Characterizing allelic regulatory variation in maize

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Amanda Jo Waters

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Nathan M. Springer

May 2016

Acknowledgements

First, I would like to thank my advisor, Nathan Springer for his continual support and encouragement during my time in his lab. I have grown so much academically and as a person throughout my Ph.D. I would also like to thank the members of my advisory committee: Robert Stupar, Peter Tiffin, and Chad Myers for input on my thesis as well as encouragement to pursue my goals. I have received so much support from additional faculty, staff, and the other members of the Springer lab during my time at the University. I would like to personally thank Candy Hirsch, Cory Hirsch, Qing Li, Steve Eichten, Ruth Swanson-Wagner, Peter Hermanson, Jackie Noshay, and Irina Makarevitch for their support and help.

I have also been fortunate enough to build a support network of wonderful friends and collogues. In particular, I would like to thank Erin Treiber, Mohamed Yakub, Beth Fallon, Amanda Gorton, and Gail Kalli for always listening, supporting, and believing in me even when I didn't believe in myself.

I would also like to thank my mom, Fran Waters for instilling a love of learning and knowledge from a very young age and always supporting me through the good times and rough patches. I owe so much of my success to her. Finally, I would like to thank my little sister, Sam, who has always given me encouragement and provided a necessary outlet of fun and laughter.

Abstract

Differential gene expression is a major source of phenotypic diversity and can provide the variation necessary for natural selection. The variable expression of different alleles of the same genes can result from complex regulatory mechanisms. Studying the variable expression of different alleles has been limited by technical challenges in comparing expression of closely related sequences. The implementation of new technologies such as RNAseq has allowed for the study of allelic expression and regulatory variation on a genome-wide scale. The goals of this thesis were to develop methods for studying allele-specific expression and to apply these methods to characterize differential expression of alleles in maize to study the mechanisms underlying two complex examples of gene regulation: genomic imprinting and abiotic stress response. The first set of experiments revealed the prevalence and conservation of imprinting within maize and between maize and other species as well as characterized expression and regulatory differences for imprinted genes. Hundreds of genes exhibit imprinted expression, which is the biased expression of alleles based on parent of origin. Imprinted expression of particular genes is highly conserved in maize, but related genes are rarely imprinted in other plant species. This work also found that examples of maternal and paternally imprinted genes have distinct expression and epigenetic profiles. The second set of experiments provided insights into regulatory variation between maize lines in response to abiotic stress. Substantial variation was observed for steady-state expression levels as well as stress-induced expression changes. Cis- regulatory variation is the predominant regulatory variation pattern for both steady-state and stress-induced expression. By documenting allelic variation for these two cases of complex regulation, imprinting and abiotic stress response, we can begin to understand the mechanisms that underlie this regulation and how complex regulatory variation may evolve.

Table of Contents

List of Tables

List of Figures

CHAPTER I

Literature review of expression and regulatory variation

ROLES OF DIFFERENTIAL EXPRESSION AND REGULATORY VARIATION IN PHENOTYPIC DIVERSITY

Changes in gene expression are a major source of variation influencing phenotypic variation within, and between, species. Examples of phenotypic diversity due to differential expression have been documented in many diverse systems (Wray et al., 2007; Carroll 2005). Variation in expression can be the result of changes in the genetic sequence between alleles. Sequence differences, including single nucleotide polymorphisms (SNPs) within the gene or in regulatory regions can result in differential expression of alleles and variation in phenotype (van Oeveren and Janssen 2009; Albert and Kruglyak 2015). Additionally, other changes such as insertions/deletions (indels), presence/absence variation (PAVs), and copy number variation (CNVs) have also been shown to result in heritable differences between genotypes (Springer et al., 2009; Swanson-Wagner et al., 2010; Hirsch et al., 2014). Changes in the DNA sequence alone do not explain all of the variation observed in phenotypes. Another source of variation that can result in differential expression of alleles and variation in traits is epigenetic variation. Epigenetic variation includes heritable changes that are not directly related to differences in the DNA sequence. Epigenetic changes including DNA methylation can also result in variation in phenotypes (Johannes et al., 2008; Vaughn et al., 2007; Richards et al., 2006). Different transposable element (TE) families have been shown to have variable levels of DNA methylation spreading past the element into regulatory regions of genes nearby (Eichten et al., 2012). Therefore, polymorphism in TE elements and variation in the amount of methylation flanking transposable elements may drive differential expression of alleles. Additionally, the presence or absence of certain histone modifications is associated with gene expression, so modification of certain amino acids on histones can also alter expression of alleles. Likely genetic, epigenetic, a combination

of genetic and epigenetic, and additional mechanisms are working in concert to regulate gene expression and phenotype.

Understanding the extent and mechanisms of regulatory variation between alleles has important applications beyond basic scientific knowledge, including technology development, predicting allele by environment interactions in terms of breeding, and local adaptation in natural populations. Assessing allelic diversity and regulation has been difficult due to the highly complex nature of most regulatory mechanisms, the repetitive structure of intergenic space, and availability of technologies that can differentiate between alleles. Through the use of next generation sequencing it is possible to differentiate between alleles and begin to study putative regulatory regions and variation for thousands of genes. Outlined in this review are two interesting cases for studying allele specific expression and regulatory variation: genomic imprinting (parent of origin effects) and allelic variation for response to stress. Although these two examples involve different biological phenomena, similar tools and methodology can be applied to both examples as the main goal is to differentiate between alleles and study complex regulatory mechanisms.

CASE I: INTRODUCTION TO GENOMIC IMPRINTING

Many plant species require genetic contribution from both a maternal and paternal parent to produce viable offspring. In plants, the double fertilization event during sexual reproduction results in the creation of diploid (2n) embryo and triploid (3n) endosperm. The embryo inherits one maternal and one paternal genome whereas the endosperm inherits two maternal genomes and one paternal genome. Imbalances in parental genomic dosage disrupt embryo and endosperm growth, may be involved in reproductive isolation of newly emerging species, and create barriers to hybridization (Jullien and Berger, 2010; Bushell et al., 2003). As a triploid tissue, the endosperm does not have equal genic representation from both parents and mechanisms are in place to maintain the 2 maternal**:** 1 paternal genomic balance of alleles (Bushell et al., 2003). Imprinted genes, on the other hand, do not exhibit the expected 2**:**1 ratio in expressed RNA, but instead

show a differential expression of maternally and paternally inherited alleles (Huh et al., 2008; Jullien and Berger, 2010; Deng and Disteche, 2010).

Imprinting was initially discovered in flowering plants, but has since been documented in mammals (Kermicle, 1970; Gutierrez-Marcos et al., 2012). Human growth diseases, including Beckwith-Wiedemann syndrome, Silver-Russell syndrome, Angleman syndrome, and Prader Willi syndrome are the result of chromosomal abnormalities at imprinted loci in humans (Chiesa et al., 2012; Knoll et al., 1994). In flowering plants, some imprinted genes are involved in regulating seed development (Grossniklaus et al., 1998; Costa et al., 2012) and imprinting is found primarily in endosperm tissue (Gehring et al. 2011; Hsieh et al. 2011; Luo et al. 2011; Waters et al. 2011, 2013; Wolff et al. 2011; Zhang et al. 2011, 2014; Pignatta et al. 2014). Despite this bias toward the endosperm, a handful of putative imprinted genes have been identified in the embryo (Jahnke and Scholten 2009; Autran et al. 2011; Hsieh et al. 2011; Luo et al. 2011; Nodine and Bartel 2012; Raissig et al. 2013; Pignatta et al. 2014). Altering the expression patterns of imprinted alleles can lead to changes in seed size and structure, and greater rates of seed abortion (Costa et al., 2012; Grossniklaus et al., 1998). Imprinting can include both gene specific and allele specific imprinting. Gene-specific imprinting describes instances when all alleles at a locus are imprinted. Allele specific imprinting (ASI) describes instances when only certain alleles are imprinted at a locus (Springer and Gutierrez-Marcos, 2009). The two types of imprinting expression are maternally expressed genes (MEGs) and paternally expressed genes (PEGs), which only express the maternal or paternal allele, respectively.

Initial discovery of imprinted genes

Imprinted genes discovered before genome wide expression analyses were identified due to differential phenotypes between reciprocal hybrids. The *R* gene in maize was the first imprinted gene to be identified and characterized through the use of differential seed pigmentation (Kermicle, 1970). Additional genes were identified through molecular studies comparing phenotypes between reciprocal crosses (Chaudhuri

and Messing 1994; Grossniklaus et al. 1998; Kinoshita et al. 1999, 2004; Luo et al. 2000; Danilevskaya et al. 2003; Gutierrez-Marcos et al. 2004, 2006; Kohler et al. 2005; Barouxet al. 2006; Gehring et al. 2006; Jullien et al. 2006; Haun et al. 2007; Hermon et al. 2007; Makarevich et al. 2008; Tiwari et al. 2008; Fitz Gerald et al. 2009). Imprinted genes MEDEA *(MEA),* FERTILIZATION-INDEPENDENT ENDOSPERM (*FIE*), and FERTILIZATION-INDEPENDENT SEED2 (*FIS2*) were among the initial imprinted genes identified and are well characterized. These genes are MEGs that encode polycomb group proteins (PcG), which putatively repress cell proliferation, prevent replication of the central cell in absence of fertilization, and regulate the expression of downstream genes (Kohler et al., 2005; Ohad et al., 1999; Luo et al., 2000). PcG proteins remodel chromatin state during development and silence genes to maintain cell differentiation (Orlando, 2003). *FIE* and *FIS2* homologs combine to form a complex that methylates histone H3 at lysine 27 (H3K27), which is an essential repressive modifications during development (Kohler and Makarevich, 2006). When individual mutated *mea, fie, and fis2* alleles are inherited maternally, both embryo and endosperm over proliferate, embryos arrest, and eventually abort. The presence of functional paternal alleles does not rescue the embryos, suggesting that maternal expression of *MEA, FIS2,* and *FIE* is essential for seed viability (Grossniklaus, 1998; Ohad et al., 1999; Luo et al., 2000).

Genome wide identification of imprinted genes

Prior to the use of allele specific expression analysis using RNA sequencing, only a handful of imprinted genes had been identified. Studies of *MEA*, *FIE*, and *FIS2* have shown the importance of imprinted expression of these alleles on seed survival. The identification of imprinted genes on a genome wide scale may advance our understanding of the potential roles of imprinting in flowering plants. Genome wide studies have been conducted in maize (Waters et al. 2011, 2013; Zhang et al., 2011, 2014), *Arabidopsis thaliana* (Gehring et al., 2011; Heish et al., 2011; Wolff et al., 2011), and rice (Luo et al., 2011). These studies identified between 100-400 imprinted genes in each species by

assessing the reciprocal hybrid crosses of inbred lines/accessions. The results from these studies highlight some key differences in imprinting between plants and animals. For instance, in mammals imprinting genes are often found in clusters within the genome and under the control of imprinting control regions (Bartolomei et al., 2011). Conversely, little to no evidence is found for clustering of imprinted genes in plants (Waters et al. 2011, 2013; Gehring et al., 2011; Heish et al., 2011; Wolff et al., 2011; Luo et al., 2011). Additionally, imprinting in plants is largely limited to one tissue, the endosperm, whereas imprinting in mammals is not tissue specific (Bartolomei et al., 2011; Pignatta and Gehring 2012). Genome wide identification of imprinted genes in flowering plants also provides a platform for assessing the functional annotation and conservation of imprinting within a species as well as between species.

Preliminary functional annotation analysis of imprinted genes has shown that imprinting may serve roles in maintaining genetic and epigenetic regulation, roles in flower development, and seed traits (Gehring et al., 2011; Waters et al., 2011; Costa et al., 2012; Luo et al., 2011). A few studies have been done to understand the function of particular imprinted genes, which are part of the PRC2 complex and are associated with seed viability and endosperm growth (Grossniklaus et al., 1998; Costa et al., 2012; Kinoshita et al., 1999; Luo et al., 2000). Additional functional analysis is needed to better understand the role of imprinting expression, the function of imprinted genes, and to understand why imprinting would persists at a locus.

Creating comprehensive sets of imprinted genes both within a species and across multiple species of flowering plants allows for the study of conservation of imprinting. An imprinting study in maize assessed the conservation of imprinting within different maize lines and found that a majority of imprinted genes are imprinted in all crosses assessed with limited evidence for allele specific imprinting (Waters et al, 2013). Multiple studies have assessed the conservation of imprinting between maize, rice, and Arabidopsis and found limited conservation of imprinting across these groups (Luo et al., 2011; Waters et al., 2011, 2013) suggesting that either imprinted expression of particular alleles is not functionally necessary or serves different purposes in these species.

Regulation of imprinted genes

Multiple regulatory mechanisms for imprinting expression have been identified. One potential mechanism is that imprinted loci may serve as templates for small interfering RNAs (siRNAs) that keep transposable elements (TEs) silent in the embryo (Lu et al., 2012). siRNAs that correspond to expressed genes in the endosperm are thought to travel to and silence TEs in the embryo (Calarco et al., 2012), which was supported in recent work in rice that shows the importance of siRNAs in imprinting expression (Rodriguez et al., 2013). It is imperative that TEs remain silenced in the embryo as the insertion of a TE into or proximal to a gene can disrupt expression or protein formation (Kidwell and Licsh, 1997). Using this information, one study identified putative imprinted genes by assessing allele specific expression of genes that were near unmethylated TEs in the endosperm in Arabidopsis (Gehring et al., 2009). In addition to silencing transposable elements in the embryo it has also been shown that siRNAs can travel from the vegetative nucleus to the sperm nuclei to ensure silencing of TEs in the nuclei that might fertilize the egg cell (Slotkin et al., 2009). This bias of silencing the paternal copy from the sperm cell and relaxed silencing of the central cell is thought to be another way genes exhibit imprinted expression.

Although the molecular mechanisms of imprinting are not fully understood, DNA methylation and histone modifications are associated with some imprinted genes. Methylation and acetylation of histones H3 and H4 are associated with silenced and activated alleles of some imprinted genes, respectively (Haun and Springer, 2008; Roudier et al., 2011). Additional work looking at H3K27me3 at imprinted loci showed that PEGs are frequently associated with this chromatin modification whereas MEGs rarely contain H3K27me3, which suggests potentially different regulatory mechanisms for PEGs and MEGs (Zhang et al., 2014; Waters et al., 2013). Recent experiments identified differentially methylated regions (DMRs) associated with imprinted genes. All of the DMRs associated with imprinted genes showed a lower level of DNA methylation in the endosperm when compared to leaf or embryo tissue, which is

consistent with previous DNA methylation studies (Gehring et al. 2011; Waters et al., 2011; Zhang et al., 2011; Gehring et al., 2009; Lauria et al., 2004). The reduced level of DNA methylation in the endosperm is hypothesized and shown to be the result of demethylation of the maternal allele (Jullien and Berger, 2009; Raissig et al., 2011; Rodrigues et al., 2013; Zhang et al., 2014). Although DNA methylation and histone modifications are associated with some of variation in expression of parental alleles, additional mechanisms may be important for parent specific expression.

Maintenance of imprinting through evolutionary time

Many ideas have been proposed to explain why imprinting of alleles would persist at a locus. One of the most popular explanations for the maintenance of imprinting is the parental conflict theory. This theory suggests that the maternal and paternal genomes are in conflict over the allocation of resources to the offspring (Haig and Westoby, 1989). The parental conflict theory predicts maternally expressed genes tend to slow or prevent growth of the offspring and would partition nutrients equally among all offspring. In contrast, paternally expressed genes would promote growth allowing offspring a competitive advantage. Some imprinted genes like *MEA, FIE, and FIS2* support this theory, but *MEG1* in maize increases seed size and nutrient transfer, which directly contradicts the parental conflict theory (Costa et al., 2012). In addition, the parental conflict theory does not explain why self-fertilizing species, such as *Arabidopsis thaliana* and rice, where the maternal and paternal investment should be the same, exhibit imprinting, unless imprinting is a remnant from outcrossing ancestors (Grossniklaus et al., 1998; Gehring et al., 2004).

Alternative hypotheses have also been proposed. One of which, suggests that dosage effects could be driving imprinting of some alleles (Dilkes and Comai 2004). The "dosage hypothesis" suggests that silencing of one parental allele may be favored by natural selection if this results in an optimal amount of product produced. This hypothesis can be applied more broadly to a variety of genes and not just those involved in resource allocation. An additional hypothesis suggests selection for imprinting

because it silences potentially deleterious expression of alleles in somatic tissues (Berger et al., 2012). This hypothesis would support imprinted genes that show preferential expression in the endosperm and low or no expression in other tissues, but not for imprinted genes that show similar expression across multiple tissues. None of these hypotheses alone fully explain the maintenance of imprinting.

CASE II: INTRODUCTION TO PLANT ABIOTIC STRESS

In addition to studying parent-of-origin expression, these tools can also be used to differentiate between genotypes. Maize exhibits considerable variation in response to abiotic stress (Makarevitch et al., 2015). Using allele specific expression will provide the platform to describe natural and regulatory variation for abiotic stress response in maize and identify putative candidates for important stress response genes. An improved understanding of the molecular basis for diversity of gene expression responses to abiotic stress can be used to inform strategies for engineering more stress-tolerant genotypes.

Plants, being sessile require a complex set of mechanisms to respond to, and survive in, many different environments. Plants may experience a myriad of abiotic stresses over the course of one life cycle including extreme temperatures, drought, and nutrient deficiencies. These stresses can have negative effects on plant health, survival, and in the case of crops a significant loss of yield. Temperature stress will be the main focus of this review. The phenotypic effects of heat stress could be the result of cell death, increased amounts of reactive oxygen species (ROS), loss of photosynthetic capacity, or loss of turgor pressure in the cell (Ashraf and Hafeex., 2004, Crafts-Brandner and Salvucci., 2002). Plants have evolved complex systems to cope with heat stress including adjusting osmotic pressure, modifying levels of antioxidants to re-establish redox balance, regulating metabolic production and flux, as well as up- regulating important heat shock proteins, enzymes, and transporters (reviewed in Hasanuzzaman et al., 2013). Although cold and heat stress can produce similar external phenotypes, the causes of internal damage and response pathways can vary. One of the major sources of

cell death in cold treated plants is damage to the plasma membrane of the cell and ice formation in the apoplastic space. The fluidity of the plasma membrane will change with the ratio of saturated to unsaturated fatty acids. Cold-resistant plants tend to have a higher proportion of unsaturated fatty acids, which means the temperature at which the membrane changes from semi-fluid to semi-crystalline is lower (Yadav, 2009). Plants have developed intricate physiological, biochemical, and transcriptional processes to cope with exposure to stresses. Many proteins, enzymes, metabolites, and hormones (some described below) have been identified as important factors in cold and heat response.

