THE MOLECULAR AND EVOLUTIONARY OUTCOMES OF TRANSGENERATIONAL
PLASTICITY IN *MIMULUS GUTTATUS*

BY

JACK COLICCHIO

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the
Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for
the degree of Doctor of Philosophy.

_____

Chair: Lena C. Hileman

_____

John K. Kelly

_____

Justin P. Blumenstiel

_____

Mark T. Holder

_____

Stuart J. MacDonald

Date Defended: 24 May 2016

The dissertation committee for JACK COLICCHIO certifies that this is the approved version of the following dissertation:

THE MOLECULAR AND EVOLUTIONARY OUTCOMES OF TRANSGENERATIONAL
PLASTICITY IN *MIMULUS GUTTATUS*

_____

Chair: Lena C. Hileman

Date Approved: 24 May 2016

Abstract

Derived from the same family line, in the same greenhouse, and self-pollinated in an identical fashion, what effect could hole-puncher induced leaf damage have on the offspring of these individuals? Across three independent experiments during my graduate career I have demonstrated the diverse array of lingering epigenetic, gene expression, phenotypic, and fitness effects that simple mechanical wounding has on the following generation. While focused initially on the epigenetic and gene expression basis of increased trichome density in the offspring of damaged plants, it became clear early on that a host of other pathways were also differentially regulated. In the end I identified hundreds of differentially expressed genes and thousands of genomic regions where DNA methylation varied depending on parental wounding. Along with identifying numerous differentially regulated hormone synthesis genes and secondary metabolism pathways, I twice confirmed the differential regulation of the previously identified transgenerationally plastic *Mimulus guttatus MYB Mixta-like 8 (MgMYBML8)*. A brief foray into molecular epigenetics and gene expression modeling provided our lab with the necessary information to utilize DNA methylation data and test hypotheses regarding the role of DNA methylation on transgenerational inheritance. Finally, through the use of a two-generation common garden experiment I demonstrated that transgenerational effects alter the development and resistance of plants in nature. While much remains to be deciphered regarding the molecular underpinnings, evolutionary role, and ecological relevance of transgenerational inheritance, the work presented herein provides a relatively comprehensive look at the complexity underlying an extremely simplified case of transgenerational inheritance.

Table of Contents

List of Figures

List of Tables

List of Appendices

Preface

On a mountain meadow freshly clear of snow, the imminent spectacle of diversity is shrouded in a uniform coat of germinating green. In a few weeks the slope will be a patchy assortment of textures and hues, but for now that potential is hidden within the cell walls of each seedling. From a common origin, what causes a single wildflower to develop as it does, unique from each other nearby speck of life? Fundamentally, genetic and environmental differences are the only factors that will inevitably differentiate any two organisms. Were earth homogenous, genetic and environmental factors would not interact, biology would be simpler, and my doctoral dissertation would be a master's thesis.

Instead, the ability to detect environmental conditions and alter growth accordingly has been selected for time and time again, leading to a third cause of natural variation: environmental x genetic factors. These factors, known as "phenotypic plasticity", have evolved to respond to countless biotic and abiotic parameters, and are present across the diversity of life. If current conditions are a reliable predictor of future conditions, phenotypic plasticity is advantageous. Following the hypothesis above, if current environmental conditions are correlated with the environment of the next generation, then the transmission of environmentally altered developmental trajectories between generations (transgenerational phenotypic plasticity; TPP) are also adaptive.

This ability for the environment experienced by an individual to alter the growth of their offspring is the focus of my thesis, and represents a unique basis of variation. What causes a single wildflower to develop as it does? An assemblage of genetic factors, environmental factors, and interactions between these two that occur within a lifetime (phenotypic plasticity) or are inherited as lingering signals from generations past (TPP).

1

Introduction

Plant-herbivore interactions represent one of the most common biological interactions on earth, positioned at a fundamental position near the base of most food webs. Genetic selection in response to local herbivore levels, plastic responses to herbivore damage, and transgenerational plasticity to conditions experienced in the prior generation allow plants to modify their growth in the presence of herbivores. While these first two classes of plant responses to herbivores have been popular avenues of research for decades (Caswell and Reed 1976; Rhoades and Cates 1976; Coppock *et al.* 1983; Coley 1988; Mole 1994; Agrawal *et al.* 2002; Wu and Baldwin 2010; Agrawal *et al.* 2012), the role of TPP in this, or any, response has only recently received attention (AGRAWAL *et al.* 1999; AGRAWAL 2002; GALLOWAY and ETTERSON 2007; HOLESKI *et al.* 2013b). Plant's detect and respond to herbivores through a myriad of mechanisms, including chemical receptors that detect insect saliva, regurgitant, and chitin (Gatehouse 2002), as well as gaseous signaling molecules (Yan *et al.* 2013), and more general responses to mechanical wounding (Mithöfer *et al.* 2005). In this thesis I investigate the gene expression, epigenetic, and evolutionary consequences of TPP to mechanical wounding in *Mimulus guttatus.*

Prior to my arrival at the University of Kansas it was discovered that numerous lines of *M. guttatus* respond to wounding (hole-puncher induced) through the production of more trichomes (leaf hairs) (Holeski 2007). Additionally, a single candidate gene was identified that appeared to be associated with this response (Scoville *et al.* 2011), and a genomic region was mapped that controls within generation trichome plasticity (Holeski *et al.* 2010). During the course of my dissertation research I have expanded upon this system by comparing whole genome epigenetic and gene expression changes in response to

parental wounding, how these two are related, and the evolutionary potential of this system of inheritance.

In the first chapter of my thesis I utilize a known transgenerationally responsive, and nearly completely homozygous line (Recombinant Inbred Line 94: RIL 94) to explore how gene expression patterns shift in response to parental wounding. By growing seedlings in common conditions and only altering the presence or absence of wounding in their parents we were able to assay the role of TPP on plant development. Prior to this experiment, candidate gene approaches had identified a few genes that responded plastically to parental environments (Bilichak *et al.* 2012; Rasmann *et al.* 2012); however this was the first analysis to look into the global patterns of differential expression in TPP. This study allowed us to discover the magnitude of the genome that shows significant regulatory shift in response to parental wounding (*ca.* 4%), the patterns of co-regulation in this response, and identify other candidate pathways and phenotypes that appear highly plastic to parental environment.

Once I discovered the vast scale of differential gene expression due to parental wounding, I began to wonder what the cause of this differential expression was. Previous work in the field had pointed at germ-line transmissive epigenetic regulatory elements as being of prime importance in TPP (HOLESKI *et al.* 2012; MCCUE *et al.* 2012; RASMANN *et al.* 2012; SHEN *et al.* 2012; SLAUGHTER *et al.* 2012; VERHOEVEN and VAN GURP 2012b; YANG *et al.* 2012; BOND and BAULCOMBE 2014b). Epigenetic regulation consists of a three-part system in which histone modification, small RNAs, and DNA methylation are regulated by developmental and environmental signals, and function in unison to regulate gene expression. Due to rapidly advancing methods(Miura *et al.* 2012), and a clear role in other

cases of plasticity (Verhoeven *et al.* 2010; Kou *et al.* 2011; Dowen *et al.* 2012) DNA methylation was of particular interest to me. However, before I began looking for patterns of differential methylation, a brief foray to Japan to learn a new method of methylome sequencing and construct a *M. guttatus* reference methylome was necessary.

Through the use of whole-genome bisulfite sequencing I was able to estimate the methylation frequency for most of cytosine nucleotides in the *M. guttatus* genome. Genome wide results confirmed the relative normalcy of this methylome compared to other angiosperms (ZHANG *et al.* 2006; LI *et al.* 2012; SAZE *et al.* 2012; EICHTEN *et al.* 2013a; WANG *et al.* 2014), but the primary focus of this study was to determine the relationship between methylation and gene expression. We utilized prior gene expression results to determine that gene size and the various classes of DNA methylation could explain approximately 20% of variation in gene expression. Then, I returned to the experimental system utilized in our differential expression study to compare patterns of methylation in the offspring of damaged plants, and see if the predicted regulatory role of DNA methylation on gene expression explains the patterns observed in the progeny of damaged plants.

We found thousands of genomic regions where the offspring of damaged and control plants were differentially methylated. These patterns of differential methylation were congruent with our hypotheses regarding the role of methylation in gene regulation, and confirm a role of differential methylation in TPP. Of significant interest we found evidence that some classes of DNA methylation seem to regulate up-stream environmentally responsive proteins, and other classes seem to be involved in the down-stream outcomes of plasticity. Additionally, the offspring of wounded plants have more variable gene body methylation and gene expression than the progeny of control plants.

From these three chapters we shed new light on the basis of TPP and the relationship between DNA methylation and gene expression in well-controlled greenhouse conditions. In my final chapter I consider TPP to leaf wounding across multiple natural populations, and the effects of this treatment on their offspring's growth in the field.

To test for the effects of TPP in nature, I assayed 16 natural *M. guttatus* populations for trichome density and herbivory, and collected seed and grew their progeny in the greenhouse. During this experimental generation the plants were split into damaged and control groups, and seed was collected from hand-pollinated plants in each group. Then I took seed form these offspring to field common gardens to explore the potential role of parental environment on offspring growth and resistance in field conditions. Very few other studies had documented transgenerational plasticity of traits related to fitness components in nature (Galloway and Etterson 2007; Galloway and Etterson 2009), and here we provide evidence for the relevance of TPP in this context.

Taken together, the observations presented in this dissertation offer a glimpse into the epigenetic underpinnings, diverse gene expression responses, and potential evolutionary consequences of a particular type of TPP in a single species. While this work needs to be greatly expanded upon both taxonomically and ecologically, it provides a foundation for such future studies, and suggests that in addition to genetic inheritance, environmental information can be directly passed between generations.

CHAPTER 1*

Gene expression plasticity resulting from parental leaf damage in *Mimulus guttatus*

Abstract

- Leaf trichome density in *Mimulus guttatus* can be altered by parent environment. In this study, we compare global gene expression patterns in progeny of damaged and control plants. Significant differences in gene expression likely explain the observed trichome response, and identify additional responsive pathways.

- Using whole transcriptome RNA sequencing, we estimate differential gene expression between isogenic seedlings whose parents had, or had not, been subject to leaf damage.

- We identify over 900 genes that are differentially expressed in response to parental wounding. These genes cluster into groups involved in cell wall and cell membrane development, stress response pathways, and secondary metabolism.

- Gene expression is modified due to parental environment in a targeted way that likely alters multiple developmental pathways, and may increases progeny fitness if they experience environments similar to that of their parents.

Introduction

Transgenerational phenotypic plasticity, whereby a component of individual phenotype is determined by parental environment, has been demonstrated for a number of environmental factors in a variety of plant species (AGRAWAL *et al.* 1999; HOLESKI *et al.* 2012; JABLONKA 2012). Transgenerational plasticity can be adaptive when the environment is variable and offspring are likely to experience environmental conditions similar to their parents (MOUSSEAU and FOX 1998; GALLOWAY and ETTERSON 2007). Indeed, many studies have linked parent pathogen and herbivore damage (simulated and real) to offspring defensive phenotypes including trichome density, growth rate, and the production of secondary metabolites (AGRAWAL 2002; HOLESKI 2007; SCOVILLE *et al.* 2011; RASMANN *et al.* 2012; SLAUGHTER *et al.* 2012; HOLESKI *et al.* 2013b). Abiotic stresses such as salt (BOYKO and KOVALCHUK 2010), UV radiation (MOLINIER *et al.* 2006), heat (BOYKO *et al.* 2010), and cold (BLÖDNER *et al.* 2007) have been shown to alter progeny phenotypes as well. Clearly, a diversity of plant-environment interactions produce meiotically heritable signals that can subsequently alter offspring development.

The nature of transgenerational signals, and how they are propagated, remains a mystery. However, recent studies suggest that epigenetic mechanisms (small RNAs (sRNA), DNA methylation, and histone modification) are essential for establishing and maintaining transgenerational phenotypic plasticity (GUTZAT and MITTELSTEN SCHEID 2012; PIETERSE 2012). Environmentally induced phytohormones (*e.g.,* jasmonic acid, salicylic acid, ethylene) can lead to epigenetic regulatory changes in functional genes and transposable elements, which may be transmissible between generations (SUNKAR *et al.* 2007; CHINNUSAMY *et al.* 2008; VERHOEVEN *et al.* 2010; SI-AMMOUR *et al.* 2011). Additionally, in

Arabidopsis thaliana, functional epigenetic machinery (in parent plants) appears necessary

for the transmission of induced herbivore resistance in subsequent generations (RASMANN

et al. 2012). Law *et al.* (2013) have recently shown that a component of the sRNA silencing

machinery (SHH1) probes the histone code for repressive marks, and identifies regions of

the genome to be transcribed into mobile small interfering RNAs (siRNA). Therefore,

targeted developmental or environmentally responsive histone modifications may

generate locus specific mobile stress response siRNAs capable of entering the germ line

and reiterating stress-induced epigenetic markings. Yet how epigenetic signals are

produced, transmitted, and affect specific phenotypes remains unclear (MIROUZE and

PASZKOWSKI 2011).

To investigate the mechanisms by which transgenerational signals propagate offspring

phenotypic changes, we have studied a subset of recombinant-inbred lines (RILs), created

through a cross of coastal (Point Reyes) and inland (Iron Mountain 767) *Mimulus guttatus*

[Phrymacae (BEARDSLEY and OLMSTEAD 2002)], which exhibit transgenerational plasticity of

trichome density in response to parental leaf damage. Offspring of leaf-damaged parents

develop leaves with higher trichome density compared to offspring of control plants

(HOLESKI 2007; HOLESKI *et al.* 2010; SCOVILLE *et al.* 2011)(Figure 1.1 and Appendix 1). We

hypothesize that this response occurs when trichome production genes are differentially

expressed due to the inheritance of a transgenerational wounding signal. Indeed, Scoville et

al. (2011) found that a putative negative regulator of trichome production, *MgMYBML*8 (a

MYB transcription factor), is down-regulated in the progeny of damage compared to

control plants. Trichomes are thought to deter herbivores in *M. guttatus* (HOLESKI *et al.*

2010), as in other species including *Arabidopsis thaliana* and *Solanum lycopersicum* (ÅGREN

9

and SCHEMSKE 1993; RASMANN *et al.* 2012). If herbivory varies in such a way that attack rate on parents is a good indicator of herbivore load in the next generation, epigenetic memory can confer a fitness advantage (MOUSSEAU and FOX 1998).

In this study we use a single RIL (RIL 94) that exhibits transgenerational phenotypic plasticity for trichome density to analyze global gene expression plasticity in response to parental leaf damage. Using a high-throughput RNA sequencing approach, we address the scale of differential gene expression in *M. guttatus* due to parental leaf damage, the extent to which differentially expressed genes appear to be regulated by a common signal, and the evidence for enrichment of specific classes of differentially expressed genes with special attention to functional categories related to plant defense responses. Additionally, we test whether our previously observed transgenerational response of *MgMYBML8* (Scoville *et al.*, 2011) is identified in this independent study that employs a whole transcriptome approach. This study does not identify the molecular mechanism responsible for transgenerational inheritance, or the gene (or genes) responsible for a specific transgenerational phenotype. However, this study is a first quantification of transgenerational gene expression plasticity in plants, and provides strong evidence that ecologically relevant information is transmitted between generations.

METHODS

Plant Material and Experimental Treatments

A single RIL 94 parent (F8 generation) was grown under non-stress conditions in the University of Kansas greenhouse. We self-fertilized this parent and grew six progeny individuals in 4-inch pots according to standard protocols (HOLESKI 2007). Three of the six

individuals were randomly assigned to the damage treatment group. For damage treatment, we punched two holes of *ca.* 6 mm diameter in each leaf at the developmental point when the next leaf pair had expanded. We began leaf damage at the 2nd leaf pair and continued through seed collection. To produce seed for the experimental generation we self-fertilized both damaged and control individuals. We sowed this seed in flats, and grew the progeny sets (12 individuals included in transcriptome experiment) under identical, non-stress conditions (Holeski, 2007). When the second leaf pair of seedlings from the experimental population were just visible, we collected, flash froze, and stored at -80° C all above ground tissue.

Trichome Phenotyping

We grew additional progeny from the experimental generation in the absence of any stress until the 5th leaf pair expanded to confirm the trichome induction phenotype. We took trichome counts by folding the 5th leaf, tip to base, and counting the number of glandular and non-glandular trichomes on the underside of the leaf that projected from the leaf fold across the full transect of leaf width. Variation in leaf size, for example due to slight differences in developmental stage, might alter leaf trichome density estimates. Trichomes are initiated early in leaf development and dispersed across the leaf during expansion (Hülskamp et al., 1994). Therefore, by quantifying trichome production across a full transect of the leaf we are likely to mitigate the effect of leaf size or developmental stage on trichome density estimates. We have previously identified an effect of parent leaf damage on trichome production using an alternative method that relied on trichome counts within a specified leaf area (Holeski, 2007; Scoville *et al.*, 2011). We find that results from

quantification across a leaf transect are consistent with those from previous quantifications within a specified leaf area. Counts for both 5th pair leaves were taken and averaged. Means were transformed using a log10(count+0.1). We then performed a t-test to compare trichome counts between individuals from damaged (n=62) and undamaged (n=67) parents. We repeated the transgenerational trichome induction experiment using progeny of four additional damaged and four additional undamaged parent plants. Significant differences in trichome production were assessed using a nested ANOVA design with treatment as a fixed effect and parent of origin as a random effect nested within treatment.

RNA-seq and Read Mapping

We isolated total RNA from above ground seedling tissue (30-40 mg per seedling) using Direct-zol RNA Mini-Prep following manufacturer's instructions (Zymo Research Corporation, Irvine, CA), and DNAse-treated total RNA with Turbo DNA-free (Applied Biosystems/Ambion, Austin, TX, USA). DNAse-treated RNA was extracted from 12 individuals for RNA-sequencing – two individuals from each of three damage-treated and control parental plants. We prepared RNA-seq libraries using Illumina TruSeq RNA Sample Preparation Kit following manufacturer's instructions (Illumina Incorporated, San Diego, CA, USA). We Illumina indexed and sequenced (100 bp paired end reads) the twelve libraries on four HiSeq 2500 lanes (six samples per lane, each sample run in two lanes). Reads from each of the twelve samples were mapped to the *M. guttatus* 2.0 reference genome (http://www.phytozome.net/mimulus_er.php) using the CLC Large Gap Read Mapping algorithm. We used genome annotations to guide the mapping of reads that spanned introns. Reads mapped to 24,919 of 27,851 annotated transcripts. Down-stream

differential expression analyses were based on quantification of reads mapped to the

24,919 annotated genes.

Differential Expression Testing

We used a consensus-method to call differential expression, employing five commonly used

software packages. We implemented the R packages, DESeq2 (ANDERS and HUBER 2010),

EdgeR (ROBINSON *et al.* 2010), SAMseq (LI and TIBSHIRANI 2011), limma(voom) (LAW *et al.*

2013a), and NOISeq (TARAZONA *et al.* 2011) to test for differential transcript abundance

between control and treatment groups (code available upon request). Briefly, DESeq2 and

EdgeR are both parametric tests that use a negative binomial model to account for over

dispersion of read counts. EdgeR assumes a common dispersion for all genes, while DESeq2

obtains dispersion estimates based on calculated mean-variance relationships in the given

data set. NOIseq and SAMseq are non-parametric approaches. NOIseq generates a noise

distribution based on within group variation and calculates a q-value based on divergence

from this distribution for each gene. SAMseq is based on Wilcoxon statistics averaged over

numerous samplings of the data. Finally, limma(voom) uses the voom package to perform a

variance stabilizing transformation, which produces "gene weights" that mimic traditional

micro-array data, which are then analyzed as such (Soneson & Delorenzi, 2013). All

programs normalize read count per gene based on total gene depth per individual, but not

by gene length (as we are not comparing gene expression between different alleles or

genes).

In order to increase power we excluded genes with very low expression levels (fewer than

60 mapped reads across all twelve sequenced libraries). This reduced the number of genes

tested for differential expression from 24,919 to 20,748. Using limma's correlation

coefficient calculation we determined that family accounted minimally for variance in gene

expression (1.02%)(LAW *et al.* 2013a). Additionally, both EdgeR and limma (but not the

other 3 software packages) allowed for us to call differential expression using our full

nested design with parent plant nested within treatment. For both, a similar set of genes

was found to be differentially expressed when the family factor was included (102

additional genes with EdgeR and 5 additional genes with limma). Therefore, we believe

that using a simple two group experimental design to test for different expression between

parental damage and control groups in each of these five packages provides a conservative

estimate on the true number of differentially expressed genes. Using p-values from each of

the five programs, we calculated the False Discovery Rate (FDR) for each gene (null

hypothesis is no difference between treatments). We called a gene differentially expressed

if three or more of the programs found the gene to be differentially expressed with a FDR of

< 0.1. Using variance stabilized gene expression data (ANDERS and HUBER 2010), we

generated expression heatmaps in R (GENTLEMAN *et al.* 2004). We used Euclidean Distance

as the distance function, and a UPGMA based agglomerative clustering algorithm.

Gene Ontology Annotation and KEGG Mapping

We extracted the coding sequences of all genes expressed in our tissue samples and used

Blast2GO (CONESA *et al.* 2005) for function annotation. Each transcript was blasted against

the nr database, assigned Gene Ontology (GO) IDs through the mapping feature, and

annotated to predict gene function. Additionally, we used GO Slim (ASHBURNER *et al.* 2000)

in Plant DB mode to remove incorrect annotations as well as to add plant specific GO

functional categories. We tested whether differentially expressed transcripts, or subsets of these transcripts, were enriched or depleted in GO Functional, Cellular Compartment, or Molecular Processes IDs. We used a Fisher's exact test to compare GO frequencies against the reference set of all transcripts in the transcriptome, and imposed a FDR < 0.05. Differentially expressed genes were additionally mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to identify enzymes with conserved metabolic functions. Using the KEGG-mapping function in Blast2GO, we mapped these genes to their catalytic position on one of 91 separate KEGG metabolic pathway maps (KANEHISA *et al.* 2012).

RT-qPCR Confirmation

To validate differential expression identified by RNA-seq, we performed RT-qPCR for 6 candidate genes: Heat Shock Protein 6ab, Strictosidine Synthase, Tyrosine Aminotransferase, Heat Shock Protein 40, Dormancy Associated Protein, and CHY Zinc Finger (transcript ID#s 1428104, 1324230, 1358627, 1444264, 1495616, and 1315072, respectively; Appendix 2). Candidate genes were chosen to represent three genes up-regulated and three down-regulated in response to parental leaf damage, across a range of mean expression values. We designed primers using transcript sequence data from the mapping experiment, and Primer 3 software (Version 0.4.0) to identify 150-350 bp fragments that span introns (Appendix 2). We converted RNA to cDNA using iScript cDNA synthesis kit following manufacturer's instructions (Bio-Rad, Hercules, CA). We performed RT-qPCR for three technical replicates per cDNA/primer combination on a StepOnePlus Real-Time PCR System (Life Technologies, Grand Island, New York, USA) using the

protocols described in Scoville et al. (2011), with the exception of substituting Fast SYBR

Green Master Mix for the standard SYBR Green Master Mix (Life Technologies). We

normalized candidate gene expression against the housekeeping gene EF1a (SCOVILLE *et al.*

2011).

Clustering by Gene Expression

We normalized and variance stabilized gene expression values for all 919 genes found to be

differentially expressed through our consensus methodology based on the mean sample

expression as implemented in DESeq2 (ANDERS and HUBER 2010). This was done in order to

prevent clustering due to similar mean expression levels; rather we were interested in

clustering based on direction and magnitude of differential expression. From the 919

(differentially expressed genes) x 12 (individuals) data matrix, we generated a 919 x 919

pairwise dissimilarity matrix ($1-r^2$). Because we used ($1-r^2$) as the dissimilarity metric,

genes that have high positive or negative correlations across the twelve individuals, have

near zero dissimilarity scores. This allowed clustering of transcriptional repressors with a

gene, or group of genes, negatively regulated by that factor.

We implemented a dynamic tree cutting approach and a comparison of cluster stability to

first identify an appropriate number of gene clusters based on dissimilarity metrics

(LANGFELDER *et al.* 2008; BROCK *et al.* 2011). We then implemented a self-organizing map

(SOM) based clustering technique using the R package *kohonen* (hexagonal nodes, 8x6

matrix)(WEHRENS and BUYDENS 2007). This generated a topologically complex mapping of

nodes, where individual nodes contain the least variation, and the distance between nodes

is a function of the dissimilarity between nodes. Similar nodes were further grouped using

a hierarchical clustering method. Based on a visual analysis of the U-matrix (inter-node distance matrix) of our 8x6 self-organizing map, as well as *post hoc* cluster analysis, we grouped the 48 resultant nodes into four clusters, excluding nodes with high levels of variation. Finally, we performed a Fischer's Exact test of GO terms for each of these four clusters to determine if they contained significantly different abundances of GO terms.

RESULTS

Increased Trichome Production in Progeny of Damaged Parents

Progeny of leaf-damaged *M. guttatus* RIL94 plants develop leaves with 22.9 percent higher trichome density compared to control progeny (p=0.025, df=126, Appendix 1a). Repeating this experiment using eight different parent plants and their progeny, we found a 44 percent increase in trichome density in progeny of damaged compared to control plants (F_2=13.9952, p=0.0003, df=150), with no significant effect of parent of origin (F_2=0.8650, p=0.523, df=150)(Appendix 1b). This confirms transgenerational phenotypic plasticity of trichome density in RIL94 as previously demonstrated in several experiments (HOLESKI 2007; HOLESKI *et al.* 2010; SCOVILLE *et al.* 2011).

Read Mapping

Between 87 and 115 million 100 bp Illumina reads were generated from each of the 12 seedling libraries (1.19 billion reads total). Of these, between 79% and 86% of reads per individual mapped to the reference genome (Appendix 3), resulting in a total of 988 million mapped reads. As paired-end reads come from a single RNA transcript, each mapped pair of reads contributed a single unit to the expression value of a given gene. Additionally, we

discovered that RIL 94 has a single region of heterozygosity, 3.75 megabases in length at the end of chromosome 14. We were able to determine the genotype across this region for all 12 individuals (Appendix 4), and confirmed that genotype in this region of heterozygosity does not correspond with observed differential expression or clustering patterns.

Differental transcript abundance in Response to Parental Leaf Damage

Of the 20,748 genes with greater than 60 reads mapped, 10.5% (2,186) were found to be differentially expressed with a FDR of <0.1 in at least one of the five tests. RT-qPCR for a subset of transcripts confirmed RNA-seq based quantification (Appendix 5) These five tests found between 164 (NOISeq) and 1,737 (SAMSeq) genes to be differentially expressed at this stringency, with a moderate level of overlap (Figure 1.2, Appendix 6). We found 919 genes differentially expressed in three or more of the five tests, corresponding to 4.4% of the transcriptome (Figure 1.2). Only EdgeR and limma allow inclusion of family structure for assessment of differential gene expression. Using parent plant as an additional factor we found 1,276 genes to be differentially expressed using EdgeR, and 483 to be differentially expressed using limma (Appendix 7a and 7b). We observed extensive overlap between the sets of genes found differentially expressed with parent plant as an additional factor, and the set of genes found differentially expressed using our consensus methodology (Appendix 7c). Focusing on the 919 genes found to be differentially expressed by our consensus method, mean expression of the differentially expressed genes was 1,278 reads per individual, but range from as few as 6 to as many as 200,000 reads per individual, representing a full range of gene expression levels. 464 genes were up-

regulated in response to parental damage, while 455 were down-regulated (Figure 1.3, Table 1.1). Up-regulated genes had an average of 2.61 fold increase in expression in the progeny of damaged plants, while down-regulated genes had an average of 1.71 fold decrease in expression.

Confirmation of *MgMYBML8* Differential Expression

We confirmed the differential expression of the transcription factor *MgMYBML8* (Transcript ID 1491824), previously identified by Scoville *et al.* (2011) as differentially expressed in response to parental leaf damage in *M. guttatus*. This gene was only found to be significantly differentially expressed with a genome wide FDR using the SAMSeq pipeline, however its single test p-value was <0.05 in both DESeq2 (p-value = 0.01) and EdgeR (p-value = 0.02). Progeny of control plants had an average of 23.4% higher expression (normalized average of 1,439 reads in the control group, and 1,166 reads in the damage group)(Appendix 1c), and the five individuals with highest expression of *MgMYBML8* were all from the control group.

Gene Ontology Analyses Indicate Transcript Function

Of the 24,919 genes expressed in seedlings from the experimental generation, 19,573 (78.5%) were successfully BLASTed (Conesa *et al.* 2005) to genes with significant sequence similarity (E-value < 1×10^{-3}) and annotation for at least one Gene Ontology ID (GOID). 707 of 919 (76.9%) differentially expressed genes were annotated and assigned GO terms. We identified 31 GOIDs as significantly enriched (9) or depleted (22) in the set of differentially expressed transcripts relative to the seedling transcriptome (FDR < 0.1; Appendix 8, and

see Table 1.2 for a subset of GO terms most relevant to the discussion). The 22 depleted GO terms represent gene ontology categories that have a lower number of differentially expressed genes than would be expected by chance, while the 9 enriched categories show the opposite trend (Appendix 8). Three GOIDs were enriched in down-regulated relative to up-regulated transcripts (cell periphery, cell wall, and carbohydrate metabolic process; Table 1.3), and 1 GOID was enriched in up-regulated relative to down-regulated transcripts (response to abiotic stimulus; Table 1.3).

KEGG Pathway Analysis Identifies Pathways Regulated By Parental Wounding

A manual survey of KEGG metabolic pathways identified 91 pathways containing enzymes that were differentially expressed in response to parental damage. We focused on five of these pathways to generate new predictions regarding the enzymatic basis of other potential functional changes in the progeny of damaged plants. Three of these pathways are involved in the production of plant hormones (ethylene, jasmonic acid, and abscibic acid), previously linked to plant wound response. Numerous enzymes in the phenylpropanoid pathways were differentially expressed providing evidence that chemical defenses are also differentially regulated in the progeny of damaged plants. Finally, we found several differentially expressed genes in pathways leading to the production and degradation of lignin, xyloglucan, and pectin (see discussion for specific genes found differentially expressed in these pathways).

Clustering By Gene Expression Generates Four Putative Co-Regulatory Groups

Through clustering analysis we grouped 754 of 919 differentially expressed genes into four distinct clusters. These clusters ranged in size from 88 to 288 genes. An analysis of expression heatmaps for genes in these clusters confirmed that these four clusters contain groups of differentially expressed genes with highly similar expression patterns (Figure 1.4). We found that these four clusters had distinct GO term distributions, and all four were significantly enriched ($p < 0.05$) for at least one GO term relative to the other differentially expressed genes (Table 1.3 and Figure 1.4).

DISCUSSION

Differential Gene Expression in Response to Parental Leaf Damage

Transgenerational phenotypic plasticity in plants has been documented in response to a number of environmental conditions (AGRAWAL *et al.* 1999; BOYKO *et al.* 2010; HOLESKI *et al.* 2012; JABLONKA 2012; RASMANN *et al.* 2012). However, the pattern and magnitude of differential gene expression associated with altered phenotypes had not previously been explored. Using high throughput sequencing we identify 919 genes (*ca.* 4% of the seedling transcriptome) to be differentially expressed in the progeny of leaf-damaged *M. guttatus* plants (Figure 1.3). By only choosing genes that were found to be differentially expressed (FDR<0.1) in three or more of five tests, we minimize the biases of any one computational approach in identifying differentially expressed genes, although the conservative nature of this procedure may underestimate the number of affected genes.

Transcriptome responses to altered environmental conditions experienced within a plant's lifetime have been well known for over a decade (SACHS and HO 1986; TSANG *et al.* 1991;

SEKI *et al.* 2001). Recent theoretical studies have pointed out numerous scenarios for which such within-generation plasticity would prove advantageous if transmitted across generations (HERMAN *et al.* 2013b), such as when environmental heterogeneity is low, or there is a lag time between environmental cue and the appropriate response. We now provide support for the biological capacity of these transgenerartional responses; specifically we demonstrate that environmental conditions affect gene expression across generations. Through an analysis of gene functional categories we are able to reject the null hypothesis that differentially expressed genes in the progeny of leaf-damaged parents reflects a random subset of the seedling transcriptome. Instead, we find enrichment of functional categories associated with cell wall production and modification that may underlie the observed transgenerational plasticity in trichome density. Additionally, we find enrichment for categories of biological function that may regulate an inherited response to stress (plant hormones), and alter phenotypes that provide benefits under stressful conditions (secondary metabolites).

Environmental Response Pathways are Regulated by Parental Wounding

Numerous stimulus response genes are differentially expressed as a result of parental wounding (Table 1.2, 3, Appendix 8, Figure 1.3). Both up- and down-regulated differentially expressed genes are enriched for the "response to stimulus" GOID. However, more transcripts belong to the "response to abiotic stimulus" than the "response to biotic stimulus" GOID (Table 1.2), and response to abiotic stimulus transcripts are primarily up-regulated (Table 1.3). Though the leaf-damage treatment is meant to mimic herbivore damage, studies in *Arabidopsis* have noted that purely mechanical damage elicits numerous

abiotic response pathways (KORTH and DIXON 1997). In *Arabidopsis*, it has been shown that salivary compounds and insect chitin elicit additional herbivore responses (WALLING 2000). Our results suggest that *Mimulus* also responds to mechanical damage through altered regulation of more pathways associated with abiotic rather than biotic stress. The signals transmitted in response to parent leaf-damage, and affecting gene expression in stress-response pathways, are likely, in turn, to alter the regulation of trichome production (in part through cell wall biosynthesis) and secondary defensive metabolites.

Plant hormones provide upstream regulation of many of the major stress response pathways, making the differential expression of genes involved in hormone metabolism particularly interesting. Ethylene, abscisic acid, and jasmonic acid are known to function in a diverse range of environmental responses, interact and co-regulate one another, and all appear to be differentially regulated in the progeny of mechanically damaged plants (ANDERSON *et al.* 2004; BRUCE *et al.* 2007). Our gene expression results predict that in the progeny of damaged parents, ethylene levels are elevated, while abscisic acid and jasmonic acid levels are reduced. The rate limiting enzyme in ethylene production, 1-aminocyclopropane-1-carboxylate synthase (ACC synthase, gene ID 1308039)(YU *et al.* 1979) is up-regulated in the progeny of damaged plants. Additionally, three enzymes involved in the methionine salvage pathway (gene IDs 1358627, 1377394, 1478307), which recycle methionine derivatives and convert them back into a precursor of ethylene (BLEECKER and KENDE 2000), are up-regulated. These results suggest elevated ethylene levels in the progeny of damaged compared to control plants. Two enzymes involved in abscisic acid metabolism are differentially expressed in progeny of damaged plants. A rate-limiting enzyme in abscisic acid synthesis, xanthoxin dehydrogenase (GONZÁLEZ-GUZMÁN *et*

al. 2002), is down-regulated (gene ID 1382403), and the pivotal enzyme in abscisic acid degradation, abscisic acid 8'-hydroxylase (Sᴀɪᴛᴏ *et al.* 2004), is up-regulated (gene ID 1418051). Taken together, these results suggest reduced abscisic acid production in the progeny of damaged compared to control plants. Finally, three enzymes involved in jasmonic acid synthesis are down-regulated in the progeny of damaged plants, two copies of a gene that codes for 13S-lipoxygenase (gene IDs 1429895 and 1477921), and one copy of phospholipase A2 (gene ID 1389850). These enzymes are rate limiting components in the synthesis of jasmonic acid (Yᴀɴ *et al.* 2013), suggesting that jasmonic acid levels are also reduced in the progeny of damaged compared to control plants. While jasmonic acid is traditionally associated with biotic stress responses and trichome production, unpublished results from both our group and Chris Ivey (*pers. comm.*) suggest that jasmonic acid (or at least its methylated form, methyl jasmonate) may be a negative regulator of trichomes in *M. guttatus.* The finding that not just one, but all three of these hormones are differentially regulated in the progeny of damaged plants supports the hypothesis that this transgenerational response is mediated, at least in part, through the hormonal regulation of stress response pathways.

Biological Functions Associated With Altered Trichome Production

Trichome development requires cell wall remodeling to allow epidermal outgrowth (Hüʟsᴋᴀᴍᴘ *et al.* 1994; Aᴘᴘʟᴇǫᴜɪsᴛ *et al.* 2001). Therefore, differentially expressed genes that function in cell wall establishment or modification are likely to affect trichome density. We found the "cell wall" GOID to be significant enriched (2.09 fold) within the set of differentially expressed genes. Of 61 differentially expressed genes with the cell wall GOID,

45 were down-regulated in response to parental damage (Table 1.3). Many of these down-regulated genes appear to function in cellular shape modification, for example, xyloglucan endotransglucosylase hydolases (XEHs), and enzymes involved in reshaping cell wall components (EKLÖF and BRUMER 2010). The down-regulation of five XEHs, associated with increased trichome production, suggests that these XEHs may repress trichome formation. KEGG map analysis (KANEHISA *et al.* 2012) suggests that carbohydrate metabolism genes, differentially expressed in the progeny of damaged plants, likely alter cellulose, xyloglucan, and pectin biosynthesis. Plant cell walls are primarily composed of cellulose fibers embedded in a matrix of hemicellulose (xyloglucan) and pectin polysaccharides (COSGROVE 1997). The pattern of differential gene expression is consistent with highly modified production of all three of these cell wall compounds; however, due to the vast number of both up- and down-regulated genes, it is difficult to predict how the abundances of these cell wall components would change in the progeny of damaged plants. Altered cell wall carbohydrate composition could be related to increased trichome production or a distinct transgenerational response modifying the flexibility and/or strength of progeny plants. Down regulation of phenylcoumaran benzylic ether reductase (gene ID 1419276) and lactoperoxidase (MgLP1; gene ID 156083), which both function in lignin production, the most rigid cell wall component, provides evidence that altered cell wall content may affect plant physiology in more ways than just modifying trichome density (GANG *et al.* 1999) (Figure 1.5).

Biological Functions Associated with Secondary Metabolite Synthesis

While trichomes (and the cell wall in general) provide a mechanical barrier to insect herbivory, chemical deterrents provide additional defense (WINK 1988). The GOID associated with secondary metabolite synthesis pathways is not enriched in our whole set of differentially expressed genes, but is enriched 2.39 fold in cluster 1 of our differentially expressed genes (Table 1.3 and Figure 1.4). Using KEGG maps to guide a metabolic pathway analysis, we identify two genes coding for 4-coumarate-CoA-ligase (MgCou1 and MgCou2; gene IDs 1434953 and 1307066) as differentially expressed in the offspring of damaged plants, and their cumulative expression is higher in the progeny of damaged individuals (Figure 1.5). This gene produces Coumaroyl-CoA, the precursor of plant phenylpropanoids. Additionally, we found two of the pathways involved in the processing of Coumaroyl-CoA to be down-regulated, and one to be up-regulated in the progeny of damaged plants. The down-regulated pathways lead to the production of lignins (discussed above) and flavonols, while the continued processing of Coumaroyl-CoA into Anthocyanins, Flavanones, and Flavones appears to be up-regulated (Figure 1.5). Through the up-regulation of enzymes involved in phenylpropanoid production and the differential expression of specific down-stream pathways, the progeny of damaged plants may modify their allocation of resources to secondary metabolite production in a way that increases their fitness in environments similar to that experienced by their parents (TREUTTER 2005).

Differential Expression of Heat Shock Proteins

Genes coding for heat shock proteins (HSPs) are differentially expressed in response to parental damage. Twenty seven of 28 are significant up-regulated, and 1 of 28 is significantly down-regulated (Appendix 9). These enzymes cannot be linked to specific

pathways, but their function in protein mobility and folding is well understood (LINDQUIST and CRAIG 1988). HSPs have previously been found to be up-regulated in responses to many plant stresses (SWINDELL *et al.* 2007), but this is the first time in which this response has been shown transgenerationally in plants.

