Epigenetic Responses of Arabidopsis to Abiotic Stress

Suzanne Laliberte

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Plant Pathology, Physiology and Weed Science

David C. Haak, Co-Chair
James Westwood, Co-Chair
Jacob Barney
Shawn Askew
Lenwood Heath

February 17, 2023
Blacksburg, Virginia

Keywords: epigenetics, weeds, bioinformatics, RNA Seq, differential expression analysis, whole genome bisulfite sequencing, data mining, k-means clustering, decision tree, random forest, multi-'omics

Epigenetic Responses of Arabidopsis to Abiotic Stress

Suzanne Laliberte

ABSTRACT

Weed resistance to control measures, particularly herbicides, is a growing problem in agriculture. In the case of herbicides, resistance is sometimes connected to genetic changes that directly affect the target site of the herbicide. Other cases are less straightforward where resistance arises without such a clear-cut mechanism. Understanding the genetic and gene regulatory mechanisms that may lead to the rapid evolution of resistance in weedy species is critical to securing our food supply. To study this phenomenon, we exposed young Arabidopsis plants to sublethal levels of one of four weed management stressors, glyphosate herbicide, trifloxysulfuron herbicide, mechanical clipping, and shading. To evaluate responses to these stressors we collected data on gene expression and regulation via epigenetic modification (methylation) and small RNA (sRNA). For all of the treatments except shade, the stress was limited in duration, and the plants were allowed to recover until flowering, to identify changes that persist to reproduction. At flowering, DNA for methylation bisulfite sequencing, RNA, and sRNA were extracted from newly formed rosette leaf tissue. Analyzing the individual datasets revealed many differential responses when compared to the untreated control for gene expression, methylation, and sRNA expression. All three measures showed increases in differential abundance that were unique to each stressor, with very little overlap between stressors. Herbicide treatments tended to exhibit the largest number of significant differential responses, with glyphosate treatment most often associated with the greatest differences and contributing to overlap. To evaluate how large datasets from methylation, gene expression, and sRNA analyses could be connected and mined to link regulatory information with changes in gene expression, the information from each dataset and for each gene was united in a single large matrix and mined with classification algorithms. Although our models were able to differentiate patterns in a set of simulated data, the raw datasets were too noisy for the models to consistently identify differentially expressed genes. However, by focusing on responses at a local level, we identified several genes with differential expression, differential sRNA, and differential methylation. While further studies will be needed to determine whether these epigenetic changes truly influence gene expression at these sites, the changes detected at the treatment level could prime the plants for future incidents of stress, including herbicides.

Epigenetic Responses of Arabidopsis to Abiotic Stress

Suzanne Laliberte

GENERAL AUDIENCE ABSTRACT

Growing resistance to herbicides, particularly glyphosate, is one of the many problems facing agriculture. The rapid rise of resistance across herbicide classes has caused some to wonder if there is a mechanism of adaptation that does not involve mutations. Epigenetics is the study of changes in the phenotype that cannot be attributed to changes in the genotype. Typically, studies revolve around two features of the chromosomes: cytosine methylation and histone modifications. The former can influence how proteins interact with DNA, and the latter can influence protein access to DNA. Both can affect each other in self-reinforcing loops. They can affect gene expression, and DNA methylation can be directed by small RNA (sRNA), which can also influence gene expression through other pathways. To study these processes and their role in abiotic stress response, we aimed to analyze sRNA, RNA, and DNA from *Arabidopsis thaliana* plants under stress. The stresses applied were sublethal doses of the herbicides, glyphosate and trifloxysulfuron, as well as mechanical clipping and shade to represent other weed management stressors. The focus of the project was to analyze these responses individually and together to find epigenetic responses to stresses routinely encountered by weeds. We tested RNA for gene expression changes under our stress conditions and identified many, including some pertaining to DNA methylation regulation. The herbicide treatments were associated with upregulated defense genes and downregulated growth genes. Shade treated plants had many downregulated defense and other stress response genes. We also detected differential methylation and sRNA responses when compared to the control plants. Changes to methylation and sRNA only accounted for about 20% of the variation in gene expression. While attempting to link the epigenetic process of methylation to gene expression, we connected all the data sets and developed computer programs to try to make correlations. While these methods worked on a simulated dataset, we did not detect broad patterns of changes to epigenetic pathways that correlated strongly with gene expression in our experiment's data. There are many factors that can influence gene expression that could create noise that would hinder the algorithms' abilities to detect differentially expressed genes. This does not, however, rule out the possibility of epigenetic influence on gene expression in local contexts. Through scoring the traits of individual genes, we found several that interest us for future studies.

Table of Contents:

Chapter 1: Describing Epigenetics Mechanisms and Their Connections to Stress Responses in Plants 1

Abstract 1

Introduction 1

Epigenetic responses to stress 2

Mechanisms of Epigenetic Activity 2

Regulation of DNA methylation 4

Examples of Epigenetics and Stress Response 7

Conclusions 8

References 9

Chapter 2: RNA Sequence Analysis of Weed Management Stress Responses in Arabidopsis thaliana 23

Abstract 23

Introduction 23

Methods 26

Results 28

Discussion 40

References 43

Chapter 3: Differential Responses of Epigenetic Pathways in Response to Weed Management Stressors 48

Abstract 48

Introduction 48

Methods 51

Results 52

Discussion 76

References 80

Chapter 4: Multi-‘omics Analysis of Arabidopsis Stress Response Analysis Combining sRNA, Methylation, and RNA Sequence Data 87

Abstract 87

Introduction 87

Methods 91

Results 95

Discussion 110

References 112

Chapter 1: Describing Epigenetics Mechanisms and Their Connections to Stress Responses in Plants

Abstract:

Epigenetics is the study of changes to a phenotype that are not due to changes in the genetic sequence of an organism. These changes can last for hours or persist across generations, and they allow plants to maintain stress response changes for varying lengths of time. Typically, epigenetics studies focus on either DNA methylation or histone modifications. Histone modifications affect enzyme access to DNA and binding of molecular machinery. They can also influence DNA methylation through self-reinforcing pathways. DNA methylation can occur in three contexts, CG, CHG, and CHH, with H as any base other than G. The contexts of methylation are maintained by distinct processes and can affect the expression of nearby features differently. RNA-directed DNA methylation can lead to methylation in all contexts. This process is complex and involves many proteins and is guided by small RNAs (sRNAs), which can also limit protein expression directly via RNA interference. All these changes have been known to occur in stress response, and they can sometimes impact gene expression. These changes could prime plants to respond more favorably to a second exposure to the same stressor.

Introduction:

Plants may encounter a variety of stressors throughout their lifetimes. Some, like bacterial pathogens and herbivorous insects, are biotic while abiotic stressors can include heat, flooding, chemical contaminants, and UV radiation.[1] Plants may evolve new mechanisms to manage these problems, or they can acclimate to handle the pressure. While evolution at the level of DNA sequence mutations can be an excellent strategy to cope with a long-term problem, many of the pressures plants face result from living in a dynamic environment where flexibility can be the key to survival. Sometimes plants can acclimate to environmental changes without relying on genomic mutations.[2] Adaptability must also be balanced with giving progeny the best chance to survive and reproduce; if progeny encounter the same stress, a hardwired response would give them the advantage.

Several well-documented examples of plant evolution in response to stress are found in response to herbicides.[3] Such a strong selective force provides great incentive for adaptation, and some plants have thus evolved to cope with these chemicals in a variety of ways. One method of surviving herbicide exposure is to overproduce the target enzyme. In response to glyphosate, there have been several reports of weed biotypes with resistance due to multiple copies of the target gene 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS).[4,5] With so much of the EPSPS enzyme available, a greater dose of the herbicide is needed for inhibition of amino acid synthesis. Acetolactate synthase and acetyl-CoA carboxylase inhibiting herbicides block the synthesis of other essential amino acids.[6,7,8] Several species of weeds acquired resistance to these two chemicals through mutations in the genomic sequence at the target site of the herbicide. They switched at least one amino acid at or near the site where the herbicide binds to render the enzyme resistant to the herbicide.[9] Another instance of this type of resistance occurred in response to triazine, which inhibits photosystem II. A common target site mutation confers resistance, but it also hinders enzyme functionality to make photosynthesis less efficient. In this case, the mutation comes at a fitness cost that negatively affects growth and reproduction. Triazine resistant weeds would have trouble competing with their susceptible counterparts in the

absence of the herbicide.[10] Sometimes evolved responses are not ideal and even outright harmful to the plant, yet they are advantageous in the face of strong selection pressure.

Epigenetic responses to stress

While structurally altered genomes and sequence mutations have helped some plant species to survive new challenges, other cases of stress response cannot be attributed to DNA mutations.[3] Holly leaves on smaller trees or that are closer to the ground are more likely to be prickly.[11] This adaptation can occur in response to mammalian herbivory and may be a defense response.[11] As these changes occur on branches of an individual plant, it is unlikely that they were caused by a change in the genomic sequence.[12] In a more transient case, Arabidopsis plants previously exposed to drought acquired some resistance to this stress for a period of 5-7 days.[13] In another study, seeds subjected to salt stress exhibited higher germination rates under salty conditions.[14] Two generations of progeny from Arabidopsis exposed to *Pseudomonas syringae* exhibited increased resistance to the pathogen.[15,16] In snapdragons, there is a gene that can affect flower shape; however, researchers discovered some flowers with this phenotype yet without changes to this gene.[17] They did find that the gene was extensively methylated in plants with the mutant phenotype but the wild-type gene. These changes may, in the case of the holly trees, be localized to an area of the individual plant while others, such as those found in the *Pseudomonas* stress study, can persist through at least a couple of generations. They are unlikely to be a result of mutations and may be caused by epigenetic processes.

Epigenetics is the study of changes to an organism's phenotype that are not reflected in a change to its genotype.[18] One of the two commonly known mechanisms for epigenetic changes is DNA methylation, in which a methyl group is added to the fifth carbon of a cytosine nucleotide.[19] DNA methylation in plants can occur in three contexts: CG, CHG, and CHH. 'H' is any nucleotide base other than guanine.[20] Patterns of DNA methylation can impact expression of nearby genomic features, such as genes and transposons.[21,22,23] Although it is accepted that many epigenetic changes are reset between generations, some of them can be inherited.[24,25]

The other commonly studied epigenetic mark affects histone proteins, which bind to DNA to form the basic unit of chromatin structure.[26] Histones contain 2 copies of four proteins, H2A, H2B, H3, and H4, that bind together to form an octamer. DNA is wrapped around this protein complex to form the nucleosome. The amino acid residues of these proteins can accept various groups to affect how histones interact with other proteins and how the DNA interacts with the nucleosome. Sometimes the groups can cause changes individually; in other cases, multiple modifications can create a pattern associated with a specific response. A few examples of histone modifications are acetylation, methylation, phosphorylation, ubiquitylation, and sumoylation.[27,28] While we will not consider histone data for this experiment, these modifications sometimes correlate with patterns of DNA methylation, chromatin structure, and expression of genomic features.[29]

Mechanisms of Epigenetic Activity

The recognized epigenetic mark we studied, DNA methylation, is believed to impact gene expression. There are a few common patterns of DNA methylation near plant genes.[30] Unmethylated genes lack methylation in all contexts. CG methylation near the transcription start

site is often associated with repression.[31,32] The most frequent class is gene body methylation, in which CG methylation increases near the center of the genes and decreases near the transcription start and stop sites. CHG enriched genes have methylation in CHG and sometimes CG contexts, and they are frequently repressed.[33] Plant genes can also be affected by RNA directed DNA methylation (RdDM), which is associated with methylation of cytosines in all contexts and gene silencing.[34]

The role of gene body methylation is unclear. Cytosine methylation increases the likelihood of its conversion to thymine by deanimation.[35] Despite this risk, methylation patterns are conserved in orthologous genes across several plant species, and these genes tend to evolve slowly.[36,37] CG gene body methylation is associated with moderately expressed genes.[31,32] A study of Arabidopsis accessions found no strong correlation between gene body methylation and gene expression.[38] Some nonvascular plant species do not have the CG gene body methylation pattern.[39] There have been two independent instances where members of the Brassicaceae family lost chromomethylase 3 (CMT3) function and associated gene body methylation with little to no impact on gene expression. This has led the authors to suggest that gene body methylation could merely be a side effect of CMT3 activity.[40] Brassicaceae also appear to have lower levels of gene body methylation when compared to other angiosperms.[33] Perhaps CMT3 function varies between species and other factors.[41] Gene body methylation may affect gene expression differently in Arabidopsis than other species.

Methylation can also occur near transposons and affect their activity. The patterns of transposon methylation have been the subject of several studies. Hypermethylation is frequently coupled with silencing, and hypomethylation often leads to activation.[42] Usually, the methylation occurs in non-CG contexts and is associated with silencing. The type of methylation near transposons is associated with several factors, including species, transposon class, and chromosomal position. For example, sugar beets had higher levels of CG and CHG methylation in retrotransposons, and higher levels of methylation in all contexts of DNA transposons.[43] Trends of higher methylation levels in transposons have also been detected in other crops. Transposon methylation patterns were affected by chromosome position and transposon family in maize. Transposons near transcribed maize genes were characterized by 24 nt small interfering RNA (siRNA) and CHH methylation while transposons near repressed regions had higher levels of CG and CHG methylation.[44] Tomato also had heavily methylated transposons in repressed regions with local increases of CHH methylation of transposons in active regions.[45] 24 nt siRNA and CHH methylation are associated with RdDM, in which regulation of a specific region of the genome can be targeted based on its sequence. It is one way to control where DNA methylation occurs.

In a transcriptionally active region of the genome, methylation can have negative consequences for the plant by reducing transcription of essential genes. Methylation near transposons can come in a variety of patterns and can impact plant fitness in a variety of ways. Methylation of transposons in rice could spread to impact nearby genes, especially when the transposons were located within genes. In this case, there did not appear to be a strong fitness cost to the reduced expression.[46] Conversely, methylation changes can cause oil palm's mantled phenotype, which greatly reduces fruit quality and fertility. This change in fruit structure is instigated by a loss of RdDM of the Karma transposons in an intron of EgDEF1 and improper

splicing of EgDEF1.[23] This illustrates how targeted methylation can affect nearby features, and represents one pattern that could impact gene expression in a manner that contributes to herbicide or other abiotic stress tolerance.

Histone proteins are affected by a variety of enzymes that usually fall into one of three categories: writers that add groups to histone residues, readers that bind to certain modifications, and erasers that remove groups.[47] These groups can impact chromatin structure and accessibility. Some marks are more common in the accessible euchromatin, while others are usually localized in the densely packed heterochromatin.[48],[49] Some modifications, such as acetylation, interfere with chromatin folding to make DNA more accessible for transcription.[50] Histone methylation, on the other hand, does not impact the charge, and its effect on DNA accessibility and chromatin structure may be more context specific.[51] Many histone modifications and associated enzymes are required for plant germination and transition to flowering states.[29] Others patterns are associated with stress response.[13]

The effects of some histone modifications have been identified in plants. Dimethylation of lysine 9 on histone 3 (H3K9me2) is associated with heterochromatin while H3K9me3 is associated with euchromatin.[48],[49] H3K27me3 is associated with repression of genes in plant euchromatin.[52] H4K20me is found in heterochromatin near transposable elements.[53] H2B monoubiquitination (H2Bub1) is associated with activation, while H2Aub1 is associated with repression.[54] H2Bub1 is required for H3K79 and H3K4 methylation, which are also associated with activation, and it blocks H3K9me2.[55] H3K9me2 is frequently associated with DNA methylation in the CHG context and silencing of genes and transposons.[56] Genes with a majority of H3K9me2 and a minority of H3K9 acetylation (H3K9ac) are usually repressed while others with both marks are expressed near the activated levels H3K9ac.[56] Histone marker H3K9ac near the 5' region of genes is associated with higher levels of expression.[56] Histone proteins can affect chromatin structure, DNA accessibility, other histone marks, and DNA methylation.

Regulation of DNA methylation

Different methylation patterns are associated with mechanisms for DNA methylation. DNA methyltransferase (MET1) is responsible for maintenance of most CG methylation in plants.[57],[58],[59] It also has some capacity for de novo methylation. Re-introducing MET1 to a met1 knockout can restore some CG methylation in Arabidopsis.[60] MET1 can work with Histone Deacetylase 6 (HDA6), which deacetylates and demethylates histones, to methylate TEs.[61] CHG methylation is maintained by a chromomethylase (CMET3) and kryptonite KYP/ SUVH4.[62],[63],[64] Homologs SUVH5 and SUVH6 also work with CMET3 although their activity is associated with transposons.[65] CHG methylation is associated with H3K9me2.[66] CMET3 binds to H3K9me2 to methylate DNA, and KYP binds to methylated DNA to methylate H3K9.[67,68] H3K9me3 is associated with silencing of transposons in the euchromatin.[69] CHH methylation is typically a result of RdDM, but nucleosome remodeler decrease in DNA methylation 1 (DDM1) can open DNA near H1 for CMET2 to maintain CHH methylation in an RdDM-independent manner. This is also associated with silencing of genes and transposons.[70] Together, these pathways allow for a variety of methylation contexts and histone modifications to regulate the genome.

DNA methylation can be removed via the base excision repair pathway, which removes methylated cytosines and replaced them with their unmethylated counterparts.[71] This is important for removing unwanted methyl groups as well as triggering methylation maintenance pathways by activating transposons in a more controlled manner. The demethylation allows sRNA production to ensure continued methylation near these sites.[72] Demethylation associated with sRNA production could indicate that the plant is remethylating a transposon. 5-meC DNA glycosylases, ROS1, DME, DML2, and DML3 are responsible for removing methylated cytosines from DNA.[73],[74] DME plays an additional role in removing methyl groups at maternal alleles in the endosperm.[75] ROS1 recruitment is associated with H3K18ac and H3K27me3.[76] After the methylated cytosines are removed, DNA 3′-phosphatase ZDP processes the 3'P, and endonucleases ARP or APE1L convert it to an OH group.[77],[78] Then, polymerase β and ligase LIG1 repairs the DNA strand.[79,80] This leaves an unmethylated cytosine, which can activate a silenced region or help with sRNA production for continued methylation at other cytosines nearby.[81]

Small RNA (sRNA) plays a role in guiding methylation in RdDM.[34],[41] These guiding RNA sequences can originate from several pathways. RNA polymerase IV (POL IV) initially synthesizes long noncoding RNA molecules, and it is assisted by SNF2 Domain-Containing Protein Classy (CLS) and SAWADEE Homeodomain Homologue 1 (SHH1).[82],[83] CLS1 and CLS2 help with SHH1 interactions while CLS3 and CLS4 appear to require CG methylation to function.[84] SHH1 also binds to H3K9me2 and unmethylated or monomethylated H3K4 to assist with POL IV binding in RdDM.[83],[85] An RNA dependent RNA polymerase (RDR2) is recruited to synthesize the other strand.[86] A dicer (DCL) protein cleaves the double stranded RNA molecule to produce 24 nt, double-stranded siRNA. In an alternative pathway, RDR6 can synthesize the complementary strand of RNA synthesized with POL II.[87] DCL2 and DCL4 can cleave 21 or 22 nt siRNA for RdDM, and DCL3 can cleave 24 nt siRNA from POL II transcripts.[88] This means we might expect to see sRNA around this length near sites of increased methylation. It could even lead to some of the detected differentially methylated regions. Methylation at RdDM targets can still be present in dcl knockouts, which suggests that dicer-independent RdDM is possible.[89] These short sequences can help guide methylation with the assistance of other proteins.

After the siRNA is synthesized, it can bind to Argonaute proteins, AGO2, AGO4, or AGO9.[86],[90] The AGO4 protein holds its guide RNA as it binds to long noncoding RNA synthesized by POL V.[91] POL V is recruited in part by the DDR complex, which is composed of Defective in Meristem Silencing 3 (DMS3), Defective in RNA Directed DNA Methylation 1 (DRD1), and RNA Directed DNA Methylation 1 (RDM1).[92] DMS3 helps bind the complex to the DNA.[93] SUVH2 and SUVH9 bind to methylated DNA and DRD1 of the DDR complex.[94] RRP6L1 helps hold the POL V transcripts to the chromatin.[95] lncRNA-binding protein IDN2 binds to the POL V transcripts and ATP-dependent nucleosome-remodeling complex SWI/SNF binds to it.[96] Domains Rearranged Methylase 2 (DRM2) methylates cytosines in all contexts on one strand after binding with a loaded AGO protein.[97] The resulting non-CG methylation can silence genes and transposons.

RdDM is just one of several processes for regulation by sRNA that fall into the category of RNA interference (RNAi), which is associated with siliencing.[98,99] In this case, we would expect to see reduced expression associated with the sRNA. Plant siRNA typically match their targets closely.[100] In some RNAi, double-stranded 21-22 bp sequences direct cleavage of mRNA within the complementary region.[101] Typically, the break occurs 10 bp from the 5' end of the siRNA.[101] RNAi starts with long noncoding RNA (lncRNA), which is cleaved into siRNA. These short sequences are amplified by RNA dependent RNA polymerases. RNAi is then facilitated by the RNA induced silencing complex (RISC), which is a cluster of proteins that can vary between plants and animals.[102],[103] One of the proteins is DICER, which cleaved lncRNA molecules into siRNA.[104] Another member is Argonaute 2 (AGO2), which cleaves the sense strand of siRNA to leave a single stranded guide RNA that is complementary to the target sequence.[105],[106],[107]

Arabidopsis has four DICER proteins (DCL).[108] DCL1 is the only one that is essential, as indicated by finding that its knockout mutant is lethal, and it generates 21 nt sRNA that often arise from hairpin structures.[108] DCL2 generates 22 nt sRNA, DCL3 generates 24 nt, and DCL4 generates 21 nt sRNA.[108] These three are associated with resistance to viruses.[109] DCL1 negatively regulates DCL2 and 4 activities.[109] DCL2-4 are involved in RdDM.[108] DCL1 has cofactors HYL and SE, double-stranded RNA binding proteins, that bind to RNA for miRNA synthesis.[110],[111],[112]

Arabidopsis has six RDRs from two classes, RDRα and RDRγ.[113] RDR1, RDR2, and RDR6 are RDRα and are important for various RNAi processes while the function of RDRγ is not known.[113] Some plants have fewer RDRs, but most have those three.[113] RDR1 is involved in generating siRNA from viral genes.[114] RDR2 helps generate 24 nt siRNA and is involved in RdDM.[115],[116] RDR6 contributes to resistance to viruses, and it works with Suppressor of Gene Silencing 3 (SGS3) to produce trans-siRNA.[117]

RNAi does not always involve cleaving the target mRNA. Sometimes mismatches prevent cleavage, but mismatches are not always the cause.[118],[119],[120] Aside from slicing, RNAi can deadenylate and de-cap mRNA to destabilize it.[121],[122],[123] AMP1 is bound to the rough endoplasmic reticulum (ER), and it can also bind to loaded AGO1.[124] miRNA associated with translation inhibition is also localized to the rough ER, which suggests that plant cells keep target mRNAs away from translation machinery by binding them to the rough ER in a post-transcriptional manner.[124]

siRNA has been classified based on its functions and origins.[125] MicroRNA (miRNA) originates from a single stranded RNA molecule with an inverted repeat that folds to form a hairpin structure.[126] It is not as specific as other siRNA and can bind to multiple targets.[126] miRNA sequences can also be modified by adding or removing bases to form isomiRs.[127] This is also true of other sRNA.[128],[129],[130] siRNA can be generated by miRNA directed cleavage of RNA at two sites.[131] Some siRNA can form from hairpin structures that fail to meet the criteria to be classified as miRNA. nat-siRNA is characterized by its origin, in which two complementary strands bind to form double-stranded siRNA. These two strands can overlap, or they can arise from completely different regions. Trans-acting (tasiRNA) works to silence regions other than its

origin. Because siRNA sequences can regulate multiple regions of the genome, it is possible for siRNA to be trans- and cis acting. Phased siRNA is generated as RNA is diced in intervals from a larger sequence. Heterochromatic siRNA is associated with epigenetic changes and RdDM.[125] There are other classification schemes that incorporate groups such as easiRNA that are generated from active retroposons.[132] siRNA can even arise from tRNA and rRNA.[133],[134] Individual siRNA may fall into multiple categories, but these classification systems are useful when sifting through sequence data for siRNA. Different software specializes in identifying different groups.[135]

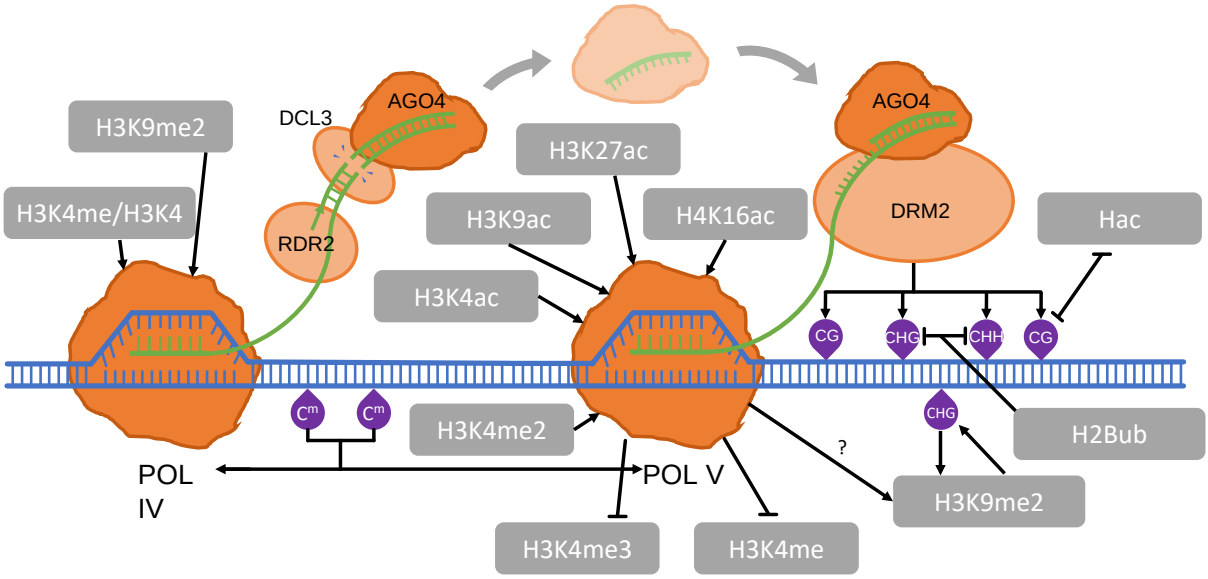


Figure 1 DNA Methylation and Histone Modifications: Some histone modifications have been linked to enzymes that support the polymerases of RdDM or are associated with other DNA methylation.[136] Some marks are known to correlate with POL V sites in rice.[137] H3K4me3 can block POL V recruitment.[138]

DNA methylation has been linked to histone modifications (Figure 1). POL IV of RdDM is bound to DNA with the assistance of Sawadee Homeodomain Homologue 1, a protein that reads H3K9me2 and unmethylated or monomethylated H3K4.[83],[85] SUVH2 and SUVH9 of the POL V complex bind to methylated DNA. Other proteins associated with this complex facilitate binding to modified histones.[96] Histone deacetylase 6, Jumonji 14, and ubiquitin-specific protease 26 can remove activating marks to make way for repressive marks, such as H3K9me by SUVH proteins.[136] In CHG methylation, CMT3 binds to H3K9me while methylating DNA, and KYP binds to methylated DNA while it methylates H3K9.[67],[68] This forms a self-reinforcing loop for CHG methylation maintenance, and H3K9me2 is associated with the silencing of genes and transposons.[56] In maize, RdDM appeared to limit a transposon's expression when heat stress compromised the silencing histone marks.[139] At this site, RdDM appeared to aid in the stable inheritance of histone modifications when stress might otherwise induce less repressive marks.

Examples of Epigenetics and Stress Response

All these mechanisms can come together to give plants the ability to respond to unique situations. For example, changes in methylation have been observed when plants are under stress. In one study, Arabidopsis plants were exposed to salt, UV, drought, cold, heat, and flooding. Global hypermethylation was detected in all but the drought stressed plants.[140] Global levels returned to normal in the progeny, but progeny of salt stressed had higher germination rates on salty media.[140] This response was lost in *dcl* knockout mutants.[140] This indicates that RNA directed DNA methylation can play a role in salt stress response. Another example of methylation and stress is found in SPCH and FAMA, two genes involved in stomatal development. Under dry conditions, they are methylated with reduced expression and the plants produce fewer stomata.[141] The phenotype was somewhat heritable in progeny crossed with wild type Arabidopsis.[142] Knockouts of RdDM proteins did not have this response and methylation of FAMA was reduced.[141] RdDM may facilitate some aspects of cold stress response. FLC expression is reduced under cold stress, and this repression is accompanied by 24 and 30 nt siRNA as well as H3K9me2.[143] Histone modifications are also affected by stress. Stalled Ser5P Pol II and H3K4me3 were found near drought induced, trainable genes.[13] Levels of H3K4me3 were increased around many drought responsive genes, and these levels remained high in "trainable" genes.[13] This stress memory only lasted 5-7 days.[13] H3K9ac has been associated with drought as well, but this modification disappears quickly as the stress is removed.[144] These examples illustrate how epigenetic changes can be responsive to stress, are sometimes heritable, and may alter plant phenotypes.

Some major plant signals with known ties to stress response are also associated with methylation and gene expression changes. Plant defense responses are often mediated by salicylic acid and jasmonic acid. Salicylic acid helps activate innate, routine defenses and disease resistance while jasmonic acid manages defense response to urgent threats.[145] Salicylic acid responses can regulate jasmonic acid responses.[146] Both are known to trigger epigenetic changes and differential expression of defense-related genes.[147,148] Some studies have linked brassinosteroids to epigenetic changes and plant defense regulation.[145,149–151] Response to drought, cold, and high salt stress is linked to abscisic acid (ABA).[145] ABA responses may alter gene expression through regulating histone writer proteins.[152] Several drought-sensitive genes have changes to DNA methylation or histone modifications.[153–155] Auxin is a plant growth regulator that is involved in many processes related to development and can inhibit ABA.[145] Auxin responses could be influenced by DNA methylation changes from methylation knockout mutants or viruses.[156–158] These are some of the most important and widely studied signaling pathways in plants involved in stress response.[145] There are many potential avenues for stress responses to trigger epigenetic changes and influence plant development.

Conclusions:

Epigenetics allows for changes to the expression of genes and transposons without requiring changes to DNA sequences. This gives plants plenty of ways to respond to environmental stimuli. These complicated processes can influence each other to provide multiple switches to allow plants to respond effectively to unique situations. DNA may bind tightly within the heterochromatin through histone modifications and thus prevent the transcription of unhelpful genes or transposons.[50] Heterochromatin may be opened again as those locked up

genes are needed. Some proteins may bind to histone marks and regulate genes, sometimes through recruiting DNA methylases.[47,97,136]

Methylation, in turn, can recruit proteins for histone modification.[56] Methylation changes can occur in three contexts and are maintained by several processes.[20] They are also modified through demethylation.[71] DNA methylation prevent the disruptive activities of transposons.[21,22,23] Maintenance of epigenetic marks at other loci can also protect the plants from harmful viral DNA.[117] A wide variety of sRNA controls gene expression and influence methylation. The major process of RdDM involves many proteins and allows targeted adjustments to the methylome to be implemented.[136] Many cases of stress response and associated signaling pathways have been linked to changes in gene expression and epigenetics. Many such changes are fluid and can allow plants to respond without the occasionally detrimental commitment of an evolved response. Some responses are less transient and will stick around to give progeny a faster alternative to evolution by mutation. Several researchers have been working on hijacking these mechanisms to further human interests in crop production and weed control.[159] Additional study is needed to understand how these mechanisms can cooperate in the myriad contexts.

There are still many questions surrounding epigenetics and stress response regulation. There are many histone marks with unknown functions, and the role of gene body methylation is still undetermined.[31,32,40] Some challenges facing researchers include the presence of other proteins, such as transcription factors, that can influence gene expression. With so many elements regulating gene expression, proving the specific changes that lead to different phenotypes can be extremely challenging.[160] More studies are needed that explore multiple types of regulation simultaneously. This gives researchers a better opportunity to clearly identify epigenetic processes at work. Being able to rule out other factors would allow them to confidently identify the source of the new phenotypes.

The heritability of epigenetic changes is inconsistent and debated.[161,162] Knowing how many generations of stress are required for the heritable changes to become fixed in the phenotype could be very useful for managing plant stress responses in agriculture. More studies are needed to determine the transience of epigenetic responses. From the study conducted in maize, it appeared that the altered transposon state took several generations before the phenotype was inherited stably.[139] Similar epigenetic changes induced by an herbicide could lead to a weed's resistant phenotype.[163] Perhaps there are other stresses challenging crop growth that could be moderated like the salt priming of seeds.[14] Knowing how these processes are triggered and how long epigenetic changes will persist could help researchers use them to improve agricultural productivity.

References:

(1) Bolaji Umar, O.; Amudalat Ranti, L.; Shehu Abdulbaki, A.; Lukman Bola, A.; Khadijat Abdulhamid, A.; Ramat Biola, M.; Oluwagbenga Victor, K. Stresses in Plants: Biotic and Abiotic. In *Current Trends in Wheat Research*; Ansari, M.-R., Ed.; IntechOpen, 2022. https://doi.org/10.5772/intechopen.100501.

(2) Kleine, T.; Nägele, T.; Neuhaus, H. E.; Schmitz-Linneweber, C.; Fernie, A. R.; Geigenberger, P.; Grimm, B.; Kaufmann, K.; Klipp, E.; Meurer, J.; Möhlmann, T.;

Mühlhaus, T.; Naranjo, B.; Nickelsen, J.; Richter, A.; Ruwe, H.; Schroda, M.; Schwenkert, S.; Trentmann, O.; Willmund, F.; Zoschke, R.; Leister, D. Acclimation in Plants – the Green Hub Consortium. *Plant J.* **2021**, *106* (1), 23–40. https://doi.org/10.1111/tpj.15144.

(3) Baucom, R. S. The Remarkable Repeated Evolution of Herbicide Resistance. *Am. J. Bot.* **2016**, *103* (2), 181–183. https://doi.org/10.3732/ajb.1500510.

(4) Gaines, T. A.; Zhang, W.; Wang, D.; Bukun, B.; Chisholm, S. T.; Shaner, D. L.; Nissen, S. J.; Patzoldt, W. L.; Tranel, P. J.; Culpepper, A. S.; Grey, T. L.; Webster, T. M.; Vencill, W. K.; Sammons, R. D.; Jiang, J.; Preston, C.; Leach, J. E.; Westra, P. Gene Amplification Confers Glyphosate Resistance in Amaranthus Palmeri. *Proc. Natl. Acad. Sci.* **2010**, *107* (3), 1029–1034. https://doi.org/10.1073/pnas.0906649107.

(5) Gaines, T. A.; Barker, A. L.; Patterson, E. L.; Westra, P.; Westra, E. P.; Wilson, R. G.; Jha, P.; Kumar, V.; Kniss, A. R. EPSPS Gene Copy Number and Whole-Plant Glyphosate Resistance Level in Kochia Scoparia. *PLOS ONE* **2016**, *11* (12), e0168295. https://doi.org/10.1371/journal.pone.0168295.

(6) Zhou, Q.; Liu, W.; Zhang, Y.; Liu, K. K. Action Mechanisms of Acetolactate Synthase-Inhibiting Herbicides. *Pestic. Biochem. Physiol.* **2007**, *89* (2), 89–96. https://doi.org/10.1016/j.pestbp.2007.04.004.

(7) Takano, H. K.; Ovejero, R. F. L.; Belchior, G. G.; Maymone, G. P. L.; Dayan, F. E. ACCase-Inhibiting Herbicides: Mechanism of Action, Resistance Evolution and Stewardship. *Sci. Agric.* **2021**, *78* (1), e20190102. https://doi.org/10.1590/1678-992x-2019-0102.

(8) Shaner, D. L. Lessons Learned From the History of Herbicide Resistance. *Weed Sci.* **2014**, *62* (2), 427–431. https://doi.org/10.1614/WS-D-13-00109.1.

(9) Heap, I. *The International Herbicide-Resistant Weed Database*. www.weedscience.org. http://www.weedscience.com/Pages/MOA.aspx?MOAID=12 (accessed 2020-03-19).

(10) Vila-Aiub, M. M.; Neve, P.; Powles, S. B. Fitness Costs Associated with Evolved Herbicide Resistance Alleles in Plants. *New Phytol.* **2009**, *184* (4), 751–767. https://doi.org/10.1111/j.1469-8137.2009.03055.x.

(11) Obeso, J. R. The Induction of Spinescence in European Holly Leaves by Browsing Ungulates. *Plant Ecol.* **1997**, *129* (2), 149–156. https://doi.org/10.1023/A:1009767931817.

(12) Herrera, C. M.; Bazaga, P. Epigenetic Correlates of Plant Phenotypic Plasticity: DNA Methylation Differs between Prickly and Nonprickly Leaves in Heterophyllous *Ilex Aquifolium* (Aquifoliaceae) Trees: Heterophylly and DNA Methylation. *Bot. J. Linn. Soc.* **2013**, *171* (3), 441–452. https://doi.org/10.1111/boj.12007.

(13) Ding, Y.; Fromm, M.; Avramova, Z. Multiple Exposures to Drought "train" Transcriptional Responses in Arabidopsis. *Nat. Commun.* **2012**, *3* (1), 740. https://doi.org/10.1038/ncomms1732.

(14) Johnson, R.; Puthur, J. T. Seed Priming as a Cost Effective Technique for Developing Plants with Cross Tolerance to Salinity Stress. *Plant Physiol. Biochem.* **2021**, *162*, 247–257. https://doi.org/10.1016/j.plaphy.2021.02.034.

(15) Stassen, J. H. M.; López, A.; Jain, R.; Pascual-Pardo, D.; Luna, E.; Smith, L. M.; Ton, J. The Relationship between Transgenerational Acquired Resistance and Global DNA Methylation in Arabidopsis. *Sci. Rep.* **2018**, *8* (1), 14761. https://doi.org/10.1038/s41598-018-32448-5.

(16) Luna, E.; Bruce, T. J. A.; Roberts, M. R.; Flors, V.; Ton, J. Next-Generation Systemic Acquired Resistance. *Plant Physiol.* **2012**, *158* (2), 844–853. https://doi.org/10.1104/pp.111.187468.
(17) Cubas, P.; Vincent, C.; Coen, E. An Epigenetic Mutation Responsible for Natural Variation in Floral Symmetry. *Nature* **1999**, *401* (6749), 157–161. https://doi.org/10.1038/43657.
(18) Weinhold, B. Epigenetics: The Science of Change. *Environ. Health Perspect.* **2006**, *114* (3), A160–A167. https://doi.org/10.1289/ehp.114-a160.
(19) Moore, L. D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2013**, *38* (1), 23–38. https://doi.org/10.1038/npp.2012.112.
(20) Pikaard, C. S.; Scheid, O. M. Epigenetic Regulation in Plants. *Cold Spring Harb. Perspect. Biol.* **2014**, 1–31. https://doi.org/10.1101/cshperspect.a019315.
(21) Miura, K.; Agetsuma, M.; Kitano, H.; Yoshimura, A.; Matsuoka, M.; Jacobsen, S. E.; Ashikari, M. A Metastable DWARF1 Epigenetic Mutant Affecting Plant Stature in Rice. *Proc. Natl. Acad. Sci.* **2009**, *106* (27), 11218–11223. https://doi.org/10.1073/pnas.0901942106.
(22) Soppe, W. J. J.; Jacobsen, S. E.; Alonso-Blanco, C.; Jackson, J. P.; Kakutani, T.; Koornneef, M.; Peeters, A. J. M. The Late Flowering Phenotype of Fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. *Mol. Cell* **2000**, *6* (4), 791–802. https://doi.org/10.1016/S1097-2765(05)00090-0.
(23) Ong-Abdullah, M.; Ordway, J. M.; Jiang, N.; Ooi, S.-E.; Kok, S.-Y.; Sarpan, N.; Azimi, N.; Hashim, A. T.; Ishak, Z.; Rosli, S. K.; Malike, F. A.; Bakar, N. A. A.; Marjuni, M.; Abdullah, N.; Yaakub, Z.; Amiruddin, M. D.; Nookiah, R.; Singh, R.; Low, E.-T. L.; Chan, K.-L.; Azizi, N.; Smith, S. W.; Bacher, B.; Budiman, M. A.; Van Brunt, A.; Wischmeyer, C.; Beil, M.; Hogan, M.; Lakey, N.; Lim, C.-C.; Arulandoo, X.; Wong, C.-K.; Choo, C.-N.; Wong, W.-C.; Kwan, Y.-Y.; Alwee, S. S. R. S.; Sambanthamurthi, R.; Martienssen, R. A. Loss of Karma Transposon Methylation Underlies the Mantled Somaclonal Variant of Oil Palm. *Nature* **2015**, *525* (7570), 533–537. https://doi.org/10.1038/nature15365.
(24) Iwasaki, M.; Paszkowski, J. Identification of Genes Preventing Transgenerational Transmission of Stress-Induced Epigenetic States. *Proc. Natl. Acad. Sci.* **2014**, *111* (23), 8547–8552. https://doi.org/10.1073/pnas.1402275111.
(25) Iwasaki, M.; Paszkowski, J. Epigenetic Memory in Plants. *EMBO J.* **2014**, *33* (18), 1987–1998. https://doi.org/10.15252/embj.201488883.
(26) Lämke, J.; Bäurle, I. Epigenetic and Chromatin-Based Mechanisms in Environmental Stress Adaptation and Stress Memory in Plants. *Genome Biol.* **2017**, *18* (1), 124. https://doi.org/10.1186/s13059-017-1263-6.
(27) Ramirez-Prado, J. S.; Piquerez, S. J. M.; Bendahmane, A.; Hirt, H.; Raynaud, C.; Benhamed, M. Modify the Histone to Win the Battle: Chromatin Dynamics in Plant–Pathogen Interactions. *Front. Plant Sci.* **2018**, *9*, 355. https://doi.org/10.3389/fpls.2018.00355.
(28) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; Lu, Z.; Ye, Z.; Zhu, Q.; Wysocka, J.; Ye, Y.; Khochbin, S.; Ren, B.; Zhao, Y. Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* **2011**, *146* (6), 1016–1028. https://doi.org/10.1016/j.cell.2011.08.008.

(29) Zhao, T.; Zhan, Z.; Jiang, D. Histone Modifications and Their Regulatory Roles in Plant Development and Environmental Memory. *J. Genet. Genomics* **2019**, *46* (10), 467–476. https://doi.org/10.1016/j.jgg.2019.09.005.
(30) Bewick, A. J.; Schmitz, R. J. Gene Body DNA Methylation in Plants. *Curr. Opin. Plant Biol.* **2017**, *36*, 103–110. https://doi.org/10.1016/j.pbi.2016.12.007.
(31) Liang, L.; Chang, Y.; Lu, J.; Wu, X.; Liu, Q.; Zhang, W.; Su, X.; Zhang, B. Global Methylomic and Transcriptomic Analyses Reveal the Broad Participation of DNA Methylation in Daily Gene Expression Regulation of Populus Trichocarpa. *Front. Plant Sci.* **2019**, *10*, 243. https://doi.org/10.3389/fpls.2019.00243.
(32) Li, X.; Zhu, J.; Hu, F.; Ge, S.; Ye, M.; Xiang, H.; Zhang, G.; Zheng, X.; Zhang, H.; Zhang, S.; Li, Q.; Luo, R.; Yu, C.; Yu, J.; Sun, J.; Zou, X.; Cao, X.; Xie, X.; Wang, J.; Wang, W. Single-Base Resolution Maps of Cultivated and Wild Rice Methylomes and Regulatory Roles of DNA Methylation in Plant Gene Expression. *BMC Genomics* **2012**, *13* (1), 300. https://doi.org/10.1186/1471-2164-13-300.
(33) Niederhuth, C. E.; Bewick, A. J.; Ji, L.; Alabady, M. S.; Kim, K. D.; Li, Q.; Rohr, N. A.; Rambani, A.; Burke, J. M.; Udall, J. A.; Egesi, C.; Schmutz, J.; Grimwood, J.; Jackson, S. A.; Springer, N. M.; Schmitz, R. J. Widespread Natural Variation of DNA Methylation within Angiosperms. *Genome Biol.* **2016**, *17* (1), 194. https://doi.org/10.1186/s13059-016-1059-0.
(34) Wassenegger, M.; Heimes, S.; Riedel, L.; Sänger, H. L. RNA-Directed de Novo Methylation of Genomic Sequences in Plants. *Cell* **1994**, *76* (3), 567–576. https://doi.org/10.1016/0092-8674(94)90119-8.
(35) Ehrlich, M.; Norris, K. F.; Wang, R. Y.; Kuo, K. C.; Gehrke, C. W. DNA Cytosine Methylation and Heat-Induced Deamination. *Biosci. Rep.* **1986**, *6* (4), 387–393. https://doi.org/10.1007/BF01116426.
(36) Takuno, S.; Gaut, B. S. Gene Body Methylation Is Conserved between Plant Orthologs and Is of Evolutionary Consequence. *Proc. Natl. Acad. Sci.* **2013**, *110* (5), 1797–1802. https://doi.org/10.1073/pnas.1215380110.
(37) Takuno, S.; Gaut, B. S. Body-Methylated Genes in Arabidopsis Thaliana Are Functionally Important and Evolve Slowly. *Mol. Biol. Evol.* **2012**, *29* (1), 219–227. https://doi.org/10.1093/molbev/msr188.
(38) Kawakatsu, T.; Huang, S. C.; Jupe, F.; Sasaki, E.; Schmitz, R. J.; Urich, M. A.; Castanon, R.; Nery, J. R.; Barragan, C.; He, Y.; Chen, H.; Dubin, M.; Lee, C.-R.; Wang, C.; Bemm, F.; Becker, C.; O'Neil, R.; O'Malley, R. C.; Quarless, D. X.; Schork, N. J.; Weigel, D.; Nordborg, M.; Ecker, J. R.; Alonso-Blanco, C.; Andrade, J.; Becker, C.; Bemm, F.; Bergelson, J.; Borgwardt, K.; Chae, E.; Dezwaan, T.; Ding, W.; Ecker, J. R.; Expósito-Alonso, M.; Farlow, A.; Fitz, J.; Gan, X.; Grimm, D. G.; Hancock, A.; Henz, S. R.; Holm, S.; Horton, M.; Jarsulic, M.; Kerstetter, R. A.; Korte, A.; Korte, P.; Lanz, C.; Lee, C.-R.; Meng, D.; Michael, T. P.; Mott, R.; Muliyati, N. W.; Nägele, T.; Nagler, M.; Nizhynska, V.; Nordborg, M.; Novikova, P.; Picó, F. X.; Platzer, A.; Rabanal, F. A.; Rodriguez, A.; Rowan, B. A.; Salomé, P. A.; Schmid, K.; Schmitz, R. J.; Seren, Ü.; Sperone, F. G.; Sudkamp, M.; Svardal, H.; Tanzer, M. M.; Todd, D.; Volchenboum, S. L.; Wang, C.; Wang, G.; Wang, X.; Weckwerth, W.; Weigel, D.; Zhou, X. Epigenomic Diversity in a Global Collection of Arabidopsis Thaliana Accessions. *Cell* **2016**, *166* (2), 492–505. https://doi.org/10.1016/j.cell.2016.06.044.