Phenotypic and tolerance variation for abiotic stress response

Substantial variation can be observed in terms of tolerance and phenotypes between species of plants that are exposed to different abiotic stresses. *Arabidopsis thaliana* is the model organism in which most of the pioneering work for abiotic stress response has been conducted. Other model plants including rice, maize, wheat, soybean, and tomato also have been studied due to their importance in agriculture (Grennan 2006; Komatsu et al., 2014; Mickelbart et al., 2015). Substantial variation for stress tolerance is observed between these species. For instance, *Arabidopsis thaliana*, is generally more tolerant to cold temperatures than *Zea mays*, which is much more tolerant to high heat stress. Phenotypic variation for stress tolerance within a species has also been observed. Differential expression and phenotypic differences were observed between different *Arabidopsis thaliana* accession that were adapted to high or low elevations when exposed to heat stress (Zhang et al., 2016). Different maize inbred lines also exhibit substantial variation in expression as well as tolerance in response to heat and cold stress (Makarevitch et al., 2015).

Abiotic stress can have many negative effects on phenotype. For instance, temperature stress has been shown to negatively affect germination, stunt growth, increase leaf chlorosis and necrosis, and reduce yield (Yadav, 2009; Hasanuzzaman et al., 2013). However, constitutive responses to adverse environments have also been shown

to limit growth and fitness, so a balance between stress-induced expression and maintenance of normal development must be achieved (Gilmour et al., 2000; Dubouzet et al., 2003). Loss of yield is a large driver for studying abiotic stress tolerance in crops. For instance, in the drought of 2012 maize yield decreased by 21%, which drove the prices up by 53% (Boyer et al., 2013). This example is not an outlier for years that have experienced similar levels of drought (Boyer et al., 2013). Understanding abiotic stress tolerance and regulatory variation will become increasingly important for breeding efforts to maintain yield in changing environments.

Wild ancestors of difference crop species are an important source of variation for abiotic stress tolerance. For example, introgression of certain regions of teosinte, the wild ancestor of maize (*Zea mays*) has provided tolerance to certain stresses including flooding tolerance (Micklebart et al., 2015). Additionally, introgression of wild wheat segments into commercial wheat has provided tolerance to high salinity soils, heat stress, and drought stress (Micklebart et al., 2015). This concept has been applied to additional cereals as well as other commercial crops like tomato, soybean, and potato (Micklebart et al., 2015). It is possible that breeding for certain desirable traits in crops also decreased variation for stress tolerance, and providing alternate alleles from increased abiotic stress tolerance. As environments become more variable and extreme due to global climate change in addition to an increasing population, it will become increasingly important to develop novel techniques for increasing and stabilizing yield across environments.

Stress-induced expression as a source of abiotic stress tolerance

Increased expression of certain genes under stress conditions can increase tolerance to cold stress. Work in *Arabidopsis thaliana* identified an integral set of transcription factors called dehydration-responsive element binding factors or C-repeat binding factors (DREB/CBF) within the AP2/ERF family of transcription factors, which increase tolerance under cold, drought, and heat stresses (Jaglo-Ottosen, 1998, Schramm et al., 2007). DREBs have since been identified in other flowering plants including maize, rice, and soybean (Liu et al., 2013, Dubouzet et al., 2003, Chen et al., 2007). Two

different subclasses of DREBs exist, DREB1/CBF, which is induced by cold conditions and DREB2, which has induced transcription when exposed to dehydration and heat conditions (Liu et al., 1998, Sakuma et al., 2006, Schramm et al., 2007). The cold-responsive (COR) group of genes discovered in *Arabidopsis thaliana* are also highly expressed under cold conditions and have been shown to be a crucial component to cold stress response (Gilmour et al., 1988 and Hajela et al., 1990). The COR genes consist of 4 gene families, each of which are comprised of two tandem genes (Thomashow, 1999). In cold stress conditions DREB1/CBF expression is induced, which in turn induces expression of the COR genes and provides cold tolerance (Jaglo-Ottosen, 1998).

Stress-induced expression of additional transcription factor families including myleocytomatosis oncogene (MYC), myeloblastosis oncogene (MYB), WKRY, NAC, and Zinc finger homeodomain (ZF-HD) have been shown to play a role in cold stress response (Abe et al., 1997; Chen et al., 2012; Nakashima et al., 2012). Both cold and heat conditions induce various mitogen-activated protein (MAP) and calcium dependent protein kinases after treatment (Abe et al, 1997; Hasanuzzaman et al., 2013). One important family of transcription factors that increase tolerance to heat stress is the heat shock factors (HSF). HSFs regulate the accumulation of heat shock proteins (Hsps), which are known to be vital for normal plant growth and development, and certain Hsp families have been shown to be necessary for tolerance to heat stress (Shi et al., 1998, Hong and Vierling, 2001; reviewed in Qu et al., 2013).

Regulation of stress-induced gene expression

Stress induced expression of key components in stress response pathways is essential for tolerance to abiotic stress. For example, in addition to many different plant maintenance functions, phytohormone abscisic acid (ABA) increases in abundance under stress conditions and is involved in transcriptional regulation of downstream genes (including MYB and MYC) when plants are exposed to drought, cold, or high salt stresses (Tuteja 2007). At least two different pathways may be involved in stress response: ABA-dependent and ABA-independent, which have both been documented

(Tuteja 2007). Additional phytohormones including cytokinin, ethylene, jasmonic acid, auxin, brassinosteroids, and salicylic acid have also been shown to be important for stress response (O'Brien and Benkova 2013; Peleg and Blumwald 2011). Cross-talk between these hormones as well as interactions with these hormones and expression of important transcription factors showcases the complex nature of regulation of stress response in plants (Munne-Bosch and Muller 2013).

Important binding sites in the regulatory regions of stress-induced genes have also been identified. Multiple studies found cis- regulatory elements 5' of COR genes including the dehydration response element (DRE), which provides a binding site for cold responsive transcription factors (Yamaguchi-Shinozaki and Shinozaki, 1994; Yamaguchi-Shinozaki and Shinozaki, 1993; Hovarth et al., 1993). Cold and drought responsive genes DREB1A and DREB2A also bind to cis- DREs upstream of responsive genes, which results in induced expression of downstream genes. The two DREB alleles increase in expression under different stresses and induce up- regulation of different downstream genes, which suggests that the DREB alleles are functioning and induced independently (Liu 1998). A cis- regulatory element, ABA responsive promoter element (ABRE) allows proteins and binding factors to bind and modulate ABA induced gene expression (reviewed in Raghavenadra et al., 2010; Yamaguchi-Shinozaki and Shinozaki, 2006). Additional cold and drought 5' regions for transcription factors MYB, MYC, NAC and ICE have also been identified (reviewed in Yamaguchi-Shinozaki and Shinozaki, 2006). Heat shock elements (HSE) are part of the promoter sequence of hsps that allow HSFs to bind and activate transcription (Mittal et al., 2011). Recent work in maize also showed that the presence of certain transposable element families was associated with the up- regulation of hundreds of genes in response to abiotic stress conditions (Makarevitch et al., 2015). The conserved regions within the TE fragments that are associated with stress-induced expression are binding regions that provide the sequence necessary for transcription factors to bind and transcribe nearby genes (Makarevtich et al., 2015).

Although examples of both cis- and trans- regulatory elements have been described, not as many examples of exclusive trans- regulatory elements have been described for stress response. General stress response likely includes master trans-regulators, which when activated turn on many genes, but identifying these master regulators is not trivial. Certain TFs and photyohormones are important stress response trans- acting elements that activate many genes across the genome (Yamaguchi-Shinozaki and Shinozaki, 2006; Hasanuzzaman et al., 2013). A recent study in *Arabidopsis thaliana* used differential and allele specific expression to identify examples of cis- and trans- regulatory variation patterns across the genome. This study identified a bias toward cis- regulatory variation when contrasting Arabidopsis accessions as well as under control and drought conditions, but also found examples of trans- regulatory variation expression patterns (Cubillos et al., 2014).

Understanding transcriptome and regulatory variation for abiotic stress response has additional applications beyond understanding stress tolerance; including aiding breeders in decisions about lines to grow in different environments as well as understanding local adaptation in natural populations. Pioneering molecular work has shed light on the roles of certain transcription factors and genes as well as cis- regulatory elements associated with abiotic stress response. Genome wide analysis of abiotic stress response can provide the framework for studying and capitalizing on natural and regulatory variation as well as identify additional important genes and regulatory regions/elements.

CONCLUSIONS

Variation in sequence, epigenetic profile, and expression is the basis of phenotypic diversity. Understanding how alleles are differentially expressed on a genomic scale can aide in our understanding of sequence and epigenetic variation in terms of regulatory variation and identify important genes that are involved in different processes.

The utilization of next generation sequencing has provided the resolution necessary to discriminate alleles and begin to understand differential expression and regulatory variation between alleles. Understanding these processes in maize, an agronomically important crop, could provide novel insights into crops and potentially be applied to understanding yield traits. Maize (Zea mays) is a model system to study differential expression and regulatory variation. Not only does maize have many genetic and genomic resources, including a reference genome and resequencing of many additional lines, but maize exhibits substantial variation between inbred lines.

This thesis outlines the experiments and methods used to study allele specific expression and regulatory variation for genomic imprinting and abiotic stress response in maize as a way to study complex regulatory systems. Chapters II and III describe efforts to understand the prevalence and conservation of imprinting within maize and between maize and rice or *Arabidopsis thaliana*. The third chapter assesses the level of regulatory variation between maize inbred lines in response to cold or heat stress. An improved understanding of the allelic diversity as well as regulatory variation will not only provide valuable information for genomic imprinting and abiotic stress response, but build the framework for applications to many other systems and questions.

REFERENCES

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., & Shinozaki, K. 1997. Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. The Plant Cell, 9(10), 1859–68. http://doi.org/10.1105/tpc.9.10.1859

Albert, F. W., & Kruglyak, L. 2015. The role of regulatory variation in complex traits and disease. Nature Reviews Genetics, 16(4), 197–212. http://doi.org/10.1038/nrg3891

Ashraf, M., & Hafeez, M. 2004. Thermotolerance of pearl millet and maize at early growth stages : growth and nutrient relations (Vol. 48, pp. 81–86.

Autran, D., Baroux, C., Raissig, M. T., Lenormand, T., Wittig, M., Grob, S., et al Grossniklaus, U. 2011. Maternal epigenetic pathways control parental contributions to Arabidopsis early embryogenesis. Cell, 145(5), 707–19. http://doi.org/10.1016/j.cell.2011.04.014

Baroux, C., Gagliardini, V., Page, D. R., & Grossniklaus, U. 2006. Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis. Genes & Development, 20(9), 1081–6. http://doi.org/10.1101/gad.378106

Bartolomei, M. S., & Ferguson-Smith, A. C. 2011. Mammalian genomic imprinting. Cold Spring Harbor Perspectives in Biology, 3(7), a002592–. http://doi.org/10.1101/cshperspect.a002592

Berger, F., Vu, T. M., Li, J., & Chen, B. 2012. Hypothesis: selection of imprinted genes is driven by silencing deleterious gene activity in somatic tissues. Cold Spring Harbor Symposia on Quantitative Biology, 77(0), 23–9. http://doi.org/10.1101/sqb.2012.77.014514

Boyer, J. S., Byrne, P., Cassman, K. G., Cooper, M., Delmer, D., Greene, T., et al Gaffney, J. 2013. The U.S. drought of 2012 in perspective: A call to action. Global Food Security, 2(3), 139–143. http://doi.org/10.1016/j.gfs.2013.08.002

Bushell C., Spielman M., and Scott R.J. 2003. The basis of natural and artificial postzygotic hybridization barriers in Arabidopsis species. The Plant Cell. 15:1430-1442.

Calarco J.P., Borges F, Donoghue M.T.A., Van Ex F., Jullien P.E., Lopes T., Gardner R., Berger F., Feijó J.A., Becker J.D., Martienssen R.A. 2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell. 151(1): 194-205.

Carroll, S. B. 2005. Evolution at two levels: on genes and form. PLoS Biology, 3(7), e245. http://doi.org/10.1371/journal.pbio.0030245

Chaudhuri, S., & Messing, J. 1994. Allele-specific parental imprinting of dzr1, a posttranscriptional regulator of zein accumulation. Proceedings of the National Academy of Sciences of the United States of America, 91(11), 4867–71. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=43890&tool=pmcentrez&rendertype=abstract

Chen, L., Song, Y., Li, S., Zhang, L., Zou, C., & Yu, D. 2012. The role of WRKY transcription factors in plant abiotic stresses. Biochimica et Biophysica Acta, 1819(2), 120–8. http://doi.org/10.1016/j.bbagrm.2011.09.002

Chen, M., Wang, Q.-Y., Cheng, X.-G., Xu, Z.-S., Li, L.-C., Ye, X.-G., et al Ma, Y.-Z. 2007. GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. Biochemical and Biophysical Research Communications, 353(2), 299–305. http://doi.org/10.1016/j.bbrc.2006.12.027

Chiesa N., De Crescenzo A., Mishra K., Perone L., Carella M., Palumbo O., Mussa A., Sparago A., Cerrato F., Russo S., Lapi E., Cubellis M.V., Kanduri C., Silengo M.C., Riccio A., and Ferrero G.B. 2012. The KCNQ10T1 imprinting control region and non-coding RNA: new properties derived from the study of Beckwidth-Wiedemann syndrome and Silver-Russell syndrome cases. Human Molecular Genetics 21:10-25.

Costa L.M., Yuan J., Rouster W.P., Dickinson H., and Gutierrez-Marcos J.F., Maternal control of nutrient allocation in plant seeds by genomic imprinting. Current Biology 22:160-165.

Crafts-Brandner, S. J., & Salvucci, M. E. 2002. Sensitivity of photosynthesis in a C4 plant, maize, to heat stress. Plant Physiology, 129(4), 1773–80. http://doi.org/10.1104/pp.002170

Cubillos, F. A., Stegle, O., Grondin, C., Canut, M., Tisné, S., Gy, I., & Loudet, O. 2014. Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis. The Plant Cell, 26(11), 4298–310. http://doi.org/10.1105/tpc.114.130310

Danilevskaya O.N., Hermon P., Hantke S., Muszynski M.G., Kollipara K., and Ananiev E.V. 2003. Duplicated fie genes in maize: expression pattern and imprinting suggest distinct functions. Plant Cell 15: 425-438.

Deng X, Disteche CM (2010) Genomic Responses to Abnormal Gene Dosage: The X Chromosome Improved on a Common Strategy. PLoS Biol 8(2.

Dilkes, B. P., & Comai, L. 2004. A differential dosage hypothesis for parental effects in seed development. The Plant Cell, 16(12), 3174–80. http://doi.org/10.1105/tpc.104.161230

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al Yamaguchi-Shinozaki, K. 2003. OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. The Plant Journal, 33(4), 751–763. http://doi.org/10.1046/j.1365-313X.2003.01661.x

Eichten, S. R., Ellis, N. A., Makarevitch, I., Yeh, C.-T., Gent, J. I., Guo, L., et al Springer, N. M. 2012. Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS Genetics, 8(12), e1003127. http://doi.org/10.1371/journal.pgen.1003127

Fitz Gerald, J. N., Hui, P. S., & Berger, F. 2009. Polycomb group-dependent imprinting of the actin regulator AtFH5 regulates morphogenesis in Arabidopsis thaliana. Development (Cambridge, England), 136(20), 3399–404. http://doi.org/10.1242/dev.036921

Gehring M., Missirian V., and Henikoff S. 2011. Genomic Analysis of Parent-of-Origin Allelic Expression in Arabidopsis thaliana Seeds. PLoS One 6: e23687.

Gehring M., Bubb K.L., and Henikoff S. 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324: 1447-1451.

Gehring, M., Huh, J. H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J. J., et al Fischer, R. L. 2006. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell, 124(3), 495–506. http://doi.org/10.1016/j.cell.2005.12.034

Gehring M., Choi Y., and Fischer R.L. 2004. Imprinting and seed development. Plant Cell 16 Suppl: S203-13.

Gilmour, S. J. 2000. Overexpression of the Arabidopsis CBF3 Transcriptional Activator Mimics Multiple Biochemical Changes Associated with Cold Acclimation. PLANT PHYSIOLOGY, 124(4), 1854–1865. http://doi.org/10.1104/pp.124.4.1854

Gilmour, S. J., Hajela, R. K., & Thomashow, M. F. 1988. Cold Acclimation in Arabidopsis thaliana. PLANT PHYSIOLOGY, 87(3), 745–750. http://doi.org/10.1104/pp.87.3.745

Grennan, A. K. 2006. Abiotic stress in rice. An “omic” approach. Plant Physiology, 140(4), 1139–41. http://doi.org/10.1104/pp.104.900188

Grossniklaus U., Vielle-Calzada J.P., Hoeppner M.A., and Gagliano W.B. 1998. Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science 280: 446-450.

Gutierrez-Marcos, J. F., Constância, M., & Burton, G. J. 2012. Maternal to offspring resource allocation in plants and mammals. Placenta, 33 Suppl 2, e3–10. http://doi.org/10.1016/j.placenta.2012.08.006

Gutiérrez-Marcos, J. F., Costa, L. M., Biderre-Petit, C., Khbaya, B., O’Sullivan, D. M., Wormald, M., et al Dickinson, H. G. 2004. maternally expressed gene1 Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. The Plant Cell, 16(5), 1288–301. http://doi.org/10.1105/tpc.019778

Gutiérrez-Marcos, J. F., Costa, L. M., & Evans, M. M. S. 2006. Maternal gametophytic baseless1 is required for development of the central cell and early endosperm patterning in maize (Zea mays. Genetics, 174(1), 317–29. http://doi.org/10.1534/genetics.106.059709

Hajela, R. K., Horvath, D. P., Gilmour, S. J., & Thomashow, M. F. 1990. Molecular Cloning and Expression of cor (Cold-Regulated) Genes in Arabidopsis thaliana. Plant Physiology, 93(3), 1246–52. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1062659&tool=pmcentrez&rendertype=abstract

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., & Fujita, M. 2013. Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. International Journal of Molecular Sciences, 14(5), 9643–84. http://doi.org/10.3390/ijms14059643

Haun W.J. and Springer N.M. 2008. Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. Plant J. 56: 903-912.