Four Clusters of Differentially Expressed Genes

Our clustering results provide evidence that the observed transgenerational plasticity in gene expression is due to transmission of a few signals derived from parental leaf damage that regulate the expression of gene clusters, and not unique transgenerational inheritance at each locus. The finding that each of the four clusters contains different types of genes (Figure 1.4, Table 1.3), as identified by GOID term distributions, supports the hypothesis that these gene clusters reflect distinct groups of genes regulated by a common signal. Individual plants appear capable of inheriting a wound response signal for some, but not all clusters of differentially expressed genes. For example, individual D2A (Figure 1.4) shows a strong gene expression damage response for clusters 1, a moderate response for cluster 2, but no response for clusters 3 and 4. It may be that certain parent plants only produce a subset of the transgenerational wound response signals. However, this appears unlikely, as even siblings often show different gene expression patterns. For example, individual D3a shows a clear gene expression damage response for cluster 2, while its sibling, D3b, does not exhibit this response. Thus, it appears that even selfed siblings from within a recombinant inbred line can inherit different transgenerational signals. Alternatively, a single transgenerational signal modifying gene expression in four distinct pathways with different dynamics could lead to the observed clustering results. Additional

27

analyses aimed at determining the transmitted signal (*e.g.,* altered DNA methylation states

and/or differentially transmitted small RNAs) will be necessary to distinguish between

these two possibilities. While there appear to be multiple corgeulatory gene groups

differentially regulated in response to parent environment, offspring may display all, or

only a subset, of the possible signals. What these signals are and why they are stochastically

expressed in the progeny generation remains unknown. It is possible that mosaic patterns

of differential expression correspond to mosaic epigenetic inheritance, as suggested by

(JABLONKA and RAZ 2009).

Identifying the Epigenetic Signal Transmitted to Progeny

Recent findings suggest that differential DNA methylation and small RNA transcription of

transposable elements and coding genes can be directed by environmental conditions and

can lead to differential gene expression in *cis* (ITO *et al.* 2011). Additionally, histone

acetylation and methylation at specific loci can be transgenerationally responsive to

parental stress (LANG-MLADEK *et al.* 2010). It is unclear how DNA methylation, histone

modifications, and small RNAs propagate environmental signals transgenerationally, but it

is clear that these three epigenetic mechanisms interact to perpetuate these signals (MCCUE

et al. 2012; LAW *et al.* 2013b). In *Arabidopsis*, the genes encoding numerous chromatin-

remodeling proteins (HKMTs) have been found to be hypermethylated, and in turn silenced

in the offspring of salt stressed individuals, leading to DNA hypomethylation elsewhere in

the genome (BILICHAK *et al.* 2012). One possibility is that histone modifications in response

to stress lead to the production of locus-specific small RNAs that invade the germline and

reprogram stress induced histone and/or DNA methylation (MOLNAR *et al.* 2010; CALARCO *et al.* 2012).

Our findings support the hypothesis that wound response signals produced in parent plants can affect the physiology and development of offspring via altered gene regulation in the absence of inductive signals experienced by offspring. If these parent-derived signals are epigenetic in nature, they should be detectable as differential small RNA profiles and/or DNA methylation patterns in progeny of damaged compared to control plants. The finding that numerous plant hormone pathways show differential expression in our progeny plants presents an alternative hypothesis; hormone loading into seeds may cause gene expression and phenotypic plasticity in progeny. Determining whether the signal transmitted from parent to offspring is hormonal, epigenetic, or belongs to an as yet unidentified mechanism is a vital next step for advancing our understanding of transgenerational phenotypic plasticity.

Conclusions

This study illustrates how RNA-seq data can be used to both identify candidate genes responsible for phenotypes of interest, as well as other genes, developmental pathways and phenotypes responsive to a stimulus of interest. We confirm results from a previous candidate gene study (Scoville *et al.* 2011) – *MgMYBML8* is down-regulated in response to parental leaf damage. However, this genome-wide survey finds differential expression of over 900 additional genes and it is yet to be determined the extent to which *MgMYBML8* regulates, or is regulated by, genes in this larger set. Many genes in this larger set appear to be co-regulated and involved in similar metabolic pathways; often in ways that allow us to

generate testable hypotheses of how evolutionarily significant phenotypes are regulated by parental leaf damage. Future studies will aim to determine whether these phenotypes are responsive to parental leaf damage in the way predicted by this transcriptome analysis. Annual plants may be incapable of transmitting information to their progeny through cis-temporal mechanisms (such as language, or airborne/water-soluble chemical cues)(Dudley & File, 2007), but it appears that they have evolved a mechanism(s) for disseminating environmental information from one generation to the next. The diversity of responsive genes and the scale of differential expression in response to a single environmental variable is exciting, but future studies in other species, in the field, and in response to other environmental conditions will be necessary to gauge the impact that this system of inheritance has on gene regulation and evolution (RICHARDS *et al.* 2012). Finally, the signals which cause transgenerational environmentally induced differential gene expression still must be identified. While much work has been done showing the importance of various epigenetic factors in transgenerational inheritance, a study linking parental epigenetic modification, offspring epigenome, offspring gene expression, and altered offspring phenotype and fitness in the same system will be necessary to produce a more complete model of transgenerational phenotypic plasticity.

Table 1.1. Expression patterns and calculated p-values of the top fifteen up-regulated, and down-regulated annotated transcripts in response to parental damage.

Gene Description	Damaged parent[1]	Control parent[1]	log$_2$ FC	EdgeR	DESeq2
Up-regulated transcripts					
dnaj protein	2,881	1,292	1.10	4.48E-09	7.88E-11
polygalacturonase	217	36	2.47	5.22E-14	3.25E-10
heat stress transcription factor 6b	602	123	2.35	5.01E-17	6.40E-10
heat shock protein	1,840	491	2.02	2.33E-10	2.89E-09
bri1 kinase inhibitor 1	419	174	1.21	4.19E-10	6.57E-09
p-hydroxybenzoic acid efflux pump subunit aaeb	152	21	2.80	3.52E-18	3.53E-08
glycerol-3-phosphate transporter	1,978	419	2.11	2.49E-14	1.17E-07
hydroxymethylglutaryl-lyase	1,041	293	1.83	2.79E-13	1.17E-07
udp-glycosyltransferase	732	403	0.79	4.83E-07	4.81E-07
outer arm dynein light chain 1	428	183	1.21	9.14E-06	7.47E-07
dnaj subfamily b	298	143	0.95	4.47E-09	2.29E-06
l-type lectin-domain containing receptor kinase	2,053	987	0.99	6.48E-13	2.34E-06
chloroplast movement under blue light 1	955	518	0.75	3.78E-08	3.06E-06
acyl- n-acyltransferases	280	158	0.73	1.28E-05	4.39E-06
heat shock protein 70	34,114	21,777	0.55	3.57E-07	5.85E-06
Down-regulated transcripts					
pollen ole-e-1	2,165	2,684	-0.43	1.55E-10	4.85E-07
protochlorophyllide chloroplastic	9,099	21,057	-1.43	1.20E-09	2.68E-06

cold acclimation protein cor413-pm1	2,989	5,200	-0.96	4.66E-09	8.10E-06
metal-nicotianamine transporter ysl7	564	848	-0.71	1.31E-08	1.58E-05
sodium calcium exchanger protein	3,592	5,063	-0.64	2.60E-08	2.49E-05
phosphoenolpyruvate carboxykinase	8	55	-2.71	6.61E-08	5.58E-05
chaperone protein dnaj 20	170	697	-2.28	2.71E-07	0.00016
ralf 34	573	854	-0.68	2.80E-07	0.00016
cytosolic sulfotransferase 12	868	1,453	-0.98	3.12E-07	0.00018
abc transporter c 5	786	1,428	-0.97	3.90E-07	0.00022
remorin protein	1,517	2,312	-0.74	7.94E-07	0.00039
early nodulin 1	295	415	-0.67	9.55E-07	0.00045
protein kinase chloroplastic	390	495	-0.46	1.34E-06	0.00058
lrr receptor serine threonine-protein kinase	451	664	-0.70	1.53E-06	0.00064
stem-specific protein tsjt1	1,157	1,867	-0.88	1.75E-06	0.00071

Table 1.2. Subset of GO terms found at significantly elevated frequencies in the set of differentially expressed (DE) genes compared to transcriptome-wide frequencies. See Appendix 8 for the full set of 31 enriched GO terms.

GO term	% transcriptome with term	% DE genes with term
cell periphery	14.3	**22.7***
membrane	25.7	**35.4***
response to stimulus	23.2	**32.3***
response to abiotic stimulus	9.6	**15.9***
cell wall	3.2	**6.7***
response to stress	15.1	**21.5***

***$p < 0.005$

Table 1.3. GO terms found at significantly elevated frequencies in the set of significantly up-regulated DE genes, the set of significantly down-regulated DE genes, and in each of the gene clusters (Figure 1.4). Each test for significant enrichment of a GO term is for the specific gene group compared to all DE genes not in that group.

GO term	% up-regulated genes with term	% down-regulated genes with term	% cluster 1 genes with term	% cluster 2 genes with term	% cluster 3 genes with term	% cluster 4 genes with term
cell periphery	18.6	**26.5***	18.1	25.4	18.2	**34.2***
membrane	33.3	36.3	33.0	**38.8**	28.6	42.5
response to stimulus	31.6	31.7	31.2	32.1	**46.8***	34.2
response to abiotic stimulus	**18.6***	12.8	**17.2***	12.5	**32.5***	12.3
cell wall	3.5	**10.2***	5.1	6.3	6.5	**16.4***
response to stress	22.6	19.5	20.0	18.3	**40.3***	23.3
carbohydrate metabolic process	7.0	**15.4**	14.0	7.6	10.4	16.4
secondary metabolic process	4.9	5.2	**9.8***	2.2	3.9	2.7

*p <0.05, **0.005 < p < 0.05, ***p < 0.005

Figure 1.1. A model for environmentally induced transgenerational phenotypic plasticity in *M. guttatus*. a) Leaf clipping to parent plants simulates mechanical damage effects of insect herbivory. b) Signals resulting from parent environment (damage/no damage) are transmitted to progeny. c) Progeny exhibit phenotypic plasticity in trichome production in response to parent environment (damage/no damage). d) Transgenerationally inherited signals are translated into plastic phenotypes through differential gene expression in progeny. *M. guttatus* image available under GNU free documentation license 1.3. Modified from Flora von Deutschland (THOMÉ 1885).
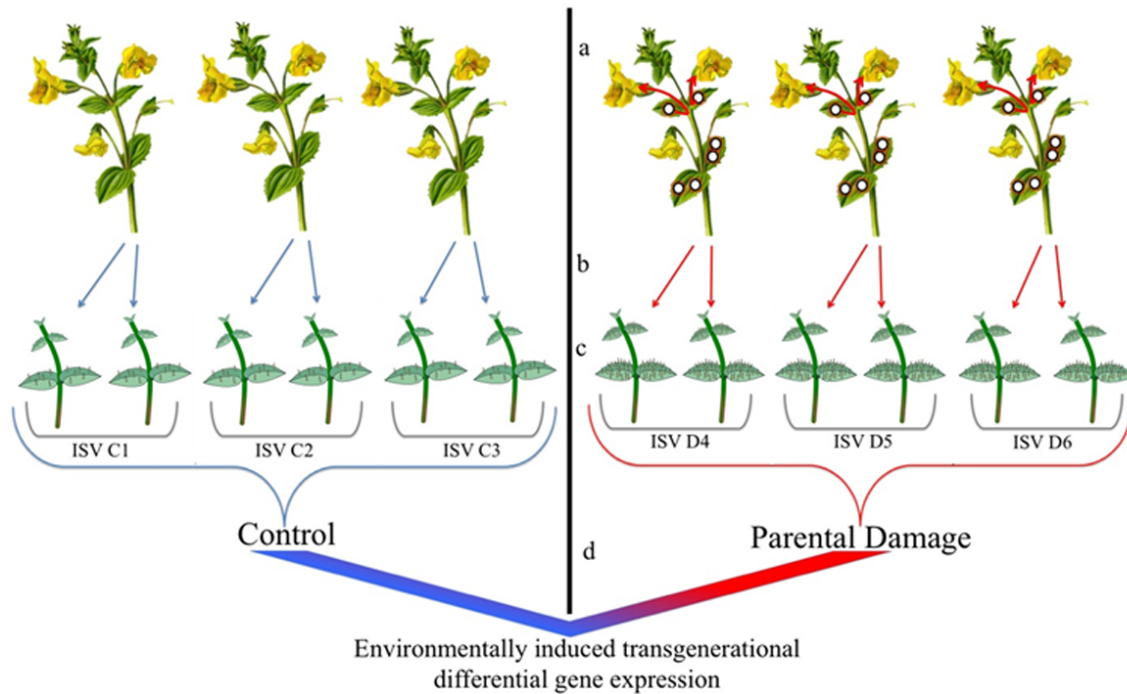
Figure 1.2. Mean and standard deviation (in parentheses) for additive and total genetic variance calculated with and without bias-correction (above and below the line, respectively), as well as with and without epistasis for both of the allele frequency distributions.
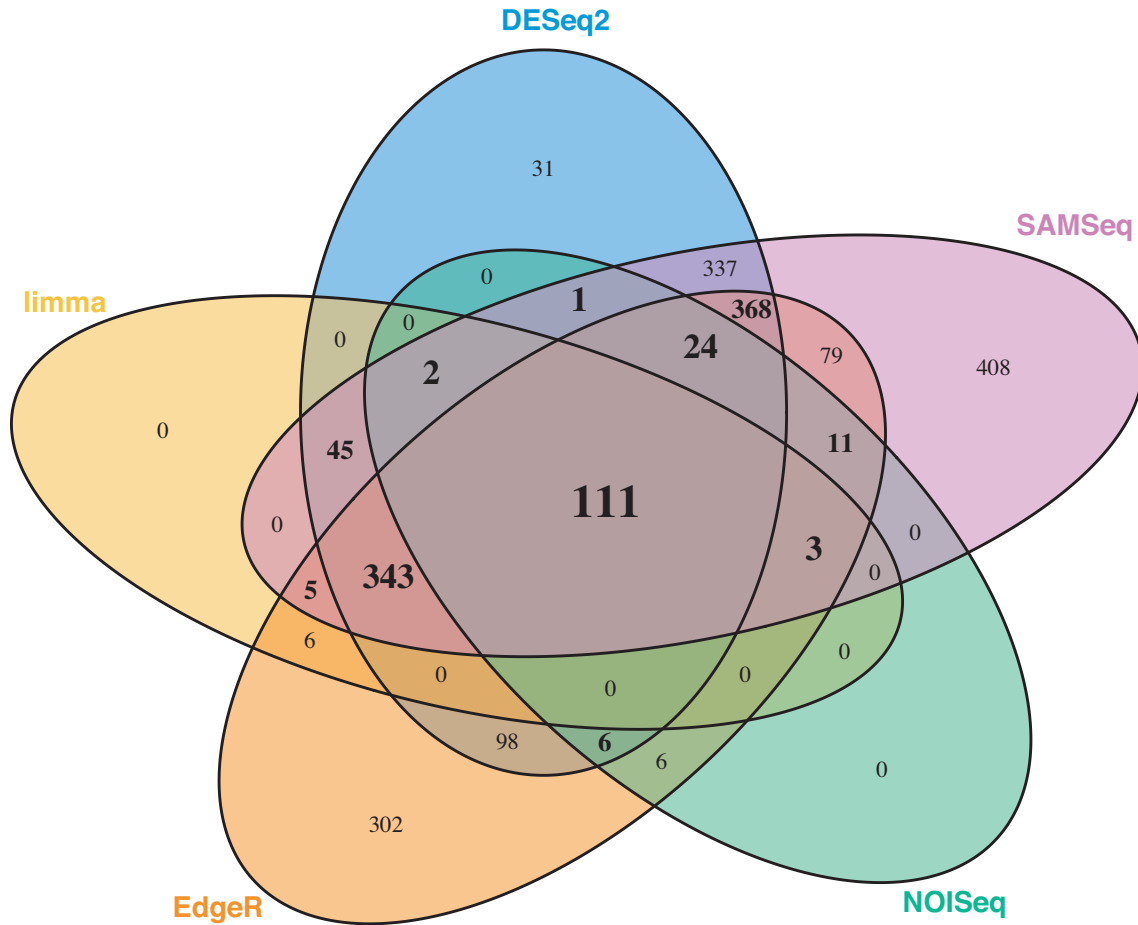
Figure 1.3. Heat map of 919 differentially expressed genes by treatment category. Individuals are coded as C or D = control group or leaf damage group, 1, 2 or 3 = family line, and a or b = one of two siblings included per family line. Differentially expressed genes identified to GOIDs cell membrane, abiotic stimulus, secondary metabolic process, and cell wall are indicated by color codes along the left-side of the heat map.
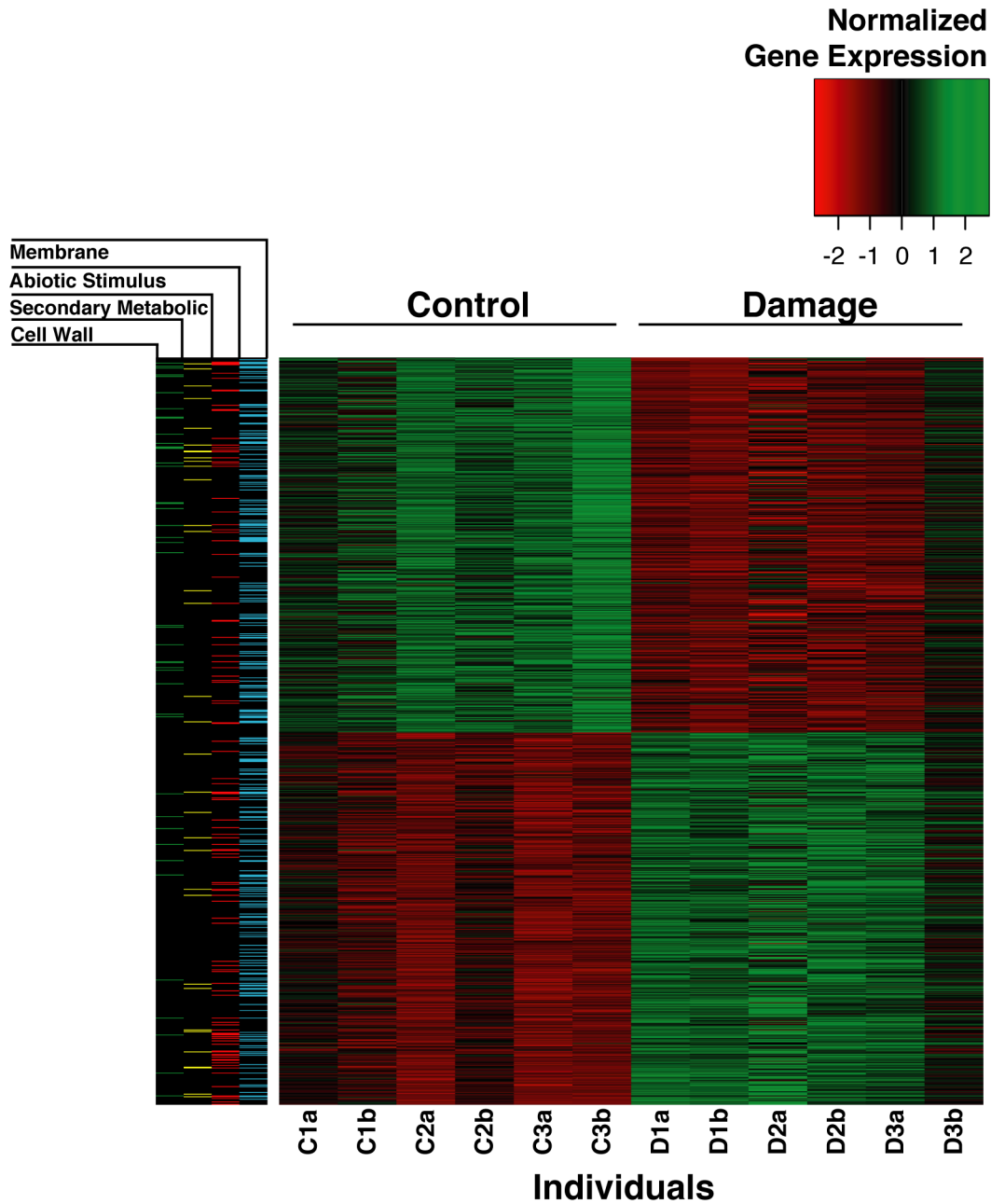
Figure 1.4. Heat maps of differentially expressed genes by cluster. Color keys for both expression level (red to green) and GOID (green, yellow, red, blue in left hand columns) are as in Figure 1.3. Asterisks indicate GOIDs significantly enriched in a given cluster relative to the full set of differentially expressed genes. Both secondary metabolic process and abiotic stimulus GOIDs are enriched in the set of cluster 1 genes, cell membrane GOID is enriched in the set of cluster 2 genes, abiotic stimulus GOID is enriched in the set of cluster 3 genes, and cell wall GOID is enriched in the set of cluster 4 genes.
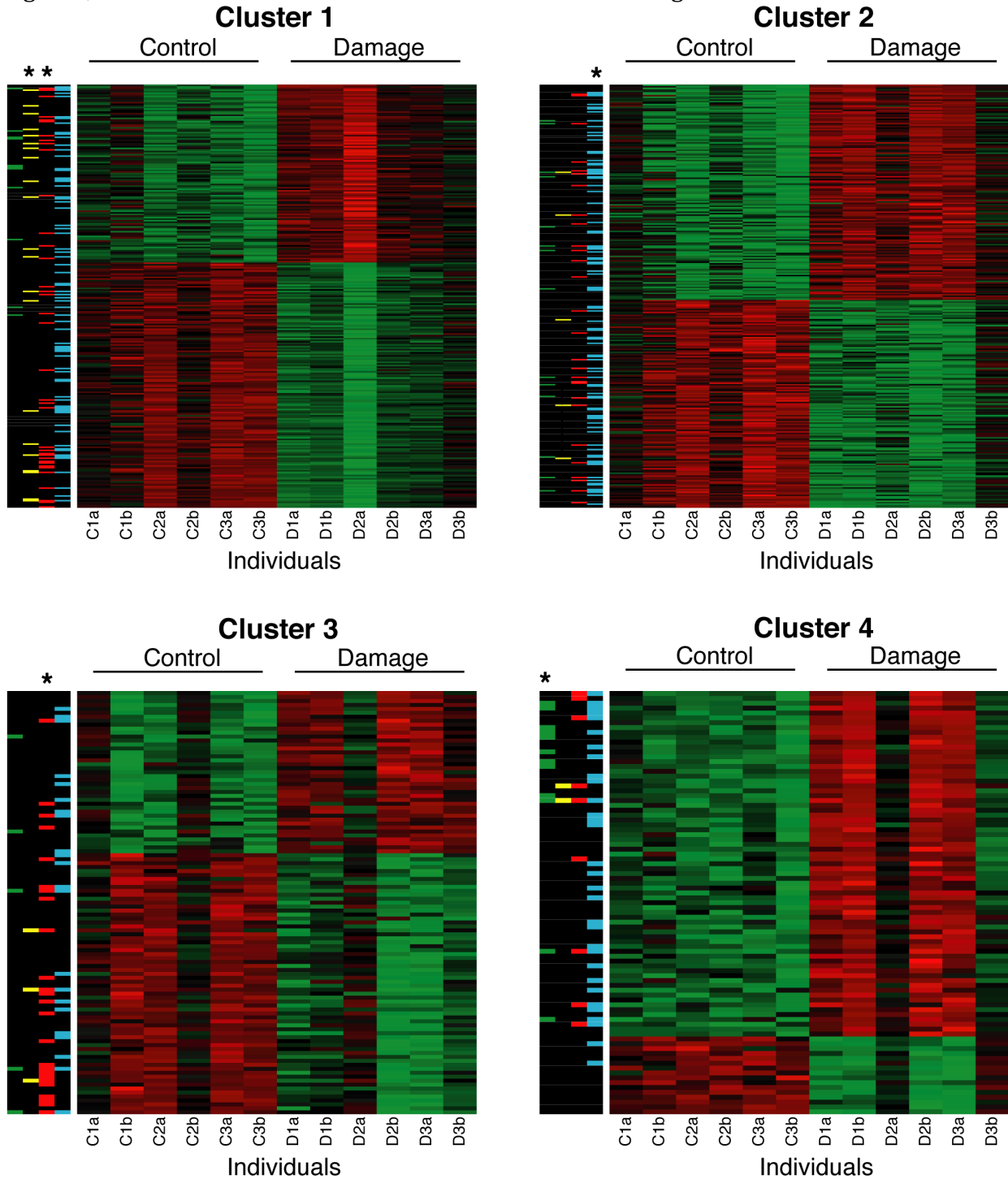
Figure 1.5. Flux through the phenylpropanoid biosynthetic pathway derived from RNA-seq gene expression data in the context of a modified KEGG map. Box plots show relative gene expression in progeny of damaged (blue) and control (red) plants. The following genes are included in the map: MgGB1 and 2 (gentiobiase; Gene IDs 1317357 and 1351622), MgCou1 and 2 (4-coumarate ligase; Gene IDs1307066 and 1434953), MgFLS1 (flavanol synthase; Gene ID 1400552), Mg3'5'H1 (3'5'-hydroxylase; Gene ID 1338195), MgCR1 and 2 (cinnamoyl-CoA reductase; Gene IDs 1399905 and 1315821), MgLP1 (lactoperoxidase; Gene ID 156083), and Mg5MAT1 (5-O-glucoside 6'-O-malonyltransferase; Gene ID 1355759). Green arrows indicate inferred increased flux through that step in the pathway in response to parent leaf damage, red arrows indicate inferred decrease flux through that step of the pathway in response to parent leaf damage, and black arrows indicate no inferred change in flux through that step of the pathway as a result of parent environment.
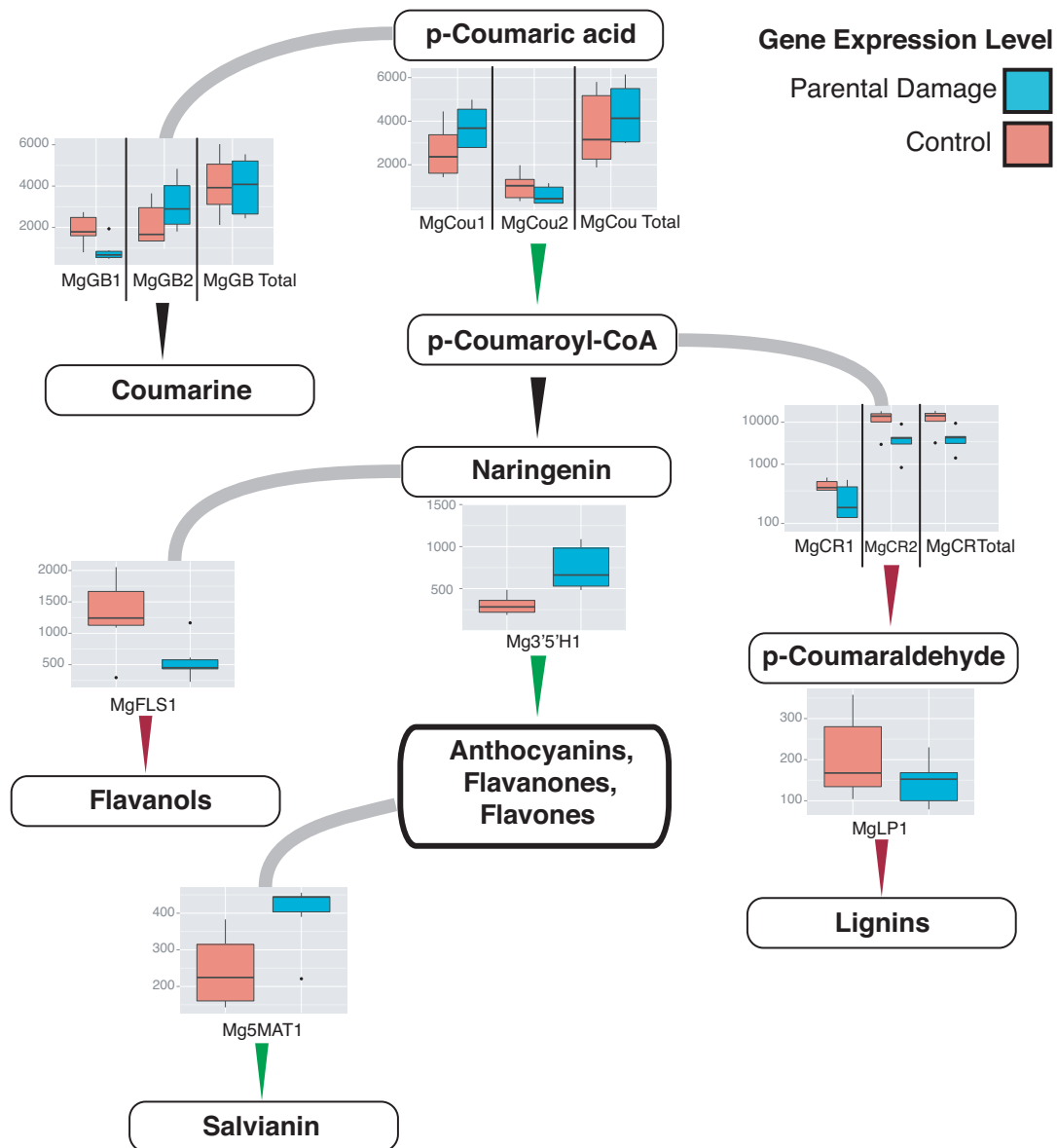
CHAPTER 2*

DNA Methylation and gene expression in *Mimulus guttatus*

* Colicchio, J. M., F. Miura, J. K. Kelly, T. Ito and L. C. Hileman, 2015a DNA methylation and gene expression in Mimulus guttatus. BMC genomics **16:** 507.

Abstract

Background: The presence of methyl groups on cytosine nucleotides across an organism's genome (methylation) is a major regulator of genome stability, crossing over, and gene regulation. The capacity for DNA methylation to be altered by environmental conditions, and potentially passed between generations, makes it a prime candidate for transgenerational epigenetic inheritance. Here we conduct the first analysis of the *Mimulus guttatus* methylome, with a focus on the relationship between DNA methylation and gene expression.

Results: We present a whole genome methylome for the inbred line Iron Mountain 62 (IM62). DNA methylation varies across chromosomes, genomic regions, and genes. We develop a model that predicts gene expression based on DNA methylation ($R^2=0.2$). *Post hoc* analysis of this model confirms prior relationships, and identifies novel relationships between methylation and gene expression. Additionally, we find that DNA methylation is significantly depleted near gene transcriptional start sites, which may explain the recently discovered elevated rate of recombination in these same regions.

Conclusions: The establishment here of a reference methylome will be a useful resource for the continued advancement of *M. guttatus* as a model system. Using a model-based approach, we demonstrate that methylation patterns are an important predictor of variation in gene expression. This model provides a novel approach for differential methylation analysis that generates distinct and testable hypotheses regarding gene expression.

Introduction

DNA cytosine methylation is an epigenetic modification that acts in conjunction with histone modification and small RNAs to regulate gene expression (TATE and BIRD 1993; FLAVELL 1994; ZILBERMAN *et al.* 2006) and control transposable elements (MIURA *et al.* 2001; LIPPMAN *et al.* 2004). In addition, DNA methylation appears to alter mutation rates (XIA *et al.* 2012) and to decrease rates of recombination (MIROUZE *et al.* 2012). It is found in organisms spanning the eukaryotic phylogeny (FENG *et al.* 2010; HUFF and ZILBERMAN 2014), and can occur in many sequence contexts. In plants, cytosine methylation can be found in CG, CHG, or CHH contexts, where H is any nucleotide besides G (GRUENBAUM *et al.* 1981). It appears that much of the methylome is stable within an individual; however, the methylome does exhibit predictable plastic responses to developmental and environmental cues (KINOSHITA and JACOBSEN 2012; BOND and BAULCOMBE 2014a).

Recent work has greatly expanded our knowledge of the mechanisms involved in maintaining and modifying DNA methylation in plants (LEONHARDT *et al.* 1992; LINDROTH *et al.* 2001; CAO and JACOBSEN 2002; LAW and JACOBSEN 2010; LAW *et al.* 2013b; BOND and BAULCOMBE 2014b; EICHTEN *et al.* 2014), yet we still do not fully understand how specific patterns of DNA methylation in and near coding sequences control gene expression. In *Arabidopsis thaliana*, CG DNA methylation in regulatory sequences is negatively correlated with gene expression (ZHANG *et al.* 2006; ZILBERMAN *et al.* 2006), possibly through limiting promoter accessibility. Contrastingly, gene body CG methylation is elevated in moderate to highly expressed genes (GRUENBAUM *et al.* 1981; ZILBERMAN *et al.* 2006; LI *et al.* 2012), potentially though the removal of histone variant H2A.Z (COLEMAN-DERR and ZILBERMAN 2012). Similar patterns of association between the distribution of plant CG methylation and

gene expression have been found in the wild rice (Lɪ *et al.* 2012), tomato (Zʜᴏɴɢ *et al.* 2013), and maize (Eɪᴄʜᴛᴇɴ *et al.* 2013a). Additionally, *Arabidopsis* genes within differentially methylated regions tended to be more highly expressed in individuals with increased CG methylation, but lower in individuals with increased non-CG (CHG and CHH) methylation (Sᴄʜᴍɪᴛᴢ *et al.* 2013c). However, the interaction between gene expression and different forms of DNA methylation in and around genes has not been fully explored. For example, the impact of non-CG methylation on gene expression is especially understudied, despite its established role in regulating transposable elements through pre- and post-transcriptional silencing (Sᴀᴢᴇ *et al.* 2012).

The standard method for characterizing genomic patterns of DNA methylation is to classify genes into methylation quantiles and then compare gene expression across these groups (Zɪʟʙᴇʀᴍᴀɴ *et al.* 2006; Lɪ *et al.* 2012; Gᴇɴᴛ *et al.* 2013; Tᴀᴋᴜɴᴏ and Gᴀᴜᴛ 2013; Zʜᴏɴɢ *et al.* 2013; Lɪ *et al.* 2014; Wᴀɴɢ *et al.* 2014). Here, we adopt an explicit model-based approach, predicting gene expression from gene methylation and other basic gene-specific features (exon length, intron length, and exon number). We compare the methylome of an inbred line, to gene expression from a distinct recombinant inbred line, and test how well DNA methylation, in combination with other stable genetic factors, predict gene expression across lines and tissue types. The explanatory power of stable epigenetic variation on gene expression is relatively unknown (although see (Yᴜᴀɴ *et al.* 2007) for model-based approaches to predicting gene expression via promoter motifs in *Saccharomyces cerevisiae,* and (Lɪ *et al.* 2008) for a Sanger sequencing approach to gene expression modeling based on histone and DNA methylation in rice). With the model-based approach presented here,

we are able to assess the scale to which constitutive epigenetic variation effects global gene expression, and the patterns of DNA methylation through which this regulation is manifest.

Previous studies of *Mimulus guttatus* have demonstrated transgenerational epigenetic inheritance (HOLESKI 2007; HOLESKI *et al.* 2010; SCOVILLE *et al.* 2011; COLICCHIO *et al.* 2014). Herbivore induced defensive traits can be transmitted between generations, and the observed transcriptional basis of this response [11], has made it a promising model system in the burgeoning field of ecological epigenetics (HOLESKI *et al.* 2012; HOLESKI *et al.* 2013b; LATZEL *et al.* 2013; KILVITIS *et al.* 2014). However, along with identifying transmissible epigenetic marks, it is vital to understand the role that stable epigenetic regulation has on gene expression. Here we present the first *M. guttatus* methylome. We utilize a novel modeling approach to untangle the complex interactions between methylation and gene expression. We show that non-CG gene body methylation may have a significant effect on gene expression despite occurring at relatively low levels. Utilizing a GO term enrichment approach, we demonstrate that certain functional categories are over-represented in genes with high gene body CG methylation. We provide evidence that there are differences in methylation and gene expression between chromosomes, such that mean gene expression is significantly lower across some chromosomes than others. Finally, we look at transcriptional start sites across the genome, where recent evidence suggests increased recombination in *M. guttatus* (HELLSTEN *et al.* 2013), and find a corresponding decrease in DNA methylation.

Materials and Methods

DNA Extraction and Bisulfite Sequencing

We germinated seeds from the *M. guttatus* Iron Mountain inbred line, IM62, the line that was sequenced to establish the *M. guttatus* reference genome (HELLSTEN *et al.* 2013)(http://phytozome.jgi.doe.gov). When the second leaf pair of seedlings was just visible we collected leaf tissue from multiple seedlings, flash froze it in liquid nitrogen, and stored it at -80° C. We performed DNA extractions using a CTAB protocol (HOLESKI *et al.* 2013a). We pooled DNA from multiple seedlings before library construction in order to limit the effects of aberrant intra-individual variation(HARDCASTLE 2013). From this pooled sample we generated sequencing template for whole genome bisulfite sequencing (WGBS) following the PBAT (Post-Bisulfite Adaptor Tagging) protocol (MIURA *et al.* 2012). With 1 ng of unmethylated lambda DNA obtained from Promega used as a spike-in control for conversion efficiency, 100 ng of genomic DNA from *M. guttatus* was treated with bisulfite using EZ DNA Methylation kit from Zymo Research. Two rounds of random primer extension for tagging bisulfite treated DNA with adaptors were performed using primers for single-end library construction as described in [42]. The concentration of templates was determined by qPCR with Library Quantification Kits from KAPA biosystems. A single lane of 100 cycle reactions on HiSeq 2500 was assigned for the library sequencing.

Read Mapping

We used the software BMap (MIURA *et al.* 2012) (http://itolab.med.kyushu-u.ac.jp/BMap/index.html) to map bisulifte treated reads to the *M. guttatus* v2.0 reference genome (http://phytozome.jgi.doe.gov). In short, BMap first searches candidate genomic

loci for each read in two duplicated genome sequences, one with every C in the genome converted to a T (C2T), and one with G to A (G2A), using an approach called adaptive seed (KIEŁBASA *et al.* 2011). Next BMap creates pairwise alignments between the read and original DNA sequence of every candidate loci, and calculates scores for each alignment allowing mismatches between T in the reads with C in the reference. Finally an alignment with the highest score is reported for each read. We used default parameters for mapping with BMap. Using alignments exported by BMap, methylation status for every cytosine in every read was called, and counts both supporting the methylated and unmethylated state are assigned for every cytosine residue of the reference genome. Methylation levels for CG, CHG and CHH contexts are exported to different files and analyzed independently.

Global Methylome Analysis

We estimated the number of total and methylated cytosines mapped across the genome on a per-nucleotide basis for the *M. guttatus* IM62 seedling methylome. Percent methylation was calculated for each 1 kb window across the genome for total methylation, as well as methylation in each of the three sequence contexts. Centromere positions were estimated from characteristic repeat sequences (FLAGEL *et al.* 2014).

Gene Methylation Analysis

Using the *M. guttatus* v2.0 annotations (GOODSTEIN *et al.* 2012), we calculated the percent methylation in each sequence context for each of the 24,130 annotated genes. Only the 17,043 for which we had gene expression data (COLICCHIO *et al.* 2014) were used for down-stream analysis. For each annotated gene we defined three regions: up-stream as the

1kb up-stream of the transcriptional start site, gene body as the transcribed portion of the gene, and down-stream as the 1kb downstream of the 3' UTR. Gene expression values were generated previously by RNAseq from seedling tissue of genetically distinct *M. guttatus* – a recombinant inbred line derived from cross between divergent populations (COLICCHIO *et al.* 2014).

In order to determine if gene methylation and expression varied across chromosomes we performed four ANOVAs with chromosome as an explanatory variable and CG, CHG, CHH, and log-transformed gene expression as response variables.

Gene ontology terms were already assigned to genes (COLICCHIO *et al.* 2014), and were utilized both to calculate the total number of GO terms per gene, as well as to perform a Fisher's Exact test to determine what, if any, types of genes were enriched or depleted in our set of highly CG methylated genes, and our set of chromosomes exhibiting significantly reduced gene expression levels.

In order to choose a predictive gene expression model, we included methylation in each of three contexts, percent methylation in gene bodies, up-stream and down-stream regions, intron length (sum of all introns for a gene), exon length (sum of all exons for a gene), number of exons, and interaction terms up to the third degree. Gene length, intron size, and intron number are all known to be positively correlated with gene expression in plants (REN *et al.* 2006), opposite the trend observed in animals (CASTILLO-DAVIS *et al.* 2002). We used a Bayesian information criterion (BIC) (POSADA and BUCKLEY 2004) to inform our restricted maximum likelihood (REML) model selection (done in order to limit the number of parameters included in our model, and in turn reduce over fitting). Additionally, genes were parsed randomly into thirds, and parameters were tunes for each

of these three groups independently. These models were then used to predict gene expression in the remaining to gene groups to provide 3-fold cross-validation (KOHAVI 1995). We Z-transformed values to make parameter estimates comparable, making a value of 0 represent the mean value for a variable, with positive or negative deviations reflecting the number of standard deviations a value is from the mean.