(39) Takuno, S.; Ran, J.-H.; Gaut, B. S. Evolutionary Patterns of Genic DNA Methylation Vary across Land Plants. *Nat. Plants* **2016**, *2* (2), 15222. https://doi.org/10.1038/nplants.2015.222.
(40) Bewick, A. J.; Ji, L.; Niederhuth, C. E.; Willing, E.-M.; Hofmeister, B. T.; Shi, X.; Wang, L.; Lu, Z.; Rohr, N. A.; Hartwig, B.; Kiefer, C.; Deal, R. B.; Schmutz, J.; Grimwood, J.; Stroud, H.; Jacobsen, S. E.; Schneeberger, K.; Zhang, X.; Schmitz, R. J. On the Origin and Evolutionary Consequences of Gene Body DNA Methylation. *Proc. Natl. Acad. Sci.* **2016**, *113* (32), 9111–9116. https://doi.org/10.1073/pnas.1604666113.
(41) Zhang, H.; Lang, Z.; Zhu, J.-K. Dynamics and Function of DNA Methylation in Plants. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (8), 489–506. https://doi.org/10.1038/s41580-018-0016-z.
(42) Lippman, Z.; Gendrel, A.-V.; Black, M.; Vaughn, M. W.; Dedhia, N.; Richard McCombie, W.; Lavine, K.; Mittal, V.; May, B.; Kasschau, K. D.; Carrington, J. C.; Doerge, R. W.; Colot, V.; Martienssen, R. Role of Transposable Elements in Heterochromatin and Epigenetic Control. *Nature* **2004**, *430* (6998), 471–476. https://doi.org/10.1038/nature02651.
(43) Zakrzewski, F.; Schmidt, M.; Van Lijsebettens, M.; Schmidt, T. DNA Methylation of Retrotransposons, DNA Transposons and Genes in Sugar Beet (*Beta Vulgaris* L.). *Plant J.* **2017**, *90* (6), 1156–1175. https://doi.org/10.1111/tpj.13526.
(44) Gent, J. I.; Ellis, N. A.; Guo, L.; Harkess, A. E.; Yao, Y.; Zhang, X.; Dawe, R. K. CHH Islands: De Novo DNA Methylation in near-Gene Chromatin Regulation in Maize. *Genome Res.* **2013**, *23* (4), 628–637. https://doi.org/10.1101/gr.146985.112.
(45) Zhong, S.; Fei, Z.; Chen, Y.-R.; Zheng, Y.; Huang, M.; Vrebalov, J.; McQuinn, R.; Gapper, N.; Liu, B.; Xiang, J.; Shao, Y.; Giovannoni, J. J. Single-Base Resolution Methylomes of Tomato Fruit Development Reveal Epigenome Modifications Associated with Ripening. *Nat. Biotechnol.* **2013**, *31* (2), 154–159. https://doi.org/10.1038/nbt.2462.
(46) Choi, J. Y.; Purugganan, M. D. Evolutionary Epigenomics of Retrotransposon-Mediated Methylation Spreading in Rice. *Mol. Biol. Evol.* **2018**, *35* (2), 365–382. https://doi.org/10.1093/molbev/msx284.
(47) Hyun, K.; Jeon, J.; Park, K.; Kim, J. Writing, Erasing and Reading Histone Lysine Methylations. *Exp. Mol. Med.* **2017**, *49* (4), e324–e324. https://doi.org/10.1038/emm.2017.11.
(48) Gent, J. I.; Madzima, T. F.; Bader, R.; Kent, M. R.; Zhang, X.; Stam, M.; McGinnis, K. M.; Dawe, R. K. Accessible DNA and Relative Depletion of H3K9me2 at Maize Loci Undergoing RNA-Directed DNA Methylation. *Plant Cell Online* **2014**, *26* (12), 4903–4917. https://doi.org/10.1105/tpc.114.130427.
(49) Fischer, A.; Hofmann, I.; Naumann, K.; Reuter, G. Heterochromatin Proteins and the Control of Heterochromatic Gene Silencing in Arabidopsis. *J. Plant Physiol.* **2006**, *163* (3), 358–368. https://doi.org/10.1016/j.jplph.2005.10.015.
(50) Tse, C.; Sera, T.; Wolffe, A. P.; Hansen, J. C. Disruption of Higher-Order Folding by Core Histone Acetylation Dramatically Enhances Transcription of Nucleosomal Arrays by RNA Polymerase III. *Mol. Cell. Biol.* **1998**, *18* (8), 4629–4638. https://doi.org/10.1128/MCB.18.8.4629.
(51) Martin, C.; Zhang, Y. The Diverse Functions of Histone Lysine Methylation. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (11), 838–849. https://doi.org/10.1038/nrm1761.

(52) Zhang, X.; Clarenz, O.; Cokus, S.; Bernatavichute, Y. V.; Pellegrini, M.; Goodrich, J.; Jacobsen, S. E. Whole-Genome Analysis of Histone H3 Lysine 27 Trimethylation in Arabidopsis. *PLoS Biol.* **2007**, *5* (5), e129. https://doi.org/10.1371/journal.pbio.0050129.
(53) Roudier, F.; Ahmed, I.; Bérard, C.; Sarazin, A.; Mary-Huard, T.; Cortijo, S.; Bouyer, D.; Caillieux, E.; Duvernois-Berthet, E.; Al-Shikhley, L.; Giraut, L.; Després, B.; Drevensek, S.; Barneche, F.; Dèrozier, S.; Brunaud, V.; Aubourg, S.; Schnittger, A.; Bowler, C.; Martin-Magniette, M.-L.; Robin, S.; Caboche, M.; Colot, V. Integrative Epigenomic Mapping Defines Four Main Chromatin States in Arabidopsis: Organization of the Arabidopsis Epigenome. *EMBO J.* **2011**, *30* (10), 1928–1938. https://doi.org/10.1038/emboj.2011.103.
(54) Feng, J.; Shen, W.-H. Dynamic Regulation and Function of Histone Monoubiquitination in Plants. *Front. Plant Sci.* **2014**, *5*. https://doi.org/10.3389/fpls.2014.00083.
(55) Sridhar, V. V.; Kapoor, A.; Zhang, K.; Zhu, J.; Zhou, T.; Hasegawa, P. M.; Bressan, R. A.; Zhu, J.-K. Control of DNA Methylation and Heterochromatic Silencing by Histone H2B Deubiquitination. *Nature* **2007**, *447* (7145), 735–738. https://doi.org/10.1038/nature05864.
(56) Zhou, J.; Wang, X.; He, K.; Charron, J.-B. F.; Elling, A. A.; Deng, X. W. Genome-Wide Profiling of Histone H3 Lysine 9 Acetylation and Dimethylation in Arabidopsis Reveals Correlation between Multiple Histone Marks and Gene Expression. *Plant Mol. Biol.* **2010**, *72* (6), 585–595. https://doi.org/10.1007/s11103-009-9594-7.
(57) Finnegan, E. J.; Peacock, W. J.; Dennis, E. S. Reduced DNA Methylation in Arabidopsis Thaliana Results in Abnormal Plant Development. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (16), 8449–8454.
(58) Ronemus, M. J.; Galbiati, M.; Ticknor, C.; Chen, J.; Dellaporta, S. L. Demethylation-Induced Developmental Pleiotropy in Arabidopsis. *Science* **1996**, *273* (5275), 654–657. https://doi.org/10.1126/science.273.5275.654.
(59) Kankel, M. W.; Ramsey, D. E.; Stokes, T. L.; Flowers, S. K.; Haag, J. R.; Jeddeloh, J. A.; Riddle, N. C.; Verbsky, M. L.; Richards, E. J. Arabidopsis MET1 Cytosine Methyltransferase Mutants. *Genetics* **2003**, *163* (3), 1109–1122.
(60) Zubko, E.; Gentry, M.; Kunova, A.; Meyer, P. De Novo DNA Methylation Activity of METHYLTRANSFERASE 1 (MET1) Partially Restores Body Methylation in Arabidopsis Thaliana: MET1 Body Methylation. *Plant J.* **2012**, *71* (6), 1029–1037. https://doi.org/10.1111/j.1365-313X.2012.05051.x.
(61) Liu, X.; Yu, C.-W.; Duan, J.; Luo, M.; Wang, K.; Tian, G.; Cui, Y.; Wu, K. HDA6 Directly Interacts with DNA Methyltransferase MET1 and Maintains Transposable Element Silencing in Arabidopsis. *Plant Physiol.* **2012**, *158* (1), 119–129. https://doi.org/10.1104/pp.111.184275.
(62) Lindroth, A. M. Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. *Science* **2001**, *292* (5524), 2077–2080. https://doi.org/10.1126/science.1059745.
(63) Malagnac, F. An Arabidopsis SET Domain Protein Required for Maintenance but Not Establishment of DNA Methylation. *EMBO J.* **2002**, *21* (24), 6842–6852. https://doi.org/10.1093/emboj/cdf687.
(64) Du, J.; Johnson, L. M.; Groth, M.; Feng, S.; Hale, C. J.; Li, S.; Vashisht, A. A.; Gallego-Bartolome, J.; Wohlschlegel, J. A.; Patel, D. J.; Jacobsen, S. E. Mechanism of DNA

Methylation-Directed Histone Methylation by KRYPTONITE. *Mol. Cell* **2014**, *55* (3), 495–504. https://doi.org/10.1016/j.molcel.2014.06.009.

(65) Inagaki, S.; Miura-Kamio, A.; Nakamura, Y.; Lu, F.; Cui, X.; Cao, X.; Kimura, H.; Saze, H.; Kakutani, T. Autocatalytic Differentiation of Epigenetic Modifications within the Arabidopsis Genome. *EMBO J.* **2010**, *29* (20), 3496–3506. https://doi.org/10.1038/emboj.2010.227.

(66) Bernatavichute, Y. V.; Zhang, X.; Cokus, S.; Pellegrini, M.; Jacobsen, S. E. Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in Arabidopsis Thaliana. *PLOS One* **2008**, *3* (9). https://doi.org/10.1371/journal.pone.0003156.

(67) Du, J.; Zhong, X.; Bernatavichute, Y. V.; Stroud, H.; Feng, S.; Caro, E.; Vashisht, A. A.; Terragni, J.; Chin, H. G.; Tu, A.; Hetzel, J.; Wohlschlegel, J. A.; Pradhan, S.; Patel, D. J.; Jacobsen, S. E. Dual Binding of Chromomethylase Domains to H3K9me2-Containing Nucleosomes Directs DNA Methylation in Plants. *Cell* **2012**, *151* (1), 167–180. https://doi.org/10.1016/j.cell.2012.07.034.

(68) Johnson, L. M.; Bostick, M.; Zhang, X.; Kraft, E.; Henderson, I.; Callis, J.; Jacobsen, S. E. The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. *Curr. Biol.* **2007**, *17* (4), 379–384. https://doi.org/10.1016/j.cub.2007.01.009.

(69) Veiseth, S. V.; Rahman, M. A.; Yap, K. L.; Fischer, A.; Egge-Jacobsen, W.; Reuter, G.; Zhou, M.-M.; Aalen, R. B.; Thorstensen, T. The SUVR4 Histone Lysine Methyltransferase Binds Ubiquitin and Converts H3K9me1 to H3K9me3 on Transposon Chromatin in Arabidopsis. *PLoS Genet.* **2011**, *7* (3), e1001325. https://doi.org/10.1371/journal.pgen.1001325.

(70) Zemach, A.; Kim, M. Y.; Hsieh, P.-H.; Coleman-Derr, D.; Eshed-Williams, L.; Thao, K.; Harmer, S. L.; Zilberman, D. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell* **2013**, *153* (1), 193–205. https://doi.org/10.1016/j.cell.2013.02.033.

(71) Parrilla-Doblas, J. T.; Roldán-Arjona, T.; Ariza, R. R.; Córdoba-Cañero, D. Active DNA Demethylation in Plants. *Int. J. Mol. Sci.* **2019**, *20* (19), 4683. https://doi.org/10.3390/ijms20194683.

(72) Calarco, J. P.; Borges, F.; Donoghue, M. T. A.; Van Ex, F.; Jullien, P. E.; Lopes, T.; Gardner, R.; Berger, F.; Feijó, J. A.; Becker, J. D.; Martienssen, R. A. Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. *Cell* **2012**, *151* (1), 194–205. https://doi.org/10.1016/j.cell.2012.09.001.

(73) Morales-Ruiz, T.; Ortega-Galisteo, A. P.; Ponferrada-Marin, M. I.; Martinez-Macias, M. I.; Ariza, R. R.; Roldan-Arjona, T. DEMETER and REPRESSOR OF SILENCING 1 Encode 5-Methylcytosine DNA Glycosylases. *Proc. Natl. Acad. Sci.* **2006**, *103* (18), 6853–6858. https://doi.org/10.1073/pnas.0601109103.

(74) Ortega-Galisteo, A. P.; Morales-Ruiz, T.; Ariza, R. R.; Roldán-Arjona, T. Arabidopsis DEMETER-LIKE Proteins DML2 and DML3 Are Required for Appropriate Distribution of DNA Methylation Marks. *Plant Mol. Biol.* **2008**, *67* (6), 671–681. https://doi.org/10.1007/s11103-008-9346-0.

(75) Gehring, M.; Huh, J. H.; Hsieh, T.-F.; Penterman, J.; Choi, Y.; Harada, J. J.; Goldberg, R. B.; Fischer, R. L. DEMETER DNA Glycosylase Establishes MEDEA Polycomb Gene Self-Imprinting by Allele-Specific Demethylation. *Cell* **2006**, *124* (3), 495–506. https://doi.org/10.1016/j.cell.2005.12.034.

(76) Tang, K.; Lang, Z.; Zhang, H.; Zhu, J.-K. The DNA Demethylase ROS1 Targets Genomic Regions with Distinct Chromatin Modifications. *Nat. Plants* **2016**, *2* (11), 16169. https://doi.org/10.1038/nplants.2016.169.
(77) Córdoba-Cañero, D.; Roldán-Arjona, T.; Ariza, R. R. Arabidopsis ZDP DNA 3′-Phosphatase and ARP Endonuclease Function in 8-OxoG Repair Initiated by FPG and OGG1 DNA Glycosylases. *Plant J.* **2014**, *79* (5), 824–834. https://doi.org/10.1111/tpj.12588.
(78) Li, Y.; Córdoba-Cañero, D.; Qian, W.; Zhu, X.; Tang, K.; Zhang, H.; Ariza, R. R.; Roldán-Arjona, T.; Zhu, J.-K. An AP Endonuclease Functions in Active DNA Demethylation and Gene Imprinting in Arabidopsis. *PLoS Genet.* **2015**, *11* (1), e1004905. https://doi.org/10.1371/journal.pgen.1004905.
(79) Córdoba-Cañero, D.; Roldán-Arjona, T.; Ariza, R. R. Arabidopsis ARP Endonuclease Functions in a Branched Base Excision DNA Repair Pathway Completed by LIG1: Roles of ARP and LIG1 in Base Excision Repair. *Plant J.* **2011**, *68* (4), 693–702. https://doi.org/10.1111/j.1365-313X.2011.04720.x.
(80) Çağlayan, M. Pol β Gap Filling, DNA Ligation and Substrate-Product Channeling during Base Excision Repair Opposite Oxidized 5-Methylcytosine Modifications. *DNA Repair* **2020**, *95*, 102945. https://doi.org/10.1016/j.dnarep.2020.102945.
(81) Li, Y.; Kumar, S.; Qian, W. Active DNA Demethylation: Mechanism and Role in Plant Development. *Plant Cell Rep.* **2018**, *37* (1), 77–85. https://doi.org/10.1007/s00299-017-2215-z.
(82) Onodera, Y.; Haag, J. R.; Ream, T.; Nunes, P. C.; Pontes, O.; Pikaard, C. S. Plant Nuclear RNA Polymerase IV Mediates SiRNA and DNA Methylation-Dependent Heterochromatin Formation. *Cell* **2005**, *120* (5), 613–622. https://doi.org/10.1016/j.cell.2005.02.007.
(83) Law, J. A.; Vashisht, A. A.; Wohlschlegel, J. A.; Jacobsen, S. E. SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV. *PLoS Genet.* **2011**, *7* (7), e1002195. https://doi.org/10.1371/journal.pgen.1002195.
(84) Zhou, M.; Palanca, A. M. S.; Law, J. A. Locus-Specific Control of the de Novo DNA Methylation Pathway in Arabidopsis by the CLASSY Family. *Nat. Genet.* **2018**, *50* (6), 865–873. https://doi.org/10.1038/s41588-018-0115-y.
(85) Law, J. A.; Du, J.; Hale, C. J.; Feng, S.; Krajewski, K.; Palanca, A. M. S.; Strahl, B. D.; Patel, D. J.; Jacobsen, S. E. Polymerase IV Occupancy at RNA-Directed DNA Methylation Sites Requires SHH1. *Nature* **2013**, *498* (7454), 385–389. https://doi.org/10.1038/nature12178.
(86) Singh, J.; Mishra, V.; Wang, F.; Huang, H.-Y.; Pikaard, C. S. Reaction Mechanisms of Pol IV, RDR2, and DCL3 Drive RNA Channeling in the SiRNA-Directed DNA Methylation Pathway. *Mol. Cell* **2019**, *75* (3), 576-589.e5. https://doi.org/10.1016/j.molcel.2019.07.008.
(87) Nuthikattu, S.; McCue, A. D.; Panda, K.; Fultz, D.; DeFraia, C.; Thomas, E. N.; Slotkin, R. K. The Initiation of Epigenetic Silencing of Active Transposable Elements Is Triggered by RDR6 and 21-22 Nucleotide Small Interfering RNAs. *Plant Physiol.* **2013**, *162* (1), 116–131. https://doi.org/10.1104/pp.113.216481.
(88) McCue, A. D.; Nuthikattu, S.; Reeder, S. H.; Slotkin, R. K. Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. *PLoS Genet.* **2012**, *8* (2), e1002474. https://doi.org/10.1371/journal.pgen.1002474.

(89) Yang, D.-L.; Zhang, G.; Tang, K.; Li, J.; Yang, L.; Huang, H.; Zhang, H.; Zhu, J.-K. Dicer-Independent RNA-Directed DNA Methylation in Arabidopsis. *Cell Res.* **2016**, *26* (1), 66–82. https://doi.org/10.1038/cr.2015.145.
(90) Mallory, A.; Vaucheret, H. Form, Function, and Regulation of ARGONAUTE Proteins. *Plant Cell* **2010**, *22* (12), 3879–3889. https://doi.org/10.1105/tpc.110.080671.
(91) Wierzbicki, A. T.; Ream, T. S.; Haag, J. R.; Pikaard, C. S. RNA Polymerase V Transcription Guides ARGONAUTE4 to Chromatin. *Nat. Genet.* **2009**, *41* (5), 630–634. https://doi.org/10.1038/ng.365.
(92) Law, J. A.; Ausin, I.; Johnson, L. M.; Vashisht, A. A.; Zhu, J.-K.; Wohlschlegel, J. A.; Jacobsen, S. E. A Protein Complex Required for Polymerase V Transcripts and RNA-Directed DNA Methylation in Arabidopsis. *Curr. Biol.* **2010**, *20* (10), 951–956. https://doi.org/10.1016/j.cub.2010.03.062.
(93) Kanno, T.; Bucher, E.; Daxinger, L.; Huettel, B.; Böhmdorfer, G.; Gregor, W.; Kreil, D. P.; Matzke, M.; Matzke, A. J. M. A Structural-Maintenance-of-Chromosomes Hinge Domain–Containing Protein Is Required for RNA-Directed DNA Methylation. *Nat. Genet.* **2008**, *40* (5), 670–675. https://doi.org/10.1038/ng.119.
(94) Johnson, L. M.; Du, J.; Hale, C. J.; Bischof, S.; Feng, S.; Chodavarapu, R. K.; Zhong, X.; Marson, G.; Pellegrini, M.; Segal, D. J.; Patel, D. J.; Jacobsen, S. E. SRA- and SET-Domain-Containing Proteins Link RNA Polymerase V Occupancy to DNA Methylation. *Nature* **2014**, *507* (7490), 124–128. https://doi.org/10.1038/nature12931.
(95) Zhang, H.; Tang, K.; Qian, W.; Duan, C.-G.; Wang, B.; Zhang, H.; Wang, P.; Zhu, X.; Lang, Z.; Yang, Y.; Zhu, J.-K. An Rrp6-like Protein Positively Regulates Noncoding RNA Levels and DNA Methylation in Arabidopsis. *Mol. Cell* **2014**, *54* (3), 418–430. https://doi.org/10.1016/j.molcel.2014.03.019.
(96) Zhu, Y.; Rowley, M. J.; Böhmdorfer, G.; Wierzbicki, A. T. A SWI/SNF Chromatin-Remodeling Complex Acts in Noncoding RNA-Mediated Transcriptional Silencing. *Mol. Cell* **2013**, *49* (2), 298–309. https://doi.org/10.1016/j.molcel.2012.11.011.
(97) Zhong, X.; Du, J.; Hale, C. J.; Gallego-Bartolome, J.; Feng, S.; Vashisht, A. A.; Chory, J.; Wohlschlegel, J. A.; Patel, D. J.; Jacobsen, S. E. Molecular Mechanism of Action of Plant DRM De Novo DNA Methyltransferases. *Cell* **2014**, *157* (5), 1050–1060. https://doi.org/10.1016/j.cell.2014.03.056.
(98) Lee, R. C.; Feinbaum, R. L.; Ambros, V. The C. Elegans Heterochronic Gene Lin-4 Encodes Small RNAs with Antisense Complementarity to Lin-14. *Cell* **1993**, *75* (5), 843–854. https://doi.org/10.1016/0092-8674(93)90529-Y.
(99) Wightman, B.; Ha, I.; Ruvkun, G. Posttranscriptional Regulation of the Heterochronic Gene Lin-14 by Lin-4 Mediates Temporal Pattern Formation in C. Elegans. *Cell* **1993**, *75* (5), 855–862. https://doi.org/10.1016/0092-8674(93)90530-4.
(100) Filipowicz, W.; Bhattacharyya, S. N.; Sonenberg, N. Mechanisms of Post-Transcriptional Regulation by MicroRNAs: Are the Answers in Sight? *Nat. Rev. Genet.* **2008**, *9* (2), 102–114. https://doi.org/10.1038/nrg2290.
(101) Elbashir, S. M. RNA Interference Is Mediated by 21- and 22-Nucleotide RNAs. *Genes Dev.* **2001**, *15* (2), 188–200. https://doi.org/10.1101/gad.862301.
(102) Hammond, S. M.; Bernstein, E.; Beach, D.; Hannon, G. J. An RNA-Directed Nuclease Mediates Post-Transcriptional Gene Silencing in Drosophila Cells. *Nature* **2000**, *404* (6775), 293–296. https://doi.org/10.1038/35005107.

(103) Zhang, R.; Jing, Y.; Zhang, H.; Niu, Y.; Liu, C.; Wang, J.; Zen, K.; Zhang, C.-Y.; Li, D. Comprehensive Evolutionary Analysis of the Major RNA-Induced Silencing Complex Members. *Sci. Rep.* **2018**, *8* (1), 14189. https://doi.org/10.1038/s41598-018-32635-4.
(104) Hutvagner, G. A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the Let-7 Small Temporal RNA. *Science* **2001**, *293* (5531), 834–838. https://doi.org/10.1126/science.1062961.
(105) Meister, G.; Landthaler, M.; Patkaniowska, A.; Dorsett, Y.; Teng, G.; Tuschl, T. Human Argonaute2 Mediates RNA Cleavage Targeted by MiRNAs and SiRNAs. *Mol. Cell* **2004**, *15* (2), 185–197. https://doi.org/10.1016/j.molcel.2004.07.007.
(106) Martinez, J.; Patkaniowska, A.; Urlaub, H.; Lührmann, R.; Tuschl, T. Single-Stranded Antisense SiRNAs Guide Target RNA Cleavage in RNAi. *Cell* **2002**, *110* (5), 563–574. https://doi.org/10.1016/S0092-8674(02)00908-X.
(107) Joga, M. R.; Zotti, M. J.; Smagghe, G.; Christiaens, O. RNAi Efficiency, Systemic Properties, and Novel Delivery Methods for Pest Insect Control: What We Know So Far. *Front. Physiol.* **2016**, *7*. https://doi.org/10.3389/fphys.2016.00553.
(108) Henderson, I. R.; Zhang, X.; Lu, C.; Johnson, L.; Meyers, B. C.; Green, P. J.; Jacobsen, S. E. Dissecting Arabidopsis Thaliana DICER Function in Small RNA Processing, Gene Silencing and DNA Methylation Patterning. *Nat. Genet.* **2006**, *38* (6), 721–725. https://doi.org/10.1038/ng1804.
(109) Qu, F.; Ye, X.; Morris, T. J. Arabidopsis DRB4, AGO1, AGO7, and RDR6 Participate in a DCL4-Initiated Antiviral RNA Silencing Pathway Negatively Regulated by DCL1. *Proc. Natl. Acad. Sci.* **2008**, *105* (38), 14732–14737. https://doi.org/10.1073/pnas.0805760105.
(110) Han, M.-H.; Goud, S.; Song, L.; Fedoroff, N. The Arabidopsis Double-Stranded RNA-Binding Protein HYL1 Plays a Role in MicroRNA-Mediated Gene Regulation. *Proc. Natl. Acad. Sci.* **2004**, *101* (4), 1093–1098. https://doi.org/10.1073/pnas.0307969100.
(111) Yang, L.; Liu, Z.; Lu, F.; Dong, A.; Huang, H. SERRATE Is a Novel Nuclear Regulator in Primary MicroRNA Processing in Arabidopsis. *Plant J.* **2006**, *47* (6), 841–850. https://doi.org/10.1111/j.1365-313X.2006.02835.x.
(112) Dong, Z.; Han, M.-H.; Fedoroff, N. The RNA-Binding Proteins HYL1 and SE Promote Accurate in Vitro Processing of Pri-MiRNA by DCL1. *Proc. Natl. Acad. Sci.* **2008**, *105* (29), 9970–9975. https://doi.org/10.1073/pnas.0803356105.
(113) Zong, J.; Yao, X.; Yin, J.; Zhang, D.; Ma, H. Evolution of the RNA-Dependent RNA Polymerase (RdRP) Genes: Duplications and Possible Losses before and after the Divergence of Major Eukaryotic Groups. *Gene* **2009**, *447* (1), 29–39. https://doi.org/10.1016/j.gene.2009.07.004.
(114) Wang, X.-B.; Wu, Q.; Ito, T.; Cillo, F.; Li, W.-X.; Chen, X.; Yu, J.-L.; Ding, S.-W. RNAi-Mediated Viral Immunity Requires Amplification of Virus-Derived SiRNAs in Arabidopsis Thaliana. *Proc. Natl. Acad. Sci.* **2010**, *107* (1), 484–489. https://doi.org/10.1073/pnas.0904086107.
(115) Xie, Z.; Johansen, L. K.; Gustafson, A. M.; Kasschau, K. D.; Lellis, A. D.; Zilberman, D.; Jacobsen, S. E.; Carrington, J. C. Genetic and Functional Diversification of Small RNA Pathways in Plants. *PLoS Biol.* **2004**, *2* (5), e104. https://doi.org/10.1371/journal.pbio.0020104.
(116) Lu, C.; Kulkarni, K.; Souret, F. F.; MuthuValliappan, R.; Tej, S. S.; Poethig, R. S.; Henderson, I. R.; Jacobsen, S. E.; Wang, W.; Green, P. J.; Meyers, B. C. MicroRNAs and

Other Small RNAs Enriched in the Arabidopsis RNA-Dependent RNA Polymerase-2 Mutant. *Genome Res.* **2006**, *16* (10), 1276–1288. https://doi.org/10.1101/gr.5530106.
(117) Peragine, A. SGS3 and SGS2/SDE1/RDR6 Are Required for Juvenile Development and the Production of Trans-Acting SiRNAs in Arabidopsis. *Genes Dev.* **2004**, *18* (19), 2368–2379. https://doi.org/10.1101/gad.1231804.
(118) Gandikota, M.; Birkenbihl, R. P.; Höhmann, S.; Cardon, G. H.; Saedler, H.; Huijser, P. The MiRNA156/157 Recognition Element in the 3′ UTR of the Arabidopsis SBP Box Gene SPL3 Prevents Early Flowering by Translational Inhibition in Seedlings: MiRNA Regulation of SPL3. *Plant J.* **2007**, *49* (4), 683–693. https://doi.org/10.1111/j.1365-313X.2006.02983.x.
(119) Axtell, M. J.; Westholm, J. O.; Lai, E. C. Vive La Différence: Biogenesis and Evolution of MicroRNAs in Plants and Animals. *Genome Biol.* **2011**, *12* (4), 221. https://doi.org/10.1186/gb-2011-12-4-221.
(120) Franco-Zorrilla, J. M.; Valli, A.; Todesco, M.; Mateos, I.; Puga, M. I.; Rubio-Somoza, I.; Leyva, A.; Weigel, D.; García, J. A.; Paz-Ares, J. Target Mimicry Provides a New Mechanism for Regulation of MicroRNA Activity. *Nat. Genet.* **2007**, *39* (8), 1033–1037. https://doi.org/10.1038/ng2079.
(121) Rehwinkel, J. A Crucial Role for GW182 and the DCP1:DCP2 Decapping Complex in MiRNA-Mediated Gene Silencing. *RNA* **2005**, *11* (11), 1640–1647. https://doi.org/10.1261/rna.2191905.
(122) Braun, J. E.; Huntzinger, E.; Izaurralde, E. The Role of GW182 Proteins in MiRNA-Mediated Gene Silencing. In *Ten Years of Progress in GW/P Body Research*; Chan, E. K. L., Fritzler, M. J., Eds.; Advances in Experimental Medicine and Biology; Springer New York: New York, NY, 2013; Vol. 768, pp 147–163. https://doi.org/10.1007/978-1-4614-5107-5_9.
(123) Brodersen, P.; Sakvarelidze-Achard, L.; Bruun-Rasmussen, M.; Dunoyer, P.; Yamamoto, Y. Y.; Sieburth, L.; Voinnet, O. Widespread Translational Inhibition by Plant MiRNAs and SiRNAs. *Science* **2008**, *320* (5880), 1185–1190. https://doi.org/10.1126/science.1159151.
(124) Li, S.; Liu, L.; Zhuang, X.; Yu, Y.; Liu, X.; Cui, X.; Ji, L.; Pan, Z.; Cao, X.; Mo, B.; Zhang, F.; Raikhel, N.; Jiang, L.; Chen, X. MicroRNAs Inhibit the Translation of Target MRNAs on the Endoplasmic Reticulum in Arabidopsis. *Cell* **2013**, *153* (3), 562–574. https://doi.org/10.1016/j.cell.2013.04.005.
(125) Axtell, M. J. Classification and Comparison of Small RNAs from Plants. *Annu. Rev. Plant Biol.* **2013**, *64* (1), 137–159. https://doi.org/10.1146/annurev-arplant-050312-120043.
(126) Wang, J.; Mei, J.; Ren, G. Plant MicroRNAs: Biogenesis, Homeostasis, and Degradation. *Front. Plant Sci.* **2019**, *10*, 360. https://doi.org/10.3389/fpls.2019.00360.
(127) Mack, G. S. MicroRNA Gets down to Business. *Nat. Biotechnol.* **2007**, *25* (6), 631–638. https://doi.org/10.1038/nbt0607-631.
(128) Ebhardt, H. A.; Tsang, H. H.; Dai, D. C.; Liu, Y.; Bostan, B.; Fahlman, R. P. Meta-Analysis of Small RNA-Sequencing Errors Reveals Ubiquitous Post-Transcriptional RNA Modifications. *Nucleic Acids Res.* **2009**, *37* (8), 2461–2470. https://doi.org/10.1093/nar/gkp093.
(129) Dunn, D. B. The Occurence of 1-Methyladenine in Ribonucleic Acid. *Biochim. Biophys. Acta* **1961**, *46* (1), 198–200. https://doi.org/10.1016/0006-3002(61)90668-0.

(130) Limbach, P. A.; Crain, P. F.; McCloskey, J. A. Summary: The Modified Nucleosides of RNA. *Nucleic Acids Res.* **1994**, *22* (12), 2183–2196. https://doi.org/10.1093/nar/22.12.2183.
(131) Axtell, M. J.; Jan, C.; Rajagopalan, R.; Bartel, D. P. A Two-Hit Trigger for SiRNA Biogenesis in Plants. *Cell* **2006**, *127* (3), 565–577. https://doi.org/10.1016/j.cell.2006.09.032.
(132) Borges, F.; Martienssen, R. A. The Expanding World of Small RNAs in Plants. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (12), 727–741. https://doi.org/10.1038/nrm4085.
(133) Li, S.; Xu, Z.; Sheng, J. TRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. *Genes* **2018**, *9* (5), 246. https://doi.org/10.3390/genes9050246.
(134) Chen, Z.; Sun, Y.; Yang, X.; Wu, Z.; Guo, K.; Niu, X.; Wang, Q.; Ruan, J.; Bu, W.; Gao, S. Two Featured Series of RRNA-Derived RNA Fragments (RRFs) Constitute a Novel Class of Small RNAs. *PLOS ONE* **2017**, *12* (4), e0176458. https://doi.org/10.1371/journal.pone.0176458.
(135) Morgado, L.; Johannes, F. Computational Tools for Plant Small RNA Detection and Categorization. *Brief. Bioinform.* **2019**, *20* (4), 1181–1192. https://doi.org/10.1093/bib/bbx136.
(136) Matzke, M. A.; Mosher, R. A. RNA-Directed DNA Methylation: An Epigenetic Pathway of Increasing Complexity. *Nat. Rev. Genet.* **2014**, *15* (6), 394–408. https://doi.org/10.1038/nrg3683.
(137) Zheng, K.; Wang, L.; Zeng, L.; Xu, D.; Guo, Z.; Gao, X.; Yang, D.-L. The Effect of RNA Polymerase V on 24-Nt SiRNA Accumulation Depends on DNA Methylation Contexts and Histone Modifications in Rice. *Proc. Natl. Acad. Sci.* **2021**, *118* (30), e2100709118. https://doi.org/10.1073/pnas.2100709118.
(138) Li, J.; Yang, D.-L.; Huang, H.; Zhang, G.; He, L.; Pang, J.; Lozano-Durán, R.; Lang, Z.; Zhu, J.-K. Epigenetic Memory Marks Determine Epiallele Stability at Loci Targeted by de Novo DNA Methylation. *Nat. Plants* **2020**, *6* (6), 661–674. https://doi.org/10.1038/s41477-020-0671-x.
(139) Guo, W.; Wang, D.; Lisch, D. RNA-Directed DNA Methylation Prevents Rapid and Heritable Reversal of Transposon Silencing under Heat Stress in Zea Mays. *PLOS Genet.* **2021**, *17* (6), e1009326. https://doi.org/10.1371/journal.pgen.1009326.
(140) Boyko, A.; Blevins, T.; Yao, Y.; Golubov, A.; Bilichak, A.; Ilnytskyy, Y.; Hollander, J.; Meins, F.; Kovalchuk, I. Transgenerational Adaptation of Arabidopsis to Stress Requires DNA Methylation and the Function of Dicer-Like Proteins. *PLoS ONE* **2010**, *5* (3), e9514. https://doi.org/10.1371/journal.pone.0009514.
(141) Tricker, P. J.; Gibbings, J. G.; Rodríguez López, C. M.; Hadley, P.; Wilkinson, M. J. Low Relative Humidity Triggers RNA-Directed de Novo DNA Methylation and Suppression of Genes Controlling Stomatal Development. *J. Exp. Bot.* **2012**, *63* (10), 3799–3813. https://doi.org/10.1093/jxb/ers076.
(142) Tricker, P.; López, C.; Gibbings, G.; Hadley, P.; Wilkinson, M. Transgenerational, Dynamic Methylation of Stomata Genes in Response to Low Relative Humidity. *Int. J. Mol. Sci.* **2013**, *14* (4), 6674–6689. https://doi.org/10.3390/ijms14046674.
(143) Swiezewski, S.; Crevillen, P.; Liu, F.; Ecker, J. R.; Jerzmanowski, A.; Dean, C. Small RNA-Mediated Chromatin Silencing Directed to the 3' Region of the Arabidopsis Gene Encoding the Developmental Regulator, FLC. *Proc. Natl. Acad. Sci.* **2007**, *104* (9), 3633–3638. https://doi.org/10.1073/pnas.0611459104.

(144) Kim, J.-M.; To, T. K.; Ishida, J.; Matsui, A.; Kimura, H.; Seki, M. Transition of Chromatin Status During the Process of Recovery from Drought Stress in Arabidopsis Thaliana. *Plant Cell Physiol.* **2012**, *53* (5), 847–856. https://doi.org/10.1093/pcp/pcs053.
(145) Wang, Y. H.; Irving, H. R. Developing a Model of Plant Hormone Interactions. *Plant Signal. Behav.* **2011**, *6* (4), 494–500. https://doi.org/10.4161/psb.6.4.14558.
(146) Caarls, L.; Pieterse, C. M. J.; Van Wees, S. C. M. How Salicylic Acid Takes Transcriptional Control over Jasmonic Acid Signaling. *Front. Plant Sci.* **2015**, *6*. https://doi.org/10.3389/fpls.2015.00170.
(147) Zhou, C.; Zhang, L.; Duan, J.; Miki, B.; Wu, K. *HISTONE DEACETYLASE19* Is Involved in Jasmonic Acid and Ethylene Signaling of Pathogen Response in Arabidopsis. *Plant Cell* **2005**, *17* (4), 1196–1204. https://doi.org/10.1105/tpc.104.028514.
(148) Jin, H.; Choi, S.-M.; Kang, M.-J.; Yun, S.-H.; Kwon, D.-J.; Noh, Y.-S.; Noh, B. Salicylic Acid-Induced Transcriptional Reprogramming by the HAC–NPR1–TGA Histone Acetyltransferase Complex in Arabidopsis. *Nucleic Acids Res.* **2018**. https://doi.org/10.1093/nar/gky847.
(149) Sui, P.; Jin, J.; Ye, S.; Mu, C.; Gao, J.; Feng, H.; Shen, W.-H.; Yu, Y.; Dong, A. H3K36 Methylation Is Critical for Brassinosteroid-Regulated Plant Growth and Development in Rice: H3K36me in Brassinosteroid-Regulation Function. *Plant J.* **2012**, *70* (2), 340–347. https://doi.org/10.1111/j.1365-313X.2011.04873.x.
(150) *Brassinosteroids: Plant Growth and Development*; Hayat, S., Yusuf, M., Bhardwaj, R., Bajguz, A., Eds.; Springer: Singapore, 2019.
(151) Liao, K.; Peng, Y.-J.; Yuan, L.-B.; Dai, Y.-S.; Chen, Q.-F.; Yu, L.-J.; Bai, M.-Y.; Zhang, W.-Q.; Xie, L.-J.; Xiao, S. Brassinosteroids Antagonize Jasmonate-Activated Plant Defense Responses through BRI1-EMS-SUPPRESSOR1 (BES1). *Plant Physiol.* **2020**, *182* (2), 1066–1082. https://doi.org/10.1104/pp.19.01220.
(152) Chinnusamy, V.; Gong, Z.; Zhu, J.-K. Abscisic Acid-Mediated Epigenetic Processes in Plant Development and Stress Responses. *J. Integr. Plant Biol.* **2008**, *50* (10), 1187–1195. https://doi.org/10.1111/j.1744-7909.2008.00727.x.
(153) Sah, S. K.; Reddy, K. R.; Li, J. Abscisic Acid and Abiotic Stress Tolerance in Crop Plants. *Front. Plant Sci.* **2016**, *7*. https://doi.org/10.3389/fpls.2016.00571.
(154) Vendramin, S.; Huang, J.; Crisp, P. A.; Madzima, T. F.; McGinnis, K. M. Epigenetic Regulation of ABA-Induced Transcriptional Responses in Maize. *G3 GenesGenomesGenetics* **2020**, *10* (5), 1727–1743. https://doi.org/10.1534/g3.119.400993.
(155) Khraiwesh, B.; Arif, M. A.; Seumel, G. I.; Ossowski, S.; Weigel, D.; Reski, R.; Frank, W. Transcriptional Control of Gene Expression by MicroRNAs. *Cell* **2010**, *140* (1), 111–122. https://doi.org/10.1016/j.cell.2009.12.023.
(156) Forgione, I.; Wołoszyńska, M.; Pacenza, M.; Chiappetta, A.; Greco, M.; Araniti, F.; Abenavoli, M. R.; Van Lijsebettens, M.; Bitonti, M. B.; Bruno, L. Hypomethylated Drm1 Drm2 Cmt3 Mutant Phenotype of Arabidopsis Thaliana Is Related to Auxin Pathway Impairment. *Plant Sci.* **2019**, *280*, 383–396. https://doi.org/10.1016/j.plantsci.2018.12.029.
(157) Li, W.; Liu, H.; Cheng, Z. J.; Su, Y. H.; Han, H. N.; Zhang, Y.; Zhang, X. S. DNA Methylation and Histone Modifications Regulate De Novo Shoot Regeneration in Arabidopsis by Modulating WUSCHEL Expression and Auxin Signaling. *PLoS Genet.* **2011**, *7* (8), e1002243. https://doi.org/10.1371/journal.pgen.1002243.

(158) Yang, L.; Meng, D.; Wang, Y.; Wu, Y.; Lang, C.; Jin, T.; Zhou, X. The Viral Suppressor HCPro Decreases DNA Methylation and Activates Auxin Biosynthesis Genes. *Virology* **2020**, *546*, 133–140. https://doi.org/10.1016/j.virol.2020.04.003.
(159) Guo, Q.; Liu, Q.; A. Smith, N.; Liang, G.; Wang, M.-B. RNA Silencing in Plants: Mechanisms, Technologies and Applications in Horticultural Crops. *Curr. Genomics* **2016**, *17* (6), 476–489. https://doi.org/10.2174/1389202917666160520103117.
(160) Suza, W.; Lee, D. *Genetics, Agriculture, and Biotechnology*; Iowa State University Digital Press, 2021. https://doi.org/10.31274/isudp.2021.113.
(161) Baulcombe, D. C.; Dean, C. Epigenetic Regulation in Plant Responses to the Environment. *Cold Spring Harb. Perspect. Biol.* **2014**, *6* (9), a019471–a019471. https://doi.org/10.1101/cshperspect.a019471.
(162) Feng, S.; Jacobsen, S. E.; Reik, W. Epigenetic Reprogramming in Plant and Animal Development. *Science* **2010**, *330* (6004), 622–627. https://doi.org/10.1126/science.1190614.
(163) Délye, C.; Jasieniuk, M.; Le Corre, V. Deciphering the Evolution of Herbicide Resistance in Weeds. *Trends Genet.* **2013**, *29* (11), 649–658. https://doi.org/10.1016/j.tig.2013.06.001.

Chapter 2: RNA Sequence Analysis of Weed Management Stress Responses in *Arabidopsis thaliana*

Abstract:

Weeds are regularly exposed to severe stress in the form of control measures that may include herbicides, physical damage, and competition from crops. Although the immediate impact of stress on plant gene expression is well documented, we wanted to investigate plant responses that persisted after a period of recovery or adaptation to the stress. Young Arabidopsis plants were treated with sublethal doses of one of four stressors associated with weed control: glyphosate, trifloxysulfuron, shade, or mechanical clipping. After the plants had recovered from the stress to the point of flowering, RNA was extracted from newly formed leaves and sequenced to characterize gene expression relative to unstressed control plants. Glyphosate-treated plants showed an enrichment of up-regulated defense and stress response genes, and down-regulation of growth and photosynthesis genes. Glyphosate and trifloxysulfuron treated plants shared some upregulated genes, but overall, the gene expression patterns were different among the stresses. The shaded plants and mechanically clipped plants had reduced expression of many defense response genes, but different genes were involved. These findings of persistent changes in gene expression indicate that plants respond specifically to different stressors and maintain new expression patterns in a way that could prime them to resist a similar, future stress.

Introduction:

In 1943, the Cooperative Mexican Agricultural Program helped Mexico to become self-sufficient in wheat production and ushered in a period of agricultural innovation known as the Green Revolution.[1] Much research would follow to support the needs of an increasing population. New varieties of wheat, rice, and maize were bred to have several useful traits, including resistance to abiotic and biotic stressors.[2] Now, researchers are struggling to similarly increase production at a rate that will support the world's population in the coming decades. Erosion, climate change, and limited fresh water add pressure to many growers.[3] Herbicide resistance in weeds, which steal resources away from crops, is another threat to food production.[4] In this study we focus on stress responses linked with weed control measures, including herbicides, clipping, and shade. Better weed management could help present day farmers have higher yields to meet the food supply demands.

Before the introduction of selective herbicides, farmers removed weeds by hand or tilling. There were some toxic substances that were known to kill plants, such as arsenic and copper salts, but these were damaging to crop plants as well.[5] 2,4-D was the first commercially available synthetic herbicide in 1946.[6] Early in the days of synthetic herbicides, there were concerns about the development of resistance through either evolution of an individual species or inadvertently promoting the growth of the most resistant weedy species. Plowing and hand removal was recommended to supplement spraying. At the time, this labor was considered inexpensive.[7]

In 1957, the first specific case of herbicide resistance was reported in wild carrot for 2,4-D and related herbicides.[8],[9] Increased tolerance to herbicides was noted in the United Kingdom as early as 1954. The rotation of herbicides was recommended to help control this issue.[10] By 1969, there were about 120 synthetic herbicides with a variety of modes of action on the market.[11] Reports of herbicide resistance were limited and largely overshadowed by the threat of

insecticide resistance until a case of atrazine resistance in 1968.[8,12] Following this, resistance to triazine, ACCase-, and ALS-inhibiting herbicides was extensively studied, but the problems would intensify greatly with the introduction of herbicide tolerant crops that allowed growers to spray weeds after planting.[13,14]

Glyphosate tolerant soybeans were released to the market in 1996 and planted across millions of acres, but glyphosate resistant ryegrass was discovered the same year.[14,15,16] Nevertheless, glyphosate usage in the United States increased with the release of additional tolerant crops.[17] Although some other herbicide tolerant crops were available previously, they were not as popular as the Roundup Ready series.[18] Resistance to glyphosate has been identified in many other weeds, including palmer amaranth, hairy fleabane, horseweed, and ragweed.[4] Usage further increased to combat increasing thresholds of resistance. Its recent decline may be due to the increase in resistance in weeds becoming so great that Roundup Ready crops are less cost effective.[17]

Alternatives to herbicides for controlling weeds include tilling and manual weeding. The former disrupts soil structure and is associated with soil erosion, and the latter is not economically feasible.[19,20] Leaving weeds to grow unchecked in fields has caused yield losses of 53% in rice, 26% in Sorghum, 20% in corn, and 25% in wheat in the United States.[20,21] There must be a method to make herbicides work, or weed management itself will have to change to protect yields. Crops tolerant to other herbicides, such as 2,4-D, glufosinate, and dicamba, are available, which allows for a rotation of herbicides that can be applied while the crops are growing.[22] A greater understanding of how resistance develops in weeds could contribute to solutions.[23] Some have suggested there could be an epigenetic component to resistance.[24,25]

The present study is focused on *Arabidopsis thaliana*, which is both a model organism and an agricultural weed. The plants are relatively easy to grow, and many stress response studies have been performed previously that set a baseline for comparison.[26] Also, its genome is the best characterized and annotated of all plant species.[27] This study includes four stressors associated with weed management: shade; mechanical clipping; and the herbicides, glyphosate and trifloxysulfuron. Weeds can encounter shade stress while they attempt to grow beneath crop plants. Clipping may occur as farmers attempt to remove weeds through mowing or other tools that inflict mechanical injury. Knowing how weeds respond to these management methods can provide insight into the best practices for controlling them.

The treatments were selected to trigger physiological responses, so we expected to see some long-term changes in gene expression in harvested tissue. Plants typically respond to clipping of leaves by growing back and initiating some defense responses to block herbivores.[28] Shade is another stress that weeds encounter in the field. When weeds emerge after the crops or compete with each other, they can have limited access to sunlight.[29] Plants typically respond to shade stress by having a higher shoot to root ratio and expanding their leaf area. They may grow quickly in a vertical direction to escape the stress, and yields are often reduced.[30] Without a chemical component, these stressors should be very different from those induced by herbicides and provide a good contrast. In all cases, the plants were stressed with sublethal doses to obtain progeny for further studies.

While we checked gene expression values, we were particularly interested in the expression of the herbicide target genes. The herbicides used in this study were glyphosate and trifloxysulfuron. Glyphosate targets the EPSPS enzyme which adds phosphoenolpyruvate's enolpyruvyl moiety to shikimate-3-phosphate[13], leading to a block in the synthesis of aromatic amino acids.[31] Typical symptoms develop over the course of several days and involve chlorosis that starts in leaves of the apex and spreads into necrosis throughout the whole plant.[32] The other herbicide, trifloxysulfuron, is an ALS inhibitor. It blocks the synthesis of branched chain amino acids, and thus reduces cell division and plant growth.[33] Symptoms typically appear as a purplish discoloration of new grass leaves or broadleaf stems. Emerging plants may be stunted.[32] There are several modes of action for herbicides, and they can have a variety of effects on plants. With our two herbicides, we want to know if there are target site changes as well as altered regulation of other genes that could contribute to herbicide resistance and recovery.

Resistance to herbicides has been grouped into two broad classes: target site resistance (TSR) and non-target site resistance (NTSR).[23] TSR typically involves either an amino acid substitution at the binding site of the herbicide to the target enzyme or overexpression of the target enzyme.[34],[35] TSR rarely involves more than one gene, and is relatively easy to screen for among resistant plants.[34] NTSR includes all other sources of resistance. So far, NTSR has been grouped into three categories: translocation, avoidance, and metabolism.[36] Altered translocation can prevent the herbicide from moving throughout the plant to restrict damage.[37] This has been observed in several weed species that have evolved resistance to glyphosate.[37] They keep the herbicide in the leaves and prevent its spread to vulnerable shoot tissue. Avoidance of the herbicide entails minimizing the effects of the herbicide. An example of this can be found in ryegrass that is resistant to paraquat, which affects its targets by generating O_2^- radicals.[38] Resistant ryegrass can overproduce superoxide dismutase to deal with the radicals instead of directly dealing with the herbicide.[39] Metabolism-based resistance encompasses instances where the herbicide is detoxified and broken down. The pathway for herbicide metabolism can be roughly divided into four steps.[23] First, the herbicide is altered by phase I enzymes to allow phase II enzymes to add a hydrophilic molecule in the second step. Phase III transporters send the herbicide to the vacuole or extracellular space where it is finally degraded by phase IV enzymes. For example, a recent study of tribenuron-methyl resistant *Myosoton aquaticum* found constitutive up-regulation of all the gene families commonly involved in metabolism-based resistance: p450s, GSTs, glucosyltransferases, and ABC transporters.[40]

While identifying TSR is fairly straightforward, mechanisms of NTSR often involve more genes and can be harder to elucidate.[23] Epigenetics could even impact resistance. Methylation is associated with gene expression levels, chromatin structure, and transposon activity.[41] It is sometimes directed by sRNA, which can also influence expression directly in RNAi.[42] Methylation and sRNA could individually or collaboratively impact the expression of just the right p450s or some other genes to allow plants to survive herbicide application. We are also looking at the expression of genes related to such epigenetic pathways, including DNA methylation, RNA-directed DNA methylation, and histone modifications. Broad changes to methylation and methylation regulation could alter gene regulation on a more global scale. Another study has found an instance of UV light stress interfering with a methylase to impact

broad methylation patterns in Arabidopsis.[43] If stress responses could trigger changes to broad or even local epigenetic changes, they could lead to altered gene expression that could impact the phenotype and perhaps even prime the plant for future stresses.[44]

Our objective was to compare and contrast the gene regulation changes of Arabidopsis exposed to several abiotic stressors involved in weed control: two herbicides, mechanical clipping, and shade. We subjected young plants to these stresses, and later harvested new leaf tissue. From the resulting RNA seq analysis, we wanted to answer several questions. First, we were curious if stress responses would persist after the weeks of recovery and how would they relate to the plant coping with the stress. We want to know if there are differentially expressed genes pertaining to herbicide resistance and epigenetic pathways. The herbicides target different pathways, and there could be some important similarities and differences between those treatments. Persistent changes to gene expression and changes to epigenetic regulation could allow long term changes to the phenotype that could influence future responses to stress.

