***A. thaliana* and *A. arenosa* Cleavage Silencing**
AlloF1	5	10	34	3
AlloF8	7	36	26	9
A. suecica	2	27	40	15

Homoeolog cleavage silencing in allotetraploids compared to parents

If both parents had more than 5 RP10M, but the hybrids had zero counts for a respective homoeolog, this was deemed homoeolog cleavage silencing (Table 3.2). Homoeolog cleavage silencing, indicates a loss of sRNA-directed cleavage of one of both homoeolog target transcripts.

The F1 and F8 allotetraploids, along with *A. suecica* all display cases of cleavage silencing for both the A. thaliana and *A. arenosa* homoeologs within the respective hybrids. The events where both homoeologs, which were prominently cleaved in both parental species, are now no longer cleaved in the hybrids could be indicating the loss of sRNA-mediated regulation acting upon the respective mRNA target. F1, F8 and *A. suecica* have 3 (0.06%), 9 (0.19%), and 15 (0.31%) events respectively showing this phenomenon. Each of the three events found in the F1 allotetraploid are under the control of ath-miR5658. However, only one event had a correlated increase in target gene expression which is expected as cleavage was lost. This gene, AT1G18740, displayed over a 2-fold increase in expression when compared to the MPV. The function of the gene is currently unknown function but is a component of the chloroplast membrane. The other two events involve KNAT3 (KNOTTED1-LIKE HOMEOBOX GENE 3) loci and though they are no longer cleaved their expression levels were additive when compared to the MPV. Interestingly, even though total expression is additive there is homoeolog bias expression with *A. arenosa* derived transcripts contributing 61.88%. KNAT3 is known to be expressed in the vegetative meristem and functions to alter leaf morphology (Lincoln et al., 1994). Therefore, the bias towards *A. arenosa* homoeolog expression could contribute to the observed *A. arenosa*-like phenotype of the allotetraploid leaf.

Seven of the cleavage events that show a loss of A. thaliana in the F1 allotetraploids were found to share a mRNA expression pattern where there is significant

bias towards *A. arenosa* transcripts. With some events showing over a 4-fold *A. arenosa* expression bias over *A. thaliana* homeologs, the observed A. thaliana cleavage loss could be a product of the *A. thaliana* transcripts not reaching the threshold concentration where sRNA-directed cleavage works to regulate gene expression. Similarly, of the 36 events where *A. arenosa* cleavage is silenced, 24 (66.67%) display significantly higher *A. thaliana* transcript expression (p-value <0.05), including 16 that have greater than 2-fold expression bias. Interestingly, many of these genes play roles in abiotic and biotic stress, as well as protein folding (Table 3.3).

Table 3.3. F1 allotetraploid stress and defense response genes displaying loss of *A. arenosa* cleavage along with *A. thaliana* expression bias

Locus	**Gene**	**Function**	**sRNA**
AT1G03380	ATATG18G	Response to starvation	ath-miR849
AT1G06460	ATACD32.1	Response to heat, peroxisome maintenance, protein homeostasis	ath-miR5996
AT1G32540	LOL1	Defense, regulation of removal of superoxide radicals	ath-miR1886.3
AT1G66980	SNC4	Defense response to damage and pathogens	ath-miR857
AT3G14200	DnaJ-domain superfamily	Heat shock protein binding;	ath-miR4221
AT3G53110	LOS4	Response to heat and cold stress	ath-miR5648-3p
AT5G24810	ABC1 family	Unfolded protein response	ath-miR837-3p
AT5G37850	ATSOS4	Hyperosmotic salinity response	ath-miR858a

Of the 9 cleavage events within the F8 allotetraploids that depict the loss of cleavage to both homoeologs, all were active events in the F1's and 5 reappeared in *A. suecica*. These genes include universal stress protein ATUSP (AT3G53990), and a CC-NBS-LRR class disease response protein (AT5G35450). The other events that are no longer cleaved involve a nucleoside triphosphate hydrolase (AT3G28520) that responds to abscisic acid stimuli, and a set of methyltransferases regulated by ath-miR163. Following the trend seen in the F1 allotetraploids, the target transcripts that have lost cleavage predominantly are found to have up to a 10-fold decrease in mRNA expression levels. The only exception is ATUSP, which displayed an approximate 2-fold increase in expression (p-value <0.0001). Interestingly, ATUSP has been shown to perform a crucial role in stress tolerance, dealing with both oxidative stress and heat shock (Jung et al., 2015). This observation follows previously reported data indicating non-additivity relating to stress response genes, along with the increased stress resistance phenotype common in interspecific hybrids (Kim and Chen, 2011).

In contrast to the F1 allotetraploids, the F8's had more *A. thaliana* cleavage silencing (36 events) than *A. arenosa* (26 events). The F8 allotetraploids maintained 22 cleavage silencing patterns from the F1's, and among these events, 9 exhibited A. thaliana cleavage silencing while 13 showed *A. arenosa* cleavage silencing. There were a number of phenotypically interesting genes found to have undergone changes in homoeolog cleavages within F8 allotetraploids, including those that function in auxin response, photosynthesis, and nuclear transport (Table 3.4).

Table 3.4. F8 allotetraploid genes exhibiting homoeolog cleavage loss

Locus	Gene	Function	sRNA
AT1G30330	ARF6	Auxin response	ath-miR167a/b
AT1G56010	NAC1	Auxin mediated signaling	ath-miR164a/b
AT1G72830	ATHAP2C	Negative regulation of transcription	ath-miR169a
AT2G46100	NTF2	Nuclear import of Ran	ath-miR865-3p
AT2G30060	Pleckstrin homology domain superfamily	Ran GTPase binding	ath-miR831
AT3G61470	LHCA2	Photosynthesis	ATsRR166927
AT3G26060	ATPRX Q	Protection of the photosynthetic apparatus	ath-miR156