We identified transposable elements across the *M. guttatus* genome from the repeat-masked genome assembly (GOODSTEIN *et al.* 2012). Genomic repeats larger than 100 base pairs were selected and percent methylation in all three sequence contexts was identified for these repeats.

Results and Discussion

Global Methylation

Of the 186 million reads generated, 126 million were mapped to the genome (67.7% mapping, mean read depth = 19, median = 6). This proportion is typical for *Mimulus* genomic studies (eg. KELLY *et al.* 2013) given the substantial proportion of the physical genome that is not contained in the v2 reference genome. Mapping to unmethylated lambda DNA confirmed that our PBAT treatment achieved 99.4% conversion of unmethylated cytosines to thymine. Methylation is most common in a CG context (72%), intermediate in a CHG context (36.5%), and lowest in a CHH context (6.1%)(Figure 2.1), with 23% of total cytosine's being methylated. The percent of genome methylation found in *M. guttatus* is higher in all contexts than *Oryza sativa* (LI *et al.* 2012), *Arabiopsis thaliana* (FENG *et al.* 2010), *Brachypodium distachyiom* (TAKUNO and GAUT 2013), lower in all contexts than *Solanum lycopersicum* (ZHONG *et al.* 2013), and both higher or lower than *Zea mays*

(GENT *et al.* 2013) and *Glycine max* (SCHMITZ *et al.* 2013a) depending on context (Figure 2.1). While CHH methylation levels are higher in *M. guttatus* than *Z. mays* and *G. max*, the opposite is true for CHG methylation. CG methylation is highest in *Z. Mays*, moderate in *M. guttatus*, and lowest in *G. max* (Figure 2.1).

Approximate positions of centromeres on *M. guttatus* chromosomes are given by the location and density of centromeric repeats (FLAGEL *et al.* 2014). We confirmed that regions of the genome with high levels of centromeric repeats also tended to have high CG, CHG, and CHH methylation (Figure 2.2). We found that gene expression and gene body CG, CHG, CHH methylation varied significantly across chromosomes (log(expression): $F_{13,17042} = 4.43$, CG: $F_{13,17042} = 10.85$, CHG: $F_{13,17042} = 19.07$, CHH: $F_{13,170423} = 6.10$, p<0.001)). Chromosomes that have on average higher levels of methylation tended to also have lower gene expression (Figure 2.3). From this result, it is unclear whether certain chromosomes are constitutively more highly methylated and transcriptionally silenced, or whether throughout development epigenetic modification at a whole chromosome scale can change the relative expression of genes across entire chromosomes. It does appear that silenced chromosomes have a higher density of heterochromatic repeats, hinting that certain chromosomes may be condensed throughout development.

Gene Methylation

Methylation was significantly depleted in gene bodies relative to both inter-genic regions and transposable elements in all three-sequence contexts (Table 2.1). While CG methylation was only modestly reduced in gene bodies relative to intergenic regions (Gene Bodies: 56%, Intergenic: 75%), CHG (Gene Bodies: 3.8%, Intergenic: 45%) and CHH (Gene

Bodies: 1.2%, Intergenic: 7.2%) methylation levels were drastically reduced (Table 2.1). Similar results were found in *Oryza sativa* (Li *et al.* 2012), *Arabidopsis thaliana* (Shen *et al.* 2012), and *Glycine max* (Schmitz *et al.* 2013a). Methylation both up-stream and down-stream of gene starts was also reduced relative to genome-wide averages. We found that up-stream regions were elevated in non-CG methylation compared to gene bodies, but that up-stream CG methylation was reduced compared to gene body CG methylation (Table 2.1).

The methylation levels in all contexts (CG, CHG, CHH) and genic positions (up-stream, down-stream, and gene body) at a given gene were significantly correlated with one another (Figure 2.4). These were positive correlations for all cases but two. The two exceptions were negative correlations between up- and down-stream CHH methylation with gene body CG methylation. The most significant positive correlations were found between CHG and CHH or CG methylation levels at both up-stream and down-stream regions, as well as between CHG and CHH gene body methylation. Interestingly, the methylation levels for all three contexts vary greatly across the three gene regions in a fairly unpredictable manner. For instance, correlation between up-stream CG methylation and gene body CG methylation is only $r = 0.14$. This highlights the disparate functions of regulatory region methylation with that of gene body methylation (Jones 2012). The extremely high correlations between CHG and CHH methylation (Figure 2.4, $r > 0.67$) in all three regions is likely due to the involvement of similar enzymatic machinery in the propagation of both types of non-CG methylation (Bartee *et al.* 2001).

Methylation Effect on Gene Expression

A stepwise cubic polynomial model was selected to predict log(gene expression) based on minimum BIC. Out of a possible 454 parameters, the minimum BIC criterion selected a model with 29 factors that explained (R^2) 20.1% of the variation in log transformed expression values (*SS* Model: 1764, SS Error: 6981, $F_{28,17042}$ =153.6, p<0.0001, Tables 2, 3 and 4, Figure 2.5, Appendix 10). Including all 454 parameters increases R^2 only marginally (to 23.3%), and the minimum calculated R^2 calculated in 3-fold cross-validation was 17.9%. Generally, there is an excess of genes predicted to be expressed at log-transformed values between 1.5 and 2.5, that were actually expressed at levels less than 1.2, as well as genes expressed above 4, which this model never predicts (Appendix 10). It is clear that while gene methylation can modify gene expression, it cannot predict the complete repression, or extremely high expression of some genes. As all parameters were Z-transformed prior to modeling, the effect estimates are comparable across variables (Table 2.4). In order to maintain both statistical and molecular consistency throughout, both Z-transformed values and raw values are reported. The inclusion of both various forms of DNA methylation and gene architecture (number of exons, exon length, intron length) have not been included in a single model explicitly testing their ability to predict gene expression, but their independent effects have often been looked at in relation to gene expression. While it is hard to compare our integrative analysis on gene expression with prior studies, we generally find the same direction of effect in our data as was found in other plant systems (ZILBERMAN *et al.* 2006). Trends are thus not *Mimulus* specific, but likely more general effects of DNA methylation on gene expression in angiosperms. Finally, when discussing the role of various forms of methylation on gene expression we often designate a specific type of methylation as having a positive or negative effect on gene expression. In

this context that indicates that there was significant predictive ability for a given type of methylation on gene expression. However, due to the nature of this experimental design we cannot definitively define the arrow of causation.

Gene Body CG Methylation

Linear Effects: $\log(GE) = 2.61 - 0.07 m_{cg} = f^1$, where m_{cg} is gene body CG methylation. Controlling for gene architecture and other forms of methylation, we observe a negative linear effect of gene body CG methylation on gene expression (Figure 2.5 and 6a. black line). The effect size of gene body CG methylation (m_{cg}) is -0.07 (Table 2.3); a gene with $m_{cg} = -1$ (32%) is predicted to have 35% higher expression than one with $m_{cg} = 1$ (80%) (Figure 2.6a, black line). Previous studies report that gene body CG methylation is positively correlated with gene expression (GRUENBAUM *et al.* 1981; ZHANG *et al.* 2006; ZILBERMAN *et al.* 2006; LI *et al.* 2012). While the linear component of the model seems to contradict these previous reports, it cannot be interpreted in isolation. The polynomial and interaction terms indicate that gene body methylation has neither universally positive nor negative effects on gene expression. Traditional methods that looked for associations between gene expression and gene body CG methylation (which find a positive correlation between the two), and modeling methods as applied here followed by only analysis of the simple linear terms (which finds a negative correlation) come up quite short in portraying the role of gene body CG methylation in transcriptional regulation. By considering non-linear effects of methylation on gene expression we can begin to increase our understanding of the role of gene body CG methylation in gene regulation.

Quadratic Effects: $f^1 - 0.1m_{cg}^2 = f^2$. The squared gene body CG methylation term

has the second largest effect size of any methylation term (after gene body CHG

methylation) on gene expression, and leads to a predicted local m_{cg} maximum for gene

expression (due to it being a negative parabola, Figure 2.6a, green line). This maximum is

found at $m_{cg} = -0.35$ (47%). As gene methylation increases or decreases relative to a

moderate 45% methylation, gene expression is expected to decrease (Figure 2.6a; green

line).

Cubic Effects: $f^2 - 0.03m_{cg}^3 = f^3$. The cubed gene body CG methylation term is also

negative; compared to our quadratic model, this leads to higher predicted gene expression

for genes with lower than average methylation, and lower for genes with higher than

average methylation. This slightly lowers the predicted local maximum of gene expression

to $m_{cg} = -0.43(45\%)$ (Figure 2.6a, blue line). These data agree with previous findings that

there is a non-linear relationship between gene body CG methylation and gene expression

with an intermediate optimum (ZILBERMAN *et al.* 2006).

Interaction Terms

Negative Promoter CG Methylation Interaction: $f^3 - .02m_{cg}u_{cg} = f^4$. The effect of

interaction terms in this model is best understood by comparing expected gene expression

across m_{cg} values for a variety of interaction term values. Changes in linear interaction

term values (in this case up-stream cg methylation u_{cg}), lead to a change in our linear m_{cg}

coefficient. For example, at $u_{cg} = 1$ (82%), $0.2m_{cg}$ is subtracted from our earlier model,

we are left with:

$$\log(GE) = 2.61 - .\boldsymbol{07m_{cg}} - .10m_{cg}^2 - 0.03m_{cg}^3 - .\boldsymbol{02m_{cg}} = f_{u=1}^3$$

$$f_{u=1}^3 = 2.61 - .\boldsymbol{09m_{cg}} - .10m_{cg}^2 - 0.03m_{cg}^3$$

At $u_{cg} = 1$ (82%) we find that the local maximum for gene expression is at

$m_{cg} = -0.62$ (40.2%), while at $u_{cg} = -1$ (24.1%) the local maximum for gene expression

is at $m_{cg} = -0.29$ (48.8%). As up-stream CG methylation decreases (Figure 2.6b, from

purple to yellow lines), gene body CG methylation is expected to have a more positive effect

on gene expression.

While it has long been noted that regulatory region methylation is linked with

reduced gene expression, here we find evidence that the difference in methylation between

these regions also appears to correlate with gene expression. The negative interaction term

between up-stream and gene body CG methylation predicts that distinctly different levels

of methylation up-stream and within genes tends to correspond with higher levels of gene

expression. When gene body CG methylation and regulatory methylation are both high,

gene expression tends to be low (Figure 2.6b, purple lines at high gene body CG values).

However, as either decreases (Figure 2.6b, purple lines at low gene body CG values, or

yellow lines at high CG methylation values), gene expression is expected to increase.

Three Positive Interaction Terms: $\boldsymbol{f^4} + \boldsymbol{m_{cg}}(.\boldsymbol{02u_{chh}} + .\boldsymbol{02m_{chh}} + .\boldsymbol{04l_{exon}}) = \boldsymbol{f^5}$:

While only up-stream CG methylation showed a negative interaction with gene body CG

methylation, three terms have positive linear interactions: Up-stream CHH methylation,

gene body CHH methylation, and exon length. These can be treated in much the same way

as our negative interaction term. Depending on the values of these terms, they can offset

each other and lead to the removal of any interaction effect. For example if exon length

$(l_{exon}) = -1$ and up-stream(u_{chh}) and gene body(m_{chh}) CHH methylation = 1 these

positive interaction terms cancel out (-.4+.2+.2=0). However, if we consider them varying

in the same direction, they can have a striking effect on the relationship between gene body

CG methylation and gene expression. At $u_{chh} = m_{chh} = l_{exon} = 1$ (and the negative

interaction term $u_{cg} = 0$), we see the local maximum is at a methylation level of $m_{cg} =$

-0.05 (55.1%). If our negative interaction term $u_{cg} = -1$, this increases to $m_{cg} = +0.05$

(57.3%) (Figure 2.6b, varying the values of our interaction terms, $u_{chh} = m_{chh} = l_{exon} =$

$-u_{cg}$ from -1.6 to 1.6, as the summed interaction term increases (lines become yellow) the

local maxima for gene expression does so as well). When $u_{chh} = m_{chh} = l_{exon} = -u_{cg} < 0$,

gene body CG methylation is almost purely repressive. At a summed interaction value less

than -0.7 there is no longer a local maximum, and CG methylation has a purely negative

effect on gene expression.

Quadratic Interaction Terms: $\log(GE) = f^5 + .03m_{cg}^2 l_{exon} = f^6$: Finally the

interaction between the quadratic gene body CG methylation term and exon length is

included in this model. As our quadratic term increases, not only does the position of the

local gene expression maximum increase, so to does the inflection point (the point at which

the function changes from concave to convex). Now, at the same linear interaction values

tested above ($u_{chh} = m_{chh} = l_{exon} = -u_{cg} = 1$), our local maximum occurs at $m_{cg} =$

0.07 (57.8%)(Figure 2.6c). As exon sizes increase, the effect of gene body CG methylation is

expected to rapidly become more positive, and peak gene expression is predicted to occur

at higher m_{cg} levels. At $l_{exon} = 3$(3.5 kb), we find that the local maximum for gene

expression occurs at $m_{cg} = 0.90$ (78.1%) and at $l_{exon} = 4$(6kb) there is no longer a local

maximum for m_{cg}, and the highest expected gene expression occurs at m_{cg} approaching

100% (largest gene size in Figure 2.6c). It appears that for genes with smaller exons,

moderately methylated genes are most highly expressed, but as genes become larger so to

does the level of gene methylation that is associated with more highly expressed genes. Our

gene size by gene body CG methylation results confirm a pattern observed by Zilberman et

al. (ZILBERMAN *et al.* 2006) in the first genome-wide methylome analysis in *Arabidopsis* in

which found only a marginal relationship between gene size and gene expression, except

for the genes in which gene bodies were methylated and then they found a positive

relationship between gene size and gene expression.

Individual effects of interaction terms: $\log(GE) = f^5 + .08l_{exon} - .02l_{exon}^2 = f^6$:

Finally, we consider the effect of multiple terms simultaneously. Up until this point we have

only included gene body CG methylation effects, and its interaction terms, while not

including the independent effects of the term with which it interacts. Independent of gene

body CG methylation, we find that gene expression tends to increase as the standardized

exon length increases from -1 (500bp) to 2 (2kb), and beyond this point we expect a

decline. In the absence of interaction terms, only considering independent effects of gene

body CG methylation and exon length, we would estimate that peak gene expression occurs

at an exon length of 2kb, and methylation of 45%. Here we show that the effect of gene

body CG methylation on expression is extremely size dependent, and that gene expression

is expected to be highest for large highly gene body CG methylated genes, but lowest for

small highly gene body CG methylated genes (Figure 2.6d). It may be that as exon length increases, gene methylation is necessary to stabilize transcription, while for smaller genes it is not necessary for this purpose, and rather plays a repressive effect due to condensing chromatin near the transcription start site.

In this same way all other independent and interaction terms could be added to this model, parameters considered, and hypotheses tested. As nine distinct parameters are included (with 27 total terms) in this model the results quickly become difficult to conceptualize or visualize, yet through full-model construction, followed by simplification methods as presented above it is possible to decipher complex higher order regulatory interactions. We briefly discuss the effects of the other significant gene size and methylation terms in this model.

Intron Length

Intron length shows significant first, second, and third order effects with a gene expression peak at an intron size of approximately 1700 base pairs. Additionally, a positive interaction term with both exon length and number of introns suggests that generally, longer genes with more introns tend to be more highly expressed. Although relatively large genes do tend to be most highly expressed, there are negative quadratic terms for both exon and intron length that suggest after a certain point, increasing exon and intron length should be associated with decreased gene expression.

Non-CG Gene Body Methylation

Gene body CHG methylation had significant linear, quadratic, and cubic independent terms, and an exon length interaction term. Gene body CHG methylation has a negative effect on gene expression across nearly its full range of possible values (Figure 2.5), and it appears that it is the increase from no CHG methylation to slight CHG methylation that reduces gene expression. After this point the effect of CHG methylation appears to be minimal. The negative exon length interaction term suggests that long genes with CHG methylation tend to be more significantly repressed than smaller genes.

Gene body CHH methylation was found to have a negative effect on gene expression (Figure 2.5), but a positive interaction with gene body CG methylation. Thus, as gene body CHH methylation increases, gene body CG methylation is expected to have a more positive effect on gene expression, but mean gene expression, independent of gene body CG methylation, is expected to decrease. Like CHG methylation, a manual inspection reveals that the jump from no CHH methylation to low levels of CHH methylation leads to a decrease in gene expression, but after this, the effects of increased methylation are minimal.

While it has been suggested that non-CG gene body methylation may be misattributed to genomic regions that are actually pseudogenes or paralogs (SCHMITZ *et al.* 2013a; SEYMOUR *et al.* 2014), here we find evidence that in at least some cases these genes are still expressed, albeit at lower levels than non-methylated genes. One possible explanation is that non-CG methylation of genes may be a first step on the path toward pseudogenization (LI *et al.* 2005), whereby genes become targeted by non-CG methylation, gene expression is reduced, mutational constraints become lightened, and eventually the gene becomes entirely non-functional. Additionally, it may be that tightly developmentally

controlled small RNAs are responsible for the majority of this methylation, and the use of identical tissue for methylation and gene expression analysis would identify a stronger role of gene body non-CG methylation on gene expression. Finally, even trace amounts of non-CG gene body methylation may be indicative of the presence of small RNAs, and RNA-directed DNA methylation (RdDM) (WASSENEGGER 2000). It could be that the methylation of just a few nucleotides by a single 24nt siRNA is enough to reduce gene expression, without significantly altering the methylation state of the whole gene.

Regulatory Region Methylation

Along with a negative interaction with gene body CG methylation, up-stream CG methylation also has a direct negative effect on gene expression (Figure 2.5) and a negative interaction with up-stream CHH methylation. Not only does up-stream CG methylation limit the positive effect of gene body CG methylation on predicted gene expression, it also directly reduces predicted expression. Up-stream CHH methylation has both a significant positive linear effect on gene expression (Figure 2.5), and a positive interaction with gene body CG methylation. The negative interaction term with up-stream CG methylation suggests that while up-stream CHH methylation generally has a positive effect on gene expression, when it is found alongside CG methylation, this effect is negated.   While down-stream CHH methylation did not interact with gene body CG methylation, it was also found to have a positive effect on gene expression (Figure 2.5).

A previous study in *Arabidopsis* similarly found that there was a positive correlation between gene expression and regulatory CHH methylation (albeit not in a regression framework) (GENT *et al.* 2013). They posit that as gene expression increases, unstable

transcripts are produced as by-products at both the 5' and 3' ends of genes. In turn, this lead to the production of small RNAs that can target and cause CHH methylation bracketing highly expressed genes through RNA directed DNA methylation (RdDM). The possibility that increased gene expression causes increased regulatory CHH methylation, and not vice-versa does not introduce bias in this framework, but rather reinforces that our interpretations do not imply causality between these variables.

Gene expression modeling overview

While the traditional method of looking for simple associations between methylation state and gene expression has provided some insight into epigenetic regulation, here we demonstrate that modeling approaches can provide additional insight into these systems. We explain a surprisingly high (20.1%) amount of the variation in log(gene expression) simply through methylation and gene architecture variation. We considered a potential 454 parameters in our model before settling on 29, but it is important to note that many other factors such as presence of enhancers within the gene body, distance to transposable elements, likely also modify the role of methylation on expression. By considering exon and intron length within this model we do take the first steps to account for these potential confounding factors of methylation on expression. It is worth stressing that the gene expression and methylome data were not only collected from different individuals, but also different genetic lines, using different vegetative tissue types, and grown under slightly different greenhouse conditions. It is certainly possible that a similar model, tuned across multiple paired methylome and gene expression samples, could predict gene expression with greater precision. This portion of gene expression

variation explained represents that which is at least relatively stable across individual genotypes, tissue, and conditions. While here we apply this model to gene expression on a gene-by-gene basis, through altering the response variable to another parameter of a gene, such as it's mutation rate, gene expression variance, or the tissues in which it is expressed, this model could be extended to look for other roles of DNA methylation on gene function and evolution.

Results from this and other (ZILBERMAN *et al.* 2006; LI *et al.* 2012; WANG *et al.* 2014; YANG *et al.* 2015) studies suggest that gene body CG methylation needs to be considered to have a quadratic effect on gene expression, and that this effect is highly dependent on exon size. Thus, genes can either be parsed according to exon length prior to estimating the role of gene body CG methylation on expression, or the interaction between exon length and methylation should be considered in the model. Other forms of methylation appear to have a more straightforward role in regulating gene expression, and in some cases it may suffice to predict that, for example, as up-stream CG methylation increases at a gene, its expression will likely decrease.

Gene Ontology Analysis of Genes with High CG Gene Body Methylation

Comparing genes in the top 10% genome wide for gene body CG methylation with the remainder of the genome, we found numerous gene categories that are either enriched or depleted in our set of highly CG methylated genes. Genes coding for proteins with kinase activity, involved in signal transduction, and nucleotide binding were among those which tended to be highly methylated, while proteins functioning in the thylakoid, plastid, and ribosome, as well as proteins involved in primary metabolism, photosynthesis, and RNA

binding tended to be lowly or moderately methylated (Figure 2.7). Similar results have

been found in *Brachypodium*, rice (WANG *et al.* 2014), and *Arabidopsis* (ZILBERMAN *et al.*

2006).

Decreased Methylation Near Transcription Start Sites

We looked for changes in methylation near gene transcription start sites. We found

that CG, CHG, and CHH methylation all were significantly depleted at and around gene start

sites (Figure 2.8). This depletion, along with the negative interaction term between up-

stream and gene body CG methylation on gene expression, points towards a role of

methylation in epigenetically labeling coding genetic regions. Additionally, recent evidence

has shown that in *M. guttatus* genetic recombination occurs at higher frequency near gene

start sites. In other systems it has been shown that DNA methylation is negatively

correlated with recombination (MIROUZE *et al.* 2012), and it may be that decreased

methylation at gene start sites is related to the increase in recombination.

Decreased methylation near transcription start sites (TSS) was one of the earliest

discovered phenomena of gene methylation (ZILBERMAN *et al.* 2006). However, new

evidence in *M. guttatus* (HELLSTEN *et al.* 2013) provides us with a novel framework in which

to view this pattern. Hellsten et al. (HELLSTEN *et al.* 2013) identified an approximately two-

fold increase in recombination near gene start sites (the beginning of exon 1 being most

enriched), and postulated that this may be related to nucleosome depleted open chromatin

at these regions as is the case in *Arabidopsis* (ZHANG *et al.* 2012) and rice (YU *et al.* 2002). At

the time of their publication however, there was no evidence for a similar trend in *Mimulus*.

Here, evidence of depleted methylation near TSS (Table 2.1; Figure 2.8) provides support

to the theory that open chromatin (unmethylated) near TSS may increase local recombination rates. It appears that at least in yeast double stranded breaks occur most frequently in open chromatin regions (PAN *et al.*), which may explain the observed increase in recombination near transcription start sites. It is likely that the increased recombination near TSS is simply a by-product of the dual forces exerted by DNA methylation, one involved in gene regulation, and another limiting double stranded breaks. The ability for DNA methylation to alter both of these processes provides an interesting link between gene regulation and DNA recombination that may or may not prove to be of evolutionary significance. Further studies linking methylation and recombination at a nucleotide level should further clarify this trend.

Transposable Element Methylation

We identified 1,411 transposable elements across the genome ranging in copy number from 1 to 2,380 (median copy number = 7). Percent methylation was calculated in each of three sequence contexts. In total, 34% of the *M. guttatus* genome was estimated to be of transposable element sequence, and methylation levels within transposable elements were significantly higher than that of genes, and at similar levels to inter-genic regions (Table 2.1). We did not find there to be a significant copy number effect on TE methylation. Of the top 25 most common transposable elements in the *Mimulus* genome, six were type 1, and 19 were type 2 transposons (Table 2.5).

We find that DNA methylation in all contexts is enriched in transposable elements relative to genes, however this is most significant for non-CG methylation (Table 2.1). This suggests that both RNA dependent DNA methylation (RdDM) is targeting and silencing

transposable elements in *M. guttatus* as is this case in other angiosperms. Found at 2,380

copies, the helB8c family of helitron elements is far and away the most common transposon

in the *Mimulus* genome (more abundant than the next seven TE families combined; Table

2.5). Helitrons are a relatively newly discovered class of type 2 transposable elements that

propagate through a rolling circle mechanism that is still somewhat mysterious (XIONG *et al.*

2014). One thing that is clear, is that these elements have been highly successful in

propagating across flowering plants, making up 2% of the *Arabidopsis* genome (HOLLISTER

and GAUT 2007); a single family of helitrons makes up 6% of the maize genome (XIONG *et al.*

2014), making it the most abundant DNA transposon identified. Here, we provide evidence

for the success of these elements across the diversity of flowering plants.

Conclusions

Much remains unknown about the gene regulatory information contained in an

organism's methylome, but here we provide further evidence of complex interactions

between gene methylation and expression. DNA methylation may actively alters gene

expression, itself be altered by gene expression, or both methylation and expression may

be jointly determined by a distinct genetic feature. Still the ability to explain over a fifth of

the variation in log transformed gene expression by local DNA methylation, and basic

genetic architecture (exon length, intron length, exon number), is promising and has

numerous potential applications. Recent efforts have shown that the plant methylome is

relatively stable throughout development (EICHTEN *et al.* 2013b), unlike gene expression. In

this way methylation at a gene likely reflects moderately stable epigenetic control of gene

expression, while developmentally activated transcription factors and small RNAs may

provide highly plastic gene expression control throughout development. Through

combining differential methylation analyses across tissue types, environmental treatments,

or genetic lines with a modeling approach as described here; our understanding of the role

of epigenetic variation in gene regulation can be greatly increased.

Table 2.1. *Mimulus guttatus* methylation across sequence contexts and genomic regions.

	Proportion of cytosines methylated		
	CG	CHG	CHH
Transposable Elements	0.73	0.36	0.063
Gene Body	0.56	0.038	0.012
1st 500bp of Gene Body	0.28	0.032	0.019
Up-stream Regulatory	0.35	0.11	0.027
Inter-Genic Regions	0.75	0.45	0.072
Total	0.72	0.365	0.061

Table 2.2. Summary of REML genetic architecture and methylation fit on log transformed gene expression.

R-Square	0.202
R-Square Adj.	0.200
Root Mean Square Error	0.641
Mean of Response	2.483

Table 2.3. Analysis of variance in gene expression predictive model.

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	28	1764.7903	63.0282	153.6
Error	17014	6981.5210	0.4070	**Prob > F**
C. Total	17042	8746.3113		<.0001*

Table 2.4. Sorted estimate of parameter effects on log transformed gene expression.

Positive Terms	Estimate	Std Error	t Ratio	Prob>|t|
Intron Length	0.3472	0.0117	29.75	<.0001
Gene Body CHG^2	0.0874	0.0082	10.64	<.0001
Number of Exons*Intron Length	0.0793	0.0081	9.74	<.0001
Exon Length	0.0767	0.0102	7.55	<.0001
Exon Length* Intron Length	0.0553	0.0070	7.86	<.0001
Gene Body CG * Exon Length	0.0392	0.0089	4.4	<.0001
Gene Body CG^2* Exon Length	0.0303	0.0069	4.37	<.0001
Up-Stream CHH	0.0275	0.0055	4.99	<.0001
Gene Body CG*Gene Body CHH	0.0244	0.0064	3.78	0.0002
Down-stream CHH	0.0185	0.0050	3.69	0.0002
Up-stream CHH* Percent CG	0.0167	0.0051	3.28	0.0011
Intron Length^3	0.0105	0.0007	14.4	<.0001
Exon Length^2 * Number of Exons	0.0074	0.0009	8.19	<.0001
Negative Terms				
Gene Body CHG	-0.3273	0.0197	-16.58	<.0001
Intron Length^2	-0.1611	0.0076	-21.18	<.0001
Gene Body CG^2	-0.0980	0.0092	-10.62	<.0001
Gene Body CG	-0.0720	0.0118	-6.09	<.0001
Exon Length * Number of Exons	-0.0662	0.0076	-8.72	<.0001
Gene Body CHH	-0.0451	0.0076	-5.93	<.0001
Number of Exons	-0.0308	0.0112	-2.75	0.0059
Percent CG^3	-0.0277	0.0059	-4.73	<.0001
Up-Stream CG	-0.0274	0.0054	-5.06	<.0001
Exon Length^2	-0.0205	0.0033	-6.28	<.0001
Gene Body CHG * Exon Length	-0.0198	0.0058	-3.41	0.0007
Up-stream CG* Up-stream CHH	-0.0188	0.0058	-3.23	0.0012
Up-stream CG* Gene Body CG	-0.0170	0.0052	-3.28	0.001
Exon Length * Intron Length * Number of Exons	-0.0118	0.0016	-7.21	<.0001
Gene Body CHG^3	-0.0063	0.0008	-7.95	<.0001

Table 2.5. Transposable element frequencies, classes, and methylation.

ID	Copies	CG	CHG	CHH	Family	Class
		Percent Methylation				
helB8c	2380	0.809	0.473	0.052	Helitron	2
MULE_MITE1c	674	0.627	0.263	0.102	MITE	2
Copia1b	424	0.780	0.437	0.058	Copia	1
helD8b	402	0.712	0.357	0.055	Helitron	2
MULE_MITE2b	245	0.633	0.285	0.077	MULE	2
pogo_MITE2b	203	0.738	0.281	0.071	MITE	2
MULE_MITE16b	200	0.713	0.207	0.070	MULE	2
hAT_MITE1	197	0.782	0.294	0.051	MITE	2
MULE_na62	165	0.768	0.359	0.064	MULE	2
MULE_MITE1a	158	0.720	0.250	0.071	MULE	2
LARD4	155	0.793	0.442	0.081	LARD	1
hAT_na66a	151	0.869	0.276	0.042	hAT	2
Tourist6c	151	0.634	0.259	0.071	MITE	2
MuDR8	150	0.791	0.492	0.089	MuDR	2
MULE_na13a	145	0.752	0.400	0.068	MULE	2
Copia1a	143	0.717	0.374	0.045	Copia	1
Copia2	137	0.685	0.494	0.085	Copia	1
SINE1a	134	0.685	0.293	0.112	SINE	1
Gypsy8	128	0.605	0.228	0.033	Gypsy	1
MULE_na13b	128	0.449	0.260	0.058	MULE	2
helF3c	119	0.737	0.362	0.067	Helitron	2
Jittery7	116	0.639	0.260	0.053	Mu	2
Toursit4c	115	0.781	0.315	0.085	MITE	2
Gypsy4	111	0.818	0.402	0.051	Gypsy	1
MULE_MITE25b	109	0.626	0.151	0.042	MULE	2