Methods:

Plant Growth Conditions:

Arabidopsis thaliana seeds of inbred ecotype Columbia were soaked in 0.1% agarose for five days in darkness and then sown onto Sunshine #1 potting media. The plants were grown in a Conviron growth chamber with a 16 hour and 22˚C day and an 8 hour 20˚C night. The light intensity during the day was 90 μmol $m^{-2}s^{-1}$. About 25-30 days after sowing, most of the plants had developed a full rosette, and plants used for treatments were selected for developmental uniformity. These plants were randomly assigned to either the control treatment or a sublethal dose of one of our four stressors that had been empirically determined to cause injury that was just short of lethal, so plants could recover, flower and set seeds. Glyphosate treated plants were sprayed with 7% of the recommended label rate of 0.80 kg ae / ha of RoundUp Pro Concentrate; Trifloxysulfuron treated plants were sprayed with 2.8% of the label rate of 0.02 kg ai /ha of Monument 75WG; shade treated plants were grown under a shade cloth that blocked 90% of the ambient light; and clipped plants had 90% of their biomass removed 3 times in 7 day intervals. Plants exposed to herbicide were treated in a spray chamber set to spray 187 L/ha at 2.09 KPH. Some of the stressors delayed flowering by several days, so under the above conditions, new control plants for them were grown that could be harvested at a similar time to match the clipped and glyphosate-treated plants. When the *Arabidopsis* plants had started to flower and had ~9 siliques, new rosette leaves were harvested for nucleic acid extraction (Fig. 1).

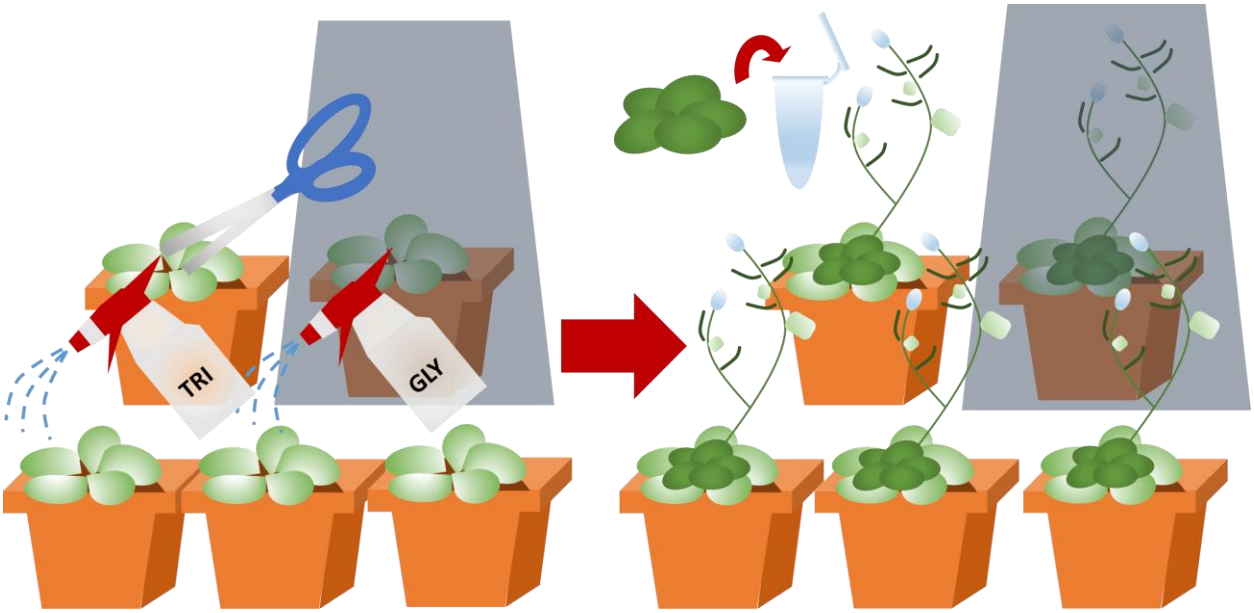


Figure 2. Experimental Design: Young Arabidopsis plants were treated with the stressors, glyphosate, trifloxysulfuron, mechanical clipping, and shade. After plants had recovered from each stress sufficiently to initiate flowering, new rosette leaves were harvested to obtain DNA and RNA for sequencing.

Tissue Collection and Sequencing:

Leaves were ground in liquid nitrogen, and the frozen ground tissue was separated into aliquots for nucleic acid extraction. RNA was extracted using the miRNeasy Mini Kit (Qiagen, Germany) according to the supplied protocol. Samples of extracted RNA were sent to Novogene for Illumina library preparation and paired-end sequencing on the HiSeq 4000 platform. There, Arabidopsis RNA libraries were sequenced to 25 million reads.

RNA Mapping and Analysis:

Adaptors were trimmed with Trimmomatic V0.33, using the default 4:15 sliding window and leading and lagging quality values of 3.[45] The minimum read length was set to 75bp to identify mRNA sequences. The sequences passing these quality filters were mapped to the TAIR10 genome with STAR V2.7.0c using the default settings.[46] These mapped reads were converted into sorted sam files and duplicates were removed with SAMtools V1.2.[47] RNA read levels were extracted with FeatureCounts V1.6.4.[48] This matrix was further analyzed with DESeq2 V1.28.1, which was recommended among other similar software packages for analyses with fewer replicates and low false positive rate.[49] DESeq2 calculates dispersion and shrinkage estimates for each gene to identify log fold changes in expression that are significant.[50] Each treatment was contrasted to the control, and differentially expressed genes were identified with adjusted p-values below .001 and a $\log_2$fold change of at least two. The differentially expressed genes identified with DESeq2 were compared to those identified with EdgeR V3.30.3 using an adjusted p value of .01 and a log2 fold change of at least 2. The DESeq2-identified, differentially expressed genes were entered into a gene ontology (GO) analysis with Panther to look for significant enrichment of GO terms among the annotations for the differentially expressed

genes.[51] It was set to the defaults with a Fisher's exact test and Bonferroni's multiple testing correction.

Results:

RNA PCA evaluation

The mapped RNA was added to a count matrix to tally the RNA that mapped to each known gene in Arabidopsis to estimate gene expression levels. To test the quality of our results, we used a principal component analysis (PCA) to see if the RNA results of each sample were similar to the others in their treatment group. The variance stabilized count matrix was used for a PCA analysis of the samples (Fig. 2). While some members of each treatment group are spread out, the replicates can be grouped by their treatment. As the clipped and glyphosate sprayed plants had delayed flowering, we grew some separate control plants to harvest at the same time. We expected their associated controls to somewhat resemble those treatments, and each pair of those controls are between their associated treatment and the control plants harvested alongside the trifloxysulfuron and shade-treated plants.

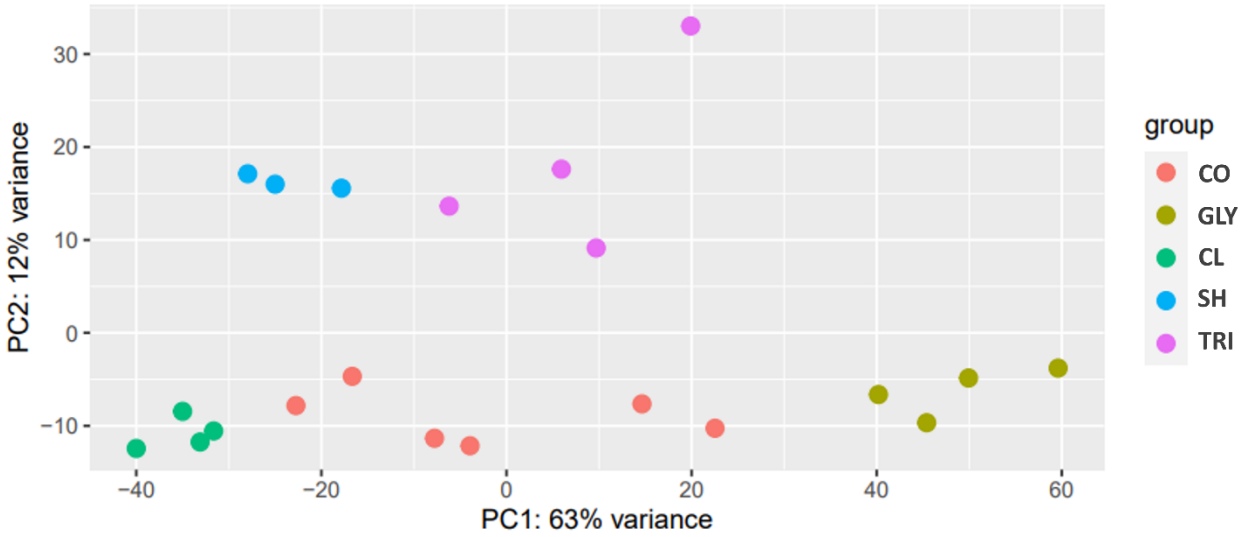


Figure 3 RNA PCA Plot: A variance stabilized count matrix of control (CO), glyphosate (GLY), clipping (CL), shade-treated (SH), and trifloxysulfuron (TRI) RNA sequences were compared with PCA.

After analyzing the count matrix with DESeq2, differentially expressed genes were identified for each treatment (Fig.3). Glyphosate treated plants had the most differentially expressed genes, followed by shaded, clipped, and trifloxysulfuron treated plants (Fig. 3). In most cases, more genes were downregulated under stress rather than upregulated when compared to the untreated control. The exception to this was the trifloxysulfuron treated plants, for which the great majority of differentially expressed genes were upregulated. The results of the DESeq2 differential expression analysis were compared to those identified using EdgeR to develop a candidate list that was robust to analysis method. The adjusted p value was relaxed to .01 to accommodate EdgeR's more stringent requirements, but generally, the genes identified were consistent (Fig. 4).[52]

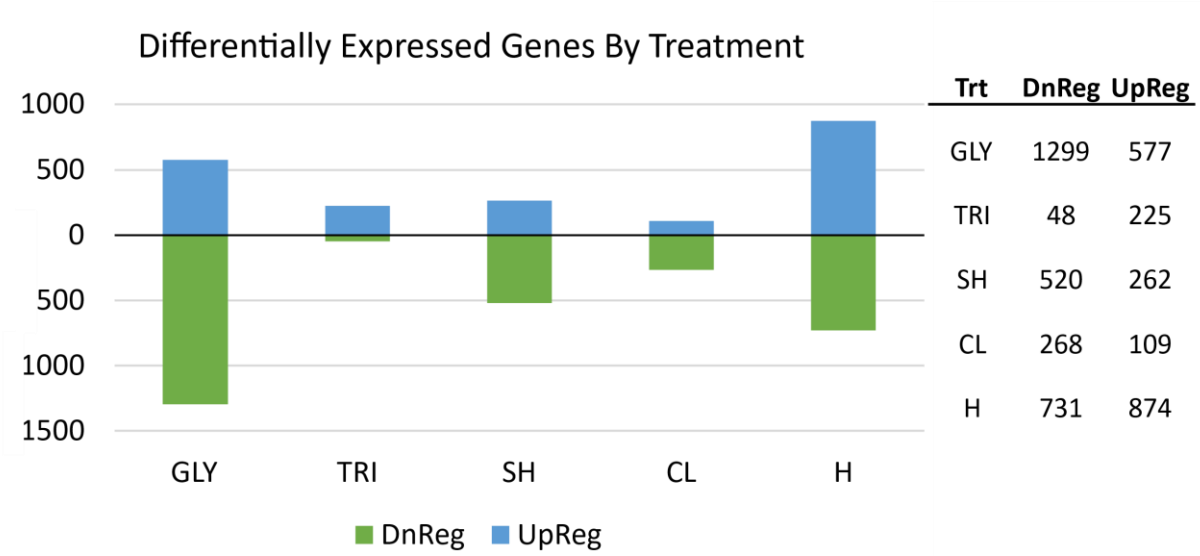
The identified differentially expressed genes for each treatment were compared to identify which DEGs were commonly up- and downregulated under different stressors. For downregulated genes, the most overlap occurred between shaded and glyphosate treated plants, closely followed by shaded and clipped plants (Fig. 5). Some terms were enriched with this set of genes, including biological processes related to defense, jasmonic acid response, aging, and anthocyanin synthesis. The most overlap between upregulated genes occurred between glyphosate and trifloxysulfuron treated plants and included enriched phosphate starvation responses (Fig. 6).

This list of differentially expressed genes was used for enrichment analysis of gene ontology (GO) terms. Most groups of differentially expressed genes returned GO terms that had significant enrichment. The trifloxysulfuron downregulated genes was the smallest group of genes. Trifloxysulfuron upregulated genes included enriched terms of drought response, abscisic acid response, and leaf senescence (Fig. 7). The glyphosate treated plants' upregulated genes held the same terms as well as biological processes associated with defense, toxin metabolism, and response to salicylic acid (Fig. 8). There was also an enrichment of transporters and glutathione transferases among the molecular processes. The downregulated genes also included some stress response genes, there was heavy enrichment of growth and photosynthesis-related genes. We also identified several downregulated DEGs in glyphosate pertaining to methylation (Table 1). Mechanically clipped plants had enrichment of stress and defense responses among the down-regulated genes, but few terms among the upregulated genes (Fig. 9). Shade-treated plants' upregulated genes had enriched stress response process, and many defense-related genes were downregulated in shade-treated Arabidopsis (Fig. 10).

The herbicide-treated Arabidopsis plants were also contrasted against the other two stressors to identify genes associated with herbicide treatment instead of those that may just be a generic stress response. In total, 731 genes were down-regulated in the presence of herbicide and 874 genes were up-regulated relative to the other stressors (Fig. 3). Together, the GO terms followed the general pattern of glyphosate treated Arabidopsis with an increase in expression of defense and other stress response genes while cell cycle and photosynthesis related genes had lower expression levels (Fig. 11). Some of the other stress responses genes were connected to salt, drought, and oxidative stress.

Some gene classes have been implicated in herbicide metabolism, including ABC transporters, glycosyltransferases, glutathione-s-transferases, and p450s.[23] The counts of those

genes were compared across treatment groups (Table 2). While the vast majority of these genes were unaffected by our treatments, some were differentially expressed. Glyphosate treated Arabidopsis had the most differentially expressed genes in all categories of herbicide metabolism genes, but more of them were upregulated than downregulated, which was against the overall trend of glyphosate DEGs. The number of herbicide metabolism DEGs decreased with shaded, mechanically clipped, and trifloxysulfuron treated Arabidopsis, respectively, and this reflected the decrease in overall DEGs with those treatments. Glyphosate and trifloxysulfuron treated Arabidopsis did not appear to have any overlapping metabolism genes.



Trt	DnReg	UpReg
GLY	1299	577
TRI	48	225
SH	520	262
CL	268	109
H	731	874

Figure 3: Differentially expressed genes detected by DESeq2 with an adjusted p-value of .001 and a $\log_2$ fold change of at least 2. Arabidopsis plants treated with glyphosate (GLY), trifloxysulfuron (TRI), mechanical clipping (CL) and 50% shade (SH) were contrasted against the untreated control. (H) is a contrast with both herbicides against the other stressors, with 874 genes upregulated under herbicide treatment when compared to the other stressors.

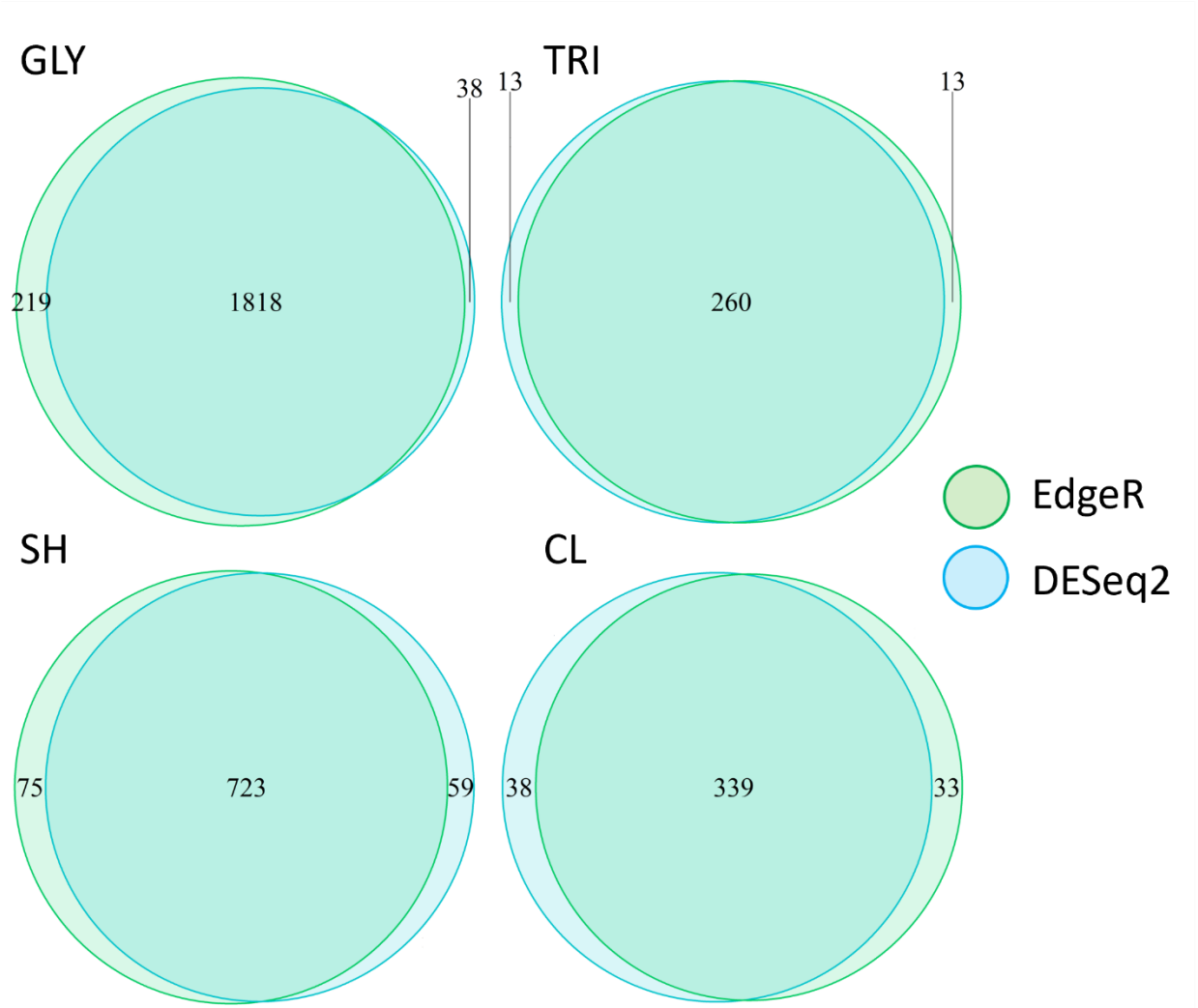


Figure 4. DESeq2 and EdgeR Differentially Expressed Genes: Most of the genes identified with DESeq2 were also identified with EdgeR general linear model contrast and a multiple comparison-adjusted p value of .01 for each treatment.

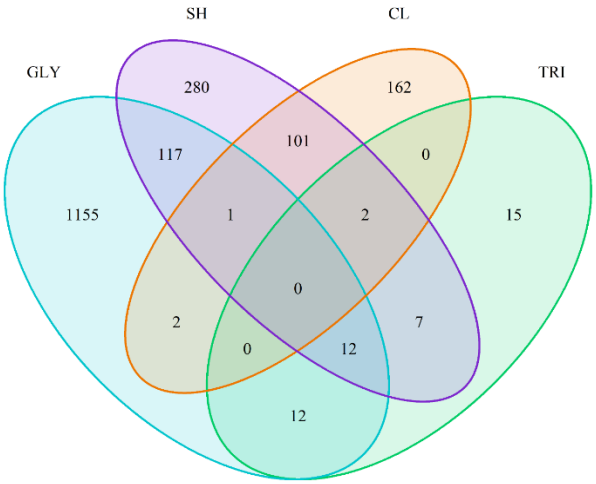


Figure 5. Overlapping Downregulated Genes: There was some overlap between genes that were downregulated under glyphosate-treated (GLY), trifloxysulfuron-treated (TRI), mechanically clipped (CL), and shaded (SH) plants.

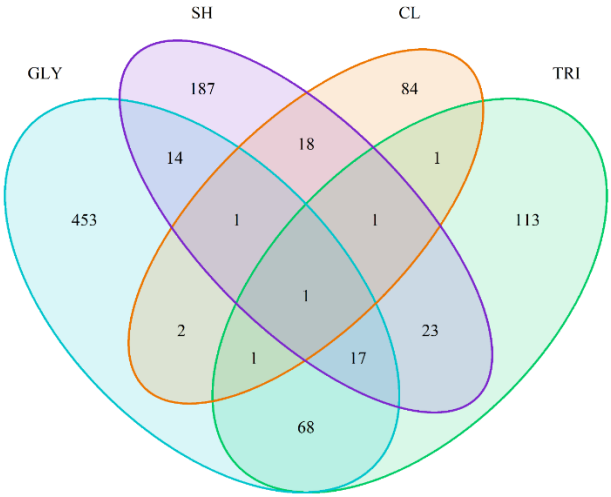


Figure 6 Overlapping Upregulated Genes: There was some overlap between genes that were upregulated under glyphosate-treated (GLY), trifloxysulfuron-treated (TRI), mechanically clipped (CL), and shaded (SH) plants.

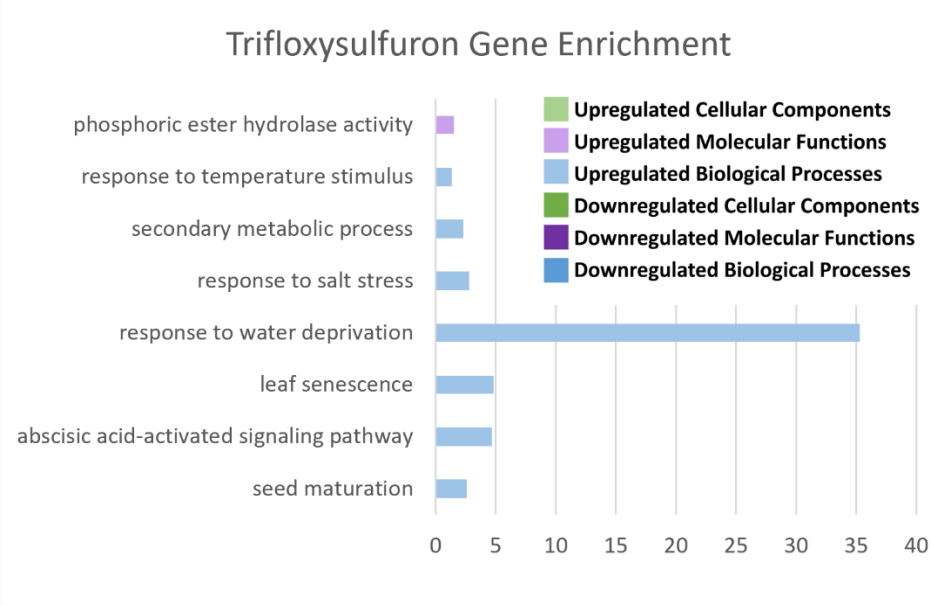


Figure 7: Trifloxysulfuron Gene Enrichment: The negative $\log_{10}$ of the expected p-values is shown for gene ontology terms that were over- and under-enriched under trifloxysulfuron herbicide stress. A higher value indicates more significant enrichment.

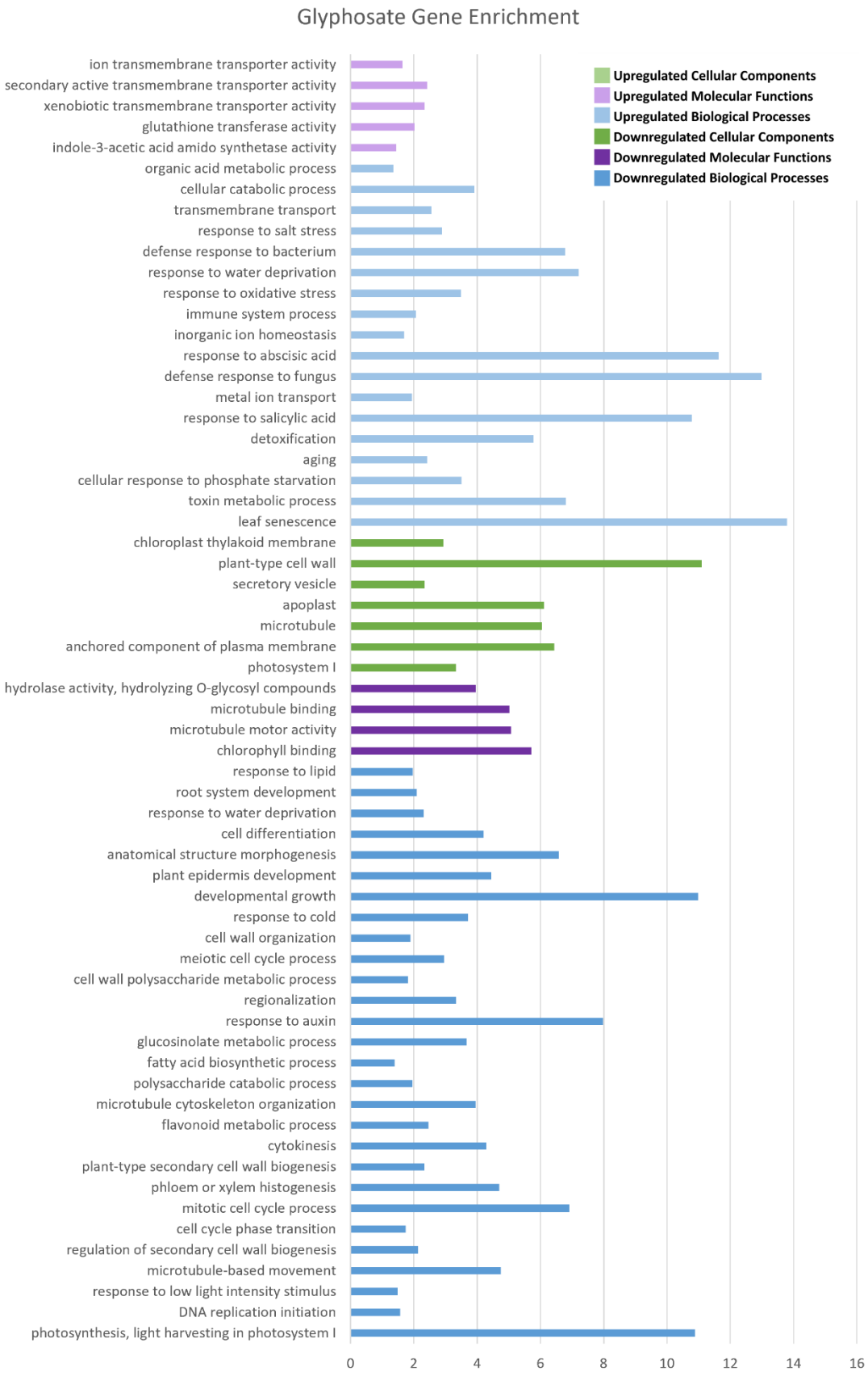

Glyphosate Gene Enrichment
Upregulated Cellular Components
Upregulated Molecular Functions
Upregulated Biological Processes
Downregulated Cellular Components
Downregulated Molecular Functions
Downregulated Biological Processes
ion transmembrane transporter activity
secondary active transmembrane transporter activity
xenobiotic transmembrane transporter activity
glutathione transferase activity
indole-3-acetic acid amido synthetase activity
organic acid metabolic process
cellular catabolic process
transmembrane transport
response to salt stress
defense response to bacterium
response to water deprivation
response to oxidative stress
immune system process
inorganic ion homeostasis
response to abscisic acid
defense response to fungus
metal ion transport
response to salicylic acid
detoxification
aging
cellular response to phosphate starvation
toxin metabolic process
leaf senescence
chloroplast thylakoid membrane
plant-type cell wall
secretory vesicle
apoplast
microtubule
anchored component of plasma membrane
photosystem I
hydrolase activity, hydrolyzing O-glycosyl compounds
microtubule binding
microtubule motor activity
chlorophyll binding
response to lipid
root system development
response to water deprivation
cell differentiation
anatomical structure morphogenesis
plant epidermis development
developmental growth
response to cold
cell wall organization
meiotic cell cycle process
cell wall polysaccharide metabolic process
regionalization
response to auxin
glucosinolate metabolic process
fatty acid biosynthetic process
polysaccharide catabolic process
microtubule cytoskeleton organization
flavonoid metabolic process
cytokinesis
plant-type secondary cell wall biogenesis
phloem or xylem histogenesis
mitotic cell cycle process
cell cycle phase transition
regulation of secondary cell wall biogenesis
microtubule-based movement
response to low light intensity stimulus
DNA replication initiation
photosynthesis, light harvesting in photosystem I
0
2
4
6
8
10
12
14
16

Figure 8: Glyphosate Gene Enrichment: The negative $\log_{10}$ of the expected p-values is shown for gene ontology terms that were over- and under-enriched under glyphosate herbicide stress.

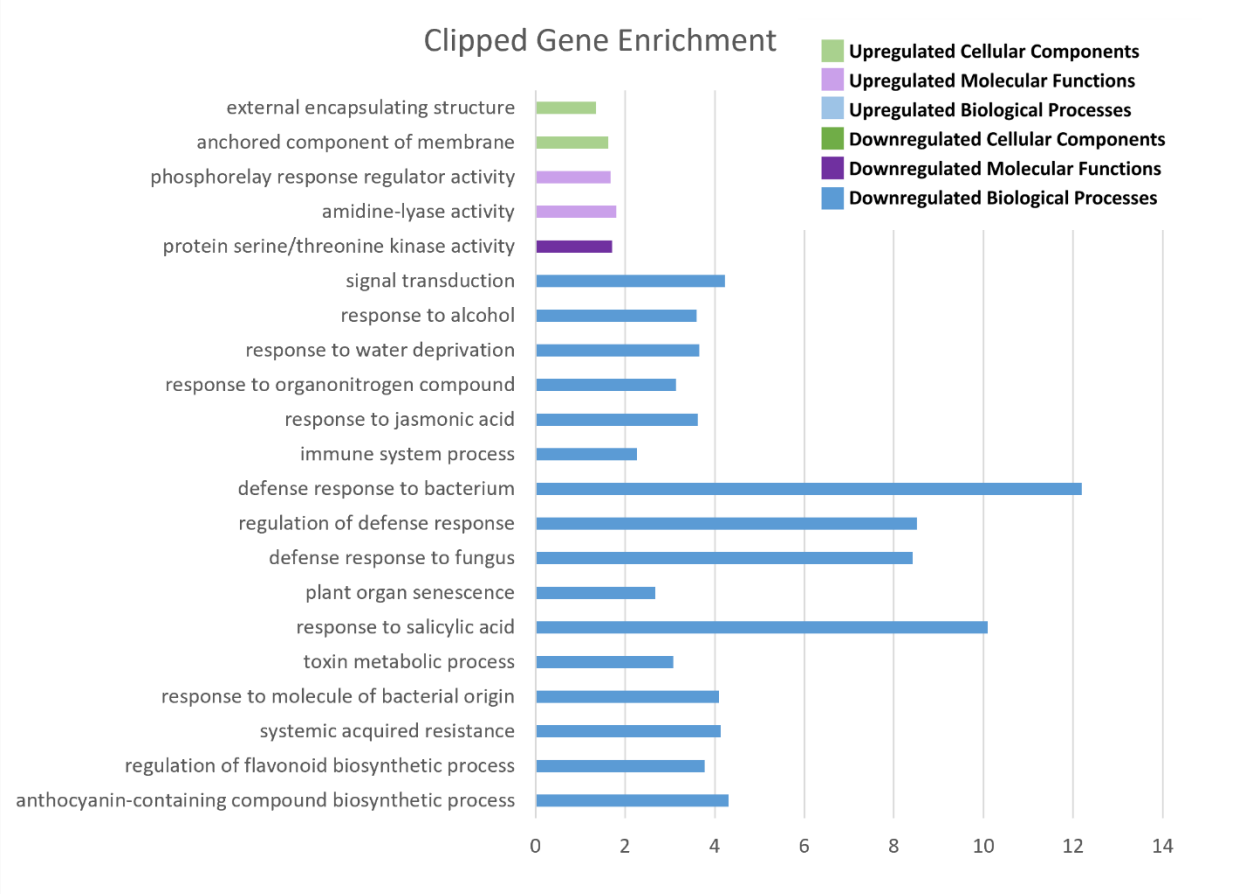


Figure 9. Mechanically Clipped Gene Enrichment: The negative $\log_{10}$ of the expected p-values is shown for gene ontology terms that were over- and under-enriched under mechanical clipping stress.

Figure 10. Shade-Treated Gene Enrichment: The negative $\log_{10}$ of the expected p-values is shown for gene ontology terms that were over- and under-enriched under shade stress.

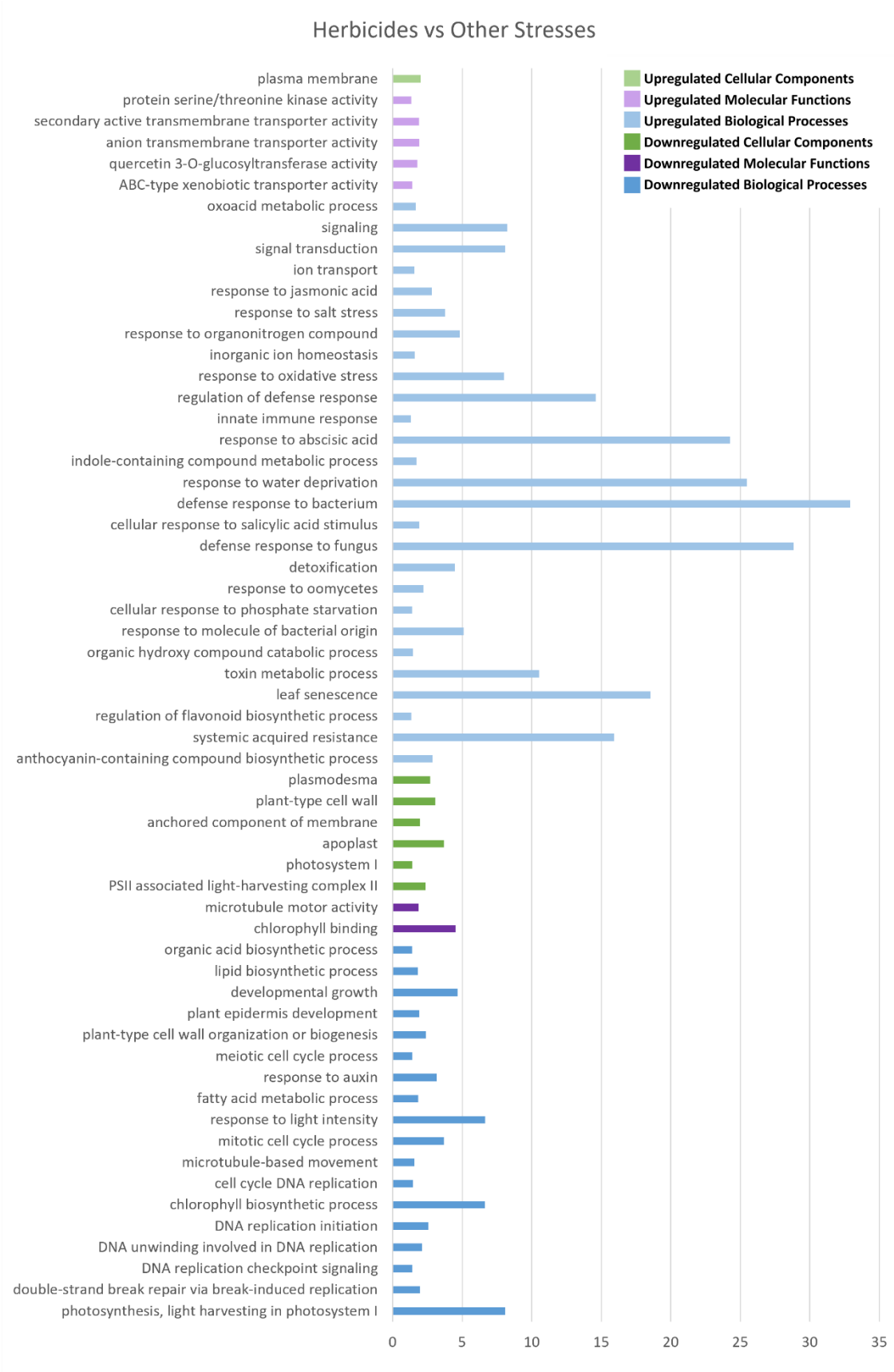

Herbicides vs Other Stresses
Upregulated Cellular Components
Upregulated Molecular Functions
Upregulated Biological Processes
Downregulated Cellular Components
Downregulated Molecular Functions
Downregulated Biological Processes
plasma membrane
protein serine/threonine kinase activity
secondary active transmembrane transporter activity
anion transmembrane transporter activity
quercetin 3-O-glucosyltransferase activity
ABC-type xenobiotic transporter activity
oxoacid metabolic process
signaling
signal transduction
ion transport
response to jasmonic acid
response to salt stress
response to organonitrogen compound
inorganic ion homeostasis
response to oxidative stress
regulation of defense response
innate immune response
response to abscisic acid
indole-containing compound metabolic process
response to water deprivation
defense response to bacterium
cellular response to salicylic acid stimulus
defense response to fungus
detoxification
response to oomycetes
cellular response to phosphate starvation
response to molecule of bacterial origin
organic hydroxy compound catabolic process
toxin metabolic process
leaf senescence
regulation of flavonoid biosynthetic process
systemic acquired resistance
anthocyanin-containing compound biosynthetic process
plasmodesma
plant-type cell wall
anchored component of membrane
apoplast
photosystem I
PSII associated light-harvesting complex II
microtubule motor activity
chlorophyll binding
organic acid biosynthetic process
lipid biosynthetic process
developmental growth
plant epidermis development
plant-type cell wall organization or biogenesis
meiotic cell cycle process
response to auxin
fatty acid metabolic process
response to light intensity
mitotic cell cycle process
microtubule-based movement
cell cycle DNA replication
chlorophyll biosynthetic process
DNA replication initiation
DNA unwinding involved in DNA replication
DNA replication checkpoint signaling
double-strand break repair via break-induced replication
photosynthesis, light harvesting in photosystem I
0
5
10
15
20
25
30
35

Figure 11. Herbicides Against Other Stressors Enrichment: The negative $\log_{10}$ of the expected p-values is shown for gene ontology terms that were over- and under-enriched under either herbicide treatment when compared to shade or clipping stress.

Table 1. DNA Methylation Genes: Some differentially expressed genes with functions pertaining to DNA methylation are listed for each treatment except mechanical clipping, which did not have DEGs with this function.

Gene ID	Treatment	DEG	Name	GO Functions[27]
AT5G43990	GLY	-	SUVR2	zinc ion binding, histone-lysine N-methyltransferase activity, histone lysine methylation, nucleolus, protein binding, protein binding, nucleus, chromosome, gene silencing by RNA-directed DNA methylation
AT2G36490	GLY	-	ROS1	base-excision repair, DNA demethylase activity, class I DNA-(apurinic or apyrimidinic site) endonuclease activity, DNA-(apurinic or apyrimidinic site) endonuclease activity, nucleus, DNA repair, DNA N-glycosylase activity, protein binding, DNA-(apurinic or apyrimidinic site) endonuclease activity, DNA methylation, DNA N-glycosylase activity, DNA demethylation, defense response to fungus
AT1G57820	GLY	-	VIM1	ubiquitin-protein transferase activity, maintenance of DNA methylation, protein ubiquitination, ubiquitin protein ligase activity, methyl-CpG binding, methyl-CpNpG binding, methyl-CpNpN binding, chromatin binding, chromocenter, double-stranded methylated DNA binding, pericentric heterochromatin assembly, DNA methylation on cytosine, histone binding, nucleus, chromatin organization, cell division, ubiquitin-protein transferase activity, protein ubiquitination, nucleus, DNA methylation on cytosine within a CG sequence, positive regulation of DNA methylation-dependent heterochromatin assembly
AT1G05490	GLY	+	CLSY3	gene silencing by RNA-directed DNA methylation, positive regulation of RNA interference
AT5G66750	GLY	-	DDM1	DNA binding, nucleus, DNA-dependent ATPase activity, regulation of gene expression by genetic imprinting, regulation of DNA methylation, positive regulation of histone H3-K9 methylation, DNA helicase activity, negative regulation of histone H4 acetylation, protein binding, transposition, RNA-mediated
AT3G54560	GLY	-	H2A11	DNA binding, protein heterodimerization activity, protein binding, flower development, defense

				response to bacterium, detection of temperature stimulus, regulation of gene expression, regulation of DNA methylation, response to osmotic stress, negative regulation of transcription, DNA-templated
AT1G13790	GLY	-	FDM4	gene silencing by RNA-directed DNA methylation
AT5G49160	GLY	-	MET1	chromatin binding, DNA (cytosine-5-)-methyltransferase activity, nucleus, C-5 methylation of cytosine, zygote asymmetric cytokinesis in embryo sac, regulation of gene expression by genetic imprinting, maintenance of DNA methylation, DNA methylation on cytosine within a CG sequence, DNA mediated transformation, protein binding, methyltransferase activity, negative regulation of flower development
AT4G08590	GLY	-	VIM6	nucleus, methyl-CpG binding, maintenance of DNA methylation, DNA methylation on cytosine within a CG sequence, methyl-CpNpG binding, methyl-CpNpN binding, histone binding, ubiquitin protein ligase activity, cytoplasm, ubiquitin-protein transferase activity, protein ubiquitination
AT5G39550	GLY	-	VIM3	maintenance of DNA methylation, histone binding, ubiquitin protein ligase activity, methyl-CpG binding, methyl-CpNpG binding, methyl-CpNpN binding, DNA methylation on cytosine, ubiquitin-protein transferase activity, vegetative to reproductive phase transition of meristem, protein ubiquitination, DNA methylation on cytosine, nucleus, DNA methylation on cytosine within a CG sequence, positive regulation of DNA methylation-dependent heterochromatin assembly
AT1G13790	TRI	-	FDM4	gene silencing by RNA-directed DNA methylation
AT5G43990	SH	-	SUVR2	zinc ion binding, histone-lysine N-methyltransferase activity, histone lysine methylation, nucleolus, protein binding, nucleus, chromosome, gene silencing by RNA-directed DNA methylation

Table 2. Overlapping Herbicide Metabolism Genes: The counts of four herbicide metabolism types of genes, p450s (p450), glutathione-S-transferases (GST), glucosyltransferases (GT), and ABC transporters (ABC), respectively, that were significantly differentially expressed are listed for each combination of conditions: glyphosate (G), trifloxysulfuron (T), clipped (L), and shaded (S). Uppercase indicates that the genes were upregulated under that condition while lowercase indicates that the genes were downregulated under that condition.

	P450	**GST**	**GT**	**ABC**
GS	0	0	1	0
GL	0	0	1	0
Gl	0	0	0	0

G	12	5	21	5
Gs	1	1	4	2
Gt	2	0	0	0
Gts	0	0	0	1
gTS	0	0	1	0
gT	5	1	1	0
gTsl	0	0	1	0
gSl	0	1	0	0
gl	0	3	1	0
g	6	4	10	3
gsl	1	1	1	0
gs	1	0	0	3
TSL	1	0	0	0
TL	1	0	0	0
T	3	0	4	2
SL	3	0	1	1
S	7	0	3	2
L	9	0	1	0
l	1	2	4	1
sl	2	1	5	2
s	8	0	15	5
tl	1	0	0	1
t	0	1	1	1
tsl	0	0	1	0
ts	1	1	1	0
NA	179	32	241	100

Discussion:

Our goal was to identify Arabidopsis' response to herbicides and other stresses by determining which genes' expression was affected by the stress. Comparing and contrasting these responses can help us understand how Arabidopsis reacts to herbicides in general and individually. We also applied two other stresses associated with weed management, shade and mechanical clipping. We identified many differentially expressed genes across all treatments, especially glyphosate. There were some similarities between the glyphosate and trifloxysulfuron responses, despite the differing modes of action. Clipped and shaded plants both had common downregulated genes. We identified several genes that are interesting for further study.

All of the samples could be separated by treatment in the PCA. The controls grown with glyphosate-treated plants and those grown with the clipped were between those treatments and the other controls. This shows that the overall RNA levels were similar between treatments and shows that there could be consistent responses to our stresses. We identified many differentially expressed genes: 1876 in glyphosate-treated, 273 in trifloxysulfuron-treated, 782 in shaded, and 376 in shaded. Under all stresses but trifloxysulfuron, there were more downregulated than upregulated genes. This gives us many genes to examine and submit for gene ontology analyses. Getting similar results from both the DESeq2 and EdgeR methods gave us more confidence in

the obtained results. Glyphosate and trifloxysulfuron-treated plants had the greatest number of overlapping DEGs at 68, while clipped and shaded plants had 101, by far the most overlapping downregulated genes.

The increase in defense and stress response genes for herbicide treated plants could be related to their exposure to toxic chemicals, which could resemble a pathogen attack. There were not many significant enrichment terms among the upregulated genes in trifloxysulfuron treated Arabidopsis, but those present did not overlap with another RNA seq analysis with Arabidopsis and a different ALS-inhibiting herbicide, which found enrichment of cell differentiation and secondary metabolite synthesis genes.[53] Glyphosate is known to impact photosynthesis, cause oxidative stress, and is associated with increased susceptibility to disease.[54] All of these processes appeared affected by the stress when we conducted the enrichment analysis. Some individual genes of interest included the herbicide target genes, EPSPS (AT1G48860) and ALS (AT3G48560).[55,56] In glyphosate-treated Arabidopsis, the EPSPS gene had a log2 fold change of -.406 and an adjusted p value of .034. It did not fall within our stringent requirements for identifying differentially expressed genes. Also, the ALS gene did not appear to be differentially regulated with a log2 fold change of .214 and an adjusted p value of .398. It is interesting that these responses persisted for weeks after the initial treatment and in tissue developed after the stress, and it raises the possibility of epigenetics driving long-term changes throughout the plant. This means that epigenetic changes could prime plants to respond better to future stresses.

The level of herbicide applied was deliberately selected to stress but not kill our plants. It is possible that some long-term responses include herbicide target and other genes related to resistance. In the field, glyphosate resistance has come about through several means, including both TSR and NTSR. In our plants, we detected differential expression from several genes in the herbicide metabolism families: p450s, ABC transporters, glutathione-s transferases, and glycosyltransferases.[23] There was also a positive enrichment of xenobiotic transmembrane transporters and other transporters, which could relate to translocation based resistance. Given that the plants had time to recover, it is interesting that so many resistance-related genes were still active. It is possible that such changes could give plants an advantage if they were exposed to the same stress and thus promote herbicide resistance in this weedy species.

Resistance to ALS inhibiting herbicides developed rapidly after their release and helped the scientific community take notice of herbicide resistance.[4,8] It was first detected for trifloxysulfuron in bluegrass in 2013 and has since been identified in several other species, including a few weeds from the *Amaranthus* genus.[57,4] Resistance to other ALS-targeting herbicides frequently develops through amino acid substitution; however, NTSR has also been identified.[58] In our plants, there was enrichment of some salt stress response genes. Trifloxysulfuron-treated Arabidopsis had the most in common with those treated with glyphosate when comparing overlapping upregulated genes. Despite having different modes of action, the two herbicides appeared to elicit similar responses. When glyphosate and trifloxysulfuron treated plants' RNA levels were contrasted against the other two stresses, the overall enrichment of DEGs had much in common with the glyphosate treated when compared to the control. One reason that glyphosate could have elicited such a strong response is that it is not readily metabolized by most plants. Trifloxysulfuron, on the other hand, is characterized by a shorter half-life.[59,60] It is possible that there is still glyphosate within the plant to trigger stress responses.