Haun W.J., Laoueille-Duprat S., O'connell M.J., Spillane C., Grossniklaus U., Phillips A.R., Kaeppler S.M., and Springer N.M. 2007. Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste (Mez) homologs. Plant J. 49: 325-337.

Hermon, P., Srilunchang, K., Zou, J., Dresselhaus, T., & Danilevskaya, O. N. 2007. Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of the maternal allele. Plant Molecular Biology, 64(4), 387–95. http://doi.org/10.1007/s11103-007-9160-0

Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., et al Buell, C. R. 2014. Insights into the maize pan-genome and pan-transcriptome. The Plant Cell, 26(1), 121–35. http://doi.org/10.1105/tpc.113.119982

Hong, S.-W., & Vierling, E. 2001. Hsp101 is necessary for heat tolerance but dispensable for development and germination in the absence of stress. The Plant Journal, 27(1), 25–35. http://doi.org/10.1046/j.1365-313x.2001.01066.x

Horvath, D. 1993. Regulation of Arabidopsis thaliana L. Heyn) cor78 in response to low temperature. PLANT PHYSIOLOGY, 103(4), 1047–1053. http://doi.org/10.1104/pp.103.4.1047

Hsieh T.F., Shin J., Uzawa R., Silva P., Cohen S., Bauer M.J., Hashimoto M., Kirkbride R.C., Harada J.J., Zilberman D., and Fischer R.L. 2011. Regulation of imprinted gene expression in Arabidopsis endosperm. Proc. Natl. Acad. Sci. U. S. A. 108: 1755-1762.
Huh J.H., Bauer M.J., Hsieh T.F., and Fischer R.L. 2008. Cellular programming of plant gene imprinting. Cell 132**:** 735-744.

Jaglo-Ottosen, K. R. 1998. Arabidopsis CBF1 Overexpression Induces COR Genes and Enhances Freezing Tolerance. Science, 280(5360), 104–106. http://doi.org/10.1126/science.280.5360.104

Jahnke, S., & Scholten, S. 2009. Epigenetic resetting of a gene imprinted in plant embryos. Current Biology : CB, 19(19), 1677–81. http://doi.org/10.1016/j.cub.2009.08.053

Johannes, F., Colot, V., & Jansen, R. C. 2008. Epigenome dynamics: a quantitative genetics perspective. Nature Reviews. Genetics, 9(11), 883–90. http://doi.org/10.1038/nrg2467

Jullien PE, Berger F (2010) Parental Genome Dosage Imbalance Deregulates Imprinting in Arabidopsis. PLoS Genet 6(3) : e1000885

Jullien P.E. and Berger F. 2009. Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr. Opin. Plant Biol. 12: 637-642.

Jullien P.E., Katz A., Oliva M., Ohad N., and Berger F. 2006. Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis. Curr. Biol. 16: 486-492.

Kermicle J.L. 1970. Dependence of the R-Mottled Aleurone Phenotype in Maize on Mode of Sexual Transmission. Genetics 66: 69-85.

Kidwell, M.G. & D. Lisch, 1997. Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. USA 94: 7704-7711.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., et al Kakutani, T. 2004. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science (New York, N.Y.), 303(5657), 521–3.

Kinoshita T., Yadegari R., Harada J.J., Goldberg R.B., and Fischer R.L. 1999. Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. Plant Cell 11: 1945-1952.

Knoll J.H., Cheng S.D., LALANDE M. 1994) Allele specificity of DNA replication timing in the Angelman/ Prader-Willi syndrome imprinted chromosomal region. Nat Genet. 6:41-6

Kohler C., and Makarevitch G. 2006. Epigenetic mechanisms governing seed development in plants. EMBO Reports. 7(12): 1223-1227.

Kohler C., Page D.R., Gagliardini V., and Grossniklaus U. 2005. The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nature Genetics. 37(1): 28-30.

Komatsu, S., Kamal, A. H. M., & Hossain, Z. 2014. Wheat proteomics: proteome modulation and abiotic stress acclimation. Frontiers in Plant Science, 5, 684. http://doi.org/10.3389/fpls.2014.00684

Lauria M., Rupe M., Guo M., Kranz E., Pirona R., Viotti A., and Lund G. 2004. Extensive maternal DNA hypomethylation in the endosperm of Zea mays. Plant Cell 16: 510-522.

Liu, Q. 1998. Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low-Temperature-Responsive Gene Expression, Respectively, in Arabidopsis. THE PLANT CELL ONLINE, 10(8), 1391–1406. http://doi.org/10.1105/tpc.10.8.1391

Liu, S., Wang, X., Wang, H., Xin, H., Yang, X., Yan, J., et al Qin, F. 2013. Genome-Wide Analysis of ZmDREB Genes and Their Association with Natural Variation in Drought Tolerance at Seedling Stage of Zea mays L. PLoS Genetics, 9(9), e1003790. http://doi.org/10.1371/journal.pgen.1003790

Lu J., Zhang C., Baulcombe D. C., and Chen J. 2012. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. PNAS. 109(14): 5529-5534.

Luo M., Bilodeau P., Dennis E.S., Peacock W.J., and Chaudhury A. 2000. Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. Proc. Natl. Acad. Sci. U. S. A. 97: 10637-10642.

Makarevich, G., Villar, C. B. R., Erilova, A., & Köhler, C. 2008. Mechanism of PHERES1 imprinting in Arabidopsis. Journal of Cell Science, 121(Pt 6), 906–12. http://doi.org/10.1242/jcs.023077

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., & Springer, N. M. 2015. Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genetics, 11(1), e1004915. http://doi.org/10.1371/journal.pgen.1004915

Mickelbart, M. V, Hasegawa, P. M., & Bailey-Serres, J. 2015. Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nature Reviews. Genetics, 16(4), 237–51. http://doi.org/10.1038/nrg3901

Mittal, D., Enoki, Y., Lavania, D., Singh, A., Sakurai, H., & Grover, A. 2011. Binding affinities and interactions among different heat shock element types and heat shock factors in rice (Oryza sativa L.. The FEBS Journal, 278(17), 3076–85. http://doi.org/10.1111/j.1742-4658.2011.08229.x

Munné-Bosch, S., & Müller, M. 2013. Hormonal cross-talk in plant development and stress responses. Frontiers in Plant Science, 4, 529. http://doi.org/10.3389/fpls.2013.00529

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. 2012. NAC transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta, 1819(2), 97–103. http://doi.org/10.1016/j.bbagrm.2011.10.005

Nodine, M. D., & Bartel, D. P. 2012. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature, 482(7383), 94–7. http://doi.org/10.1038/nature10756

O'Brien, J. A., & Benková, E. 2013. Cytokinin cross-talking during biotic and abiotic stress responses. Frontiers in Plant Science, 4, 451. http://doi.org/10.3389/fpls.2013.00451

Ohad N., Yadegari R., Margossian L., Hannon M., Michaeli D., Harada J.J., Goldberg R.B., and Fischer R.L. 1999. Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell 11: 407-416.

Orlando V. 2003. Polycomb, epigenomes, and control of cell identity. Cell 112: 599-606.

Peleg, Z., & Blumwald, E. 2011. Hormone balance and abiotic stress tolerance in crop plants. Current Opinion in Plant Biology, 14(3), 290–295. http://doi.org/10.1016/j.pbi.2011.02.001

Pignatta, D., Erdmann, R. M., Scheer, E., Picard, C. L., Bell, G. W., & Gehring, M. 2014. Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. eLife, 3, e03198. http://doi.org/10.7554/eLife.03198

Pignatta D., Gehring M. 2012) Imprinting meets genomics: new insights and new challenges. Curr Opin. In Plant Bio. 15(5): 530-535

Qu, A.-L., Ding, Y.-F., Jiang, Q., & Zhu, C. 2013. Molecular mechanisms of the plant heat stress response. *Biochemical and Biophysical Research Communications*, *432*(2), 203–7. http://doi.org/10.1016/j.bbrc.2013.01.104

Raghavendra, A. S., Gonugunta, V. K., Christmann, A., & Grill, E. 2010. ABA perception and signalling. Trends in Plant Science, 15(7), 395–401. http://doi.org/10.1016/j.tplants.2010.04.006

Raissig, M. T., Bemer, M., Baroux, C., & Grossniklaus, U. 2013. Genomic imprinting in the Arabidopsis embryo is partly regulated by PRC2. PLoS Genetics, 9(12), e1003862. http://doi.org/10.1371/journal.pgen.1003862

Richards, E. J. 2006. Inherited epigenetic variation--revisiting soft inheritance. Nature Reviews. Genetics, 7(5), 395–401. http://doi.org/10.1038/nrg1834

Rodrigues, J. A., Ruan, R., Nishimura, T., Sharma, M. K., Sharma, R., Ronald, P. C., et al Zilberman, D. 2013. Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proceedings of the National Academy of Sciences of the United States of America, 110(19), 7934–9. http://doi.org/10.1073/pnas.1306164110

Roudier F., Ahmed I., Berard C., Sarazin A., Mary-Huard T.,Cortijo S., Bouyer D., Caillieux E., Duvernois-Berthet E., Al-Shikhley L., Giraut L., Despres B., Drevensek S., Barneche F., Derozier S., Brunaud V., Aubourg S., Schnittger A., Bowler C., Martin-Magniette M-L., Robin S., Caboche M., and Colot V. 2011. Integrative epigenomic mapping defines four main chromatin states in Arbidopsis. EMBO J. 30(10): 1928-1938

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., & Yamaguchi-Shinozaki, K. 2006. Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. Proceedings of the National Academy of Sciences of the United States of America, 103(49), 18822–7. http://doi.org/10.1073/pnas.0605639103

Schramm, F., Larkindale, J., Kiehlmann, E., Ganguli, A., Englich, G., Vierling, E., & von Koskull-Döring, P. 2008. A cascade of transcription factor DREB2A and heat stress

transcription factor HsfA3 regulates the heat stress response of Arabidopsis. The Plant Journal : For Cell and Molecular Biology, 53(2), 264–74. http://doi.org/10.1111/j.1365-313X.2007.03334.x

Shi, Y., Mosser, D. D., & Morimoto, R. I. 1998. Molecularchaperones as HSF1-specific transcriptional repressors. Genes & Development, 12(5), 654–666. http://doi.org/10.1101/gad.12.5.654

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J. D., Feijó, J. A., & Martienssen, R. A. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell, 136(3), 461–72. http://doi.org/10.1016/j.cell.2008.12.038

Springer, N. M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., et al Schnable, P. S. 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genetics, 5(11), e1000734. http://doi.org/10.1371/journal.pgen.1000734

Springer N.M., and Gutierrez-Marcos F. 2009. Imprinting in Maize. Maize Handbook-Genetics and Genomics. 2:429-440.

Stelpflug, S. C., Rajandeep, S., Vaillancourt, B., Hirsch, C. N., Buell, C. R., Leon, N. De, & Kaeppler, S. M. 2015. An expanded maize gene expression atlas based on RNA-sequencing and its use to explore root development. *The Plant Genome*, (608), 314–362. http://doi.org/10.3835/plantgenome2015.04.0025

Swanson-Wagner, R. A., Eichten, S. R., Kumari, S., Tiffin, P., Stein, J. C., Ware, D., & Springer, N. M. 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Research, 20(12), 1689–99. http://doi.org/10.1101/gr.109165.110

Thomashow, M. F. 1999. PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology, 50, 571–599. http://doi.org/10.1146/annurev.arplant.50.1.571

Tiwari, S., Schulz, R., Ikeda, Y., Dytham, L., Bravo, J., Mathers, L., et al Scott, R. J. 2008. MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in Arabidopsis, encodes the conserved C-terminal domain of polyadenylate binding proteins. The Plant Cell, 20(9), 2387–98. http://doi.org/10.1105/tpc.108.061929

Tuteja, N. 2007. Abscisic Acid and abiotic stress signaling. Plant Signaling & Behavior, 2(3), 135–8.

van Oeveren, J., & Janssen, A. 2009. Single Nucleotide Polymorphisms. In Single Nucleotide Polymorphisms (Vol. 578, pp. 73–91. http://doi.org/10.1007/978-1-60327-411-1

Vaughn, M. W., Tanurdzić, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P. D., et al Martienssen, R. A. 2007. Epigenetic natural variation in Arabidopsis thaliana. PLoS Biology, 5(7), e174. http://doi.org/10.1371/journal.pbio.0050174

Waters, A. J., Bilinski, P., Eichten, S. R., Vaughn, M. W., Ross-Ibarra, J., Gehring, M., & Springer, N. M. 2013. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proceedings of the National Academy of Sciences of the United States of America, 110(48), 19639–44. http://doi.org/10.1073/pnas.1309182110

Waters A.J., Makarevitch I., Eichten S.R., Swanson-Wagner R.A., Yeh C-T, Xu W., Schnable P.S., Vaughn M.W., Gehring M., and Springer N.M. 2011. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell. 23(12):4221-4233.

Wolff P., Weinhofer I., Seguin J., Roszak P., Beisel C., Donoghue M.T., Spillane C., Nordborg M., Rehmsmeier M., and Kohler C. 2011. High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. PLoS Genet. 7: e1002126.

Wray, G. A. 2007. The evolutionary significance of cis-regulatory mutations. Nature Reviews. Genetics, 8(3), 206–16. http://doi.org/10.1038/nrg2063

Yadav, S. K. 2010. Cold stress tolerance mechanisms in plants. A review. Agronomy for Sustainable Development, 30(3), 515–527. http://doi.org/10.1051/agro/2009050

Yamaguchi-Shinozaki, K., & Shinozaki, K. 1993. Characterization of the expression of a desiccation-responsive rd29 gene of Arabidopsis thaliana and analysis of its promoter in transgenic plants. Molecular & General Genetics : MGG, 236(2-3), 331–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8437577

Yamaguchi-Shinozaki, K., & Shinozaki, K. 1994. A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. The Plant Cell, 6(2), 251–64. http://doi.org/10.1105/tpc.6.2.251

Yamaguchi-Shinozaki, K., & Shinozaki, K. 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annual Review of Plant Biology, 57, 781–803. http://doi.org/10.1146/annurev.arplant.57.032905.105444

Zhang, M., Xie, S., Dong, X., Zhao, X., Zeng, B., Chen, J., et al Lai, J. 2014. Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. Genome Research, 24(1), 167–76. http://doi.org/10.1101/gr.155879.113

Zhang M., Zhao H., Xie S., Chen J., Xu Y., Wang K., Zhoa H., Guan H., Hu X., Jiao Y., Song W., and Lai J. 2011) Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. PNAS. 108(50): 20042-20047.

Zhang, N., Vierling, E., & Tonsor, S. 2016. Adaptive divergence in transcriptome response to heat and acclimation in Arabidopsis thaliana plants from contrasting climates. bioRxiv. Cold Spring Harbor Labs Journals.

Chapter II: **Context Statement**

SUMMARY

Imprinting describes the differential expression of alleles based upon their parent of origin. Deep sequencing of RNAs from maize endosperm and embryo tissue 14 days after pollination was used to identify imprinted genes among a set of ~12,000 genes that were expressed and contained sequence polymorphisms between the B73 and Mo17 genotypes. The analysis of parent-of-origin patterns of expression resulted in the identification of 100 putative imprinted genes in maize endosperm including 54 maternally expressed genes (MEGs) and 46 paternally expressed genes (PEGs). Three of these genes have been previously identified as imprinted while the remaining 97 genes represent novel imprinted maize genes. A genome-wide analysis of DNA methylation identified regions with reduced endosperm DNA methylation in, or near, 19 of the 100 imprinted genes. The reduced levels of DNA methylation in endosperm are caused by hypomethylation of the maternal allele for both MEGs and PEGs in all cases tested. Many of the imprinted genes with reduced DNA methylation levels also show endosperm-specific expression patterns. The imprinted maize genes were compared with imprinted genes identified in genome-wide screens of rice and Arabidopsis and at least 10 examples of conserved imprinting between maize and each of the other species were identified.

Chapter II entitled ‘Imprinting in maize endosperm’ has been adapted from work in the publication:

Amanda J. Waters*, Irina Makarevitch*, Steve R. Eichten*, Ruth A. Swanson-Wagner, Mary Gehring, Cheng-Ting Yeh, Patrick S. Schnable, Matthew W. Vaughn, Nathan M. Springer (2011) Parent-of-Origin Effects on Gene Expression and DNA Methylation in the Maize Endosperm. The Plant Cell. 23(12): 4221-4233 *These authors contributed equally to this publication.

During the course of this work many authors contributed to this work. Amanda J. Waters, Irina Makarevitch, Steve R. Eichten, Ruth A. Swanson-Wagner, and Cheng-Ting Yeh performed research. In particular, I helped collect and process samples and data. Cheng-Ting Yeh, Patrick S. Schnable, Matthew W. Vaughn contributed analytic/ computational tools. Amanda J. Waters, Irina Makarevitch, Steve R. Eichten, Mary Gehring, Cheng-Ting Yeh, Patrick S. Schnable, Matthew W. Vaughn, Nathan M. Springer helped analyzed data. I also aided in validating the imprinting expression of genes through multiple methods. Figures and text were contributed by Amanda J. Waters, Irina Makarevitch, and Nathan M. Springer. I have removed author contact information and acknowledgments as well as formatted references to be consistent throughout my thesis.

CHATPER II

Imprinting in the maize endosperm

INTRODUCTION

During sexual reproduction both parents contribute one haploid genome to their offspring and for the majority of genes the alleles inherited from the mother and father exhibit similar expression levels. However, a small proportion of genes exhibit parent-of-origin differences in the expression of the maternal and paternal alleles, which is termed imprinting (Huh et al., 2008). Imprinting reflects epigenetic control of two identical, or nearly identical, alleles that are present in the same nucleus. In plants, most examples of imprinting have been observed in the endosperm tissue and not in vegetative tissues (reviewed by Gehring et al., 2004; Huh et al., 2008; Jullien and Berger 2009; Springer and Gutierrez-Marcos 2009) although there is evidence that imprinting may also occur in embryos (Jahnke and Scholten 2009). The endosperm is a triploid (2 maternal:1 paternal genomes) product of the double fertilization event in plant reproduction and there is evidence that some imprinted genes play a role in regulation of endosperm development (Grossniklaus et al., 1998; Kinoshita et al., 1999; Luo et al., 2000; Kohler et al., 2003; Jullien et al., 2006). The kinship theory (Haig and Westoby 1989) has been proposed as a model to explain why imprinting may provide adaptive value for parental control over allocation of resources to offspring but does not provide a mechanism for imprinted gene expression (Haig and Westoby 1989). The kinship theory proposes that imprinting would be maintained when a gene's expression level in one individual affects the fitness of other individuals who have different probabilities for carrying the maternal and paternal alleles (Haig 2004).