Figure 2.1. Interspecific comparison of plant DNA methylation levels. A comparison of global DNA methylation levels in CG (red), CHG (green), and CHH (blue) sequence contexts found in *Mimulus guttatus* compared with those of *Arabidopsis thaliana*, *Glycine max* (Schmitz *et al.* 2013a), *Brachypodium distachyiom* (Takuno and Gaut 2013), *Oryza sativa* (Li *et al.* 2012), *Solanum lycopersicum* (Zhong *et al.* 2013), and *Zea mays* (Gent *et al.* 2013).
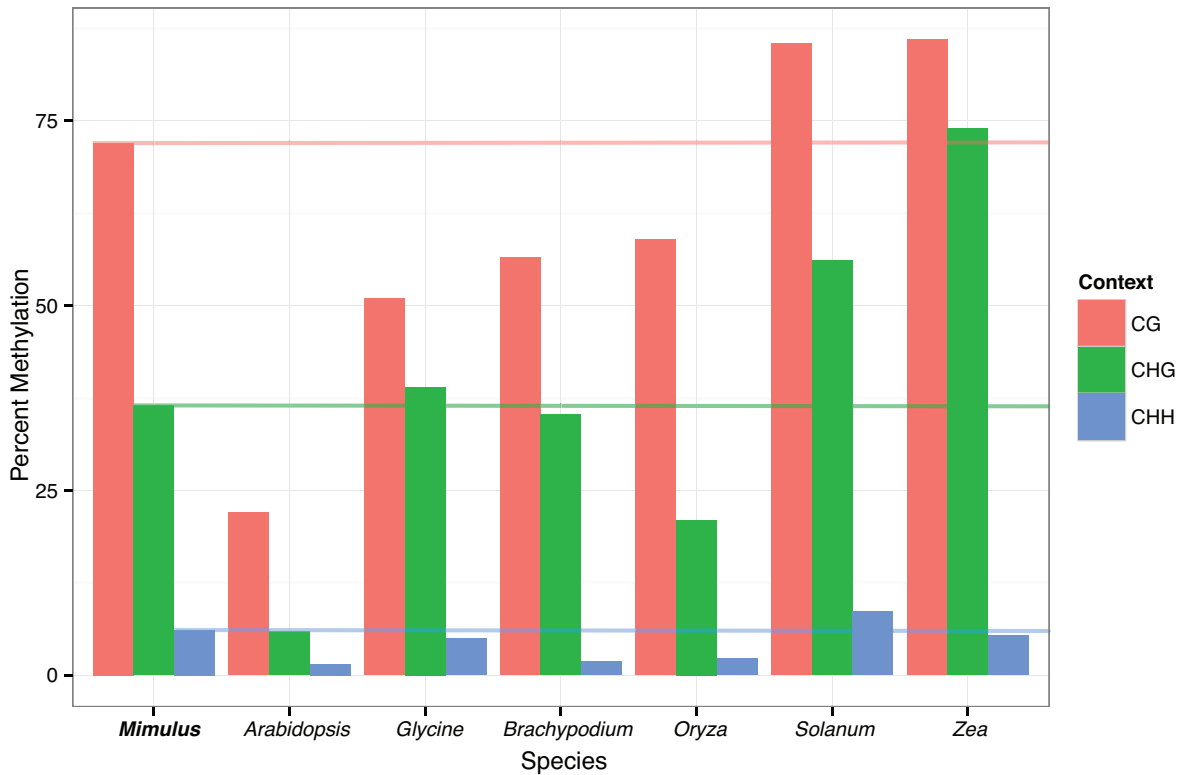
Figure 2.2. DNA methylation across the *Mimulus guttatus* genome. DNA methylation across the 14 *Mimulus guttatus* linkage groups (putative chromosomes) in all three sequence contexts: CG (red), CHG (green), and CHH (blue). Centromeric repeat densities, adapted from (FLAGEL *et al.* 2014), are shown along the X-axis (darker bars indicate higher repeat density). Areas with higher repeat density tend to also have higher DNA methylation. A smoother line (WICKHAM 2009) was fit across 1kb methylation averages.
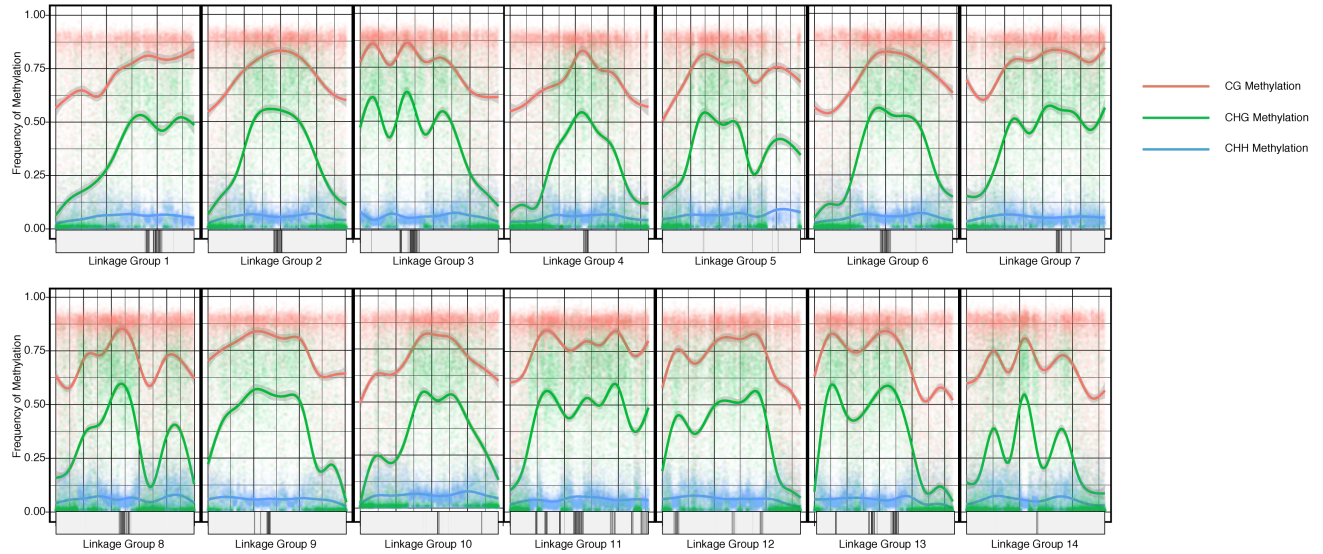
Figure 2.3. Variation in methylation and expression across chromosomes. A heatmap showing variation in gene expression and methylation across the 14 *Mimulus guttatus* putative chromosomes. The 14 chromosomes clustered into two large groups, those with generally high methylation and low gene expression (top cluster, red dendogram), and those exhibiting the opposite pattern (bottom cluster, green dendogram). On the heat map, red indicates high values and blue indicates low values of methylation or gene expression.
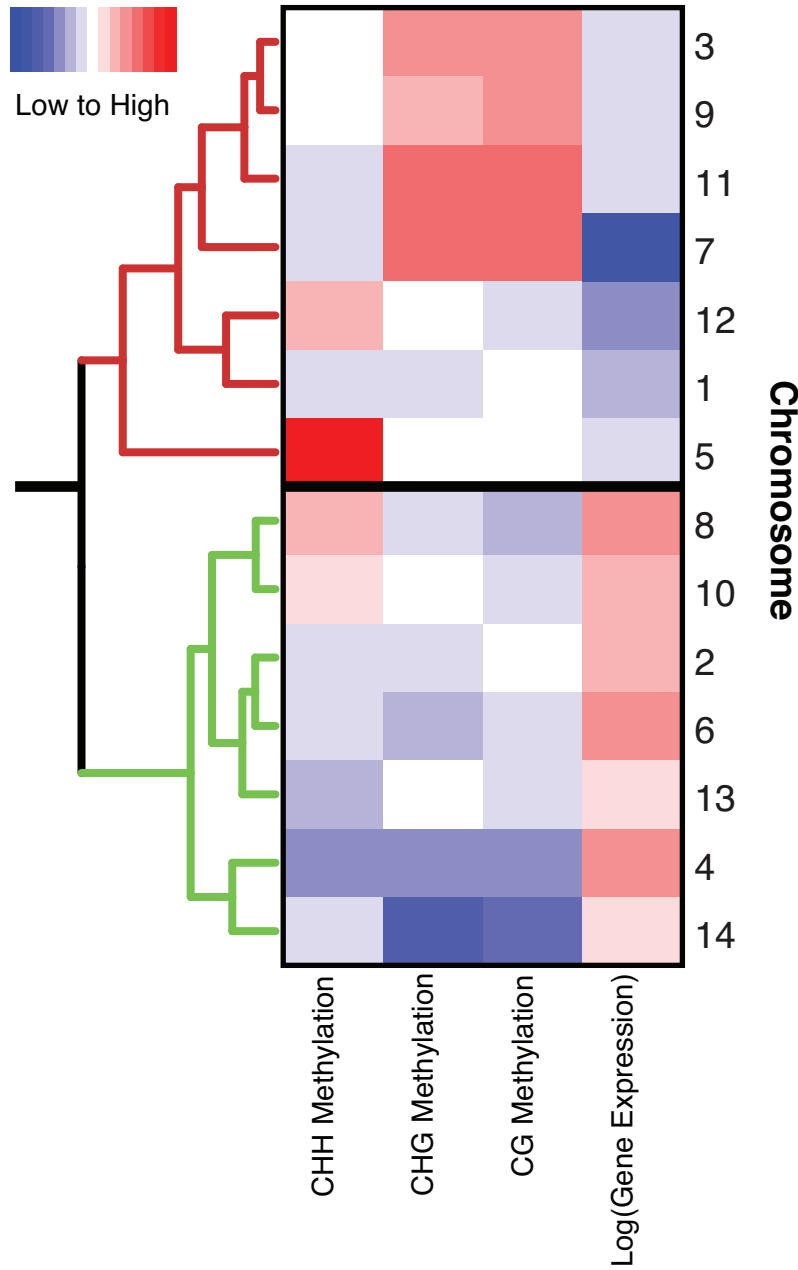
Figure 2.4. Correlation matrix between forms of methylation at individual genes.
Clouds represent density, and lines show the slope of the correlation. Green lines indicate forms of methylation with a positive correlation, while red represents negative correlation. Numbers represent the Pearson correlation (*r*) value, bolded numbers highlight correlations with an r > 0.35. All correlations were found to be statistically significant (n=17,038, p<0.05).
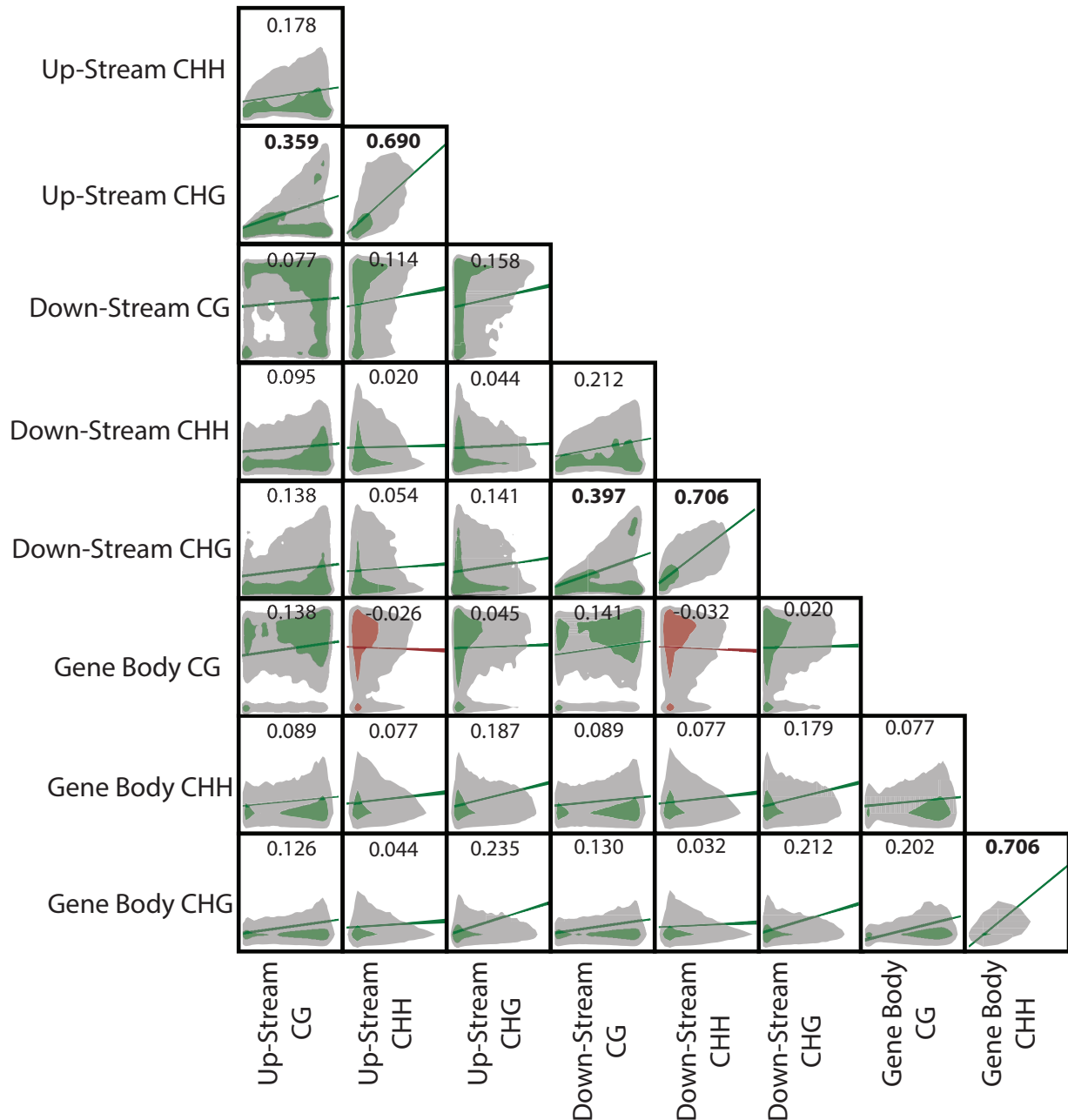
Figure 2.5. Correlations between DNA methylation and gene expression. A single star represents a significant linear correlation, two stars a significant second-order correlation, and three stars a third order correlation. The red dashed lines represent the means, the black line represents the regression line, and the blue line represents 95% confidence intervals.
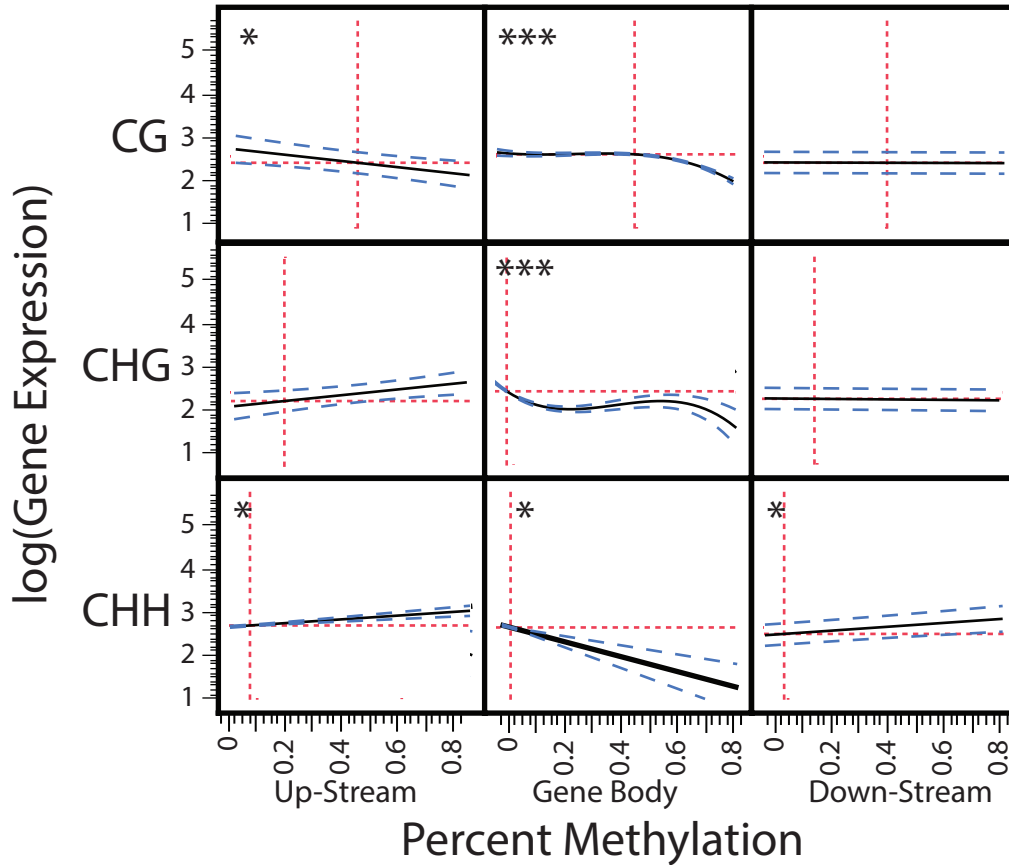
Figure 2.6. DNA methylation modeling to predict gene expression. A visual depiction of our simplified model showing the effect of gene body CG methylation and an increasing complexity of interaction terms on gene expression. A) A scatterplot comparing Z-transformed gene body CG methylation values with log(gene expression) values. The black line shows the linear term, green line includes both the linear and quiadratic term, and the blue line includes linear, quadratic, and cubic terms. B) Interaction plot depicting the interaction between gene CG methylation and exonlength, up-stream CHH methylation, gene body CHH methylation, and gene body CG methylation on gene expression. Summed terms across these four terms are considered ranging from -1.6 (dark purple) to 1.6 (yellow). Points represent actual genes CG gene body methylation, gene expression, and their color represents their interaction sum on the same scale as the model colors. C) The second order interaction term of squared gene body CG methylation by exon length is added to the model depicted in B. As exon length increases (goes from red to blue) gene body CG methylation is found to have a more positive effect on gene expression. Points represent genes, and colors represent the exon length of these genes on the same scale as the model colors. D) The independent effect of exon length on gene expression is added to the model depicted in C. The shape of the lines does not change, however predicted gene expression is altered (the lines move up or down on the y-axis) depending on the predicted effects of exon length on gene expression.
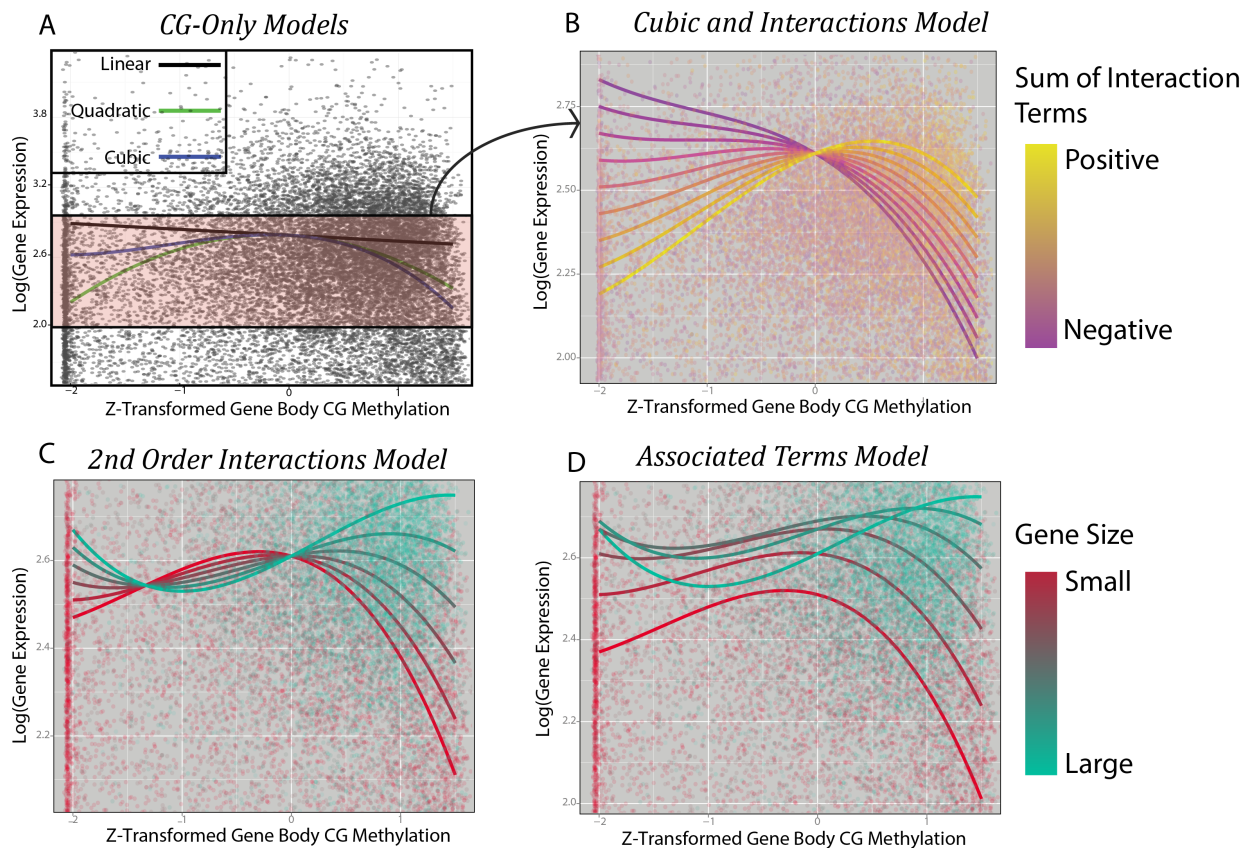
Figure 2.7. Gene Ontology classes over and underrepresented in highly gene body methylated genes. Genes Ontology terms significantly enriched and depleted in genes in the top 10% for gene body CG methylation. X-axis shows the percent of genes in both the high CG methylated portion (blue) as well as the remainder of the transcriptome (red) that contained the given GO terms. Text color represents the class of GO term: blue-molecular functions, grey-biological processes, and green-cellular component.
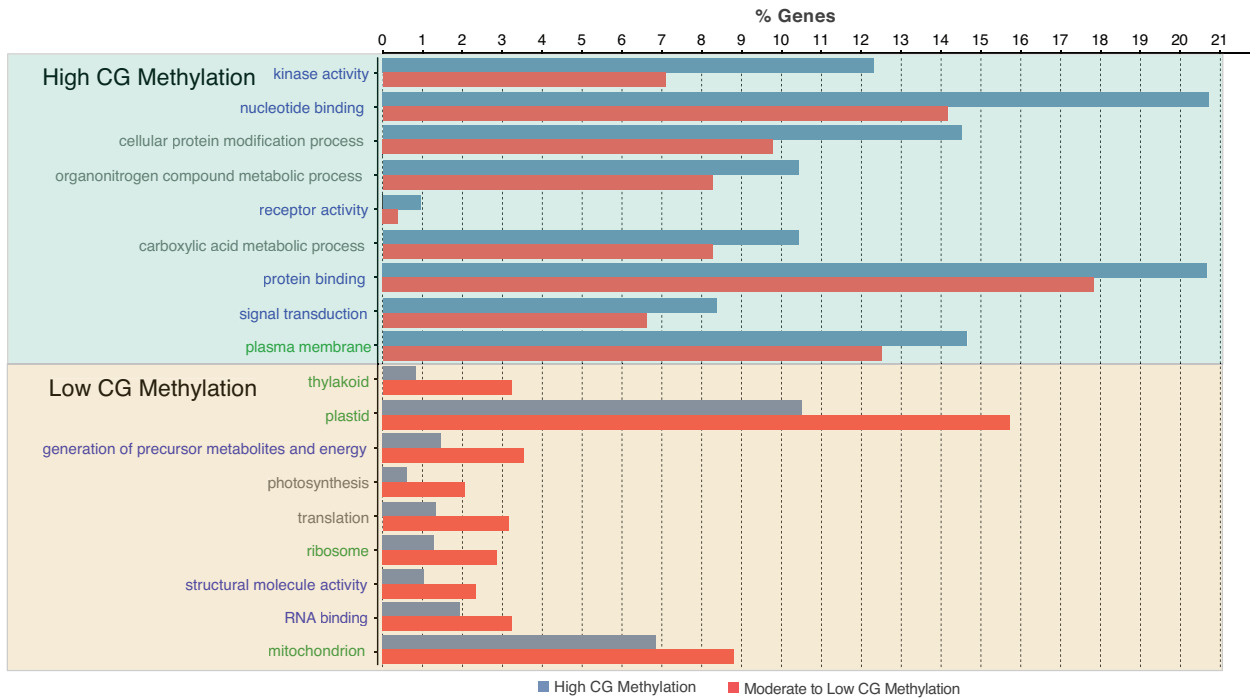
Figure 2.8. DNA methylation around transcriptional start sites. Around gene start sites, and persisting into the first 500 base pairs of the gene body, we observe a significant drop in DNA methylation. For CG (p-value=6.45×10^{-55}), CHG (p=3.38×10^{-100}), and CHH (p=4.61×10^{-308}) methylation was significantly reduced in the first 500bp of the gene relative to the up-stream regions. Both CG (p=3.55×10^{-138}) and CHG (p=4.04×10^{-11}) methylation then significantly increases over the next 500 bp, while for CHH it continued to decline (p=8.93×10^{-10}).
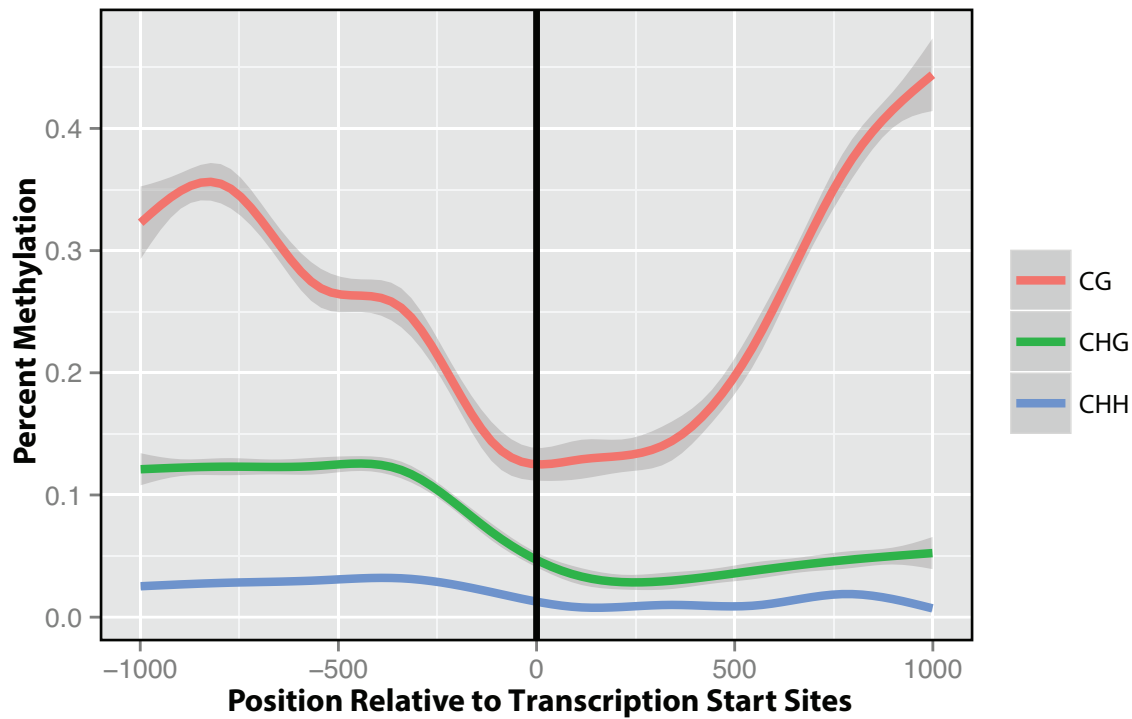
CHAPTER 3

Differential methylation in the offspring of wounded *Mimulus guttatus*

Abstract

The offspring of wounded *Mimulus guttatus* plants differentially express hundreds of genes and produce leaves with greater numbers of trichomes. Here we utilize whole genome bisulfite sequencing in the progeny of damaged and control plants to assess the potential role of differential methylation on the previously observed patterns of gene expression. We identify thousands of genomic regions of differential methylation, as well as an increase in gene methylation variation in the progeny of damaged plants. A significant overabundance of these differentially methylated regions overlap with differentially expressed genes coding and up-stream regions, suggesting a role of differential methylation in transgenerationally plastic gene expression. Through differentiating between CG, CHG, and CHH methylation we find evidence that different classes of methylation regulate different classes of transgenerationally plastic genes, and identify likely candidate genes that function in transgenerational plasticity.

Introduction

Phenotypic plasticity, or the ability to modify development according to environmental cues, is of vital importance in the success of life in response to the constantly changing abiotic and biotic world around us. While the molecular mechanisms, evolutionary implications, and diversity of plastic phenotypic responses to the environment have been well-studied, the ability of a subset of these signals to be transmitted to the next generation remains poorly understood and skeptically viewed (HEARD and MARTIENSSEN 2014). A great deal of this skepticism is due its ties with "Lamarckianism", but other, more scientifically grounded concerns, also linger. One of these is that their contribution to fitness and phenotypic variation in nature is poorly understood. Through the budding field of ecological epigenetics (BOSSDORF *et al.* 2008; KILVITIS *et al.* 2014) these questions must be addressed using a combination of field, greenhouse, and common garden techniques. Another weakness in the field is the lack of a demonstrated molecular mechanism to explain this inheritance. In this study we identify thousands of differentially methylated regions that provide evidence for an epigenetic role in the transmission of environmental information

In response to varying biotic (AGRAWAL *et al.* 1999; VAN DAM and BALDWIN 2001; AGRAWAL 2002; HOLESKI 2007; HOLESKI *et al.* 2010; SCOVILLE *et al.* 2011; HOLESKI *et al.* 2012; LAU 2012; LUNA *et al.* 2012; RASMANN *et al.* 2012; SLAUGHTER *et al.* 2012; COLICCHIO *et al.* 2014), and abiotic (DURRANT 1962; BLÖDNER *et al.* 2007; GALLOWAY and ETTERSON 2007; BOYKO *et al.* 2010; VERHOEVEN *et al.* 2010; HERMAN and SULTAN 2011; HERMAN *et al.* 2012; VERHOEVEN and VAN GURP 2012a) conditions in the parent generation, offspring fitness, phenotype, gene expression, and methylation have all been found to vary. For example, in

Mimulus guttatus, progeny plants increase trichome production and differentially express over one thousand genes in response to parental wounding (HOLESKI 2007; SCOVILLE *et al.* 2011; COLICCHIO *et al.* 2014). The progeny of drought stressed *Polygonum persicaria* individuals alter seedling growth to increase fitness in dry conditions (HERMAN *et al.* 2012). Maternal light environment influences offspring growth and increases offspring fitness in similar environments in *Campanulastrum americanum* (GALLOWAY and ETTERSON 2007). Finally, response to a wide variety of environmental signals in the progeny of apomictic dandelions (*Taraxacum officinale*) increases DNA methylation variation (VERHOEVEN *et al.* 2010) and growth (VERHOEVEN and VAN GURP 2012a). Additionally, evidence in *Solanum lycopersicum* and *Arabidopsis thaliana* suggests that key players in the plant hormone response and epigenetic regulatory pathways are vital for at least a portion of these transgenerational effects (RASMANN *et al.* 2012). While this is a promising start, the specific loci that control transgenerational inheritance and the timing and mechanism of germ line reiteration remain unknown.

DNA cytosine methylation acts alongside histone modifications and small RNAs as key epigenetic regulators through a number of linked pathways in which an ever-growing number of enzymes and molecules (EICHTEN *et al.* 2014; HUFF and ZILBERMAN 2014; MATZKE and MOSHER 2014; HOLOCH and MOAZED 2015; LI *et al.* 2015) alter the transcription and translation of genes into proteins (along with traditional protein transcription factors). DNA methylation can occur on any cytosine nucleotide across the genome, but the mechanisms that propagate methylation and the effect they have on genome regulation vary between those in a CG, CHG, and CHH (where H is any nucleotide that is not G) context (BARTEE *et al.* 2001; CAO and JACOBSEN 2002; KANKEL *et al.* 2003; CHAN *et al.* 2006; LAW and

JACOBSEN 2010; JONES 2012; LI *et al.* 2012; LAW *et al.* 2013b; HUFF and ZILBERMAN 2014; STROUD *et al.* 2014). While all three of these types of methylation play a role in silencing transposable elements (TEs), their role in regulating gene expression is much more variable. Not only does the context of the methylation (CG, CHG, CHH) impact their role in gene regulation, so to does the location of methylation relative to the gene (upstream, within an intron or exon, down-stream, etc.) (YUAN *et al.* 2007; LI *et al.* 2008; COLICCHIO *et al.* 2015a; KARIÑHO-BETANCOURT *et al.* 2015). CG methylation is the most prevalent form of DNA methylation (occurring on over 50% of CG sequences in many cases), and within gene bodies it has a very complex relationship with gene expression in which surrounding DNA methylation, gene length, and other factors alter the role of CG methylation in gene regulation (ZILBERMAN *et al.* 2006; WANG *et al.* 2014; COLICCHIO *et al.* 2015a). However, when methylation occurs in up-stream regulatory regions it appears to directly suppress expression (KILBY *et al.* 1992; METTE *et al.* 2000; ZILBERMAN *et al.* 2006; COLICCHIO *et al.* 2015a).

CHG and CHH methylation are often grouped as "non-CG" methylation, and tend to be found at significantly lower levels across the genome than CG methylation (ZHANG *et al.* 2006; CHODAVARAPU *et al.* 2012; LI *et al.* 2012; TAKUNO and GAUT 2013; ZHONG *et al.* 2013; SEYMOUR *et al.* 2014; WANG *et al.* 2014; COLICCHIO *et al.* 2015a). CHG and CHH methylation are propagated and reiterated by a partially overlapping set of enzymes and molecules, often initiated by 24-nt small interfering (si) RNAs(LAW and JACOBSEN 2010; MOLNAR *et al.* 2010; GENT *et al.* 2013; LAW *et al.* 2013b; MATZKE and MOSHER 2014; STROUD *et al.* 2014; HOLOCH and MOAZED 2015). Both types of non-CG gene body methylation are associated with transcriptional silencing and decreased expression. The majority of epigenetic marks

appear to be reset during gamete formation in plants (LANG-MLADEK *et al.* 2010), but the recent discovery that si-RNAs are loaded into pollen granules (CALARCO *et al.* 2012) and are phloem mobile and can mediate methylation in recipient cells (LEWSEY *et al.* 2016) presents one possible mechanism through which environmentally induced epigenetic marks could be reinstated and transmitted between generations.

In this paper, I investigate the role of differential methylation in transgeneration induction in response to leaf damage in *M. guttatus* (HOLESKI 2007). Since Holeski's initial studies of trichome induction, the induction effect has been shown to be transmitted both paternally and maternally (In Press, Scoville et al., 2016), partially dependent on DNA methylation (In Press, Scoville et al. 2016), persist through at least two generations (In Press, Scoville et al., 2016), and involves differential expression of over 900 genes (COLICCHIO *et al.* 2015b). Trangenerational induction also has significant effects on plant resistance to herbivory in the field in *M. guttatus* (In Submission, Colicchio 2016). Additionally, a recently sequenced methylome has allowed us to study the relationship between methylation and expression in *M. guttatus* (COLICCHIO et al. 2015a). With this newfound molecular understanding of the interplay between methylation and gene expression we seek to identify how parental environment alters offspring methylome. Here, I present whole-genome methylome data from both the progeny of damaged and control plats, and examine patterns of differential methylation alongside previously published gene expression results.

Methods

Plant material, experimental design, and trichome phenotyping

A single *Mimulus guttatus* RIL 94 (F$_8$ generation) parent plant was grown under non stress conditions in the University of Kansas greenhouse in the same manner as described in the complementary gene expression study (COLICCHIO *et al.* 2014). Briefly, selfed seed from this individual was split randomly into damage and control treatment groups. For damage treatment, we punched two holes of c. 6mm diameter in each at the developmental point when the next leaf had expanded. Five wounded and 5 control plants were self fertilized to propagate seed for the transgenerational effect study described here. The progeny of these individuals were grown until the second leaf pair had expanded to one centimeter in width, at which point each leaf from the second node was collected and flash frozen in liquid nitrogen within one minute of removal from the plant.

DNA extraction and library preparation

We extracted DNA from the progeny of five damaged and five control individuals using a CTAB protocol (HOLESKI *et al.* 2013a). From these samples, we generated sequencing template for whole genome bisulfite sequencing (WGBS) following the PBAT (Post-Bisulfite Adaptor Tagging) protocol revision 12 (MIURA *et al.* 2012). With 1 ng of unmethylated lambda DNA obtained from Promega used as a spike-in control for conversion efficiency, between 55 and 100 ng of genomic DNA from each individual was treated with bisulfite using EZ DNA Methylation kit from Zymo Research. Two rounds of random primer extension for tagging bisulfite treated DNA with adaptors were performed

using primers for paired-end library construction. Unique adaptors were ligated onto each library to allow for down-stream de-multiplexing. The concentration of templates was determined by qPCR with Library Quantification Kits from KAPA biosystems, and due to low library quantities, one control individual was omitted from sequencing. Two lanes of 100 cycle reactions on HiSeq 2500 rapid-run were assigned for the library sequencing at the University of Kansas Genome Sequencing Core for these and three other PBAT WGBS samples (11 samples per lane). Each sample was sequenced in both lanes.

Read Mapping

We used the software BMap (MIURA *et al.* 2012) (http://itolab.med.kyushu-u.ac.jp/BMap/index.html) to map bisulifte treated reads to a resequenced *M. guttatus* RIL 94 reference genome created v2.0 reference genome. In short, BMap first searches candidate genomic loci for each read in two duplicated genome sequences, one with every C in the genome converted to a T (C2T), and one with G to A (G2A), using an approach called adaptive seed (KIEŁBASA *et al.* 2011). Next BMap creates pairwise alignments between the read and original DNA sequence of every candidate loci, and calculates scores for each alignment allowing mismatches between T in the reads with C in the reference. Finally an alignment with the highest score is reported for each read. We used default parameters for mapping with BMap, as previously found ideal and described previously (COLICCHIO *et al.* 2015a). Using alignments exported by BMap, methylation status for every cytosine in every read was called, and counts both supporting the methylated and unmethylated state are assigned for every cytosine residue of the reference genome.

Methylation levels for CG, CHG and CHH contexts are exported to different files to allow for independent analysis.

Global Methylation Analysis

In order to compare global methylation, we used a change point detector to scan the genome of each individual, to identify regions where average CG methylation changes(YOKOYAMA *et al.* 2015). This results in the partitioning of the genome into blocks ranging from tens to hundreds of KBs long in which methylation rates are relatively constant. Using ggplot we constructed plots showing the length and size distribution of these methyl-regions (MDM plots) for each individual(YOKOYAMA *et al.* 2015).

Gene Methylation Analysis

Using the same methodology as previously described we calculated percent methylation within upstream and gene body regions for each of the 24,130 genes in the *M. guttatus* v2.0 annotation(COLICCHIO *et al.* 2015a). For each gene we calculated the within-group (Control and Damage) standard deviation of gene body methylation for all three methylation types using average methylation across all nucleotides within a gene for a given individual. A paired t-test was performed to compare differences between standard deviation within the damage and control treatment group. Methylation values for each gene of the nine individuals were used to perform PCA analysis of the individuals. This was done distinctly for CG and CHH methylation. Additionally, within treatment group standard deviation was calculated for each gene for each type of methylation. A paired t-test was

performed on these standard deviation values to test if one group had more variable gene body methylation than the other. Plots were constructed using ggplot to compare the amount of methylation variation for each gene in damaged vs. control individuals.

To test for a relationship between changes in methylation between gene body CG, CHG, and CHH, as well as up-stream CG methylation types we calculated the pairwise correlation between difference in mean methylation for each gene between the offspring of damaged and control plants. Methylation values for each sequence context were Z-transformed prior to this analysis to normalize the data.

Using transposable element position information for previous analyses, we determined percent methylation across cytosine sequence contexts for the transposable elements in each individual. We merged the read counts that mapped to the same species of transposable element, and used a GLM with logit link function to call differential methylation between individuals in the damage vs. control treatment group. Different families of TE were separated after analysis, and patterns of differential methylation were analyzed for each family.

Identification of Differentially Methylated Regions:

We used the R package DSS to identify differentially methylated loci between damaged and treatment groups. We performed this procedure separately for CG, CHG, and CHH methylation, as well as a combined file to identify regions that showed general differential methylation across the three methylation types. We used the function DML to identify individual loci, and then the "callDMR" function to identify differentially methylated regions, and merge nearby differentially methylated regions into larger ones.

Default settings were used, smoothing was turned on, and a minimum region length was set to 20bp, and p.threshold to 0.01. We only considered regions from which the absolute value of the sum of the test statistics within the differentially methylated region was greater than or equal to 30. This was done to enrich our set of differentially methylated regions for both larger regions, and those for which there was more evidence that the region was differentially methylated between the two groups.

Custom scripts along with BEDtools were used to identify genomic features that overlap, or were within 1kb down-stream of differentially methylated regions. When differentially methylated regions were located within genes we broke the differential methylation position down further by whether introns or exons composed of over 50% of the region.

Contingency tables were constructed with grouping based on the presence/absence of an overlapping differentially methylated region and differential expression. Chi-square tests were performed for these tables, and follow up contingency tests were performed based on direction of significant expression and methylation. Gene Ontology enrichment analyses and KEGG mapping were performed for genes overlapping differentially methylated regions using the software Blast2GO.

Results

In the offspring of damaged and control plants, the average methylation was 71.8% and 67.7% in a CG context, 38.7% and 32.5% in CHG, and 14.6% and 13.2% in CHH respectively (Appendix 11). The offspring of one control plant (OC2) had particularly low inter-genic methylation in all sequence contexts, lowering control group average

methylation (Appendix 11). Similar MDM plots for CG methylation across the all plants suggest that global epigenetic regulation is not altered in the offspring of damaged plants (Appendix 12). In both damaged and control individuals, large 1kb-20kb regions of 75%-95% methylation, and smaller 250bp-5kb regions of less than 20% methylation predominate across the genome (Appendix 12).

Across treatment groups, gene body variation was higher in the offspring of damaged plants for all forms of methylation. The most pronounced shift was found for CG methylation (22.8% higher standard deviation for gene body methylation, Figure 3.1a), followed by CHG (17.8%, Figure 3.1b), and CHH (11.8%, Figure 3.1c) methylation. This may explain the observed increase in gene expression variation in the offspring of damaged vs. undamaged plants (mean standardized variation increase of 6.7%, Figure 3.1d). This shift seems to be the result of a general trend towards elevated variance rather than a subset of genes showing large scale variance increases (Figure 3.1 a:d).

The direction and magnitude of percent change in methylation for the three contexts of methylation were significantly correlated with each other (CHG/CHH: $r=0.361$, $p<0.0001$, CG/CHG: $r=0.280$, $p<0.0001$, CG/CHH: $r=0.195$, $p<0.0001$). Up-stream CG methylation had a slightly positive correlation with gene body CG methylation ($r=0.0233$, $p=0.008$), and a slightly positive, but non-significant correlation with both types of non-CG methylation. There was no general relationship between the direction of change of methylation of a gene and the gene expression response of that gene.

Transposable Element Methylation

Transposable element methylation was quite plastic in response to parental wounding. 205 of 1385 transposable elements considered in this study were differentially methylated in at least one sequence context. Change in direction was positively correlated for all three sequence contexts (CG/CHG: r=0.317, p<0.0001, CG/CHH: r=0.132, p<0.0001, CHG/CHH: r=0.373, p<0.0001). For CG, CHG, and CHH 72, 63, and 101 transposable elements respectively were differentially methylation with 48, 53, and 71 having increased methylation in the offspring of damaged plants, and 24, 10, and 30 having decreased methylation (Figure 3.2). In 29/34 cases in which a TE was differentially methylated for multiple sequence contexts, the direction of change was the same between contexts. Standard least squares suggested that there was a significant effect of transposon family on the change of CG methylation (DF=15, SS=0.48, F=1.84, p=0.026) in the offspring of wounded plants, but not CHG or CHH methylation.

The offspring of wounded plants had on average slightly but significantly lower CG TE methylation (OD: 71.5%, OC: 71.7%, t-ratio=3.13, p=0.0018), and a highly significant increase in standard deviation of methylation between individuals (OD:0.041, OC:0.024, t=-24.81, p<0.0001) . After performing an arcsin square root transformation (angular transformation) further analyses were performed on the transformed TE methylation levels. Least squares regression identified significant negative first and second order effects of average methylation on change in variance in the progeny of damaged plants (F=113.4, p<0.0001, Figure 3.3a). This leads to maximum predicted increases in variation at mean methylation levels of 42%, while more highly and lowly methylated TEs tend to have similar levels of variation between the offspring of wounded and control plants. Additionally, lowly methylated TEs tend to have decreased methylation, while highly

methylated TEs have similar or slightly higher methylation in the offspring of damaged plants (F=54.13, p<0.0001, Figure 3.3B). Lastly, overall standard deviation of methylation had a significant negative effect on the change of direction of total methylation in the progeny of damaged plants (F=75.54, p<0.0001, Figure 3.3C). TEs with more variable methylation tended to decrease in methylation in the offspring of wounded plants, while those with less variable methylation tended to increase in methylation.

For CHG methylation, the offspring of damaged plants had slightly higher mean TE methylation (OD: 37.3%, OC: 36.9%, t= 6.53, p<0.0001), and much higher variation between (OD: 0.047, OC: 0.031, t=25.85, p<0.0001). For CHH methylation there was not a significant change in either mean (OD: 14.1%, OC: 14.0%, t=1.69, p=0.09) or variation (OD: 0.0204, OC: 0.0197, t=1.78, p=0.075) of TE methylation between the offspring of damaged and undamaged plants. There was not a significant relationship between average methylation and change in direction of methylation or change in variation of methylation in response to damage for either CHH or CHG methylation. TE family had a significant effect on change of direction for CG methylation (p<0.001), but not on either CHG or CHH methylation.

Differentially Methylated Regions

We identified 600 CG differentially methylated regions (DMRs), 304 CHG DMRs, and 8,877 CHH DMRs between the offspring of damaged and control individuals. DMRs ranged in size from 20-713bp and contained between 6 and 198 cytosines in the context considered. Within these regions percent methylation changes between the offspring of damaged and control plants was 22.8% for CG, 23.4% for CHG, and 17.8% for CHH. Both CG

and CHG DMRs were more likely to have increased methylation in the offspring of damaged (63.1% for CG and 69.1% for CHG) plants, but a similar amount of CHH DMRs were up and down-regulated in the offspring of damaged plants (49.3% up-regulated). Of the 600 CG DMRs, 213 overlapped with gene bodies (35.5%), as did 162 CHG DMRs (53.3%), and 2,316 of 8,877 CHH DMRs (26.1%). Additionally, 153 CG DMRs, 132 CHG DMRs, and 1,614 CHH DMRs fell within 1kb upstream of a gene, a hot spot of epigenetic regulation occurs.

12.2% (26/213) of the genes that contained a CG DMR within their gene body were differentially expressed, significantly more than the 7.5% of differentially expressed used genome wide (Table 3.1). Genes overlapping CHH DMRs, up-stream CG and up-stream CHH DMRs were also differentially expressed more frequently than expected by chance (Table 3.1). These results provide evidence that the inheritance of altered methylation states is associated with coinciding differential gene expression, but do not confirm or reject prior hypotheses regarding the role of direction of differential methylation in gene regulation. To test these hypotheses genes were classified by the direction of change in both methylation and gene expression.

There was not a strong relationship between direction of methylation change in gene body CG DMRs and direction of differential expression. However, CG DMRs primarily located within exons of DE genes tended to have increased expression in the offspring of wounded plants (9 of 12 cases), while DMRs within introns tended to be down-regulated (Figure 3.4.a, 8 of 12). Differential CG methylation in up-stream regions was associated with differential gene expression when there was a methylation increase in the offspring of damaged plants, but not a decrease in up-stream methylation. In 9 of 16 cases increased up-stream CG methylation coincided with decreased gene expression, not significantly

more than the 7 that had increased expression. For both up-stream and gene body CG DMRs methylation tended to be slightly lower than genome-wide methylation levels (63%) in the group with higher methylation and get further more depleted in the group with lower methylation to 40%. These patterns hold up whether the methylation levels increased or decreased in the offspring of damaged plants.

For both gene body and up-stream CHH DMRs there was an overabundance of regions with increased methylation and decreased gene expression in the offspring of damaged plants (Table 3.2). This is predicted by prior gene methylation modeling results, and supports the hypothesis that CHH methylation is associated with transcriptional gene silencing. Considering DE genes overlapping a CHH DMR, 77 of 206 genes were up-methylated and down-expressed compared to the 49 expected if all classes were present at background levels (Chi-Square=17.546, p=0.0015, Chi-Square contribution of 15.4 by up-methylated, down-expressed class). These DMRs primarily overlapped with exon regions, but the pattern of increased methylation coinciding with decreased expression was detectable in intron regions as well (Figure 3.4.b). Considering DE genes that contained CHH DMRs within upstream regions, 56 of 147 genes fell into this up-methylation down expression class, also significantly more than the 34 expected by chance (Chi-Square=15.09, p=0.0045, Chi-Square contribution of 13.6 by up-methylated, down-expressed class). Within gene body CHH DMRs for which there was increased methylation in the offspring of damaged plants we see that the mean methylation increases on average from 26% to 44%. Interestingly, this is significantly higher than the background average of 12% gene body CHH methylation across genes as a whole (p<0.001). Up-stream CHH DMRs

show a similar pattern wherein methylation increases from an average of 27% to 44% in the offspring of damaged plants.

Genes overlapping CG DMRs were enriched for the GO terms growth and nucleotide binding were the two most significantly enriched classes of genes (Figure 3.5.b). CG DMRs within upstream were enriched for enzyme regulator activity and other GO terms (Figure 3.5.a). CHH DMRs were enriched for cell cycle, cytoskeleton, and DNA metabolic process (Figure 3.5.d), and when considering only up-methylated gene body CHH DMRs, the most significantly enriched GO term was response to abiotic stimulus (Figure 3.5.e, p=0.00076), followed by transportation and catalytic activity. Up-stream CHH DMRs were not enriched for any class of GO term. Gene body CHG DMRs were enriched for GO terms transport and abiotic stimulus (Figure 3.5.c).