Additionally, 11 events present in the F1 were found to have changed the type of homoeolog cleavage silencing influencing respective sRNA-target pairs. Changes in silencing included 10 events previously identified as no longer cleaving the *A. arenosa* homoeolog in F1 allotetraploids, with 8 exhibiting *A. thaliana* homoeolog cleavage loss and 2 losing cleavage of both homoeologs in the F8 generation. The 2 events that had cleavage of neither homoeolog in F8 allotetraploids, were ath-miR2112-5p targeting of AT3G28520, a nucleoside-triphosphatase located in mitochondrion, and ath-miR5998a/b targeting AT1G80180, a MAPK kinases substrate found within stomatal complex. Both target genes had over 2-fold repressed expression of the target gene transcript (p-value >

0.05), which was driven by primarily decreasing the *A. arenosa* homoeolog expression and enhancing the biased expression of the *A. thaliana* homoeolog. This bias homoeolog expression pattern was already present within the F1 allopolyploids and resembles the differential expression observed between the parental species, *A. thaliana* and *A. arenosa* (p-value = 0.00025). This suggests that other transcriptional regulation mechanisms, most likely a dominant *A. thaliana trans* factor, are accounting for the expression variation between the homoeologs, which in turn decreases the need for sRNA-directed regulation of these genes.

For the 8 cleavage events that were found to have switched from *A. arenosa* to *A. thaliana* cleavage loss between the F1 and F8 allotetraploids, 7 had *A. thaliana* homoeolog was expressed at least 2-fold more (p-value < 0.0005) than the *A. arenosa* homoeolog and only the *A. thaliana* homoeolog was detected as being cleaved. These events include ath-miR773b targeting ATRBX1 (AT5G20570), ath-miR857 targeting SNC4 (AT1G66980), ath-miR858a targeting SOS4 (AT5G37850) and ath-miR169 targeting NF-YA10 (AT5G06510). In the F8 allotetraploids, the homoeolog expression changes in each of the 7 target genes, with *A. arenoas* homoeolog increasing and the *A. thaliana* homoeolog decreasing in expression. While the biased expression of *A. thaliana* transcripts over *A. arenosa* is maintained within F8, the degree of bias decreased. It is possible that because of the decrease in A. thaliana homoeolog expression, the *A. thaliana* transcript concentration is now below the threshold for sRNA-directed regulation and thus cleavage is lost in the F8. Conversely, the observed increase of *A. arenosa* homoeolog expression in the F8 allotetraploids compared the F1 could be the cause of *A. arenosa* homoeologs being found to be cleaved. However, the amount of *A. arenosa* homoeolog cleavage is low for this set of events and does not represent a statistically significant difference from the absence of cleavage seen in the F1

allotetraploids. Therefore, these events are likely to be experiencing similar regulation to the 2 previously discussed cleavage events of AT3G28520 and AT1G80180, where a dominant *A. thaliana trans* factor could be accounting for the variation in expression between homoeologs.

Homoeolog cleavage silencing within *A. suecica* was found to include 82 events in total with 27 involving lost cleavage of the *A. thaliana* homoeolog, 40 for the *A. arenosa* homoeolog and 15 events where both homoeologs were no longer cleaved when compared to the parental species. Half of these events were also found within the F8 allotetraploids, 24 displaying the same pattern of cleavage as in F8 and 17 having altered homoeolog cleavage loss. The maintained cleavage events include the regulation of ARF6, LHCA2 and ATPRX Q, with each gene showing an increase in expression. Phenotypically the increase in LHCA2 and ATPRX Q follow the trend previously reported that *Arabidopsis* allotetraploids have their *A. thaliana* homoeologs upregulated in genes involved in photosynthesis (Shi et al., 2012). In this study, we see a loss of A. thaliana cleavage for these two photosynthesis related genes, which in turn have over 2-fold increases in expression bias towards the *A. thaliana* homoeolog (p-value <0.05). *A. suecica* also was also found to silence the cleavage of *A. thaliana* homoeologs of GUN5 (AT5G13630) and NYC1 (AT4G13250) which both function in chlorophyll biosynthesis. In relation to auxin response pathways, in addition to ARF6 cleavage, A. suecica has lost the cleavage of ARF3 by ta-siR2141. Together, these auxin response genes contribute to organ identity, particularly leaf morphology. ARF3 is essential for leaf polarity specification and functions, along with ARF6, in the determinacy of the floral meristem and subsequently flower maturation (Li et al., 2016). These regulation changes could be contributing to the later flowering time observed in *Arabidopsis* allotetraploids.

Transmission of cleavage pattern through subsequent generations

To test whether the pattern of differential cleavage between the parent species changed in resynthesized F1 allotetraploids and persisted over generations, we compared sRNA-directed cleavage in F1, F8 and the natural allotetraploid *A. suecica.* In order to focus in on cleavage events that were more likely to be biologically relevant, sRNA-target pairs that contained a sRNA expressed greater than 100 RP10M in atleast one of the sample species (*A. thaliana*, *A. arenosa*, F1, F8 or *A. suecica*) were selected for. Out of the 4795 total cleavage events across all samples, 1618 were mediated by sRNAs with expression levels that met this requirement (Table 3.5). In the following analysis, for a cleavage event to be determined as significant it had to display at least a 2-fold increase/decrease in the number of respective cleaved targets (p-value ≤ 0.1).

Table 3.5. Differential cleavage of targets between the parents or in response to genome merger, selfing or wild evolution for sRNAs with expression greater than 100 RP10M (p-value ≤ 0.1).

	Cleavage Pattern	**Number of Events**
Parents – Aa vs At	Aa = At	1449
	Aa < At	76
	Aa > At	93
F1 vs MPV	F1 = MPV	1103
	F1 < MPV	234
	F1 > MPV	281
F8 vs F1	F8 = F1	1530
	F8 < F1	61
	F8 > F1	27
As vs F8	As = F8	1527
	As < F8	44
	As > F8	47

Between the two parental species, *A. arenosa* and A. thaliana, 1029 (63.60%) events were differentially cleaved at a 2-fold level. There were 76 (4.70%) events that are statistically significantly decreased, while 93 (5.75%) display significant increased cleavage in *A. arenosa* (Table 3.5). Within the events that are significantly different between the parents, when cleavage depicts a pattern of *A. arenosa* < *A. thaliana*, 30 showed a non-additive decrease of cleavage in the F1 allotetraploids, while 11 were found to have a non-additive increase of cleavage in F1. In contrast, when *A. arenosa* > *A. thaliana* differential cleavage was found between the parents, 40 events show a nonadditive decrease of cleavage in the F1 allotetraploids and 16 events had a nonadditive increase of cleavage in the F1 allotetraploids. Therefore, out of the 169 differentially cleaved events between the parents, 97 (60.63%) are non-additively cleaved within the F1 allopolyploid. This follows a similar trend where many non-additively expressed proteins and mRNA transcripts in F1 allotetraploids are also differentially expressed between the parents (Ng et al., 2012). The hypothesis is that genes that display the greatest divergence between the parents contribute proportionally more to the non-additivity found in allotetraploid progenitors.