Shade and clipping are also interesting because they depict other mechanisms of weed control in the field. These treatments also provide another set of stress responses to which we can compare herbicide responses to separate those specific to herbicides. Mechanically clipped plants showed an enrichment of stress and defense response genes among the down-regulated genes, including responses to pathogenic organisms, jasmonic acid defense, and drought. Another study found that defense response and water deprivation genes were initially upregulated after mechanical wounding although the response decreased after a few hours.[28] The weeks of recovery time for our mechanically clipped plants could explain why we did not detect the same defense responses if the responses in that Arabidopsis study decreased after a few hours. Our mechanically clipped plants had far fewer DEGs than those treated with glyphosate. The clipped plants and shaded plants had the most overlap with 101 downregulated genes. In both cases, the downregulated genes had enriched defense-related functions. Perhaps the clipped plants were shifting resources towards a more important plant function for recovery. The shaded plants may have been conserving resources for growth, as they were stressed continuously.

The shade treatment was unique in that it was continuously applied, so there was no recovery time with this stress. This could affect which genes were active in those plants. Shade-treated plants only had some significant GO terms associated with the up-regulated genes: a few that were defense related, osmotic stress response, and drought response genes. The decrease in expression of stress response and defense related genes may have been the plant reducing resources allocated to genes not essential for reproduction to compensate for the lack of light. This would concur with other studies, but this is not how all plants species respond to shade.[61],[62],[63] A recent study has identified two genes (AT3G22170 and AT4G15090) that regulate such a balance in Arabidopsis, and our results appear to support this balance but not the gene expression values.[64] The actual log2 fold change of AT3G22170 was -.284 and adjusted p value was .269, which does not indicate differential expression, and AT4G15090 similarly had a log2 fold change of -.315 and adjusted p value of .148. The shade-treated plants appeared to have reduced expression of many defense response genes, but a few stress response genes were upregulated. There were a few pertaining to defense and some related to oxidative stress. These upregulated processes could have been involved in the plant's immediate needs to cope with its current stress.

There are several steps that could be taken to follow up this experiment. First, the expression levels of these genes of interest should be verified with quantitative PCR. This can help confirm that the genes that appear differentially expressed through this analysis are truly significantly different between treatments. Many Arabidopsis genes have already been studied, but sometimes further analysis is needed to determine the function of lesser-known genes or their role in stress response. While differential expression analysis is an excellent tool for selecting genes to spend resources investigating, expression that correlates with a stress can be a coincidence. Not all transcripts create functional protein products, and post-transcriptional regulation can influence the effect of the transcripts. A common technique for analyzing the effect of genes on plant physiology is to use knockout plant.[65] Mutant plants are generated to have reduced or no expression of the gene in question, and the effect is observed. This technique may be useful for further characterizing genes identified with this analysis.

Still, we identified several genes that could be relevant to stress and particularly herbicide response. If gene expression can still be so different after the stress was imposed, it is possible that the plants might react differently if they were exposed to the stress again. In the glyphosate treated plants, there were many differentially expressed genes with relevance to herbicide response and resistance. There were also several downregulated genes related to methylation regulation and RdDM. Glyphosate and trifloxysulfuron-treated plants had more upregulated genes in common than the other treatments while clipped and shaded plants had more downregulated defense genes in common. Our treatments elicited a long-term response of disrupted gene regulation, which allows for a possibility of stress memory in Arabidopsis.

References:

(1) *The Future Development of Maize and Wheat in the Third World*; International Maize and Wheat Improvement Center, Ed.; CIMMYT: Mexico, D.F, 1987.

(2) Khush, G. S. Green Revolution: The Way Forward. *Nat. Rev. Genet.* **2001**, *2* (10), 815–822. https://doi.org/10.1038/35093585.

(3) Kendall, H. W. Constraints on the Expansion of the Global Food Supply. In *A Distant Light*; Kendall, H. W., Ed.; Springer New York: New York, NY, 2000; pp 202–223. https://doi.org/10.1007/978-1-4419-8507-1_20.

(4) Heap, I. *The International Herbicide-Resistant Weed Database*. www.weedscience.org. http://www.weedscience.com/Pages/MOA.aspx?MOAID=12 (accessed 2020-03-19).

(5) Edwards, C. A. The Impact of Pesticides on the Environment. In *The Pesticide Question*; Pimentel, D., Lehman, H., Eds.; Springer US: Boston, MA, 1993; pp 13–46. https://doi.org/10.1007/978-0-585-36973-0_2.

(6) Vats, S. Herbicides: History, Classification and Genetic Manipulation of Plants for Herbicide Resistance. In *Sustainable Agriculture Reviews*; Lichtfouse, E., Ed.; Springer International Publishing: Cham, 2015; Vol. 15, pp 153–192. https://doi.org/10.1007/978-3-319-09132-7_3.

(7) Salisbury, E. J. THE WEED PROBLEM*. *Nature* **1942**, *149* (3787), 594–597. https://doi.org/10.1038/149594a0.

(8) Shaner, D. L. Lessons Learned From the History of Herbicide Resistance. *Weed Sci.* **2014**, *62* (2), 427–431. https://doi.org/10.1614/WS-D-13-00109.1.

(9) Whitehead, C. W.; Switzer, C. M. THE DIFFERENTIAL RESPONSE OF STRAINS OF WILD CARROT TO 2,4-D AND RELATED HERBICIDES. *Can. J. Plant Sci.* **1963**, *43* (3), 255–262. https://doi.org/10.4141/cjps63-052.

(10) Abel, A. L. The Rotation of Weed Killers. In *Proceedings of the British Weed Control Conference*; London, 1954.

(11) Timmons, F. L. A History of Weed Control in the United States and Canada. *Weed Sci.* **1970**, *18* (2), 294–307. https://doi.org/10.1017/S0043174500079807.

(12) Ryan, G. F. Resistance of Common Groundsel to Simazine and Atrazine. *Weed Sci.* **1970**, *18* (5), 614–616. https://doi.org/10.1017/S0043174500034330.

(13) Steinrücken, H. C.; Amrhein, N. The Herbicide Glyphosate Is a Potent Inhibitor of 5-Enolpyruvylshikimic Acid-3-Phosphate Synthase. *Biochem. Biophys. Res. Commun.* **1980**, *94* (4), 1207–1212. https://doi.org/10.1016/0006-291X(80)90547-1.

(14) Charles, D. *Lords of the Harvest: Biotech, Big Money, and the Future of Food*; Perseus Pub: Cambridge, Mass, 2001.
(15) Pratley, J.; Baines, P.; Eberbach, P.; Incerti, M.; Broster, J. Glyphosate Resistance in Annual Ryegrass. *Proc. 11th Annu. Conf. Grassl. Soc. New South Wales* **1996**, 122.
(16) Powles, S. B.; Debrah F. Lorraine-Colwill; James J. Dellow; Preston, C. Evolved Resistance to Glyphosate in Rigid Ryegrass (Lolium Rigidum) in Australia. *Weed Sci.* **1998**, *46* (5), 604–607.
(17) Benbrook, C. M. Trends in Glyphosate Herbicide Use in the United States and Globally. *Environ. Sci. Eur.* **2016**, *28* (1), 3. https://doi.org/10.1186/s12302-016-0070-0.
(18) Green, J. M. The Benefits of Herbicide-Resistant Crops. *Pest Manag. Sci.* **2012**, *68* (10), 1323–1331. https://doi.org/10.1002/ps.3374.
(19) Liu, Z.; Cao, S.; Sun, Z.; Wang, H.; Qu, S.; Lei, N.; He, J.; Dong, Q. Tillage Effects on Soil Properties and Crop Yield after Land Reclamation. *Sci. Rep.* **2021**, *11* (1), 4611. https://doi.org/10.1038/s41598-021-84191-z.
(20) Gianessi, L. P.; Reigner, N. P. The Value of Herbicides in U.S. Crop Production. *Weed Technol.* **2007**, *21* (2), 559–566. https://doi.org/10.1614/WT-06-130.1.
(21) Gianessi, L. P.; Sankula, S. *The Value of Herbicides in U.S. Crop Production*; National Center for Food and Agricultural Policy, 2003.
(22) *ISAAA's GM Approval Database*. http://www.isaaa.org/gmapprovaldatabase/.
(23) Yuan, J. S.; Tranel, P. J.; Stewart, C. N. Non-Target-Site Herbicide Resistance: A Family Business. *Trends Plant Sci.* **2007**, *12* (1), 6–13. https://doi.org/10.1016/j.tplants.2006.11.001.
(24) Kim, G.; Clarke, C. R.; Larose, H.; Tran, H. T.; Haak, D. C.; Zhang, L.; Askew, S.; Barney, J.; Westwood, J. H. Herbicide Injury Induces DNA Methylome Alterations in Arabidopsis. *PeerJ* **2017**, *5*, e3560. https://doi.org/10.7717/peerj.3560.
(25) Markus, C.; Pecinka, A.; Karan, R.; Barney, J. N.; Merotto, A. Epigenetic Regulation - Contribution to Herbicide Resistance in Weeds?: Epigenetic Regulation of Herbicide Resistance. *Pest Manag. Sci.* **2018**, *74* (2), 275–281. https://doi.org/10.1002/ps.4727.
(26) Kant, P.; Kant, S.; Gordon, M.; Shaked, R.; Barak, S. *STRESS RESPONSE SUPPRESSOR1* and *STRESS RESPONSE SUPPRESSOR2* , Two DEAD-Box RNA Helicases That Attenuate Arabidopsis Responses to Multiple Abiotic Stresses. *Plant Physiol.* **2007**, *145* (3), 814–830. https://doi.org/10.1104/pp.107.099895.
(27) *The Arabidopsis Information Resource (TAIR)*. https://www.arabidopsis.org/download/index-auto.jsp%3Fdir%3D%252Fdownload_files%252FGenes%252FTAIR10_genome_release.
(28) Reymond, P.; Weber, H.; Damond, M.; Farmer, E. E. Differential Gene Expression in Response to Mechanical Wounding and Insect Feeding in Arabidopsis. *Plant Cell* **2000**, *12* (5), 707–719. https://doi.org/10.1105/tpc.12.5.707.
(29) Borger, C. P. D.; Hashem, A.; Pathan, S. Manipulating Crop Row Orientation to Suppress Weeds and Increase Crop Yield. *Weed Sci.* **2010**, *58* (2), 174–178. https://doi.org/10.1614/WS-09-094.1.
(30) Patterson, D. T. Effects of Environmental Stress on Weed/Crop Interactions. *Weed Sci.* **1995**, *43* (3), 483–490. https://doi.org/10.1017/S0043174500081510.

(31) Dill, G. M.; Sammons, R. D.; Feng, P. C. C.; Kohn, F.; Kretzmer, K.; Mehrsheikh, A.; Bleeke, M.; Honegger, J. L.; Farmer, D.; Wright, D.; Haupfear, E. A. Glyphosate: Discovery, Development, Applications, and Properties. In *Glyphosate Resistance in Crops and Weeds*; Nandula, V. K., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010; pp 1–33. https://doi.org/10.1002/9780470634394.ch1.

(32) Baumann, P. A.; Dotray, P. A.; Prostko, E. P. *Herbicides: How They Work and the Symptoms They Cause*. https://agrilifeextension.tamu.edu/library/gardening/herbicides-how-they-work-and-the-symptoms-they-cause/.

(33) *Weeds 2001: The BCPC Conference: Proceedings of an International Conference Held at the Brighton Metropole Hotel, Brighton, UK, 12-15 November 2001*; British Crop Protection Council, Ed.; British Crop Protection Council: Farnham, Surrey, 2001.

(34) *Herbicide Resistance in Plants: Biology and Biochemistry*; Powles, S. B., Holtum, J. A. M., Eds.; Lewis Publishers: Boca Raton, 1994.

(35) Pline-Srnic, W. Physiological Mechanisms of Glyphosate Resistance. *Weed Technol.* **2006**, *20* (2), 290–300. https://doi.org/10.1614/WT-04-131R.1.

(36) Sammons, R. D.; Gaines, T. A. Glyphosate Resistance: State of Knowledge. *Pest Manag. Sci.* **2014**, *70* (9), 1367–1377. https://doi.org/10.1002/ps.3743.

(37) Preston, C.; Wakelin, A. M. Resistance to Glyphosate from Altered Herbicide Translocation Patterns: Resistance to Glyphosate Owing to Altered Translocation. *Pest Manag. Sci.* **2008**, *64* (4), 372–376. https://doi.org/10.1002/ps.1489.

(38) Dodge, A. D.; Harris, N. The Mode of Action of Paraquat and Diquat. *Biochem. J.* **1970**, *118* (3), 43P-44P. https://doi.org/10.1042/bj1180043P.

(39) Harper, D. B.; Harvey, B. M. R. Mechanism of Paraquat Tolerance in Perennial Ryegrass. II. Role of Superoxide Dismutase, Catalase and Peroxidase. *Plant Cell Environ.* **1978**, *1* (3), 211–215. https://doi.org/10.1111/j.1365-3040.1978.tb00763.x.

(40) Bai, S.; Liu, W.; Wang, H.; Zhao, N.; Jia, S.; Zou, N.; Guo, W.; Wang, J. Enhanced Herbicide Metabolism and Metabolic Resistance Genes Identified in Tribenuron-Methyl Resistant *Myosoton Aquaticum* L. *J. Agric. Food Chem.* **2018**, *66* (37), 9850–9857. https://doi.org/10.1021/acs.jafc.8b02740.

(41) He, X.-J.; Chen, T.; Zhu, J.-K. Regulation and Function of DNA Methylation in Plants and Animals. *Cell Res.* **2011**, *21* (3), 442–465. https://doi.org/10.1038/cr.2011.23.

(42) Montgomery, M. K.; Fire, A. Double-Stranded RNA as a Mediator in Sequence-Specific Genetic Silencing and Co-Suppression. *Trends Genet.* **1998**, *14* (7), 255–258. https://doi.org/10.1016/S0168-9525(98)01510-8.

(43) Jiang, J.; Liu, J.; Sanders, D.; Qian, S.; Ren, W.; Song, J.; Liu, F.; Zhong, X. UVR8 Interacts with de Novo DNA Methyltransferase and Suppresses DNA Methylation in Arabidopsis. *Nat. Plants* **2021**. https://doi.org/10.1038/s41477-020-00843-4.

(44) Johnson, R.; Puthur, J. T. Seed Priming as a Cost Effective Technique for Developing Plants with Cross Tolerance to Salinity Stress. *Plant Physiol. Biochem.* **2021**, *162*, 247–257. https://doi.org/10.1016/j.plaphy.2021.02.034.

(45) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

(46) Dobin, A.; Davis, C. A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T. R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29* (1), 15–21. https://doi.org/10.1093/bioinformatics/bts635.

(47) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinforma. Oxf. Engl.* **2009**, *25* (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

(48) Liao, Y.; Smyth, G. K.; Shi, W. FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30* (7), 923–930. https://doi.org/10.1093/bioinformatics/btt656.

(49) Seyednasrollah, F.; Laiho, A.; Elo, L. L. Comparison of Software Packages for Detecting Differential Expression in RNA-Seq Studies. *Brief. Bioinform.* **2015**, *16* (1), 59–70. https://doi.org/10.1093/bib/bbt086.

(50) Love, M. I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15* (12), 550. https://doi.org/10.1186/s13059-014-0550-8.

(51) Mi, H.; Muruganujan, A.; Huang, X.; Ebert, D.; Mills, C.; Guo, X.; Thomas, P. D. Protocol Update for Large-Scale Genome and Gene Function Analysis with the PANTHER Classification System (v.14.0). *Nat. Protoc.* **2019**, *14* (3), 703–721. https://doi.org/10.1038/s41596-019-0128-8.

(52) Robinson, M. D.; McCarthy, D. J.; Smyth, G. K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26* (1), 139–140. https://doi.org/10.1093/bioinformatics/btp616.

(53) Markus, C.; Pecinka, A.; Merotto, A. Insights into the Role of Transcriptional Gene Silencing in Response to Herbicide-Treatments in Arabidopsis Thaliana. *Int. J. Mol. Sci.* **2021**, *22* (7), 3314. https://doi.org/10.3390/ijms22073314.

(54) Gomes, M. P.; Smedbol, E.; Chalifour, A.; Hénault-Ethier, L.; Labrecque, M.; Lepage, L.; Lucotte, M.; Juneau, P. Alteration of Plant Physiology by Glyphosate and Its By-Product Aminomethylphosphonic Acid: An Overview. *J. Exp. Bot.* **2014**, *65* (17), 4691–4703. https://doi.org/10.1093/jxb/eru269.

(55) Klee, H. J.; Muskopf, Y. M.; Gasser, C. S. Cloning of an Arabidopsis Thaliana Gene Encoding 5-Enolpyruvylshikimate-3-Phosphate Synthase: Sequence Analysis and Manipulation to Obtain Glyphosate-Tolerant Plants. *Mol. Gen. Genet. MGG* **1987**, *210* (3), 437–442. https://doi.org/10.1007/BF00327194.

(56) Miki, B. L.; Labbé, H.; Hattori, J.; Ouellet, T.; Gabard, J.; Sunohara, G.; Charest, P. J.; Iyer, V. N. Transformation of Brassica Napus Canola Cultivars with Arabidopsis Thaliana Acetohydroxyacid Synthase Genes and Analysis of Herbicide Resistance. *Theor. Appl. Genet.* **1990**, *80* (4), 449–458. https://doi.org/10.1007/BF00226744.

(57) Cross, R. B.; McCarty, L. B.; Tharayil, N.; Whitwell, T.; Bridges, W. C. Detecting Annual Bluegrass (*Poa Annua*) Resistance to ALS-Inhibiting Herbicides Using a Rapid Diagnostic Assay. *Weed Sci.* **2013**, *61* (3), 384–389. https://doi.org/10.1614/WS-D-12-00172.1.

(58) Zhao, B.; Fu, D.; Yu, Y.; Huang, C.; Yan, K.; Li, P.; Shafi, J.; Zhu, H.; Wei, S.; Ji, M. Non-Target-Site Resistance to ALS-Inhibiting Herbicides in a Sagittaria Trifolia L. Population.

Pestic. Biochem. Physiol. **2017**, *140*, 79–84. https://doi.org/10.1016/j.pestbp.2017.06.008.
(59) Kanissery, R.; Gairhe, B.; Kadyampakeni, D.; Batuman, O.; Alferez, F. Glyphosate: Its Environmental Persistence and Impact on Crop Health and Nutrition. *Plants* **2019**, *8* (11), 499. https://doi.org/10.3390/plants8110499.
(60) Bhowmik, P. C. Introduction to the Symposium: History of Sulfonylurea Herbicide Use in Turfgrass Environments. *Weed Technol.* **2012**, *26* (3), 575–578. https://doi.org/10.1614/WT-D-11-00165.1.
(61) Gommers, C. M. M.; Keuskamp, D. H.; Buti, S.; van Veen, H.; Koevoets, I. T.; Reinen, E.; Voesenek, L. A. C. J.; Pierik, R. Molecular Profiles of Contrasting Shade Response Strategies in Wild Plants: Differential Control of Immunity and Shoot Elongation. *Plant Cell* **2017**, *29* (2), 331–344. https://doi.org/10.1105/tpc.16.00790.
(62) de Wit, M.; Spoel, S. H.; Sanchez-Perez, G. F.; Gommers, C. M. M.; Pieterse, C. M. J.; Voesenek, L. A. C. J.; Pierik, R. Perception of Low Red:Far-Red Ratio Compromises Both Salicylic Acid- and Jasmonic Acid-Dependent Pathogen Defences in Arabidopsis. *Plant J.* **2013**, *75* (1), 90–103. https://doi.org/10.1111/tpj.12203.
(63) Romero-Montepaone, S.; Sellaro, R.; Esteban Hernando, C.; Costigliolo-Rojas, C.; Bianchimano, L.; Ploschuk, E. L.; Yanovsky, M. J.; Casal, J. J. Functional Convergence of Growth Responses to Shade and Warmth in *Arabidopsis*. *New Phytol.* **2021**, *231* (5), 1890–1905. https://doi.org/10.1111/nph.17430.
(64) Liu, Y.; Wei, H.; Ma, M.; Li, Q.; Kong, D.; Sun, J.; Ma, X.; Wang, B.; Chen, C.; Xie, Y.; Wang, H. Arabidopsis FHY3 and FAR1 Regulate the Balance between Growth and Defense Responses under Shade Conditions. *Plant Cell* **2019**, *31* (9), 2089–2106. https://doi.org/10.1105/tpc.18.00991.
(65) Aklilu, E. Review on Forward and Reverse Genetics in Plant Breeding. *Life* **2021**, *14* (1), 127–135. https://doi.org/10.1080/26895293.2021.1888810.

Chapter 3: Differential Responses of Epigenetic Pathways in Response to Weed Management Stressors

Abstract:

Plants can cope with stress in a variety of ways, including modulation of the regulatory mechanisms associated with changes in gene expression. DNA methylation, which is sometimes directed by sRNA, is a widely studied epigenetic process. While fluctuations in methylation and sRNA have been linked to stress response, it is unknown which changes will persist after the plants have recovered from stress and if the changes are similar under different stresses. DNA and sRNA were extracted from Arabidopsis plants that had been exposed to no stress (controls) or one of four stresses: glyphosate, trifloxysulfuron, mechanical clipping, or shade. DNA methylation and sRNA levels of genes in stressed plants were compared to those of an untreated control to identify differential activity. We found many responses involving methylation and sRNA such as differentially methylated regions (DMRs), differentially methylated cytosines (DMCs), and differentially expressed sRNA. Glyphosate-treated plants had the largest number of differentially methylated sites, including 66 hypomethylated DMRs, and differentially expressed sRNA. There was almost no overlap among the differentially expressed sRNA between treatments. These results show that plants undergo stress-specific changes to systems associated with epigenetic mechanisms, suggesting that such systems may be involved in plant adaptation to stress.

Introduction:

Epigenetics is the study of changes to physiology that are not due to changes in the genetic sequence.[1] This focuses on the forces that drive gene expression. Such a system was first proposed by Waddington, who also coined the word "epigenetics" in 1942 as he remarked on differences between phenotypes that are not linked to the genetic sequence.[2] The term was related to a 17th century discussion about epigenesis and the processes that may underly tissue differentiation.[3] Epigenetics is believed to drive sometimes heritable gene expression changes that alter the phenotype. There are some processes, including histone modifications and DNA methylation, that can facilitate these changes.[1] sRNA can also be relevant in post-transcriptional regulation and RNA-directed DNA methylation (RdDM).[4]

In this study, we were interested in long-term stress responses in Arabidopsis, with a focus on weed control measures. We treated young plants with glyphosate, trifloxysulfuron, shade, and mechanical clipping. Then, the stress was removed in all treatments but shade until flowering, which was when we extracted tissue for nucleic acid sequencing. In the previous chapter, we identified many differentially expressed genes from this experiment, showing that the stressors elicited long term responses for all treatments. Stress responses in plants are not limited to changes in gene expression. We want to study how these changes could be regulated in Arabidopsis, a weedy species, to better understand how weeds cope with stress and thus move towards better management strategies.[5] This chapter focuses on the epigenetic responses that were detected through DNA methylation and sRNA levels.

Two pathways suspected of controlling epigenetics are DNA methylation and histone modifications. DNA methylation involves the addition of a methyl group to the fifth carbon of cytosine.[6] Histone modifications are the addition of functional groups, such as methylation and acetylation, to the amino acids of histone proteins, which DNA coils around to form chromatin.[7]

Chromatin has two forms, heterochromatin and euchromatin. The former is a tightly wound form that is condensed and not readily accessible for transcription, while the latter is a looser form that allows genes to be expressed.[8] Some histone modifications are associated with each form, and others are believed to help DNA become accessible for transcription or help other proteins bind near genes to regulate their expression.[8,9] Although this study does not focus on histone modifications, they are an important epigenetic process, and a factor that can influence both gene expression and DNA methylation.[10]

In mammals, DNA methylation primarily occurs in the CG context, a cytosine that is followed by a guanine.[11] Plants also have methylation in the CHH and CHG context, where "H" is any base other than guanine.[11] In plants, CG methylation is maintained by DNA methylase 1 (MET1).[12,13,14] CHG methylation acts with H3K9me in a self-reinforcing loop facilitated by chromomethylase 3 (CMT3). CHH methylation is maintained by (CMT2) in some cases, but it is often associated with a process known RNA-directed DNA methylation (RdDM), which methylates DNA in all contexts.[15,16] Both are associated with transposon silencing. Some patterns of DNA methylation coincide with gene expression levels, suggesting a causal relationship between DNA methylation and gene expression.[17] While methylation levels are typically reset with new generations, some broad patterns of methylation appear highly conserved.[18] As methylation levels are known to change under some stresses, it is interesting to know whether these changes could potentially be passed to future generations and thereby prime them for potential encounters with those same stresses.[19,20]

Hotchkiss first discovered DNA methylation in 1948.[21] In bacteria, it has a role in the restriction enzyme defense system against bacteriophages.[22] Some restriction enzymes are sensitive to the presence of methylation and will thus only cut unmethylated viral DNA. This principle was used in an early method for detecting methylation levels in extracted DNA.[23] Some microarrays were developed to detect the resulting different-sized fragments.[24] There are several routes to study the effects of reduced DNA methylation in plants. One method involves chemicals, such as 5-azacytidine, procainamide, and 5-aza-2′-deoxycytidine, that are known to remove cytosine methylation and reactivate some genes.[25],[26] Knockout mutant plants for various combinations of DNA methylases and associated processes have reduced methylation.[27–29] Some target methylases, such as cytosine methylase 1 (MET1) and proteins that assist methylases, such as decrease in DNA methylation 1 (DDM1). These provide opportunities for researchers to conduct experiments and determine the importance of methylation in various growth conditions.[30],[31] Methylation had suspected involvement in differential expression and tissue differentiation.[32] Other studies found that organism development proceeded abnormally in the absence of methylation.[23] Today, DNA methylation is detected through a method known as bisulfite sequencing. Bisulfite treatment converts unmethylated cytosines to uracil, which is then converted to thymine as fragments are amplified during sequencing.[33] Mapping these altered sequences to the genome with settings that tolerate conversions of cytosine to thiamine and then checking the counts of each can give an estimate of the methylation level at a particular base.[34]

Several patterns of DNA methylation near genes have been identified.[35] Some genes are unmethylated and do not have methylation in any context. A common pattern of methylation is known as gene body methylation, which involves higher levels of CG methylation in the center

of genes but lower levels at the transcription start and termination sites.[36] The role of gene body methylation, particularly in Arabidopsis, is unclear.[37] Promoter methylation involves higher levels of methylation at the start sites and is associated with transcriptional suppression.[38],[39] CHG methylation can occur in some genes and is associated with lower levels of transcription.[17] RdDM involves higher levels of methylation in all contexts and is associated with silencing.[16] Methylation changes have been linked to stress responses in a variety of studies.[40],[41],[42] It has also been connected to phenotype changes.[43,44]

Epigenetics can also involve small RNA (sRNA). Plant sRNA is synthesized in the nucleus from long noncoding RNA (lncRNA).[45] The primary known functions for sRNA is regulation of the expression of genomic features, such as genes and transposons; and recruitment of proteins to a specific target.[46] There are three primary groups in RNAi: microRNA (miRNA), Piwi-interacting RNA, and small interfering RNA (siRNA).[47] miRNA has been connected to development, differentiation, and response to salt stress.[48],[49] With miRNA, Argonuate 1 binds with other proteins to form a RISC complex for post transcriptional regulation.[47],[50],[51] Then, messenger RNA with the right complementary bases is sliced with the RISC complex to prevent gene expression, or the protein complex remains on its target to prevent translation of the mRNA into protein.[48,52] miRNA is one class of sRNA that can regulate gene expression post-transcriptionally, and some of our sRNA could be miRNA.

While plants do not have Piwi-interacting small RNA, the role of siRNA has expanded to fill part of that niche.[53] 21-22 bp siRNA are involved in gene silencing.[54] In RdDM, siRNA guides a large protein complex to bind to a specific target within the genome and methylates DNA through Domains Rearranged Methylase 2.[4] The guiding siRNA for this process is typically 24 base pairs (bp) long, but it can also be 21 or 22 bp.[45] This siRNA is typically heterochromatic and thus cis-acting, which means that its target is the surrounding DNA of the original lncRNA.[47] Some siRNA is trans-acting, which is also known as ta-siRNA.[47] It is generated by the cleavage of other RNA by miRNA.[55] There are siRNA known to come from transfer RNA, ribosomal RNA, and retrotranposons.[56],[57],[58] This has led to several overlapping classification schemes based on sRNA targets, origins, and functions, and these have varying usefulness depending on the objective of the study.

sRNA levels can be challenging to study. One problem commonly encountered is multimapping while aligning sRNA sequence data to a genome. Shorter sequences will naturally map to more locations in the genome. This makes downstream analyses challenging. Many programs will automatically discard multimapping reads. Discarding all of them would be a big loss of data. With ta-siRNA acting on sites other than its origin, it could easily align to both locations. The high fidelity of siRNA sequences to their targets could mean that the sRNA would map to both locations, and the source and target of the siRNA would be important. There are several programs available that can attempt to sort where the multimapping sRNAs should be placed in the genome.[59],[60],[61] Despite these challenges, some other studies have linked differentially expressed sRNA to stress responses in plants.[45,62,63] Differentially expressed sRNA associated with potential non-target site resistance genes for the fenoxaprop-P-ethyl herbicide has been detected in an herbicide resistant variety of slough grass.[64] sRNA has the potential to influence gene expression in ways that can help plants cope with stress.

Between these two regulatory forces, small RNA and DNA methylation, we wanted to see whether there were long-term differential methylation and sRNA changes associated with our treatments, glyphosate, trifloxysulfuron, clipping, and shade, when they are compared to the untreated control. Both epigenetic systems have suspected involvement in gene expression, so we wanted to test for differential expression at any sites of differential methylation and sRNA. Perhaps some of these changes may prime the plant in a way that could influence how it reacts to future instances of the stresses.[65] These persistent changes could give weeds an advantage if they are exposed to the same stress again. Growers could consider this while selecting the appropriate weed control strategies and perhaps avoid using the same method repeatedly.

Methods:

Plant Growth Conditions:

Arabidopsis thaliana seeds of inbred ecotype Columbia were soaked in 0.1% agarose for five days in darkness and then sown onto Sunshine #1 potting media. The plants were grown in a Conviron growth chamber with a 16 hour and 22˚C day and an 8 hour 20˚C night. The light intensity during the day was 90 μmol $m^{-2}s^{-1}$. About 25-30 days after sowing, most of the plants had developed a full rosette, and plants used for treatments were selected for developmental uniformity. These plants were randomly assigned to either the control treatment or a sublethal dose of one of our four stressors that had been empirically determined to cause injury that was just short of lethal, so plants could recover, flower and set seeds. Glyphosate treated plants were sprayed with 7% of the recommended label rate of 0.80 kg ae / ha of RoundUp Pro Concentrate; Trifloxysulfuron treated plants were sprayed with 2.8% of the label rate of 0.02 kg ai /ha of Monument 75WG; shade treated plants were grown under a shade cloth that blocked 90% of the ambient light; and clipped plants had 90% of their biomass removed 3 times in 7 day intervals. Plants exposed to herbicide were treated in a spray chamber set to spray 187 L ha^{-1} at 2.09 KPH. Some of the stressors delayed flowering by several days, so under the above conditions, new control plants for the were grown that could be harvested at a similar time to match the clipped and glyphosate-treated plants. When the *Arabidopsis* plants had started to flower and had ~9 siliques, new rosette leaves were harvested for nucleic acid extraction (Fig. 1).

Tissue Collection and Sequencing:

Leaves were ground in liquid nitrogen, and the ground tissue was split for RNA and DNA extraction. RNA was extracted using Qiagen's miRNeasy Mini Kit according to the manufacturer's protocol. DNA was extracted using Qiagen's DNeasy Plant Mini Kit according to the manufacturer's protocol. Nucleic acid samples were sent to Novogene for library preparation and sequencing. The DNA was also subjected to bisulfite treatment for methylation sequencing by Novogene. There, Arabidopsis sRNA libraries were sequenced to 3 million reads and whole-genome bisulfite libraries to 30X.

sRNA Mapping and Analysis:

Adaptors were trimmed with Trimmomatic, using the default settings.[66] The minimum base length was set to 18 to identify sRNA sequences. The remaining reads were mapped with Manatee V1.0, which uses Bowtie V1.0.1 and a GTF file, which includes known sRNA, to identify the position of sRNA reads.[59] Manatee preferentially maps sRNA to established sites

and develops a count matrix of sRNA.[59] This matrix was further analyzed with DESeq2 V1.32.[67] Each treatment was contrasted to the control, and differentially expressed sRNA were selected for having an adjusted p-value below .01 and an absolute $\log_2$ fold change of at least two. Sites with sRNA that were not linked to the GFF file features were also identified with Manatee and the differential expression was calculated with DESeq2.[67]

Bisulfite Mapping and Analysis:

Adaptors were trimmed with Trimmomatic V0.33, using the default 4:15 sliding window and leading and lagging quality values of 3. The minimum base length was set to 75 to identify mRNA sequences. The remaining sequences were mapped to the TAIR10 genome through Bismark V0.20.0 using Bowtie2 using the default settings.[34] The output files were deduplicated and further analyzed with Methylkit.[68] Differentially methylated cytosines (DMCs) were identified for all three contexts with a difference of at least 10% and an adjusted p-value of less than 0.01. Sites were only included if they were covered in all of the replicates for the sample. The samples were also clustered with Methylkit V1.14.2, using a correlation distance metric and ward agglomeration. DMRs were identified using DMRCaller V1.20.0 with the neighborhood method on the Bismark CX report files.[69]

lncRNA and Transposon Identification:

Potential transposons were identified by running RepeatModeler on the TAIR10 genome.[70] These were merged with previously identified transposons in the TAIR10 gff. Priority in determining the bounds of the transposons was given to the previously annotated ones. The Cantata database was used to identify sites of lncRNA.[71]

Results:

sRNA Analysis:

The count matrices, one of previously annotated features and one of newly detected sRNA sites, of sRNA mapped with manatee were visualized via PCA to determine how similar the samples were to one another (Figures 1- 3). In both PCA analyses, glyphosate treated plants and associated controls had more in common with each other than with the other samples. Clipped Arabidopsis samples also clustered together, but the remaining treatments' samples were mixed (Figures 2 and 3). With DESeq2, the count matrices were evaluated to identify sRNA that were differentially expressed when contrasted to an untreated control. These values indicated the number of sites in the Arabidopsis genome with varying levels of sRNA that were consistent in a treatment group. Glyphosate treated plants exhibited the greatest number of differentially expressed sRNA, with 758 differentially expressed sRNA from established sequences and Manatee-detected sRNA regions. This was followed by 223 from clipped, 156 from shaded, and 156 from trifloxysulfuron-treated plants. (Figure 4) The glyphosate treatment appears to have elicited the strongest sRNA response. When the identified differentially express sRNA that were mapped to known features were compared, there was little overlap of differentially transcribed sRNA between the treatment groups (Figure 5). There were only 15 total differentially expressed sRNA that were common to some of the treatments, but none were common to all treatments. The treatments elicited unique, long -term sRNA responses in the plants.

Next, we attempted to determine whether the unique responses were random or associated with different genomic features, such as genes and transposons. Comparing the

locations of the differentially expressed sRNA to the classes of these features in the annotation file showed which types of features were most strongly influenced by each treatment (Table 1). 92% of the differentially expressed sRNA in glyphosate associated with genes was downregulated, while the other treatments had an average of 12% downregulated gene sRNA. 42% of the differentially expressed sRNA in glyphosate-treated plants mapped to unannotated regions, while the other treatments had an average of 64% of their differentially expressed sRNA map to unannotated regions. There were 62 differentially expressed sRNA that mapped to transposons, while the other treatments had a combined total of 27. Glyphosate-treated plants' differentially expressed sRNA were found in a unique distribution of features. The sRNA may be regulating different systems for each treatment.

24 nt sRNA is frequently associated with maintenance RdDM, and sRNA of about 21 nt can be relevant for other processes.[4,45] Given the relationship between sRNA length and function, the proportion of sRNA lengths from each experimental group were counted (Figure 6). This did not, however, reveal any differences between the treatments. The standard deviation of the sRNA levels overlapped for the different lengths. Although glyphosate-treated Arabidopsis has a higher proportion of 24 nt sRNA at 25% than the other experimental groups, at about 7%, they were grown under different conditions and the levels overlapped with the control plants grown at the same time (see Methods).

Next, we tested for relationships between sRNA location and methylation, both of which have roles in RdDM, which is typically silences transposons.[4] The average methylation levels of TAIR10 GFF features with aligned sRNA were determined for control and treated plants. Methylation of transposons with sRNA appeared higher than that of genes in all conditions (Figure 7 and 8), but it did not appear to vary when comparing treated levels and those of the control plants. The methylation of sites with downregulated sRNA did not appear decreased, and that of sites with upregulated sRNA did not appear increased when compared to the methylation in the control plants. (Figure 9 and 10) Overall sRNA methylation in genes and transposons did not appear to change when compared to the control.

When changes in sRNA levels and gene expression were compared, the correlation between them significantly accounted for about 20% of the variation using the Pearson's correlation coefficent(Figure 11). 19%, 21%, 14%, and 24% of the variance was explained by this in glyphosate treated, trifloxysulfuron treated, shaded, and mechanically clipped plants, respectively. The p-values were 3.2729E-220, 3.6113E-281, 1.8212E-125 and 0. While the correlation for these relationships was consistently positive across treatments, the R^2 values indicate that it was very weak.

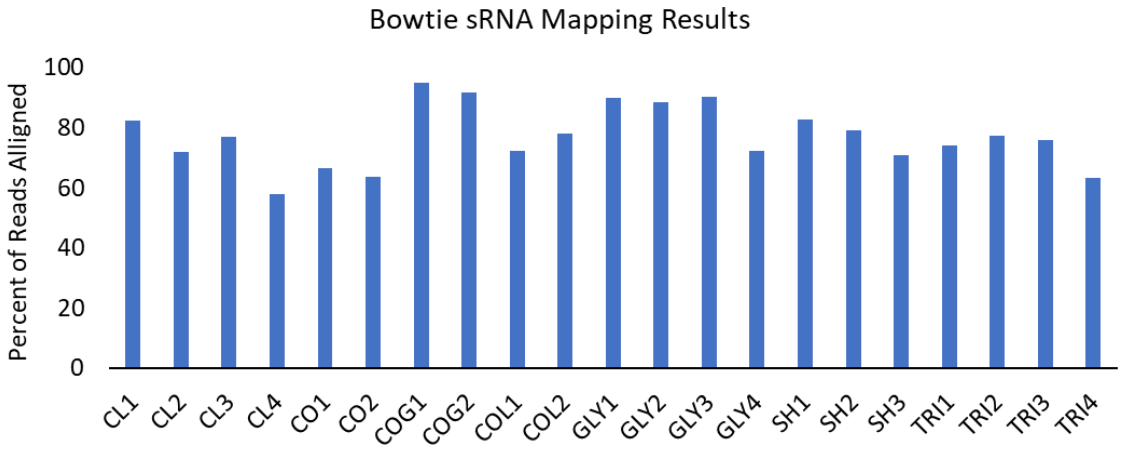


Figure 1 sRNA mapping efficiency: When the sRNA reads were alligned to the TAIR10 genome using Bowtie that allowed 50 multimaps per read, these were the resulting percentages of sRNA that successfully aligned to the genome. The numbers indicate the replicates, and and the letters stand for the treatments: clipping (CL), control (CO), glphosate control (COG), clipping control (COL), glyphosate (GLY), shade (SH), and trifloxysulfuron (TRI).

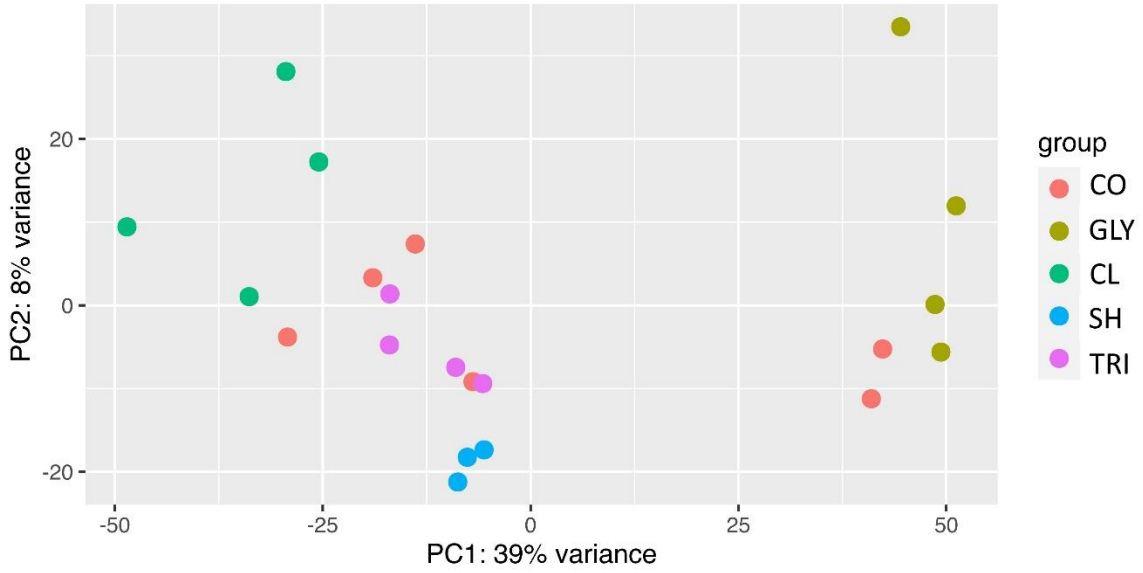


Figure 2 DESeq2 PCA of TAIR10 GFF Feature sRNA levels: A PCA chart is displayed for the count matrix of sRNA that aligned to GFF features using the Manatee mapping software. The treatment groups are as follows: control (CO), glyphosate (GLY), clipping (CL), shade (SH), and trifloxysulfuron (TRI).

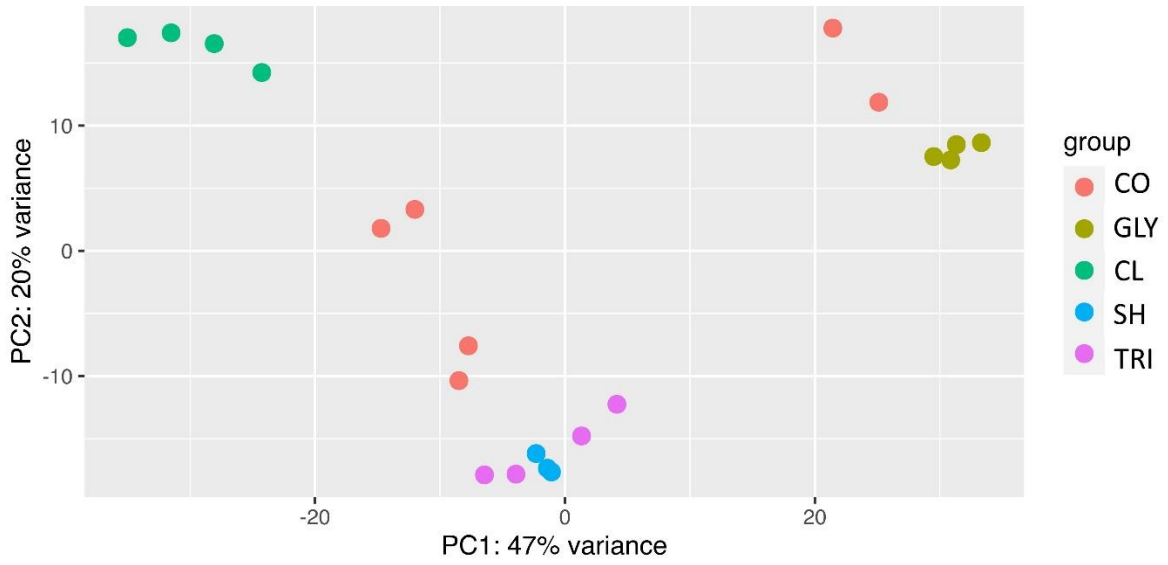


Figure 3 DESeq2 PCA of Other sRNA Clusters: The count matrix of sRNA clusters that did not align to GFF features was subjected to a PCA test. The treatment groups are as follows: The treatment groups are as follows: control (CO), glyphosate (GLY), clipping (CL), shade (SH), and trifloxysulfuron (TRI).

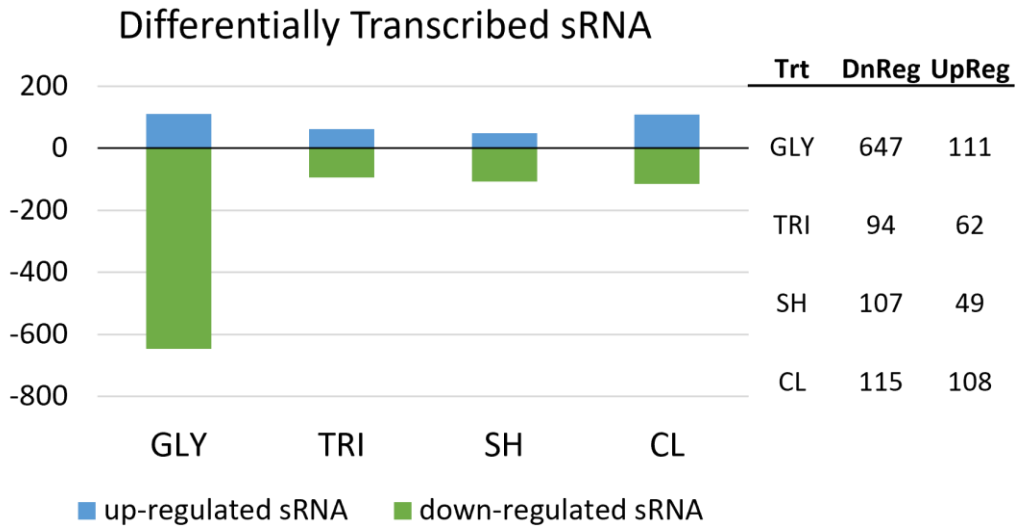


Figure 4 Counts of Differentially Transcribed sRNA: The counts of the up and down-regulated sRNA under stress is shown for each treatment: glyphosate (GLY), trifloxysulfuron (TRI), shade (SH), and clipping (CL).

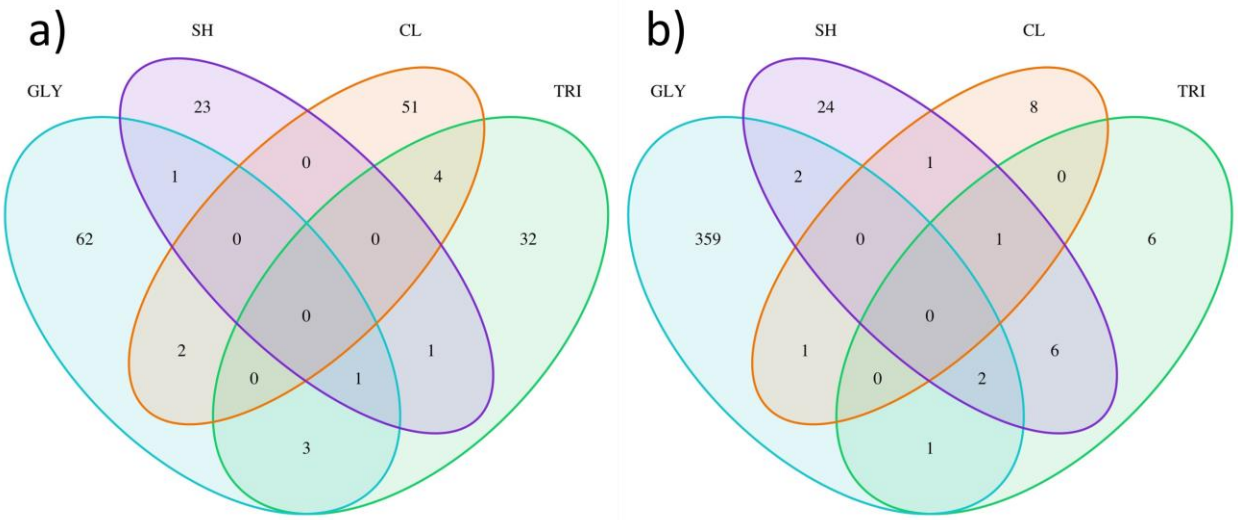


Figure 5 Overlapping sRNA: Some feature-mapping sRNA were differentially upregulated (a) and downregulated (b) in different treatments.

Table 1 sRNA and GFF features: Manatee used the GFF file to identify features with differentially expressed sRNA. These features were sorted by class to determine where sRNA associated with each stress was acting. Differentially expressed clusters of sRNA that were identified by Manatee but not associated with any GFF feature are listed under unannotated.