KEGG mapping identified numerous metabolic pathways that contained a high number of differentially methylated genes for various classes. Of the 5 candidate pathways identified during our prior transgenerational gene expression work in this system(COLICCHIO *et al.* 2015b), four of them were found to contain a substantial number of genes with differential methylation matching the previously observed patterns of differential gene expression (COLICCHIO *et al.* 2015b). The same enzymes that were differentially expressed and involved in the production of jasmonic acid, ethylene, and GA synthesis were also differentially CG methylated. Additionally, 6 enzymes involved in phenylpropanoid metabolism were differentially CHH methylated (5 up-methylated, 1 down-methylaed), as were 20 involved in starch and sucrose metabolism, with many of these involved in the synthesis of pectin and xyloglucans, vital cell wall components, that are also differentially expressed.

Transcription factor *Mimulus guttatus MYB Mixta-like 8 (MgMYBML8)*, was originally

identified in a qPCR candidate gene screen as a potential regulator of transgenerational

trichome induction (SCOVILLE *et al.* 2011), and is yet again found to be differentially

regulated due to parental wounding. In the offspring of damaged plants there is both a 34

nucleotide DMR containing 4 CHG cytosines with increased methylation (18%) and a 106

nucleotide region containing 16 CHH cytosines with increased methylation (14.5%) within

the Mixta-8 gene body. Both of these forms of differential methylation predict decreased

gene expression, the same pattern observed in both q-PCR and RNA-seq experiments

across numerous lines, separate grow ups, and different developmental stages.

Discussion

In response to parental wounding a signal is produced that leads to altered gene

expression and phenotypes in the offspring. Here we provide evidence for the epigenetic

basis of this transgenerational response, as well as an increase in variation in the offspring

of wounded individuals. By performing whole genome bisulfite sequencing we were able to

gauge the magnitude and scale of transgenerational epigenetic inheritance. Using a sliding

window approach we were able to identify differentially methylated regions without

regard to *a priori* expectations. We find relatively small regions of targeted differential

methylation, and a general trend towards increased methylation variation. Both of these

findings parallel prior gene expression results: and get us one step closer to deciphering

the mechanism through which parental environment affects offspring development. In the

offspring of damaged plants gene expression was also more variable across individuals,

and genes that were previously identified as being down regulated were found here to have overlapping repressive methylation marks.

While all forms of cytosine methylation play a role in the controlling gene expression and chromatin state in plants, the context of a methylated cytosine greatly affects its role in gene regulation (COLICCHIO *et al.* 2015a). Determining the role of CG, CHG, and CHH methylation in mediating transgenerational inheritance was a major goal of this study, and the results presented above shed new light upon the flexibility and variation of these different classes of methylation. Along with altering gene expression, DNA methylation also plays a large role in the control and suppression of transposable elements. The effect of environmental stresses on altering the regulation of transposable elements is well known, but the persistence of these changes between generations is still poorly understood. Here we generate a snapshot of the epigenetic regulation of transposable elements one generation removed from leaf damage and find that there are still small but significant epigenetic changes due to parental environment.

Variance Increase

While some genomic regions had an increase or decrease in methylation in the offspring of wounded plants, many more simply had increased variation in response to parental damage. This genome wide trend of increased variation in the offspring of damage plants was most obvious in the CG methylation context, where within treatment variation increased by over 22% across gene bodies. Previous gene expression work identified a coinciding increase in gene expression variation in the progeny of damaged plants. Studies in asexual dandelions previously demonstrated that the offspring of plants

exposed to a salt stress as well as a number of plant hormones express a greater degree of epigenetic diversity than the progeny of control plants (VERHOEVEN *et al.* 2010). Taken together these results provide substantial evidence that the offspring of plants exposed to stressful environments may exhibit more epigenetic, gene expression, and potentially phenotypic variation than the offspring of plants grown in controlled environments. In our case we find that rather than a loss of epigenetic regulation of a set of highly variable genes, there is a general increase in variability that suggests that epigenetic deregulation is a genome-wide trend. The evolutionary significant of this increase in variation in response to environmental perturbations is unclear, but it could be that through increasing epigenetic variation, environmental stressors facilitate rapid evolution (ROBERTSON and RICHARDS 2015).

Transposable Elements

In the offspring of damaged plants we identify substantial shifts in the methylation of a large number of transposable elements. Our results suggest that many transposable elements are significantly up-methylated in all three contexts, and that TE CG methylation has a number of nuanced shifts in response to parental wounding. Moderately CG methylated genes showed a significant increase in methylation variation in the offspring of wounded plants, and TEs with more variable methylation tended to decrease in methylation in the offspring of wounded plants, while those with less variable methylation tended to increase in methylation. Additionally, there was a significant effect of transposon family on the direction change in TE CG methylation. Non-CG methylation did not show any of these patterns, but did show general increases in TE methylation. This explains why our

general linear model identified significantly (p<0.05) more CG up-methylated than down-methylated TEs even though average TE CG methylation was higher in the offspring of control plants.

While CG methylation is the most stably transmitted form, recent results suggest that non-CG methylation is the first responder in reprogramming disrupted methylation states (KORTH and DIXON 1997; LAW and JACOBSEN 2010; STROUD et al. 2014). We hypothesize that in response to damage, there is a loss of CG methylation in lowly and moderately methylated TEs. Through a negative feedback loop this would likely lead to increased transcription of these TEs, leading to the activation of RdDM silencing mechanisms and increased non-CG methylation at these loci (MARÍ-ORDÓÑEZ et al. 2013; NUTHIKATTU et al. 2013). Overtime, this should lead to the return of full CG methylation, but this process may take numerous generations. We hypothesize that the high level of CG methylation variation in the offspring of damaged plants represents an intermediate step in the recovery of CG methylation at these initially epigenetically released transposable elements.

Differentially Methylated Regions

The discovery of hundreds of regions with differential CG and CHG methylation, and thousands of regions with differential CHH methylation demonstrates that parental environment leads to targeted epigenetic changes in the next generation. The size of these regions averaged between 177 basepairs, with a median size of 135. Potentially meaningfully, this size range is similar to the approximately 147 basepairs that wrap around a histone octamer to form a nucleosome. It could be that targeted histone

modifications in response to wounding trigger the methylation or demethylation of associated DNA. The finding that there are over 20 times as many CHH DMRs as either CHG or CG DMRs suggests that non-symmetrical DNA methylation is much more plastic to the environment than more stable CG or CHG methylation. Most methylation changes in *Arabidopsis thaliana* plants exposed to bacterial pathogens were also found to occur in a CHH context, and compared to other forms of differential methylation, CHH methylation DMRs varied the most between different classes of stressors (Dowen *et al.* 2012). Additionally, differential CHH methylation has recently been shown to be the primary form of methylation to vary across plant cell and tissue types (Kawakatsu *et al.* 2016). These results suggest that CHH methylation is the most flexible for of cytosine methylation in an evolutionary, developmental, and transgenerational context.

Gene Associated DMRs

Approximately 35% of CG, 53% of CHG DMRs and 27% of CHH DMRs overlapped with gene bodies, with thousands of other DMRs located within 1kb upstream of genes. These DMRs located in, or near, genic regions provide us with obvious candidates as the loci up-stream of the observed patterns of differential gene expression. We tested this hypothesis by performing a series of chi-square analyses on the overlap between our sets of differentially methylated and differentially expressed genes. Genes with either CG or CHH DMRs overlapping or up-stream of them were more likely to be differentially expressed than genes without nearby DMRs. The increase in likelihood of being classified as differentially expressed due to proximity to a DMR ranged from 20% (Gene body CHH) to 80% (Up-Stream CG). Gene body CHG overlapping DMRs were 47% more likely to be

differentially expressed than background levels, but due to the small sample size (only 162 gene body CHG DMRs), this was not found to be significant (p=0.089). While a high number of CHG DMRs overlapped with differentially expressed genes, they did not overlap significantly more frequently than expected by chance. Still, there was a trend towards increased DE in the offspring of CHG DMR genes, and it is quite likely that CHG DMRs also play a role in the observed transgenerational response.

The overlap between differentially expressed and methylated genes is notable given that this contrast is "underpowered" for two important reasons. First, different tissue, individuals, and experimental grow-ups were used for the DNA methylation and gene expression experiments. While the same recombinant inbred line was used for both analyses, whole seedling tissue was used for differential expression testing, while only leaf tissue was used for the differential methylation analysis. This was done to reduce the variation in methylome profiles within the tissue used for methylome analysis, but may have also lead to us missing regions of the genome that are only differentially regulated in stem or meristem tissue. Additionally, both parent and progeny individuals were unique between experiments. Because of this, any differential regulation in either parent or progeny generation that was due to treatment*environment or treatment*individual effects would only be identified in one experiment but not the other.

While the use of different individuals and experimental generations limits the power of this experiment to predict patterns of gene expression based on methylation data, it also adds credence to the hypothesis that genes found differentially regulated in both experiments are responsive to parental wounding and not other environmental factors. Because of this, the 447 differentially expressed genes for which there was also found to be

an overlapping DMR represent prime candidates as genes that show relatively consistent differential regulation in the generation following parental wounding.

Although our ability to predict gene expression patterns from methylation patterns is lost through this experimental design, within the hundreds of genes that were differentially expressed and methylated the predicted patterns relating methylation and gene expression hold up. In particular we find a highly significant overabundance of genes for which there was an up-stream or gene body increase in CHH methylation, and a decrease in gene expression. The link between CHH methylation and RdDM or repressive histone modifications is an obvious explanation for this finding. The result that CHH methylation in these regions were significantly higher than genome wide average levels even in the offspring of control plants, suggests that these regions experience increased silencing in the offspring of damaged plants, but that these regions may be predisposed to CHH based transcriptional gene silencing.

For gene body CG DMRs, the direction of methylation was not consistently associated with a specific direction of gene expression change, but rather a general trend toward differential expression. A likely explanation for this finding stems from the complex relationship between gene body methylation and gene expression. In our prior analysis of DNA methylation in *M. guttatus* we identified that along with significant first, second, and third order gene body CG methylation effects on gene expression, it also interacted with 6 other terms in our model to alter gene expression. In future studies with the same plants analyzed for both gene expression and methylation in response to various stressors, we will be able to test these complex predictive patterns, but here we can only

postulate that the lack of clear direction for CG methylation's effect on gene expression is due to this complexity.

Up-stream CG methylation within genes tended to have a significant effect on gene expression only when there was an increase in methylation in the offspring of wounding. When this was the case, we found an overabundance of both up and down regulated genes. Previous genome wide scans have repeatedly identified that up-stream CG methylation has a negative effect on gene expression, so this result is a bit surprising (KILBY *et al.* 1992; METTE *et al.* 2000). It could be that in some cases elevated up-stream CG methylation prevents the binding of repressive transcription factors, or that the observed instances of up-methylated DMRs and elevated gene expression are highly variable regulatory regions which were altered in opposite directions in these two experiments.

Genes and Pathways of Interest

The distinct ontology terms enriched in our various sets of differentially methylated genes suggest that different classes of DNA methylation play different roles in mediating the transgenerational response. The enrichment of nucleotide binding proteins in the set of genes overlapping CG DMRs suggests that these genes may have a role in the regulation of up-stream players in the transgenerational response pathways. Additionally, enzymes involved in the synthesis of Jasmonic Acid, Ethylene, and Gibberellin that were previously identified as differentially expressed in the offspring of damaged plants, also show up here as containing gene body CG DMRs. Spermine synthase acts to shunt S-adenosyl-L-methionine out of the Yang cycle directly upstream of the synthesis of 1-aminocyclopropane-1-carboxylate, the precursor to ethylene (YANG and HOFFMAN 1984).

This gene was significantly down regulated in the offspring of damage plants, and here we provide evidence for co-occuring decreased gene body CC methylation. The decreased expression of spermine synthase should lead to elevated levels of ethylene synthesis in the progeny of damaged plants; in turn alter their physiology and gene expression in a host of ways (BLEECKER and KENDE 2000; ANDERSON *et al.* 2004).

The synthesis of jasmonic acid begins with the hydrolysis of a phospholipid to form alpha-linoleic acid (YAN *et al.* 2013). Phospholipase A1 is one of two enzymes capable of synthesizing this reaction (ISHIGURO *et al.* 2001), and is up-regulated in the offspring of damaged *M guttatus* plants. Here we find that this gene also has increased methylation in the offspring of damaged plants, potentially leading to it's increased expression, and potentially an increase in the conversion of linoleic acid into jasmonic acid. Lastly, 3-beta-dioxygenase has increased gene body CG as well as CHH methylation in the offspring of damaged plants along with decreased gene expression. This gene converts the inactive gibberelin precursors GA9 and GA20 to bioactive GA1 and GA4 (LANGE *et al.* 1997). These three plant hormones play a substantial role in regulating numerous developmental processes as well as biotic and abiotic stress responses. Through the differential methylation of these three plant hormones, hundreds of other genes could have their expression altered through the complex network of transcription factors controlled by these hormones. The differential CG methylation of nucleotide binding proteins along with enzymes involved in hormone synthesis suggests that differential CG methylation may represent an up-stream facilitator of transgenerational phenotypic plasticity.

In contrast, differential gene body CHH methylation appears to function in either a more specialized or downstream capacity within the transgenerational response to

parental wounding. Rather than being enriched for nucleotide binding proteins and hormone synthesizing enzymes, differentially CHH methylated tend to be involved in stress response, the cytoskeleton, cellular transport, starch and sucrose metabolism, and phenylpropanoid production.   Numerous intriguing candidate genes exhibit the most common pattern, increased CHH methylation and decreased gene expression, providing a glimpse into the role of CHH methylation in transgenerational plasticity.

The decreased expression of *Mimulus guttatus MYB Mixta-like 8 (MgMYBML8)* in the offspring of wounded plants has been shown repeatedly across methods (qPCR and RNA-seq), experimental generations, and tissue types. Here we find evidence that regions of *MgMYBML8* have elevated CHH and CHG methylation in the offspring of damaged plants. MIXTA–like genes have been implicated as positive regulators of trichome production in *Antirrhinum,* and have been shown to be both positive and negative regulators of trichome development in *Arabidopsis* and cotton. Increases in CHH and CHG gene body methylation suggest that RdDM and coinciding transcriptional silencing of *MgMYBML8* may play a role in the increased trichome density in the offspring of damaged plants.

We also find this same pattern on increased CHH methylation and decreased gene expression in previously identified candidate genes involved in cell wall synthesis, breakdown, and rearrangement (a xyloglucan endotransglucosylase hydrolase, beta-glucosidase, and pectin methyltransferase) and secondary compound metabolism (aspartic proteinase 1 and cinnamyl dehydrogenase). In Arabidopsis, differential CG methylation was found relatively consistent across a host of pathogens, yet CHH methylation was heavily pathogen dependent (DOWEN *et al.* 2012). Additionally, CHH methylation varies the most across populations of *Arabidopsis* (SCHMITZ *et al.* 2013b) and across the diversity of

angiosperms (NIEDERHUTH *et al.* 2016), while CG methylation is much more universally conserved (TAKUNO and GAUT 2013). Our findings support the more flexible role of CHH methylation in regulating the downstream outcomes and specific developmental shifts due to parental environment, while CG methylation appears to function primarily in the regulation of more conserved up-stream pathways.

Conclusions

This transmission of environmentally responsive epigenetic markings between generations represents a mechanism through which outside information can be integrated into the genome and alter gene expression for some length of time after the initial signal recedes. Here we assay DNA methylation variation dependent on parental environment, and find evidence that it plays a role in linking parental environment with altered offspring gene expression. The association between differentially methylated regions and nearby differentially expressed genes strengthens the hypothesis that differential methylation plays a role in mediating transgenerational inheritance and sheds light onto the contrasting roles of CG and non-CG methylation in transgenerational inheritance. Additionally, genome wide increases in methylome variation and the differential methylation of certain TE classes suggests that aside from targeted differential gene regulation, parental conditions can alter an organisms epigenetic profile in a host of other ways. Of particular evolutionary interest, increased epigenetic diversity may be an unattended side effect of a stressful environment, but may also increase the speed through which plants can adapt to rapidly changing environments.

Table 3.1. Chi-Square contingency tables of differentially expressed genes and those overlapping differentially methylated regions.

Gene Body CG					
	Sig DE	NS DE		Chi-Sq	p-value
Sig Meth	**26**	187	12.20%	6.54	0.0105
NS Meth	1852	22690	7.50%		

Up-Stream CG					
	Sig DE	NS DE		Chi-Sq	p-value
Sig Meth	**21**	132	13.70%	8.28	0.004
NS Meth	1857	22751	7.52%		

Gene Body CHG					
	Sig DE	NS DE		Chi-Sq	p-value
Sig Meth	18	144	11.10%	2.882	0.089
NS Meth	1861	22732	7.56%		

Up-Stream CHG					
	Sig DE	NS DE		Chi-Sq	p-value
Sig Meth	13	119	9.80%	0.97	0.33
NS Meth	1866	22757	7.58%		

Gene Body CHH					
	Sig DE	NS DE		Chi-Sq	p-value
Sig Meth	**206**	2110	8.90%	6.19	0.013
NS Meth	1673	20766	7.40%		

Up-Stream CHH					
	Sig DE	NS DE		Chi-Sq	p-value
Sig Meth	**147**	1467	9.10%	5.67	0.018
NS Meth	1732	21409	7.48%		

Table 3.2. Chi-Square contingency tables addressing the relationship of direction of change in gene expression, with direction of change in methylation for overlapping DMRs.

Gene Body CG						
		Up	Down	NS	Chi-Sq	p-value
Meth	Up	7	8	105	7.31	0.12
	Down	6	5	82		
	NS	824	1028	22690		

Gene Body CHH						
		Up	Down	NS	Chi-Sq	p-value
Meth	Up	43	**77**	1054	17.48	0.0015
	Down	37	49	1056		
	X	757	916	20766		

Up-Stream CG						
		Up	Down	NS	Chi-Sq.	p-value
Meth	Up	**7**	**9**	75	14.33	0.006
	Down	1	4	57		
	NS	829	1029	22744		

Up-Stream CHH						
		Up	Down	NS	Chi-Sq	P-Valu
Meth	Up	28	**56**	733	15.0871	0.00452
	Down	30	33	734		
	NS	779	953	21409		

Figure 3.1. Density plot of gene-by-gene variation within the of damaged vs. control individuals. A) CG Methylation standard deviation in the damage progeny group as a function of control standard deviation for the same gene. B) CHG, C) CHH, D) Standard deviation divided by mean gene expression in the damage progeny group as a function of control variation for the same gene. Shifts in all four above the 1:1 line demonstrate a general increase in variance in the progeny of wounded plants.
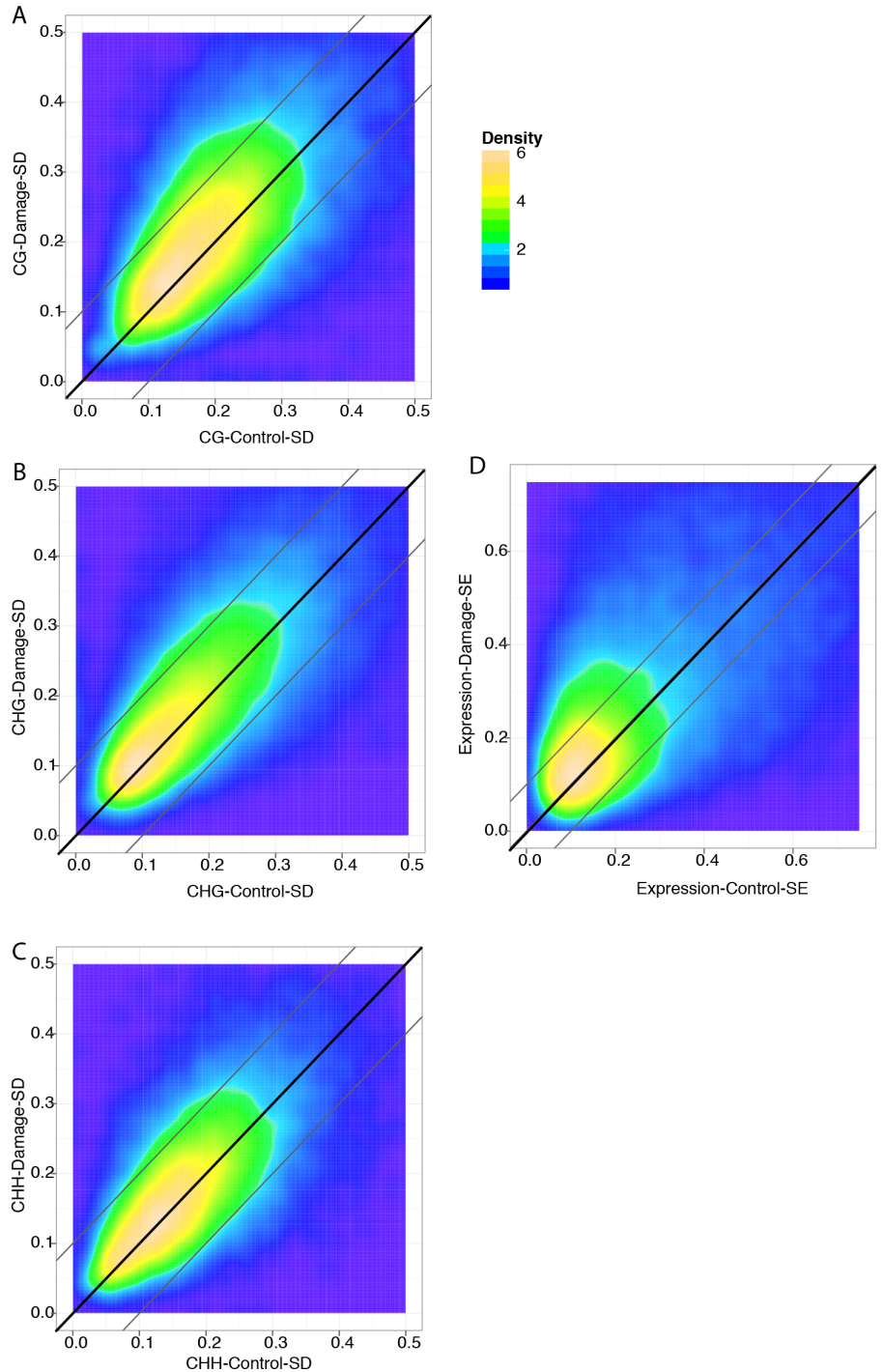
Figure 3.2. Differentially methylated transposable elements. Green numbers represent transposable elements in which methylation for the given class increases significantly (p<0.05) in the offspring of wounded plants. The first number in overlapping circles represents the number of TEs that changes in the same direction in response to damaged, and the second number represents those that changes in opposite directions.
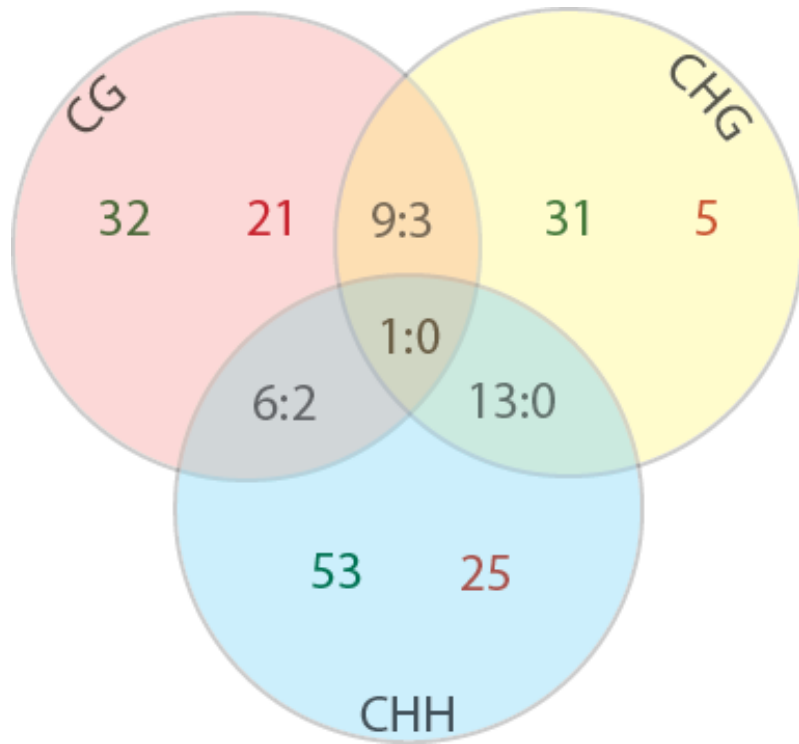
Figure 3. 3. Patterns of differential TE CG methylation in the offspring of damaged plants. Numbers given on all axes represent angular angular transformed values. A) Significant negative first and second order effects of average methylation on change in variance. Moderately methylated TEs had increased variability in the progeny of wounded plants, while this was not seen for lowly or highly methylated TEs. B) More highly methylated TEs tended to have increased CG methylation in the offspring of damaged plants, but lowly methylated TEs tended to have decreased methylation. TEs with highly variabile methylation tended to have lower methylation in the offspring of damaged plants.
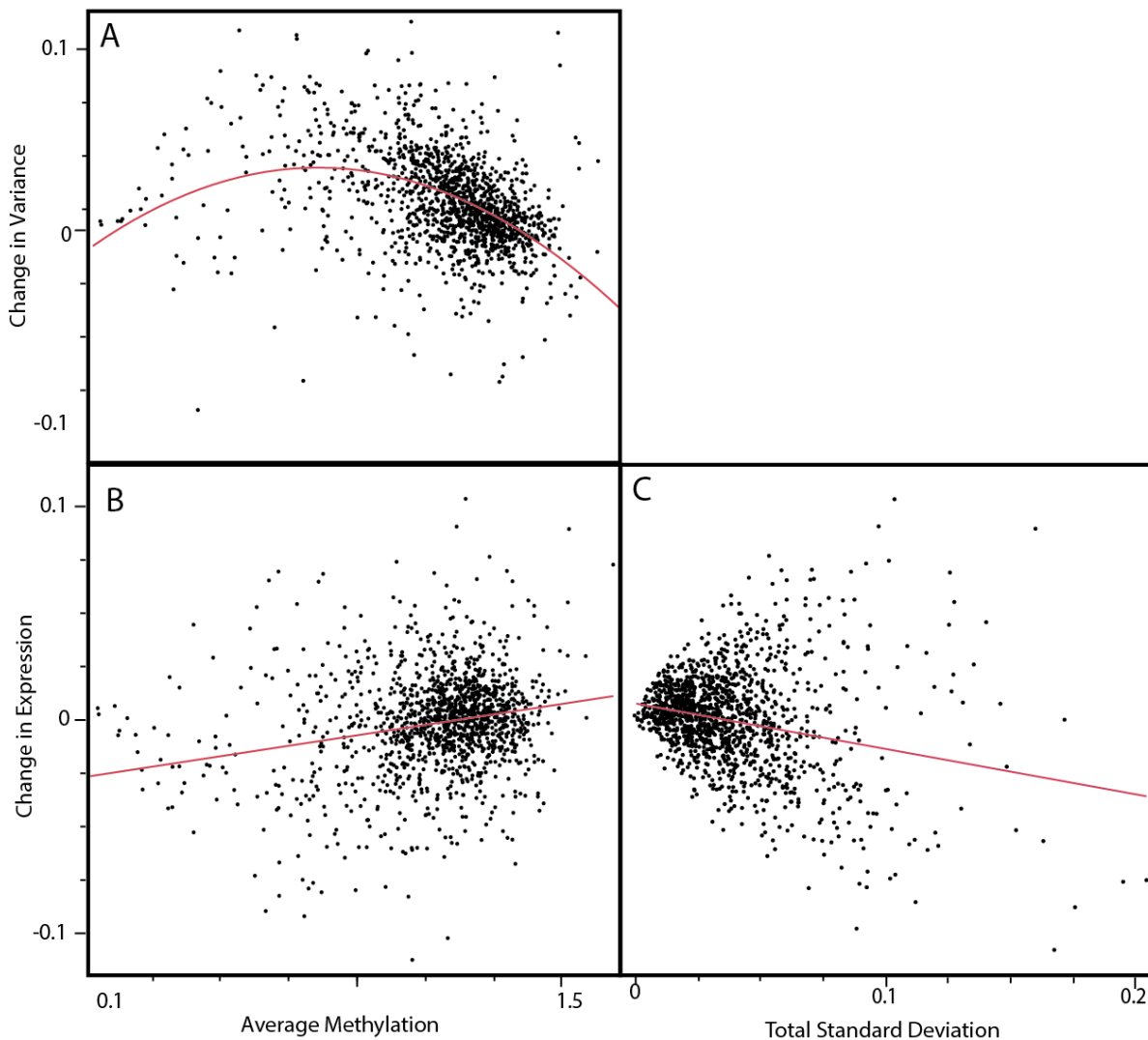
Figure 3.4. Heatmap of differentially methylated regions overlapping differentially expressed genes. A) CG DMRs overlapping differentially expressed genes, separated by whether the majority of the DMR overlaps with an exon or intron sequence. B) CHH DMRs, separated in the same manner as part 3.4.A.
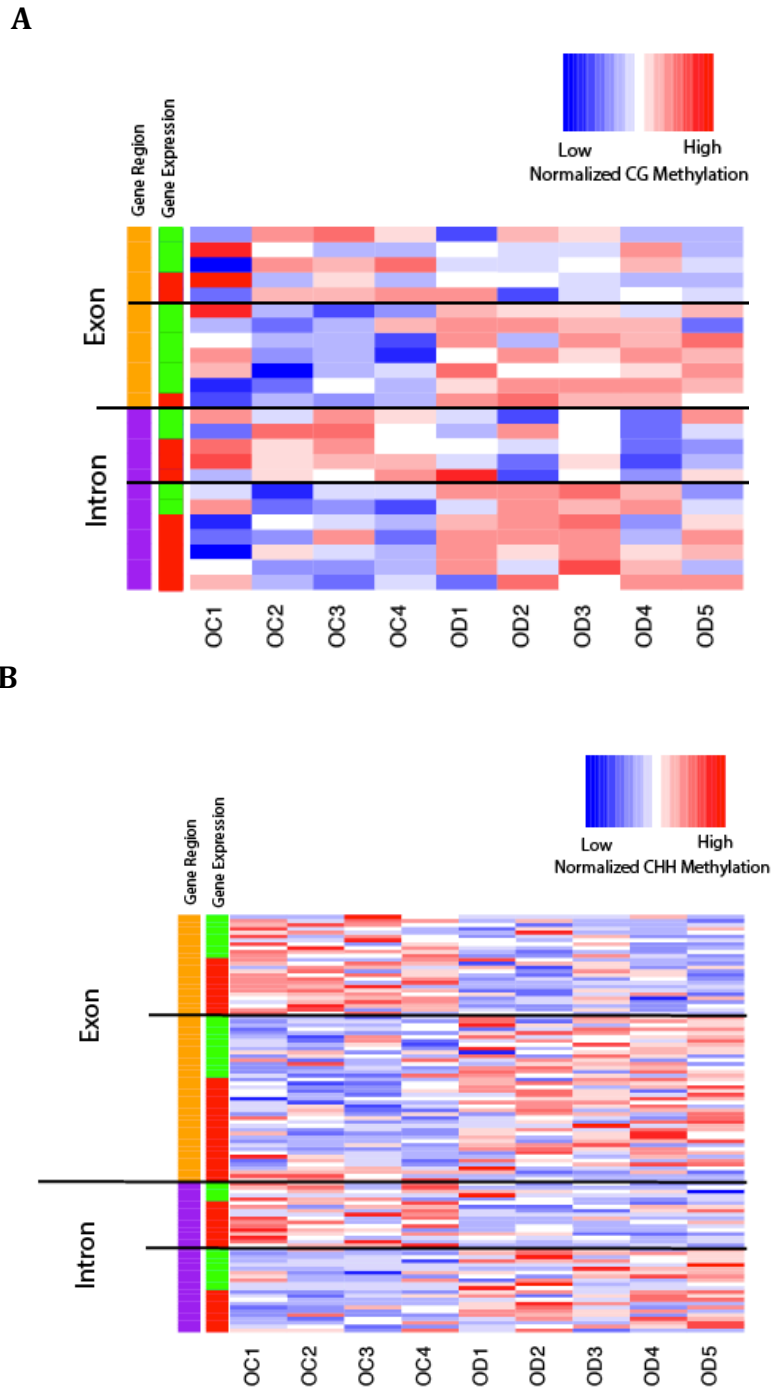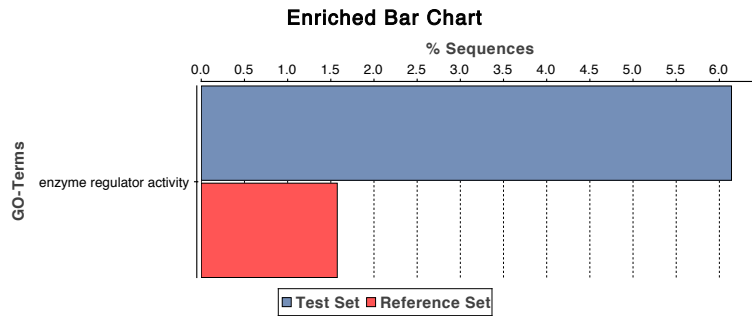
Figure 3.5. Enriched GO term in sets of genes overlapping various classes of DMRs. A) Up-stream CG DMRs, B) Gene body CG DMRs, C) Gene body CHG DMRs, D) Gene body CHH DMRs, and E) Gene body CHH DMRs of increased methylation in the progeny of damaged plants.
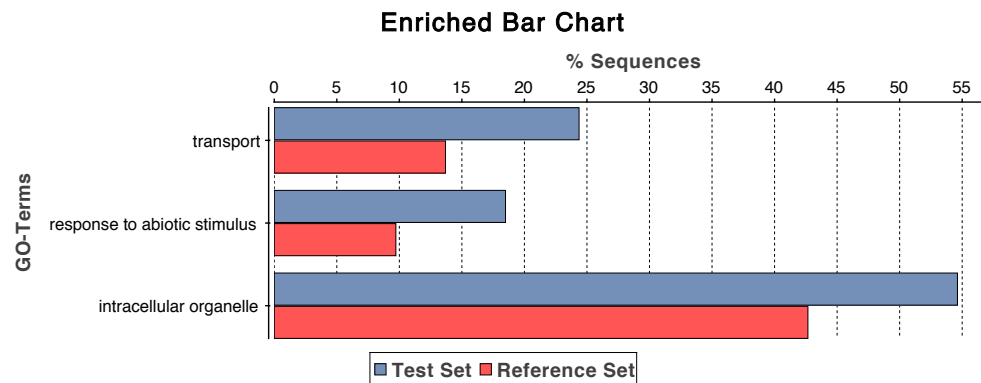
A

B

C

D

Enriched Bar Chart

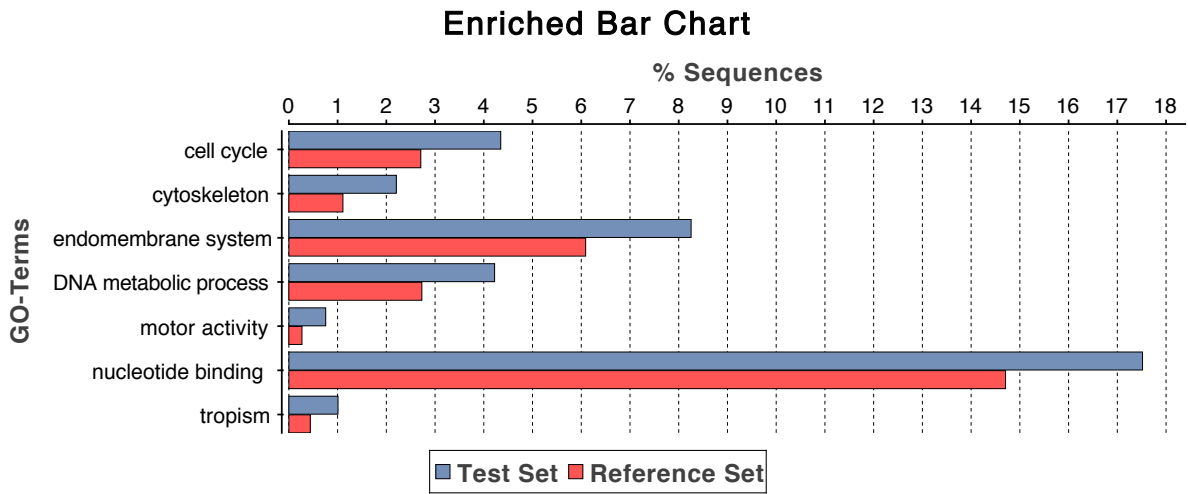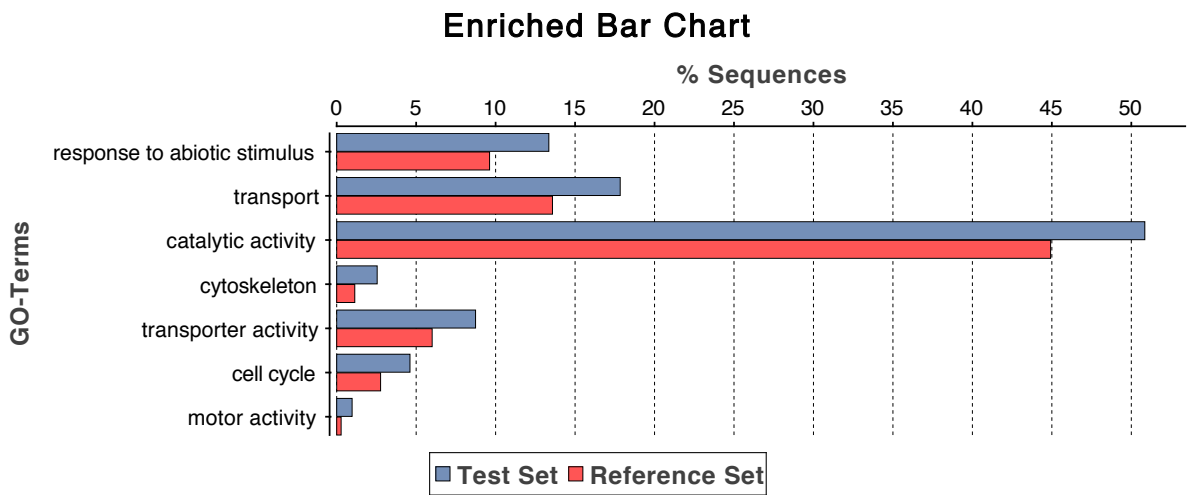**% Sequences**

E

Enriched Bar Chart

% Sequences
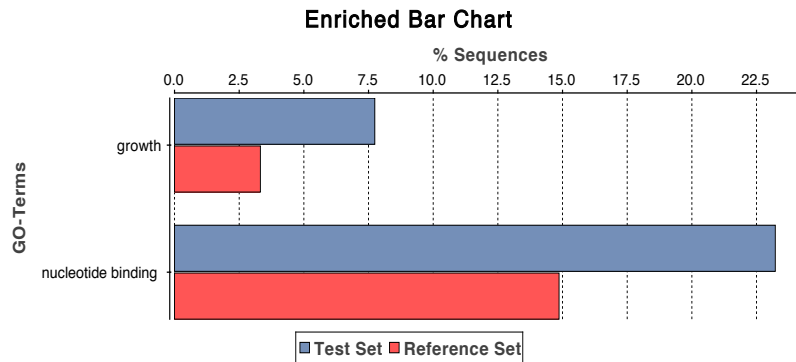
CHAPTER 4

Transgenerational Effects Alter Plant Defense and Resistance in Nature

Abstract

Trichomes, or leaf hairs, are epidermal extensions that take a variety of forms and perform many functions in plants, including herbivore defense. In this study, I document genetic variation in trichome density, within-generation plasticity, and a direct role in herbivore defense for *Mimulus guttatus*. After establishing the relationship between trichomes and herbivory, I test for transgenerational effects on trichome density and herbivore resistance. Variation in herbivore density and the high cost of plant defense makes plant-herbivore interactions a system in which transgenerational phenotypic plasticity (TPP) seems apt to evolve. Here, I demonstrate that parental wounding increases trichome density and reduces herbivory in the offspring of damaged plants in natural populations. Moreover, this response varies between populations. This is among the first studies to demonstrate that TPP contributes to variation in nature, and also suggests that selection can modify TPP in response to local conditions.