Several lines of evidence suggest that DNA methylation and chromatin changes are important components of the imprinting mechanism. Genome-wide hypomethylation of the maternal genome in endosperm tissue has been observed in multiple plant species (Lauria et al., 2004; Gehring et al., 2009; Hsieh et al., 2009; Zemach et al., 2010). Endosperm hypomethylation in Arabidopsis is at least partially an active process

controlled by DME, a 5-methylcytosine DNA glycosylase that is expressed in the central cell before fertilization (Gehring et al., 2006). Many regions of endosperm hypomethylation correspond to fragments of transposable elements (Gehring et al., 2009; Zemach et al., 2010). In addition to being important for establishing imprinted expression, endosperm hypomethylation has been proposed to provide a mechanism to reinforce silencing of transposable elements in the adjacent embryo tissue (Mosher and Melnyk, 2010; Springer, 2009; Kohler and Weinhofer-Molisch 2010; Bauer and Fischer 2011). In addition to the potential involvement of DNA methylation there is also evidence that chromatin changes mediated by Polycomb group proteins contribute to imprinted regulation (Kohler and Weinhofer-Molisch 2010; Raissig et al., 2011). At several imprinted maize genes there is evidence for both DNA methylation and chromatin modification differences between the maternal and paternal alleles (Haun and Springer 2008).

A handful of imprinted genes were discovered based on parent-of-origin specific phenotypes associated with mutant alleles (reviewed by Gehring et al., 2004; Huh et al., 2008; Jullien and Berger 2009; Springer and Gutierrez-Marcos 2009; Raissig et al., 2011). Several groups have used modified differential display or SNP-based assays to screen allelic expression to identify additional imprinted genes (Guo et al, 2003; Gutierrez-Marcos et al, 2004; Stupar et al, 2007; McKeown et al., 2011). Gehring et al. 2009) profiled DNA methylation patterns in endosperm tissue and identified numerous regions of endosperm-specific hypomethylation. A number of additional candidate imprinted genes were identified by finding examples of genes with endosperm-specific expression located near regions of endosperm hypomethylation (Gehring et al., 2009). The development of RNA-seq technologies has enabled several recent-genome wide scans of imprinting in Arabidopsis (Hsieh et al 2011; Wolff et al., 2011; Gehring et al., 2011) and rice (Luo et al., 2011). These "genome-wide" scans can survey allelic expression patterns in the roughly 10,000 genes that contain polymorphisms between the two parental alleles and are expressed in endosperm. In each case there have been roughly 100 maternally expressed genes (MEGs) or paternally expressed genes (PEGs)

identified. The exact tissue, parental genotypes and informatics-filters used appear to substantially affect the identification of imprinted genes as the genes identified in the three Arabidopsis surveys differ substantially (Hsieh et al 2011; Wolff et al., 2011; Gehring et al., 2011). Generally, there is evidence for imprinting at a number of genes with putative regulatory or signaling roles.

Maize provides an excellent system to study parent-of-origin effects. The maize endosperm is relatively large and persists throughout seed development. In addition, the maize genome has a complex organization of interspersed genes and transposable elements. Given the hypothesis that imprinting may frequently arise via influences of nearby transposable elements (Kohler and Weinhofer-Molisch 2010; Mosher and Melnyk, 2010; Bauer and Fischer, 2011) there may be numerous imprinted genes in the maize endosperm. In this study very deep sequencing of transcripts in maize embryo and endosperm tissue was used to identify genes with parent-of-origin effects. A set of 100 putatively imprinted genes was identified. The identification of a large set of imprinted maize genes provides opportunities to further understand the mechanisms and role of imprinting in plants and to further understand seed development.

RESULTS

High-throughput sequencing of RNA provides an opportunity to study both gene expression and allele-specific expression levels in hybrid samples. We collected three biological replicates of B73 and Mo17 endosperm tissue 14 days after pollination (DAP), sequencing a relatively small number of single end 76 bp reads (~5-9 million per sample) (Figure 1). In addition, we pooled tissue from three biological replicates of B73 x Mo17 and Mo17 x B73 reciprocal hybrids. We performed deep sequencing of hybrid tissues from pooled replicates instead of from biological replicates because the relevant comparison is between alleles and not between samples. Therefore, our statistics are applied to the sampling of reads for two alleles in the same sample and not to reads between samples. Substantially more sequence reads were recovered from the hybrid samples (Figure 1). In total we recovered ~28 billion base pairs from >270 million reads

of RNA-seq data from 14 DAP endosperm tissue (Table 1). The sequence reads were mapped to the 39,656 filtered gene models derived from the AGPv2 reference genome (AGPv2_FGSv5b) (Figure 1). These alignments were used to develop the estimates of relative gene expression levels based on the reads per kilobase per million reads (RPKM) using Tophat – Cufflinks (Trapnell et al., 2010). Slightly over half (21,910 out of 39656) of the genes have a RPKM value over 1 in both reciprocal hybrids and were retained as "expressed" genes for further analyses. We also gathered parental and reciprocal hybrid RNA-seq data from 14DAP embryo tissue. The analysis of genes with endosperm-specific and embryo-specific gene expression patterns revealed that we could find examples of genes with moderate to high levels of expression in embryo, but without transcripts in endosperm. However, we did not find genes with endosperm expression but no transcripts in embryo (Figure 2).

Identification of imprinted maize genes:

Allele-specific gene expression was determined by mapping the sequence reads to a set of 1.55 million B73-Mo17 SNPs. Over 70% (28,104 / 39,656) of the Zm5a.60 FGS sequences have at least one B73-Mo17 SNP and 58% (22,916) genes have at least three SNPs. The number of reads that correspond to the B73 and Mo17 alleles were determined for each SNP position and were summed across the gene to determine the number of maternal and paternal reads per gene. In endosperm there are 12,571 genes for which there are at least 10 reads that could be assigned to a specific allele in both reciprocal hybrids (Figure 1). This analysis enabled us to assess allele-specific expression for 57% of the genes that are expressed in endosperm tissue. Similar criteria identified 11,606 genes in embryo tissue with robust allele-specific expression data.

The relative allele-specific expression levels from the reciprocal hybrids can be used to assess the impact of allelic variation (B73 vs. Mo17) (Figure 3) and parent-of-origin effects (maternal vs. paternal) (Figure 4A). If the two alleles are equivalently expressed in endosperm tissue we would expect to observe 66% of the transcripts derived from the B73 allele in the B73 x Mo17 endosperm tissue and 33% of the transcripts

derived from the B73 allele in Mo17 x B73 endosperm tissue. In cases where one allele (B73 or Mo17) is preferentially expressed, presumably due to *cis*-regulatory variation, it is expected that expression would be biased towards the same allele in both reciprocal hybrids and will follow a curve towards complete B73 or Mo17 allele expression. In contrast, examples of parent-of-origin effects would be expected to exhibit a bias towards one allele in one hybrid and bias towards the other allele in the reciprocal hybrid. The observed data (Figure 4A) show that most genes are distributed along the curve indicating common allelic variation for gene expression. A contrast of the allelic expression bias in hybrid samples with differential expression in the parental genotypes (Figure 3) provides further evidence for common *cis*-regulatory variation in B73 and Mo17 alleles. It should be noted that there is some evidence for preferential detection of the B73 allele (center of the distribution is shifted up and to the right) that is likely the result of more efficient mapping of sequence reads to the reference B73 allele than to the Mo17 allele rather than an actual expression bias. The genes were categorized based on whether they exhibit biased allelic expression in both hybrids (chi-square <0.01) relative to the expectation of 2:1 maternal to paternal transcripts. Nearly 5,000 genes showed a consistent bias towards the B73 or Mo17 allele in both reciprocal hybrids (Figure 4A). A smaller number of genes showed evidence for consistent maternal (n=103) or paternal (n=365) bias in both reciprocal hybrids (Figure 4A).

We focused on the relative expression of the maternal and paternal alleles of genes with parent-of-origin effects. The implementation of chi-square tests to find consistent maternal or paternal bias provides one mechanism to identify maternally expressed genes (MEGs) or paternally expressed genes (PEGs) (Figure 1). However, these criteria can identify genes with high expression levels that have parental effects of small magnitude. Therefore, we further filtered the list of MEGs and PEGs by requiring at least 90% of the transcripts to be maternal or paternal in both reciprocal hybrids (Figure 1). This resulted in the identification of 72 MEGs and 46 PEGs. A similar set of allele-specific expression analyses were performed on the RNA-seq data obtained from embryo tissue in order to identify MEGs or PEGs in embryo. These analyses identified

29 MEGs and 9 PEGs. However, all of these genes were also identified as imprinted in endosperm and all show much higher levels of expression in endosperm than in embryo. This finding has several potential explanations. First, it is possible that there is endosperm contaminating our embryo samples. However, while it is relatively easy to separate endosperm and embryo at 14 DAP we can't rule out contamination. A second explanation for our finding is that there is trafficking of transcripts produced in the endosperm to the embryo tissue. This would result in apparent parent-of-origin bias for these transcripts without active expression in embryos. A third explanation is that these transcripts exhibit active parent-of-origin specific transcription in both embryo and endosperm, similar to *mee1* (Jahnke and Scholten 2009). A fourth possible explanation is that these are relatively stable transcripts that are inherited from the sperm or egg. With information only on steady state transcript levels, rather than on active transcription in embryo tissue, it is difficult to separate these potential explanations. We thus focused the remaining analyses on the endosperm-imprinted genes.

Genes could be identified as MEGs for several reasons in addition to true imprinted expression, including contamination by maternal tissues and the presence of stable maternally inherited transcripts. In contrast, PEGs could be due to the presence of transcripts provided by the male gamete or due to imprinted expression in endosperm. While it is possible that some of the MEGs and PEGs are the result of stable transcripts provided by the parental gametes these are likely to be rare at 14 days after pollination. To assess whether some of the MEGs might result from contamination with maternal tissues we assessed the relative expression patterns for the 58 MEGs that are represented in a maize gene expression atlas (Sekhon et al., 2011) (the remaining 14 MEGs were not assessed by Sekhon et al., 2011). Hierarchical clustering for these genes was performed in a series of tissues representing different developmental stages of embryo, endosperm, whole seed and pericarp (Figure 5). A subset of the MEGs exhibited at least five-fold higher expression in whole seeds relative to endosperm tissue (18/58) and 13 of the 18 genes were not highly expressed in endosperm tissue. Thus, the maternal bias for these genes may be due to presence of small amounts of contamination from maternal tissues

(Figure 6). The 18 genes were therefore omitted from further analyses, leaving 54 MEGs that were classified as putatively imprinted genes. Expression data indicated higher levels of expression in endosperm tissue than in whole seeds for all 40 of the remaining genes with expression atlas data (Sekhon et al., 2011).

The 54 MEGs and 46 PEGs resulted in 100 genes classified as putatively imprinted. There are three well-characterized imprinted genes that are polymorphic in B73 and Mo17: *Fie1* (Gutierrez-Marcos et al., 2003; Danilevskaya et al., 2003), *Nrp1* (Guo et al., 2003) and *Mez1* (Haun et al., 2007). Imprinting for each of these three loci was well supported by the RNA-seq data (Figure 4B). The cross between B73 and Mo17 did not include polymorphic or imprinted alleles for some other known imprinted loci and therefore we could not assess imprinting at *R* (Kermicle 1970), *B* (Selinger and Chandler 2001), *mee1* (Jahnke and Scholten 2009) or *Meg1* (Gutierrez-Marcos et al., 2004). Stupar et al. 2007) reported imprinting for two other EST sequences in B73xMo17. Only one of these genes, GRMZM2G099295, was within the filtered gene set and although this gene did not contain enough allele-specific reads to pass our filtering criteria the observed reads, nine maternal reads and zero paternal reads, did support imprinting.

We proceeded to confirm the allele-specific expression pattern observed in the RNA-seq analysis for a number of the novel imprinted genes. The imprinted expression of 24 of the novel genes was assessed by either cleaved amplified polymorphic sequence (CAPS) assays (Figure 7) or sequencing of RT-PCR products (Figure 8) using an independent biological replicate sample of hybrid 14 DAP endosperm tissue. Imprinted expression was confirmed for 23 of the 24 genes that were tested. Together with the three previously identified examples of imprinting we have confirmed imprinting for 27 of the genes. For several of these genes (Figure 7) we further confirmed the imprinting at additional time-points earlier and later in endosperm development.

Previous studies of small numbers of imprinted genes have noted that many are expressed exclusively in endosperm tissue (Danilevskaya et al., 2003; Kinoshita et al., 2004). Many of our putative imprinted genes (71/100) were surveyed in a recent maize

expression atlas (Sekhon et al., 2011). The normalized expression levels in all of these tissues were compared to identify genes with endosperm-preferred expression (Figure 5). The average expression in endosperm was compared to the average expression level in all other tissues (excluding whole seeds which include endosperm tissue) in order to identify genes with high levels of endosperm expression relative to other tissues. Nearly half of the genes (31/71) exhibit at least five-fold higher expression in endosperm than in other tissues (Figure 5). The genes with endosperm preferred expression include both MEGs (20) and PEGs (11). The remaining genes exhibit moderate levels of expression in other tissues or developmental stages.

Potential examples of allele-specific imprinting

Imprinting describes unbalanced expression of two alleles based on their parent of origin. In our identification of MEG and PEG genes we required >90% uniparental transcripts in both reciprocal hybrids samples. This requires that both the B73 and Mo17 alleles exhibit imprinted expression patterns. Given the hypothesis that imprinting may target transposable element sequences (Kohler and Weinhofer-Molisch 2010; Mosher and Melynk 2010; Bauer and Fischer 2011) there may be instances of imprinting in which certain alleles are imprinted but other alleles at the same locus do not exhibit imprinting due to epigenetic or genetic variation (allele-specific imprinting). Indeed, there are examples of allele-specific imprinting at the *R*, *B* and *dzr* loci of maize (Kermicle 1970; Chaudhuri et al., 1994; Selinger and Chandler 2001) and several putative examples of allele-specific imprinting have been identified in Arabidopsis (Hsieh et al., 2011; Wolff et al., 2011; Gehring et al., 2011). We identified potential allele-specific imprinting candidates as genes with at least 20 allele-specific reads in both reciprocal hybrids that had >98% of the reads from one parent in one of the hybrids but bi-allelic (40-80% maternal expression) in the other hybrid. A total of 53 potential allele-specific imprinting candidates were identified in our data. These results would need to be tested in additional crosses in order to further evaluate allele-specific imprinting in maize.

Endosperm-specific hypomethylation at many imprinted maize genes:

For many previously identified imprinted plant genes there is evidence for reduced DNA methylation of the maternal allele (Jullien and Berger 2009; Raissig et al., 2011). We generated whole-genome methylation profiling data for embryo, endosperm and leaf tissue for both B73 and Mo17. The methylation profiling was performed by immunoprecipitation with a 5-methylcytosine antibody followed by hybridization to a 2.1 million feature long oligonucleotide microarray that contains probes every ~200bp across the low-copy space of the maize genome. This array platform included 3,058 probes within or in regions surrounding the 100 imprinted genes. We searched for examples of multiple adjacent probes near the imprinted genes that showed lower levels of DNA methylation in endosperm tissue relative to embryo and leaf in both B73 and Mo17. There are 19 genes that contain a region of endosperm-specific hypomethylation (Figure 5). The 19 regions of reduced endosperm methylation include 12 genes with preferred maternal expression and 7 genes with preferred paternal expression.

The reduced methylation in endosperm tissue is hypothesized to be the result of demethylation of the maternal allele while the paternal allele remains methylated. Indeed, one of these hypomethylated regions was found at the same site as the previously identified differentially methylated region (DMR) of *ZmFie1* (Gutierrez-Marcos et al., 2006; Hermon et al., 2007) (Figure 9). A previously identified DMR at *Mez1* (Haun et al., 2007) was not identified using these criteria. However, closer inspection found that the single probe at the DMR showed evidence for lower methylation levels in endosperm (Figure 9). The *Mez1* DMR is near repetitive sequence so there are no other probes covering this region. It is likely that some other genes may also have DMRs that were not identified since they were not interrogated by multiple adjacent probes and did not fit our stringent criteria for DMR identification.

We proceeded to assess the allele-specific methylation patterns at several of the DMRs identified by the MeDIP-chip analysis. These assays were limited to 9 DMR regions that include SNPs that allow for discrimination of the maternal and paternal alleles. For each region, we amplified undigested genomic DNA and genomic DNA that

had been subjected to a restriction digest using the methylation dependent restriction enzyme *FspEI*. The amplified products were then sequenced to assess the relative abundance of the two alleles. In all cases, the undigested DNA revealed a mixture of the maternal and paternal alleles (Figure 10 and 11). However, following digestion with the methylation-dependent enzyme only the maternal SNP was observed. This provides evidence for differential methylation of the paternal allele and maternal allele at these regions with lower levels of methylation at the maternal allele. In all 9 examples tested (6 MEGs and 3 PEGs) we found evidence for maternal hypomethylation.

Our data provided evidence for DMRs near some imprinted genes but not others. We hypothesized that DMRs may be more prevalent near imprinted genes that exhibit endosperm-preferred expression. These genes could be silenced by DNA methylation in other tissues but then exhibit increased expression in endosperm concomitant with reduced methylation. Expression atlas data (Sekhon et al., 2011) was available for 14/19 genes that contain DMRs. The majority of these genes (10/14) have at least 5-fold higher expression levels in endosperm tissue than in other plant tissues (Figure 5). This provides evidence for prevalent DMRs in imprinted genes with endosperm-preferred expression.

Characterization of imprinted maize genes:

The chromosomal positions of the remaining 100 imprinted genes were examined (Figure 4C). While clustering of imprinted genes is prevalent in mammals there is little evidence of clustering in plant species. The imprinted genes were distributed across all ten maize chromosomes. We found two examples in which adjacent genes show maternal specific expression including GRMZM2G354558 / GRMZM2G354579 on chromosome 3 and GRMZM2G144594 / GRMZM2G569356 on chromosome 7. In both cases the two genes show primarily maternal expression and are expressed from the same strand. Other than these two examples, we did not observe evidence for clusters of multiple imprinted genes located within 250 kb of each other.