When homoeolog cleavage bias is found within F1 allotetraploids for the differentially cleaved events between the parents, the bias mostly followed that parental pattern. For example for the 41 *A. arenosa* (Aa) < *A. thaliana* (At) parent events, 19 (46.34%) maintained this pattern F1(Aa) < F1(At) with the homoeologs within F1 and only 1 (2.44%) swapped homoeolog cleavage bias to F1(Aa) > F1 (At). Moreover, with the 56 Aa > At differential parental events 26 (42.86%) F1(Aa) > F1(At) and 7 (12.50%) F1(Aa) < F1(At) events showed bias. In total, approximately half of the cleavage events that are differentially cleaved between the parents also exhibit non-additive cleavage in F1 allotetraploids, driven by homoeolog cleavage bias that maintains the parental pattern

(53, 54.64%). An additionally interesting trend from this data is that for the 56 events which are differentially expressed between the parents in the direction of increased *A. arenosa* cleavage, while also found to be non-additive in F1 hybrids, the occurrence of homoeologous bias cleavage is more often found exhibiting *A. arenosa* dominant cleavage. This indicates that parental *A. arenosa* dominance in cleavage patterning can be transferred to the F1 allotetraploid homoeologs, which leads to the observed non-additive change in total cleavage between the F1 and MPV.

Further investigation of the F1 allotetraploid compared to the expected MPV found 1054 (65.14%) of the sRNA-directed cleavage events exhibited at least a 2-fold change difference. Of these events, 515 were statistically significant (p-value ≤ 0.1), with 234 (14.46%) representing non-additive repression of cleavage and 281 (17.38%) showing non-additive enhancement of cleavage. This left 1073 (66.32%) of the cleavage events being additively regulated. Interestingly, for events that displayed at least a 2-fold increase in cleavage (F1 > MPV) the majority (85.93%) were found to be statistically significant. This could mean that the enhanced sRNA-directed target cleavage events play an important role maintaining genomic stability in response to the genomic shock occurring during the early stages of allopolyploid formation (Ha et al., 2009).

To investigate how the regulation of sRNA-directed cleavage changes during the selfing of F1 allotetraploid to produce a more stable heterosis phenotype, F8 allotetraploids were compared to the F1 generation. It was found that 845 (52.22%) events within the F8 allotetraploids displayed 2-fold changes in cleavage. Of these events, 88 were statistically significant (p-value ≤ 0.1), with 61 (3.77%) being non-additively repressed, 27 (1.67%) non-additively enhanced, while 1530 (94.56%) events were equally active compared to the F1. The proportion of cleavage events that were significantly different between F8 and F1 allotetraploids is much less than we observed

when comparing F1 directly to its parents and the expected MPV. Moreover, when *A. suecica* was used to represent fixed heterosis, which it gained through wild evolution, the comparison between F8 allotetraploids and the natural *A. suecica* found a smaller number of significantly different events than F1. Overall, within *A. suecica*, 804 (49.69%) sRNA-directed cleavage events had 2-fold changes in cleavage, with 44 (2.72%) decreased and 47 (2.90%) increased at statistically significant levels compared to F8. Therefore the majority of events, 1527 (94.38%) were equally cleaved between the F8 allotetraploids and *A. suecica*. Together, the F1, F8 and *A. suecica* data suggests that because of the genome shock that occurs in the newly formed resynthesized interspecific F1 hybrid, the extent of non-additive differential cleavage is rapidly established in the F1. Furthermore, the differential cleavage from F1 to F8 and onto *A. suecica* acts to help stabilize to heterosis phenotype, as is observed. A similar trend has been reported in the expression ratio changes between *A. thaliana* and *A. arenosa* homeologs during this allotetraploid evolution (Shi et al., 2012). In addition, F8 versus *A. suecica* cleavage differences could be the result of adaptation and diversification undergone by *A. suecica* during its wild evolution, which derived some 12,000 to 300,000 years ago (Jakobsson et al., 2006).

Correlation between cleavage and transcript expression

Conventional thought is that an observed increase in target gene cleavage by a particular sRNA should be correlated with the decrease in expression of the target mRNA transcript. Likewise, a decrease in target cleavage is expected to follow with an increase in transcript expression. However, as shown many times during this investigation, this canonical relationship is not the prominent observation. For example, if we select the significant non-additively cleaved events within F1 and graph the log2(F1/MPV) for gene

expression against log2(F1/MPV) for cleavage, the resulting plot shows two clusters, which represent increased cleavage and decreased cleavage in F1 allotetraploids respectively (Figure 3.1A). When F1 decreased cleavage and increased cleavage are plotted separately the resulting trendlines show no correlation with R2 = 0.0054 and R2 = 0.0001 respectively (Figure 3.1B and C).