	Up-Regulated sRNA				Down-Regulated sRNA			
	GLY	TRI	SH	CL	GLY	TRI	SH	CL
gene	24	33	13	48	261	3	4	2
ncRNA	2	0	0	0	6	1	2	0
pseudogene	4	0	2	0	5	1	1	0
rRNA	0	0	0	0	2	0	1	0
snoRNA	0	0	0	0	22	0	1	0
snRNA	0	0	0	1	0	0	0	0
transposable element	31	2	7	0	31	2	9	7
transposable element gene	2	2	2	0	8	0	1	0
tRNA	6	4	2	8	30	9	17	2
unannotated	42	21	23	51	282	78	71	104

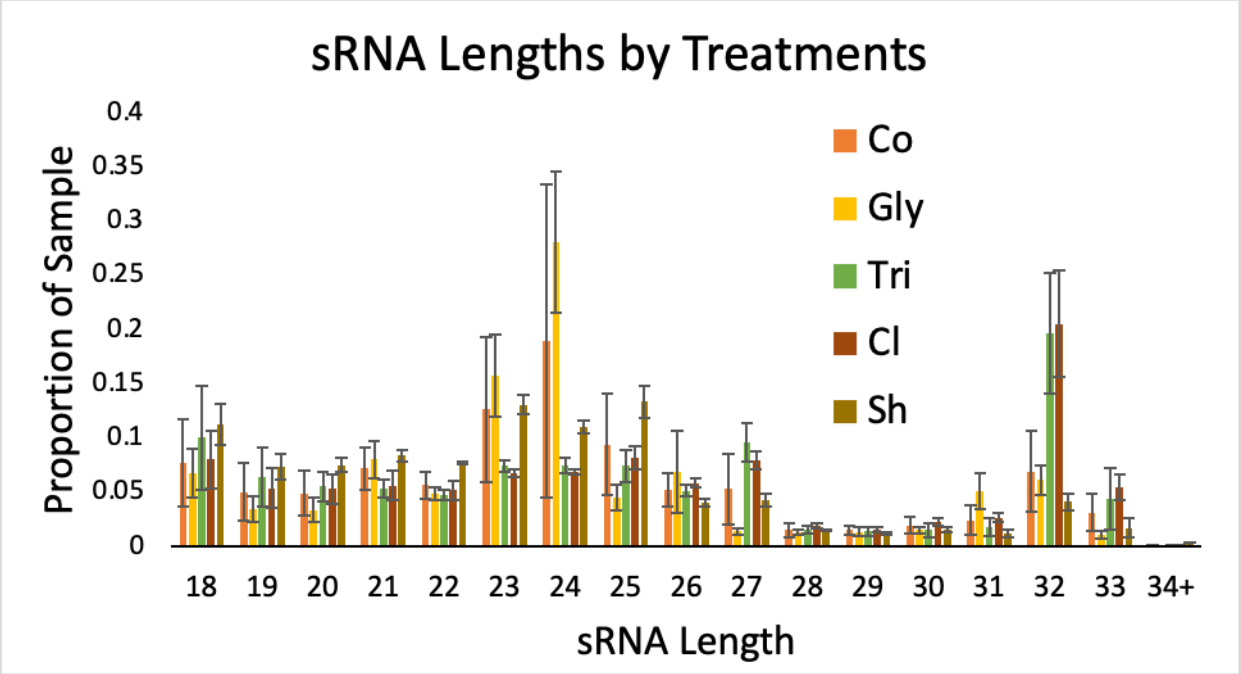


Figure 6 sRNA Lengths by Treatment: The length of sRNA associated with each treatment was compared across the samples. The error bars denote the standard deviation.

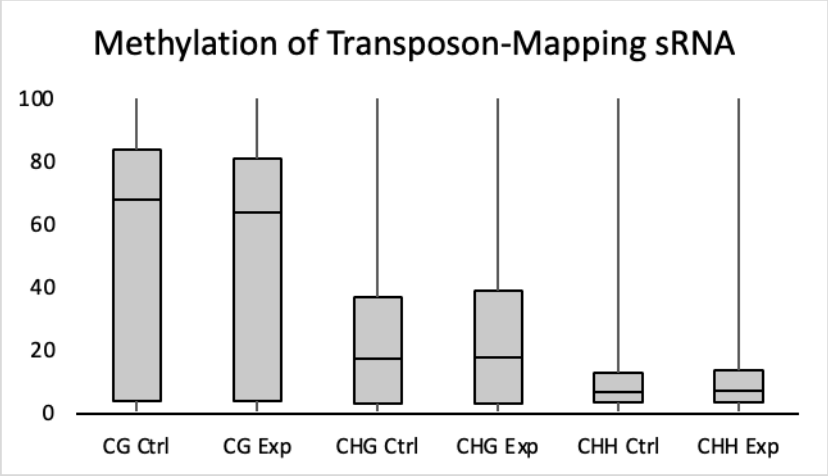


Figure 7 Methylation of Transposons and sRNA: A box plot of the methylation levels in each context for sRNA that mapped to transposons is shown for control (Ctrl) and treated (Exp) plants.

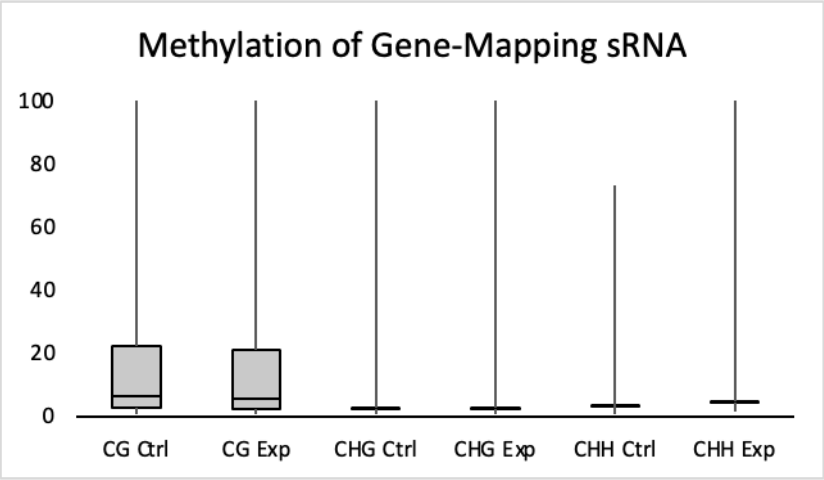


Figure 8 Methylation of Genes and sRNA: A box plot of the methylation levels in each context for sRNA that mapped to genes is shown for control (Ctrl) and treated (Exp) plants.

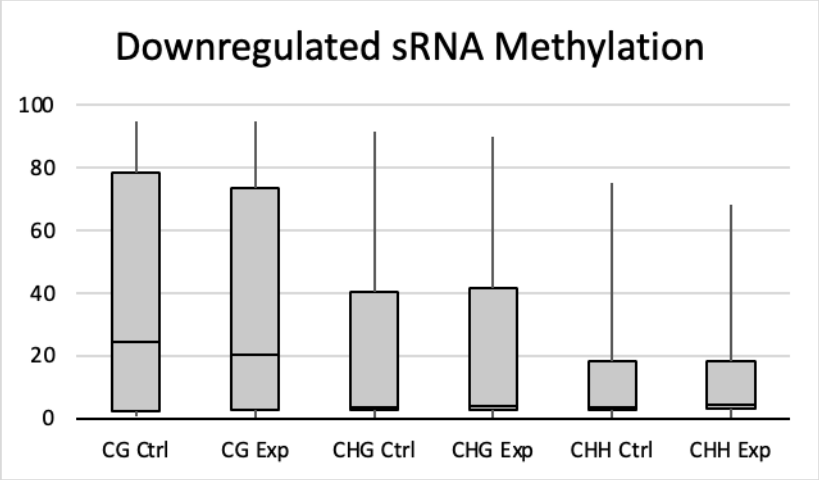


Figure 9 Methylation of Downregulated sRNA Sites: A box plot of methylation at sRNA mapping sites that were downregulated under stress treatments.

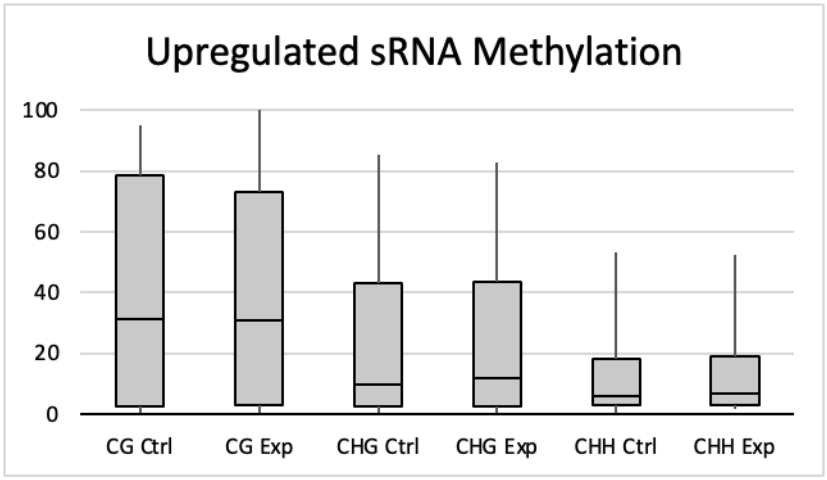


Figure 10 Methylation of Upregulated sRNA Sites: A box plot of methylation at sRNA mapping sites that were upregulated under stress treatments.

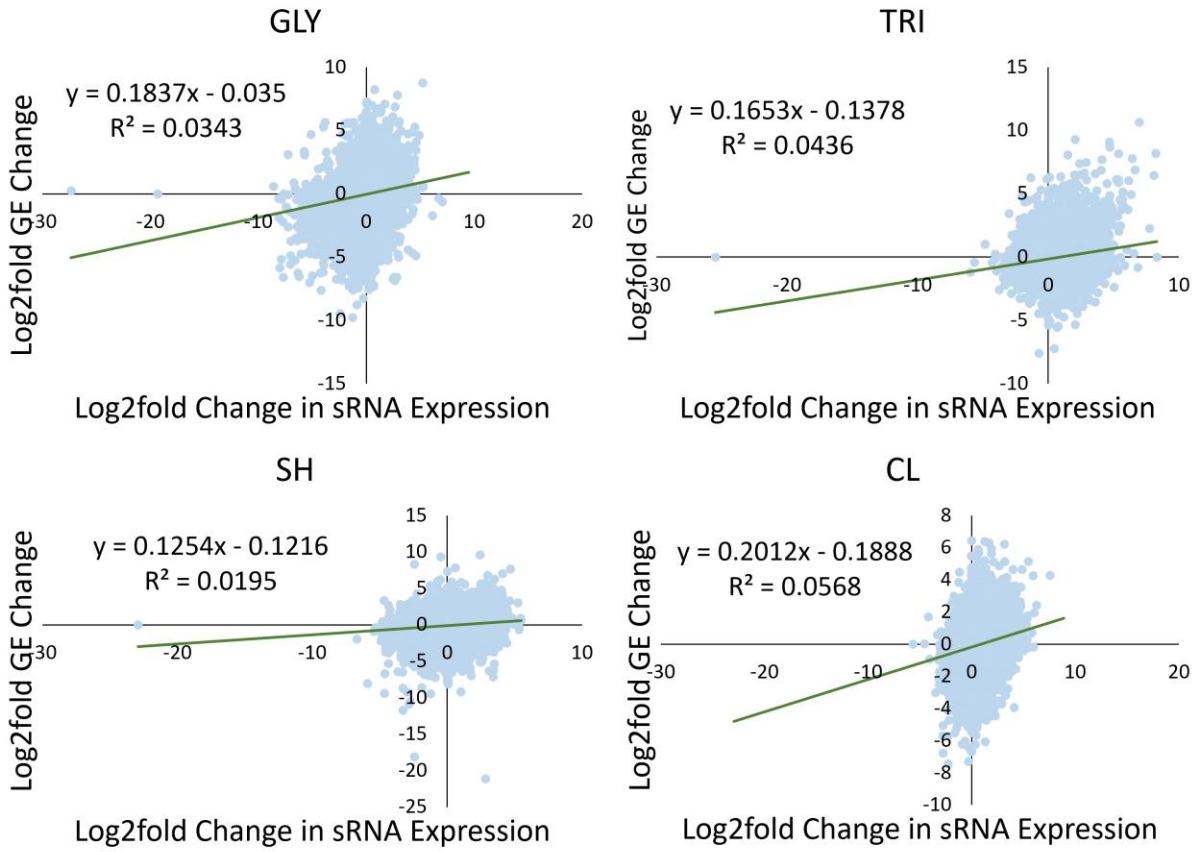


Figure 11 Changes in sRNA and Gene Expression: The log_2fold change in expression of sRNA and that of genes was compared. A simple linear trendline and associated R^2 values were generated.

Methylation Analysis:

The overall methylation values were similar for all the treatment groups. (Figure 12) Glyphosate-treated plants had higher CHH methylation and lower levels of CG methylation, and shade-treated plants also had higher levels of CHH and CHG methylation than the other treatment groups. Methylation profiles were clustered with Methylkit based on each context against the control plants (Figures 13-Figure 15). The glyphosate-treated plants (GLY) typically clustered with their associated controls. As the treatments caused delays in development, new controls were planted to flower at the same time as the experimental groups. The glyphosate control plants (COG) were grown at the same time as the glyphosate treated plants. They clustered together in the CG and CHG context, which reflected the sRNA PCA results, but there was more dissimilarity with CHH methylation. Mechanically clipped plants' (CL) methylation in the CG context was most similar to the other controls, and both shaded (SH) and trifloxysulfuron treated plants (TRI) were mixed together. SH was the only treatment where all replicates remained grouped in the CHG context. In the CHH context, GLY and SH replicates were similar to their treatment group, but the other samples did not group by treatment.

The methylation levels were then compared for the genomic features: genes, promoters, transposons, lncRNA, and sRNA. The methylation levels of each treatment for each context were merged with the features, including the various combinations in which they overlap. It was hypothesized that the levels of methylation for these features would follow the overall methylation levels with potential for changes in the contexts with DMRs. The methylation for each class of feature and combination generally followed the overall methylation values, but there did not appear to be broad shifts in the methylation of groups of features between treatments. The highest methylation levels typically occurred within bases that were associated with transposons and sRNA, where RdDM has expected involvement.[72]

When the methylation levels were compared at an individual base level between treated and the untreated controls, GLY had the most differentially methylated cytosines (DMCs) that were primarily hypomethylated with reduced methylation levels. (Table 2) TRI followed this trend and had the third most DMCs. SH had more hypermethylated sites with increased methylation levels than any other treatment. Another software package was used to find differentially methylated regions (DMRs), and all of the significant sites with changes of at least 10% were hypomethylated. (Table 3) Most occurred in herbicide-treated plants. Across chromosomes, the methylation levels followed consistent patterns with all treatments, with slight variations in the overall levels that reflected the differences in the overall methylation levels. These differences were most pronounced near the centromere (Figure 16). The pattern of methylation was similar throughout the mitochondria, and the low levels of methylation in the chloroplast were expected. (Figure 17)

When the base expression of genes was compared to the control methylation levels, higher levels of CG methylation had a positive trend with higher levels of expression, and non-CG methylation had a negative trend (Figure 18). The variance of CG, CHG, and CHH methylation significantly accounted for 9%, 17%, and 16% of the variance of gene expression, respectively. There was a cluster of high CG methylated genes with lower levels of expression. This seemed interesting. When changes in DNA methylation were compared to changes in expression, little relationship could be determined (Figure 19). CG methylation significantly accounted for 4% in mechanically clipped, 13% in glyphosate treated, and 4% in shaded plants.

CHG methylation accounted for 5% of the variation in trifloxysulfuron treated plants, and CHH accounted for 7% of the variation in gene expression of glyphosate treated plants.

To attempt to see if there were overall larger shifts in the pattern of methylation within genes, the methylation of all genes in all treatment groups was broken into segments and clustered based on their pattern of methylation. The elbow method was used on the distortion for each cluster number to determine the number of clusters (Figure 20).[73],[74] Four, three, and four values of were selected as the k-value for kMeans clustering of CG, CHG, and CHH methylation, respectively. The number of clusters were chosen from a similar study that used methylation to differentiate pseudogenes from expressed genes, but the authors did not find a strong relationship between patterns of methylation and expression.

For CG methylation, there was one cluster with high methylation in and near the gene, and another had high CG methylation in the gene body, with lower levels at the promoter. (Figure 21a) The other two groups had mostly lower CG methylation, but one had high levels of methylation in the promoter. Of the genes that clustered with different groups between treatments, only one involved a shift from a group promoter CG methylation to high CG methylation, and there were no transitions between the high and low groups. (Figure 21b) The average log2 fold change in expression of genes remained between -1 and 1, with larger standard deviations indicating that they are around zero. (Figure 21c) The one gene that shifted from promoter CG methylation to high CG methylation in clipped plants had a large log2fold change (-2.85) in expression. The overall base mean expression of genes within each cluster varied widely for all clusters, but that of high CG methylated genes was consistently lower.

The three CHG methylation clusters involved promoter CHG methylation, high CHG methylation, and low CHG methylation. (Figure 22a) Most of the genes that changed clusters between the low and promoter CHG methylation clusters. (Figure 22b) The log2fold change in expression of genes that changed clusters again remained close to zero with high variation (Figure 22c). Genes with high levels of CHG methylation were typically expressed at lower levels in all treatments, but genes that had lower or promoter CHG methylation varied widely in their expression levels (Figure 22d).

Another cluster with promoter methylation was identified for CHH methylation, as well as clusters with high, mid, and low levels of CHH methylation (Figure 23a). Again, there were few genes that transitioned from a promoter or low CHH methylation cluster to the highest CHH methylation cluster, and the log2fold changes in expression were close to zero (Figure 23b and c). Members of high and mid levels of CHH methylation clusters had consistent lower levels of expression, but there was wide variation in the expression of genes in the promoter and low CHH methylation clusters. (Figure 23d)

When the average methylation of other features was compared across treatments, higher methylation averages were consistently observed for regions containing transposons identified with RepeatModeler and sRNA sites with a total length of at least 10 base pairs and at least 21 reads mapping fully to the region (Figure 24). This was also observed consistently when the regions of overlap were compared (Figure 25). Again, average methylation levels were

consistently higher in regions where sRNA and transposons overlapped. There did not appear to be broad shifts in methylation between features among the treatments.

RepeatModeler was used to identify potential transposons and repeat regions as well as classify them.[70] There are different families of transposons, and some transposons have multiple copies in the genome, so we wanted to know if similar transposons were methylated differently in response to our treatments.[75] The methylation levels of the transposon families were identified for each treatment to determine if there were transposon families that were broadly differentially methylated in a long-term response to stress. Transposon families with fewer than 20 members were excluded, and then the difference in methylation against the control was compared in each context for each transposon family. The five transposon families with the greatest difference were selected to look for larger differences in methylation that could be associated with a transposon's sequence. (Table 5-7)

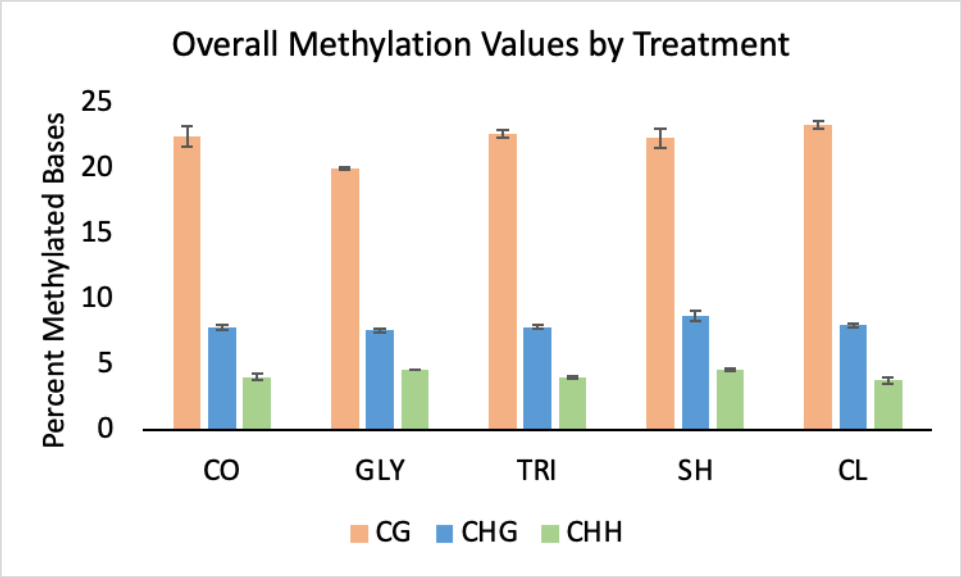


Figure 12 Overall Methylation by Treatment: The average methylation values of all bases for each methylation context

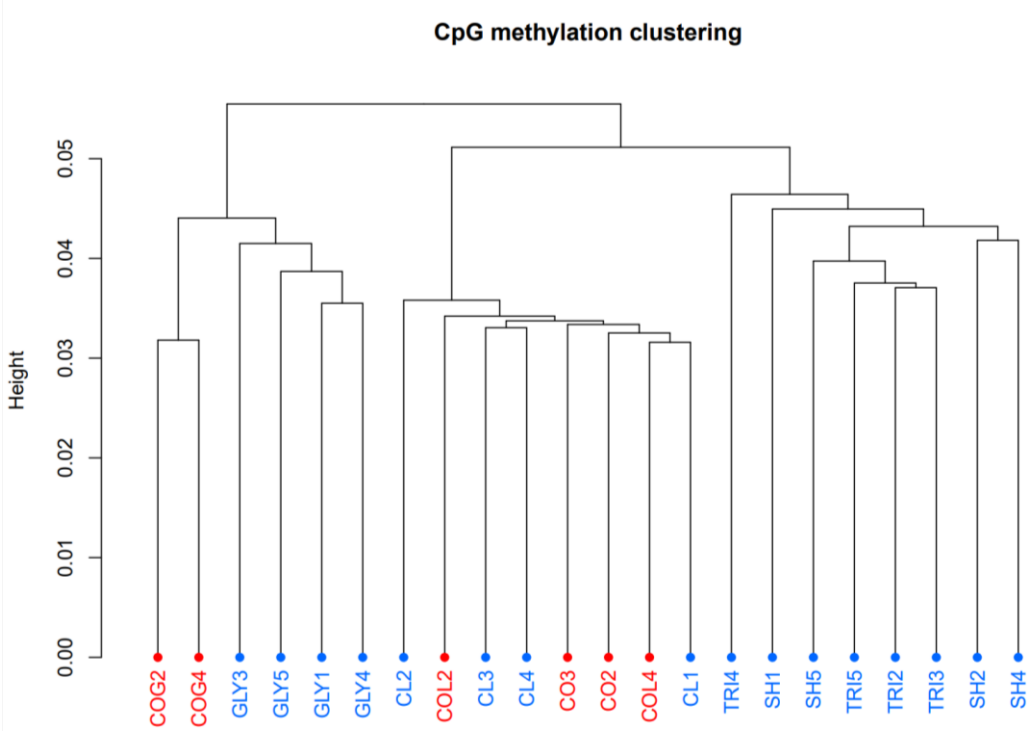


Figure 13 CG Methylation Clustering: The Bismark cx report files were clustered with Ward's hierarchical method via Methylkit using the correlation distance metric. The treated plants (blue) are glyphosate-treated (GLY), mechanically clipped (CL), trifloxysulfuron-treated (TRI), and shaded (SH). COG is the controls (red) that matured alongside the glyphosate treated plants; COL matured with clipping; and CO with the other treatments.

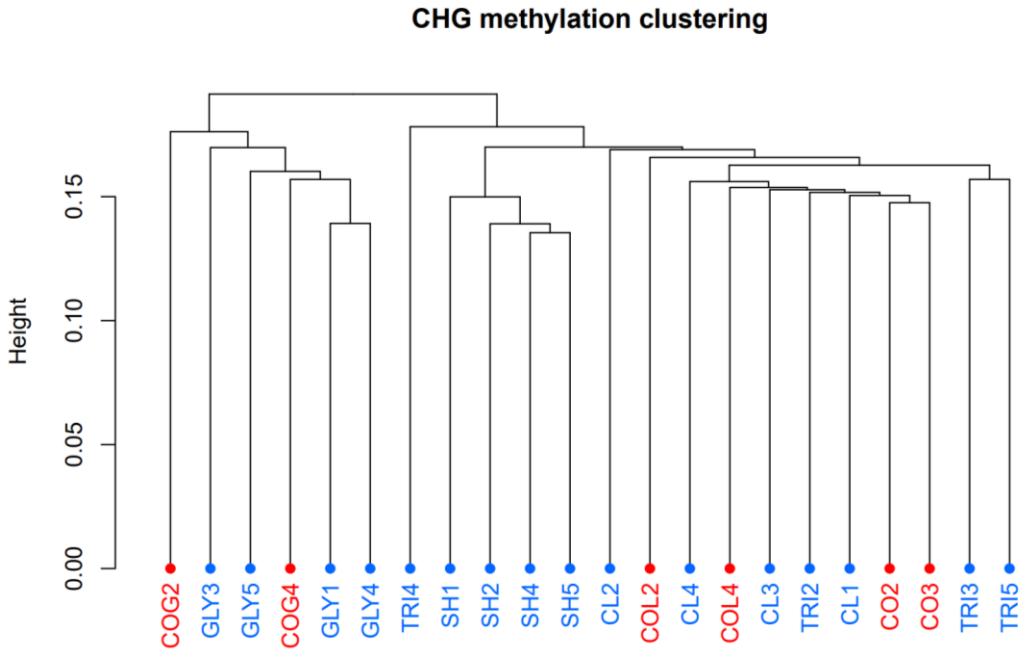


Figure 14 CHG Methylation Clustering: The Bismark cx report files were clustered with Ward's hierarchical method via Methylkit using the correlation distance metric. The treated plants (blue) are glyphosate-treated (GLY), mechanically clipped (CL), trifloxysulfuron-treated (TRI), and shaded (SH). COG is the controls (red) that matured alongside the glyphosate treated plants; COL matured with clipping; and CO with the other treatments.

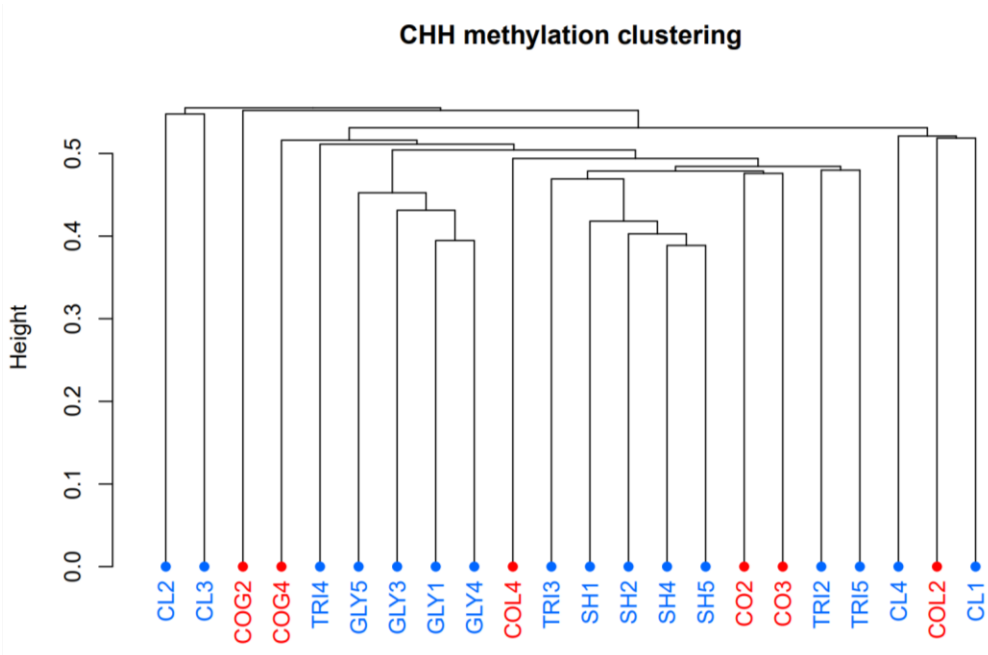


Figure 15 CHH Methylation Clustering: The Bismark cx report files were clustered with Ward's hierarchical method via Methylkit using the correlation distance metric. The treated plants (blue) are glyphosate-treated (GLY), mechanically clipped (CL), trifloxysulfuron-treated (TRI),

and shaded (SH). COG is the controls (red) that matured alongside the glyphosate treated plants; COL matured with clipping; and CO with the other treatments.

Table 2 DMCs: DMCs identified using Methylkit in all contexts for each treatment.

	hypermethylated			**hypomethylated**		
	CpG	CHG	CHH	CpG	CHG	CHH
GLY	36	13	573	2026	23	21
TRI	42	1	4	810	9	2
CL	1	0	0	0	0	1
SH	8	68	879	121	1	3

Table 3 DMRs: DMRs identified using DMRCaller in all contexts for each treatment.

	hypermethylated			**hypomethylated**		
	CpG	CHG	CHH	CpG	CHG	CHH
GLY	0	0	0	54	4	8
TRI	0	0	0	48	0	0
CL	0	0	0	0	0	0
SH	0	0	0	6	0	0

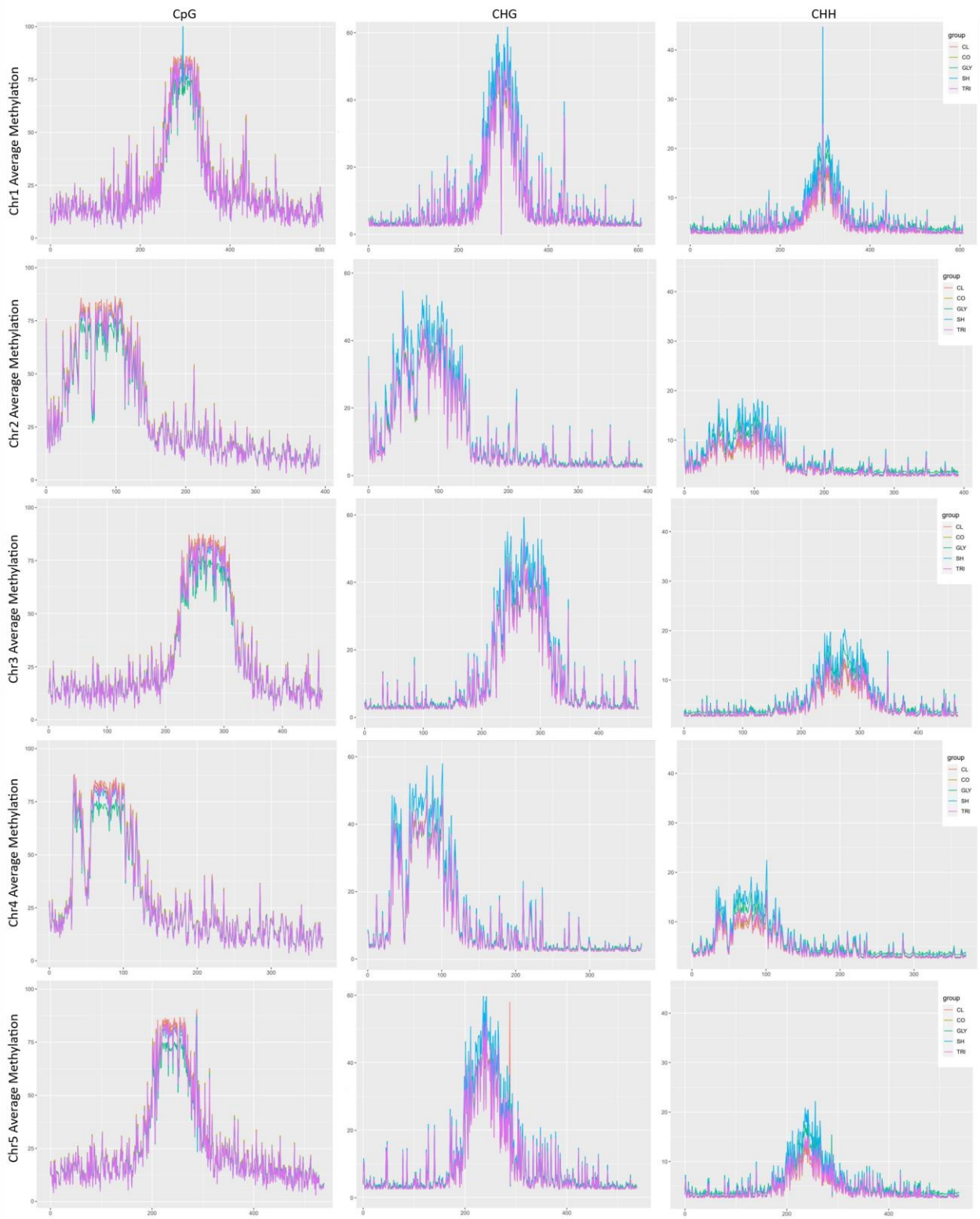


Figure 6 Chromosome Methylation by Treatments: Average percent methylation of chromosomes was calculated for in 50000 bp intervals in each cytosine context for each treatment: control (gold), glyphosate (green), trifloxysulfuron (purple), shade (blue), and clipped (red). The densely methylated centromeres are visible for each chromosome.[76]

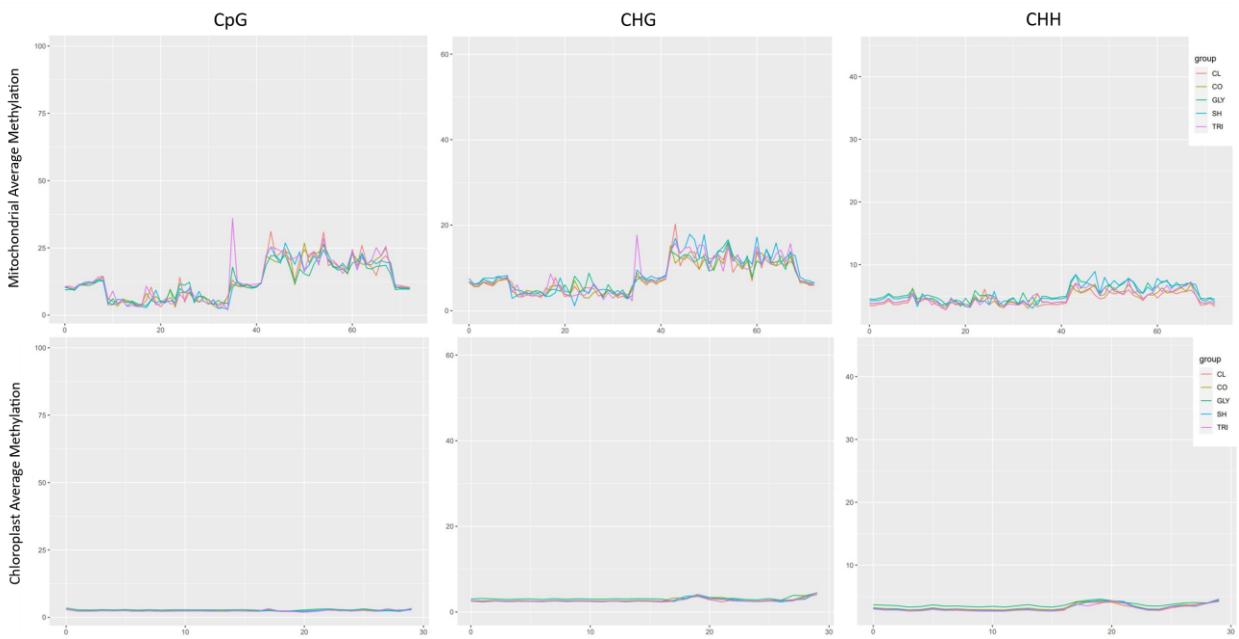


Figure 17 Plastid Chromosome Methylation by Treatments: Average percent methylation of plastids (mitochondria above and chloroplast below) was calculated for in 5000 bp intervals in each cytosine context for each treatment: control (gold), glyphosate (green), trifloxysulfuron (purple), shade (blue), and clipped (red).

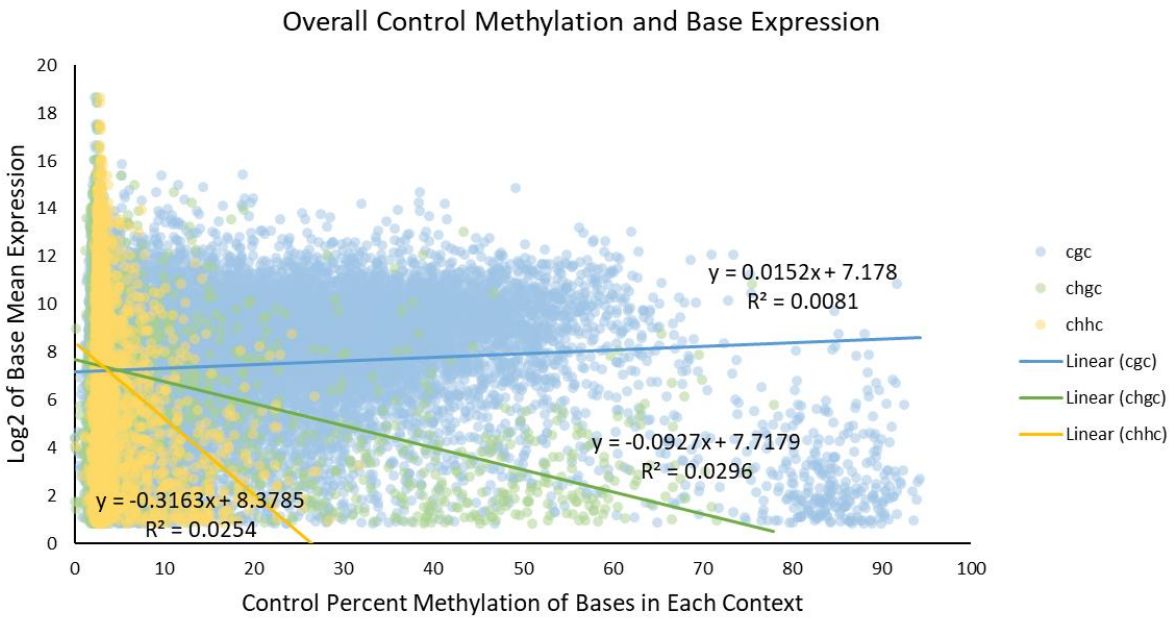


Figure 18 Methylation Levels and Expression: The base mean expression of genes was compared to the methylation levels of the control plant genes.

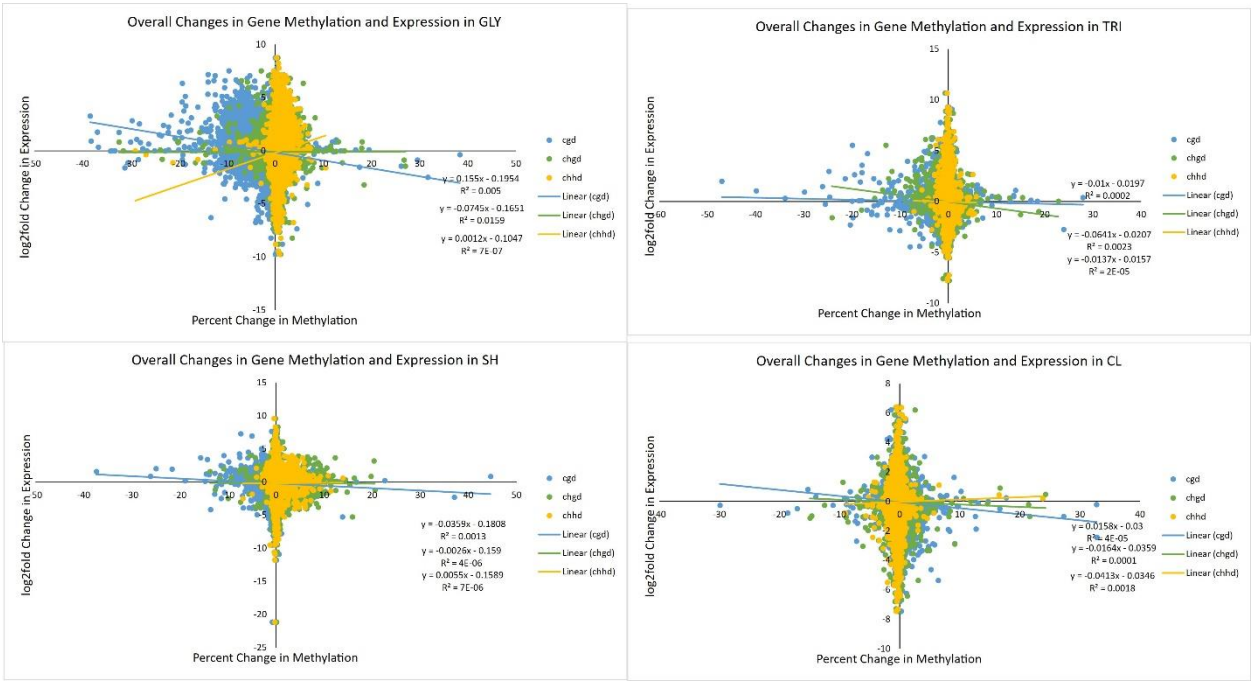


Figure 19 Changes in Methylation and Gene Expression: The percent change in genes' methylation in each context was compared to the change in their expression. A linear trendline is displayed for each context with the gene expression. This was repeated for all four treatments: glyphosate (GLY), trifloxysulfuron (TRI), shade (SH), and clipping (CL).

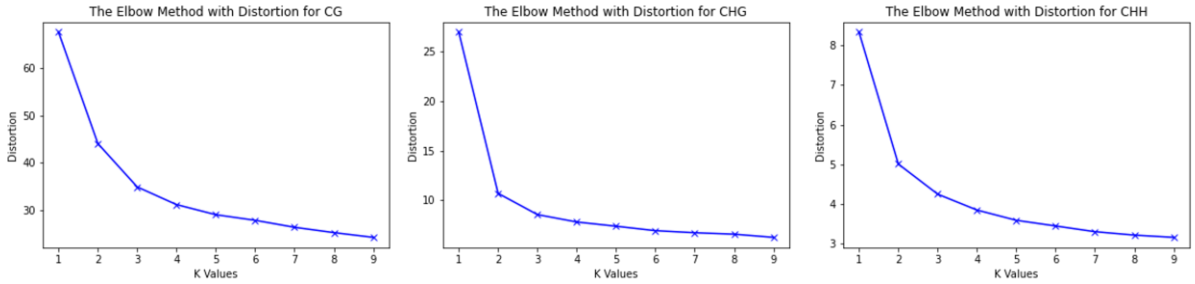
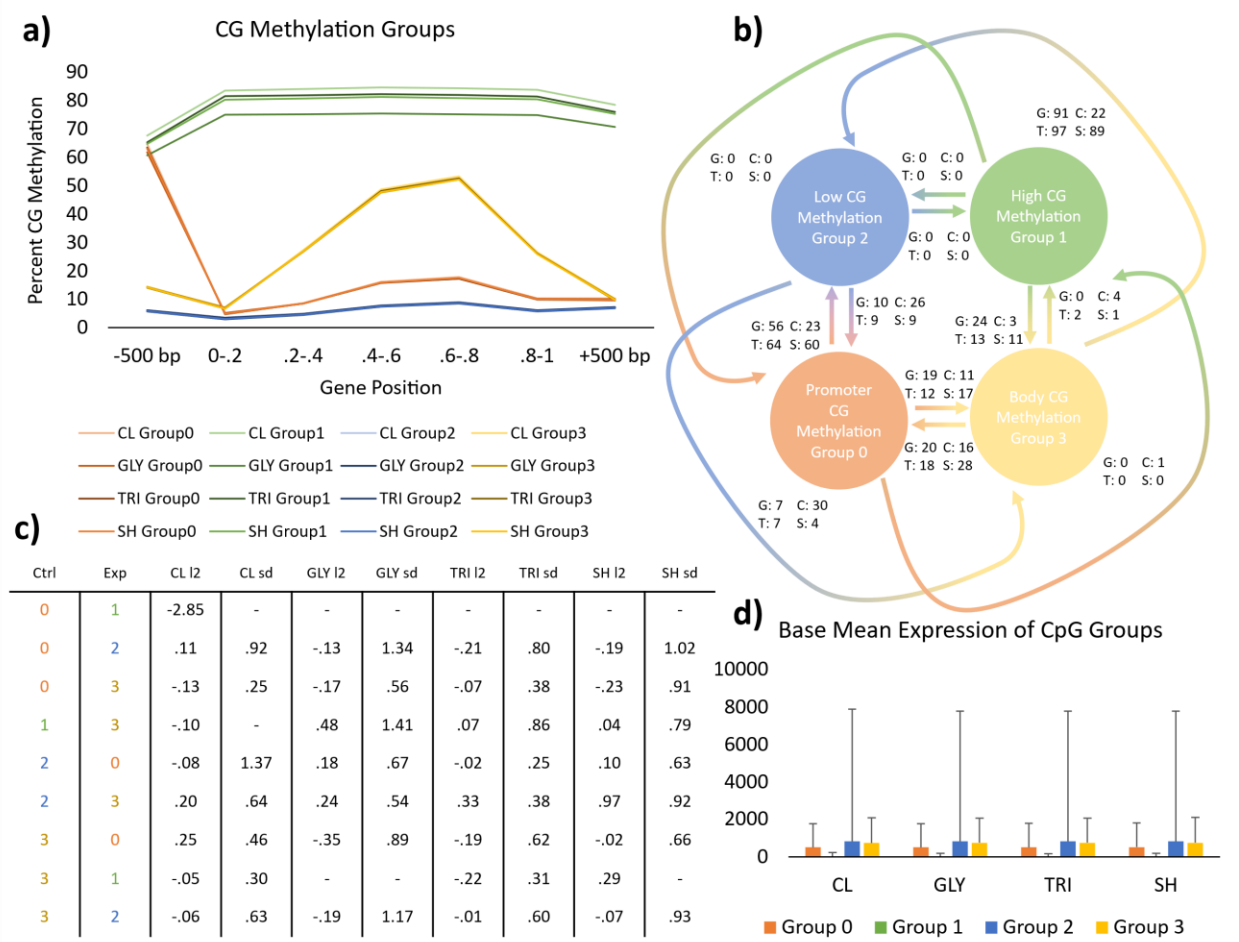
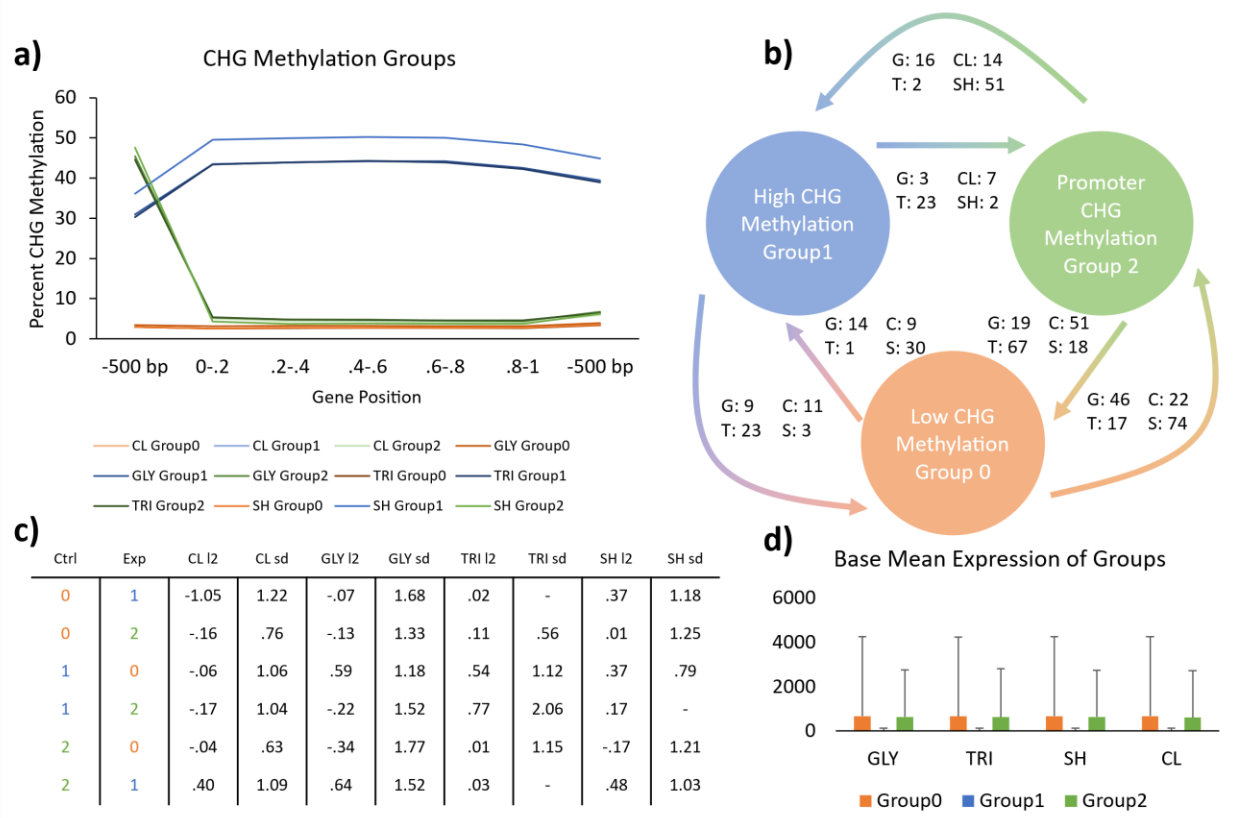


Figure 20 Distortion Elbow Graphs for kMeans: The distortion was plotted for different potential values of “k” for kMeans cluster numbers for seven sections of methylation in genes for each methylation context.