BLASTP alignments of the protein sequences from the maize imprinted genes identified a related Arabidopsis protein for 40/54 of the MEGs and 40/46 of the PEGs. The annotations of these Arabidopsis genes were analyzed to identify functional categories of genes (based on GO_slim plant annotations) that are over-represented using BinGO (Maere et al., 2005). The PEGs are enriched for binding activities based on nucleic acid and DNA binding proteins. In addition, these genes are enriched for kinase activities and for genes involved in flower development (Figure 12). The MEGs are enriched for several processes including multicellular development and response to stimulus (Table 2). In addition, when the GO Biological process annotations of the imprinted genes were assessed there are enrichments for chromatin modification and organization terms in both MEGs and PEGs. In particular, maize orthologs of VIM1 (Woo et al., 2007), VRN5 (Greb et al., 2007), FIE (Ohad et al., 1999) and CLF (Goodrich et al., 1997) exhibit imprinted expression patterns. The over-representation of chromatin proteins and transcription factors suggest that many targets of imprinting have important regulatory roles.

Several recent studies in Arabidopsis and rice have utilized RNA-seq to identify imprinted genes (Hsieh et al., 2011; Wolff et al., 2011; Luo et al., 2011; Gehring et al., 2011). We sought to determine whether similar genes are imprinted in multiple species. This comparison is complicated by the fact that different sets of genes were assessed in each species due to SNP availability and read coverage. Imprinting was assessed for roughly one-half to one-third of the genes in each of the species being studied. The 100 maize imprinted genes were used as queries against all rice or Arabidopsis protein sequences and up to three close matches in the target species were identified. There are eleven examples of maize genes where one of the closest matches in rice exhibits imprinting (Luo et al., 2011) (Table 3). Luo et al. 2011) did not report which rice genes contained polymorphisms between the parental lines and therefore it is not clear whether the absence of detected imprinting was due to lack of imprinting or lack of polymoprhisms. Therefore, it is possible that the rice ortholog for some of the other imprinted maize genes is actually imprinted but could not be surveyed by Luo et al.

2011). The best Arabidopsis matches (up to three) were identified for 80 of the 100 imprinted maize genes and were compared with data from Arabidopsis studies. Similar to our study, each of the Arabidopsis imprinting studies identified a larger non-stringent list of imprinted genes and a subsequent filtered list that aimed to remove genes due to potential maternal tissue contamination or other false positives. We identified 10 maize imprinted genes for which Arabidopsis homologues were stringently identified as imprinted (Table 4). We also noted instances in which some of these genes were identified in the stringent list for one study but in the non-stringent list for another study and genes for which Gehring et al (2009) identified endosperm DMRs near these Arabidopsis genes. There are two examples in which two different maize genes are both related to the same imprinted Arabidopsis genes. There are two genes for which imprinting is detected in all three species, AC19534.3FG003 / Os04g22240 / AT1G57820 (VIM1) and GRMZM2G365731 / Os10g30944 / AT4G11400 (ARID/BRIGHT DNA binding domain). Given that less than 1% of the genes in any one species were identified as imprinted, the overlap of imprinting among maize and Arabidopsis or rice was highly significant ($p<0.001$). In addition, many examples that lack conservation for imprinting may reflect false-negatives due to lack of ability to assess imprinting in all three species.

DISCUSSION

We used RNA-seq to identify MEGs and PEGs that likely represent examples of imprinting in maize. The use of genome-wide allele-specific expression profiling has greatly increased the number of putatively imprinted maize genes. Previous studies documented ~10 maize imprinted genes (Springer and Gutierrez-Marcos 2009; Raissig et al., 2011). Our analysis uncovered several of these same loci and identified many novel imprinted genes (the others did not have SNPs in B73 / Mo17 and could not be assessed). Some known imprinted genes exhibit parent-of-origin effects during early endosperm development but show bi-allelic expression at later stages (Danilevskaya et al., 2003; Grimanelli 2005). We intentionally profiled a relatively late stage of endosperm

development to identify genes with consistent imprinting throughout endosperm development rather than delayed paternal activation. Indeed, each of the genes that we tested exhibited imprinting from 10 to 22 DAP. These genes are putatively classified as imprinted based on parent-of-origin bias in gene expression. However, it is formally possible that some of these allelic biases may represent stable transcripts inherited from parental cells rather than active mono-allelic transcription in the F_1 (Raissig et al., 2011). At least 30 of 71 genes surveyed in the expression atlas exhibit increasing transcript levels during endosperm development, which supports the classification of imprinted expression. The analysis of the expression patterns and DNA methylation patterns for the imprinted maize genes provided an opportunity to test several hypotheses about the expression patterns of imprinted genes and mechanisms of imprinting. We were also able to compare imprinting in maize and other species to identify examples of conserved imprinting between species.

Imprinting is more prevalent in endosperm than in embryo

The majority of examples of imprinting in plants are documented in endosperm tissue (Huh et al., 2008; Jullien and Berger 2009; Springer and Gutierrez-Marcos 2009). There is one documented example of a maize gene (*mee1*) that is maternally expressed in both embryo and endosperm (Jahnke and Scholten 2009). Several of the recent RNA-seq based analyses of imprinting have compared embryo and endosperm tissue and have found few or no examples of imprinting in the embryo of Arabidopsis (Hsieh et al., 2011, Gehring et al., 2011) and rice (Luo et al., 2011). In our study we assessed allele-specific expression in both endosperm and embryo tissue of maize. There were substantially fewer examples of MEGs or PEGs in embryo compared to endosperm. We also noted that each example of MEGs or PEGs in embryo was also observed in endosperm. This may reflect low levels of endosperm contamination of our embryo tissue with endosperm or maternal tissues, movement of transcripts from endosperm to embryo, or active imprinted transcription in both endosperm and embryo. It is difficult to fully evaluate these three potential explanations using data on steady-state transcript levels. However,

we did not observe evidence for imprinting of genes in embryo that are not imprinted in endosperm. Further studies such as allele-specific occupancy for RNA polymerase II or nuclear run-on assays in embryo tissue will be necessary to evaluate whether there is active allele-specific transcription for these genes.

Multiple classes of imprinted genes in plants

Previous studies in Arabidopsis have provided evidence for multiple molecular mechanisms of imprinting in plants. There is evidence that both DNA methylation and chromatin modifications can both play roles in imprinting and there are likely additional mechanisms as well (Hsieh et al., 2011; Wolff et al., 2011). Our analysis of expression patterns and methylation levels provides further evidence for a role for DNA methylation regulating imprinting for a subset of imprinted genes. We noted that some of the maize imprinted genes exhibit endosperm-preferred expression such that expression is low in all tissues except for endosperm. This same set of genes is highly enriched for DMRs near the imprinted genes. DNA methylation may play an important role in regulating expression and imprinting for this set of genes. It is noteworthy that the set of genes that have DMRs include both MEGs and PEGs and in both cases the maternal allele exhibits hypomethylation. The specific role of DNA methylation in regulating imprinting at PEGs is not clear but has been noted in Arabidopsis as well (Gehring et al., 2009; Gehring et al., 2011, Wolff et al., 2011, Hsieh et al., 2011). Hsieh et al (2011) provided evidence that maternal demethylation at PEGs may be important for proper silencing of the maternal allele. Maternal demethylation at PEGs may allow for targeting of PcG proteins to the maternal allele (Weinhofer et al., 2010; Hsieh et al., 2011).

Conservation of imprinting among species

High-resolution allele-specific expression profiling has now identified imprinted genes in three species. In all three species, there are ~50-150 imprinted genes identified from a set of 10,000 to 15,000 testable genes. It seems that imprinting is relatively infrequent in plants but does affect several hundred genes in each species, similar to the

frequency of imprinted genes in mammals (Feil and Berger 2007). Given these low rates we would expect to observe very few examples of imprinting of the same genes in both species by chance. However, there are ~10 examples of conserved imprinting between each of the species. It should be noted that comparing imprinted genes between species is complicated by the requirement for SNPs between the two alleles and therefore the set of genes tested is not the same between species. For example, the conservation for imprinting between GRMZM2G356731 and AT4G11400 was identified in one Arabidopsis study (Wolff et al., 2011) but could not be tested in the other studies due to lack of polymorphisms in AT4G11400 in the Ler and Col ecotypes. Interestingly, we noted that the genes with conserved imprinting between species tend to be genes with endosperm-preferred expression. There are eleven examples of conserved imprinting that show endosperm-preferred expression in maize and only two examples of conserved imprinting that do not have endosperm-preferred expression (the other four genes with conserved imprinting do not have expression atlas data). This provides evidence that evolutionarily conserved imprinting is primarily restricted to genes that are regulated by DNA methylation and are exclusively expressed in the endosperm.

Genes with conserved imprinting in maize and either rice or Arabidopsis are potentially key regulators of seed development. The analysis of the GO annotations for genes that are imprinted in both maize and Arabidopsis identified over-representation for a number of biological process annotation. Half of these genes are annotated as being involved in regulation of biological processes. There are a number of genes with putative chromatin or DNA binding annotations as well. For example, genes related to VIM proteins, which are involved in regulating DNA methylation patterns in Arabidopsis (Woo et al., 2007) are imprinted in all three species. It is quite likely that the genes with conserved imprinting play important roles in regulation of seed development. Further functional analyses of imprinted plant genes will be necessary to shed light on the functional consequences of imprinting in plants.

MATERIAL AND METHODS

RNA-seq analysis: B73 and Mo17 plants were grown in Saint Paul at the University of Minnesota Agricultural experiment station during the summer of 2010. Reciprocal crosses and self-pollinations for all genotypes were performed on August 2nd, 3rd and 4th and several ears representing each cross were harvested 14 days after pollination (DAP). The endosperm and embryo tissue were dissected from at least five ears for each genotype and were pooled together and frozen in liquid nitrogen. RNA was isolated using Qiagen (Valencia CA) RNAeasy kits and subsequently purified by LiCl precipitations. These RNA samples were submitted to the University of Minnesota Biomedical Genomics Center for sequencing using the Illumina GAIIX (Inbred samples) or HiSeq (Hybrid samples) platforms.

Prior to alignment to the reference genome, the nucleotides of each raw read were scanned for low quality bases and individual bases with PHRED quality values <15 out of 40 (Ewing and Green 1998; Ewing *et al.* 1998) or having an error rate of □0.03% were removed by our trimming pipeline, which is a variant of previously published software but adapted for Next-Generation Sequencing (NGS) data (Li and Chou 2004). Expression levels for 3 biological replicates of endosperm and embryo tissue from B73 and Mo17 were developed independently using the Tophat-Cufflinks transcript assembly pipeline (Trapnell et al., 2010). Reads were aligned to the B73 AGPv2 reference assembly using Tophat. Transcript quantification and expression levels in the form of reads per mapped kilobase (RPKM) for the B73 working gene set were calculated using Cufflinks under default parameters.

Trimmed reads were aligned to the reference genome using GSNAP (Wu and Nacu 2010) and reads mapping uniquely or to single loci (≤2 mismatches every 36 bp and ≤3 bp tails allowed) were used for subsequent analysis. Polymorphisms detected from the alignment of B73, Mo17 and reciprocal hybrid samples against the reference B73 genome were used to develop read counts of B73 and Mo17 alleles using previously identified SNP sites.

The number of reads containing the B73 or Mo17 allele was summed for all SNPs within the same gene. The 12,571 genes that contain at least 10 reads that could be assigned to a particular allele were used to perform chi-square tests (relative to an expected 2 maternal : 1 paternal) ratio. The 468 genes that exhibit a significant (χ^2 <0.01) difference from the 2:1 ratio were further filtered to identify genes for which at least 90% of the reads were derived from one parent in both reciprocal hybrids. The RNA-seq reads are deposited at the NCBI SRA under accession SRA048055.2.

CAPS and RT-PCR sequencing validation assays: Imprinted expression patterns were validated and further characterized using sequencing or CAPS assays for several genes. RT-PCR was performed on first-strand cDNA that was derived from an independent set of 14 DAP RNA samples or from a developmental series of endosperm time points (Haun et al., 2007) using the primers listed in Supplemental Table 5. In some cases the RT-PCR products were directly sequenced to assess the contribution of the parental alleles in the hybrid samples. For other genes the amplification products were digested with restriction enzymes that differentiate the B73 and Mo17 alleles (specific enzymes listed in Supplemental Table 5).

DNA methylation profiling: DNA methylation was profiled according to the same methods described in Eichten et al (2011). Endosperm DNA was isolated from three replicates of 14 DAP endosperm from both B73 and Mo17 (same tissue used for RNA-seq) using DNAeasy kits from Qiagen (Valencia CA). Methylated DNA was immunoprecipitated with an anti-5- methylcytosine monoclonal antibody from 400ng of sonicated DNA using the Methylated DNA IP kit (Zymo Research, Orange, CA; Cat # D5101). For each replication and genotype, whole genome amplification was conducted on 50-100ng IP DNA and also 50-100ng of sonicated DNA (input control) using the Whole Genome Amplification kit (Sigma Aldrich, St. Louis, MO, Cat # WGA2-50RXN). For each amplified IP input sample, 3 μg amplified DNA were labeled using the Dual-Color Labeling Kit (Roche NimbleGen, Madison WI; Cat # 05223547001) according to

the array manufacturer's protocol (Roche NimbleGen Methylation UserGuide v7.0). Each IP sample was labeled with Cy5 and each input/control sonicated DNA was labeled with Cy3. Samples were hybridized to the custom 2.1M probe array (GEO Platform GPL13499) for 16-20hrs at 42°C. Slides were washed and scanned according to NimbleGen's protocol for the GenePix4000B scanner. Images were aligned and quantified using NimbleScan software (Roche NimbleGen), which produced raw data reports for each probe on the array.

Normalization and linear modeling for DNA methylation data*:* Pair files exported from NimbleScan were imported into the Bioconductor statistical environment (http://bioconductor.org/). Sample-dependent MeDIP enrichments were estimated for each probe by fitting a fixed linear model accounting for array, dye, and sample effects to the data using the limma package (Smyth, 2004). The following contrasts were then fit: B73 IP vs B73 input; Mo17 input vs B73 input; Mo17 IP vs (Mo17 input vs B73 input); B73 IP vs (Mo17 IP vs (Mo17 input vs B73 input)). Moderated t-statistics and the log-odds score for differential MeDIP enrichment was computed by empirical Bayes shrinkage of the standard errors with the False Discovery Rate controlled to 0.05. Microarray results were deposited with NCBI GEO under accession GSE33730.

Identification of DMRs*:* Putative DMRs were identified by comparing DNA methylation levels in endosperm and leaf tissue (from Eichten et al., 2011). All probes within 3 kilobases of the imprinted genes were identified and the examples in which multiple adjacent probes had significantly lower ($q<0.05$) methylation levels in endosperm relative to leaf tissue. The allele-specific methylation difference was tested for 8 of these regions that included SNPs by amplification with and without digestion with the DNA methylation dependent enzyme FspEI (New England Biolabs, Ipswich, MA). 1ug genomic DNA was digested for 16hr with FspEI according to manufacturer's recommendations. Mock digestions were performed substituting glycerol for restriction enzyme. Following digestion, PCR amplification of the target regions was performed

(primers listed in Supplemental table 6). The amplified products were sequenced using di-deoxy sequencing at the University of Minnesota Biomedical Genomics Center.

Informatics: The protein sequences from all of the filtered genes from annotation AGPv2_FGSv5b.60 were used as queries against the Arabidopsis (TAIR.10) or rice Osa1 (Ouyang et al., 2007) protein sequences. The closest match that had an eValue <1e-20 was retained for each of the maize genes. Annotations for maize genes were inferred based on the annotations for the closest sequence in Arabidopsis using the TAIR annotations.

REFERENCES

Bauer M.J. and Fischer R.L. 2011). Genome demethylation and imprinting in the endosperm. Curr. Opin. Plant Biol. 14: 162-167.

Chaudhuri S. and Messing J. 1994). Allele-Specific Parental Imprinting of dzr1, a Posttranscriptional Regulator of Zein Accumulation. PNAS 91: 4867-4871.

Danilevskaya O.N., Hermon P., Hantke S., Muszynski M.G., Kollipara K., and Ananiev E.V. 2003). Duplicated fie genes in maize: expression pattern and imprinting suggest distinct functions. Plant Cell 15: 425-438.

Eichten S.E., Swanson-Wagner R.A., Schnable J.C., Waters A.J., Hermanson P.J., Liu S., Yeh C-T., Jia Y., Gendler K., Freeling M., Schnable P.S., Vaughn M.W., and Springer N.M. 2011) Heritable epigenetic variation among maize inbreds. PLoS Genetics. In press.

Ewing B. and Green P. 1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8: 186-194.

Ewing B., Hillier L., Wendl M.C., and Green P. 1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8: 175-185.

Feil R. and Berger F. 2007). Convergent evolution of genomic imprinting in plants and mammals. Trends Genet. 23: 192-199.

Gehring M., Choi Y., and Fischer R.L. 2004). Imprinting and seed development. Plant Cell 16 Suppl: S203-13.

Gehring M., Huh J.H., Hsieh T.F., Penterman J., Choi Y., Harada J.J., Goldberg R.B., and Fischer R.L. 2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124: 495-506.

Gehring M., Bubb K.L., and Henikoff S. 2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324: 1447-1451.

Gehring M., Missirian V., and Henikoff S. 2011). Genomic Analysis of Parent-of-Origin Allelic Expression in Arabidopsis thaliana Seeds. PLoS One 6: e23687.

Goodrich J., Puangsomlee P., Martin M., Long D., Meyerowitz E.M., and Coupland G. 1997). A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature 386: 44-51.

Greb T., Mylne J.S., Crevillen P., Geraldo N., An H., Gendall A.R., and Dean C. 2007). The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. Curr. Biol. 17: 73-78.

Grimanelli D., Perotti E., Ramirez J., and Leblanc O. 2005). Timing of the maternal-to-zygotic transition during early seed development in maize. Plant Cell 17: 1061-1072.

Grossniklaus U., Vielle-Calzada J.P., Hoeppner M.A., and Gagliano W.B. 1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science 280: 446-450.

Guo M., Rupe M.A., Danilevskaya O.N., Yang X., and Hu Z. 2003). Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. Plant J. 36: 30-44.

Gutierrez-Marcos J.F., Pennington P.D., Costa L.M., and Dickinson H.G. 2003). Imprinting in the endosperm: a possible role in preventing wide hybridization. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 358: 1105-1111.