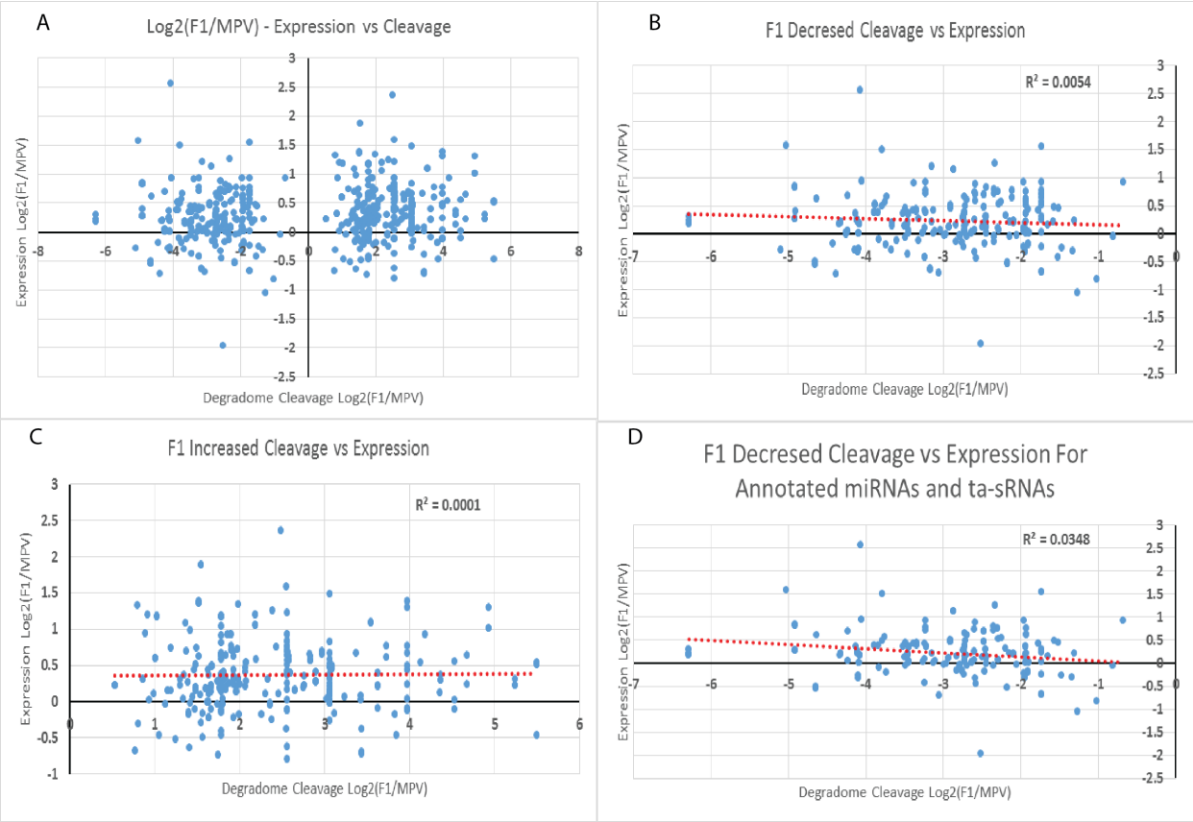

Figure 3.1. Correlation between cleavage and expression

(**A-D**) Log2 fold change values for cleavage and expression of sRNA-target pairs; comparing F1 allotetraploids to MPV. Significant cleavage events that are non-additive in F1 (p-value 0.1) were selected for correlation analysis. (**A**) All nonadditive cleavage events were plotted. (**B**) Events with decreased cleavage in F1 compared to MPV. (**C**) Events with increased cleavage in F1 compared to MPV. (**D**) Annotated miRNA and ta-siRNA-directed cleavage events exhibiting decreased cleavage in F1 compared to MPV.

Interestingly, if F1 vs MPV cleavage events mediated only by annotated miRNAs and ta-siRNAs are selected for, the plot of non-additive decreased cleavage within the F1 allotetraploids has log-fold changes in cleavage and target mRNA levels that are negatively correlated (r=-0.187, p-value = 0.0108, R2 = 0.0348) (Figure 3.1D). Though the correlation was low, it is still significant. Interestingly, for the cleavage events that increase non-additively in F1, there is no correlation with the target mRNA levels. This trend is also displayed in F8 allotetraploids and *A. suecica*. Together with the uncorrelated cleavage-expression pairs, this provides evidence that there must be other mechanisms, such as transcriptional regulation of miRNA and ta-siRNA targets, to fully account for target expression variations.

Differential cleavage events contributing to allopolyploid heterosis

With 866 sRNA-directed cleavage events being found to be differentially active at significant levels between all samples in this study, it can be said that sRNAs are playing a role in interspecific gene expression observed in *Arabidopsis* allotetraploids. The extent of which events have stronger influence over homoeolog variations and the heterosis phenotype within the allotetraploid remains to be fully elucidated. This study though, can supply candidate sRNA-target pairs that may be important drivers of heterosis.

To illustrate the differentially cleaved sRNA-target interactions, volcano plots were produced (Figure 3.2). The cleavage data used to produce these plots are the total number of cleaved fragments for a particular event. Therefore, the homoeolog bias within the allotetraploids is not being taken into account. The significant cleavage events are the

blue dots, while red dots represent cleavage events that did not meet the P-value cut off of 0.05 or had a log2[fold change] value between -1 and 1. In these instances the fold change is calculated by (hybrid cleavage) / (MPV cleavage). This means that because the majority of the significant cleavage events are off to the left of the volcano plot, then those events represent depressed cleavage values for the hybrids compared to the MPV. The number of significant events is the greatest within F1 allotetraploids as heterosis is rapidly established. As heterosis stabilizes through generations to F8 and onto *A. suecica* through wild evolution, the number of significant differences is less. This trend is visualized in the Figure 3.3 heatmap. Not only is differential cleavage less prevalent after heterosis is established in the F1 generation, most of the significant differential events within F8 and *A. suecica* are examples of decreased cleavage.

Most sRNA binding sites are highly conserved between *A. thaliana* and *A. arenosa* and because of this, we can more confidently look at bias homoeolog cleavage events (Rhoades, et al., 2002). There was one cleavage event found that was found to have sequence variability at the site of sRNA binding. This event was the targeting of a pentatricopeptide repeat (PPR) superfamily protein (AT1G12775) by the trans-acting siRNA TAS2. It was discovered that there is a SNP in A. arenosa two nucleotides downstream from the actual site of cleavage (Figure 3.4). Interestingly, this SNP increased the complementarity between TAS2 and the target, which also led to 100% *A. arenosa* bias cleavage. This gene also showed cleavage directed by TAS1c, which had a binding site upstream of TAS2. TAS1c had bias cleavage of *A. thaliana* homeologs. This secondary site of cleavage may be *A. thaliana* bias because TAS2 was cleaving only *A. arenosa* transcripts and thus there would be an excess of *A. thaliana* transcripts available.