Introduction

Through common garden experiments, Clausen, Keck, and Heisey demonstrated the role of environmental and genetic factors on phenotypic and fitness variation in nature (CLAUSEN *et al.* 1948). By exposing plants to a variety of common conditions they were able to detect not only a role of stable genetic and general environmental factors on plant growth, but also that there is natural variation in how plants respond to environmental conditions. This ability to alter one's phenotype in response to environmental cues is labeled "phenotypic plasticity", and the molecular mechanisms and adaptive significance of this have been demonstrated across a broad array of circumstances. When an environmental cue is a reliable predictor of future ecological conditions, phenotypic plasticity is advantageous (WEST-EBERHARD 1989; AGRAWAL 2001; HERMAN *et al.* 2013a; KUIJPER and HOYLE 2015).

Following the hypotheses regarding with-in generation plasticity, if current environmental conditions are a good predictor of the conditions experienced by the next generation (positive autocorrelation), then the transmission of altered developmental trajectories between generations (transgenerational phenotypic plasticity, TPP) should also be adaptive (HERMAN and SULTAN 2011; JABLONKA 2012; HERMAN *et al.* 2013a; LEIMAR and MCNAMARA 2015). For example, if yearly mean temperature patterns exhibit positive autocorrelation, selection should favor genotypes of plants that, when exposed to particularly warm environments, produce offspring transgenerationally primed for another warm season. However, if environmental conditions tend to cycle rapidly then the conditions experienced in one generation will likely be very different than those experienced in the next generation (negative autocorrelation). In this case

transgenerational effects are not expected to evolve, or they are expected to evolve in the negative direction. If yearly mean temperature patterns exhibited negative inter-annual autocorrelation, selection would favor genotypes of plants that when exposed to warm conditions produce offspring primed for a cool season.

Since the demonstration of transgenerationally induced defenses in wild radish (AGRAWAL *et al.* 1999), herbivore response has become a model system for studying TPP in plants. Numerous studies have demonstrated that the offspring of wounded plants produce more chemical and physical defenses than the offspring of unwounded plants. In environments where herbivore activity in one generation has a positive autocorrelation with herbivore activity in the next generation, TPP to herbivory should be adaptive and this type of transgenerational induction should evolve. On the other hand, if there is no autocorrelation in herbivore abundance this form of transgenerational plasticity is not expected to evolve, and if high herbivore years tend to be followed by low herbivore years (negative autocorrelation) than it would be beneficial for the offspring of wounded plants to expended less energy on herbivore defense than the offspring of unwounded plants.

The transcriptional basis of TPP (COLICCHIO *et al.* 2015b), its epigenetic origin (BOYKO *et al.* 2010; LANG-MLADEK *et al.* 2010; VERHOEVEN *et al.* 2010; CALARCO *et al.* 2012; RASMANN *et al.* 2012; HERRERA and BAZAGA 2013), and taxonomic prevalence (HOLESKI *et al.* 2012) have been major foci of herbivory driven TPP research. In *Mimulus guttatus* (yellow monkeyflower)*,* simulated herbivory (leaf wounding) of parental plants produces offspring with more trichomes—defensive hair-like epidermal structures—than offspring of undamaged parental plants (HOLESKI 2007). The offspring of damaged plants differ from the offspring of control plants via a broad, multifaceted transcriptional response (COLICCHIO *et*

al. 2015b). Molecular epigeneticists have recently demonstrated the presence of an intricate three part system through which histone modifications (GREENBERG *et al.* 2013), DNA methylation (MATZKE and MOSHER 2014), and small-RNAs (SUNKAR *et al.* 2007; McCUE *et al.* 2012) are responsive to environmental conditions, alter gene expression, and persist into the following generations(VERHOEVEN and VAN GURP 2012b). These epigenetic mechanisms appear likely to represent the underlying molecular basis of transgenerational plasticity (RASMANN *et al.* 2012).

While examples of transgenerational plasticity and our molecular understanding of TPP continue to build, relatively little is still known about the role of transgenerational plasticity in nature. In the *M. guttatus* system previous studies demonstrated the presence of transgenerational trichome induction, and variation in this response across recombinant inbred lines, and their progenitor inbred lines. However, until this point no one has tested for a role of trichomes in herbivore defense in *M. guttatus,* variation among natural populations in transgenerational plasticity, or a role of parental environment on plant development or fitness components in nature.

In *Campanulastrum americanum* field studies demonstrated that maternal light conditions alter the probability that an offspring will assume a biennial life history strategy, that this is appears to be adaptive, and that there is variation in this response (GALLOWAY and ETTERSON 2007; GALLOWAY and ETTERSON 2009). Recently, researchers studying *Phaseolus lunatus* (wild lima bean) utilized a novel approach to study transgenerational chemical defense induction and the role of this induction on plant survival in nature(BALLHORN *et al.* 2016).   These studies provide complementary insight into transgenerational phenotypic plasticity in nature, and this study expands upon previous

work in the *M. guttatus* and other systems by addressing inter-population variation in transgenerational plasticity in trichome production and herbivore resistance in nature.

Methods

Natural Population Phenotyping, Herbivory assays, and Collection

During the summer of 2014 I identified 16 natural populations of *Mimulus guttatus* within a 150 by 50 mile area in Central Oregon. These sites ranged in elevation from 89 to 1,481 meters (Appendix 13). When over 50% of plants at a given site began setting seed (between June 7th and August 5th), 12-20 plants were collected per site and brought to the Plant Biology lab at HJ Andrews Experimental forest. Here, I assayed herbivory on every leaf of the primary axis on a 0 to 5 ranking (0: no leaf damage, 1: 1-10% leaf area removed, 2: 11-20%, 3: 21-30%, 4: 31-40%, 5: >41%) using a visual estimation of leaf damage that has been found accurate (JOHNSON *et al.* 2015). I also measured plant height and width, and counted the number of flowers produced by each plant. In addition, I counted glandular and non-glandular trichomes from three leaves per plant (one of which was always of the second leaf pair, and two of which were later leaves) as described previously (COLICCHIO *et al.* 2015b). At the end of the growing season (August 10th- September 2nd) I revisited these sites and randomly selected and collected seed from 10 plants per site.

Experimental Greenhouse Generation

In the fall of 2014, I grew seed from 6 maternal lines for eight of the sixteen populations sampled in the field (chosen to represent a wide-range of ecological conditions) at the University of Kansas Greenhouse. Seeds were germinated individually in

1-inch cells, before being transplanted to 2-inch containers for continued growth. I

phenotyped one third of the plants for second leaf trichome density at the third leaf pair

expanded stage. Of the remaining plants, half were subject to wounding through the hole-

punch method from the third through the sixth leaf pair which was previously shown to

induce increases in trichome production in inbred *M. guttatus* lines (HOLESKI 2007;

COLICCHIO *et al.* 2015b). Within each population, I randomly selected one damaged and one

control plant from each maternal line to use as parent plants for the next generation, while

the remainders were phenotyped for 2nd and 7th leaf pair glandular and non-glandular

trichomes. Plants derived from TBR did not continue leaf development to the 7th leaf pair,

and were therefore excluded from the analysis of within generation trichome induction. I

performed crosses using a "circular crossing design" for both damaged and control

breeding individuals (Appendix 14). From this I generated six paired lines for each of the

eight populations, deriving from either the offspring of two control or two damaged plants

(Appendix 14).

Field Common Garden Design

During the summer of 2015, I germinated seeds from each of these populations at

the University of Oregon greenhouse in 1" flats and transplanted one to two week old

seedlings into two common garden sites in the Cascade mountain region of Central Oregon.

At both sites, individuals were planted in a randomized design across the site. One common

garden, HJ Andrews Experimental Forest (HJA), was located >2 mi away from any native *M.*

guttatus population and differed from typical *M. guttatus* habitat in that it contained a

greater composition of soil organic matter and received fewer hours per day of direct

sunlight (personal observations). The other site, Browder Ridge (BR), features a large

native population and has been the site of numerous prior common garden experiments (MOJICA *et al.* 2012; MONNAHAN and KELLY 2015).

At HJA, 1,232 seeds from 44 lines (4 were excluded due to insufficient seed set), each replicated with either damaged or control parents, were planted on April 18th, 799 (64.8%) germinated and were planted into the field on May 12th, with 224 of these plants (28.0%) eventually flowering. 1,056 seeds from these same lines were planted for the BR garden on May 4th, 573 (54.3%) germinated and were transplanted into this field site on May 25th, of these 94 reached flowering (16.4%). During the growing season, rapid dry down due to the drought conditions in the Cascades during the summer of 2015 necessitated the addition of supplemental water at both sites at the rate of 5 gallons per common garden site every other day for 3 weeks during the growing season. Of the surviving 318 plants, 153 were the offspring of damaged parent plants, while 165 were the offspring of control plants. Trichome counts were completed for 271 of 318 plants.

I surveyed plants every other day, and on the day that a plant produced its first flower, the following traits were scored: largest leaf length and width, number of leaves, plant height, node of flower, peduncle length. I assayed herbivory by the same method as described above, and one second node leaf was collected for trichome phenotyping.

Statistical Analysis

Incidence of herbivory, elevation, aridity (ZOMER *et al.* 2008), latitude and longitude were all considered as possible explanatory variables in a standard least squares regression of population trichome density in JMP v10 (SAS Institute Inc., Cary, NC). Significance in this model was determined through t-tests. The linear relationship between

population incidence of herbivory and trichome density was carried forward using linear regression.

Nominal logistic regression was used to determine if trichome densities have a significant effect on a leaf being classified as experiences minor damage compared to moderate/severe damage. An effect likelihood ratio test was used within this model to test for a significant role of trichomes on damage severity.

Standard least squares regression was used to determine the relationship between population, trichome density, a "population x trichome" interaction term, and stem width on plant flower production. Stem width was included in this model as a covariate to help partition out variation due to general plant vigor, and get more directly at the relationship between trichome density and flower production for a plant of a given size. Effect significance was determined by F-ratio tests based on factor sum of squares.

To detect signatures of phenotypic plasticity, 7th leaf log trichome density was considered the response variable with population, family nested within population, wounding treatment, and a "population x family" interaction term considered as explanatory variables. Using a GLM framework in JMP v10 (SAS Institute Inc., Cary, NC) log-ratio chi-square tests were preformed between models to determine the significance of the various terms.

Population of origin, line nested within population, growth environment (field or greenhouse), parental wounding, and all possible two interaction effects between population, parental wounding, and growth environment, as well as their three way interaction were considered in standard least squares model with log-transformed trichome density as the response variable. Using a GLM framework in JMP v10 (SAS

Institute Inc., Cary, NC) log-ratio chi-square tests were preformed between models to determine the significance of the various terms.

Considering plants from which there was at least some sign of insect damage I created a model to test which, if any, factors limit insect herbivore damage. Using a stepwise model selection method in JMP v10 (SAS Institute Inc., Cary, NC), with severity of damage coded as minor, moderate, or high as the response variable. Possible explanatory variables to include were population, site, parental treatment, days to flower, leaf width, plant height, trichomes, treatment*population, site*treatment, site*trichomes, and site*population as possible explanatory factors. I then selected a model using both a minimum BiC and AicC criterion.

Results

Natural Population Surveys

Central Oregon *M. guttatus* exhibit vast natural variation in trichome density and herbivory (Appendix 15). Incidence of herbivory at a given site was the only factor correlated with population mean trichome density (t=3.96, p=0.0027) when considered alongside latitude, longitude, aridity, and elevation. Incidence of herbivory was not significantly correlated with any of the preceding environmental variables (Appendix 16). The significant positive correlation between population incidences of herbivory and trichome density (R^2 = 0.735, Figure 4.2) suggests that herbivory driven natural selection may play a role in population trichome density variation.

Trichome densities were higher in leaves that were classified as having minor herbivore damage (1%-10%) compared to those that experienced moderate to severe

herbivore damage (10%-50%) (n=311, L-R Chi-square: 13.98, p=0.0002, Figure 4.3b).

While leaves that experienced minor herbivory had on average 93 trichomes, those that

experienced moderate to severe damage had only 58 trichomes on average. Considering

the effect of trichome on plant fitness, there was not a general relationship between plant

trichome density and plant flower production (df=1, SS=1.36, F=0.025, p=0.872), but the

"trichome x population" interaction term was found to significantly affect flower

production. Of the populations carried forward for common garden experiments:

trichomes were positively associated with fitness in WC, CSR, IM, and TBR, and negatively

associated with fitness in BR, HOL, LPD, and MWL. This suggests that while trichomes were

found to generally reduce herbivory, their effect on fitness varies across populations. In

sites where trichomes were found negatively associated with fitness, it could be that

trichomes are ineffective deterrents to herbivores, or plants at those sites tended to

produce more trichomes than optimal under current herbivore conditions.

Within Generation Plasticity

A significant "population x wounding " interaction term on plant trichome density in

response to wounding (n=851, df=6, X^2=14.36, p=0.026) suggests that there is natural

variation in within-generation plasticity, or "Genetic x Environmental" variation. In

response to wounding, individuals from MWL and LPD both showed significant decreases

in 7th leaf pair trichome density, while TBR, CSR, IM, and WC showed increases, and HOL

remained at very low levels before and after wounding. Of note, in both populations where

trichome density decreased in response to wounding (MWL and LPD) there was a negative

relationship between trichome density and herbivory in natural sites, while in all four

populations where trichome density increased due to wounding there was a positive

relationship between trichome density and flower production in the field.

Transgenerational Phenotypic Plasticity

Parental wounding had a significant population dependent effect on offspring

trichome density (Pop*PT, df= 7, X^2=21.96, p=0.0026) in field common garden and

greenhouse conditions. The offspring of wounded plants from WC, HOL, TBR, CSR, and BR

had increased trichome density, while offspring trichome production declined in the

offspring of wounded plants from MWL, LPD, and IM. Considering populations for which

there were signs of both with-in generation and between generation plasticity, we find that

in 5/6 cases the directions of plasticity match between with-in and between generation

induction. Plants derived from MWL and LPD showed within generation and

transgenerational decreases in trichome density in response to wounding, while WC, CSR,

and TBR showed with-in generation and transgenerational positive trichome induction in

response to wounding.   Within IM (Iron Mountain) we found evidence for significant

within generation positive induction of trichome density, but negative transgenerational

effects in which the offspring of wounded plants produced fewer trichomes than the

offspring of control plants.

We also identified a significant effect of growth environment on trichome

production (df=1, X^2=184.73, p<0.0001) with plants grown in the greenhouse producing

many more trichomes than their siblings grown in the field. Different populations varied in

the scale to which they produced more trichomes in the field ("Environment x Population",

df=7, X^2=48.61, p<0.0001), but plants from all populations produced more trichomes in

greenhouse compared to field environments. There was not a significant "Environment x Parental Treatment" effect on trichome density (df=1, X^2=0.074, p=0.785), but a significant "Environment x Parental Treatment x Population" term (df=7, X^2=16.16, p=0.024) suggests that populations vary in how they express transgenerational induction in field vs. greenhouse environments.

At common gardens plants producing more trichomes tend to experience less severe herbivory (n=135, df=1, X^2=5.14, p=0.023). Plants experiencing minor levels of herbivory had an average trichome density of 9.8 trichomes, moderate had 5.8, and severe had 2.7.

To further analyze the factors effecting plant herbivore damage in the field, two models were considered. Using a minimum BiC criterion a model in which only common garden site, parental wounding, and their interaction term were selected as explanatory factors explaining herbivore severity in the field (n=115, df=3, BiC=158.2, AiCc=146.2, X^2=25.29, p<0.0001). The offspring of wounded plants experienced less severe herbivore damage than the offspring of control plants. Of the 25 offspring of damaged plants grown at HJA that experienced herbivory, all of them experienced only minor leaf wounding (less than 10% leaf area removal), while 6/19 offspring of undamaged plants at this same site experienced moderate herbivory (between 10%-30% leaf area removal, Figure 4.3c). At BR, the most striking difference was found in the severe herbivory category; only 2/27 (7.5%) offspring of damaged plants experienced severe wounding (>30% leaf area removal), while 13/43 (30%) of offspring from control plants experienced this level of wounding.

Under the minimum AiC selected model, those same three terms and eight additional terms were found to influence plant herbivory (df=11, BiC= 167.9, AiCc=139.2, X^2=51.2, p<0.0001). This less conservative approach identified significant differences in herbivory between plants of different populations, significant parental wounding*population and trichome*site effects on wounding, and trichome density. Of particular interest we find evidence that while there is a general trend of increased herbivore resistance in the offspring of wounded plants across all populations, some populations show this to a great extent, while others show the opposite response. The offspring of wounded plants from WC showed the greatest increase in wound resistance, while the offspring of wounded plants from MWL were less resistant than the offspring of control plants. Log-ratio Chi-Square tests within this model confirm that the effect of parental wounding on offspring herbivore resistance varies significantly between individuals from MWL and WC. Of the other populations, BR, CSR, and IM showed similar resistance in the offspring of control and wounded individuals, while HOL, LPD, and TBR showed increased resistance in the offspring of damaged plants.

Discussion

The experiments above reiterate the role of plastic responses to the environment, genetic differentiation, and the inferred role of selection on phenotypic differentiation in plants using an extension of the common garden technique popularized by Clausen, Keck, and Heisey (CLAUSEN *et al.* 1948). The addition of an intermediate greenhouse generation, in which plants were either exposed to mechanical wounding or not, extends the framework to consider transgenerational effects on phenotype and fitness. I demonstrate

128

that (i) trichomes limit herbivory in the field and , (ii) parental environment alters

offspring herbivore resistance and (iii) defense phenotypes, and that (iv) there is natural

variation in these responses. Additionally, we find that there are patterns of inter-annual

temperature autocorrelation in central Oregon that may favor genotypes that exhibit

transgenerational plasticity.

Even in the relatively minute geographical range (150 x 80 miles) we consider in

this study, there is vast natural variation in trichome density both within and between

populations suggesting that this trait may have a fickle association with plant fitness.

Supporting this hypothesis, at some sites plants with more trichomes produced a greater

number of flowers, while at others there was a negative relationship between trichome

density and flower number. Depending on the type and quantity of herbivores present in a

given year the cost/benefit of trichomes may vary widely, selecting for a trait that is

particularly environmentally responsive in comparison to core metabolic traits in which

canalization is likely to limit plasticity.

Trichomes: Natural Variation and Role in Herbivore Defense

Population level mean trichome densities are strongly correlated with population

level incidences of herbivory (R^2=0.735, Figure 4.2), and at both common gardens and

natural field sites plants with more trichomes tend to experience less severe herbivory.

Additionally, at some natural sites trichomes were associated with increased fitness, while

at other they are associated with decreased fitness. Taken together, this suggests that *M.

guttatus* trichomes play a role in herbivore defense, and they are plastically induced in

response to herbivores. Both of these possibilities are supported in the literature; some

genotypes of *Mimulus guttatus* induce trichome production in response to wounding (HOLESKI 2007), and there are costs and benefits associated with herbivore defense(MAURICIO and RAUSHER 1997; MAURICIO 1998) that could lead to population dependent selection on trichome densities.

Supporting plasticity's role in altering trichome density, there was a significant with-in generation effect of wounding (Figure 4.1c, Appendix 18, Appendix 19) on plant trichome density. Additionally, there was significant inter-population variation in this plasticity, demonstrating the presence of "Genetic x Environmental" trichome variation. In response to wounding four of the seven populations considered increased trichome production, one showed no pattern of change, and two populations produced fewer trichomes. The two populations (LPD and MWL) that produced fewer trichomes in response to wounding were also the only two of seven populations to have a negative correlation between trichomes and fitness in the field, while the four populations had positive relationships between trichome density and fitness in the field. As only living, flower producing plants were considered in this study, high constitutive trichome production at these two sites may increase the probability of survival to flowering, but given that a plant does flower have a negative effect on fecundity.

Transgenerational Effects on Trichomes and Herbivory

In both greenhouse and field experiments leaf trichome density was affected by wounding conditions in the previous generation. While the majority of lines showed increased trichome density in the offspring of wounded individuals, plants derived from LPD, MWL, and IM wounded offspring produced fewer trichomes. This represents an

example of negative transgenerational effects, which are expected to evolve under negative environmental autocorrelations. The parental stress imposed in this experiment, mechanical wounding though hole-punching, does not perfectly mimic the response of herbivory (REYMOND *et al.* 2000), yet wounding responsive genes seem to partially overlap with herbivory induced genes and represent a variety of general stress-response pathways (REYMOND *et al.* 2004).

While trichome densities were found to vary in response to parental wounding in both field and greenhouse environments, there was a significant three way "Current Environment x Parental Environment x Population" effect on trichome density. This demonstrates that the phenotypic outcomes of transgenerational plasticity vary depending on the current environment and genotype of an individual and represents an example of "genetic x environmental x transgenerational" variation. While the current environment effects the magnitude of the transgenerational response, in this study the same populations that exhibit transgenerational induction in nature did so in the greenhouse as well.

Using both minimum BiC and AiCc model selection approaches we compared models to identify factors influencing herbivore resistance in field common gardens. The minimum BiC selected approach identified only parental wounding, common garden site, and their interaction term were found to significantly influence wounding severity. As herbivory is known to impose strong selective pressures, increased herbivore resistance in the offspring of wounded plants suggests that experiences in the prior generation contribute to variation in fitness in nature. Transgenerational induction of increased herbivore defense in response to parental wounding should be adaptive when patterns of year-to-year variation in herbivore density exhibit at least a minor degree of temporal

autocorrelation. In order to directly test whether these patterns of herbivore variation

exist in nature, long-term herbivore abundance studies would need to be preformed. While

this data set is not available at these locations, patterns of abiotic environmental variation

can provide a first glimpse into environmental autocorrelations in nature.

Climate data from 1895-2014 exhibits patterns of year-to-year autocorrelation for

both annual mean temperature and precipitation. Both temperature and moisture

availability impact herbivore activity, and so these climatic variables should shed light unto

patterns of inter-annual herbivore variation. At all four locations there were positive

autocorrelations in mean annual temperature, with autocorrelation values varying from

0.12 to 0.41. While the intricacies underlying this autocorrelation were not studied here, it

is possible that patterns such as the El Nino Southern Oscillation and Pacific Decadal

Oscillation are responsible for inter-annual temperature autocorrelations (NEWMAN *et al.*

2003). If mean annual temperature influences herbivore activity, then this observed

pattern of temperature autocorrelation may give rise to positive autocorrelations in

herbivore abundance. This should select for genotypes where the offspring of wounded

plants are better defended than the offspring of control plants, as observed in this study.

Annual autocorrelation of precipitation was generally less prominent than

temperature autocorrelations, and varied not only in terms of magnitude, but also

direction. In three of the four sites considered there was a slight negative autocorrelation

in annual precipitation (from -0.05 to -0.11), and in one site there was a positive

correlation of 0.16. Over the past 120 years, at some locations particularly wet years have

tended to be followed by another wet year, in other locations wet years tend to be followed

by dry years.

Using the more complex minimum AiCc selected model we find evidence for natural variation in transgenerationally induced resistance to herbivory. Considered alongside the natural variation in TPP for trichome production, there is substantial support for positive transgenerational induction within the WC, TBR, and HOL populations (induced herbivore resistance and trichome production), and moderate support for BR and CSR (induced trichome production but not herbivore resistance). Within LPD we have evidence of induced herbivore resistance, but decreased trichome production, potentially implicating other forms of defense in their transgenerational response. In MWL there is significant evidence of negative transgenerational induction (reduced trichome production and herbivore resistance in the progeny of wounded plants), with moderate support for negative induction in IM (reduced trichome density but no change in herbivore resistance). Out of the eight populations considered, the two that showed signs of negative transgenerational induction, IM (Iron Mountain) and MWL (Mount Washington Lookout), were derived from the highest elevation sites.

IM and MWL are not located particularly close geographically, but they do represent the two northern most populations surveyed in this study. It could be that they represent a genetic clade that does not have the genetic machinery necessary to exhibit transgenerational plasticity.   While this is possible, IM is located only 8 miles from the BR, which, along with the next most northern population, CSR, both show significant signs of transgenerational plasticity. It therefore seems possible that a certain aspect of high elevation life selects against transgenerational plasticity for trichome induction and herbivore resistance.

Assuming rates of herbivory are both temperature and moisture dependent as suggested in the literature(BALE *et al.* 2002), the observed positive autocorrelations in temperature may explain the lack of transgenerational plasticity at high elevations. At alpine sites such as IM and MWL, both insect herbivores and plants are most active during the relatively short growing season between when the ground clears of snow and the soil completely dries. While high temperatures and ample rain during the growing season should increase herbivore activity in the short-term, it will also reduce nearby snowpack going into winter (WALKER *et al.* 1999; PEDERSON *et al.* 2011). During the following growing season decreased snowpack will reduce moisture availability, a vital factor determining herbivore success (Bale *et al.* 2002). Thus, while at high elevations inter-annual temperature autocorrelations exist, the relationship between temperature and snowpack may generate a negative autocorrelation in inter-annual herbivore activity.

While the contribution of transgenerational effects on fitness in nature has numerous ecological implications, the evolutionary significance of TPP depends upon natural variation in the underlying molecular mechanism. This study demonstrates the presence of significant inter-population variation for TPP (Figure 4.1d), suggesting that evolution in response to different regimes of environmental variation could in turn lead to variation in TPP. Population level variation in TPP was considered in this experiment due to the assumption that the environmental patterns selecting for or against TPP should be relatively constant within a population but vary considerably across the species range. Evidence suggests that there is also with-in population variation for TPP(GALLOWAY and ETTERSON 2007), which could be acted upon by natural selection to favor genotypes whose capacity to transgenerationally respond are favorable in the local environment.

Recently, there has been controversy over the evolutionary significance of TPP, yet theoretical work continues to demonstrate that under certain patterns of year-to-year environmental variation TPP should be adaptive (HERMAN *et al.* 2013a; LEIMAR and MCNAMARA 2015), and here I provide empirical support for the adaptive potential of this system and the presence of TPP variation in nature. In environments where parental conditions are highly correlated with offspring conditions, natural selection should favor alleles that lead to a high degree of TPP, but in areas with low year-to-year serial correlation, these alleles should be selected against (Figure 1d, FURROW and FELDMAN 2014). Based on patterns of year-to-year temperature variation we confirm that the necessary ingredients necessary for transgenerational plasticity to evolve are present in this system.

Conclusions

Aside from imparting selective pressures, these experiments suggest that environmental conditions can directly manipulate the growth and success of future generations through transgenerational effects. One possible explanation for this system of inherited environmental information is that a portion of the environmentally induced epigenetic changes (such as DNA methylation or histone modifications) are not reset, but rather passed between generations (VERHOEVEN *et al.* 2010). Further work is necessary to determine the mechanism through which epigenetic effects are reiterated in the germ line, but evidence for the epigenetic basis of TPP is mounting. Methylation changes in response to environmental stress (WANG *et al.* 2010; DOWEN *et al.* 2012), stably transmitted epigenetic markings (RASMANN *et al.* 2012; SLAUGHTER *et al.* 2012; SCHMITZ *et al.* 2013a; LI *et al.* 2014), and epigenetic effects on gene expression (COLICCHIO *et al.* 2015a) all point towards epigenetic inheritance as the source of TPP.

Adaptive TPP in nature was first demonstrated as a transition between annual and biennial life history strategies in response to maternal light conditions in *Campanulastrum americanum* (GALLOWAY and ETTERSON 2007; GALLOWAY and ETTERSON 2009). Here, utilizing the system of plant herbivore defense, this work gets expanded upon by considering multiple populations, natural TPP variation, and measures of offspring fitness and phenotype in common gardens. The DNA passed from parent to offspring is insensitive to the experiences of a generation, but rather a compendium of the mutation, selection, and migration of ones ancestors. It is clear that in addition to the stochastic voyage towards fitness peaks, the transmission of environmentally induced signals also plays a role in the success and development of an organism. While this and other recent experiments have confirmed the potential implications of TPP, future studies across a diversity of species and involving herbivory, light conditions, and other environmental factors will be necessary to demonstrate it's relative ecological and evolutionary significance.

Figure 4.1. Four mechanisms that explain natural variation in plant trichome density. (A) Transgenerational induction: In response to wounding plants transmit a signal to their offspring that leads to increased trichome production in the next generation. In nature the offspring of damaged plants have an increased likelihood of producing glandular trichomes, leading to less severe herbivory in these offspring of damaged plants. (B)Within-generation induction: In response to wounding plants produce more trichomes on later leaf pairs. Within-generation trichome induction has been documented in a number of species, and within this study there was natural variation between populations for the induction of glandular trichomes. It is currently unclear as to whether similar signals and mechanisms are responsible for within and between generation induction. (C) Herbivore-driven selection. Between population variation in herbivore abundance leads to coinciding variation in selective forces acting upon plant defense traits, such as trichomes. There was a strong correlation between incidence of herbivore damage and trichome density that was maintained in common garden conditions. This suggests that along with plastic trichome regulation, stable genetic selection due to local herbivore abundance has adjusted baseline trichome densities on a population scale. (D) Variation-driven TPP selection. When year-to-year variation in herbivore abundance has high positive autocorrelation the transmission of environmental signals should be selected for; however when autocorrelation is low or negative TPP should be selected against. Other factors, such as the periodicity of the transitions between high and low herbivore years and the cost of TPP should also affect the evolution and maintenance of transgenerational inheritance. This study demonstrates that there is natural variation in TPP, but future work will be necessary to demonstrate the environmental parameters that select on this variation.
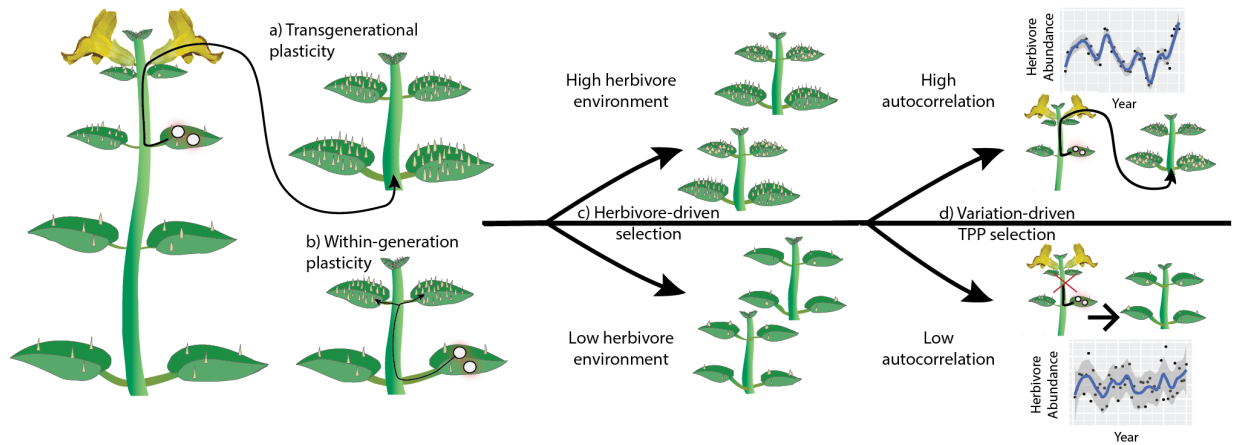
Figure 4.2. Positive correlation (R^2=0.735) between percent leaves damaged and the number of trichomes per leaf in sixteen natural *M. guttatus* populations. Elevation and longitude of each population are demarcated by size and color of the point respectively. These factors represented alternative ecological correlates that may have, but were not, found to be associated with plant trichome production.
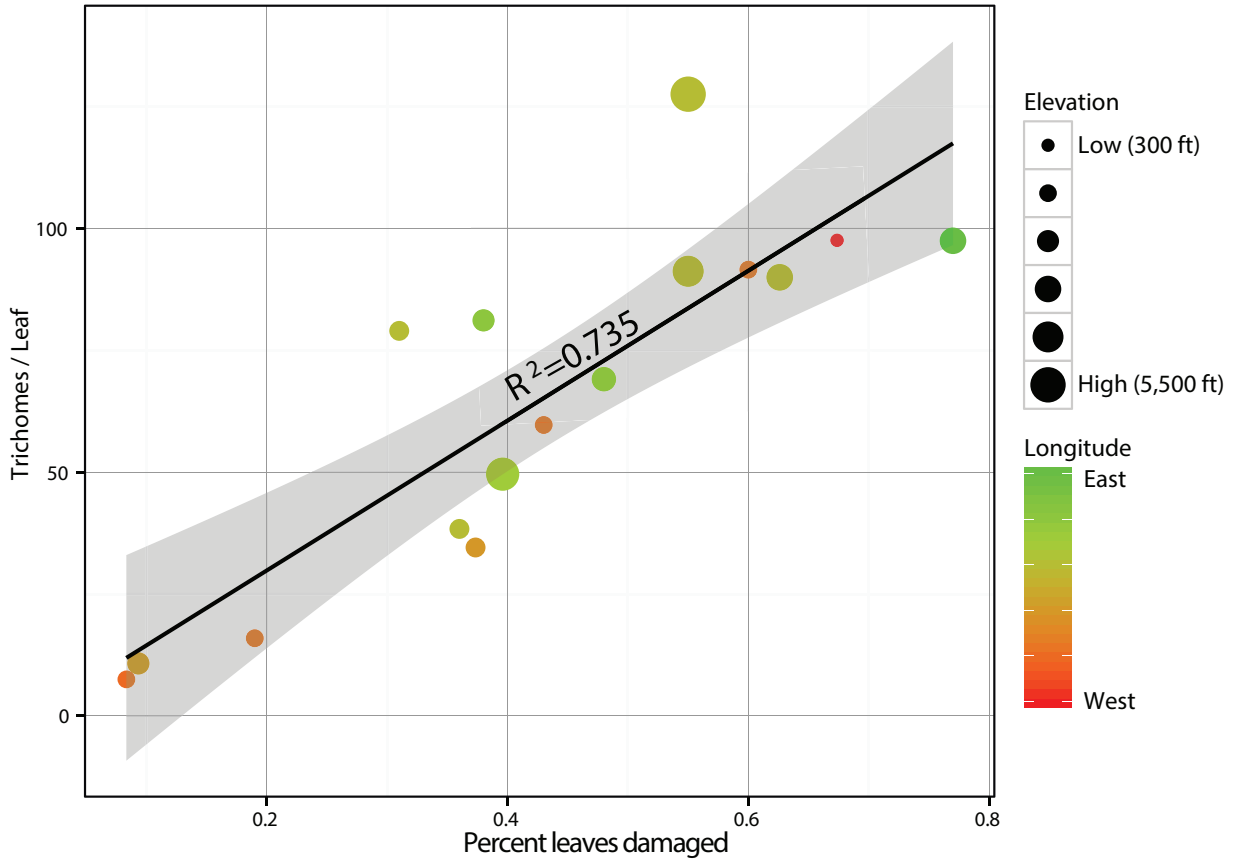
Figure 4.3. (A) Mean log transformed glandular trichome counts for *M. guttatus* across common garden sites that experienced varying levels of herbivory. Error bars represent standard error. Severely wounded plants had significantly fewer glandular trichomes than plants with no or minor wounding. (B) Mean log transformed trichome counts at natural populations for plants that experienced varying levels of herbivory. Error bars represent standard error. Severely wounded plants had significantly fewer trichomes than plants with minor wounding. (C) Damage severity at BR (n=70) and HJA (n=44) common garden sites in the offspring of control and damaged plants. The width of columns scales to the number of individuals in that treatment class for a given common garden site.
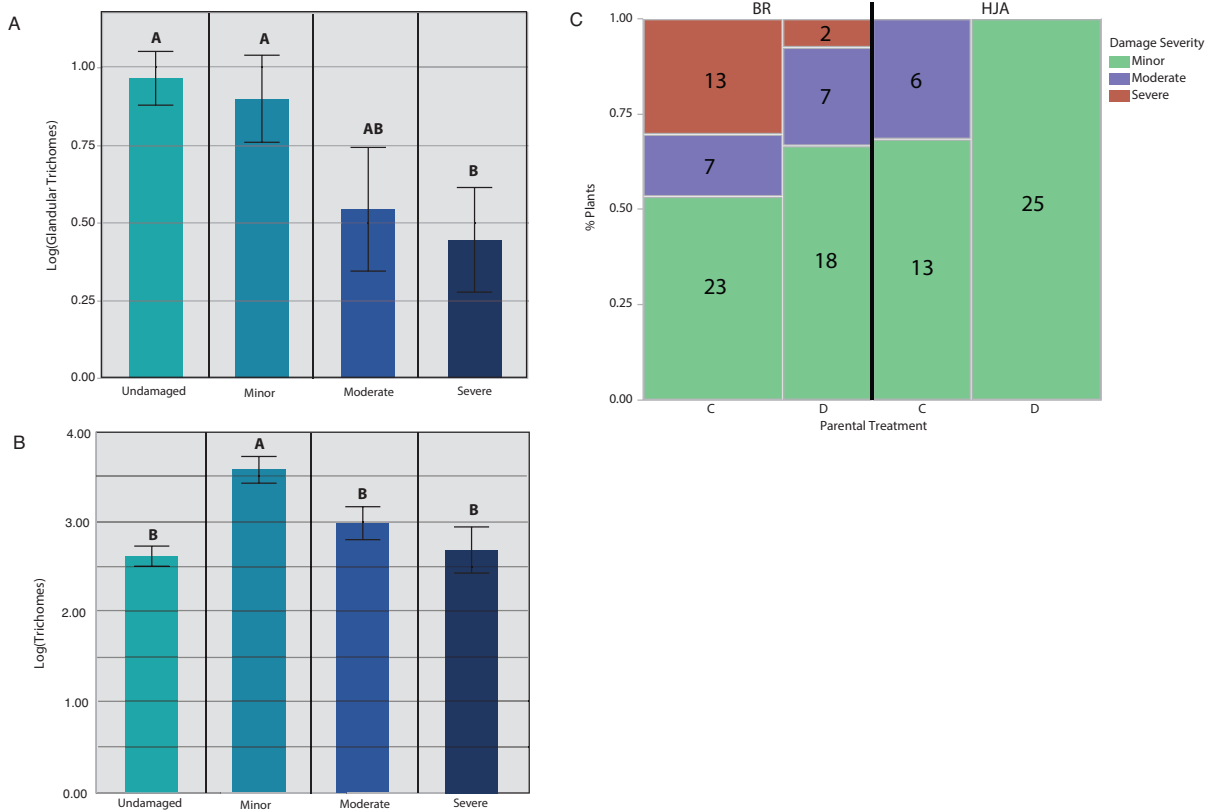
Conclusions

In at least one relatively unassuming wildflower it appears that the simple act of hole-punching a leaf leads to epigenetic, gene expression, phenotypic, and resistance differences in it's offspring. In this dissertation I demonstrate that not only are genes involved in trichome production differentially expressed, but so to are those in a host of other metabolic pathways. Additionally, many differentially expressed genes overlap with regions of differential methylation, suggesting a key role of epigenetics in TPP. When taken to the field, not only do the offspring of wounded plants produce more trichomes than the offspring of control plants, but they also resist herbivory to a greater extent. Taken together, these results suggest that the mechanisms capable of promoting adaptive transgenerational inheritance are in place, and may play a prominent role in adaptation.

As the scientific community begins to devote more energy to the study of TPP, and the "Lamarckian" stigma in the field begins to fade, its true place in the extended evolutionary synthesis model of life will be discovered. Whatever this place may be, as a brief aside or a core concept, the transmission of environmental information between generations demonstrates the remarkable power of natural selection. The fact that a complementary system of inheritance can evolve from within our primary system of genetic inheritance is a marvelous thing, and a prime example of how seemingly novel mechanisms can emerge from fundamental laws. The next time you find yourself in an alpine meadow in mid-bloom; delight in the fact that the vast natural variation is not simply the product of environmental and genetic differences, but also reflects lingering environmental responses of generations past.

Literature Cited

AGRAWAL, A. A., 2001 Phenotypic Plasticity in the Interactions and Evolution of Species. Science **294:** 321-326.

AGRAWAL, A. A., 2002 Herbivory and maternal effects: mechanisms and consequences of transgenerational induced plant resistance. Ecology **83:** 3408-3415.

AGRAWAL, A. A., J. K. CONNER, M. T. JOHNSON and R. WALLSGROVE, 2002 Ecological genetics of an induced plant defense against herbivores: additive genetic variance and costs of phenotypic plasticity. Evolution **56:** 2206-2213.

AGRAWAL, A. A., A. P. HASTINGS, M. T. JOHNSON, J. L. MARON and J.-P. SALMINEN, 2012 Insect herbivores drive real-time ecological and evolutionary change in plant populations. Science **338:** 113-116.

AGRAWAL, A. A., C. LAFORSCH and R. TOLLRIAN, 1999 Transgenerational induction of defences in animals and plants. Nature **401:** 60-63.

ÅGREN, J., and D. W. SCHEMSKE, 1993 The cost of defense against herbivores: an experimental study of trichome production in Brassica rapa. The American Naturalist **141:** 338-350.