Gutierrez-Marcos J.F., Costa L.M., Biderre-Petit C., Khbaya B., O'Sullivan D.M., Wormald M., Perez P., and Dickinson H.G. 2004). maternally expressed gene1 Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. Plant Cell 16: 1288-1301.

Gutierrez-Marcos J.F., Costa L.M., Dal Pra M., Scholten S., Kranz E., Perez P., and Dickinson H.G. 2006). Epigenetic asymmetry of imprinted genes in plant gametes. Nat. Genet. 38: 876-878.

Haig D. and Westoby M. 1989). Parent-specific gene-expression and the triploid endosperm. Am. Nat. 134: 147-155.

Haig D. 2004). Genomic imprinting and kinship: how good is the evidence? Annu. Rev. Genet. 38: 553-585.

Haun W.J., Laoueille-Duprat S., O'connell M.J., Spillane C., Grossniklaus U., Phillips A.R., Kaeppler S.M., and Springer N.M. 2007). Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste (Mez) homologs. Plant J. 49: 325-337.

Haun W.J. and Springer N.M. 2008). Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. Plant J. 56: 903-912.

Hermon P., Srilunchang K.O., Zou J., Dresselhaus T., and Danilevskaya O.N. 2007). Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of the maternal allele. Plant Mol. Biol. 64: 387-395.

Hsieh T.F., Ibarra C.A., Silva P., Zemach A., Eshed-Williams L., Fischer R.L., and Zilberman D. 2009). Genome-wide demethylation of Arabidopsis endosperm. Science 324: 1451-1454.

Hsieh T.F., Shin J., Uzawa R., Silva P., Cohen S., Bauer M.J., Hashimoto M., Kirkbride R.C., Harada J.J., Zilberman D., and Fischer R.L. 2011). Regulation of imprinted gene expression in Arabidopsis endosperm. Proc. Natl. Acad. Sci. U. S. A. 108: 1755-1762.

Huh J.H., Bauer M.J., Hsieh T.F., and Fischer R.L. 2008). Cellular programming of plant gene imprinting. Cell 132: 735-744.

Jahnke S. and Scholten S. 2009). Epigenetic resetting of a gene imprinted in plant embryos. Curr. Biol. 19: 1677-1681.

Jullien P.E., Katz A., Oliva M., Ohad N., and Berger F. 2006). Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis. Curr. Biol. 16: 486-492.

Jullien P.E. and Berger F. 2009). Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr. Opin. Plant Biol. 12: 637-642.

Kermicle J.L. 1970). Dependence of the R-Mottled Aleurone Phenotype in Maize on Mode of Sexual Transmission. Genetics 66: 69-85.

Kinoshita T., Yadegari R., Harada J.J., Goldberg R.B., and Fischer R.L. 1999). Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. Plant Cell 11: 1945-1952.

Kinoshita T., Miura A., Choi Y., Kinoshita Y., Cao X., Jacobsen S.E., Fischer R.L., and Kakutani T. 2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science 303: 521-523.

Kohler C., Hennig L., Spillane C., Pien S., Gruissem W., and Grossniklaus U. 2003). The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes Dev. 17: 1540-1553.

Kohler C. and Weinhofer-Molisch I. 2010). Mechanisms and evolution of genomic imprinting in plants. Heredity 105: 57-63.

Lauria M., Rupe M., Guo M., Kranz E., Pirona R., Viotti A., and Lund G. 2004). Extensive maternal DNA hypomethylation in the endosperm of Zea mays. Plant Cell 16: 510-522.

Li S. and Chou H.H. 2004). LUCY2: an interactive DNA sequence quality trimming and vector removal tool. Bioinformatics 20: 2865-2866.

Luo M., Bilodeau P., Dennis E.S., Peacock W.J., and Chaudhury A. 2000). Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. Proc. Natl. Acad. Sci. U. S. A. 97: 10637-10642.

Luo M., Taylor J.M., Spriggs A., Zhang H., Wu X., Russell S., Singh M., and Koltunow A. 2011). A Genome-Wide Survey of Imprinted Genes in Rice Seeds Reveals Imprinting Primarily Occurs in the Endosperm. PLoS Genet. 7: e1002125.

Maere S., Heymans K., and Kuiper M. 2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21: 3448-3449.

McKeown P.C., Laouielle-Duprat S., Prins P., Wolff P., Schmid M.W., Donoghue M.T., Fort A., Duszynska D., Comte A., Lao N.T., Wennblom T.J., Smant G., Kohler C., Grossniklaus U., and Spillane C. 2011). Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds. BMC Plant. Biol. 11: 113.

Mosher R.A. and Melnyk C.W. 2010). siRNAs and DNA methylation: seedy epigenetics. Trends Plant Sci. 15: 204-210.

Ohad N., Yadegari R., Margossian L., Hannon M., Michaeli D., Harada J.J., Goldberg R.B., and Fischer R.L. 1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell 11: 407-416.

Ouyang S., Zhu W., Hamilton J., Lin H., Campbell M., Childs K., Thibaud-Nissen F., Malek R.L., Lee Y., Zheng L., Orvis J., Haas B., Wortman J., and Buell C.R. 2007). The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. 35: D883-7.

Raissig M.T., Baroux C., and Grossniklaus U. 2011). Regulation and flexibility of genomic imprinting during seed development. Plant Cell 23: 16-26.

Sekhon R.S., Lin H., Childs K.L., Hansey C.N., Buell C.R., de Leon N., and Kaeppler S.M. 2011). Genome-wide atlas of transcription during maize development. Plant J. 66: 553-563.

Selinger D.A. and Chandler V.L. 2001). B-Bolivia, an allele of the maize b1 gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream. Plant Physiol. 125: 1363-1379.

Smyth G.K. 2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3: Article3.

Springer N.M. and Gutierrez-Marcos J. 2009). Imprinting in maize. In *the Maize Handbook*, S. Hake and J. Bennetzen, eds (New York: Springer).

Springer N.M. 2009). Small RNAs: how seeds remember to obey their mother. Curr. Biol. 19: R649-51.

Stupar R.M., Hermanson P.J., and Springer N.M. 2007). Nonadditive expression and parent-of-origin effects identified by microarray and allele-specific expression profiling of maize endosperm. Plant Physiol. 145: 411-425.

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J., and Pachter L. 2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28: 511-515.

Weinhofer I., Hehenberger E., Roszak P., Hennig L., and Kohler C. 2010). H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS Genet. 6: e1001152.

Wolff P., Weinhofer I., Seguin J., Roszak P., Beisel C., Donoghue M.T., Spillane C., Nordborg M., Rehmsmeier M., and Kohler C. 2011). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. PLoS Genet. 7: e1002126.

Woo H.R., Pontes O., Pikaard C.S., and Richards E.J. 2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev. 21: 267-277.

Wu T.D. and Nacu S. 2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873-881.

Zemach A., Kim M.Y., Silva P., Rodrigues J.A., Dotson B., Brooks M.D., and Zilberman D. 2010). Local DNA hypomethylation activates genes in rice endosperm. Proc. Natl. Acad. Sci. U. S. A. 107: 18729-18734.

Table 1. RNA-seq reads for B73 and Mo17 14 DAP endosperm

Genotype	Read type	# Trimmed reads	Trimmed Base Pairs (bp)	# Reads aligned to Zm5b.60 FGS genes
B73 replicate 1	1 x 76bp	6,489,633	478,481,544	4,092,833
B73 replicate 2	1 x 76bp	7,247,795	534,805,319	4,918,597
B73 replicate 3	1 x 76bp	8,999,189	670,198,080	6,055,606
Mo17 replicate 1	1 x 76bp	7,513,825	556,416,464	3,897,504
Mo17 replicate 2	1 x 76bp	7,295,965	507,649,188	4,962,175
Mo17 replicate 3	1 x 76bp	5,013,980	362,013,476	2,995,521
B73xMo17 pool	2 x 76bp	209,874,775	12,633,236,001	126,297,384
Mo17xB73 pool	2 x 76bp	201,851,318	12,239,426,296	118,754,499
Total		454,286,480	27,982,226,368	271,974,119

Table 2. GO Slim enrichments imprinted genes

Paternal expressed genes (n=40 genes with Arabidopsis homologs)

GO_ID	Description	Genes	p-value*
3677	DNA binding	11	0.01
9908	flower development	4	0.01
3676	nucleic acid binding	14	0.02
16301	kinase activity	7	0.02
5488	binding	24	0.02

Maternal expressed genes (n=40 genes with Arabidopsis homologs)

GO_ID	Description	Genes	p-value*
7275	multicellular organismal development	11	0.00
32501	multicellular organismal process	11	0.00
50896	response to stimulus	13	0.02
50789	regulation of biological process	11	0.04

*FDR corrected P-value

Table 3. Examples of conserved imprinting in maize and rice

Maize Gene ID	Chr	Maternal read#	Paternal read#	Maize pattern	Rice Gene ID	Rice description	Luo et al	Arabidopsis Gene ID
AC191534.3_FG003	7	23	2397	PEG	Os04g22240	C3HC4 type zinc-finger	PEG	AT1G57820
GRMZM2G000404	7	0	160	PEG	Os09g28940	ubiquitin-specific protease	PEG	
GRMZM2G028366	5	15	13763	PEG	Os02g12840	DEAD-box ATP-dependent RNA helicase	PEG	AT1G54270
GRMZM2G073700	6	15858	592	MEG	Os06g11730	RNA-binding (RRM/RBD/RNP motifs)	MEG	AT1G78260
GRMZM2G110306	1	0	33	PEG	Os08g27240	ARID/BRIGHT DNA-binding domain	PEG	AT3G43240
GRMZM2G118205	4	9320	17	MEG	Os08g04290	OsFIE	MEG	AT3G20740
GRMZM2G170099	3	2891	20	MEG	Os01g18810	hypothetical protein	MEG	
GRMZM2G365731	1	2	396	PEG	Os10g30944	ARID/BRIGHT DNA-binding domain	PEG	AT4G11400
GRMZM2G379898	1	72	0	MEG	Os03g01320	Protease inhibitor/seed storage/LTP	MEG	AT1G62500
GRMZM2G447406	8	100	3155	PEG	Os01g70060	hypothetical protein	PEG	
GRMZM5G845175	10	24	708	PEG	Os04g32880	CBS domain - membrane protein	PEG	AT1G09020

Table 4. Examples of conserved imprinting in maize and Arabidopsis

Maize Gene ID	Chr	Maternal read#	Paternal read#	Maize pattern	Arabidopsis Gene ID	Arabidopsis description	Gehring et al (2011)[a]	Hsieh et al (2011)[a]	Wolff et al (2011)[a]
AC191534.3_FG003	chr7	23	2397	PEG-DMR	AT1G57820	VARIANT IN METHYLATION 1 (VIM1)	PEG-DMR[b]		MEG[fil]
GRMZM2G064905	chr8	3	62	PEG	AT5G53150	DNAJ heat shock domain	PEG		MEG[fil]
GRMZM2G127160	chr6	1354	84630	PEG	AT1G70560	tryptophan aminotransferase of Arabidopsis 1 (TAA1)	PEG-DMR		PEG
GRMZM2G365731	chr1	2	396	PEG	AT4G11400	ARID/BRIGHT DNA-binding domain	DMR		PEG
GRMZM2G014119	chr6	6869	10	MEG-DMR	AT5G03240	polyubiquitin 3 (UBQ3)	MEG	MEG[fil]	MEG[fil]
GRMZM2G044440	chr8	2795	9	MEG-DMR	AT5G03240	polyubiquitin 3 (UBQ3)	MEG	MEG[fil]	MEG[fil]
GRMZM2G103247	chr9	98	2	MEG-DMR	AT2G28890	poltergeist like 4 (PLL4)	MEG-DMR		MEG[fil]
GRMZM2G112925	chr1	492	49	MEG	AT2G28890	poltergeist like 4 (PLL4)	MEG-DMR		MEG[fil]
GRMZM2G370991	chr5	15027	22	MEG-DMR	AT5G03280	ETHYLENE INSENSITIVE 2 (EIN2)	MEG-DMR	MEG	MEG[fil]
GRMZM2G088020	chr8	106	0	MEG	AT5G53870	early nodulin-like protein 1 (ENODL1)	MEG[fil]	MEG	MEG[fil]

[a]The genes listed as MEG or PEG passed all filters in the cited reference. Genes listed as MEG[fil] were filtered due to potential seed coat expression or because they did not pass stricter statistical cut-offs for imprinting in these studies but may actually be imprinted.
[b]The presence of a differentially methylated region (DMR) from Gehring et al (2009) is indicated.

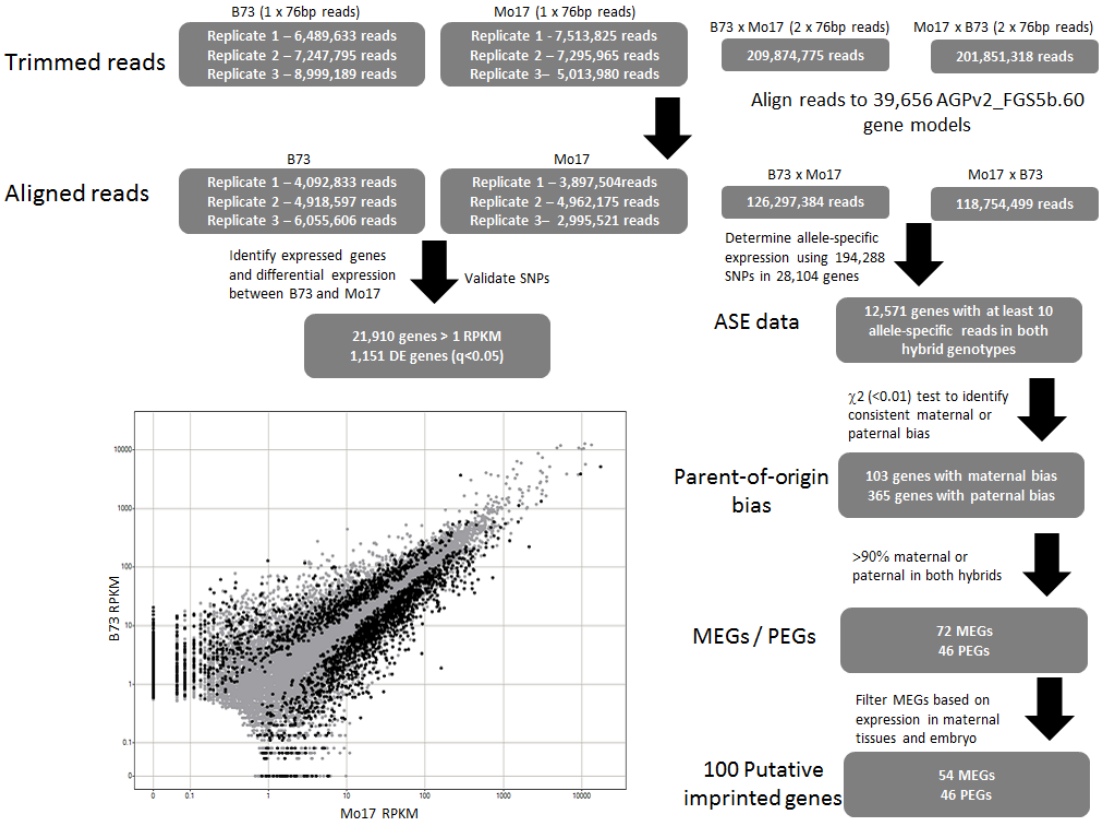


Figure 1. Flow-chart for analyses of endosperm RNA-seq data. The number of trimmed reads is shown for each of the sequenced samples. These reads were aligned to the AGPv2_FGS5b gene models. The data from the inbreds was used to confirm SNPs and to evaluate differential expression in B73 relative to Mo17. The graph shows the RPKM values for the two genotypes and the black data points indicate genes with significantly (q<0.05) different expression in the two genotypes. The RNA-seq data from the hybrids was used for allele-specific expression analyses in order to identify imprinted genes

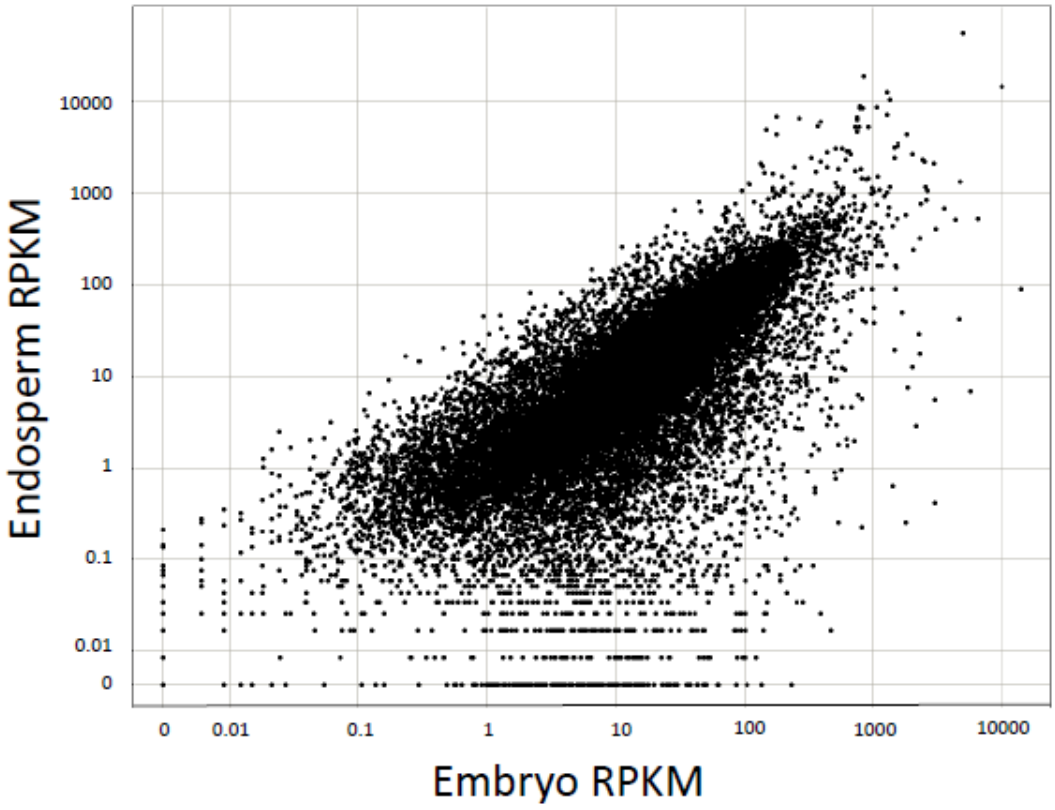


Figure 2. Expression levels in embryo and endosperm tissue. The reads per kilobase per million reads were determined for the Zm5b.60 filtered gene set for both embryo and endosperm tissue. There are a number of genes that exhibit embryo-specific gene expression (plot along the x-axis). However, we did not find examples of genes that exhibit high endosperm expression levels but not embryo expression.