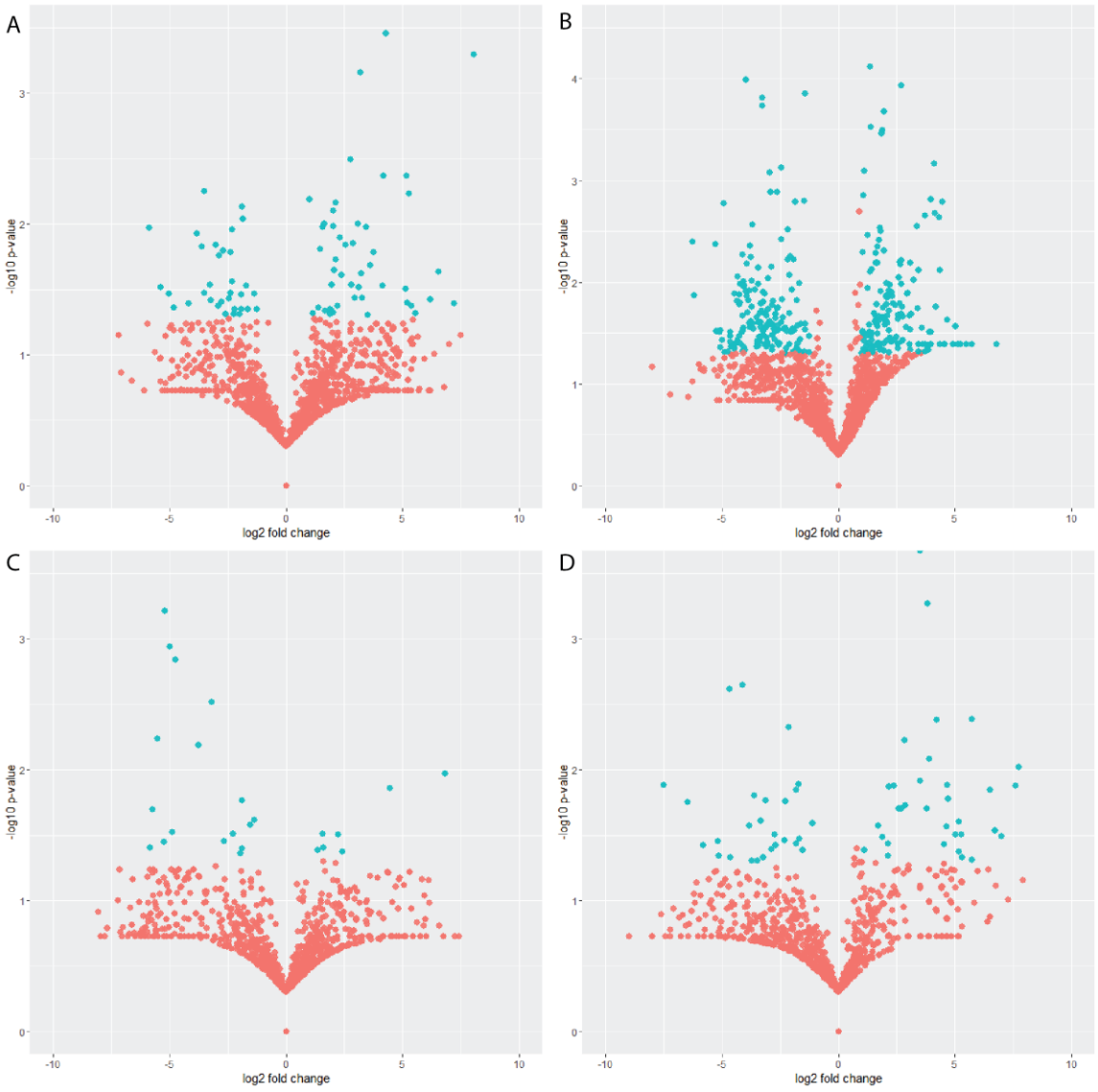

Figure 3.2. Differentially cleaved sRNA-target interactions

(**A-D**) Log2(Fold Change) of cleavage (x-axis) plotted against –Log10(p-value). (**A**) Differential events between the parental species *A .arenosa* and *A. thaliana (*Aa/At). (**B**) Non-additive cleavage events in F1 allotetraploids (F1/MPV). (**C**) Differential cleavage between F8 and F1 allotetraploids (F8/F1). (**D**) Differential cleavage between natural allotetraploid *A. suecica* and the resynthesized F8 allotetraploid (As/F8). Blue data points indicate significant events (p-value <0.05; fold change > 2).