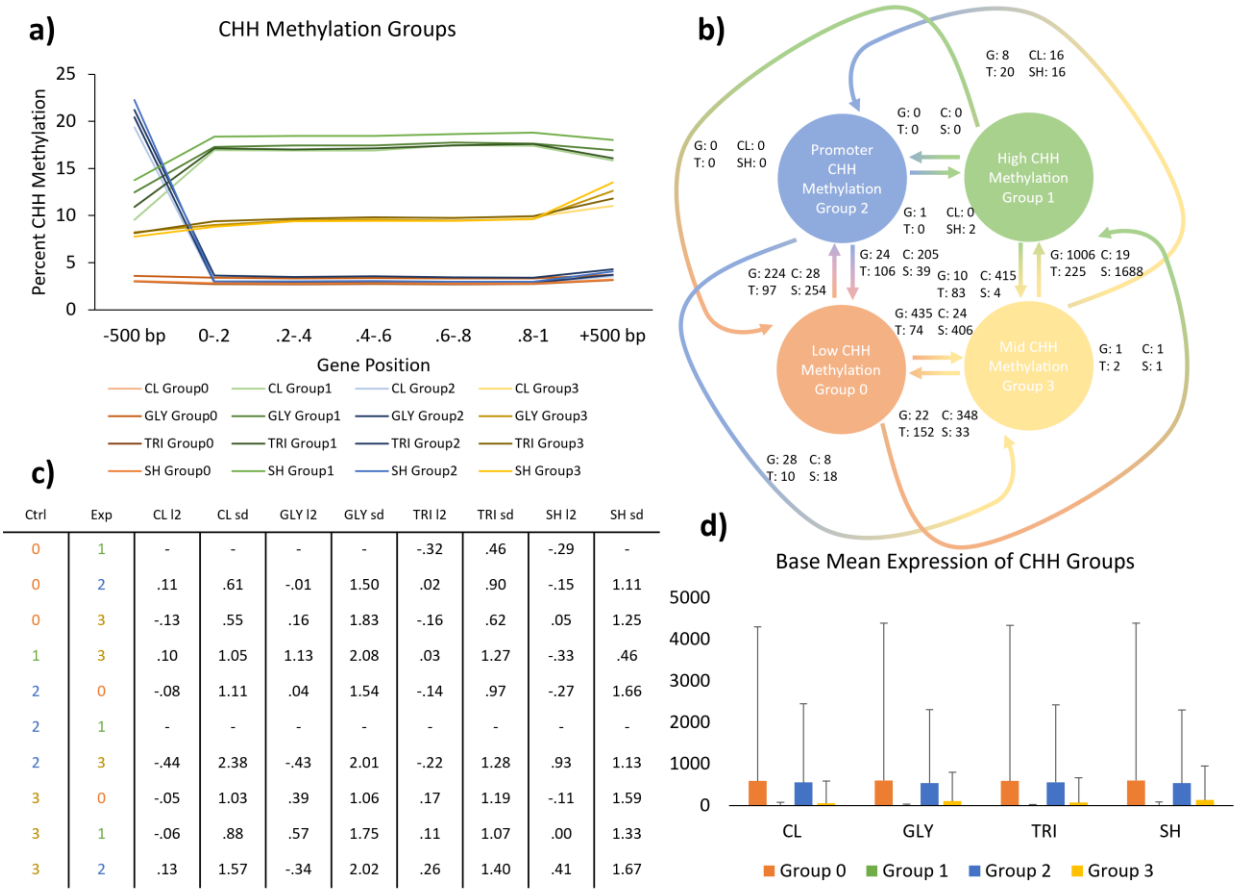
Ctrl	Exp	CL l2	CL sd	GLY l2	GLY sd	TRI l2	TRI sd	SH l2	SH sd
0	1	-2.85	-	-	-	-	-	-	-
0	2	.11	.92	-.13	1.34	-.21	.80	-.19	1.02
0	3	-.13	.25	-.17	.56	-.07	.38	-.23	.91
1	3	-.10	-	.48	1.41	.07	.86	.04	.79
2	0	-.08	1.37	.18	.67	-.02	.25	.10	.63
2	3	.20	.64	.24	.54	.33	.38	.97	.92
3	0	.25	.46	-.35	.89	-.19	.62	-.02	.66
3	1	-.05	.30	-	-	-.22	.31	.29	-
3	2	-.06	.63	-.19	1.17	-.01	.60	-.07	.93

Figure 21 CG Methylation Clusters: The average CG methylation of four kMeans clusters in the 500 bp upstream of the gene, 20% intervals throughout the gene, and 500 bp downstream is shown **(a)**. The number of genes that moved from one cluster in control treated plants to a different cluster in treated plants is shown, with glyphosate-treated (G), clipped (C), trifloxysulfuron-treated (T), and shaded (S) **(b)**. The average logfold2 change in gene expression (l2) for the genes that moved between clusters and standard deviation (sd) is shown for glyphosate-treated (GLY), clipped (CL), trifloxysulfuron-treated (TRI), and shaded (SH) **(c)**. The base mean expression of the genes in each cluster for each treatment group is shown with the standard deviation denoted by error bars **(d)**.



Ctrl	Exp	CL l2	CL sd	GLY l2	GLY sd	TRI l2	TRI sd	SH l2	SH sd
0	1	-1.05	1.22	-.07	1.68	.02	-	.37	1.18
0	2	-.16	.76	-.13	1.33	.11	.56	.01	1.25
1	0	-.06	1.06	.59	1.18	.54	1.12	.37	.79
1	2	-.17	1.04	-.22	1.52	.77	2.06	.17	-
2	0	-.04	.63	-.34	1.77	.01	1.15	-.17	1.21
2	1	.40	1.09	.64	1.52	.03	-	.48	1.03

Figure 22 CHG Methylation Clusters: The average CHG methylation of three kMeans clusters in the 500 bp upstream of the gene, 20% intervals throughout the gene, and 500 bp downstream is shown **(a)**. The number of genes that moved from one cluster in control treated plants to a different cluster in treated plants is shown, with glyphosate-treated (G), clipped (C), trifloxysulfuron-treated (T), and shaded (S) **(b)**. The average logfold2 change in gene expression (l2) for the genes that moved between clusters and standard deviation (sd) is shown for glyphosate-treated (GLY), clipped (CL), trifloxysulfuron-treated (TRI), and shaded (SH) **(c)**. The base mean expression of the genes in each cluster for each treatment group is shown with the standard deviation denoted by error bars **(d)**.



Ctrl	Exp	CL l2	CL sd	GLY l2	GLY sd	TRI l2	TRI sd	SH l2	SH sd
0	1	-	-	-	-	-.32	.46	-.29	-
0	2	.11	.61	-.01	1.50	.02	.90	-.15	1.11
0	3	-.13	.55	.16	1.83	-.16	.62	.05	1.25
1	3	.10	1.05	1.13	2.08	.03	1.27	-.33	.46
2	0	-.08	1.11	.04	1.54	-.14	.97	-.27	1.66
2	1	-	-	-	-	-	-	-	-
2	3	-.44	2.38	-.43	2.01	-.22	1.28	.93	1.13
3	0	-.05	1.03	.39	1.06	.17	1.19	-.11	1.59
3	1	-.06	.88	.57	1.75	.11	1.07	.00	1.33
3	2	.13	1.57	-.34	2.02	.26	1.40	.41	1.67

Figure 23 CHH Methylation Clusters: The average CG methylation of four kMeans clusters in the 500 bp upstream of the gene, 20% intervals throughout the gene, and 500 bp downstream is shown **(a)**. The number of genes that moved from one cluster in control treated plants to a different cluster in treated plants is shown, with glyphosate-treated (G), clipped (C), trifloxysulfuron-treated (T), and shaded (S) **(b)**. The average logfold2 change in gene expression (l2) for the genes that moved between clusters and standard deviation (sd) is shown for glyphosate-treated (GLY), clipped (CL), trifloxysulfuron-treated (TRI), and shaded (SH) **(c)**. The base mean expression of the genes in each cluster for each treatment group is shown with the standard deviation denoted by error bars **(d)**.

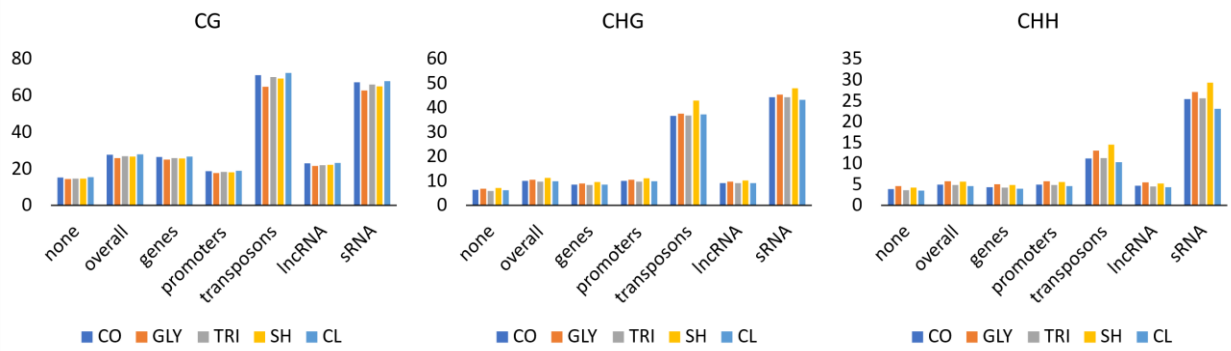


Figure 24 Average Methylation of Features: Overall methylation by genomic feature for Arabidopsis. None indicates that there is no feature there, while overall is the average methylation for all sites in that context. Each treatment group, control (CO), glyphosate-treated (GLY), trifloxysulfuron-treated (TRI), shade treated (SH), and clipped (CL) is listed for each feature type.

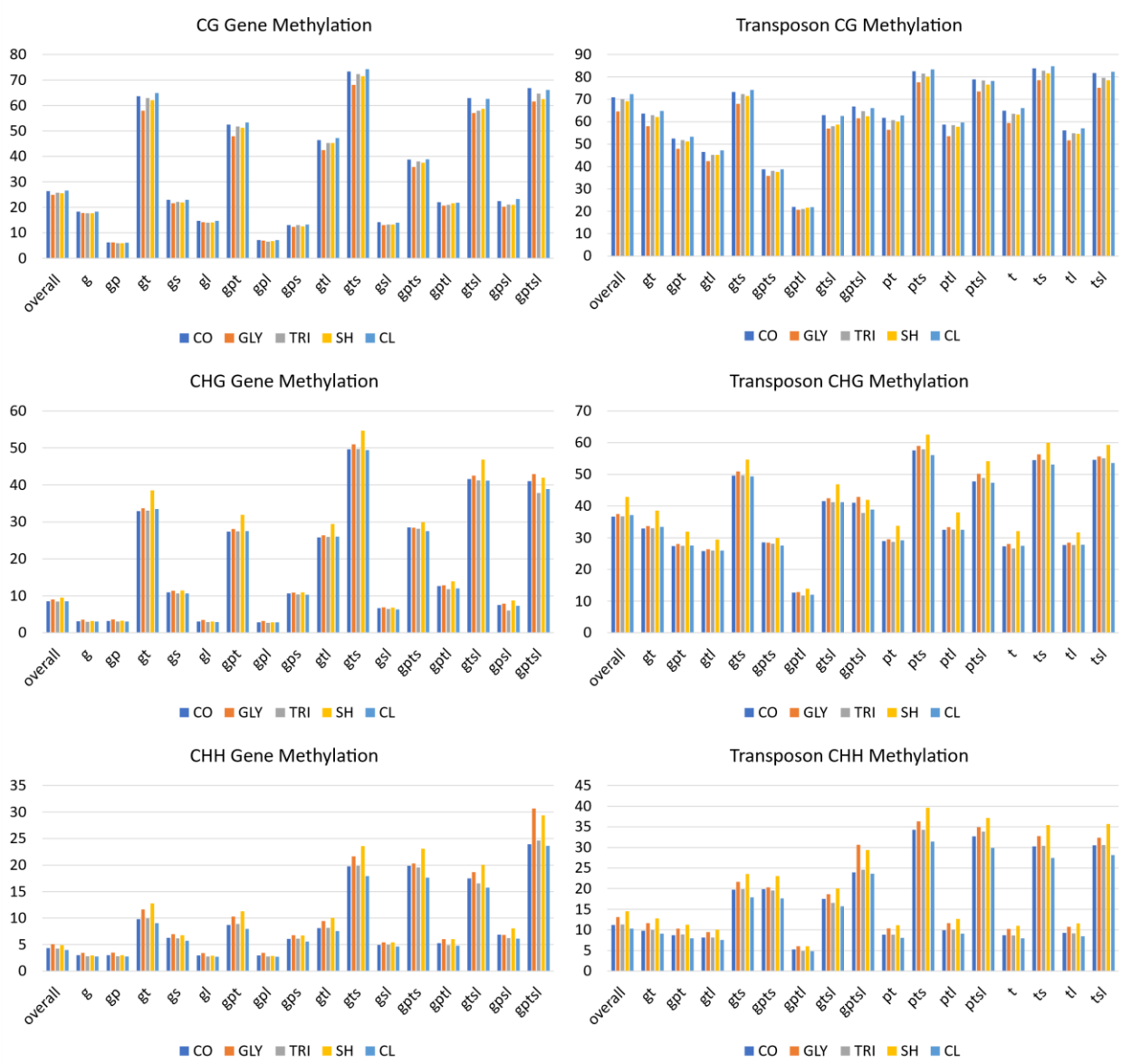


Figure 25 Gene and Transposon Overlapping Feature Methylation: Overall methylation for genes (g) as they overlap with other features: promoters (p), transposons (t), sRNA (s), and lncRNA (l). Overall is the average methylation for all sites in each context. Each treatment group, control (CO), glyphosate-treated (GLY), trifloxysulfuron-treated (TRI), shade treated (SH), and clipped (CL) is listed for each feature type.

Table 4 GLY Transposon classes with Methylation Differences: The top five transposon classes identified with RepeatModeler were identified in glyphosate-treated Arabidopsis.

CG			CHG			CHH		
transposon class	count	diff	transposon class	count	diff	transposon class	count	diff
(CTTTCT)n	22	2.609	(TTGTT)n	52	4.866	rnd-1_family-	55	5.102

						204		
(GGA)n	65	1.665	rnd-5_family-4829	33	4.118	(TTGTT)n	52	3.697
rnd-1_family-344	43	1.616	rnd-5_family-585	25	4.094	rnd-1_family-208	38	3.623
(CTTCT)n	33	1.402	rnd-5_family-17471	61	4.090	rnd-1_family-38	76	3.555
(AGG)n	27	1.390	(CTTCT)n	33	4.074	rnd-1_family-249	24	3.421
rnd-1_family-181	62	-9.005	(TGG)n	76	-1.598	(TGTTTT)n	28	-0.012
rnd-5_family-17471	61	-9.155	(AACA)n	74	-2.136	(GTTT)n	20	-0.276
rnd-1_family-239	30	-9.222	(TTCTTT)n	28	-2.166	rnd-4_family-653	160	-0.491
rnd-1_family-93	27	-9.226	(ACC)n	59	-3.224	(AGAA)n	30	-0.645
rnd-1_family-10	205	-9.241	(CCA)n	95	-4.404	(GAAA)n	23	-1.335

Table 5 TRI Transposon classes with Methylation Differences: The top five transposon classes identified with RepeatModeler were identified in trifloxysulfuron-treated Arabidopsis.

CG			CHG			CHH		
transposon class	count	diff	transposon class	count	diff	transposon class	count	diff
(C)n	40	1.527	(CCA)n	95	3.913	(ACC)n	59	1.877
(CTTCTT)n	25	1.505	(GGA)n	65	2.972	rnd-5_family-585	25	1.507
(AAAG)n	50	1.263	(ACC)n	59	2.842	(TTTGTT)n	22	1.393
rnd-5_family-532	93	1.133	rnd-5_family-4233	32	2.606	(GAG)n	38	1.133
rnd-1_family-344	43	1.077	(TTTTT)n	24	2.494	rnd-1_family-168	101	1.046
rnd-1_family-312	83	-4.978	(GAC)n	26	-3.377	(TCTTC)n	32	-1.354
(TATTTT)n	47	-5.066	(ATAAT)n	26	-3.546	rnd-1_family-312	83	-1.436
(TCTTT)n	40	-5.690	(AAAC)n	29	-4.391	(GAAA)n	23	-1.444
rnd-5_family-7698	38	-5.763	rnd-5_family-3943	26	-5.097	rnd-5_family-5702	55	-1.792
(TTTATT)n	54	-12.611	rnd-1_family-312	83	-7.980	(TCTCTT)n	32	-2.065

Table 6 SH Transposon classes with Methylation Differences: The top five transposon classes identified with RepeatModeler were identified in shade-treated Arabidopsis.

CG	CHG	CHH

transposon class	count	diff	transposon class	count	diff	transposon class	count	diff
(TTGTTT)n	25	2.837	rnd-1_family-31	45	10.730	rnd-1_family-31	45	6.468
(AAAAC)n	22	1.711	rnd-1_family-21	35	9.403	rnd-5_family-585	25	6.440
rnd-5_family-532	93	1.391	rnd-5_family-17471	61	9.308	rnd-5_family-4233	32	5.889
(TTTT)n	28	1.349	(TTCATC)n	28	9.149	rnd-1_family-93	27	5.732
(TTCTC)n	33	1.239	rnd-1_family-39	27	9.055	rnd-1_family-21	35	5.517
(AAAC)n	29	-4.632	(TGTTTT)n	28	-1.719	(GTTT)n	20	-0.754
rnd-4_family-653	160	-4.665	(TAAA)n	27	-1.843	(TCTTCA)n	41	-0.759
(TTTATT)n	54	-4.829	(AAAC)n	29	-1.981	(AGT)n	27	-0.772
rnd-5_family-7008	50	-5.070	rnd-5_family-3943	26	-2.178	(TCTCT)n	24	-1.045
rnd-5_family-1036	32	-5.127	(TTCTTT)n	28	-2.636	(TTCTC)n	33	-1.302

Table 7 CL Transposon classes with Methylation Differences: The top five transposon classes identified with RepeatModeler were identified in clipped Arabidopsis.

CG			CHG			CHH		
transposon class	count	diff	transposon class	count	diff	transposon class	count	diff
rnd-1_family-312	83	3.290	rnd-5_family-17471	61	5.0256	(ACC)n	59	1.091
(CTTCT)n	33	3.273	(CCA)n	95	3.082	(CTTTCT)n	22	0.796
(TTCTTT)n	28	3.169	(AGG)n	27	2.486	(GAG)n	38	0.670
(C)n	40	2.608	(GAA)n	384	2.155	(GCA)n	30	0.465
rnd-1_family-39	27	2.543	(ACC)n	59	1.991	(CCT)n	60	0.456
(AAAAC)n	22	-1.921	(TTGTTT)n	25	-2.805	rnd-1_family-167	58	-2.786
(TGTTTT)n	28	-2.034	rnd-1_family-163	104	-2.905	rnd-5_family-4233	32	-2.890
(TCTTCT)n	22	-2.088	rnd-5_family-3943	26	-3.187	rnd-1_family-192	75	-3.157
(TCTCTT)n	32	-2.351	(TTTTT)n	24	-3.346	rnd-1_family-163	104	-3.391
(TCTTT)n	40	-3.023	rnd-1_family-338	27	-3.797	rnd-1_family-309	56	-3.469

Discussion:

Herbicide resistance is a growing problem facing agriculture, and it may have an epigenetic component.[77] Epigenetics can involve changes to gene expression, and methylation is believed to be a major facilitator of these changes.[45] We sought to identify long-term sRNA and methylation changes in response to herbicides and other abiotic stresses used in weed control. Although the sRNA did not cluster perfectly by treatment in the PCA analysis, we still identified many differentially expressed regions of sRNA. There was little overlap of the differentially

expressed sRNA between the treatments, and it mapped to different features in the genome as well as some unknown regions. Despite its well-known role in post-transcriptional regulation, the general trend of sRNA levels with gene expression was positive.[78] This did not account for most of the variation. The methylation values in each context also did not cluster perfectly by treatment, but there were still individual differentially methylated cytosines and differentially methylated regions. The overall methylation levels of genes did not correlate well with expression. When genes were clustered by the pattern of methylation, the overall clusters matched those found in a previous study: gene body methylation, promoter methylation, high methylation, and low methylation.[35] However, the genes that changed clusters did not have an overall strong change in expression. There were many long-term methylation and sRNA changes associated with our treatments, but they did not correlate well with changes to gene expression.

We conducted a principal component analysis on the sRNA count matrix of the sRNA totals from the plants subjected to our treatments: glyphosate, trifloxysulfuron, clipping, and shade. There were also 6 controls samples, 2 that were grown to flower alongside the glyphosate-treated plants, 2 alongside the clipping, and 2 with the trifloxysulfuron and shade-treated plants. We wished to harvest tissue from plants at a similar developmental stage at the same time, so we had to grow additional control plants to account for that. The sRNA samples were expected to cluster with their associated treatments in the PCA (Fig 1-2), but the similarities between some of the samples from different treatment groups could be explained by the transient nature of sRNA.[79] It is possible that some sRNA responses to stress occurred before the samples were taken. The plants, aside from those treated with shade, had weeks to recover, which would only leave long-term signals. Much of the current sRNA could be directed towards development and other processes unrelated to our stresses.[80] This would be supported by the proximity of the glyphosate-treated samples to their associated controls in the PCA. Interestingly, there were higher levels of 24-nt sRNA, which is associated with maintenance RdDM, in glyphosate-treated Arabidopsis and its associated controls.[4] However, while those samples clustered together in the CG and CHG methylation context, the glyphosate control samples were dissimilar in the CHH context, which is heavily influenced by RdDM. If the 24-nt sRNA was maintenance RdDM, we would have expected CHH methylation to have overlap with those samples. This result was confusing, and it did not appear to support our hypothesis that active sRNAs may contribute to long-term stress responses.

Even so, some differentially expressed siRNA were identified for the treatment groups. This reflects other studies, which have identified some differentially expressed sRNA in stressed plants.[81,82] One found mostly down-regulated miRNA under glyphosate stress in rice.[81] Another study focused on profiling Arabidopsis under heat and salt stress and found that there were more up-regulated miRNA under stress.[82] Our glyphosate-treated plants had much more downregulated sRNA than upregulated, and this was different from the other treatments, which had comparatively similar levels of up- and downregulated sRNA. The level and function of sRNA in stress response may vary depending on the stress.

When the overall expression of genes was compared to sRNA levels, the sRNA did not explain most of the variation in gene expression, but the trend was consistently positive for all treatments. RNAi is a post-transcriptional regulation system, so some uncut transcripts were expected. Many sRNAs are cis-acting, so there would need to be some base level of expression

at that site to produce the sRNA.[83] Some level of gene transcription is required to produce most of its associated sRNA, and mRNA may but cut or sequestered to prevent later translation. Checking the resulting protein levels, which would indicate the strength of the post-transcriptional regulation, was beyond the scope of this study.

We expected to see higher levels of methylation, particularly in the CHH context, at sites of RdDM. RdDM adds methylation in all contexts, including CHH, which occurs infrequently in the euchromatin relative to the percent methylated cytosines in other contexts.[4,84] When the methylation of up- and downregulated sRNA sites was compared, there did not appear to be large changes in methylation between the treated plants and the controls. Methylation of sRNA sites associated with genes was lower than that near transposons, which was expected given that RdDM is often used for maintenance methylation of transposons.[4] However, there did not appear to be a great difference in methylation near sites of up- or downregulated sRNA. Perhaps much of the sRNA was not associated with methylation, or the demethylation preceding maintenance RdDM masked the effects of RdDM at those sites.[85]

The replicates were then subjected to hierarchical clustering based on their methylation data. It was hypothesized that the replicates would have the most in common with their associated treatment groups, but it seems that other processes within the plant have a stronger effect on Arabidopsis methylation. Another study attempted to form clusters based on methylation profiles in Arabidopsis under pathogen stress and also found clusters not associated with treatments in all contexts.[40] Perhaps the broad patterns of methylation are often dominated by processes unrelated to stress response, especially in the long term. Local changes are still possible. Even though the samples did not always cluster with their treatment groups, there were still differentially methylated cytosines for each treatment when compared to the control group.

Glyphosate-treated Arabidopsis had 2692 DMCs, far more differentially methylated cytosines than any other treatment. Perhaps the strength of this response to glyphosate could be influenced by the difficultly of metabolizing it within the plant; some herbicide could still be active within the plant and causing a continuous stress response.[86,87] Shaded and trifloxysulfuron-treated Arabidopsis also had several, with 1080 and 868, respectively. We detected 2 in mechanically clipped Arabidopsis. Aside from the overall numbers, it is worth noting that the types of DMCs varied between treatments. For example, shade-treated plants had over 80% of their DMCs in the hypermethylated CHH context. The majority of GLY and TRI DMCs were in the hypomethylated CG context. CL only had two DMCs, so it did not appear to have such a strong methylation response. Somewhat surprisingly, all the shade-treated CHH DMCs were not connected to the larger DMRs. Glyphosate and trifloxysulfuron treated plants had several hypomethylated DMRs in the CG context, and only the former had non-CG DMRs. None of the DMRs were connected to hypermethylation. On an even broader scale, across chromosomes, there did not appear to be any changes associated with the treatments. Methylation levels followed the same overall patterns across treatments even though some treatments had slightly different base levels of methylation. Those differences were most pronounced around the centromere, which does not have many genes. Methylation shifting away from the centromere under pathogen stress has been observed in *A. thaliana*.[88]

When the percent change in methylation was compared to the log2fold change in expression, there did not appear to be a strong relationship with any of the treatments. When overall methylation levels were compared to the log of the base mean expression, CG methylation tended to increase slightly with expression, but there was a cluster of highly methylated genes with lower levels of expression. Non-CG methylation tended to decrease with expression, which is consistent with the literature, but none of these trends were significant.[35] There are many factors that can affect gene expression, including histone modifications, binding proteins, and transcription factors.

Clustering the genes by methylation levels across them revealed patterns consistent with the literature: promoter methylation, gene body methylation, no methylation, and high methylation.[35] The gene expression of those individual clusters loosely followed trends reported in the literature, with low levels of expression in genes with high methylation. However, there was great variation within these groups, which indicates that methylation is far from the only factor determining expression levels. The pattern of methylation was classified into clusters for each gene under each treatment, and the expression of genes that switched from one grouping to another was determined. Most of the genes that changed clusters shifted between similar groups. For example, genes with low or promoter methylation, which has lower methylation levels for every other region, almost never transferred to the highest methylation cluster.

An experiment that could test the influence of methylation on expression could focus on the expression of some genes with the methylation processes knocked out.[31] The expression of those genes could be quantified, and then compared to that of a control plant to see if expression is affected by the absence of methylation. Also, there is a newer technique that involves hijacking the RdDM RNAi pathway to direct methylation in a manner similar to CRISPR Cas9.[89] If methylation is induced to a new area, it would be interesting to see if the gene expression is affected. However, our data does not support broad changes in gene methylation in a long-term response to stress, so the sites would ideally be chosen from genes that appeared to have the largest methylation differences and the largest different in expression to potentially identify sites where DNA methylation influences gene expression in *A. thaliana.*

Our data did not support the hypothesis of directed methylation broadly influencing gene expression of *A. thaliana* plants following a period of stress. There are some transcription factors and other regulatory proteins whose binding is influenced by DNA methylation.[90] They could create a link between methylation and expression in individual sites. Another study linked changes in methylation states to changes in histone proteins and the activity of a transposon.[42] At that site, it appeared that methylation could fluctuate to an extent, but if the histone modifications were compromised, it took several crosses to return to its original state. CHG methylation and histone modification are responsible for some regulation in a self-reinforcing loop. Perhaps the methylation changes must persist long enough or be drastic enough to influence histones to create a long-term effect on expression. While it is possible that methylation alone does not impact expression, many studies have shown altered expression in plants treated with demethylases or knockout mutations for major methylation processes have had altered gene expression and response to stress.[31,91]

References:

(1) Weinhold, B. Epigenetics: The Science of Change. *Environmental Health Perspectives* **2006**, *114* (3), A160–A167. https://doi.org/10.1289/ehp.114-a160.

(2) Waddington, C. H. The Epigenotype. *Int. J. Epidemiol.* **2012**, *41* (1), 10–13. https://doi.org/10.1093/ije/dyr184.

(3) Deichmann, U. Epigenetics: The Origins and Evolution of a Fashionable Topic. *Developmental Biology* **2016**, *416* (1), 249–254. https://doi.org/10.1016/j.ydbio.2016.06.005.

(4) Matzke, M. A.; Mosher, R. A. RNA-Directed DNA Methylation: An Epigenetic Pathway of Increasing Complexity. *Nat Rev Genet* **2014**, *15* (6), 394–408. https://doi.org/10.1038/nrg3683.

(5) Meyerowitz, E. M. Arabidopsis, a Useful Weed. *Cell* **1989**, *56* (2), 263–269. https://doi.org/10.1016/0092-8674(89)90900-8.

(6) Moore, L. D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacol* **2013**, *38* (1), 23–38. https://doi.org/10.1038/npp.2012.112.

(7) Lämke, J.; Bäurle, I. Epigenetic and Chromatin-Based Mechanisms in Environmental Stress Adaptation and Stress Memory in Plants. *Genome Biol* **2017**, *18* (1), 124. https://doi.org/10.1186/s13059-017-1263-6.

(8) Tamaru, H. Confining Euchromatin/Heterochromatin Territory: *Jumonji* Crosses the Line. *Genes Dev.* **2010**, *24* (14), 1465–1478. https://doi.org/10.1101/gad.1941010.

(9) Hyun, K.; Jeon, J.; Park, K.; Kim, J. Writing, Erasing and Reading Histone Lysine Methylations. *Exp Mol Med* **2017**, *49* (4), e324–e324. https://doi.org/10.1038/emm.2017.11.

(10) Zhao, T.; Zhan, Z.; Jiang, D. Histone Modifications and Their Regulatory Roles in Plant Development and Environmental Memory. *Journal of Genetics and Genomics* **2019**, *46* (10), 467–476. https://doi.org/10.1016/j.jgg.2019.09.005.

(11) Feng, S.; Cokus, S.; Zhang, X.; Chen, P.; Bostick, M.; Goll, M.; Hetzel, J.; Jain, J.; Struass, S.; Halpern, M.; Ukomadu, C.; Sadler, K.; Pradhan, S.; Pellegrini, M.; Jacobsen, S. Conservation and Divergence of Methylation Patterning in Plants and Animals. *PNAS* **2010**, *107* (19), 8689–8694. https://doi.org/10.1073/pnas.1002720107.

(12) Finnegan, E. J.; Peacock, W. J.; Dennis, E. S. Reduced DNA Methylation in Arabidopsis Thaliana Results in Abnormal Plant Development. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (16), 8449–8454.

(13) Ronemus, M. J.; Galbiati, M.; Ticknor, C.; Chen, J.; Dellaporta, S. L. Demethylation-Induced Developmental Pleiotropy in Arabidopsis. *Science* **1996**, *273* (5275), 654–657. https://doi.org/10.1126/science.273.5275.654.

(14) Kankel, M. W.; Ramsey, D. E.; Stokes, T. L.; Flowers, S. K.; Haag, J. R.; Jeddeloh, J. A.; Riddle, N. C.; Verbsky, M. L.; Richards, E. J. Arabidopsis MET1 Cytosine Methyltransferase Mutants. *Genetics* **2003**, *163* (3), 1109–1122.

(15) Zemach, A.; Kim, M. Y.; Hsieh, P.-H.; Coleman-Derr, D.; Eshed-Williams, L.; Thao, K.; Harmer, S. L.; Zilberman, D. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell* **2013**, *153* (1), 193–205. https://doi.org/10.1016/j.cell.2013.02.033.

(16) Wassenegger, M.; Heimes, S.; Riedel, L.; Sänger, H. L. RNA-Directed de Novo Methylation of Genomic Sequences in Plants. *Cell* **1994**, *76* (3), 567–576. https://doi.org/10.1016/0092-8674(94)90119-8.

(17) Niederhuth, C. E.; Bewick, A. J.; Ji, L.; Alabady, M. S.; Kim, K. D.; Li, Q.; Rohr, N. A.; Rambani, A.; Burke, J. M.; Udall, J. A.; Egesi, C.; Schmutz, J.; Grimwood, J.; Jackson, S. A.; Springer, N. M.; Schmitz, R. J. Widespread Natural Variation of DNA Methylation within Angiosperms. *Genome Biol* **2016**, *17* (1), 194. https://doi.org/10.1186/s13059-016-1059-0.

(18) Takuno, S.; Gaut, B. S. Gene Body Methylation Is Conserved between Plant Orthologs and Is of Evolutionary Consequence. *Proceedings of the National Academy of Sciences* **2013**, *110* (5), 1797–1802. https://doi.org/10.1073/pnas.1215380110.

(19) Iwasaki, M.; Paszkowski, J. Identification of Genes Preventing Transgenerational Transmission of Stress-Induced Epigenetic States. *Proceedings of the National Academy of Sciences* **2014**, *111* (23), 8547–8552. https://doi.org/10.1073/pnas.1402275111.

(20) Iwasaki, M.; Paszkowski, J. Epigenetic Memory in Plants. *EMBO J* **2014**, *33* (18), 1987–1998. https://doi.org/10.15252/embj.201488883.

(21) Hotchkiss, R. D. The Quantitative Separation of Purines, Pyrimidines, and Nucleosides by Paper Chromatography. *J. Biol. Chem.* **1948**, *175* (1), 315–332.

(22) Arber, W. Host-Controlled Modification of Bacteriophage. *Annu. Rev. Microbiol.* **1965**, *19* (1), 365–378. https://doi.org/10.1146/annurev.mi.19.100165.002053.

(23) Singer, J.; Roberts-Ems, J.; Riggs, A. D. Methylation of Mouse Liver DNA Studied by Means of the Restriction Enzymes Msp I and Hpa II. *Science* **1979**, *203* (4384), 1019–1021. https://doi.org/10.1126/science.424726.

(24) Schumacher, A. Microarray-Based DNA Methylation Profiling: Technology and Applications. *Nucleic Acids Research* **2006**, *34* (2), 528–542. https://doi.org/10.1093/nar/gkj461.

(25) Jones, P. A.; Taylor, S. M. Cellular Differentiation, Cytidine Analogs and DNA Methylation. *Cell* **1980**, *20* (1), 85–93. https://doi.org/10.1016/0092-8674(80)90237-8.

(26) Chuang, J. C.; Yoo, C. B.; Kwan, J. M.; Li, T. W. H.; Liang, G.; Yang, A. S.; Jones, P. A. Comparison of Biological Effects of Non-Nucleoside DNA Methylation Inhibitors versus 5-Aza-2′-Deoxycytidine. *Molecular Cancer Therapeutics* **2005**, *4* (10), 1515–1520. https://doi.org/10.1158/1535-7163.MCT-05-0172.

(27) Anastasiadi, D.; Esteve-Codina, A.; Piferrer, F. Consistent Inverse Correlation between DNA Methylation of the First Intron and Gene Expression across Tissues and Species. *Epigenetics & Chromatin* **2018**, *11* (1), 37. https://doi.org/10.1186/s13072-018-0205-1.

(28) Dubin, M. J.; Zhang, P.; Meng, D.; Remigereau, M.-S.; Osborne, E. J.; Paolo Casale, F.; Drewe, P.; Kahles, A.; Jean, G.; Vilhjálmsson, B.; Jagoda, J.; Irez, S.; Voronin, V.; Song, Q.; Long, Q.; Rätsch, G.; Stegle, O.; Clark, R. M.; Nordborg, M. DNA Methylation in Arabidopsis Has a Genetic Basis and Shows Evidence of Local Adaptation. *eLife* **2015**, *4*, e05255. https://doi.org/10.7554/eLife.05255.

(29) He, L.; Huang, H.; Bradai, M.; Zhao, C.; You, Y.; Ma, J.; Zhao, L.; Lozano-Durán, R.; Zhu, J.-K. DNA Methylation-Free Arabidopsis Reveals Crucial Roles of DNA Methylation in Regulating Gene Expression and Development. *Nat Commun* **2022**, *13* (1), 1335. https://doi.org/10.1038/s41467-022-28940-2.

(30) Popova, O. V.; Dinh, H. Q.; Aufsatz, W.; Jonak, C. The RdDM Pathway Is Required for Basal Heat Tolerance in Arabidopsis. *Molecular Plant* **2013**, *6* (2), 396–410. https://doi.org/10.1093/mp/sst023.

(31) Markus, C.; Pecinka, A.; Merotto, A. Insights into the Role of Transcriptional Gene Silencing in Response to Herbicide-Treatments in Arabidopsis Thaliana. *IJMS* **2021**, *22* (7), 3314. https://doi.org/10.3390/ijms22073314.
(32) Riggs, A. D. X Inactivation, Differentiation, and DNA Methylation. *Cytogenet Genome Res* **1975**, *14* (1), 9–25. https://doi.org/10.1159/000130315.
(33) Herman, J. G.; Graff, J. R.; Myöhänen, S.; Nelkin, B. D.; Baylin, S. B. Methylation-Specific PCR: A Novel PCR Assay for Methylation Status of CpG Islands. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (18), 9821–9826. https://doi.org/10.1073/pnas.93.18.9821.
(34) Krueger, F.; Andrews, S. R. Bismark: A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications. *Bioinformatics* **2011**, *27* (11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.
(35) Bewick, A. J.; Schmitz, R. J. Gene Body DNA Methylation in Plants. *Current Opinion in Plant Biology* **2017**, *36*, 103–110. https://doi.org/10.1016/j.pbi.2016.12.007.
(36) Takuno, S.; Gaut, B. S. Gene Body Methylation Is Conserved between Plant Orthologs and Is of Evolutionary Consequence. *Proceedings of the National Academy of Sciences* **2013**, *110* (5), 1797–1802. https://doi.org/10.1073/pnas.1215380110.
(37) Kawakatsu, T.; Huang, S. C.; Jupe, F.; Sasaki, E.; Schmitz, R. J.; Urich, M. A.; Castanon, R.; Nery, J. R.; Barragan, C.; He, Y.; Chen, H.; Dubin, M.; Lee, C.-R.; Wang, C.; Bemm, F.; Becker, C.; O'Neil, R.; O'Malley, R. C.; Quarless, D. X.; Schork, N. J.; Weigel, D.; Nordborg, M.; Ecker, J. R.; Alonso-Blanco, C.; Andrade, J.; Becker, C.; Bemm, F.; Bergelson, J.; Borgwardt, K.; Chae, E.; Dezwaan, T.; Ding, W.; Ecker, J. R.; Expósito-Alonso, M.; Farlow, A.; Fitz, J.; Gan, X.; Grimm, D. G.; Hancock, A.; Henz, S. R.; Holm, S.; Horton, M.; Jarsulic, M.; Kerstetter, R. A.; Korte, A.; Korte, P.; Lanz, C.; Lee, C.-R.; Meng, D.; Michael, T. P.; Mott, R.; Muliyati, N. W.; Nägele, T.; Nagler, M.; Nizhynska, V.; Nordborg, M.; Novikova, P.; Picó, F. X.; Platzer, A.; Rabanal, F. A.; Rodriguez, A.; Rowan, B. A.; Salomé, P. A.; Schmid, K.; Schmitz, R. J.; Seren, Ü.; Sperone, F. G.; Sudkamp, M.; Svardal, H.; Tanzer, M. M.; Todd, D.; Volchenboum, S. L.; Wang, C.; Wang, G.; Wang, X.; Weckwerth, W.; Weigel, D.; Zhou, X. Epigenomic Diversity in a Global Collection of Arabidopsis Thaliana Accessions. *Cell* **2016**, *166* (2), 492–505. https://doi.org/10.1016/j.cell.2016.06.044.
(38) Liang, L.; Chang, Y.; Lu, J.; Wu, X.; Liu, Q.; Zhang, W.; Su, X.; Zhang, B. Global Methylomic and Transcriptomic Analyses Reveal the Broad Participation of DNA Methylation in Daily Gene Expression Regulation of Populus Trichocarpa. *Front. Plant Sci.* **2019**, *10*, 243. https://doi.org/10.3389/fpls.2019.00243.
(39) Li, X.; Zhu, J.; Hu, F.; Ge, S.; Ye, M.; Xiang, H.; Zhang, G.; Zheng, X.; Zhang, H.; Zhang, S.; Li, Q.; Luo, R.; Yu, C.; Yu, J.; Sun, J.; Zou, X.; Cao, X.; Xie, X.; Wang, J.; Wang, W. Single-Base Resolution Maps of Cultivated and Wild Rice Methylomes and Regulatory Roles of DNA Methylation in Plant Gene Expression. *BMC Genomics* **2012**, *13* (1), 300. https://doi.org/10.1186/1471-2164-13-300.
(40) Stassen, J. H. M.; López, A.; Jain, R.; Pascual-Pardo, D.; Luna, E.; Smith, L. M.; Ton, J. The Relationship between Transgenerational Acquired Resistance and Global DNA Methylation in Arabidopsis. *Sci Rep* **2018**, *8* (1), 14761. https://doi.org/10.1038/s41598-018-32448-5.
(41) Luna, E.; Bruce, T. J. A.; Roberts, M. R.; Flors, V.; Ton, J. Next-Generation Systemic Acquired Resistance. *Plant Physiol.* **2012**, *158* (2), 844–853. https://doi.org/10.1104/pp.111.187468.

(42) Guo, W.; Wang, D.; Lisch, D. RNA-Directed DNA Methylation Prevents Rapid and Heritable Reversal of Transposon Silencing under Heat Stress in Zea Mays. *PLoS Genet* **2021**, *17* (6), e1009326. https://doi.org/10.1371/journal.pgen.1009326.
(43) Ong-Abdullah, M.; Ordway, J. M.; Jiang, N.; Ooi, S.-E.; Kok, S.-Y.; Sarpan, N.; Azimi, N.; Hashim, A. T.; Ishak, Z.; Rosli, S. K.; Malike, F. A.; Bakar, N. A. A.; Marjuni, M.; Abdullah, N.; Yaakub, Z.; Amiruddin, M. D.; Nookiah, R.; Singh, R.; Low, E.-T. L.; Chan, K.-L.; Azizi, N.; Smith, S. W.; Bacher, B.; Budiman, M. A.; Van Brunt, A.; Wischmeyer, C.; Beil, M.; Hogan, M.; Lakey, N.; Lim, C.-C.; Arulandoo, X.; Wong, C.-K.; Choo, C.-N.; Wong, W.-C.; Kwan, Y.-Y.; Alwee, S. S. R. S.; Sambanthamurthi, R.; Martienssen, R. A. Loss of Karma Transposon Methylation Underlies the Mantled Somaclonal Variant of Oil Palm. *Nature* **2015**, *525* (7570), 533–537. https://doi.org/10.1038/nature15365.
(44) Cubas, P.; Vincent, C.; Coen, E. An Epigenetic Mutation Responsible for Natural Variation in Floral Symmetry. *Nature* **1999**, *401* (6749), 157–161. https://doi.org/10.1038/43657.
(45) McCue, A. D.; Nuthikattu, S.; Reeder, S. H.; Slotkin, R. K. Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. *PLoS Genet* **2012**, *8* (2), e1002474. https://doi.org/10.1371/journal.pgen.1002474.
(46) Borges, F.; Martienssen, R. A. The Expanding World of Small RNAs in Plants. *Nat Rev Mol Cell Biol* **2015**, *16* (12), 727–741. https://doi.org/10.1038/nrm4085.
(47) Li, L.; Liu, Y. Diverse Small Non-Coding RNAs in RNA Interference Pathways. In *Therapeutic Oligonucleotides*; Goodchild, J., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2011; Vol. 764, pp 169–182. https://doi.org/10.1007/978-1-61779-188-8_11.
(48) Bartel, D. P. MicroRNAs. *Cell* **2004**, *116* (2), 281–297. https://doi.org/10.1016/S0092-8674(04)00045-5.
(49) Ding, D.; Zhang, L.; Wang, H.; Liu, Z.; Zhang, Z.; Zheng, Y. Differential Expression of MiRNAs in Response to Salt Stress in Maize Roots. *Annals of Botany* **2009**, *103* (1), 29–38. https://doi.org/10.1093/aob/mcn205.
(50) Hammond, S. M.; Bernstein, E.; Beach, D.; Hannon, G. J. An RNA-Directed Nuclease Mediates Post-Transcriptional Gene Silencing in Drosophila Cells. *Nature* **2000**, *404* (6775), 293–296. https://doi.org/10.1038/35005107.
(51) Vaucheret, H.; Vazquez, F.; Crété, P.; Bartel, D. P. The Action of *ARGONAUTE1* in the MiRNA Pathway and Its Regulation by the MiRNA Pathway Are Crucial for Plant Development. *Genes Dev.* **2004**, *18* (10), 1187–1197. https://doi.org/10.1101/gad.1201404.
(52) Li, S.; Liu, L.; Zhuang, X.; Yu, Y.; Liu, X.; Cui, X.; Ji, L.; Pan, Z.; Cao, X.; Mo, B.; Zhang, F.; Raikhel, N.; Jiang, L.; Chen, X. MicroRNAs Inhibit the Translation of Target MRNAs on the Endoplasmic Reticulum in Arabidopsis. *Cell* **2013**, *153* (3), 562–574. https://doi.org/10.1016/j.cell.2013.04.005.
(53) Chen, X. Small RNAs in Development – Insights from Plants. *Current Opinion in Genetics & Development* **2012**, *22* (4), 361–367. https://doi.org/10.1016/j.gde.2012.04.004.
(54) Elbashir, S. M. RNA Interference Is Mediated by 21- and 22-Nucleotide RNAs. *Genes & Development* **2001**, *15* (2), 188–200. https://doi.org/10.1101/gad.862301.

(55) Allen, E.; Xie, Z.; Gustafson, A. M.; Carrington, J. C. MicroRNA-Directed Phasing during Trans-Acting SiRNA Biogenesis in Plants. *Cell* **2005**, *121* (2), 207–221. https://doi.org/10.1016/j.cell.2005.04.004.
(56) Li, S.; Xu, Z.; Sheng, J. TRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. *Genes* **2018**, *9* (5), 246. https://doi.org/10.3390/genes9050246.
(57) Chen, Z.; Sun, Y.; Yang, X.; Wu, Z.; Guo, K.; Niu, X.; Wang, Q.; Ruan, J.; Bu, W.; Gao, S. Two Featured Series of RRNA-Derived RNA Fragments (RRFs) Constitute a Novel Class of Small RNAs. *PLoS ONE* **2017**, *12* (4), e0176458. https://doi.org/10.1371/journal.pone.0176458.
(58) Borges, F.; Martienssen, R. A. The Expanding World of Small RNAs in Plants. *Nat Rev Mol Cell Biol* **2015**, *16* (12), 727–741. https://doi.org/10.1038/nrm4085.
(59) Handzlik, J. E.; Tastsoglou, S.; Vlachos, I. S.; Hatzigeorgiou, A. G. Manatee: Detection and Quantification of Small Non-Coding RNAs from next-Generation Sequencing Data. *Sci Rep* **2020**, *10* (1), 705. https://doi.org/10.1038/s41598-020-57495-9.
(60) Axtell, M. J. ShortStack: Comprehensive Annotation and Quantification of Small RNA Genes. *RNA* **2013**, *19* (6), 740–751. https://doi.org/10.1261/rna.035279.112.
(61) Johnson, N. R.; Yeoh, J. M.; Coruh, C.; Axtell, M. J. Improved Placement of Multi-Mapping Small RNAs. *G3* **2016**, *6* (7), 2103–2111. https://doi.org/10.1534/g3.116.030452.
(62) Ito, H.; Gaubert, H.; Bucher, E.; Mirouze, M.; Vaillant, I.; Paszkowski, J. An SiRNA Pathway Prevents Transgenerational Retrotransposition in Plants Subjected to Stress. *Nature* **2011**, *472* (7341), 115–119. https://doi.org/10.1038/nature09861.
(63) McCue, A. D.; Nuthikattu, S.; Reeder, S. H.; Slotkin, R. K. Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. *PLoS Genet* **2012**, *8* (2), e1002474. https://doi.org/10.1371/journal.pgen.1002474.
(64) Pan, L.; Wang, Z.; Cai, J.; Gao, H.; Zhao, H.; Dong, L. High-Throughput Sequencing Reveals Differential Regulation of MiRNAs in Fenoxaprop-P-Ethyl-Resistant Beckmannia Syzigachne. *Sci Rep* **2016**, *6* (1), 28725. https://doi.org/10.1038/srep28725.
(65) Johnson, R.; Puthur, J. T. Seed Priming as a Cost Effective Technique for Developing Plants with Cross Tolerance to Salinity Stress. *Plant Physiology and Biochemistry* **2021**, *162*, 247–257. https://doi.org/10.1016/j.plaphy.2021.02.034.
(66) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
(67) Love, M. I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol* **2014**, *15* (12), 550. https://doi.org/10.1186/s13059-014-0550-8.
(68) Akalin, A.; Kormaksson, M.; Li, S.; Garrett-Bakelman, F. E.; Figueroa, M. E.; Melnick, A.; Mason, C. E. MethylKit: A Comprehensive R Package for the Analysis of Genome-Wide DNA Methylation Profiles. *Genome Biol* **2012**, *13* (10), R87. https://doi.org/10.1186/gb-2012-13-10-r87.
(69) Catoni, M.; Tsang, J. M.; Greco, A. P.; Zabet, N. R. DMRcaller: A Versatile R/Bioconductor Package for Detection and Visualization of Differentially Methylated Regions in CpG and Non-CpG Contexts. *Nucleic Acids Research* **2018**. https://doi.org/10.1093/nar/gky602.