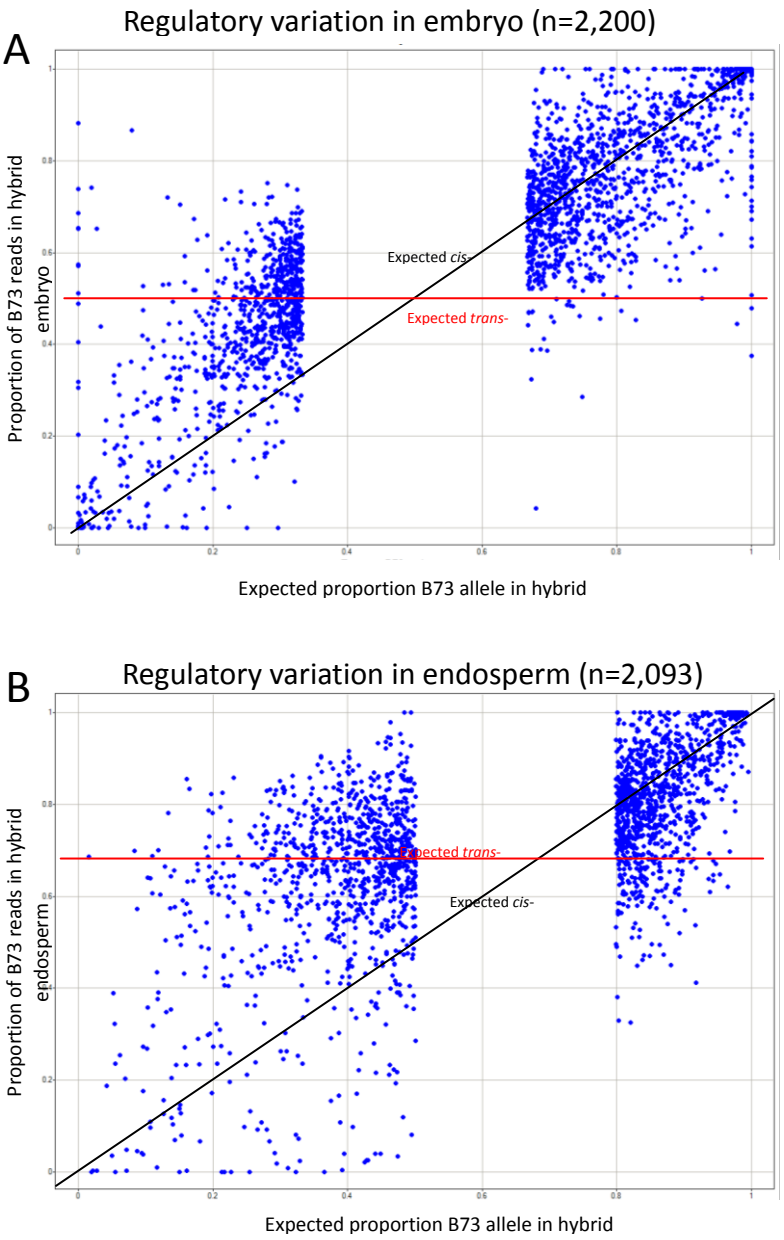


Figure 3. Allele-specific expression indicates that the majority of genes with >2-fold expression changes are due to cis-regulation. The RPKM values from the three replicates of B73 and Mo17 embryo (A) and endosperm (B) tissue were used to identify genes with at least 2-fold difference in expression between the parental genotypes and at least 1RPKM in one of the two parents. For each of these gene we calculated the expected proportion of the B73 allele if the two alleles were *cis*-regulated (for embryo = [B73 / (B73 +Mo17)] and for B73xMo17 endosperm = [(2 x B73) / ((2 x B73) + Mo17)] and this value was plotted on the x-axis. The proportion of allele-specific reads from the B73 allele in hybrid samples (both BxM and MxB for embryo and only BxM for endosperm) was plotted on the y-axis for all genes with at least 50 allele-specific reads. If genes are differentially expressed due to *cis*-regulatory variation they are expected to plot near the diagonal expected line. In contrast, *trans*-regulatory variation will result in genes that plot along the horizontal expected line (note that the 2:1 dosage in endosperm means that the expected trans line is at 0.67 rather than 0.5).

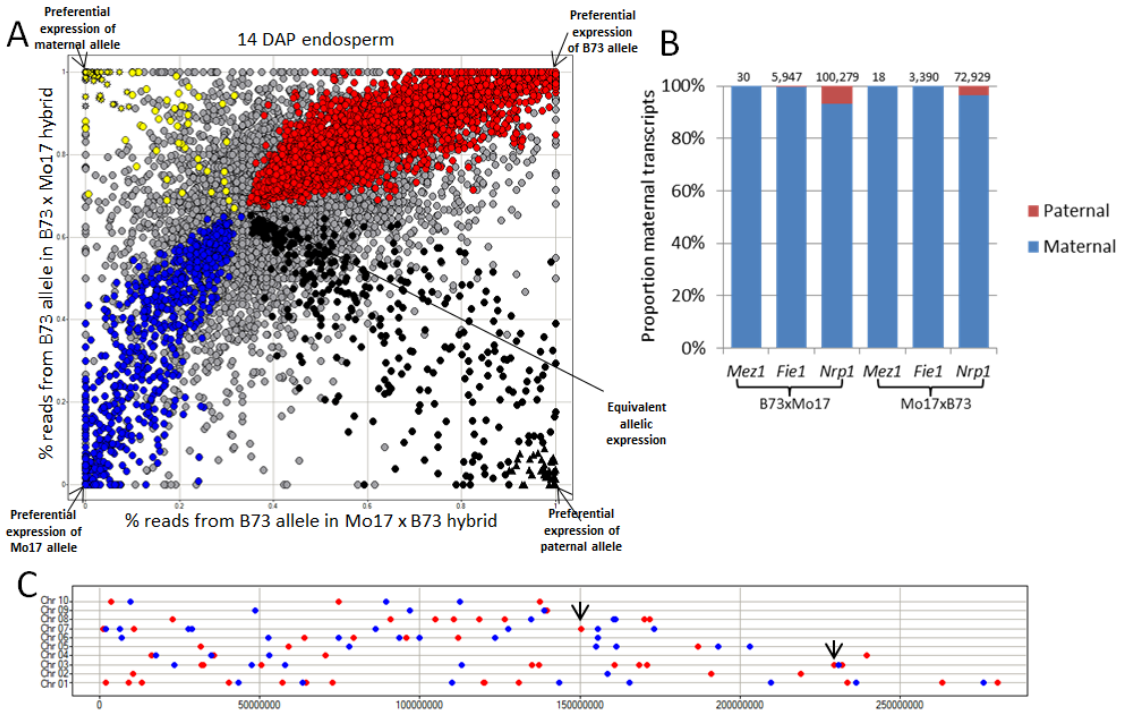


Figure 4. Identification of imprinted genes in maize endosperm. (A) The relative proportion of transcripts derived from the B73 allele was determined for both reciprocal hybrids (limited to only genes with at least 10 reads assigned to one allele in both hybrids). Genes indicated with a color are those that exhibit a significant (chi-square < 0.01) bias from the expected 2:1 ratio in both hybrids. Red coloration indicates genes that are biased towards higher expression of the B73 allele while blue indicates genes that are biased towards higher expression of the Mo17 allele. The yellow and black spots represent genes with maternal or paternal bias, respectively. The triangles represent the 46 genes that have at least 90% of the transcripts derived from the paternal allele while the star symbols represent the 72 genes with at least 90% of the transcripts derived from the maternal allele. B) The proportion of maternal (blue) and paternal (red) transcripts are shown for three genes previously identified imprinted genes in both hybrids. The total number of allele-specific reads for each gene in each sample is indicated above the bar. C) Chromosomal distribution of imprinted genes. The chromosomal positions are shown for each of the 100 imprinted genes (red spots indicate maternal expression while blue spots indicate paternal expression). The black arrows indicate two loci where adjacent genes are both expressed from the maternal allele.

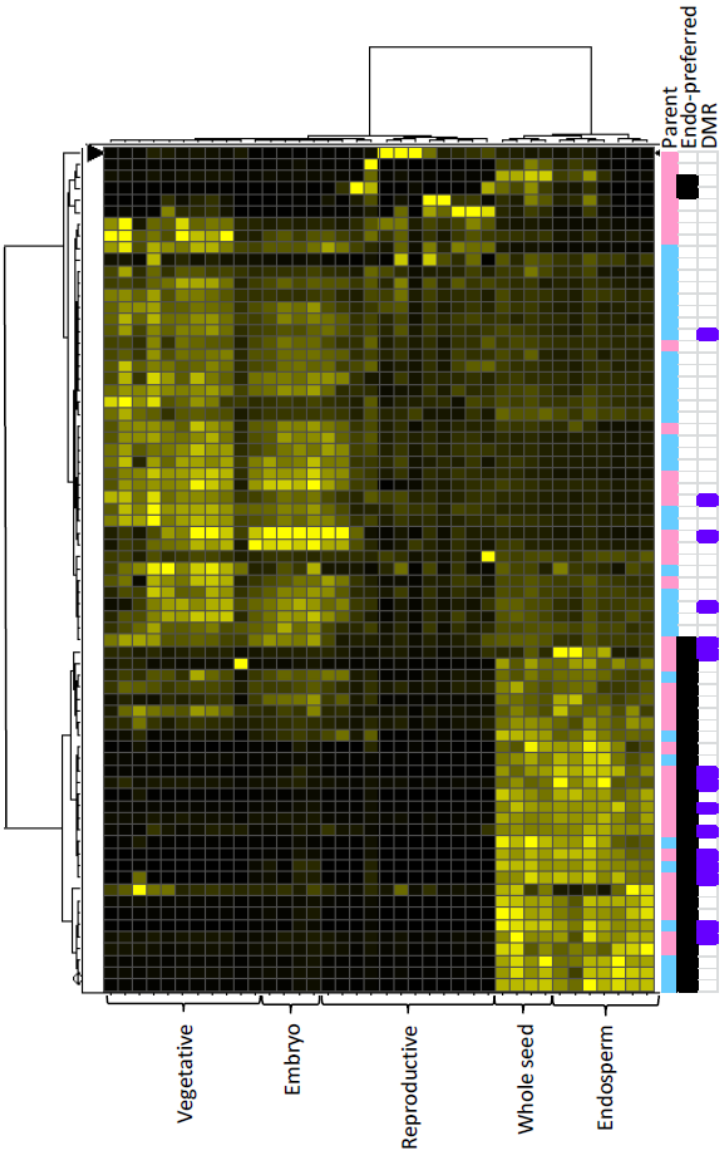


Figure 5. Developmental expression pattern of imprinted genes. Expression atlas data (Sekhon et al., 2011) was available for 71 of the 100 imprinted genes. The expression atlas data was normalized such that each gene had an average expression level of 1 across all tissues. These patterns of expression were then assessed by hierarchical clustering using Ward's method. The resulting heat map illustrates expression of each gene across the different tissues with red indicating higher expression (values above 2), green indicating low expression (values near 0) and black indicates moderate expression (values near 1). The three columns to the right of the dendrogram illustrate features of each imprinted gene. The first column indicates whether the gene is a MEG (pink) or a PEG (blue). The second column indicates which genes were classified as having endosperm-preferred expression (black). The final column indicates the genes that had a differentially methylated region (DMR) in purple.

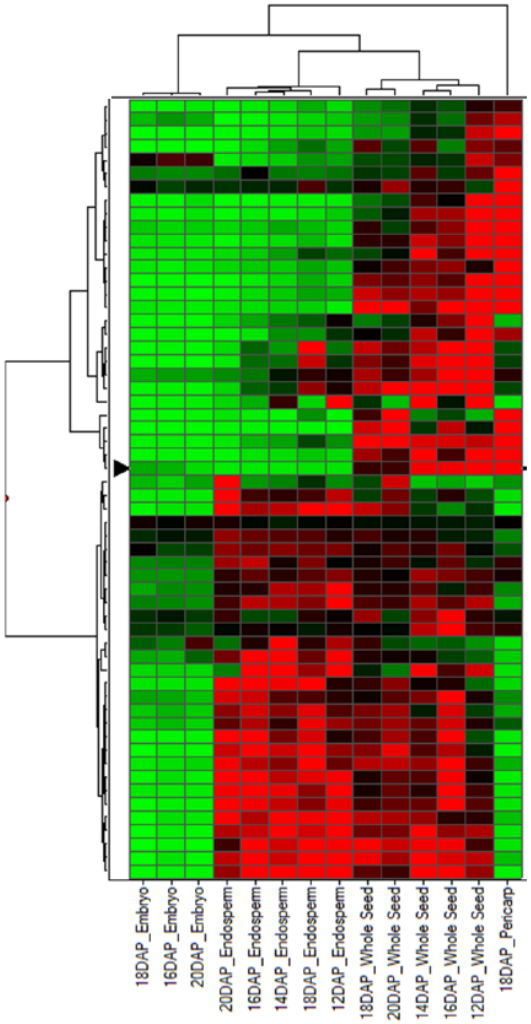


Figure 6. Relative expression levels in different seed tissues for MEGs. Expression atlas data (Sekhon et al., 2011) was available for 58 of the 72 MEGs that were originally identified. The expression levels for each gene were normalized across all seed tissues analyzed (to an average value of 1) and then used for hierarchical clustering (Ward's method) to assess the pattern of expression in different seed tissues. Red color indicates expression level of 2 or higher while green values indicate normalized expression levels of near 0. Black indicates average expression (values near 1). While some genes are much higher in endosperm than other tissues there are a subset of 18 of the 58 genes that exhibit higher at least 5-fold higher expression in whole seed or pericarp relative to endosperm. These genes were omitted from further analysis due to the potential that apparent parent-of-origin effects may be the result of contamination of maternally derived seed coat tissues.

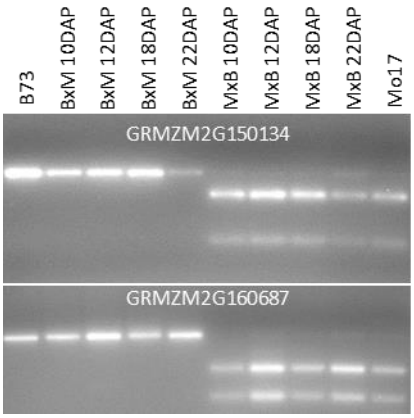


Figure 7. Persistent imprinting throughout endosperm development. For several genes we developed CAPS assays to assess allele-specific expression at multiple stages of endosperm development including 10, 12, 18 and 22 days after pollination (DAP). The outer lanes provide the expected patterns for the B73 and Mo17 alleles (data is from 14 DAP B73 and Mo17 tissue). The inner lanes show the digests of products derived from reciprocal hybrids. The BxM samples represent hybrids in which B73 was the maternal parent while the MxB samples had Mo17 as the maternal parent.

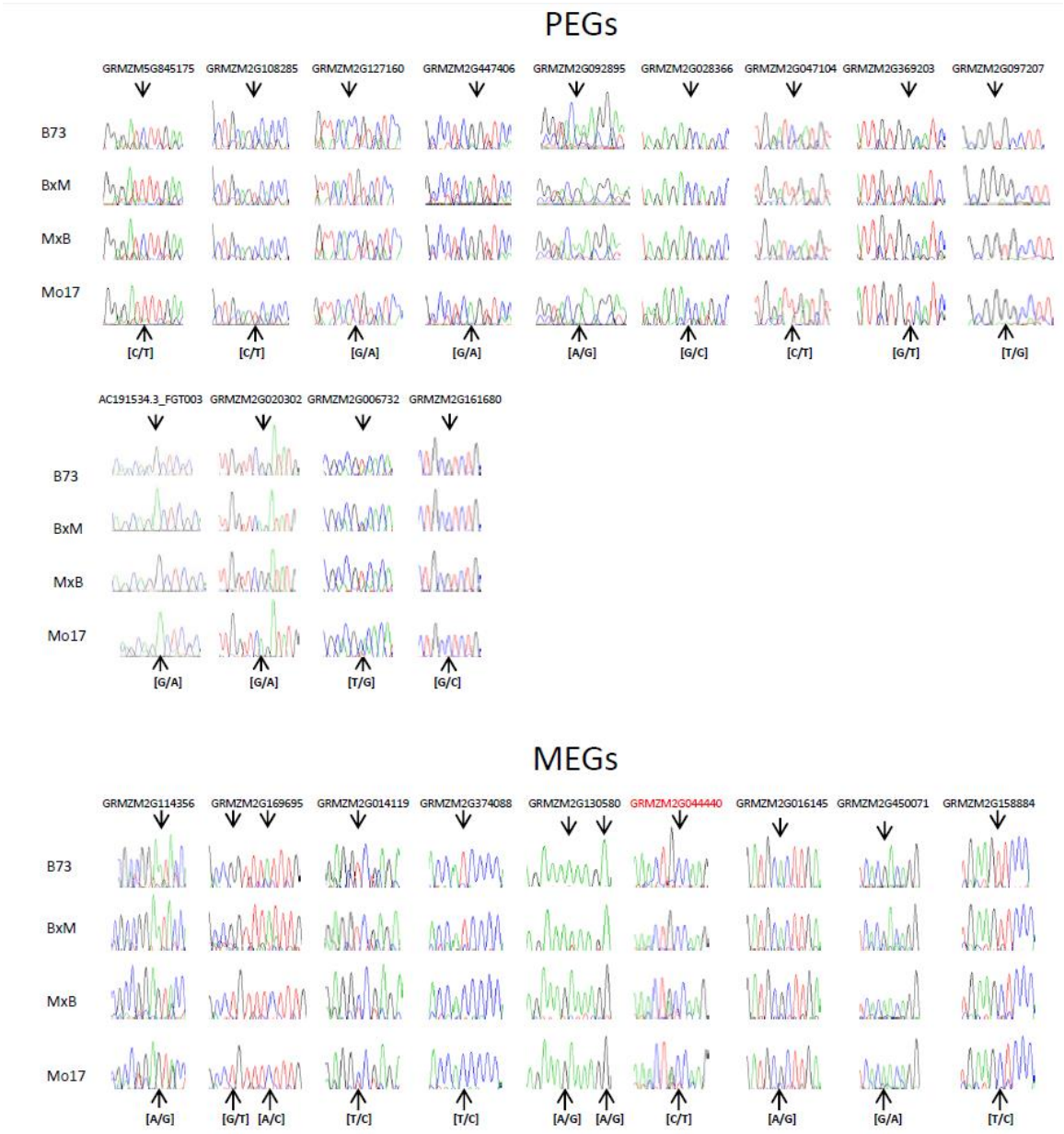


Figure 8. Validation of imprinting. RT-PCR was performed on a randomly selected set of 22 of the novel imprinted genes to amplify a region that would contain at least one SNP from cDNA from an independent biological replicate of 14 DAP endosperm tissue. The PCR amplicon obtained from the two parents and both reciprocal hybrids was then used for sequencing. In each plot the position of the SNP is indicated. The monoalleleic expression in a maternal or paternal fashion was confirmed for 21 of these 22 genes. We did not observe the expected pattern for one gene (GRMZM2G044440 – highlighted in red).