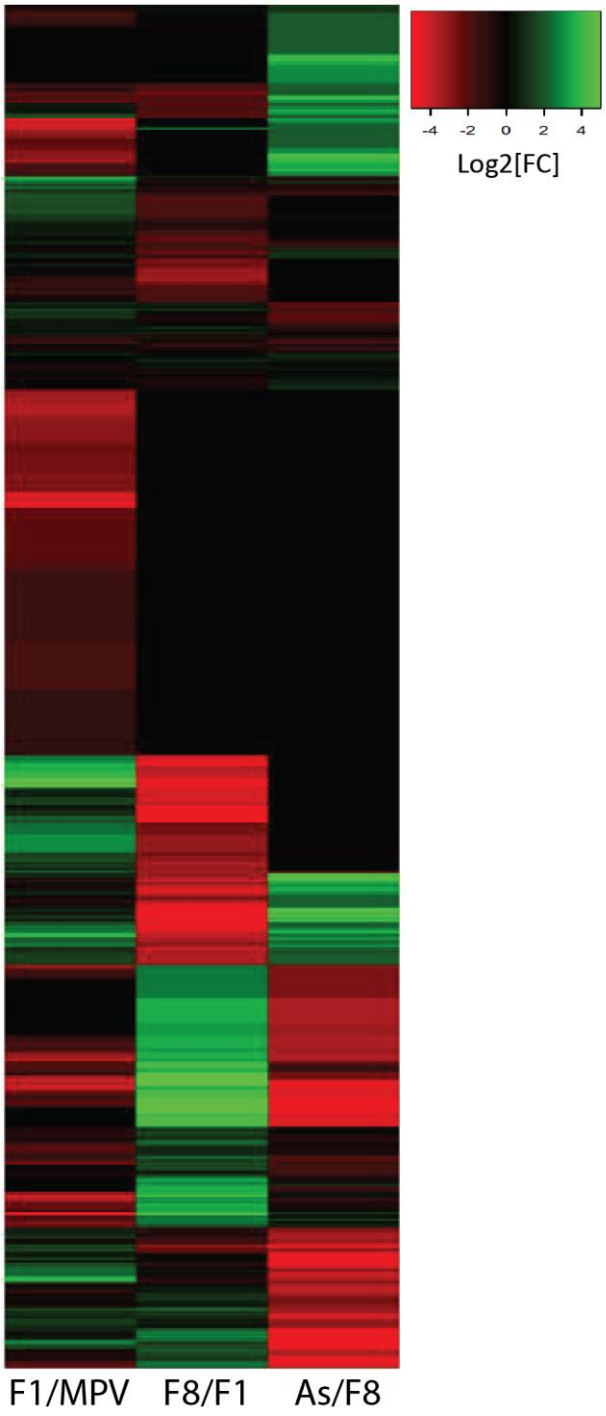

Figure 3.3 Log2(FC) of cleavage between F1/MPV, F8/F1 and *A. suecica*/F8.

Ward's method of hierarchical clustering of differential cleavage variations in F1, F8 and *A. suecica* leaves.

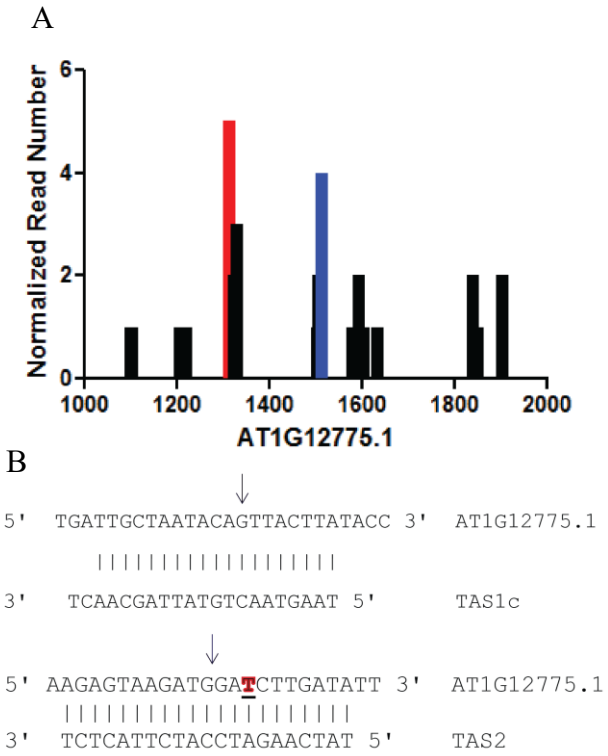

Figure 3.4. AT1G12775 targeted cleavage by TAS1c and TAS2

(**A-B**) AT1G12775 encodes a PPR gene which is cleaved by two trans-acting siRANs. TAS1c binds at position 1302 on the mRNA, while TAS2 binds at 1502. (**A**) Histogram density plot for degradome reads that align to the AT1G12775 transcript. Each column represents the 5' position of a degradome library read. The black lines are not sRNA-directed cleavage events. The red histogram column indicates *A. thaliana* homeolog cleavage, while the blue histogram column indicates *A. arenosa* homeolog cleavage. (**B**) Diagram showing the sRNA bindings sites and point of cleavage indicated by the arrow. The SNP position at the TAS2 bindings site is indicated by a red highlighted "T".

The allotetraploids also exhibited differential cleavage of some genes that function in leaf development. These are interesting cleavage events because of the allotetraploid leaf phenotype is very different when compared to *A. thaliana* and is considered to be *A. arenosa*-like. These genes include SPL2, SPL15, SPL10, ARF2, RAP2.7, AP2, and LCD1. The SPL's, which are targeted by ath-miR156 function in flowering and leaf development, while also acting as a control of the vegetative phase change.