(70) Flynn, J. M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A. G.; Feschotte, C.; Smit, A. F. RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families. *Proc Natl Acad Sci USA* **2020**, *117* (17), 9451–9457. https://doi.org/10.1073/pnas.1921046117.
(71) Szcześniak, M. W.; Rosikiewicz, W.; Makałowska, I. CANTATAdb: A Collection of Plant Long Non-Coding RNAs. *Plant Cell Physiol* **2016**, *57* (1), e8–e8. https://doi.org/10.1093/pcp/pcv201.
(72) Nuthikattu, S.; McCue, A. D.; Panda, K.; Fultz, D.; DeFraia, C.; Thomas, E. N.; Slotkin, R. K. The Initiation of Epigenetic Silencing of Active Transposable Elements Is Triggered by RDR6 and 21-22 Nucleotide Small Interfering RNAs. *Plant Physiol.* **2013**, *162* (1), 116–131. https://doi.org/10.1104/pp.113.216481.
(73) Syakur, M. A.; Khotimah, B. K.; Rochman, E. M. S.; Satoto, B. D. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conf. Ser.: Mater. Sci. Eng.* **2018**, *336*, 012017. https://doi.org/10.1088/1757-899X/336/1/012017.
(74) AlindGupta. *Elbow Method for optimal value of k in KMeans*. Geeks for Geeks. https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/.
(75) Munoz-Lopez, M.; Garcia-Perez, J. DNA Transposons: Nature and Applications in Genomics. *CG* **2010**, *11* (2), 115–128. https://doi.org/10.2174/138920210790886871.
(76) Naish, M.; Alonge, M.; Wlodzimierz, P.; Tock, A. J.; Abramson, B. W.; Schmücker, A.; Mandáková, T.; Jamge, B.; Lambing, C.; Kuo, P.; Yelina, N.; Hartwick, N.; Colt, K.; Smith, L. M.; Ton, J.; Kakutani, T.; Martienssen, R. A.; Schneeberger, K.; Lysak, M. A.; Berger, F.; Bousios, A.; Michael, T. P.; Schatz, M. C.; Henderson, I. R. The Genetic and Epigenetic Landscape of the *Arabidopsis* Centromeres. *Science* **2021**, *374* (6569), eabi7489. https://doi.org/10.1126/science.abi7489.
(77) Markus, C.; Pecinka, A.; Karan, R.; Barney, J. N.; Merotto, A. Epigenetic Regulation - Contribution to Herbicide Resistance in Weeds?: Epigenetic Regulation of Herbicide Resistance. *Pest. Manag. Sci* **2018**, *74* (2), 275–281. https://doi.org/10.1002/ps.4727.
(78) Floris, M.; Mahgoub, H.; Lanet, E.; Robaglia, C.; Menand, B. Post-Transcriptional Regulation of Gene Expression in Plants during Abiotic Stress. *IJMS* **2009**, *10* (7), 3168–3185. https://doi.org/10.3390/ijms10073168.
(79) Tiwari, B.; Habermann, K.; Arif, M. A.; Top, O.; Frank, W. Identification of Small RNAs During High Light Acclimation in Arabidopsis Thaliana. *Front. Plant Sci.* **2021**, *12*, 656657. https://doi.org/10.3389/fpls.2021.656657.
(80) Singh, A.; Gautam, V.; Singh, S.; Sarkar Das, S.; Verma, S.; Mishra, V.; Mukherjee, S.; Sarkar, A. K. Plant Small RNAs: Advancement in the Understanding of Biogenesis and Role in Plant Development. *Planta* **2018**, *248* (3), 545–558. https://doi.org/10.1007/s00425-018-2927-5.
(81) Zhai, R.; Ye, S.; Zhu, G.; Lu, Y.; Ye, J.; Yu, F.; Chu, Q.; Zhang, X. Identification and Integrated Analysis of Glyphosate Stress-Responsive MicroRNAs, LncRNAs, and MRNAs in Rice Using Genome-Wide High-Throughput Sequencing. *BMC Genomics* **2020**, *21* (1), 238. https://doi.org/10.1186/s12864-020-6637-6.
(82) Pegler, J.; Oultram, J.; Grof, C.; Eamens, A. Profiling the Abiotic Stress Responsive MicroRNA Landscape of Arabidopsis Thaliana. *Plants* **2019**, *8* (3), 58. https://doi.org/10.3390/plants8030058.

(83) Carrier, M.-C.; Ng Kwan Lim, E.; Jeannotte, G.; Massé, E. Trans-Acting Effectors Versus RNA Cis-Elements: A Tightly Knit Regulatory Mesh. *Front. Microbiol.* **2020**, *11*, 609237. https://doi.org/10.3389/fmicb.2020.609237.
(84) Niederhuth, C. E.; Schmitz, R. J. Putting DNA Methylation in Context: From Genomes to Gene Expression in Plants. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2017**, *1860* (1), 149–156. https://doi.org/10.1016/j.bbagrm.2016.08.009.
(85) Calarco, J. P.; Borges, F.; Donoghue, M. T. A.; Van Ex, F.; Jullien, P. E.; Lopes, T.; Gardner, R.; Berger, F.; Feijó, J. A.; Becker, J. D.; Martienssen, R. A. Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. *Cell* **2012**, *151* (1), 194–205. https://doi.org/10.1016/j.cell.2012.09.001.
(86) Kanissery, R.; Gairhe, B.; Kadyampakeni, D.; Batuman, O.; Alferez, F. Glyphosate: Its Environmental Persistence and Impact on Crop Health and Nutrition. *Plants* **2019**, *8* (11), 499. https://doi.org/10.3390/plants8110499.
(87) Bhowmik, P. C. Introduction to the Symposium: History of Sulfonylurea Herbicide Use in Turfgrass Environments. *Weed technol.* **2012**, *26* (3), 575–578. https://doi.org/10.1614/WT-D-11-00165.1.
(88) Pavet, V.; Quintero, C.; Cecchini, N. M.; Rosa, A. L.; Alvarez, M. E. *Arabidopsis* Displays Centromeric DNA Hypomethylation and Cytological Alterations of Heterochromatin Upon Attack by *Pseudomonas Syringae*. *MPMI* **2006**, *19* (6), 577–587. https://doi.org/10.1094/MPMI-19-0577.
(89) Gebremichael, D. E.; Haile, Z. M.; Negrini, F.; Sabbadini, S.; Capriotti, L.; Mezzetti, B.; Baraldi, E. RNA Interference Strategies for Future Management of Plant Pathogenic Fungi: Prospects and Challenges. *Plants* **2021**, *10* (4), 650. https://doi.org/10.3390/plants10040650.
(90) O'Malley, R. C.; Huang, S. C.; Song, L.; Lewsey, M. G.; Bartlett, A.; Nery, J. R.; Galli, M.; Gallavotti, A.; Ecker, J. R. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **2016**, *165* (5), 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038.
(91) Liu, S.; Bao, Y.; Deng, H.; Liu, G.; Han, Y.; Wu, Y.; Zhang, T.; Chen, C. The Methylation Inhibitor 5-Aza-2′-Deoxycytidine Induces Genome-Wide Hypomethylation in Rice. *Rice* **2022**, *15* (1), 35. https://doi.org/10.1186/s12284-022-00580-6.

Chapter 4: Multi-'omics Analysis of Arabidopsis Stress Response Analysis Combining sRNA, Methylation, and RNA Sequence Data

Abstract:

Epigenetic processes can involve many proteins, methylation, and sRNA to affect gene expression. We previously subjected Arabidopsis plants to stressors associated with weed management and then extracted RNA, sRNA, and DNA for methylation sequencing. While we had some interesting individual findings, this chapter focuses on merging these data types and analyzing them together. This could help us understand how sRNA and the various contexts of methylation coordinate in processes, such as RNA directed DNA methylation, to influence gene expression of stressed plants. This began with building and analyzing a single matrix based on our DNA methylation, mRNA, and sRNA data. While there is new software for handling multi-omics analyses, there was no package for analyzing our three data types as a group, so we had to assemble it ourselves and find way to analyze it. First, genomic regions with significant differential activity among stresses were identified for further study. Then data mining algorithms were used to check for overall patterns of epigenetic activity that may influence gene expression. The classification models were not able to consistently predict differentially expressed genes on a global scale. Using a separate scoring system, some individual genes were identified with differential methylation, sRNA expression, and mRNA expression. The noise resulting from the many means by which gene expression can be regulated made identifying trends ineffective. Focusing on local changes was more productive for locating genomic regions for further study of epigenetic regulation.

Introduction

As biotechnology has advanced, researchers have been able to obtain much more information about various aspects of molecular biology. Termed 'Omics studies, there are many branches that focus on particular macromolecules.[1] Genomics, for example, is the study of genomes, which has broad applications to biology. Researchers can compare changes to a previously sequenced genome to identify mutations, and genomes of different species can be compared to study evolution and taxonomic lineages.[2] People may also study the transcriptome, the messenger RNA (mRNA) of an organism, to identify genes from a genome as well as how genes are expressed under different circumstances. Identifying the methylated cytosine bases of DNA is studying the methylome.[3] The proteome and metabolome involve the study of proteins and other metabolites, respectively.[1] An individual 'Omics study may involve large amounts of information itself, but there are now multi-omics studies that attempt to combine two or more of these large datasets and interpret results spanning them.

For many years, the cost of sequencing was the limiting factor for nucleic acid 'omics studies. The development of high throughput sequencing platforms in 2004 and third generation sequencing technology developed in 2011 greatly reduced this cost.[4,5] Sequencing a human genome cost $3413.81 per Mb in 2001 while it cost $0.008 per Mb in May 2020.[6] As a result, multi-omics studies have become much more accessible for a variety of applications, such as connecting metabolite levels to gene expression or using histone marks to estimate gene expression.[12,18] Now, the limiting factor is the analyses that follow. Making sense of these sequences requires new software, computing power, and trained personnel.[7] Analyzing the data correctly can yield insight to overall biological trends as well as sites of interest for further study. Mistakes at this stage can waste time and laboratory resources in attempts to confirm features

that may be nonexistent. While there is established software available to assist with processing some of these 'omics datasets individually, there are some newer options for integrating them in a meaningful way.[8,9]

With so many accessible nucleic acid sequence datasets, researchers are now trying to mine the data to identify trends and draw biological conclusions.[10] There are other software packages that are designed for testing some of these features, but to our knowledge there are none that can simultaneously analyze all of them. For example, the web-based program, MethGET, can correlate gene expression and methylation, but it does not work with ncRNA or transposons.[11] ShallowChrome links histone modifications with gene expression in humans.[12] Transcriptome, metabolome, and histone modifications are combined in BRANEnet, using a machine learning algorithm known as a neural network.[13] There are a few other available pipelines that have been developed to integrate various combinations of genomes, histone modifications, transcriptomes, proteomes, miRNA, and phenotypes.[14–18] Another study in maize used a random forest classifier to link methylation and intron positions of genes to help separate pseudogenes from genes with a protein product.[19] Pipelines and software tools are very useful for beginners to process their data. Developing a reliable, accessible suite of tools to this end would help researchers analyze their data from multi-omics experiments when they would not be able to write their own scripts. This also helps with experiment reproducibility by streamlining the analysis and reducing the potential for human errors.

Unfortunately, there was no currently available tool for us to integrate our experimental results. Our data involved methylation, RNA, and sRNA sequence data as well as information from the gene annotation. It was from an experiment that used weed management stressors, glyphosate, trifloxysulfuron, shade, and clipping. While we analyzed the results individually in chapters 2 and 3 and made some 2-way comparisons, we also wanted to understand how these processes worked as a whole to facilitate stress responses. An example that integrates all of these data types can be found in a complex process known as RNA-directed DNA methylation (RdDM).[20] Guide sRNAs synthesized from lncRNA direct a large protein complex to methylate DNA in all contexts primarily to repress transposons. If methylation impacts gene expression, then it also bears relevance to the RNA seq dataset. To identify this process amid the myriad means of regulating gene expression, we could combine all this data. Other undiscovered connections and pathways are possible as well. Perhaps the surrounding features of genes create contexts in which sRNA and methylation could regulate gene expression differently. Knowing which mechanisms are used to regulate gene expression after the stressors are removed could help develop strategies for weed management.

In our case, we are interested in how these mechanisms participate in stress responses. We want to know what role, if any, these pathways have in mediating long term changes to gene regulation that could result in an altered weed phenotype. This could involve broad, overarching patterns of regulation or individual, targeted sites of regulation. Any connections found would lead to more experiments to better understand epigenetic regulation and stress response. To combine the datasets, we had to write some unique scripts. First, we converted the information of all data types into a numerical representation centered around genes to build a matrix of genes and their epigenetic features. Then, we had to test the data matrix for patterns of epigenetic regulation and identify sites of interest. We decided to use established data mining algorithms to

see if we could build a reliable classifier for identifying differentially regulated genes based on the genomic context and epigenetic information. Converting our data to numerical representations allows us to work with a variety of classifying algorithms.

These algorithms can be used for identifying patterns in our nucleic acid sequence data. This makes them useful for our objective of identifying epigenetic patterns associated with abiotic stress. This project combines methylation, sRNA, and RNA sequence data. One group has already used a random forest algorithm in a fairly successful attempt to predict functional genes out of a maize predicted gene set based on methylation levels in introns and exons across the gene.[19] However, it was not very effective for predicting expression levels based on their features. Combining the information in all of these can create many noisy features, making overfitting a major concern. Overfitting occurs when a model becomes too attuned to individual members of the training set and will miss the overall pattern to best classify data. For example, the methylation status of each cytosine context (CG, CHG, or CHH) within five equally spaced sections of a gene would generate fifteen features. Adding the distances to other overlapping or nearby features, in this case DEGs, transposons, sRNA, DMRs, lncRNA, and promoters, generates even more features. Honing in on the most important features will not only improve the predictive model; it will also help determine which epigenetic pathways are relevant for the plant's response to its treatment.

An important function of data mining that can be helpful for identifying patterns in the data is classification.[21] Classification algorithms are useful for determining how new data might fit into the previous data structure as well as developing predictive models. They are used to determine group placement of an object based on its attributes. k-nearest-neighbors (kNN) is a supervised model that records the class of a training set as well as all the training set's quantitative attributes.[22] New points are classified based on the user specified k number of neighbors.[22] The nearest neighbors are those with the shortest distance, whichever metric is used, from the new point (Fig. 1a).[22] These models thus weight all features equally and can be sensitive to noisy features that do not correlate well with the true class of the point. Decision trees are also useful classification algorithms. They split data into groups with the objective of sorting them into purer branches that contain more objects from a single class (Fig. 1b).[23] This method can accept qualitative variables, and it can detect non-linear clusters in the data.[21] It can be sensitive to patterns within the data that distance measures can struggle to detect, but the rigid splitting methods do not always form the best clusters, sometimes over- or underfitting with small datasets.[24] Aside from these two, there are many algorithms with varying strengths and weaknesses for building classification models. The best one depends on the data presented.

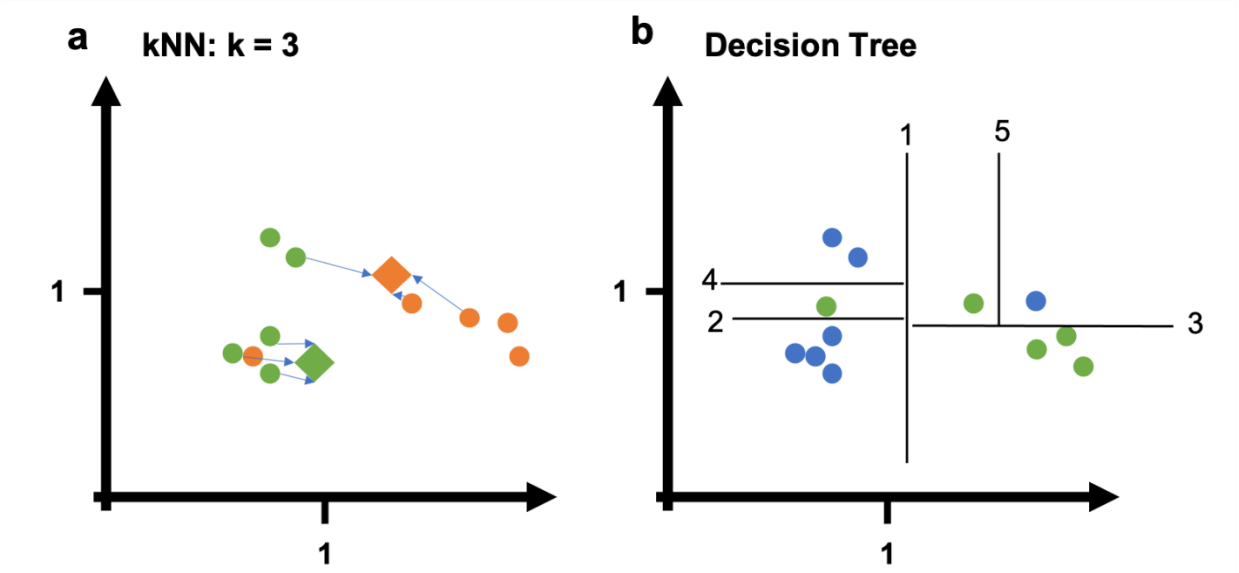


Figure 4: kNN is shown (a) classifying two new points (diamonds) according to the majority class of the nearest 3 points. Decision trees (b) use rules to split the training set into leaves, in this case regions. New rules are added to create classes of increasing purity. One such rule might be points with an x value greater than 1.1 are classified as green, while others are classified as blue. New points will be classified according to the majority class of the training set in their region.

For our dataset, we have bisulfite sequences, RNA sequences, sRNA sequences, and a GFF file filled with a variety of features. The data was obtained from an experiment with *Arabidopsis thaliana*. Young plants were exposed to one of four stresses involved in weed control, trifloxysulfuron herbicide, glyphosate herbicide, shade, and mechanical clipping. Additionally, an untreated control batch of plants was grown. In all treatments except shade, the plants were allowed to recover and flower, so we could test for long-term responses in the extracted RNA and DNA. Combining all the sequencing outputs in this experiment may give us new insight for data analysis that could not be achieved through examining the datasets individually. The workflow of processing the data into a large matrix is depicted in Figure 2. Using data mining algorithms on this matrix, we tested the influence of broad epigenetic patterns on gene expression. We assembled a simulated dataset to represent RdDM to test the ability of these algorithms to identify such trends. Also, we used a scoring system to identify hotspots with significant gene expression changes associated with overlapping epigenetic processes such as RNAi or RdDM. This would help us narrow down sites for further study. We wanted to know if there was a broad, overarching pattern of methylation and sRNA dictating gene regulation. We also wanted to evaluate local regions of significant epigenetic activity to determine whether there could be a more targeted epigenetic control of gene expression in response to our abiotic stresses.

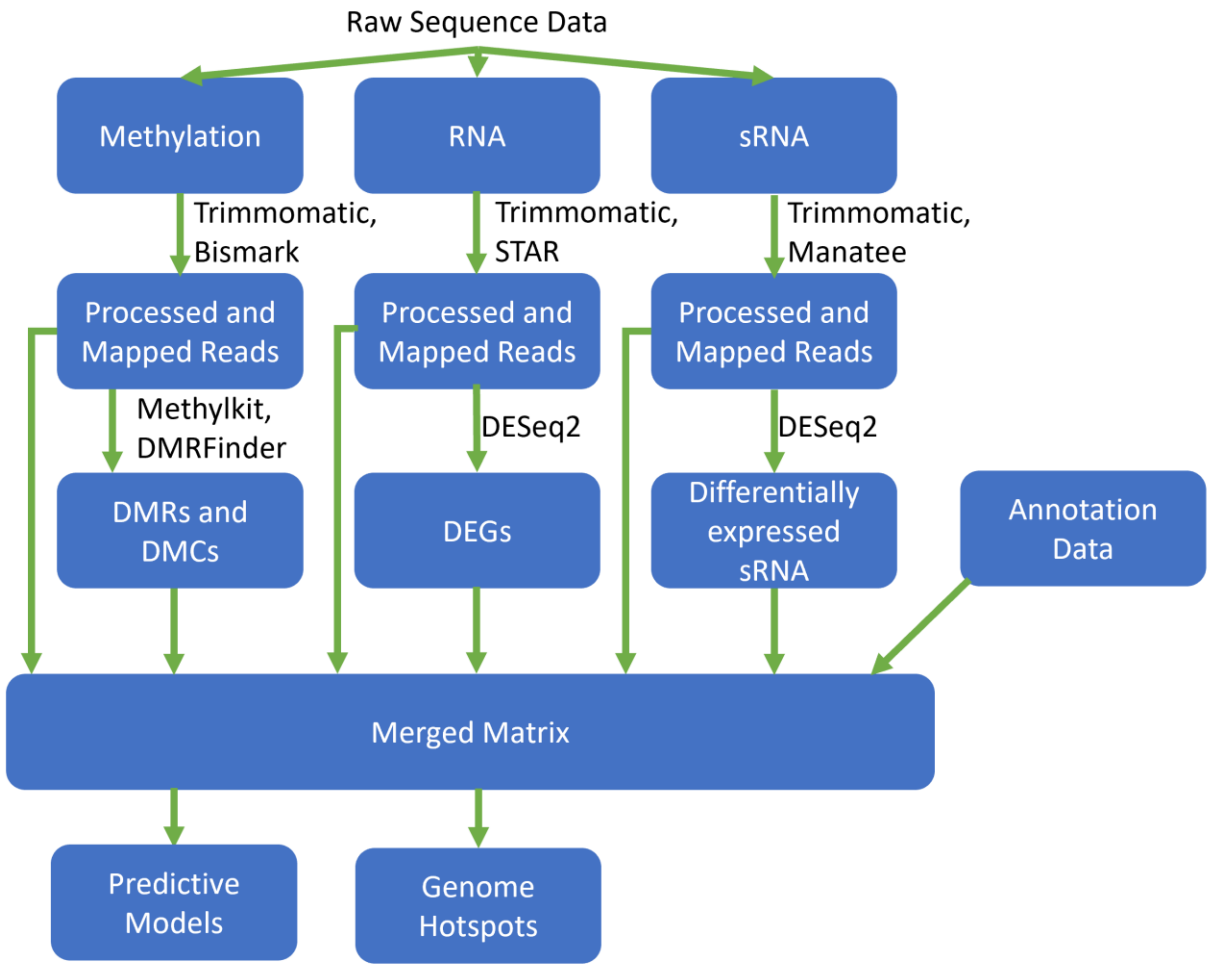


Figure 5 Data Workflow: Our scripts processed raw sequence data into three outputs that summarize some of the findings. Established software was used for the initial processing, and we merged the outputs into a larger matrix and created our final outputs with data mining algorithms and a scoring system.

Methods:

Creating the Merged Matrix:

Initially, this pipeline requires whole-genome bisulfite sequence fastq files, RNA sequence fastq files, sRNA sequence fastq files, a fasta file of the complete genome, and a gff or gtf file containing gene and lncRNA locations within the fasta file. The fasta file contains the complete genomic sequence of the organism separated by chromosomes. It serves as a reference for many downstream analyses. The gff file is a data matrix of the chromosomal location and attributes of features in the genome, such as genes and transposons. The bisulfite, RNA, and sRNA sequences started in fastq format directly from the sequencing facility and were then trimmed. This processed sequence data was aligned to the genomic fasta file. This means that each read's genomic origin will be determined based on its sequence. The gff file is important for establishing the levels of RNA sequence reads in known genes. These known genes can be further checked for RNA expression levels and epigenetic activity. The mapped reads were

converted into a count matrix and analyzed with DESeq2. This pipeline uses the output of DESeq2 to represent the RNA sequence dataset.

The RNA fastq files were preprocessed and mapped using standard tools for RNA seq studies in Chapter 2 while the sRNA and methylation fastq files were similarly processed and aligned in Chapter 3.[9,25,26] These mapped reads were further analyzed with DMRCaller to identify differentially methylated regions (DMRs) and Methylkit for differentially methylated cytosines (DMCs).[27,28] The resulting DMCs, DMRs, and methylation values from the Bismark V0.20.0 cx report files are used to generate the methylation features for the larger matrix.

The sRNA coordinates were obtained by aligning the trimmed sRNA fastq files to the TAIR10 genome with Bowtie while allowing multimapping at up to 50 locations.[8,29] Regions of the genome that were at least 20 bp long with a sRNA level of at least 21, the number of replicates in all treatment groups, were added. The sRNA reads were also mapped to known genomic features with Manatee, software tailored to multimapping sRNA.[30] These mapped reads and Manatee-identified clusters were analyzed with DESeq2 similarly to the RNA and genes.[31] This output was used to identify differentially expressed sRNA. The differentially expressed sRNA and other sRNA levels are used to generate sRNA features for the larger matrix.

While the TAIR10 Arabidopsis gff file also lists some transposons, RepeatModeler was used to identify other transposons and repeat regions.[29,32] These lists were combined to create a larger list of transposons. The proximity of transposons is relevant to epigenetics because RNA directed DNA Methylation (RdDM) primarily works to silence transposons.[20] The lncRNA coordinates from the Cantata database were used to denote lncRNA genomic positions.[33] These other attributes provide some context for genes' neighboring features in the larger matrix.

Features of the merged matrix:

Several features were chosen to represent CG methylation: the average methylation levels of CG cytosines, the coverage of CG, and difference between the experimental CG methylation and the control. Those were also calculated for the region 500 base pairs upstream of the genes. The presence/absence of CG hyper and hypomethylated DMCs and DMRs were also added. Together, this made a total of 10 features for each gene pertaining to CG methylation. These values were also calculated for the CHG and CHH cytosine methylation contexts. The number of DMRs that overlap with the gene, the number of DMRs within 2000 bp of the gene, the proximity to DMRs, the number of DMCs that overlap with the gene, the number of DMCs within 200 bp of the gene, the proximity to DMCs. The maximum value of the proximity of DMRs and DMCs was set to 2000 bp. This added up to a total of 36 methylation features.

The average sRNA levels for both genes and the 500 bp upstream was added as a feature. This was calculated for both the control group and experimental group. The highest sRNA level in each gene was determined alongside the number of overlapping sRNA, the number of sRNA within 2000 bp of the gene, the presence of significantly differentially up- and downregulated sRNA, the proximity to the nearest sRNA, and the proximity to the nearest significantly differentially expressed sRNA. The maximum value for the proximity of sRNA features was 2000 bp. This added up to a total of 15 sRNA features per gene.

A few other features were also added to represent the surrounding properties of the gene. The presence/absence of nearby up and downregulated genes, the proximity to the nearest differentially expressed gene, and the proximity to the nearest gene were added as features to represent nearby gene activity. A feature was added to denote the proximity to the nearest lncRNA as well as the number of lncRNA within 2000 bp of the gene. Features were also added to denote the proximity to the nearest transposon in base pairs and number of transposons with 2000 bp. If no genes, transposons, and/or lncRNA were within the 2000 bp range, the proximity value was set to 5000 bp. These added 9 more features to describe the context of each gene.

Predictive Modeling:

The large matrix of features was used to develop predictive models to determine whether the epigenetic features could be used to predict differentially expressed genes in either direction. Genes were given an upregulated label if their adjusted p value from the DESeq2 analysis was less than .01 and the log2 fold change was greater than 2. Downregulated genes had the same p value range but a $\log_2$ fold change value less than -2.

We used a variety of available algorithms to build classification models with our data, including the previously discussed k nearest neighbors and decision tree. In kNN, points are classified based on the classification of nearby points in the training set. Decision trees form splits to increase the purity of each branch. We also used a few other algorithms. Random forest classification uses smaller decision trees that form splits randomly.[34] The majority vote of these classification schemes is used to determine the class of a new point in the test set. This method can identify unusual patterns that a decision tree might miss, but it is vulnerable to overfitting if the smaller trees are allowed to grow too large.[34] A fourth classifying algorithm we tried was a C support vector classifier. This algorithm can map data into a higher dimensional space to create non-linear boundaries for classification.[35]

Machine learning is another algorithm-based approach to problem solving that allows a computer to respond to data without user guidance. Unsupervised machine learning does not have a training dataset, so the computer adjusts its algorithm without any user input.[36] In supervised machine learning, the software is given a training data set with known values to adjust its algorithm before it is tested on data with unknown values.[36] A potential issue that stems from training the data as well as feature selection is overfitting, in which the model is so attuned to the training set that it fails to classify new data correctly.[37] Some common data analysis tasks completed by machine learning include classifying and clustering data based on features; developing predictive models; and reducing the dimensionality of a dataset.[38] There are several algorithms readily available for completing these tasks; the best one depends on the data provided.[38] In our case, we used a multi-layer perceptron, also known as a neural network, to attempt to build our fifth classifier.[39]

The following Scikit learn classification models were used with a training set of 80% of the data: decision tree, random forest, k nearest neighbors, support vector, and a multi-layer perceptron.[40] The decision trees had a maximum depth of 8 and formed splits based on the gini index. The class weight was set to "balanced," and the other parameters were left at the default settings. The random forest classifier was given a maximum depth of 3 and balanced class weights, but the other settings were left at their defaults. A k of 3 neighbors was used for the

kNN classifier on data standardized with SciKit's StandardScaler function. The other parameters were left at their defaults, including using the Minkowski distance and allowing the algorithm to automatically detect the method for computing the nearest neighbors. The remaining default parameters included an rbf kernel and a degree of three. The support vector machine classifier was run with the 'lbfgs' solver and default settings.

A smaller matrix was also used. In an attempt to reduce the noise of so many variables, this one only used the differences in methylation for each context in both the gene and the promoter, the base mean expression of the gene, the highest level of sRNA, the gene length, the difference in sRNA levels, and the difference in the level of sRNA in the promoter. The parameters were kept the same except for the maximum depth of the decision trees, which was set to 6. This matrix included only a few basic features of the gene and the differences in the epigenetic pathways we studied between the experimental groups and the control.

Simulated Dataset:

To test the algorithms' ability to identify patterns in this type of data, a simulated dataset was created from the methylation clusters in Chapter 3. These clusters reflected gene methylation patterns identified in another study: gene body CG methylation, high methylation in all contexts, high promoter methylation, and low methylation. The mean and standard deviations were used to generate a dataset to represent genes with the 4 groups of methylation types. This matrix had 46 features, 42 for methylation and 4 to represent the transposons and sRNA. Methylation in each context for the promoter, five segments of the gene, and the 500 nucleotides downstream was generated for a control and experimental group based on the inverse of the error function.[41] Upregulated genes were generated with an experimental group resembling gene body or no methylation and a control with promoter or high methylation levels. 2000 genes of each combination were simulated for a total of 8000 upregulated genes. 8000 downregulated genes were generated with promoter or high methylation in the experimental group and gene body or no methylation in the control. A set of 18000 genes that were neither up nor downregulated was also simulated. There were 3000 for each methylation type that had a consistent pattern between the experimental group and the control. There were also 3000 genes with gene body methylation in either the experimental group or the control and no methylation as the other methylation type. This added up to a total of 34,000 genes in the simulated dataset.

Some similar data with randomly assigned labels was added to the simulated uniform dataset to perturb the model. In the simulated dataset, genes could have any combination of gene methylation and other features. The class was assigned at random with weights matching the original dataset, which had a ratio of 8/34 upregulated, 8/34 downregulated, and 18/34 unchanged genes.

Identifying interesting genes:

When the models did not reveal broad trends of epigenetic activity, we started looking at a local level with the matrix of features used to develop the models. The matrix of genes with their epigenetic information was used to identify regions of high differential activity. It was opened in a Jupyter Notebook, and a scoring system was developed to help determine which genes had the most interesting features. Columns were filtered by their values and used to award points based on the presence of nearby features, and differential activity, such as the presence of

differentially methylated cytosines or differentially expressed sRNA. Scoring these features would help us identify regions with the most probable epigenetic activity. Significantly differentially expressed genes, with an adjusted p value of less than .001 and a $\log_2$ fold change greater than 2 or less than -2, were given 3000 initial points. 500 points were added to genes with a nearby DMR, 200 points for non-CG DMCs, and 100 points for CG DMCs. The ratio of the difference in percent methylation to the average percent methylation in each context was added with a maximum value of 100 points for both the gene and the promoter. 300 points were given for genes with nearby differentially expressed small RNA and 150 were added for sRNA with a difference in expression level greater than the larger of the control and experimental standard deviation. 1000 points were added to the score of genes with points for both significant differential sRNA and methylation features. Sorting the genes by their point values helped identify which genes were interesting for potential follow-up studies of epigenetics.

Results:

Creating the matrix that combined methylation, sRNA, gene expression, and other features was successful. The features were converted into a numerical format acceptable to some Scikit learn algorithms for building classifiers.[42] kNN, C support vector, and multi-layer perception algorithms were used. To try to minimize the problem of too many features in the original dataset, we attempted to pare them down to the most interesting, and we tried to weigh the DEGs in the dataset to balance it. Even the small matrix had 11 features, which is still high for some datasets, and while changing the weight of the DEGs balanced the dataset, it can contribute to overfitting. None of the data mining algorithms' models produced with our experimental data could reliably identify DEGs, regardless of treatment. Still, we found some local sites that could be interesting for further epigenetic studies.

Across the 20 tested models for each treatment, we were not able to reliably use epigenetic information to identify differentially expressed genes. This is evidenced by low recall scores (Table 1), which indicate that the model identified many false positives for differentially expressed genes. While the receiver operating characteristic (ROC) curves do not indicate a great model either, the problems are apparent with the precision-recall (PR) curves (Fig. 3-10).[43] Figure 11 depicts ideal ROC and PR curves. An ideal PR curve would be the inverse of these, with the line following $y = 1$ before dropping to 0 at $x=1$.[44] This is also indicative of the unbalanced nature of the dataset.[45] While balancing measures were taken when available, the unbalanced nature of the dataset still caused problems. This was apparent in the low F1 values for all of the models. A good F1 score would be closer to 1 and indicates that the model classifies fewer false positives and false negatives.[46] The model based on glyphosate-treated plants performed slightly better than the others, with a highest F1 score of .1998 in a decision tree model, and it was the most balanced of the sets because it had the most DEGs. In general, the random forest and decision tree models had higher F1 scores of .0884 and .0789, respectively, and a greater area under the ROC curve than the other models. The kNN and perceptron models performed the worst, with F1 scores averaging .0125 and .0325. There did not appear to be a broad, overarching pattern of epigenetic activity among our features that would allow these classifiers to identify differentially expressed genes.

With the low F1 scores of the models built with our actual data, we decided to create a simulated dataset to reflect a pattern we might expect to see in the event of broad gene regulation

through RNA directed DNA methylation.[20] This would help us test the algorithms' abilities to build a model based on the patterns we might expect. We expected hypermethylation of the gene and promoter region to be associated with repression, and hypomethylation with upregulation. Using the means and standard deviations of the clusters previously identified, we created methylation features to simulate the RdDM changes alongside sRNA levels and proximity to transposons. The simulated dataset was easier to classify (Table 2 and Figure 11). The features were different but focused entirely on how RNA-directed DNA methylation might shift the pattern of methylation to a repressive state and thus alter expression. An estimate of the number of transposons and sRNA was added, as well as sRNA levels and the distance to the nearest transposon. Methylation in a control and experimental context for regions in and around the gene was also estimated. While the classification models were nearly perfect with just the simulated data, adding some noise, in the form of random unbalanced data, disrupted the model. We tested the models built with the simulated data for their ability to classify actual data from our experiment. We chose the glyphosate-treated plant data because it was the most balanced, and the models performed the best with it. When the same features were obtained for the actual glyphosate-treated plant data, the models created using the simulated dataset performed poorly (Table 3). We were not able to detect a broad pattern of RdDM altering the expression of many genes from our full, epigenetic matrix.

Because we could not find patterns through the noisy, full dataset, we focused on individual areas of overlap where epigenetics could play a role in influencing gene expression. Some genes with interesting epigenetic features were identified through the scoring system on the actual data (Table 4). This system helped us isolate genes with features we were interested in, such as significant differential methylation, differential sRNA expression, and differential gene expression. In glyphosate-treated plants, there were 11 genes with altered methylation, sRNA, and expression when compared to the control. 4 were identified from trifloxysulfuron-treated Arabidopsis and 3 from shaded. None were identified from the mechanically clipped plants, which only had two DMCs. Additionally, the transposable element gene AT2G05240 was identified from the glyphosate treated plants. Thus, from our weighted scoring system we were able to identify 18 genes and 1 transposable element with correlated altered expression, methylation, and sRNA.

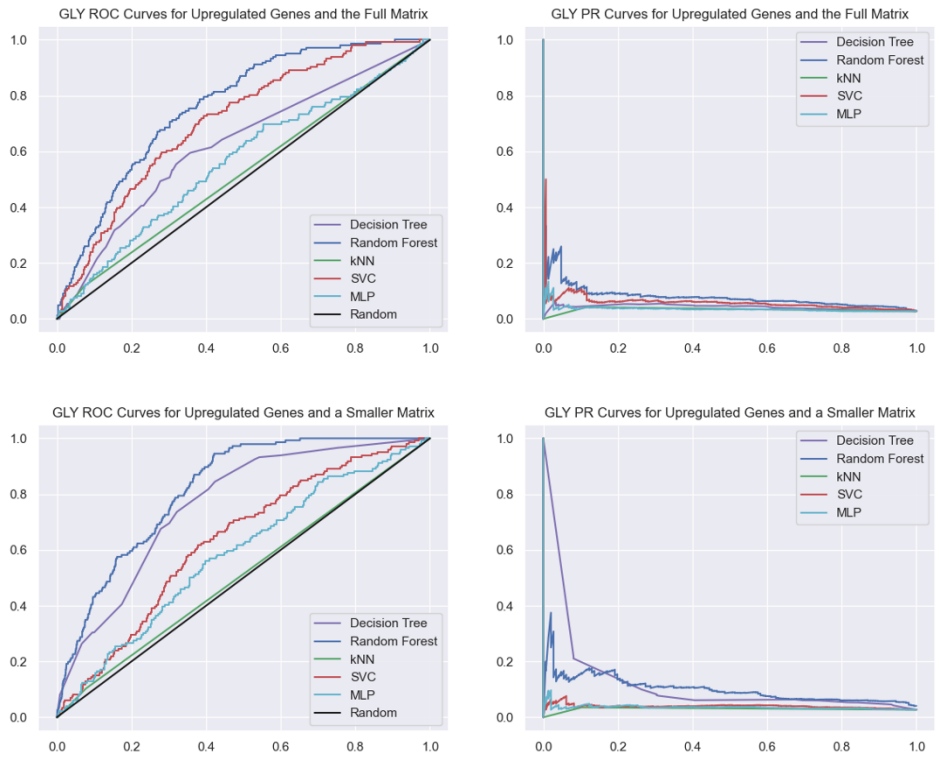


Figure 3 Glyphosate-Treated Upregulated Gene Matrices' Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify upregulated genes with the epigenetic responses in glyphosate-treated Arabidopsis.

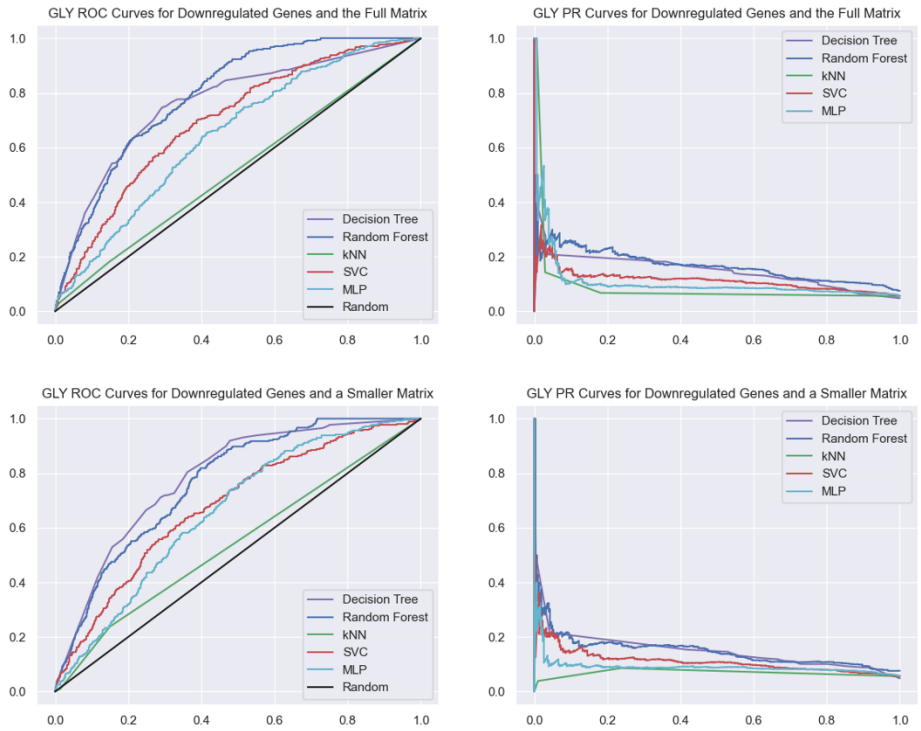


Figure 4 Glyphosate-Treated Upregulated Gene Matrices' Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify downregulated genes with the epigenetic responses in glyphosate-treated Arabidopsis.

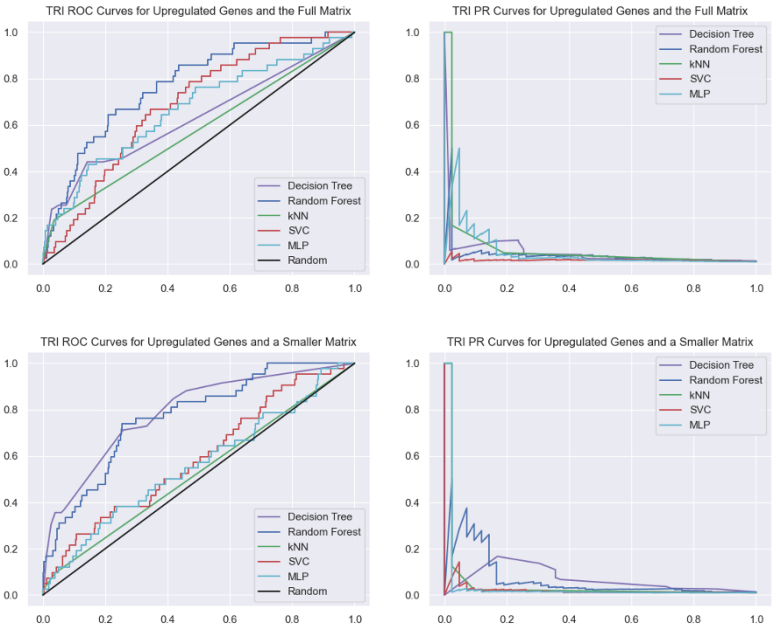


Figure 5 Trifloxysulfuron-Treated Upregulated Gene Matrices' Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify upregulated genes with the epigenetic responses in trifloxysulfuron-treated Arabidopsis.

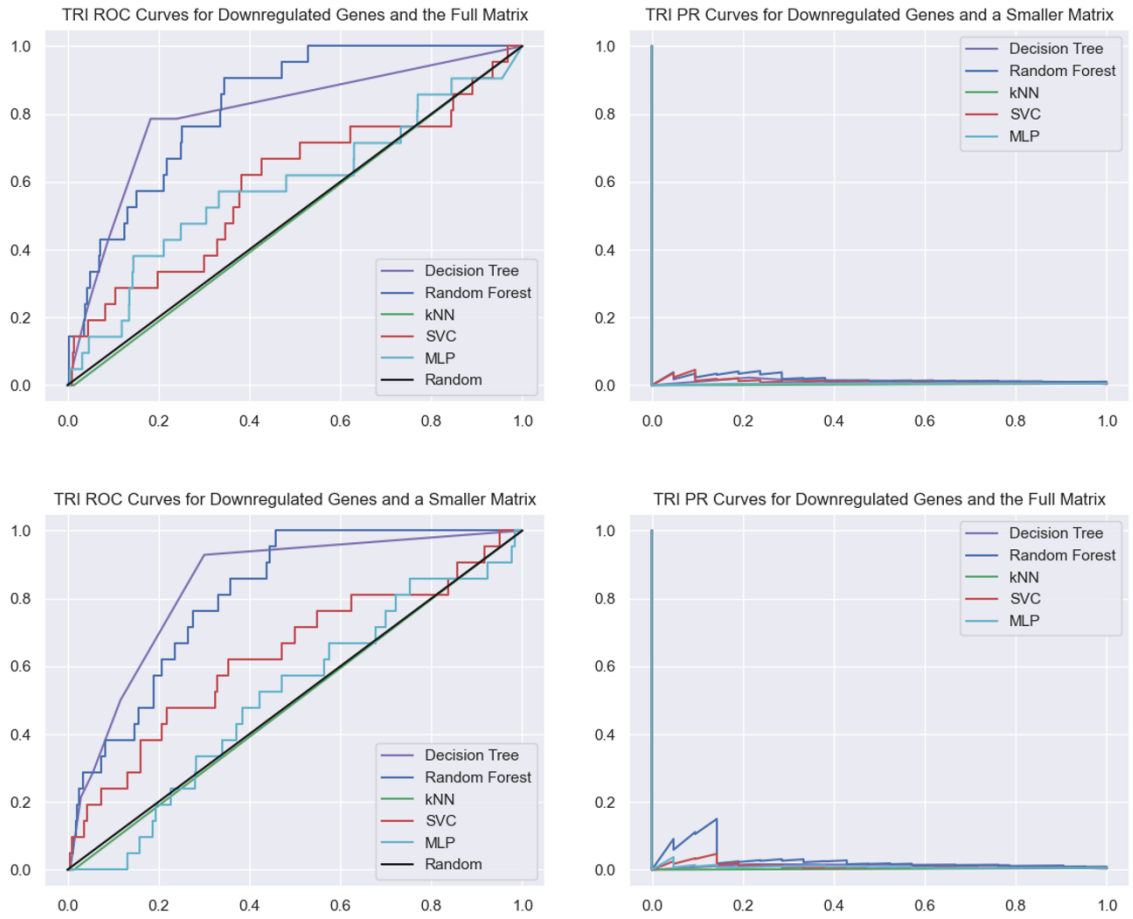


Figure 6 Trifloxysulfuron-Treated Downregulated Gene Matrices' Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify downregulated genes with the epigenetic responses in trifloxysulfuron-treated Arabidopsis.

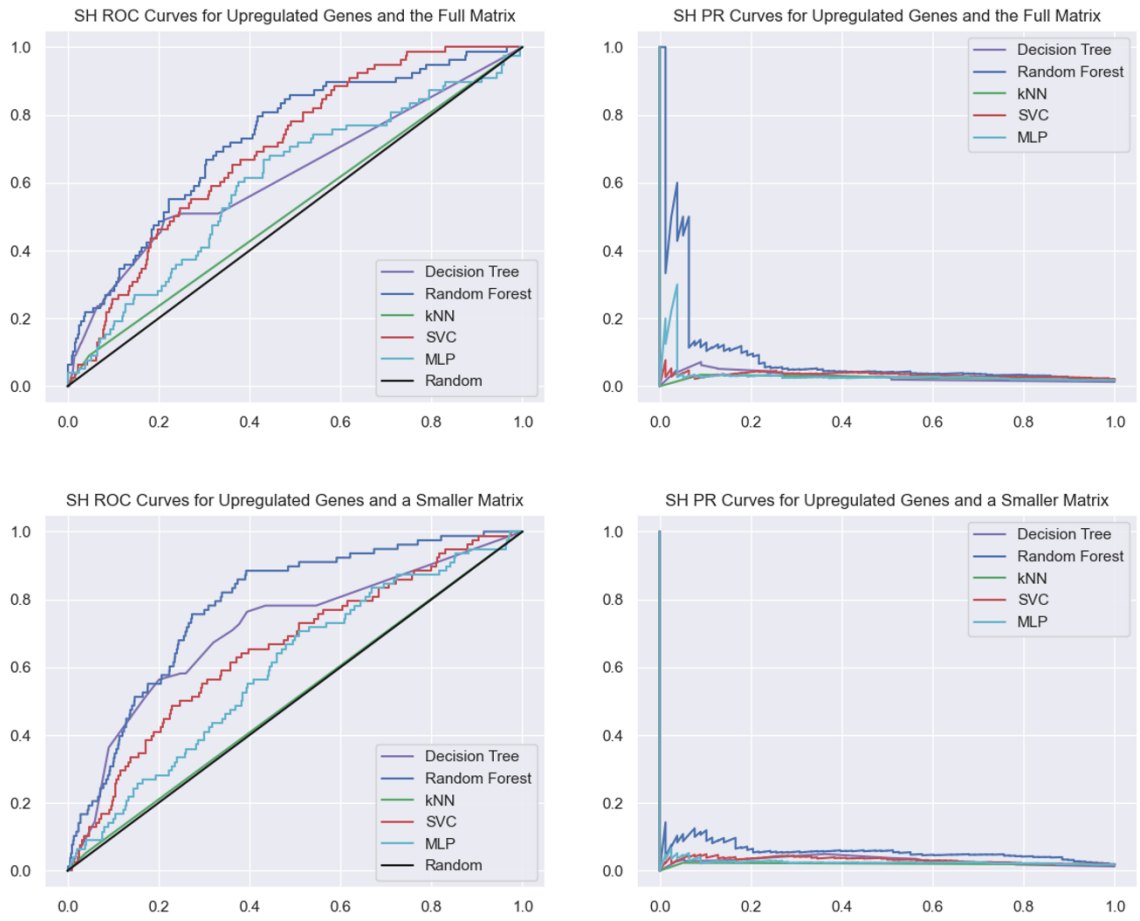


Figure 7 Shaded Upregulated Gene Matrices' Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify upregulated genes with the epigenetic responses in shaded Arabidopsis.