ANDERS, S., and W. HUBER, 2010 Differential expression analysis for sequence count data. Genome Biol **11:** R106.

ANDERSON, J. P., E. BADRUZSAUFARI, P. M. SCHENK, J. M. MANNERS, O. J. DESMOND *et al.*, 2004 Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. The Plant Cell Online **16:** 3460-3479.

APPLEQUIST, W. L., R. CRONN and J. F. WENDEL, 2001 Comparative development of fiber in wild and cultivated cotton. Evolution & development **3:** 3-17.

ASHBURNER, M., C. A. BALL, J. A. BLAKE, D. BOTSTEIN, H. BUTLER *et al.*, 2000 Gene Ontology: tool for the unification of biology. Nature genetics **25:** 25-29.

BALE, J. S., G. J. MASTERS, I. D. HODKINSON, C. AWMACK, T. M. BEZEMER *et al.*, 2002 Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. Global Change Biology **8:** 1-16.

BALLHORN, D. J., S. KAUTZ and J. M. LAUMANN, 2016 Herbivore damage induces a transgenerational increase of cyanogenesis in wild lima bean (Phaseolus lunatus). Chemoecology **26:** 1-5.

BARTEE, L., F. MALAGNAC and J. BENDER, 2001 Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes & Development **15:** 1753-1758.

BEARDSLEY, P. M., and R. G. OLMSTEAD, 2002 Redefining Phrymaceae: the placement of Mimulus, tribe Mimuleae, and Phryma. American Journal of Botany **89:** 1093-1102.

BILICHAK, A., Y. ILNYSTKYY, J. HOLLUNDER and I. KOVALCHUK, 2012 The progeny of Arabidopsis thaliana plants exposed to salt exhibit changes in DNA methylation, histone modifications and gene expression. PloS one **7:** e30515.

BLEECKER, A. B., and H. KENDE, 2000 Ethylene: a gaseous signal molecule in plants. Annual review of cell and developmental biology **16:** 1-18.

BLÖDNER, C., C. GOEBEL, I. FEUSSNER, C. GATZ and A. POLLE, 2007 Warm and cold parental reproductive environments affect seed properties, fitness, and cold responsiveness in Arabidopsis thaliana progenies. Plant, cell & environment **30:** 165-175.

BOND, D. M., and D. C. BAULCOMBE, 2014a Small RNAs and heritable epigenetic variation in plants. Trends Cell Biol **24:** 100-107.

BOND, D. M., and D. C. BAULCOMBE, 2014b Small RNAs and heritable epigenetic variation in plants. Trends in cell biology **24:** 100-107.

BOSSDORF, O., C. L. RICHARDS and M. PIGLIUCCI, 2008 Epigenetics for ecologists. Ecol Lett **11**.

BOYKO, A., T. BLEVINS, Y. YAO, A. GOLUBOV, A. BILICHAK *et al.*, 2010 Transgenerational adaptation of Arabidopsis to stress requires DNA methylation and the function of Dicer-like proteins. PLoS One **5:** e9514.

BOYKO, A., and I. KOVALCHUK, 2010 Transgenerational response to stress in Arabidopsis thaliana. Plant signaling & behavior **5:** 995-998.

BROCK, G., V. PIHUR, S. DATTA and S. DATTA, 2011 clValid, an R package for cluster validation. Journal of Statistical Software (Brock et al., March 2008).

BRUCE, T. J., M. C. MATTHES, J. A. NAPIER and J. A. PICKETT, 2007 Stressful "memories" of plants: evidence and possible mechanisms. Plant Science **173:** 603-608.

CALARCO, J. P., F. BORGES, M. T. DONOGHUE, F. VAN EX, P. E. JULLIEN *et al.*, 2012 Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell.

CAO, X., and S. E. JACOBSEN, 2002 Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proceedings of the National Academy of Sciences **99:** 16491-16498.

CASTILLO-DAVIS, C. I., S. L. MEKHEDOV, D. L. HARTL, E. V. KOONIN and F. A. KONDRASHOV, 2002 Selection for short introns in highly expressed genes. Nature genetics **31:** 415-418.

CASWELL, H., and F. C. REED, 1976 Plant-herbivore interactions. Oecologia **26:** 151-156.

CHAN, S. W., I. R. HENDERSON, X. ZHANG, G. SHAH, J. S. CHIEN *et al.*, 2006 RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in Arabidopsis. PLoS Genet **2:** e83-e83.

CHINNUSAMY, V., Z. GONG and J. K. ZHU, 2008 Abscisic Acid‐mediated Epigenetic Processes in Plant Development and Stress Responses. Journal of integrative plant biology **50:** 1187-1195.

CHODAVARAPU, R. K., S. FENG, B. DING, S. A. SIMON, D. LOPEZ *et al.*, 2012 Transcriptome and methylome interactions in rice hybrids. Proc Natl Acad Sci U S A **109:** 12040-12045.

CLAUSEN, J. C., D. D. KECK and W. M. HIESEY, 1948 *Experimental Studies on the Nature of Species: Environmental Responses of Climatic Races of Archillea. III*. Carnegie Institution of Washington.

COLEMAN-DERR, D., and D. ZILBERMAN, 2012 Deposition of histone variant H2A. Z within gene bodies regulates responsive genes. PLoS genetics **8:** e1002988.

COLEY, P., 1988 Effects of plant growth rate and leaf lifetime on the amount and type of anti-herbivore defense. Oecologia **74:** 531-536.

COLICCHIO, J. M., F. MIURA, J. K. KELLY, T. ITO and L. C. HILEMAN, 2015a DNA methylation and gene expression in Mimulus guttatus. BMC genomics **16:** 507.

COLICCHIO, J. M., P. J. MONNAHAN, J. K. KELLY and L. C. HILEMAN, 2015b Gene expression plasticity resulting from parental leaf damage in Mimulus guttatus. New Phytologist **205:** 894-906.

CONESA, A., S. GÖTZ, J. M. GARCÍA-GÓMEZ, J. TEROL, M. TALÓN *et al.*, 2005 Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21:** 3674-3676.

COPPOCK, D. L., J. DETLING, J. ELLIS and M. DYER, 1983 Plant-herbivore interactions in a North American mixed-grass prairie. Oecologia **56:** 1-9.

COSGROVE, D. J., 1997 Assembly and enlargement of the primary cell wall in plants. Annual review of cell and developmental biology **13:** 171-201.

DOWEN, R. H., M. PELIZZOLA, R. J. SCHMITZ, R. LISTER, J. M. DOWEN *et al.*, 2012 Widespread dynamic DNA methylation in response to biotic stress. Proceedings of the National Academy of Sciences **109:** E2183-E2191.

DURRANT, A., 1962 The environmental induction of heritable change in Linum. Heredity **17:** 27-61.

EICHTEN, S. R., R. BRISKINE, J. SONG, Q. LI, R. SWANSON-WAGNER *et al.*, 2013a Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell **25:** 2783-2797.

EICHTEN, S. R., R. J. SCHMITZ and N. M. SPRINGER, 2014 Epigenetics: Beyond Chromatin Modifications and Complex Genetic Regulation. Plant Physiol **165:** 933-947.

EICHTEN, S. R., M. W. VAUGHN, P. J. HERMANSON and N. M. SPRINGER, 2013b Variation in DNA methylation patterns is more common among maize inbreds than among tissues. The Plant Genome **6**.

EKLÖF, J. M., and H. BRUMER, 2010 The XTH gene family: an update on enzyme structure, function, and phylogeny in xyloglucan remodeling. Plant physiology **153:** 456-466.

FENG, S., S. J. COKUS, X. ZHANG, P.-Y. CHEN, M. BOSTICK *et al.*, 2010 Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences **107:** 8689-8694.

FLAGEL, L. E., J. H. WILLIS and T. J. VISION, 2014 The standing pool of genomic structural variation in a natural population of Mimulus guttatus. Genome biology and evolution **6:** 53-64.

FLAVELL, R., 1994 Inactivation of gene expression in plants as a consequence of specific sequence duplication. Proceedings of the National Academy of Sciences **91:** 3490-3496.

FURROW, R. E., and M. W. FELDMAN, 2014 Genetic variation and the evolution of epigenetic regulation. Evolution **68:** 673-683.

GALLOWAY, L. F., and J. R. ETTERSON, 2007 Transgenerational plasticity is adaptive in the wild. Science **318:** 1134-1136.

GALLOWAY, L. F., and J. R. ETTERSON, 2009 Plasticity to canopy shade in a monocarpic herb: within‐ and between‐generation effects. New Phytologist **182:** 1003-1012.

GANG, D. R., H. KASAHARA, Z.-Q. XIA, K. VANDER MIJNSBRUGGE, G. BAUW *et al.*, 1999 Evolution of plant defense mechanisms relationships of phenylcoumaran benzylic ether reductases to pinoresinol-lariciresinol and isoflavone reductases. Journal of Biological Chemistry **274:** 7516-7527.

GATEHOUSE, J. A., 2002 Plant resistance towards insect herbivores: a dynamic interaction. New phytologist **156:** 145-169.

GENT, J. I., N. A. ELLIS, L. GUO, A. E. HARKESS, Y. YAO *et al.*, 2013 CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome Res **23:** 628-637.

GENTLEMAN, R. C., V. J. CAREY, D. M. BATES, B. BOLSTAD, M. DETTLING *et al.*, 2004 Bioconductor: open software development for computational biology and bioinformatics. Genome biology **5:** R80.

GONZÁLEZ-GUZMÁN, M., N. APOSTOLOVA, J. M. BELLÉS, J. M. BARRERO, P. PIQUERAS *et al.*, 2002 The short-chain alcohol dehydrogenase ABA2 catalyzes the conversion of xanthoxin to abscisic aldehyde. The Plant Cell Online **14:** 1833-1846.

GOODSTEIN, D. M., S. SHU, R. HOWSON, R. NEUPANE, R. D. HAYES *et al.*, 2012 Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40:** D1178-D1186.

GREENBERG, M. V., A. DELERIS, C. J. HALE, A. LIU, S. FENG *et al.*, 2013 Interplay between active chromatin marks and RNA-directed DNA methylation in Arabidopsis thaliana. PLoS Genet **9:** e1003946.

GRUENBAUM, Y., T. NAVEH-MANY, H. CEDAR and A. RAZIN, 1981 Sequence specificity of methylation in higher plant DNA.

GUTZAT, R., and O. MITTELSTEN SCHEID, 2012 Epigenetic responses to stress: triple defense? Current opinion in plant biology.

HARDCASTLE, T., 2013 High-throughput sequencing of cytosine methylation in plant DNA. Plant Methods **9:** 16.

HEARD, E., and R. A. MARTIENSSEN, 2014 Transgenerational Epigenetic Inheritance: Myths and Mechanisms. Cell **157:** 95-109.

HELLSTEN, U., K. M. WRIGHT, J. JENKINS, S. SHU, Y. YUAN *et al.*, 2013 Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing. Proceedings of the National Academy of Sciences **110:** 19478-19482.

HERMAN, J. J., H. G. SPENCER, K. DONOHUE and S. E. SULTAN, 2013a How stable 'should' epigenetic modifications be? Insights from adaptive plasticity and bet hedging. Evolution.

HERMAN, J. J., and S. E. SULTAN, 2011 Adaptive transgenerational plasticity in plants: case studies, mechanisms, and implications for natural populations. Frontiers in plant science **2**.

HERMAN, J. J., S. E. SULTAN, T. HORGAN-KOBELSKI and C. RIGGS, 2012 Adaptive transgenerational plasticity in an annual plant: grandparental and parental drought stress enhance performance of seedlings in dry soil. Integrative and comparative biology **52:** 77-88.

HERRERA, C. M., and P. BAZAGA, 2013 Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous Ilex aquifolium (Aquifoliaceae) trees. Botanical Journal of the Linnean Society **171:** 441-452.

HOLESKI, L., 2007 Within and between generation phenotypic plasticity in trichome density of Mimulus guttatus. Journal of Evolutionary Biology **20:** 2092-2100.

HOLESKI, L., K. KEEFOVER-RING, M. D. BOWERS, Z. HARNENZ and R. LINDROTH, 2013a Patterns of Phytochemical Variation in Mimulus guttatus (Yellow Monkeyflower). Journal of Chemical Ecology **39:** 525-536.

HOLESKI, L. M., R. CHASE‐ALONE and J. K. KELLY, 2010 The genetics of phenotypic plasticity in plant defense: trichome production in Mimulus guttatus. The American Naturalist **175:** 391-400.

HOLESKI, L. M., G. JANDER and A. A. AGRAWAL, 2012 Transgenerational defense induction and epigenetic inheritance in plants. Trends in ecology & evolution.

HOLESKI, L. M., M. S. ZINKGRAF, J. J. COUTURE, T. G. WHITHAM and R. L. LINDROTH, 2013b Transgenerational effects of herbivory in a group of long-lived tree species: maternal damage reduces offspring allocation to resistance traits, but not growth. Journal of Ecology **101:** 1062-1073.

HOLLISTER, J. D., and B. S. GAUT, 2007 Population and Evolutionary Dynamics of Helitron Transposable Elements in Arabidopsis thaliana. Molecular Biology and Evolution **24:** 2515-2524.

HOLOCH, D., and D. MOAZED, 2015 RNA-mediated epigenetic regulation of gene expression. Nature Reviews Genetics.

HUFF, J. T., and D. ZILBERMAN, 2014 Dnmt1-Independent CG Methylation Contributes to Nucleosome Positioning in Diverse Eukaryotes. Cell **156:** 1286-1297.

HÜLSKAMP, M., S. MISÉRA and G. JÜRGENS, 1994 Genetic dissection of trichome cell development in Arabidopsis. Cell **76:** 555-566.

ISHIGURO, S., A. KAWAI-ODA, J. UEDA, I. NISHIDA and K. OKADA, 2001 The DEFECTIVE IN ANTHER DEHISCENCE1 gene encodes a novel phospholipase A1 catalyzing the initial step of jasmonic acid biosynthesis, which synchronizes pollen maturation, anther dehiscence, and flower opening in Arabidopsis. The Plant Cell **13:** 2191-2209.

ITO, H., H. GAUBERT, E. BUCHER, M. MIROUZE, I. VAILLANT *et al.*, 2011 An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature **472:** 115-119.

JABLONKA, E., 2012 Epigenetic inheritance and plasticity: the responsive germline. Progress in Biophysics and Molecular Biology.

JABLONKA, E., and G. RAZ, 2009 Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. The Quarterly review of biology **84:** 131-176.

JOHNSON, M. T. J., J. A. BERTRAND and M. M. TURCOTTE, 2015 Precision and accuracy in quantifying herbivory. Ecological Entomology**:** n/a-n/a.

JONES, P. A., 2012 Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics **13:** 484-492.

KANEHISA, M., S. GOTO, Y. SATO, M. FURUMICHI and M. TANABE, 2012 KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research **40:** D109-D114.

KANKEL, M. W., D. E. RAMSEY, T. L. STOKES, S. K. FLOWERS, J. R. HAAG *et al.*, 2003 Arabidopsis MET1 cytosine methyltransferase mutants. Genetics **163:** 1109-1122.

KARIÑHO‐BETANCOURT, E., A. A. AGRAWAL, R. HALITSCHKE and J. NÚÑEZ‐FARFÁN, 2015 Phylogenetic correlations among chemical and physical plant defenses change with ontogeny. New Phytologist.

KAWAKATSU, T., T. STUART, M. VALDES, N. BREAKFIELD, R. J. SCHMITZ *et al.*, 2016 Unique cell-type-specific patterns of DNA methylation in the root meristem. Nature Plants **2:** 16058.

KELLY, J. K., B. KOSEVA and J. P. MOJICA, 2013 The genomic signal of partial sweeps in Mimulus guttatus. Genome biology and evolution **5:** 1457-1469.

KIEŁBASA, S. M., R. WAN, K. SATO, P. HORTON and M. C. FRITH, 2011 Adaptive seeds tame genomic sequence comparison. Genome research **21:** 487-493.

KILBY, N. J., H. O. LEYSER and I. J. FURNER, 1992 Promoter methylation and progressive transgene inactivation inArabidopsis. Plant molecular biology **20:** 103-112.

KILVITIS, H., M. ALVAREZ, C. FOUST, A. SCHREY, M. ROBERTSON *et al.*, 2014 Ecological Epigenetics, pp. 191-210 in *Ecological Genomics*, edited by C. R. LANDRY and N. AUBIN-HORTH. Springer Netherlands.

KINOSHITA, T., and S. E. JACOBSEN, 2012 Opening the Door to Epigenetics in PCP. Plant and Cell Physiology **53:** 763-765.

KOHAVI, R., 1995 A study of cross-validation and bootstrap for accuracy estimation and model selection, pp. 1137-1145 in *IJCAI*.

KORTH, K. L., and R. A. DIXON, 1997 Evidence for chewing insect-specific molecular events distinct from a general wound response in leaves. Plant Physiology **115:** 1299-1305.

KOU, H. P., Y. LI, X. X. SONG, X. F. OU, S. C. XING *et al.*, 2011 Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (Oryza sativa L.). J Plant Physiol **168**.

KUIJPER, B., and R. B. HOYLE, 2015 When to rely on maternal effects and when on phenotypic plasticity? Evolution **69:** 950-968.

LANG-MLADEK, C., O. POPOVA, K. KIOK, M. BERLINGER, B. RAKIC *et al.*, 2010 Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in Arabidopsis. Molecular plant **3:** 594-602.

LANGE, T., S. ROBATZEK and A. FRISSE, 1997 Cloning and expression of a gibberellin 2 beta, 3 beta-hydroxylase cDNA from pumpkin endosperm. The Plant Cell **9:** 1459-1467.

LANGFELDER, P., B. ZHANG and S. HORVATH, 2008 Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics **24:** 719-720.

LATZEL, V., E. ALLAN, A. BORTOLINI SILVEIRA, V. COLOT, M. FISCHER *et al.*, 2013 Epigenetic diversity increases the productivity and stability of plant populations. Nat Commun **4:** 2875.

LAU, J. A., 2012 Direct and indirect transgenerational effects alter plant-herbivore interactions. Evolutionary Ecology **26:** 1469-1480.

LAW, C. W., Y. CHEN, W. SHI and G. K. SMYTH, 2013a Voom! Precision weights unlock linear model analysis tools for RNA-seq read counts, pp. Technical Report 1 May 2013, Bioinformatics Division, Walter and Eliza Hall Institute of Medical Reseach, Melbourne, Australia. http://www. statsci. org/smyth/pubs/VoomPreprint. pdf.

LAW, J. A., J. DU, C. J. HALE, S. FENG, K. KRAJEWSKI *et al.*, 2013b Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature.

LAW, J. A., and S. E. JACOBSEN, 2010 Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics **11:** 204-220.

LEIMAR, O., and J. M. MCNAMARA, 2015 The evolution of transgenerational integration of information in heterogeneous environments. The American Naturalist **185:** E55-E69.

LEONHARDT, H., A. W. PAGE, H.-U. WEIER and T. H. BESTOR, 1992 A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell **71:** 865-873.

LEWSEY, M. G., T. J. HARDCASTLE, C. W. MELNYK, A. MOLNAR, A. VALLI *et al.*, 2016 Mobile small RNAs regulate genome-wide DNA methylation. Proceedings of the National Academy of Sciences**:** 201515072.

Li, J., and R. Tibshirani, 2011 Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-seq data. Statistical methods in medical research.

Li, Q., S. R. Eichten, P. J. Hermanson and N. M. Springer, 2014 Inheritance Patterns and Stability of DNA Methylation Variation in Maize Near-Isogenic Lines. Genetics **196:** 667-676.

Li, X., W. Li, H. Wang, J. Cao, K. Maehashi et al., 2005 Pseudogenization of a sweet-receptor gene accounts for cats' indifference toward sugar. PLoS genetics **1:** e3.

Li, X., X. Wang, K. He, Y. Ma, N. Su et al., 2008 High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. The Plant Cell Online **20:** 259-276.

Li, X., J. Zhu, F. Hu, S. Ge, M. Ye et al., 2012 Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC Genomics **13:** 300.

Li, Y., I. Mukherjee, K. E. Thum, M. Tanurdzic, M. S. Katari et al., 2015 The histone methyltransferase SDG8 mediates the epigenetic modification of light and carbon responsive genes in plants. Genome biology **16:** 79.

Lindquist, S., and E. Craig, 1988 The heat-shock proteins. Annual review of genetics **22:** 631-677.

Lindroth, A. M., X. Cao, J. P. Jackson, D. Zilberman, C. M. McCallum et al., 2001 Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science **292:** 2077-2080.

Lippman, Z., A.-V. Gendrel, M. Black, M. W. Vaughn, N. Dedhia et al., 2004 Role of transposable elements in heterochromatin and epigenetic control. Nature **430:** 471-476.

Luna, E., T. J. Bruce, M. R. Roberts, V. Flors and J. Ton, 2012 Next-generation systemic acquired resistance. Plant Physiology **158:** 844-853.

Marí-Ordóñez, A., A. Marchais, M. Etcheverry, A. Martin, V. Colot et al., 2013 Reconstructing de novo silencing of an active plant retrotransposon. Nature genetics **45:** 1029-1039.

Matzke, M. A., and R. A. Mosher, 2014 RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nature Reviews Genetics **15:** 394-408.

Mauricio, R., 1998 Costs of resistance to natural enemies in field populations of the annual plant Arabidopsis thaliana. The American Naturalist **151:** 20-28.

Mauricio, R., and M. D. Rausher, 1997 Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. Evolution: 1435-1444.

McCue, A. D., S. Nuthikattu, S. H. Reeder and R. K. Slotkin, 2012 Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. PLoS genetics **8:** e1002474.

Mette, M., W. Aufsatz, J. Van der Winden, M. Matzke and A. Matzke, 2000 Transcriptional silencing and promoter methylation triggered by double‐stranded RNA. The EMBO Journal **19:** 5194-5201.

Mirouze, M., M. Lieberman-Lazarovich, R. Aversano, E. Bucher, J. Nicolet et al., 2012 Loss of DNA methylation affects the recombination landscape in Arabidopsis. Proceedings of the National Academy of Sciences **109:** 5880-5885.

MIROUZE, M., and J. PASZKOWSKI, 2011 Epigenetic contribution to stress adaptation in plants. Current opinion in plant biology **14:** 267-274.

MITHÖFER, A., G. WANNER and W. BOLAND, 2005 Effects of feeding Spodoptera littoralis on lima bean leaves. II. Continuous mechanical wounding resembling insect feeding is sufficient to elicit herbivory-related volatile emission. Plant Physiology **137:** 1160-1168.

MIURA, A., S. YONEBAYASHI, K. WATANABE, T. TOYAMA, H. SHIMADA *et al.*, 2001 Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature **411:** 212-214.

MIURA, F., Y. ENOMOTO, R. DAIRIKI and T. ITO, 2012 Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Research **40:** e136.

MOJICA, J. P., Y. W. LEE, J. H. WILLIS and J. K. KELLY, 2012 Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade‐off in Mimulus guttatus. Molecular ecology **21:** 3718-3728.

MOLE, S., 1994 Trade-offs and constraints in plant-herbivore defense theory: a life-history perspective. Oikos**:** 3-12.

MOLINIER, J., G. RIES, C. ZIPFEL and B. HOHN, 2006 Transgeneration memory of stress in plants. Nature **442:** 1046-1049.

MOLNAR, A., C. W. MELNYK, A. BASSETT, T. J. HARDCASTLE, R. DUNN *et al.*, 2010 Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. Science **328:** 872-875.

MONNAHAN, P. J., and J. K. KELLY, 2015 Naturally segregating loci exhibit epistasis for fitness. Biology Letters **11**.

MOUSSEAU, T. A., and C. W. FOX, 1998 The adaptive significance of maternal effects. Trends in Ecology & Evolution **13:** 403-407.

NEWMAN, M., G. P. COMPO and M. A. ALEXANDER, 2003 ENSO-forced variability of the Pacific decadal oscillation. Journal of Climate **16:** 3853-3857.

NIEDERHUTH, C. E., A. J. BEWICK, L. JI, M. ALABADY, K. DO KIM *et al.*, 2016 Widespread natural variation of DNA methylation within angiosperms. bioRxiv**:** 045880.

NUTHIKATTU, S., A. D. MCCUE, K. PANDA, D. FULTZ, C. DEFRAIA *et al.*, 2013 The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. Plant physiology **162:** 116-131.

PAN, J., M. SASAKI, R. KNIEWEL, H. MURAKAMI, HANNAH G. BLITZBLAU *et al.*, A Hierarchical Combination of Factors Shapes the Genome-wide Topography of Yeast Meiotic Recombination Initiation. Cell **144:** 719-731.

PEDERSON, G. T., S. T. GRAY, C. A. WOODHOUSE, J. L. BETANCOURT, D. B. FAGRE *et al.*, 2011 The unusual nature of recent snowpack declines in the North American Cordillera. Science **333:** 332-335.

PIETERSE, C. M., 2012 Prime time for transgenerational defense. Plant physiology **158:** 545-545.

POSADA, D., and T. R. BUCKLEY, 2004 Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic biology **53:** 793-808.

RASMANN, S., M. DE VOS, C. L. CASTEEL, D. TIAN, R. HALITSCHKE *et al.*, 2012 Herbivory in the previous generation primes plants for enhanced insect resistance. Plant physiology **158:** 854-863.

REN, X.-Y., O. VORST, M. W. FIERS, W. J. STIEKEMA and J.-P. NAP, 2006 In plants, highly expressed genes are the least compact. Trends in Genetics **22:** 528-532.

REYMOND, P., N. BODENHAUSEN, R. M. VAN POECKE, V. KRISHNAMURTHY, M. DICKE *et al.*, 2004 A conserved transcript pattern in response to a specialist and a generalist herbivore. The Plant Cell **16:** 3132-3147.

REYMOND, P., H. WEBER, M. DAMOND and E. E. FARMER, 2000 Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. The Plant Cell **12:** 707-719.

RHOADES, D. F., and R. G. CATES, 1976 Toward a general theory of plant antiherbivore chemistry, pp. 168-213 in *Biochemical interaction between plants and insects*. Springer.

RICHARDS, C. L., K. J. VERHOEVEN and O. BOSSDORF, 2012 Evolutionary significance of epigenetic variation, pp. 257-274 in *Plant Genome Diversity Volume 1*. Springer.

ROBERTSON, M., and C. RICHARDS, 2015 Non-genetic inheritance in evolutionary theory-the importance of plant studies. Non-Genetic Inheritance **2**.

ROBINSON, M. D., D. J. MCCARTHY and G. K. SMYTH, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26:** 139-140.

SACHS, M. M., and T.-H. D. HO, 1986 Alteration of gene expression during environmental stress in plants. Annual review of plant physiology **37:** 363-376.

SAITO, S., N. HIRAI, C. MATSUMOTO, H. OHIGASHI, D. OHTA *et al.*, 2004 Arabidopsis CYP707As encode (+)-abscisic acid 8′-hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid. Plant Physiology **134:** 1439-1449.

SAZE, H., K. TSUGANE, T. KANNO and T. NISHIMURA, 2012 DNA Methylation in Plants: Relationship to Small RNAs and Histone Modifications, and Functions in Transposon Inactivation. Plant and Cell Physiology **53:** 766-784.

SCHMITZ, R. J., Y. HE, O. VALDES-LOPEZ, S. M. KHAN, T. JOSHI *et al.*, 2013a Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res **23:** 1663-1674.

SCHMITZ, R. J., M. D. SCHULTZ, M. A. URICH, J. R. NERY, M. PELIZZOLA *et al.*, 2013b Patterns of population epigenomic diversity. Nature **495:** 193-198.

SCOVILLE, A. G., L. L. BARNETT, S. BODBYL‐ROELS, J. K. KELLY and L. C. HILEMAN, 2011 Differential regulation of a MYB transcription factor is correlated with transgenerational epigenetic inheritance of trichome density in Mimulus guttatus. New Phytologist **191:** 251-263.

SEKI, M., M. NARUSAKA, H. ABE, M. KASUGA, K. YAMAGUCHI-SHINOZAKI *et al.*, 2001 Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. The Plant Cell Online **13:** 61-72.

SEYMOUR, D. K., D. KOENIG, J. HAGMANN, C. BECKER and D. WEIGEL, 2014 Evolution of DNA Methylation Patterns in the Brassicaceae is Driven by Differences in Genome Organization. PLoS genetics **10:** e1004785.

SHEN, H., H. HE, J. LI, W. CHEN, X. WANG *et al.*, 2012 Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids. Plant Cell **24:** 875-892.

SI-AMMOUR, A., D. WINDELS, E. ARN-BOULDOIRES, C. KUTTER, J. AILHAS *et al.*, 2011 miR393 and secondary siRNAs regulate expression of the TIR1/AFB2 auxin receptor clade and auxin-related development of Arabidopsis leaves. Plant physiology **157:** 683-691.

SLAUGHTER, A., X. DANIEL, V. FLORS, E. LUNA, B. HOHN *et al.*, 2012 Descendants of primed Arabidopsis plants exhibit resistance to biotic stress. Plant physiology **158:** 835-843.

STROUD, H., T. DO, J. DU, X. ZHONG, S. FENG *et al.*, 2014 Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nature structural & molecular biology **21:** 64-72.

SUNKAR, R., V. CHINNUSAMY, J. H. ZHU and J. K. ZHU, 2007 Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends in Plant Science **12:** 301-309.

SWINDELL, W. R., M. HUEBNER and A. P. WEBER, 2007 Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC genomics **8:** 125.

TAKUNO, S., and B. S. GAUT, 2013 Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. Proceedings of the National Academy of Sciences **110:** 1797-1802.

TARAZONA, S., F. GARCÍA-ALCALDE, J. DOPAZO, A. FERRER and A. CONESA, 2011 Differential expression in RNA-seq: a matter of depth. Genome research **21:** 2213-2223.

TATE, P. H., and A. P. BIRD, 1993 Effects of DNA methylation on DNA-binding proteins and gene expression. Current opinion in genetics & development **3:** 226-231.

THOMÉ, O. W., 1885 Flora von Deutschland.

TREUTTER, D., 2005 Significance of flavonoids in plant resistance and enhancement of their biosynthesis. Plant Biology **7:** 581-591.

TSANG, E., C. BOWLER, D. HÉROUART, W. VAN CAMP, R. VILLARROEL *et al.*, 1991 Differential regulation of superoxide dismutases in plants exposed to environmental stress. The Plant Cell Online **3:** 783-792.

VAN DAM, N., and I. BALDWIN, 2001 Competition mediates costs of jasmonate‐induced defences, nitrogen acquisition and transgenerational plasticity in Nicotiana attenuata. Functional Ecology **15:** 406-415.

VERHOEVEN, K. J., J. J. JANSEN, P. J. VAN DIJK and A. BIERE, 2010 Stress‐induced DNA methylation changes and their heritability in asexual dandelions. New Phytologist **185:** 1108-1118.

VERHOEVEN, K. J., and T. P. VAN GURP, 2012a Transgenerational effects of stress exposure on offspring phenotypes in apomictic dandelion. PloS one **7:** e38605.

WALKER, M., D. WALKER, J. WELKER, A. ARFT, T. BARDSLEY *et al.*, 1999 Long-term experimental manipulation of winter snow regime and summer temperature in arctic and alpine tundra. Hydrological processes **13:** 2315-2330.

WALLING, L. L., 2000 The myriad plant responses to herbivores. Journal of Plant Growth Regulation **19:** 195-216.

WANG, J., N. C. MAROWSKY and C. FAN, 2014 Divergence of Gene Body DNA Methylation and Evolution of Plant Duplicate Genes. PLoS ONE **9:** e110357.

WANG, W.-S., Y.-J. PAN, X.-Q. ZHAO, D. DWIVEDI, L.-H. ZHU *et al.*, 2010 Drought-induced site-specific DNA methylation and its association with drought tolerance in rice (Oryza sativa L.). Journal of experimental botany**:** erq391.

WASSENEGGER, M., 2000 RNA-directed DNA methylation, pp. 83-100 in *Plant Gene Silencing*. Springer.

WEHRENS, R., and L. M. BUYDENS, 2007 Self-and super-organizing maps in R: the Kohonen package. Journal of Statistical Software **21:** 1-19.

WEST-EBERHARD, M. J., 1989 Phenotypic Plasticity and the Origins of Diversity. Annual Review of Ecology and Systematics **20:** 249-278.

WICKHAM, H., 2009 *ggplot2: elegant graphics for data analysis*. Springer.

WINK, M., 1988 Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. Theoretical and Applied Genetics **75:** 225-233.

WU, J., and I. T. BALDWIN, 2010 New insights into plant responses to the attack from insect herbivores. Annual review of genetics **44:** 1-24.

XIA, J., L. HAN and Z. ZHAO, 2012 Investigating the relationship of DNA methylation with mutation rate and allele frequency in the human genome. BMC Genomics **13:** 1-9.

XIONG, W., L. HE, J. LAI, H. K. DOONER and C. DU, 2014 HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proceedings of the National Academy of Sciences **111:** 10263-10268.

YAN, Y., E. BORREGO and M. V. KOLOMIETS, 2013 Jasmonate biosynthesis, perception and function in plant development and stress responses.

YANG, H., F. CHANG, C. YOU, J. CUI, G. ZHU *et al.*, 2015 Whole‐genome DNA methylation patterns and complex associations with gene structure and expression during flower development in Arabidopsis. The Plant Journal **81:** 268-281.

YANG, S. F., and N. E. HOFFMAN, 1984 Ethylene biosynthesis and its regulation in higher plants. Annual review of plant physiology **35:** 155-189.

YANG, X., Y. YU, L. JIANG, X. LIN, C. ZHANG *et al.*, 2012 Changes in DNA methylation and transgenerational mobilization of a transposable element (mPing) by the Topoisomerase II inhibitor, Etoposide, in rice. BMC Plant Biology **12:** 1-16.

YOKOYAMA, T., F. MIURA, H. ARAKI, K. OKAMURA and T. ITO, 2015 Changepoint detection in base-resolution methylome data reveals a robust signature of methylated domain landscape. BMC genomics **16:** 594.

YU, J., S. HU, J. WANG, G. K.-S. WONG, S. LI *et al.*, 2002 A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. indica). Science **296:** 79-92.

YU, Y.-B., D. O. ADAMS and S. F. YANG, 1979 1-Aminocyclopropanecarboxylate synthase, a key enzyme in ethylene biosynthesis. Archives of biochemistry and biophysics **198:** 280-286.

YUAN, Y., L. GUO, L. SHEN and J. S. LIU, 2007 Predicting Gene Expression from Sequence: A Reexamination. PLoS Comput Biol **3:** e243.

ZHANG, W., T. ZHANG, Y. WU and J. JIANG, 2012 Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. The Plant Cell Online **24:** 2719-2731.

ZHANG, X., J. YAZAKI, A. SUNDARESAN, S. COKUS, S. W. CHAN *et al.*, 2006 Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell **126:** 1189-1201.

ZHONG, S., Z. FEI, Y. R. CHEN, Y. ZHENG, M. HUANG *et al.*, 2013 Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat Biotechnol **31:** 154-159.

ZILBERMAN, D., M. GEHRING, R. K. TRAN, T. BALLINGER and S. HENIKOFF, 2006 Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature genetics **39:** 61-69.

ZOMER, R. J., A. TRABUCCO, D. A. BOSSIO and L. V. VERCHOT, 2008 Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. Agriculture, ecosystems & environment **126:** 67-80.