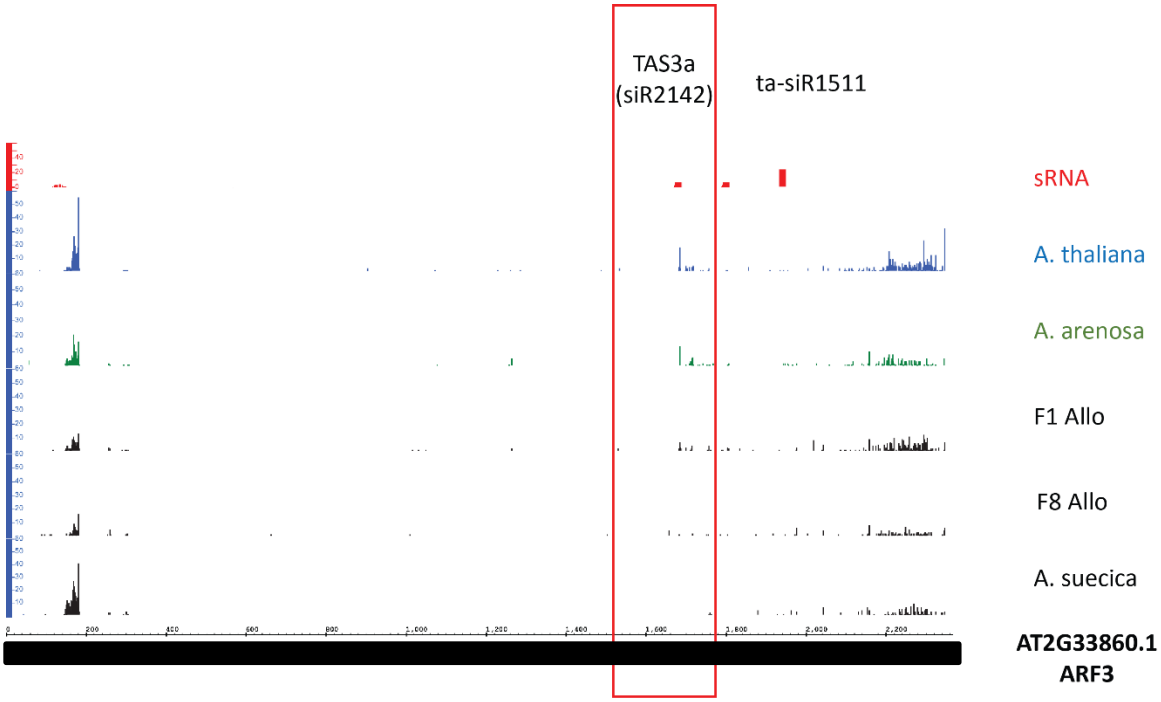

Figure 3.5. Distribution of degradome fragments along ARF3 transcript and pattern of lost cleavage.

The ARF3 gene is regulated by the trans-acting siRNA TAS3a and cleavage of the gene decreases in the succession from F1 to F8 to *A. suecica.* This pattern of decreasing cleavage is common for a majority of target genes that are regulated by ta-siRNAs. This result indicates that as heterosis stabilizes and is fixed there is a genome-wide decrease in ta-siRNA activity. In contrast ta-siRNAs are active in both the *A. thaliana* and *A. arenosa* parents. These observations are supported by the sRNA sequencing library, which shows a decreased expression of ta-siRNAs in the allotetraploids.

Future

The integrated and genome-wide nature of degradome analysis has allowed for a variety of differentially regulated sRNA-target pairs to be identified. Additionally, the degradome shows allelic cleavage bias better than any other experimental method. This study could influence a movement towards being able to modify gene networks in wildtype plant species for the purpose of reproducing the vigor phenotype.

Future work would need to be done with transgenic plants in order to confirm how sets of sRNA-target pairs are influencing heterosis. Candidate sRNA-targets pairs from the degradome analysis and validation steps that are believed to be playing the most significant role in influencing the heterosis phenotype will need to be manipulated within the tetraploid *A. thaliana* for the purpose of reproducing the observed hybrid vigor. The molecular manipulation techniques that could be utilized include: artificial microRNAs (amiRNA), short tandem target mimic (STTM) constructs, overexpression vectors of target mRNA, and gene editing of target mRNA via RNA-based CRISPR/Cas9 technology. Together these techniques will work to upregulate and/or downregulate the

sRNA and its target mRNA. If the sRNA-target pair's significant non-additivitly regulated profile determined during degradome analysis is truly an influencer of heterosis, then we should see some hybrid vigor phenotype during manipulation.

Furthermore, this could lead to the ability to generate more productive agricultural crops or disease resistant crops could spur increasing food production to keep up with the growing world population.

References

Addo-Quaye C., Eshoo T.W., Bartel D.P. and Axtell M.J. (2008), Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Current Biol, 18, 758–762.

Ahmed, F., Senthil-Kumar, M., Lee, S., Dai, X., Mysore, K. S., & Zhao, P. X. (2014). Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in Arabidopsis thaliana. RNA Biology, 11(11), 1414–1429.

An FM, Chan MT, (2012), Transcriptome-Wide Characterization of miRNA-Directed and Non-miRNA-Directed Endonucleolytic Cleavage Using Degradome Analysis Under Low Ambient Temperature in Phalaenopsis aphrodite subsp. Formosana. Plant Cell Physiol. 53(10): 1737-1750.

Axtell MJ, Bartel DP. (2005), Antiquity of microRNAs and their targets in land plants. Plant Cell, Vol 17, p1658–1673.

Birchler JA. (2015) The genetic basis of hybrid vigour. Nature Plants 1, 1-2.

Bracken CP et al. (2011), Global analysis of the mammalian RNA degradome reveals widespread miRNA-dependent and miRNA-independent endonucleolytic cleavage. Nucleic Acids Research, Vol 39, No 12, p5658-5668

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O. (2008), Widespread translational inhibition by plant miRNAs and siRNAs. Science, 320, 1185–1190

Brousse, C., Liu, Q., Beauclair, L., Deremetz, A., Axtell, M. J., & Bouché, N. (2014). A non-canonical plant microRNA target site. Nucleic Acids Research, 42(8), 5270–5279.

Chen C., Ridzon D.A., Broomer A.J., Zhou Z., Lee D.H., Nguyen J.T., et al. (2005). Real-time quantification of microRNAs by stem–loop RT–PCR. Nucleic Acids Res., 33:e179.

Chen, Z. J. (2013) Genomic and epigenetic insights into the molecular bases of heterosis. Nature Reviews Genetics 14, 471-482

Cheng-Yu Hou et al. (2014), Beyond cleaved small RNA targets: unraveling the complexity of plant RNA degradome data. BMC Genomics 15:15

Clough SJ, Bent AF (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16(6):735-43.

Doyle JJ, Flagel L, Paterson AH, Rapp RA, Soltis DE, Soltis PS et al. (2008). Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42: 443–461.

Ebert, M.S., Neilson, J.R. & Sharp, P.A. (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nat. Methods 4, 721–726

German M. A., Pillay M., Jeong D.H.H., Hetawal A., Luo S., Janardhanan P., Kannan V., Rymarquis L.A., Nobuta K., German R. et al. (2008), Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat. Biotechnol., 26, 941–946.

German MA, Luo S, Schroth G, Meyers BC, Green PJ (2009) Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat Protoc 4: 356-362. doi:10.1038/nprot.2009.8. PubMed: 19247285.

Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, Millar AH, Ecker JR (2008), A Link between RNA Metabolism and Silencing Affecting Arabidopsis Development. Developmental Cell 14, 854–866

Grover CE, Gallagher JP, Szadkowski E, Yoo M-J, Flagel LE, Wendel JF (2012). Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol 196: 966–971.