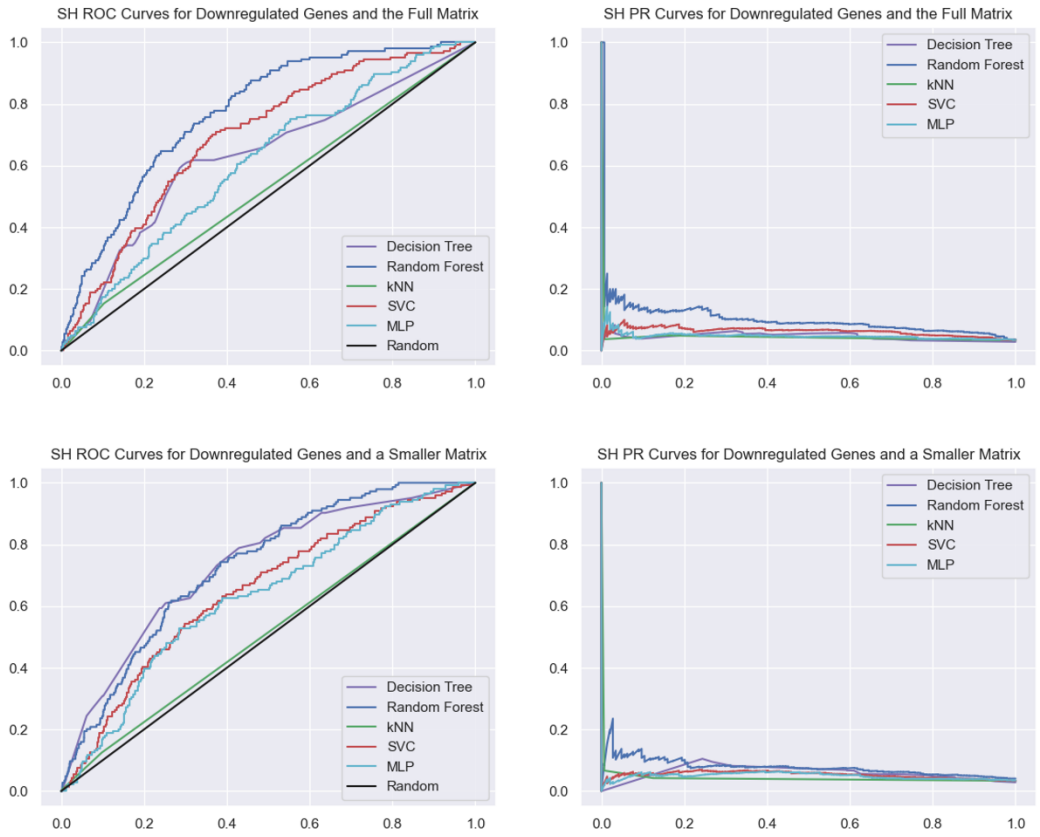


Figure 8 Shaded Downregulated Gene Matrices' Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify downregulated genes with the epigenetic responses in shaded Arabidopsis.

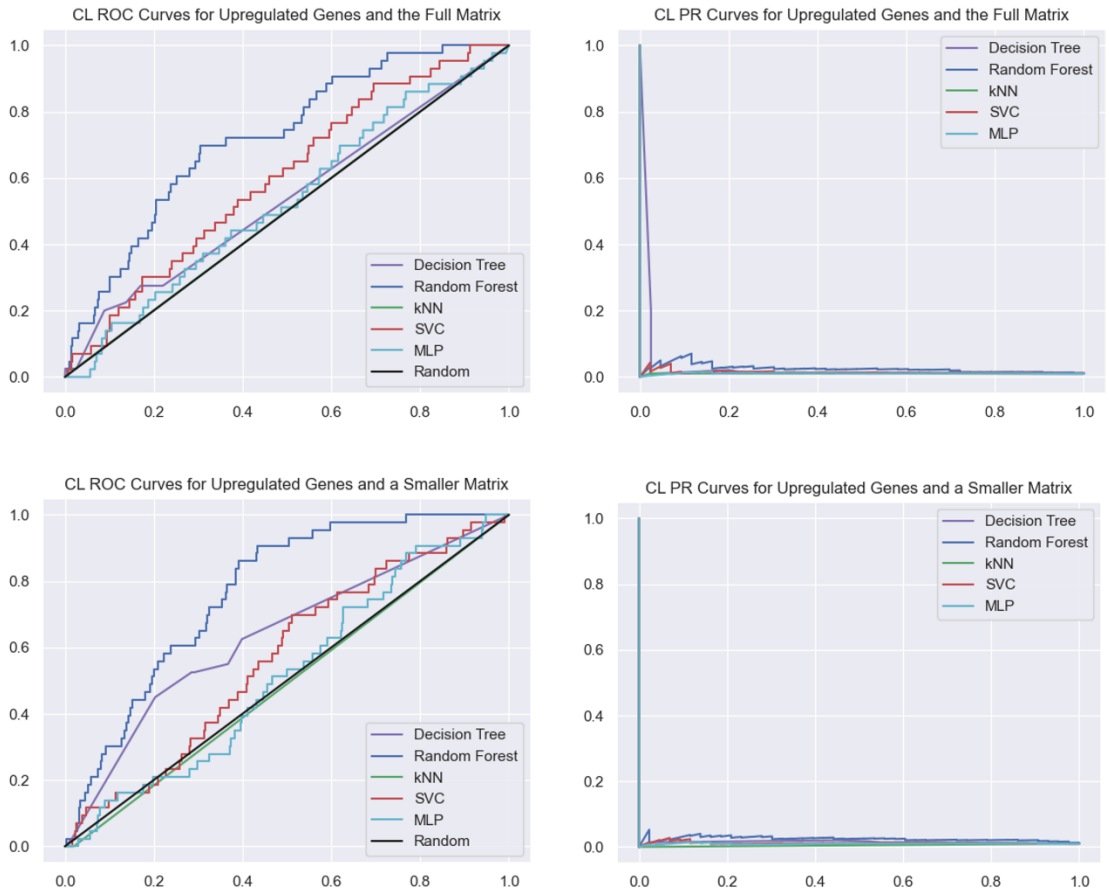


Figure 9 Clipped Upregulated Gene Matrices' Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify upregulated genes with the epigenetic responses in clipped Arabidopsis.

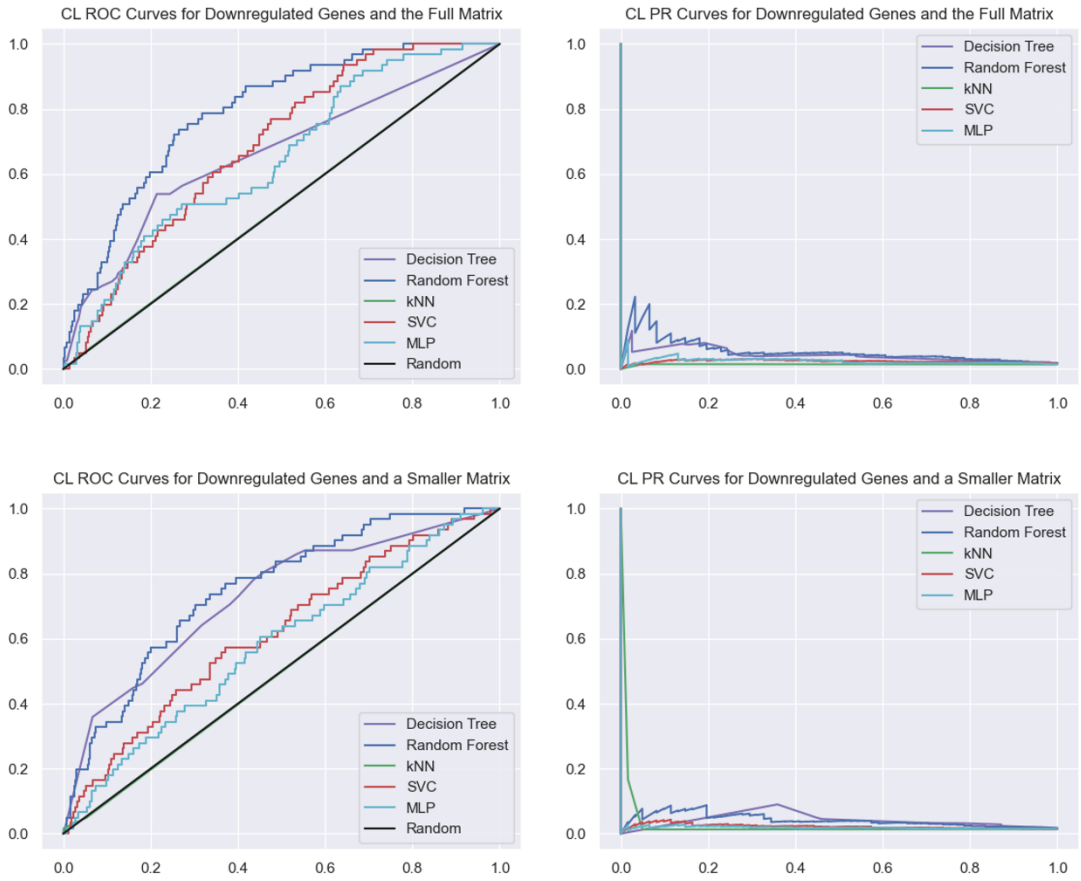


Figure 10 Clipped Downregulated Gene Matrices' Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify downregulated genes with the epigenetic responses in clipped Arabidopsis.

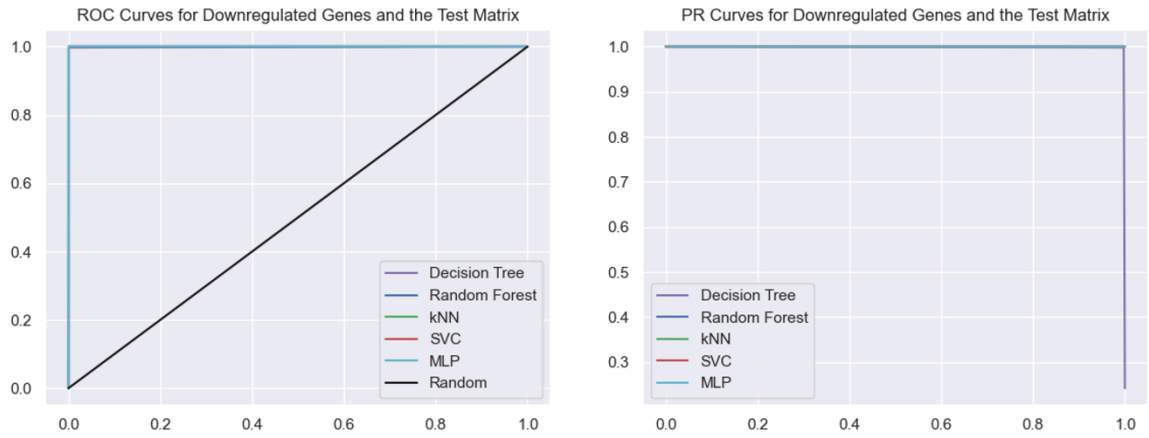


Figure 11 Test Matrix Evaluation: The receiver operating characteristic and precision recall curves were generated for a decision tree (purple), random forest (blue), k nearest neighbors (green), C support vector machine (red), and a multi-layer perceptron classifier (cyan) attempting to identify downregulated genes in the simulated dataset with 0% noisy data. This shows ideal ROC and PR curves.

Table 5 Model Performance Metrics: Some evaluation metrics for the models are listed.

Treatment	DEG Direction	Model	Matrix Size	Accuracy	Precision	Recall	F1 Score
GLY	Up	Tree	Big	0.7170	0.0475	0.4932	0.0867
GLY	Up	Forest	Big	0.7190	0.0610	0.6575	0.1117
GLY	Up	kNN	Big	0.9711	0.0000	0.0000	0.0000
GLY	Up	SVC	Big	0.6381	0.0481	0.6644	0.0898
GLY	Up	MLP	Big	0.9619	0.0933	0.0479	0.0633
GLY	Down	Tree	Big	0.7063	0.1139	0.7577	0.1980
GLY	Down	Forest	Big	0.6096	0.1128	0.8746	0.1998
GLY	Down	kNN	Big	0.9360	0.1429	0.0297	0.0492
GLY	Down	SVC	Big	0.5873	0.0914	0.7162	0.1621
GLY	Down	MLP	Big	0.9332	0.1667	0.0495	0.0763
GLY	Up	Tree	Small	0.6799	0.0602	0.7365	0.1113
GLY	Up	Forest	Small	0.6896	0.0615	0.7397	0.1135
GLY	Up	kNN	Small	0.9707	0.0000	0.0000	0.0000
GLY	Up	SVC	Small	0.5190	0.0385	0.7055	0.0730
GLY	Up	MLP	Small	0.9700	0.0952	0.0137	0.0240
GLY	Down	Tree	Small	0.7003	0.1073	0.7192	0.1867
GLY	Down	Forest	Small	0.5761	0.1022	0.8482	0.1824
GLY	Down	kNN	Small	0.9310	0.0385	0.0099	0.0157
GLY	Down	SVC	Small	0.4736	0.0796	0.7987	0.1447
GLY	Down	MLP	Small	0.9426	0.2353	0.0132	0.0250

TRI	Up	Tree	Big	0.8083	0.0311	0.4407	0.0581
TRI	Up	Forest	Big	0.7894	0.0278	0.6190	0.0532
TRI	Up	kNN	Big	0.9895	0.1667	0.0238	0.0417
TRI	Up	SVC	Big	0.6482	0.0179	0.6667	0.0349
TRI	Up	MLP	Big	0.9770	0.0845	0.1429	0.1062
TRI	Down	Tree	Big	0.8176	0.0136	0.7857	0.0267
TRI	Down	Forest	Big	0.8042	0.0150	0.6190	0.0293
TRI	Down	kNN	Big	0.9950	0.0000	0.0000	0.0000
TRI	Down	SVC	Big	0.7506	0.0064	0.3333	0.0126
TRI	Down	MLP	Big	0.9902	0.0000	0.0000	0.0000
TRI	Up	Tree	Small	0.7419	0.0362	0.7119	0.0689
TRI	Up	Forest	Small	0.7631	0.0265	0.6667	0.0510
TRI	Up	kNN	Small	0.9891	0.1250	0.0238	0.0400
TRI	Up	SVC	Small	0.6669	0.0110	0.3810	0.0214
TRI	Up	MLP	Small	0.9886	0.1667	0.0476	0.0741
TRI	Down	Tree	Small	0.6999	0.0098	0.9286	0.0193
TRI	Down	Forest	Small	0.7397	0.0130	0.7143	0.0255
TRI	Down	kNN	Small	0.9952	0.0000	0.0000	0.0000
TRI	Down	SVC	Small	0.6289	0.0079	0.6190	0.0157
TRI	Down	MLP	Small	0.9907	0.0000	0.0000	0.0000
SH	Up	Tree	Big	0.8040	0.0287	0.4364	0.0539
SH	Up	Forest	Big	0.7308	0.0361	0.5385	0.0677
SH	Up	kNN	Big	0.9809	0.0000	0.0000	0.0000
SH	Up	SVC	Big	0.6331	0.0319	0.6538	0.0608
SH	Up	MLP	Big	0.9690	0.0339	0.0256	0.0292
SH	Down	Tree	Big	0.7092	0.0574	0.5935	0.1047
SH	Down	Forest	Big	0.6750	0.0730	0.7431	0.1329
SH	Down	kNN	Big	0.9607	0.0370	0.0069	0.0117
SH	Down	SVC	Big	0.6380	0.0621	0.6944	0.1140
SH	Down	MLP	Big	0.9511	0.0875	0.0486	0.0625
SH	Up	Tree	Small	0.7378	0.0282	0.5818	0.0538
SH	Up	Forest	Small	0.7728	0.0501	0.6410	0.0929
SH	Up	kNN	Small	0.9811	0.0000	0.0000	0.0000
SH	Up	SVC	Small	0.5569	0.0270	0.6667	0.0518
SH	Up	MLP	Small	0.9795	0.0000	0.0000	0.0000
SH	Down	Tree	Small	0.5278	0.0471	0.8049	0.0889
SH	Down	Forest	Small	0.6440	0.0614	0.6736	0.1126
SH	Down	kNN	Small	0.9634	0.0667	0.0069	0.0126
SH	Down	SVC	Small	0.5353	0.0483	0.6875	0.0902
SH	Down	MLP	Small	0.9632	0.0625	0.0069	0.0125

CL	Up	Tree	Big	0.8244	0.0149	0.2750	0.0284
CL	Up	Forest	Big	0.8324	0.0279	0.4651	0.0526
CL	Up	kNN	Big	0.9895	0.0000	0.0000	0.0000
CL	Up	SVC	Big	0.7353	0.0133	0.3488	0.0257
CL	Up	MLP	Big	0.9853	0.0455	0.0233	0.0308
CL	Down	Tree	Big	0.7811	0.0444	0.5385	0.0820
CL	Down	Forest	Big	0.7711	0.0436	0.7213	0.0822
CL	Down	kNN	Big	0.9849	0.0000	0.0000	0.0000
CL	Down	SVC	Big	0.6750	0.0249	0.5738	0.0477
CL	Down	MLP	Big	0.9707	0.0149	0.0164	0.0156
CL	Up	Tree	Small	0.7069	0.0167	0.5250	0.0323
CL	Up	Forest	Small	0.7516	0.0224	0.5581	0.0430
CL	Up	kNN	Small	0.9895	0.0000	0.0000	0.0000
CL	Up	SVC	Small	0.5846	0.0123	0.5116	0.0241
CL	Up	MLP	Small	0.9828	0.0000	0.0000	0.0000
CL	Down	Tree	Small	0.6009	0.0326	0.7308	0.0624
CL	Down	Forest	Small	0.7437	0.0341	0.6230	0.0646
CL	Down	kNN	Small	0.9849	0.1667	0.0164	0.0299
CL	Down	SVC	Small	0.6808	0.0212	0.4754	0.0406
CL	Down	MLP	Small	0.9814	0.0000	0.0000	0.0000

Table 2 Simulated Data Metrics: Evaluation metrics for the performance of the models developed and tested on simulated data.

% Random Data	Direction	Classifier	Accuracy	Precision	Recall	F1 Score
0	Up	Tree	1.0	1.0	1.0	1.0
0	Down	Tree	.9990	.9982	.9976	.9979
0	Up	Forest	1.0	1.0	1.0	1.0
0	Down	Forest	1.0	1.0	1.0	1.0
0	Up	kNN	.9999	1.0	.9994	.9997
0	Down	kNN	.9997	1.0	.9988	.9994
0	Up	SVC	1.0	1.0	1.0	1.0
0	Down	SVC	1.0	1.0	1.0	1.0
0	Up	MLP	1.0	1.0	1.0	1.0
0	Down	MLP	.9999	1.0	.9994	.9997
50	Up	Tree	.8354	.6885	.5179	.5912
50	Down	Tree	.8296	.6672	.5581	.6078
50	Up	Forest	.7759	.5092	.6735	.5799
50	Down	Forest	.7690	.5085	.7060	.5912
50	Up	kNN	.8228	.6960	.3969	.5055
50	Down	kNN	.8147	.6891	.3953	.5024
50	Up	SVC	.8079	.5675	.6662	.6129
50	Down	SVC	.7787	.525	.6787	.5920

50	Up	MLP	.8016	.5699	.5335	.5511
50	Down	MLP	.7834	.5439	.5239	.5337
90	Up	Tree	.7572	.1429	.0025	.0048
90	Down	Tree	.7603	.4819	.0246	.0468
90	Up	Forest	.6088	.2920	.4427	.3519
90	Down	Forest	.6076	.2892	.4388	.3486
90	Up	kNN	.7397	.3343	.0675	.1123
90	Down	kNN	.7406	.2879	.0571	.0954
90	Up	SVC	.5972	.2933	.4617	.3587
90	Down	SVC	.6024	.2815	.4266	.3392
90	Up	MLP	.7265	.3056	.0952	.1452
90	Down	MLP	.7340	.3339	.1125	.1683

Table 3 Results of Real Data on Simulated Data Models with 0% Bad Data: The same data was assembled into the same format as the simulated data matrix with 7 regions of methylation calculated for the control and experimental groups, glyphosate (GLY), trifloxysulfuron (TRI), shade (SH), and clipped (CL). Genes that did not have all of these numbers were dropped. The models, decision tree (Tree), random forest (Forest), k-nearest neighbors (kNN), C support vector (SVC), and multi-layer perceptron (MLP), were developed with the simulated data and tested with the real data.

Treatment	DEG Direction	Classifier	Accuracy	Precision	Recall	F1 Score
GLY	Up	Tree	0.9215	0.0217	0.0621	0.0322
TRI	Up	Tree	0.9304	0.0099	0.0615	0.0170
SH	Up	Tree	0.9292	0.0129	0.0610	0.0212
CL	Up	Tree	0.9352	0.0101	0.1176	0.0186
GLY	Up	Forest	0.9215	0.0217	0.0621	0.0322
TRI	Up	Forest	0.9304	0.0099	0.0615	0.0170
SH	Up	Forest	0.9292	0.0129	0.0610	0.0212
CL	Up	Forest	0.9352	0.0101	0.1176	0.0186
GLY	Up	kNN	0.9507	0.0178	0.0248	0.0207
TRI	Up	kNN	0.9566	0.0088	0.0308	0.0137
SH	Up	kNN	0.9547	0.0224	0.0610	0.0328
CL	Up	kNN	0.9642	0.0099	0.0588	0.0169
GLY	Up	SVC	0.9295	0.0227	0.0559	0.0323
TRI	Up	SVC	0.9484	0.0174	0.0769	0.0284
SH	Up	SVC	0.9477	0.0114	0.0366	0.0173
CL	Up	SVC	0.9544	0.0112	0.0882	0.0198
GLY	Up	MLP	0.9087	0.0213	0.0745	0.0331
TRI	Up	MLP	0.8930	0.0091	0.0923	0.0166
SH	Up	MLP	0.8920	0.0186	0.1463	0.0330
CL	Up	MLP	0.8986	0.0094	0.1765	0.0179
GLY	Down	Tree	0.9199	0.0361	0.1184	0.0554

TRI	Down	Tree	0.9289	0.0168	0.1014	0.0288
SH	Down	Tree	0.9172	0.0211	0.0692	0.0323
CL	Down	Tree	0.9268	0.0095	0.0606	0.0165
GLY	Down	Forest	0.9335	0.0339	0.0855	0.0485
TRI	Down	Forest	0.9448	0.0225	0.1014	0.0368
SH	Down	Forest	0.9313	0.0211	0.0538	0.0304
CL	Down	Forest	0.9438	0.0098	0.0455	0.0161
GLY	Down	kNN	0.9545	0.0332	0.0461	0.0386
TRI	Down	kNN	0.9632	0.0219	0.0580	0.0317
SH	Down	kNN	0.9537	0.0169	0.0231	0.0195
CL	Down	kNN	0.9608	0.0104	0.0303	0.0154
GLY	Down	SVC	0.9286	0.0353	0.0987	0.0520
TRI	Down	SVC	0.9391	0.0173	0.0870	0.0288
SH	Down	SVC	0.9279	0.0171	0.0462	0.0249
CL	Down	SVC	0.9390	0.0089	0.0455	0.0149
GLY	Down	MLP	0.9135	0.0346	0.1250	0.0542
TRI	Down	MLP	0.8996	0.0146	0.1304	0.0263
SH	Down	MLP	0.8895	0.0227	0.1077	0.0375
CL	Down	MLP	0.8963	0.0065	0.0606	0.0117

Table 4 Genes of Interest: Differentially expressed genes with DMCs and sRNA differences were identified for further study.

Gene ID	Treatment	DEG	DMC	sRNA	GO Terms[29]
AT3G29320	GLY	-	CG-	-	cytoplasm, glycogen catabolic process, glycogen phosphorylase activity, transferase activity, transferring glycosyl groups, pyridoxal phosphate binding, linear malto-oligosaccharide phosphorylase activity, SHG alpha-glucan phosphorylase activity, 1, 4-alpha-oligoglucan phosphorylase activity, response to water deprivation, plastid, response to temperature stimulus, chloroplast, cytosol
AT1G53480	GLY	+	CG-	+	
AT3G61470	GLY	-	CG-	-	response to light stimulus, chloroplast thylakoid membrane, photosynthesis, photosynthesis, light harvesting in photosystem I, photosystem I antenna complex, chlorophyll binding, chloroplast thylakoid membrane, chloroplast thylakoid membrane, response to cold, response to high light intensity, response to low light intensity stimulus, protein binding, photosynthesis,

					cytosol, chlorophyll binding
AT5G62360	GLY	-	CHH+	-	enzyme inhibitor activity, negative regulation of catalytic activity, cytoplasm, plasma membrane, plant-type cell wall modification
AT3G50460	GLY	+	CG-	+	defense response to fungus, response to fungus
AT5G62350	GLY	-	CHH+	-	enzyme inhibitor activity, negative regulation of catalytic activity, plastid
ATCG01180	GLY	-	CHH+	-	plastid large ribosomal subunit, structural constituent of ribosome, translation
AT1G34060	GLY	+	CG-	+	cellular amino acid metabolic process, transaminase activity, carbon-sulfur lyase activity
AT3G28510	GLY	+	CG-	+	endoplasmic reticulum
AT4G04540	GLY	+	CG-	-	protein kinase activity, protein serine, threonine kinase activity, plasma membrane, protein phosphorylation, protein phosphorylation, kinase activity
AT3G25010	GLY	+	CG-	+	kinase activity, signal transduction, leaf senescence, cellular response to abscisic acid stimulus, leaf senescence, cellular response to abscisic acid stimulus
AT4G01985	TRI	+	CG-	+	
AT1G34440	TRI	+	CG-	-	
AT5G35375	TRI	+	CG-	-	
AT1G62975	TRI	-	CG-	-	RNA polymerase II transcription regulatory region sequence-specific DNA binding, DNA-binding transcription factor activity, RNA polymerase II-specific, DNA-binding transcription factor activity, regulation of transcription by RNA polymerase II, regulation of transcription by RNA polymerase II, protein dimerization activity, RNA polymerase II transcription regulator complex, DNA-binding transcription factor activity, regulation of transcription, DNA-templated
AT3G47340	SH	+	CG-	+	asparagine synthase (glutamine-hydrolyzing) activity, cytosol, asparagine biosynthetic process, L-asparagine biosynthetic process, cellular amino acid catabolic process, asparagine biosynthetic process, cellular response to sucrose starvation, response to sucrose, response to glucose, response to fructose, response to absence of light, cytosol
AT2G05380	SH	+	CHH+	+	mitochondrion
AT5G19120	SH	+	CG-	+	aspartic-type endopeptidase activity, cellular response to hypoxia

Discussion:

We wanted to compare sRNA, gene expression, and DNA methylation to identify connections between them that would show the role of these epigenetic mechanisms in Arabidopsis' stress response. In part, that meant searching for overall patterns of methylation and sRNA driving gene expression changes on a global scale, and it also meant checking individual genes on a local scale. We combined all types of data from our stress-response study and used this matrix to try to identify genes with differential epigenetic activity. We also used established data mining classifiers to identify links between gene expression and epigenetic responses involving sRNA and DNA methylation. These algorithms can be used to sift through high-dimensional, noisy data and make connections with features not easily observed.[21] Initially the models could not consistently identify differentially expressed genes with the features we selected. Although we had many features, there are still aspects of gene regulation that could not be captured in this study. Due to concerns about overfitting, we tried reducing the matrix to what we perceived to be key features, and we tried weighting the differentially expressed genes to balance the dataset.

One interesting outcome of this work was a set of scripts and methods to link all the different data types in a way that was useful for detecting potential broad patterns of epigenetic activity. Without existing software to do this, we had to come up with our own methods and find ways to represent the information in a numerical format. Connecting all of our sequence types and seeing whether these patterns could be detected by the classification algorithms was an important step in the analysis. Even though the resulting models did not detect broad patterns of epigenetic regulation influencing gene expression in our experimental dataset, such techniques could prove useful in another system or with a different stress. If other researchers happen to run similar experiments and have the same data types, they too might wish to look for patterns.

While we did not detect an overall broad pattern of methylation or sRNA levels that allowed us to predict expression, this does not rule out the possibility of local methylation or sRNA processes influencing gene transcription or transposon activity. Using a weighted scoring system, we identified some genes for further study using the scoring system. AT3G29320 (PHS1) is involved in starch granule synthesis.[47] AT1G53480 (MRD1) was upregulated in glyphosate treated Arabidopsis, and it is believed to promote salicylic acid defenses.[48] AT3G61470 was downregulated in glyphosate-treated plants and is a subunit of the photosystem I light harvesting complex.[49] ATCG01180 codes for the 23S ribosomal subunit of the chloroplast.[50] Interestingly, the chloroplast was mostly unmethylated when the methylation of the plastid was determined in Chapter 3. AT1G62975 is an upregulated transcription factor in trifloxysulfuron-treated Arabidopsis that could regulate other genes.[51] A dark inducible transcription factor (AT3G47340) was upregulated in the shaded plants.[52] Further study could verify the role of epigenetics, if any in the differential expression cases.

Methylation and sRNA are believed to have some influence over gene expression levels. The methylation patterns are linked to varying level of expression.[53] Higher levels of non-CG methylation are associated with repression of genes and transposons.[54,55] However, broad changes to methylation patterns did not appear to correlate well with changes in gene expression with these data. When our simulated dataset model was given increasing amounts of randomly assigned data, it quickly lost its ability to reliably detect DEGs. We may have been unable to sift

through the noise to detect any existing patterns in our actual data. There are many mechanisms that are believed to influence gene expression and therefore create noise that interferes with efforts to estimate expression based on methylation. For example, histone marks are also an epigenetic mark associated with expression changes, but we did not have data on histone status. Other proteins also regulate expression levels.[56] Negative regulators may bind to the promoter of genes to prevent transcription, and positive regulators can assist transcription machinery to produce more RNA from a gene. We added some noise to our simulated dataset to mimic this. The resulting models lost the ability to reliably distinguish DEGs from other genes. It is possible that the bulk of the long-term responses were mediated by these other regulatory mechanisms, and that would explain why we did not detect them.

Another group studied stress responses and epigenetics in plants with a clearer example of epigenetic regulation involving methylation, but histone modifications played an important role. Their study found that in maize long-term perturbation of methylation near a transposon was associated with changes to the histone mark, H3K27me3.[57] Both remained disrupted for 8 generations when the RdDM pathway was disrupted. Methylation has several self-reinforcing loops with histone marks. They hypothesized that the long-term disruption of methylation caused the histone marks to change and thus break the loop. This ultimately caused long-term increased transposon activity. Perhaps the primary role of DNA methylation is to help regulate histone modifications, which could be the major driver of regulation difference.[57] With our data, the changes to methylation may not have had adequate time to alter the histones in a way that would influence expression after a long recovery, or the stress did not cause many changes in the RdDM regulation that would affect expression. The herbicide treated plants were only sprayed once, and the clipped plants were cut 3 times. In those cases, the plants were exposed to a sudden stressor, which was then alleviated. However, the continuously shaded plants did not have a strong correlation between changes to methylation and gene expression.

There is also the possibility that the epigenetic pathways we studied did not broadly influence long-term gene expression in response to the stresses imposed. In chapter 3, patterns of gene methylation were clustered with k means, and the resulting clusters resemble those obtained in another study.[53] They featured one with high methylation in all contexts, low methylation in all contexts, CG body methylation, and high methylation in the promoter. It is worth noting that very few genes switched from the lowest CHH methylation groups, promoter and no methylation, to the highest methylation group. Perhaps there may have been a stronger methylation response when the clipping or herbicide stresses were initially imposed, but the levels returned to normal as the plants recovered. In contrast, shade treated plants were continuously stressed until harvest, so this acclimation may not have occurred. However, the continuous nature of this stress may have resulted in different acclimations as a result of this long-term exposure. There was no dramatic change in conditions that required a long recovery period; they just grew slowly and steadily.

Overall, we found that to identify genomic scale patterns between regulation and gene expression we had to reduce the noise found in typical ‘omics’ datasets. For example, we found no pattern when mining the data from weed control treatments, but when we mined simulated data we were able to detect the predicted patterns associated with regulation of gene expression. As discussed, the noise that perturbs the models can arise from various sources, including the

possibility that DNA methylation levels and gene expression are not tightly connected. This relationship was recently tested, where researchers generated Arabidopsis plants without the ability to methylate DNA and noted, among a plethora of non-regulatory impacts, a dose-dependent connection between methylation and gene expression.[58] In the case of the present study, this could be further tested by altering the methylation state of the individual sites we identified with the scoring system. If an altered DNA methylation state throughout these genes causes changes to their expression levels, that would be evidence that DNA methylation can, at least at a local level, influence gene expression and facilitate stress response. It would also suggest that researchers should carefully consider the scale at which their biological question impacts methylation and gene expression.

References:

(1) Vailati-Riboni, M.; Palombo, V.; Loor, J. J. What Are Omics Sciences? In *Periparturient Diseases of Dairy Cows*; Ametaj, B. N., Ed.; Springer International Publishing: Cham, 2017; pp 1–7. https://doi.org/10.1007/978-3-319-43033-1_1.

(2) van Belkum, A.; Struelens, M.; de Visser, A.; Verbrugh, H.; Tibayrenc, M. Role of Genomic Typing in Taxonomy, Evolutionary Genetics, and Microbial Epidemiology. *Clin Microbiol Rev* **2001**, *14* (3), 547–560. https://doi.org/10.1128/CMR.14.3.547-560.2001.

(3) Pelizzola, M.; Ecker, J. R. The DNA Methylome. *FEBS Letters* **2011**, *585* (13), 1994–2000. https://doi.org/10.1016/j.febslet.2010.10.061.

(4) Metzker, M. L. Sequencing Technologies — the next Generation. *Nat Rev Genet* **2010**, *11* (1), 31–46. https://doi.org/10.1038/nrg2626.

(5) Xiao, T.; Zhou, W. The Third Generation Sequencing: The Advanced Approach to Genetic Diseases. *Transl Pediatr* **2020**, *9* (2), 163–173. https://doi.org/10.21037/tp.2020.03.06.

(6) Wetterstrand, K. A. *DNA Sequencing Costs*; NHGRI Genome Sequencing Program. www.genome.gov/sequencingcostsdata.

(7) Shendure, J.; Aiden, E. L. The Expanding Scope of DNA Sequencing. *Nat Biotechnol* **2012**, *30* (11), 1084–1094. https://doi.org/10.1038/nbt.2421.

(8) Langmead, B. Aligning Short Sequencing Reads with Bowtie. *Curr. Protoc. Bioinform.* **2010**, *32* (1). https://doi.org/10.1002/0471250953.bi1107s32.

(9) Krueger, F.; Andrews, S. R. Bismark: A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications. *Bioinformatics* **2011**, *27* (11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.

(10) Ward, R. M.; Schmieder, R.; Highnam, G.; Mittelman, D. Big Data Challenges and Opportunities in High-Throughput Sequencing. *Systems Biomedicine* **2013**, *1* (1), 29–34. https://doi.org/10.4161/sysb.24470.

(11) Teng, C.-S.; Wu, B.-H.; Yen, M.-R.; Chen, P.-Y. MethGET: Web-Based Bioinformatics Software for Correlating Genome-Wide DNA Methylation and Gene Expression. *BMC Genomics* **2020**, *21* (1), 375. https://doi.org/10.1186/s12864-020-6722-x.

(12) Frasca, F.; Matteucci, M.; Leone, M.; Morelli, M. J.; Masseroli, M. Accurate and Highly Interpretable Prediction of Gene Expression from Histone Modifications. *BMC Bioinformatics* **2022**, *23* (1), 151. https://doi.org/10.1186/s12859-022-04687-x.

(13) Jagtap, S.; Pirayre, A.; Bidard, F.; Duval, L.; Malliaros, F. D. BRANEnet: Embedding Multilayer Networks for Omics Data Integration. *BMC Bioinformatics* **2022**, *23* (1), 429. https://doi.org/10.1186/s12859-022-04955-w.

(14) Zoppi, J.; Guillaume, J.-F.; Neunlist, M.; Chaffron, S. MiBiOmics: An Interactive Web Application for Multi-Omics Data Exploration and Integration. *BMC Bioinformatics* **2021**, *22* (1), 6. https://doi.org/10.1186/s12859-020-03921-8.
(15) Jin, T.; Rehani, P.; Ying, M.; Huang, J.; Liu, S.; Roussos, P.; Wang, D. ScGRNom: A Computational Pipeline of Integrative Multi-Omics Analyses for Predicting Cell-Type Disease Genes and Regulatory Networks. *Genome Med* **2021**, *13* (1), 95. https://doi.org/10.1186/s13073-021-00908-9.
(16) Muñoz-Benavent, M.; Hartkopf, F.; Van Den Bossche, T.; Piro, V. C.; García-Ferris, C.; Latorre, A.; Renard, B. Y.; Muth, T. GNOMO: A Multi-Omics Pipeline for Integrated Host and Microbiome Analysis of Non-Model Organisms. *NAR Genomics and Bioinformatics* **2020**, *2* (4), lqaa083. https://doi.org/10.1093/nargab/lqaa083.
(17) Fisch, K. M.; Meißner, T.; Gioia, L.; Ducom, J.-C.; Carland, T. M.; Loguercio, S.; Su, A. I. Omics Pipe: A Community-Based Framework for Reproducible Multi-Omics Data Analysis. *Bioinformatics* **2015**, *31* (11), 1724–1728. https://doi.org/10.1093/bioinformatics/btv061.
(18) Fondi, M.; Liò, P. Multi -Omics and Metabolic Modelling Pipelines: Challenges and Tools for Systems Microbiology. *Microbiological Research* **2015**, *171*, 52–64. https://doi.org/10.1016/j.micres.2015.01.003.
(19) Sartor, R. C.; Noshay, J.; Springer, N. M.; Briggs, S. P. Identification of the Expressome by Machine Learning on Omics Data. *Proc Natl Acad Sci USA* **2019**, *116* (36), 18119–18125. https://doi.org/10.1073/pnas.1813645116.
(20) Matzke, M. A.; Mosher, R. A. RNA-Directed DNA Methylation: An Epigenetic Pathway of Increasing Complexity. *Nat Rev Genet* **2014**, *15* (6), 394–408. https://doi.org/10.1038/nrg3683.
(21) Tan, P.-N.; Steinbach, M.; Karpatne, A.; Kumar, V. *Introduction to Data Mining*, Second edition.; Pearson: NY NY, 2019.
(22) Zhang, Z. Introduction to Machine Learning: K-Nearest Neighbors. *Ann. Transl. Med.* **2016**, *4* (11), 218–218. https://doi.org/10.21037/atm.2016.03.37.
(23) Safavian, S. R.; Landgrebe, D. A Survey of Decision Tree Classifier Methodology. *IEEE Trans. Syst., Man, Cybern.* **1991**, *21* (3), 660–674. https://doi.org/10.1109/21.97458.
(24) Song, Y.-Y.; Lu, Y. Decision Tree Methods: Applications for Classification and Prediction. *Shanghai Arch Psychiatry* **2015**, *27* (2), 130–135. https://doi.org/10.11919/j.issn.1002-0829.215044.
(25) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
(26) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
(27) Catoni, M.; Tsang, J. M.; Greco, A. P.; Zabet, N. R. DMRcaller: A Versatile R/Bioconductor Package for Detection and Visualization of Differentially Methylated Regions in CpG and Non-CpG Contexts. *Nucleic Acids Research* **2018**. https://doi.org/10.1093/nar/gky602.
(28) Akalin, A.; Kormaksson, M.; Li, S.; Garrett-Bakelman, F. E.; Figueroa, M. E.; Melnick, A.; Mason, C. E. MethylKit: A Comprehensive R Package for the Analysis of Genome-

Wide DNA Methylation Profiles. *Genome Biol* **2012**, *13* (10), R87. https://doi.org/10.1186/gb-2012-13-10-r87.
(29) *The Arabidopsis Information Resource (TAIR)*. https://www.arabidopsis.org/download/index-auto.jsp%3Fdir%3D%252Fdownload_files%252FGenes%252FTAIR10_genome_release.
(30) Handzlik, J. E.; Tastsoglou, S.; Vlachos, I. S.; Hatzigeorgiou, A. G. Manatee: Detection and Quantification of Small Non-Coding RNAs from next-Generation Sequencing Data. *Sci Rep* **2020**, *10* (1), 705. https://doi.org/10.1038/s41598-020-57495-9.
(31) Love, M. I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol* **2014**, *15* (12), 550. https://doi.org/10.1186/s13059-014-0550-8.
(32) Flynn, J. M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A. G.; Feschotte, C.; Smit, A. F. RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families. *Proc Natl Acad Sci USA* **2020**, *117* (17), 9451–9457. https://doi.org/10.1073/pnas.1921046117.
(33) Szcześniak, M. W.; Rosikiewicz, W.; Makałowska, I. CANTATAdb: A Collection of Plant Long Non-Coding RNAs. *Plant Cell Physiol* **2016**, *57* (1), e8–e8. https://doi.org/10.1093/pcp/pcv201.
(34) Hastie, T.; Tibshirani, R.; Friedman, J. Random Forests. In *The Elements of Statistical Learning*; Springer Series in Statistics; Springer New York: New York, NY, 2009; pp 587–604. https://doi.org/10.1007/978-0-387-84858-7_15.
(35) Shi, Y. *Optimization Based Data Mining: Theory and Applications*; Advanced information and knowledge processing; Springer-Verlag: London ; New York, 2011.
(36) Alzubi, J.; Nayyar, A.; Kumar, A. Machine Learning from Theory to Algorithms: An Overview. *J. Phys.: Conf. Ser.* **2018**, *1142*, 012012. https://doi.org/10.1088/1742-6596/1142/1/012012.
(37) Dietterich, T. Overfitting and Undercomputing in Machine Learning. *ACM Comput. Surv.* **1995**, *27* (3), 326–327. https://doi.org/10.1145/212094.212114.
(38) Murphy, K. P. *Machine Learning: A Probabilistic Perspective*; Adaptive computation and machine learning series; MIT Press: Cambridge, MA, 2012.
(39) *Geomatic Approaches for Modeling Land Change Scenarios*, 1st ed. 2018.; Camacho Olmedo, M. T., Escobar, F., Mas, J.-F., Paegelow, M., Eds.; Lecture Notes in Geoinformation and Cartography; Springer International Publishing : Imprint: Springer: Cham, 2018. https://doi.org/10.1007/978-3-319-60801-3.
(40) Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; Layton, R.; Vanderplas, J.; Joly, A.; Holt, B.; Varoquaux, G. API Design for Machine Learning Software: Experiences from the Scikit-Learn Project. **2013**. https://doi.org/10.48550/ARXIV.1309.0238.
(41) dankal444. *Use mean, min, max and stddev to create 2D dataset*. Stack Overflow. https://stackoverflow.com/questions/69932063/use-mean-min-max-and-stddev-to-create-2d-dataset/.
(42) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(43) Li, W.; Guo, Q. Plotting Receiver Operating Characteristic and Precision–Recall Curves from Presence and Background Data. *Ecology and Evolution* **2021**, *11* (15), 10192–10206. https://doi.org/10.1002/ece3.7826.
(44) DHARWAL, A. *Complete Guide to Understanding Precision and Recall Curves*. Developer's Corner. https://analyticsindiamag.com/complete-guide-to-understanding-precision-and-recall-curves/.
(45) Tam D Tran-The. *Precision-Recall Curve is More Informative than ROC in Imbalanced Data: Napkin Math & More*. Towards Data Science. https://towardsdatascience.com/precision-recall-curve-is-more-informative-than-roc-in-imbalanced-data-4c95250242f6.
(46) Alakus, T. B.; Turkoglu, I. Comparison of Deep Learning Approaches to Predict COVID-19 Infection. *Chaos, Solitons & Fractals* **2020**, *140*, 110120. https://doi.org/10.1016/j.chaos.2020.110120.
(47) Mérida, A.; Fettke, J. Starch Granule Initiation in *Arabidopsis Thaliana* Chloroplasts. *Plant J* **2021**, *107* (3), 688–697. https://doi.org/10.1111/tpj.15359.
(48) Singh, A.; Sharma, A.; Singh, N.; Nandi, A. K. MTO1-RESPONDING DOWN 1 (MRD1) Is a Transcriptional Target of OZF1 for Promoting Salicylic Acid-Mediated Defense in Arabidopsis. *Plant Cell Rep* **2022**, *41* (5), 1319–1328. https://doi.org/10.1007/s00299-022-02861-2.
(49) Varotto, C.; Pesaresi, P.; Jahns, P.; Leßnick, A.; Tizzano, M.; Schiavon, F.; Salamini, F.; Leister, D. Single and Double Knockouts of the Genes for Photosystem I Subunits G, K, and H of Arabidopsis. Effects on Photosystem I Composition, Photosynthetic Electron Flow, and State Transitions. *Plant Physiology* **2002**, *129* (2), 616–624. https://doi.org/10.1104/pp.002089.
(50) *Biochemistry & Molecular Biology of Plants*, Second edition.; Buchanan, B. B., Gruissem, W., Jones, R. L., Eds.; John Wiley & Sons Inc: Chichester, West Sussex ; Hoboken, NJ, 2015.
(51) Toledo-Ortiz, G.; Huq, E.; Quail, P. H. The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family[W]. *The Plant Cell* **2003**, *15* (8), 1749–1770. https://doi.org/10.1105/tpc.013839.
(52) Ufaz, S.; Shukla, V.; Soloveichik, Y.; Golan, Y.; Breuer, F.; Koncz, Z.; Galili, G.; Koncz, C.; Zilberstein, A. Transcriptional Control of Aspartate Kinase Expression during Darkness and Sugar Depletion in Arabidopsis: Involvement of BZIP Transcription Factors. *Planta* **2011**, *233* (5), 1025–1040. https://doi.org/10.1007/s00425-011-1360-9.
(53) Bewick, A. J.; Schmitz, R. J. Gene Body DNA Methylation in Plants. *Current Opinion in Plant Biology* **2017**, *36*, 103–110. https://doi.org/10.1016/j.pbi.2016.12.007.
(54) Niederhuth, C. E.; Bewick, A. J.; Ji, L.; Alabady, M. S.; Kim, K. D.; Li, Q.; Rohr, N. A.; Rambani, A.; Burke, J. M.; Udall, J. A.; Egesi, C.; Schmutz, J.; Grimwood, J.; Jackson, S. A.; Springer, N. M.; Schmitz, R. J. Widespread Natural Variation of DNA Methylation within Angiosperms. *Genome Biol* **2016**, *17* (1), 194. https://doi.org/10.1186/s13059-016-1059-0.
(55) Wassenegger, M.; Heimes, S.; Riedel, L.; Sänger, H. L. RNA-Directed de Novo Methylation of Genomic Sequences in Plants. *Cell* **1994**, *76* (3), 567–576. https://doi.org/10.1016/0092-8674(94)90119-8.
(56) Raven, P. H. *Biology*, Tenth edition.; McGraw-Hill: New York, NY, 2014.

(57) Guo, W.; Wang, D.; Lisch, D. RNA-Directed DNA Methylation Prevents Rapid and Heritable Reversal of Transposon Silencing under Heat Stress in Zea Mays. *PLoS Genet* **2021**, *17* (6), e1009326. https://doi.org/10.1371/journal.pgen.1009326.
(58) He, L.; Huang, H.; Bradai, M.; Zhao, C.; You, Y.; Ma, J.; Zhao, L.; Lozano-Durán, R.; Zhu, J.-K. DNA Methylation-Free Arabidopsis Reveals Crucial Roles of DNA Methylation in Regulating Gene Expression and Development. *Nat Commun* **2022**, *13* (1), 1335. https://doi.org/10.1038/s41467-022-28940-2.