Appendix 1 – A) Based on experimental population for RNAseq analysis in this study, progeny of damaged plants produce significantly more trichomes than progeny of control plants (p<0.025). B) Trichome density determined in a follow-up experiment accounting for parent plant of origin as a nested variable in ANOVA. Treatment e ect (p=0.0003), parent e ect (p=0.523). C) Based on RNAseq, progeny of damaged plants expressMGMYBML-8 at significantly lower levels than the progeny of control plants (DESeq2: p-value=0.01, EdgeR: p-value=0.02).
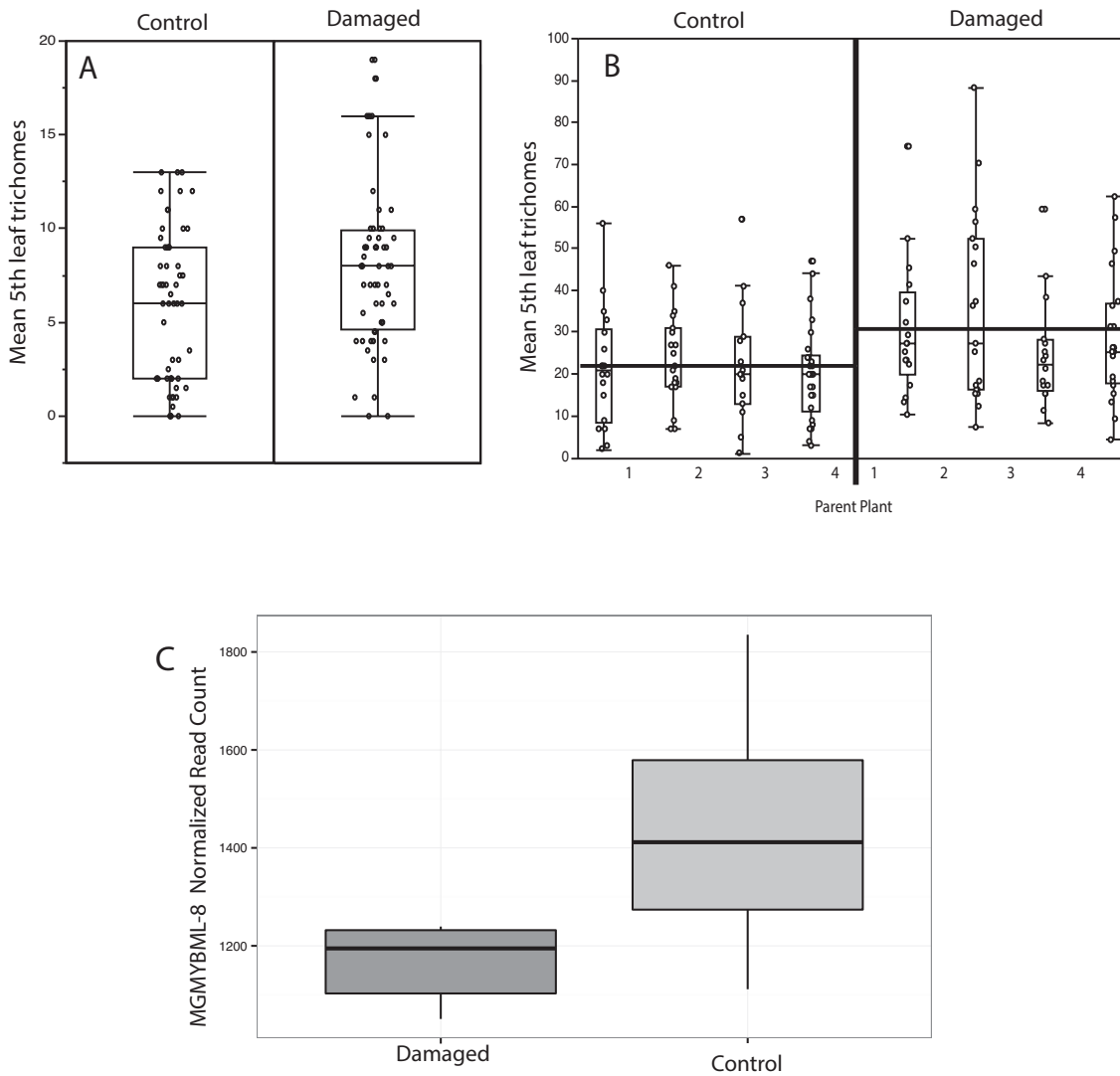
Appendix 2 – Primers used for RT-qPCR transcript expression confirmation.

| DE transcript ID	Gene ID	Position in M. guttatus genome V 2.0	Primer sequence
1428104	Heat Shock 6AB	Scaffold_4 1671299- 1673148	Primer 1
			Forward: GTTGGCAAACTCCCATTTGT
			Reverse: CGGAAGCTTTGTTGTGTGG
			Product size mRNA: 125
			Product size DNA: 420
1324230	Adipocyte plasma membrane-associated (Strictosidine synthase like-2)	Scaffold_10 18088684-18094197	Forward: TCTGCAAGTTGTGGTCTCCA
			Reverse: CGATCAAAATAGCGGAGTCG
			Product size mRNA: 150
			Product size DNA: 1,500
1358627	Tyrosine Aminotransferase	Scaffold_13 15225541- 15228255	Forward: CATCGGCACAAAAGGGTTAT
			Reverse: AATCCGTGTGGGAACGTTTA
			Product size mRNA: 130
			Product size DNA: 245
1444264	DNAJ HSP-40	Scaffold_4 20756516-20757629	Forward: ATGCACCAGGCTTACGAAAC
			Reverse: ACTGCCATTTCGTTTTCCAC
			Product size mRNA: 150
			Product size DNA: 262
1495616	Dormancy/ Auxin Associated Protein 1A	Scaffold_8 16248497-16249443	Forward: CGCGTAAGGATAACGTGTGG
			Reverse: CACCGGTGACTCCAAATCTT
			Product size mRNA: 137
			Product size DNA: 255
1315072	CHY RING Zinc Protein	Scaffold_10 2108572- 2111820	Forward: TCCTCTAGTCTGCCGCGTAT
			Reverse: TGCGGTCACACCATACATCT
			Product size mRNA: 172
			Product size DNA: 560

Appendix 3 – Number of RNaseq reads generated and mapped for each of twelve individuals.

| Plant ID	Number of Reads Mapped to Refrence Genome	Total Number of Reads	Percent of Total Reads Mapped to Refrence Genome
C1a	75,195,595.00	87,406,770.00	86.03
C1b	79,573,890.00	99,489,738.00	79.98
C2a	90,106,672.00	104,062,180.00	86.59
C2b	74,475,264.00	93,290,834.00	79.83
C3a	72,256,647.00	90,261,332.00	80.05
C3B	82,777,367.00	95,900,734.00	86.32
D1a	90,915,696.00	105,509,888.00	86.17
D1b	78,848,660.00	98,351,402.00	80.17
D2a	75,934,492.00	95,053,668.00	79.89
D2b	100,060,101.00	115,663,080.00	86.51
D3a	83,642,660.00	96,900,302.00	86.32
D3b	85,171,859.00	106,859,744.00	79.7
	988,958,903.00	1,188,749,672.00	83.13

Appendix 4 – Genotype in chromosome 14 region of heterozygosity.

| Individual	Genotype in Region of Heterozygosity
D1A	IM
D1B	Het
D2A	PR
D2B	PR
D3A	Het
D3B	IM
C1A	IM
C1B	Het
C2A	PR
C2B	IM
C3A	Het
C3B	Het

Appendix 5 - Relative expression, with minimum individual expression for each gene set to equal one, based on RNAseq and RT-qPCR gene expression results for 6 differentially expressed genes. Three up regulated transcripts, and three down-regulated transcripts were confirmed for differential expression in two offspring each from three damaged and three control parental plants. Heat Shock Protein 6ab, Strictosidine Synthase, Tyrosine Aminotransferase, Heat Shock Protein 40, Dormancy Associated Protein, and CHY Zinc Finger (transcript ID#s 1428104, 1324230, 1358627, 1444264, 1495616, and 1315072, respectively). Bars represent group averages, and points represent individual sample expression.
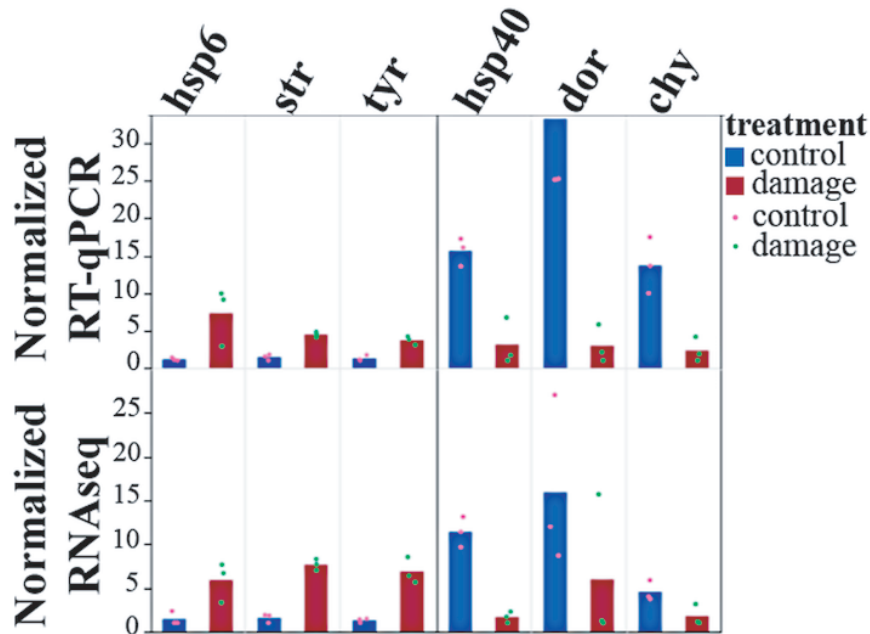
Appendix 6 – Scatterplots showing fold change by mean expression for genes differentially expressed in at least one of five programs (FDR<0.1). a-e) Of the genes found differentially expressed in at least one program, scatterplots showing those genes found to be differentially expressed in individual programs (FDR<0.1); DESeq2, limma, EdgeR, NOISeq and SAMSeq, respec- tively. f) Scatterplot showing fold change by mean expression for individual genes found to be differentially expressed in at least one program, color coded by the number of programs that found an individual gene to be differentially expressed.
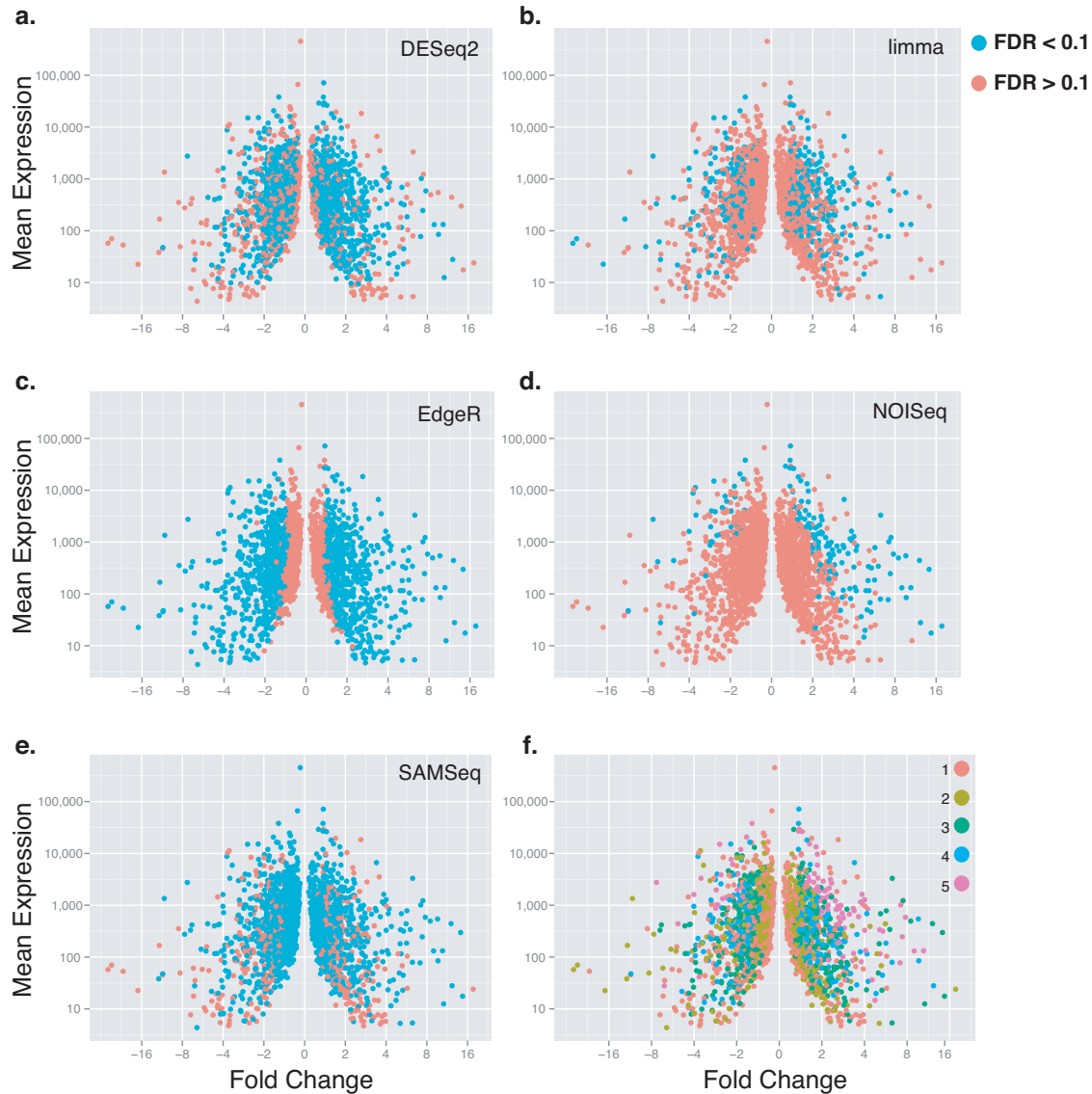
Appendix 7 – A) 483 genes found to be differentially expressed (FDR<0.1) with Limma using family structured analysis. B) 1276 genes found to be differentially expressed (FDR<0.1) with EdgeR using family structured analysis. C) Venn diagram showing the significant overlap between genes found differentially expressed using consensus methodology, and the two programs which account for parent line in the analysis.
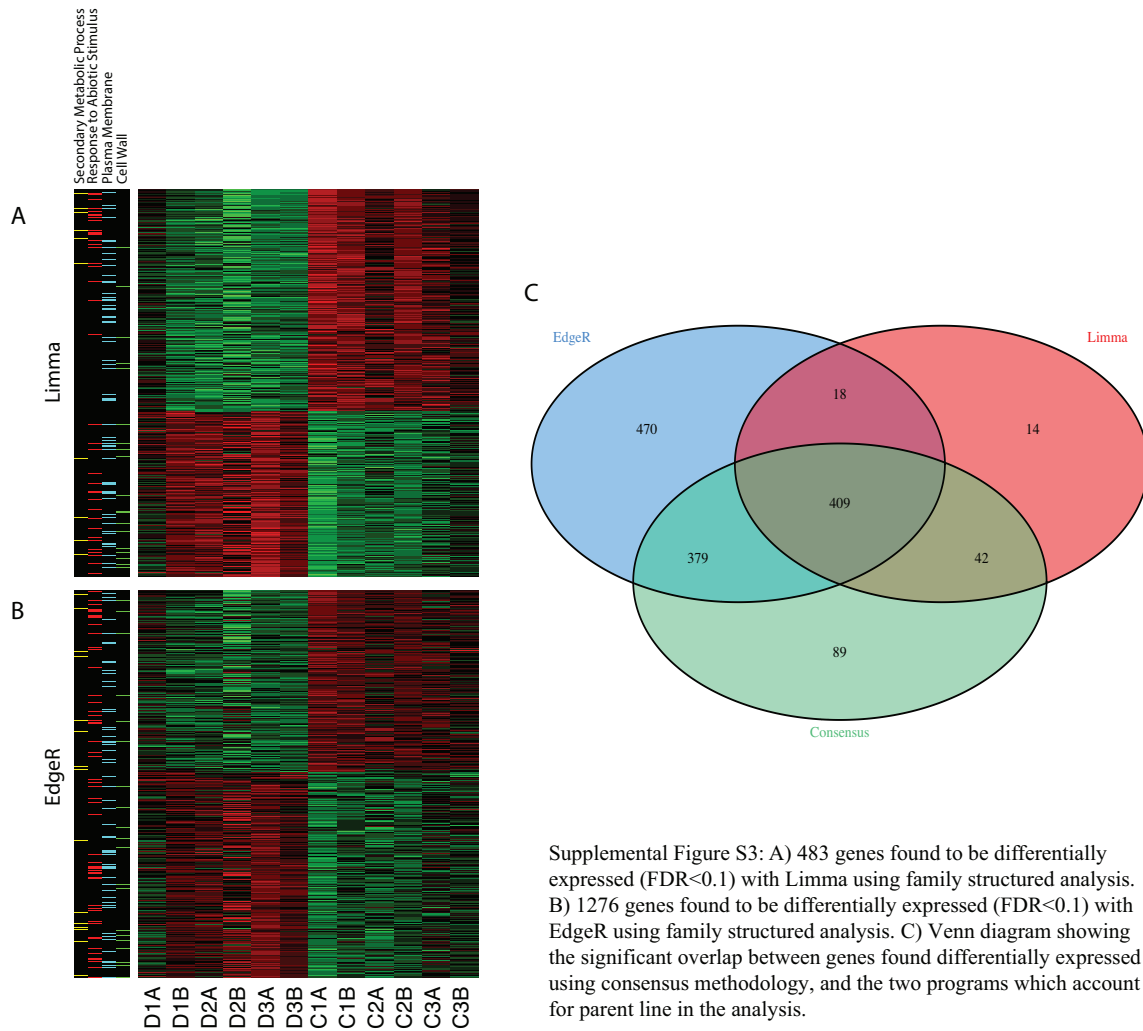
Supplemental Figure S3: A) 483 genes found to be differentially expressed (FDR<0.1) with Limma using family structured analysis. B) 1276 genes found to be differentially expressed (FDR<0.1) with EdgeR using family structured analysis. C) Venn diagram showing the significant overlap between genes found differentially expressed using consensus methodology, and the two programs which account for parent line in the analysis.

Appendix 8 – GO terms that show significant enrichment or depletion in the set of differentially expressed genes relative to the seedling transcriptome.

GO Term	GO Category	Fisher's Exact Test p-value for Enrichment/Depletion	Fisher's Exact Test FDR for Enrichment/Depletion	Over/Under	Differentially Expressed with GO Term	Whole Genome with GO Term
cell periphery	Cellular Component	1.92E-08	1.19E-05	Enriched	0.22	0.14
membrane	Cellular Component	4.74E-07	1.30E-04	Enriched	0.34	0.26
response to stimulus	Biological Process	6.31E-07	1.30E-04	Enriched	0.32	0.23
response to abiotic stimulus	Biological Process	1.55E-06	2.12E-04	Enriched	0.15	0.10
plasma membrane	Cellular Component	1.71E-06	2.12E-04	Enriched	0.19	0.13
cellular aromatic compound	Biological Process	1.42E-05	8.05E-04	Depleted	0.06	0.11
organic cyclic compound	Biological Process	1.42E-05	8.05E-04	Depleted	0.06	0.11
heterocycle metabolic process	Biological Process	1.42E-05	8.05E-04	Depleted	0.06	0.11
cellular nitrogen compound	Biological Process	1.42E-05	8.05E-04	Depleted	0.06	0.11
nucleobase-containing compound	Biological Process	1.42E-05	8.05E-04	Depleted	0.06	0.11
nitrogen compound	Biological Process	1.43E-05	8.05E-04	Depleted	0.06	0.11
gene expression	Biological Process	3.31E-05	1.71E-03	Depleted	0.01	0.04
cell wall	Cellular Component	3.85E-05	1.84E-03	Enriched	0.06	0.03
external encapsulating structure	Cellular Component	4.34E-05	1.92E-03	Enriched	0.06	0.03
extracellular region	Cellular Component	2.28E-04	9.43E-03	Enriched	0.08	0.05
cell cycle	Biological Process	3.08E-04	1.19E-02	Depleted	0.01	0.03
response to stress	Biological Process	3.48E-04	1.27E-02	Enriched	0.20	0.15
mitochondrion	Cellular Component	7.72E-04	2.66E-02	Depleted	0.05	0.09
RNA binding	Mollecular Function	8.30E-04	2.71E-02	Depleted	0.01	0.03
macromolecular complex	Cellular Component	1.99E-03	5.04E-02	Depleted	0.01	0.03
translation	Biological Process	2.00E-03	5.04E-02	Depleted	0.01	0.03
macromolecule	Biological	2.02E-03	5.04E-02	Depleted	0.01	0.03

biosynthetic process	Process			d		
cellular biosynthetic process	Biological Process	2.02E-03	5.04E-02	Depleted	0.01	0.03
cellular macromolecule	Biological Process	2.02E-03	5.04E-02	Depleted	0.01	0.03
organic substance biosynthetic process	Biological Process	2.03E-03	5.04E-02	Depleted	0.01	0.03
ribosome	Cellular Component	2.72E-03	6.24E-02	Depleted	0.01	0.03
ribonucleoprotein complex	Cellular Component	2.72E-03	6.24E-02	Depleted	0.01	0.03
nuclease activity	Mollecular Function	3.22E-03	6.98E-02	Depleted	0.00	0.01
hydrolase activity, ester bonds	Mollecular Function	3.31E-03	6.98E-02	Depleted	0.00	0.01
cellular metabolic process	Biological Process	3.38E-03	6.98E-02	Depleted	0.19	0.24
cytoplasmic part	Cellular Component	4.55E-03	9.10E-02	Depleted	0.27	0.32

Appendix 9 – Heat map showing twenty-eight differentially expressed genes coding for heat shock proteins resulting from parent leaf damage. Asterisk denotes the one heat shock gene downregulated in response to parent leaf damage.
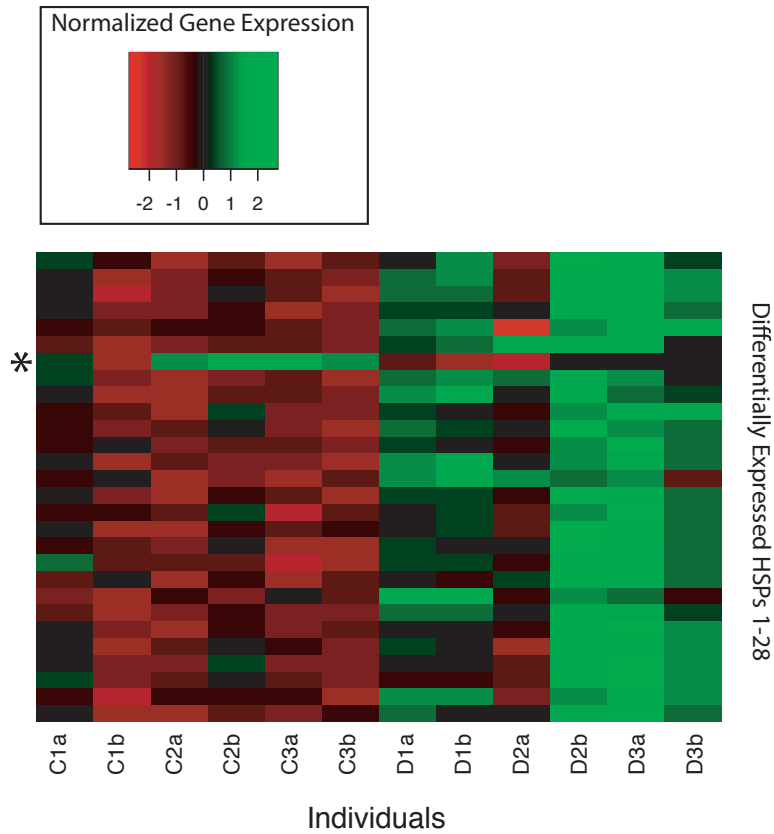
Appendix 10 – Predicted log(gene expression) from cubic polynominal REML model compared to actual log(gene expression). Slope =1.02, R^2=0.201, df=28.
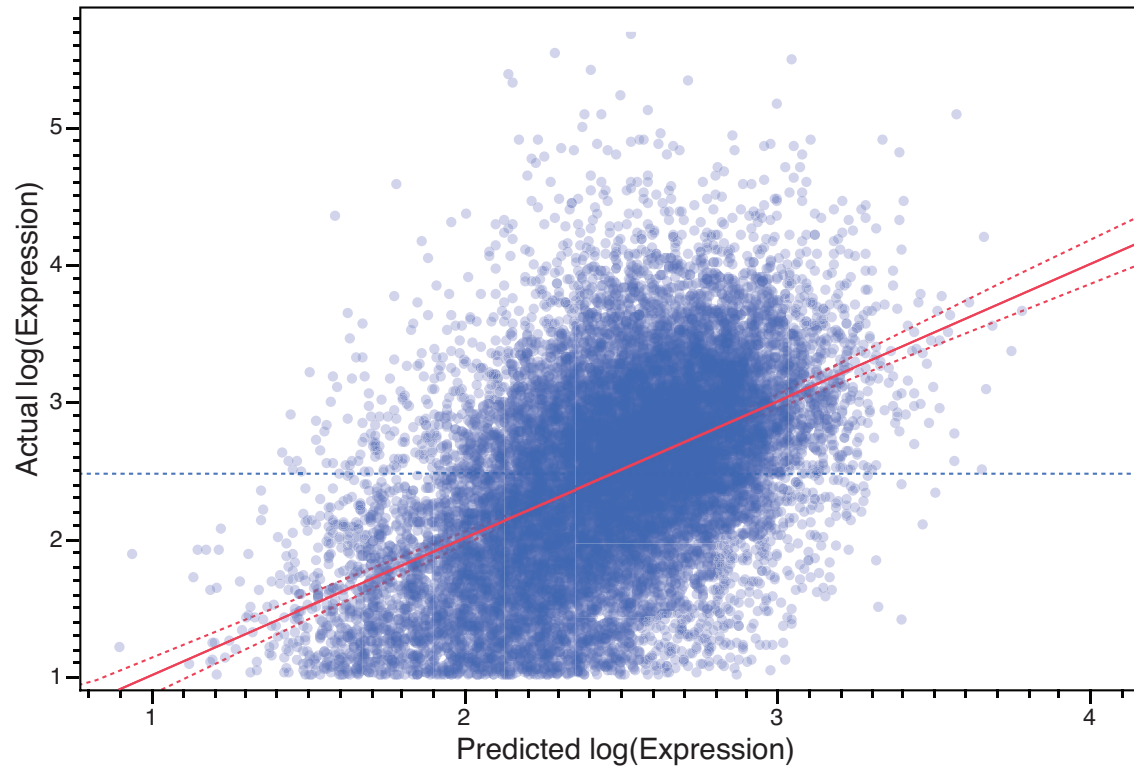
Appendix 11 – Read depth and percent methylation in the nine individuals sequenced in this study.

Treatment	Number of Reads	Average Depth	CG	CHG	CHH
OC1	1888902727	6.4	70.45	35.90	14.95
OC2	1877864952	6.4	58.85	26.80	11.60
OC3	1867678993	6.3	67.60	30.60	12.30
OC4	1778153365	6	70.00	36.80	13.90
OD1	1219264220	4.2	73.10	40.30	14.20
OD2	1860074372	6.3	72.76	39.93	14.28
OD3	2386764313	8.1	72.90	39.50	14.15
OD4	1941206754	6.6	71.25	39.40	16.30
OD5	2797016391	9.5	69.05	34.25	14.00

Appendix 12 – MDM plots for two representative individuals in this study. (A) OD1, (B) OD1.

A
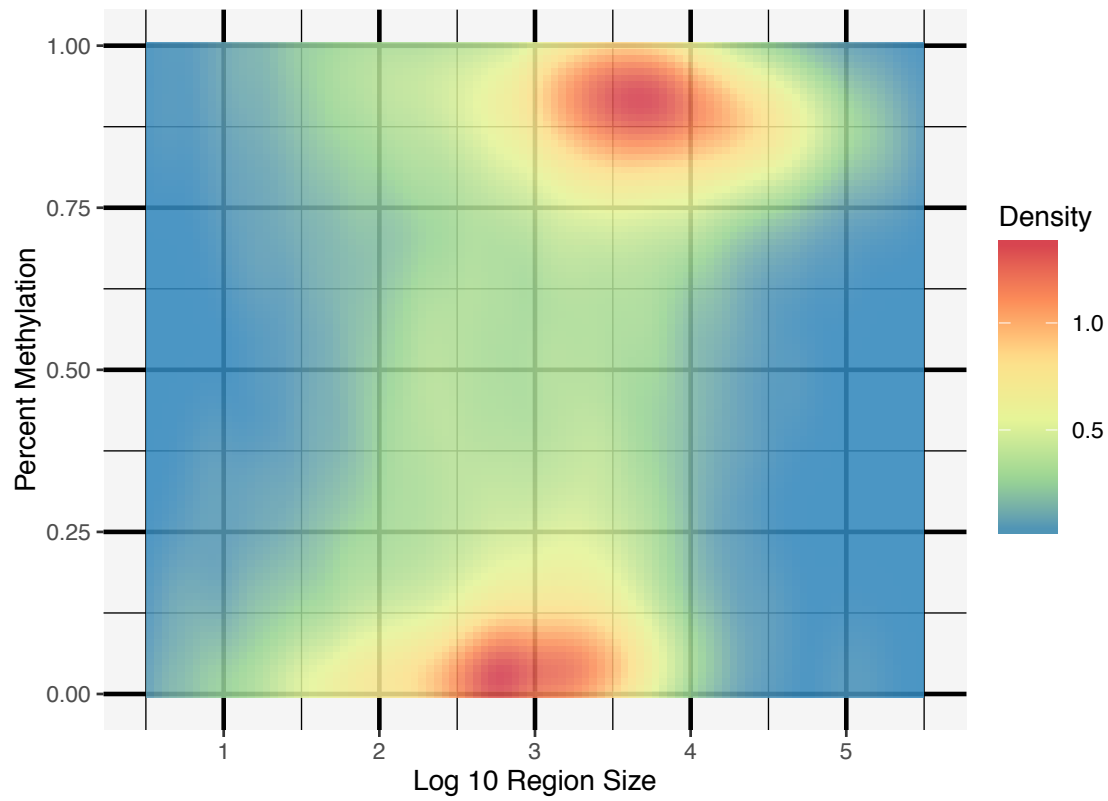
B
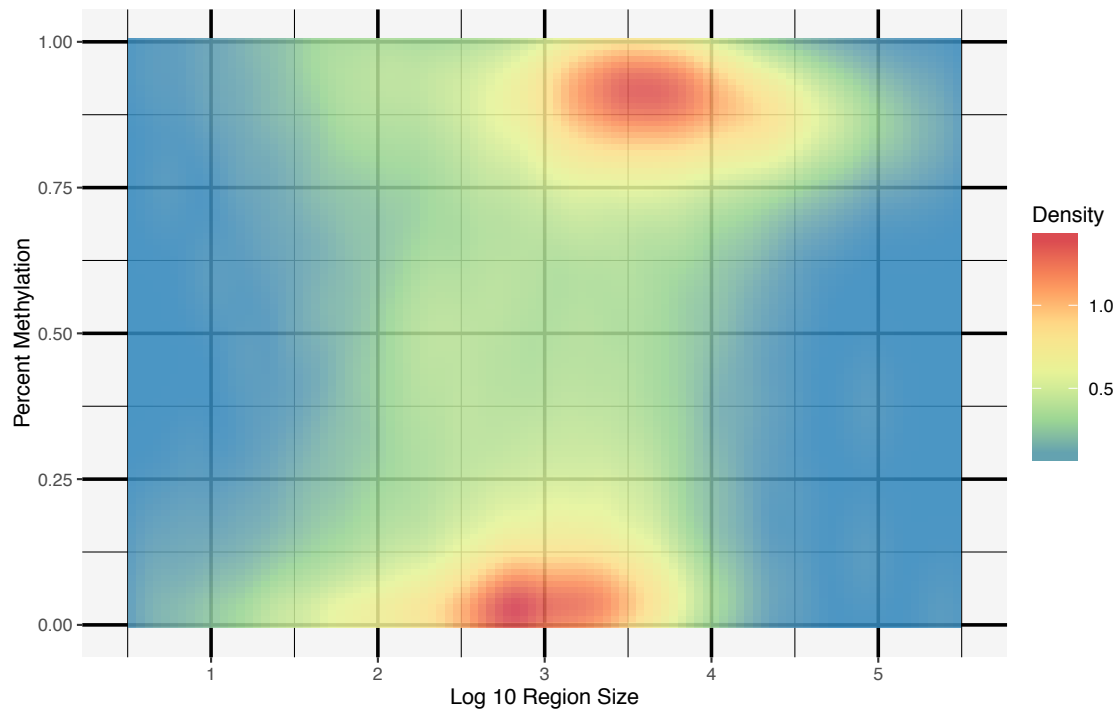
Appendix 13 – Geographical and environmental variables alongside trichome production and herbivory for each of the study populations. Also, correlation between variables and trichome production.

Population	Aridity	Elevation	Lat	Long	Herbivory	Trichomes
LPM	12095	288.3	-122.61	43.83	0.08	7.51
CSR	16418	236.2	-122.48	44.40	0.09	10.78
MO	12277	314.9	-122.57	43.81	0.19	15.94
HOL	14180	211.5	-122.72	44.09	0.37	34.55
Mona	16961	449.9	-122.27	44.19	0.36	38.36
BR2	23475	1219.2	-122.13	44.37	0.40	49.58
SX	11413	307.8	-122.54	43.79	0.43	59.67
TBR	17492	675.1	-122.05	44.29	0.48	69.06
CGR	16593	528.8	-122.25	44.13	0.31	78.97
MRT	16260	526.4	-122.05	44.21	0.38	81.12
BKM	25072	1200.9	-122.30	44.23	0.63	89.94
MTC	25886	1481.0	-122.14	44.28	0.55	91.19
LPD	12284	265.8	-122.76	43.92	0.60	91.52
MWL	26171	1223.8	-121.78	44.19	0.77	97.44
WC	16670	89.0	-123.63	44.00	0.67	97.53
IM	27094	1432.6	-122.15	44.40	0.55	127.50
Cor. with Trichomes	**0.59**	**0.55**	**0.3**	**0.14**	**0.86**	

Appendix 14 - Crossing design. Six families from each of eight populations were brought back to the greenhouse and crossed. For one version of each family parents were wounded throughout seed set (red circles), while another set of parents were not damaged to use as a control group (blue circles).
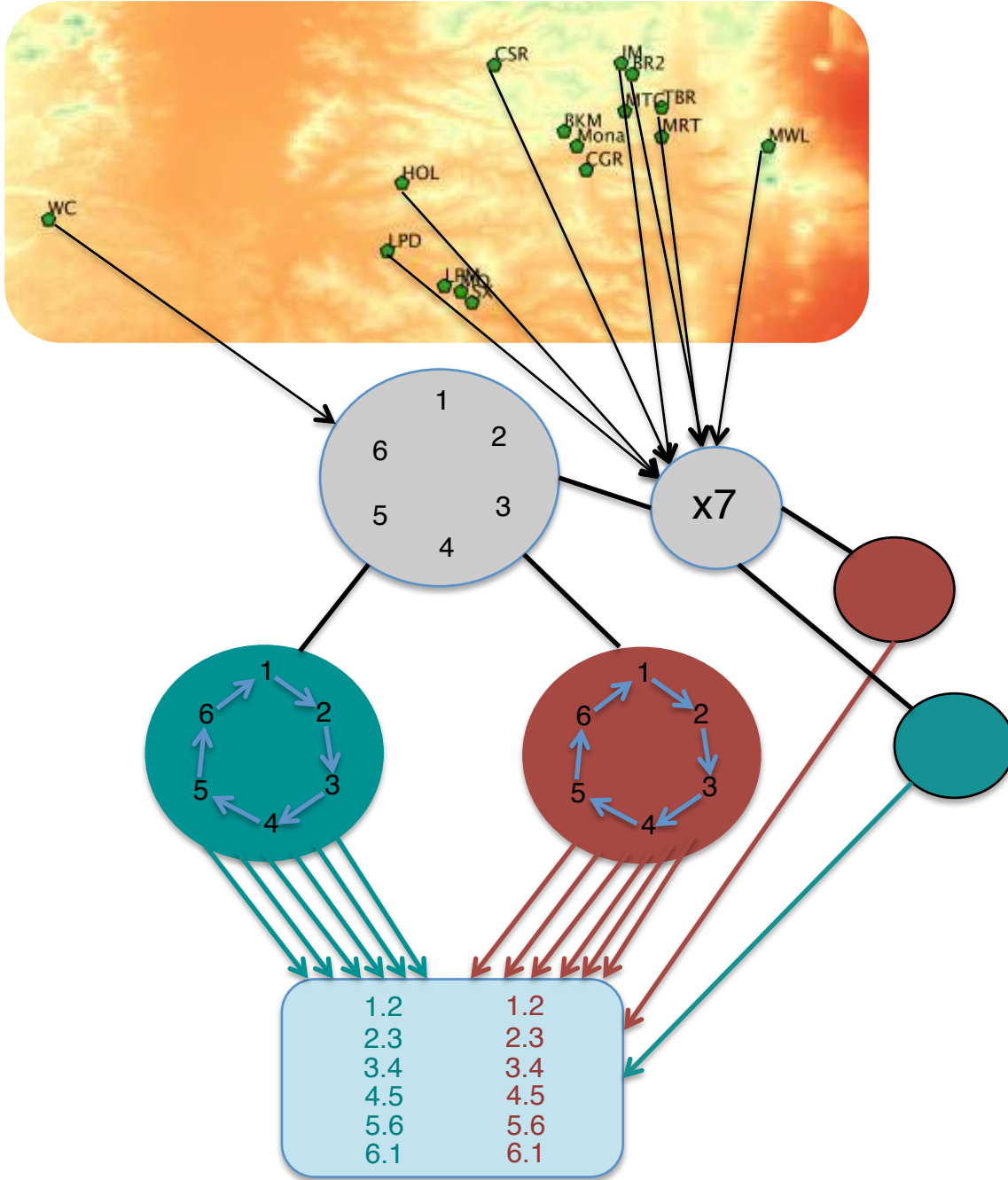
Appendix 15 - Mean glandular and non-glandular trichome counts as well as herbivory (0 represents no herbivory, 1 represents minor, 2 represents moderate) for sixteen populations across three leaf pairs. Error Bar is standard error.
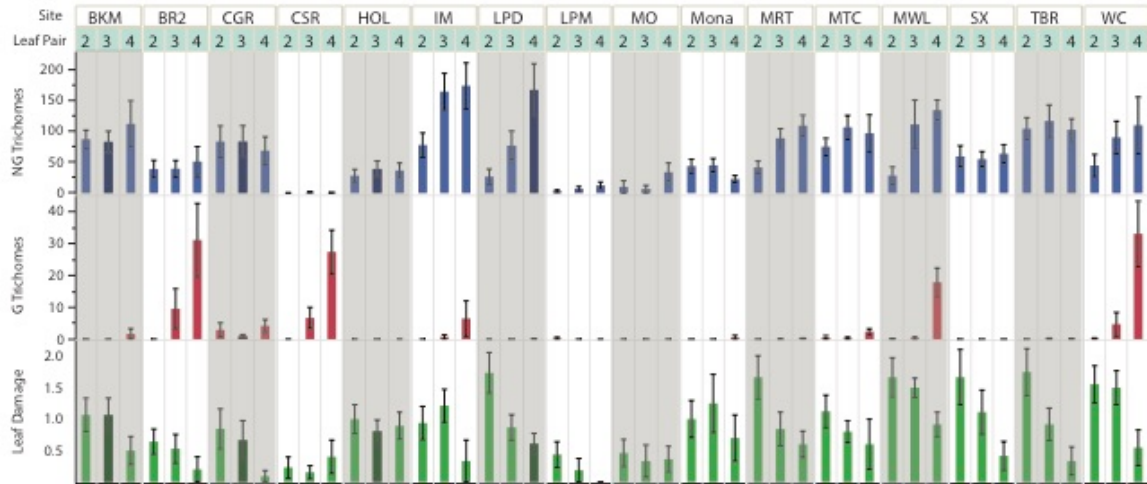
Appendix 16 - Effect of various factors on leaf trichome density and herbivory across sixteen populations

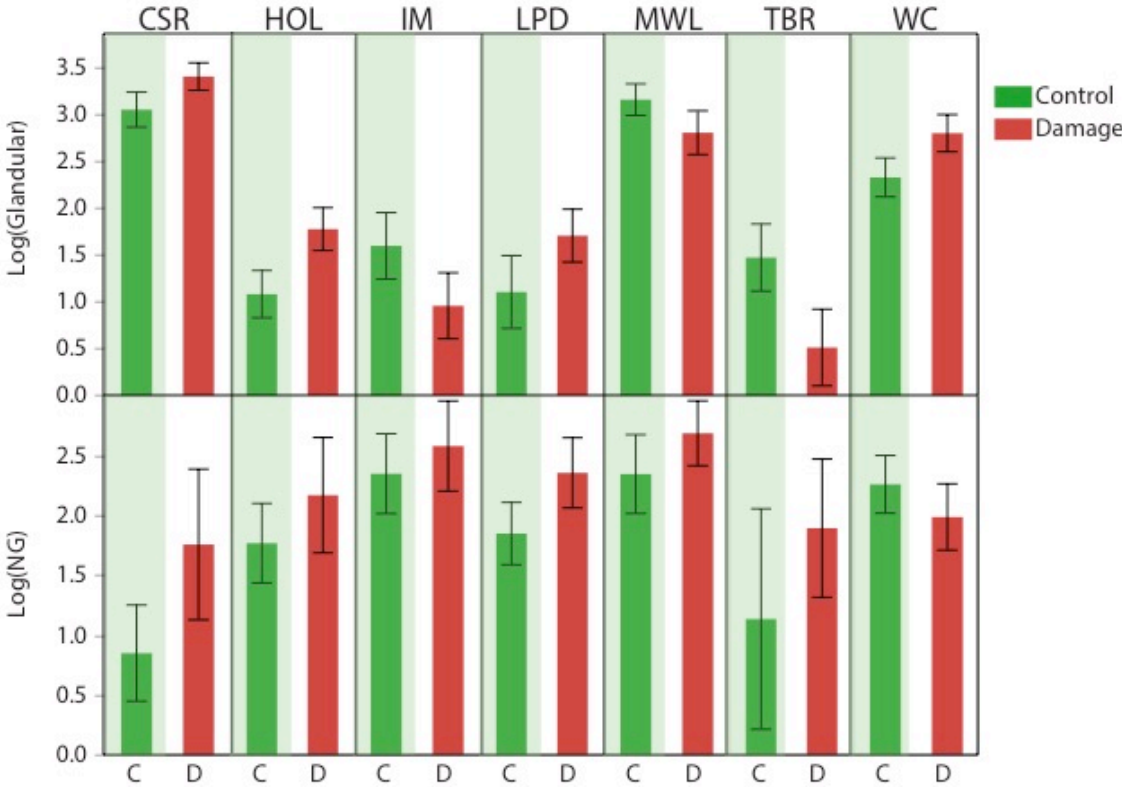|	Log(NG Trichomes)			Log(G Trichomes)			Leaf Damage		
	Df	ChiSq	P-Value	Df	ChiSq	P-Value	Df	ChiSq	P-Value
Population	15	147.26***	$<2.2 \times 10^{-16}$	12	28.59	0.0045	15	54.38***	2.28×10^{-6}
Node	2	49.79***	1.54×10^{-11}	2	18.43	9.97×10^{-5}	2	25.97***	2.30×10^{-6}
Pop*Node	30	121.01***	6.89×10^{-13}	11	27.24	0.0043	30	47.8*	0.021
Plant(Pop)	1			1			1		
Residual	459			58			605		

Appendix 17 - Effect of growth environment (field vs. greenhouse) on 2nd leaf trichome density.

	Df	Log(NG Trichomes)		Log(Glan)	
		F Ratio	P-Value	F-Ratio	P-Value
Pop	7	9.08***	<.0001	94.07***	<.0001
Envi	1	1.06	0.303	26.93***	<.0001
Pop*Envi	7	3.60**	0.0011	X	X
Residual	218	N=234		N=79	

Appendix 18 - Effect of leaf wounding plasticity on 7th leaf trichome density.

	Df	Log(NG Trichomes)		Log(Glan)	
		ChiSq	P-Value	ChiSq	P-Value
Pop	6	10.68	0.099	55.05***	<4.5x10^{-10}
Dam	1	3.16	0.074	1.43	0.265
Pop*Dam	6	4.3	0.635	14.90*	0.021
Model	13	N=172		N=190	

Appendix 19 – Phenotypic plasticity of 7th leaf trichome density in response to control or wounding treatment.

Appendix 20 – Effect of various factors on the trichome density, or presence absence of trichomes.

	Df	Log(NG Trichomes)		Log(Glan)		P/A NG		P/A G	
		ChiSq	P-Val	ChiSq	P-Value	ChiSq	P-Val	ChiSq	PVal
Pop	7	19.5*	0.0068	6.59	0.306	13.74	0.056	38.75***	2.2×10^{-6}
Site	1	0.15	0.7	0.025	0.874	5.72*	0.017	0.43	0.51
P-Dam	1	0.269	0.604	2.39	0.124	4.16*	0.042	7.1**	0.0077
Pop*P-Dam	7	20.25**	0.0025	10.45	0.107	8.23	0.32	6.2	0.52

*: p<0.05
**:p<0.005
***:p<0.0005