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ, (2009). Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. Proc Natl Acad Sci USA, 106:17835-17840.

Haussecker,D., Huang,Y., Lau,A., Parameswaran,P., Fire,A.Z. and Kay,M.A. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA, 16, 673–695

Jackson S, Chen ZJ (2010). Genomic and expression plasticity of polyploidy. Curr Opin Plant Biol 13: 153–159.

Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C, Nordborg M. (2006), A unique recent origin of the allotetraploid species Arabidopsis suecica: Evidence from nuclear DNA markers. Mol Biol Evol. 23(6):1217-31

Jones-Rhoades, M.W. & Bartel, D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell 14, 787–799

Jones-Rhoades M.W., Bartel D.P., Bartel B. (2006), MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol., 57, 19–53.

June Hyun Park et al. (2013), Degradome sequencing reveals an endogenous microRNA target in C. elegans FEBS Letters 587, 964-969

Jung, Y. J., Melencion, S. M. B., Lee, E. S., Park, J. H., Alinapon, C. V., Oh, H. T., Yun, D-J., Chi, Y. H., Lee, S. Y. (2015). Universal Stress Protein Exhibits a Redox-

Dependent Chaperone Function in Arabidopsis and Enhances Plant Tolerance to Heat Shock and Oxidative Stress. Frontiers in Plant Science, 6, 1141.

Kim, E.-D., & Chen, Z. J. (2011). Unstable Transcripts in Arabidopsis Allotetraploids Are Associated with Nonadditive Gene Expression in Response to Abiotic and Biotic Stresses. PLoS ONE, 6(8), e24251.

Langmead B, Salzberg S. (2012). Fast gapped-read alignment with Bowtie 2.Nature Methods, 9:357-359.

Li T, Ma L, Geng Y, Hao C, Chen X, Zhang X. (2015), Small RNA and Degradome Sequencing Reveal Complex Roles of miRNAs and Their Targets in Developing Wheat Grains. PLoS One. 1;10(10)

Li Y.F., Zheng Y., Addo-Quaye C., Zhang L., Saini A., Jagadeeswaran G., Axtell M.J., Zhang W., Sunkar R. (2010), Transcriptome-wide identification of microRNA targets in rice. The Plant Journal, 62, 742–759.

Li SB., Xie ZZ., Hu CG., Zhang JZ. (2016), A Review of Auxin Response Factors (ARFs) in Plants. Frontiers in Plant Science, 7:47

Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K., & Hake, S. (1994). A knotted1-like homeobox gene in Arabidopsis is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. The Plant Cell, 6(12), 1859–1876.

Liu H, Qin C, Chen Z, Zuo T, Yang X, Zhou H, Xu M, Cao S, Shen Y, Lin H, He X, Zhang Y, Li L, Ding H, Lübberstedt T, Zhang Z1, Pan G. (2014), Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. BMC Genomics. 14;15:25

Llave C, Xie Z, Kasschau KD, Carrington JC. (2002), Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science 297: 2053–2056

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res., 18: 1509-1517

Marguerat S, Bähler J. (2010), RNA-seq: from technology to biology. Cell Mol Life Sci. 67(4):569-79.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, 17,(1):10-12,

Ni Z, Kim E-D, Ha M, et al. Altered circadian rhythms regulate growth vigor in hybrids and allopolyploids. Nature. 2009;457(7227):327-331.

Ng DW, Lu J, Chen ZJ. (2012), Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. Curr Opin Plant Biol. 15(2):154-61.

Ng DW, Zhang C, Miller M, Palmer G, Whiteley M, Tholl D, Chen ZJ. (2011), cis- and trans-Regulation of miR163 and target genes confers natural variation of secondary metabolites in two Arabidopsis species and their allopolyploids. Plant Cell, 23 pp. 1729–1740

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841–842.

Ramarao, SD (2012) Bioinformatic analysis of degradome data from African oil palm (Elaeis guineensis Jacq.) inflorescence (Master's thesis). Retrieved from http://studentsrepo.um.edu.my/id/eprint/3718

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge,, C.B., Bartel, D.P. (2002), Prediction of plant microRNA targets. Cell 110(4), 513–520

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control of leaf morphogenesis by microRNAs. Nature 425: 257–263

Puchta H, Fauser F. (2014)Synthetic nucleases for genome engineering in plants: prospects for a bright future. THE PLANT JOURNAL. 78(5), p727–741

Salomé PA, McClung RC. (2005). PSEUDO-RESPONSE REGULATOR 7 and 9 Are Partially Redundant Genes Essential for the Temperature Responsiveness of the Arabidopsis Circadian Clock. Plant Cell. 17(3): 791–803.

Shi X, Ng DW-K, Zhang C, Comai L, Ye L, Chen ZJ. (2012). Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. Nature Communications, 3:950.

Shen H, He H, Li J, et al. (2012). Genome-Wide Analysis of DNA Methylation and Gene Expression Changes in Two Arabidopsis Ecotypes and Their Reciprocal Hybrids. The Plant Cell. 24(3):875-892.

Tang G, Yan J, Gu Y, Qiao M, Fan R, Mao Y, Tang X. (2012) Construction of short tandem target mimic (STTM) to block the functions of plant and animal microRNAs. Methods 58(2):118-25.

Tang G. and Tang X. (2013) Short tandem target mimic: A long journey to the engineered molecular landmine for selective destruction/blockage of microRNAs in plants and animals. Journal of Genetics and Genomics. 40 (6): 291–296

Teotia S, Singh D, Tang X, Tang G. (2016), Essential RNA-Based Technologies and Their Applications in Plant Functional Genomics. Trends in Biotechnology. 34(2), p106–123,

Thatcher SR, Burd S, Wright C, Lers A, Green PJ, (2014), Differential expression of miRNAs and their target genes in senescing leaves and siliques: insights from deep sequencing of small RNAs and cleaved target RNAs. Plant, Cell & Environment. 38(1): 188–200.

Vaucheret H. (2006), Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes Dev., 20, 759–771.