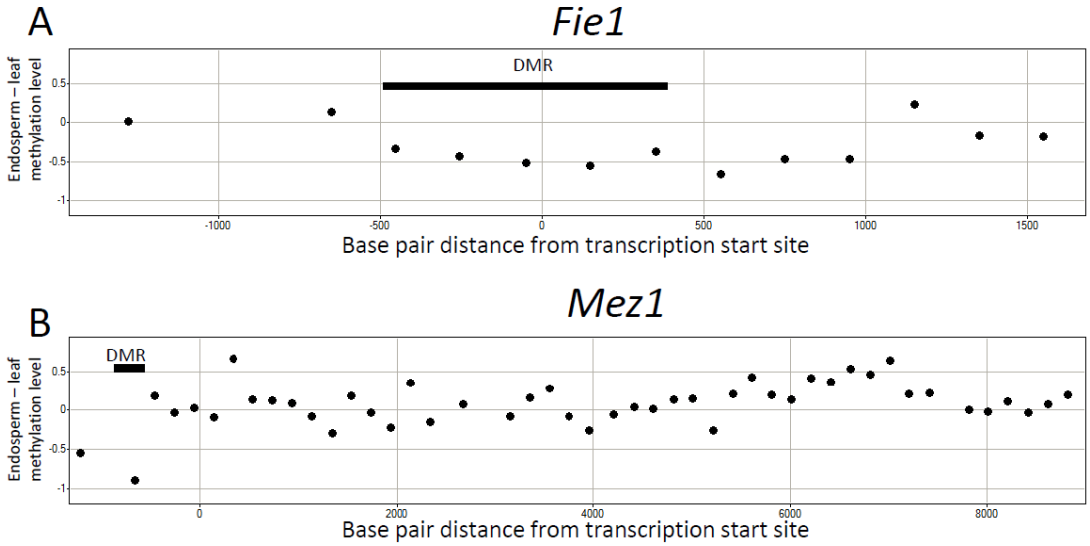


Figure 9. Identification of regions of reduced methylation at previously characterized differentially methylated regions (DMRs). The plots show the relative level of DNA methylation in endosperm compared to leaf tissue. Each data point represents the average value for a single probe in replicates of B73 and Mo17. The black bars indicate the position of previously identified DMRs (for maize *Fie1* - Gutierrez-Marcos et al., 2006 and Hermon et al., 2007 and for *Mez1* – Haun et al., 2007). The y-axis represents the differences in methylation levels detected by meDIP-chip for endosperm and leaf tissue. A value of zero indicates no differences in methylation levels while negative values represent lower methylation levels in endosperm relative to leaf.

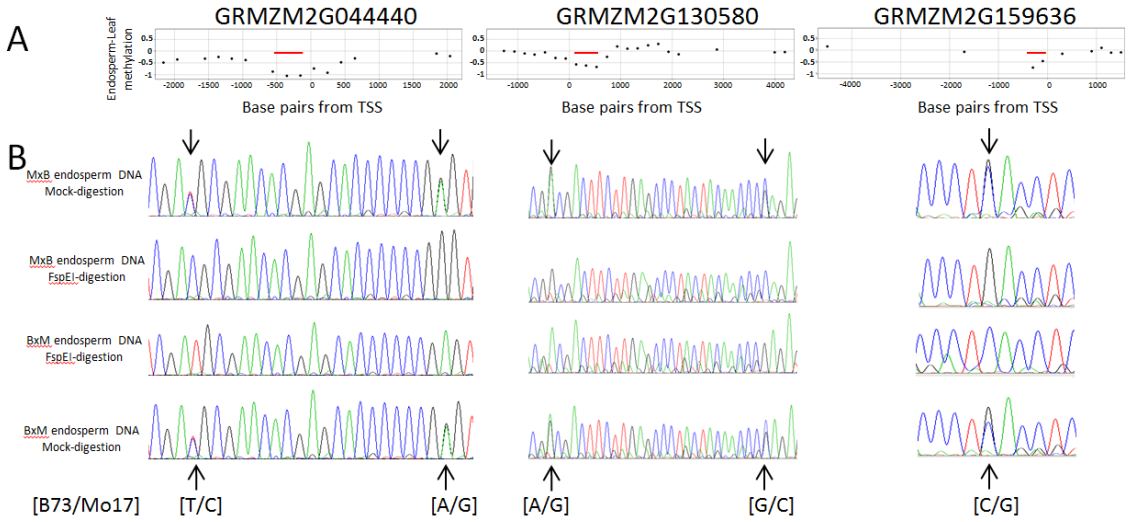


Figure 10. Identification of differentially methylated regions (DMRs) at imprinted genes. (A) The panels display array based methylation profiling data for three genes. The y-axis represents the relative differences in methylation levels in endosperm and leaf tissue (negative values indicate lower methylation in endosperm). The x-axis indicates the base pair distance of the probe from the transcription start site (TSS). Each data point represents a single probe and the values are determined from the average of three replicates of B73 and Mo17 tissues. The red lines indicate regions that were identified as hypomethylated in endosperm. (B) For each of these three genes we identified SNPs within the hypomethylated region (listed under traces in [B73/Mo17] format). The SNP-containing regions were amplified and sequence in both reciprocal hybrids. The sequencing plots from the undigested DNA from the reciprocal hybrids (first and last traces) reveal presence of both alleles. However, following digestion with the methylation-dependent restriction enzyme FspEI (middle traces) only the maternal allele is detected confirming specific hypomethylation of the maternal allele.

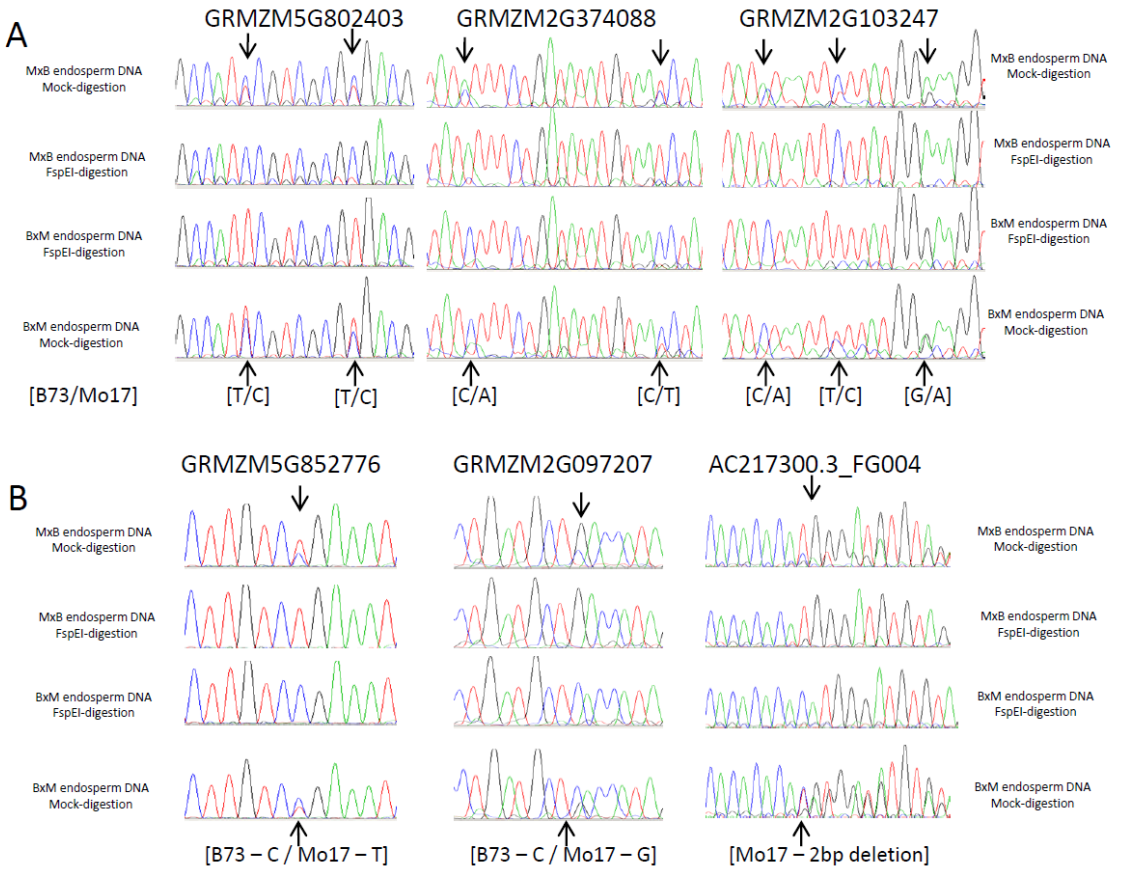


Figure 11. Hypomethylation of the maternal allele at DMRs. SNPs were identified within regions of endosperm hypomethylation near six additional imprinted genes including three with maternal expression (A) and three with paternal expression (B). The SNPs (or a 2 base pair deletion in one case) are indicated below the sequence traces. PCR amplification was performed on endosperm DNA (from either B73 x Mo17 or Mo17x B73 tissue) for undigested DNA (top and bottom traces). These traces show the presence of both alleles with higher levels of the maternal allele due to the 2:1 dosage. Following digestion with FspEI, a restriction enzyme that only cleaves methylated DNA) only the maternal allele is detected (middle traces). This provides evidence for loss of methylation for the maternal allele in endosperm tissue.

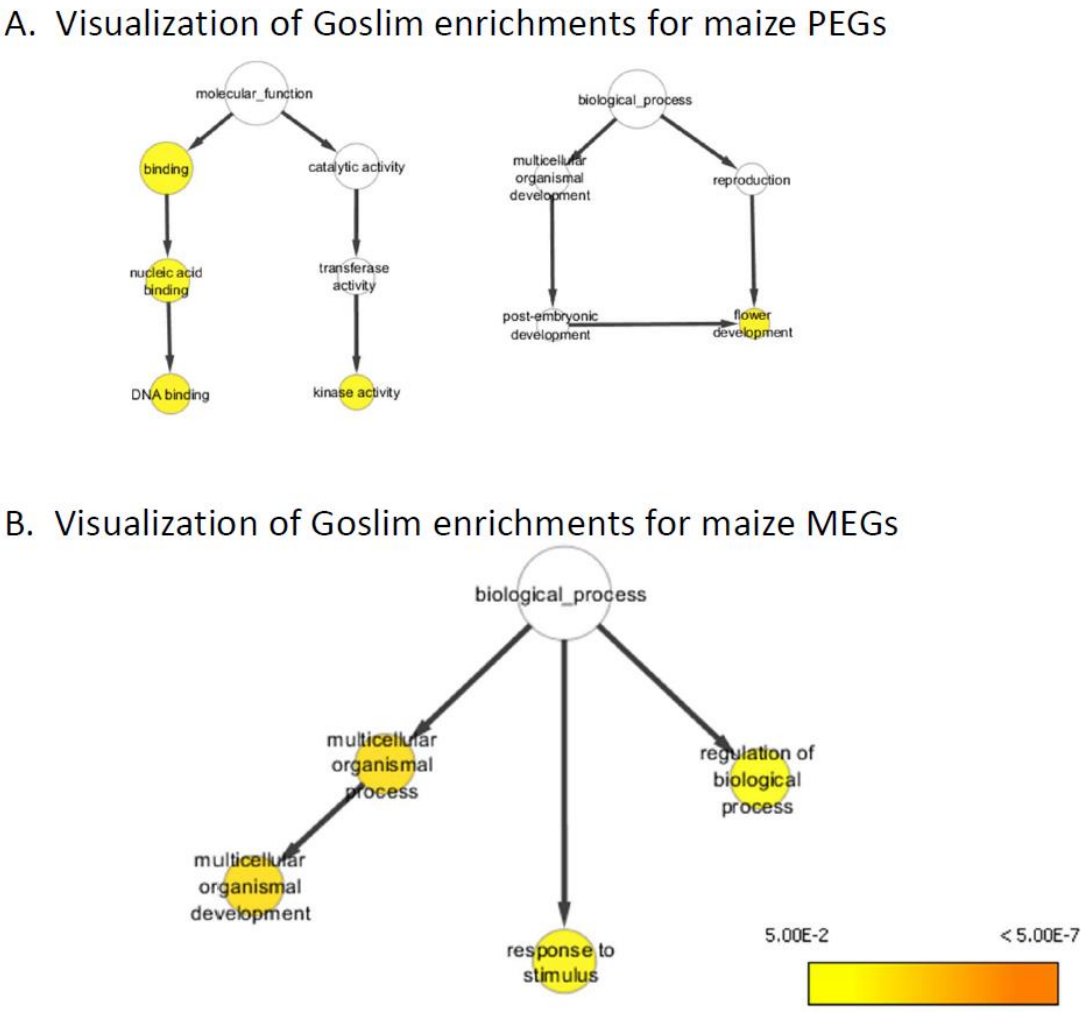


Figure 12. Annotation enrichments for MEG and PEGs. Each of the maize genes was used as a query to identify related Arabidopsis genes. We identified a related Arabidopsis gene for 40 of the PEGs and 40 of the MEGs. These Arabidopsis genes were used to search for enrichments of GO terms using the GO slim annotations in BinGO (Maere et al., 2005). The resulting enrichments are visualized for the PEGs (A) and MEGs (B). The increasing shading colors represent higher significance for the over-representation of any particular classification term.

Chapter III: **Context Statement**

SUMMARY

In plants, a subset of genes exhibit imprinting in endosperm tissue such that expression is primarily from the maternal or paternal allele. Imprinting may arise as a consequence of mechanisms for silencing of transposons during reproduction, and in some cases imprinted expression of particular genes may provide a selective advantage such that it is conserved across species. Separate mechanisms for the origin of imprinted expression patterns and maintenance of these patterns may result in substantial variation in the targets of imprinting in different species. Here we present deep sequencing of RNAs isolated from reciprocal crosses of four diverse maize genotypes, providing a comprehensive analysis that allows evaluation of imprinting at more than 95% of endosperm-expressed genes. We find that over 500 genes exhibit statistically significant parent-of-origin effects in maize endosperm tissue, but focused our analyses on a subset of these genes that had >90% expression from the maternal allele (69 genes) or from the paternal allele (108 genes) in at least one reciprocal cross. Over 10% of imprinted genes show evidence of allelic variation for imprinting. A comparison of imprinting in maize and rice reveals that 13% of genes with syntenic orthologs in both species exhibit conserved imprinting. Together, these data suggest that imprinting only has functional relevance at a subset of loci that currently exhibit imprinting in maize.

Chapter III entitled 'Comprehensive analysis of imprinting in maize' has been adapted from my work in the publication:

Waters, A. J., Bilinski, P., Eichten, S. R., Vaughn, M. W., Ross-Ibarra, J., Gehring, M., & Springer, N. M. 2013). Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proceedings of the National Academy of Sciences of the United States of America, 110(48), 19639–44. http://doi.org/10.1073/pnas.1309182110

During the course of this experiment the other researchers contributed to data management and analysis. Mary Gehring provided assistance with experimental design, analysis, and finical support. I collected and prepared the samples for RNA sequencing, analyzed the expression data, and completed the conservation of imprinting within the maize samples and between my samples and *Arabidopsis thaliana* and rice. Mathew Vaughn and Steve Eichten provided computational resources and support. Paul Bilinski and Jeffrey Ross-Ibarra completed the population genetic analysis. Creation of figures and text for the manuscript was developed by Nathan Springer and myself.

Portions of text, tables, and figures have been omitted from the original manuscript for this chapter to focus on my contributions and analyses. All omitted data can be accessed through the published manuscript. Additionally, I have removed author contact information and acknowledgments as well as formatted references to be consistent throughout my thesis.

CHAPTER III

Comprehensive analysis of imprinting in maize

INTRODUCTION

Imprinting describes a biased expression of alleles that depends upon the parent of origin. Imprinting is observed in both flowering plants and mammals (Bartolomei and Ferguson-Smith, 2011; Gutierrez-Marcos et al., 2012; Pignatta and Gehring, 2012) but there are differences in the mechanisms and organization of imprinted genes in these organisms (Bartolomei and Ferguson-Smith, 2011 and Das et al., 2009). In plants, imprinting is most prevalent in the endosperm, a triploid tissue that contains two maternal genomes and a single paternal genome (Li and Berger, 2012). The endosperm provides an energy source for germinating seeds and, as the majority of harvested grain consists of endosperm tissue, a major source of calories in the human diet. A better understanding of imprinting will shed further light on epigenetic gene regulation and endosperm development and could provide an avenue for altering plant reproductive processes or seed quality.

Despite a widespread interest in imprinting and its potential importance, the function of most imprinted genes is not well-characterized in plants and imprinting has only recently been assayed on a genome-wide level. Imprinting is reflected in parentally biased allele-specific expression in the endosperm tissue of intraspecific reciprocal hybrids. A quantitative method for detecting the relative expression of two alleles that have nearly identical sequences is required to find such an effect, traditionally limiting analysis to a handful of imprinted genes identified based on phenotype or through targeted analyses (Gehring et al., 2009; Kermicle 1970; Kohler et al., 2012). The implementation of deep sequencing of RNA molecules (RNAseq) has allowed detection of additional imprinted genes (Hsieh et al., 2011; Gehring et al., 2011; Luo et al., 2011; Waters et al., 2011; Wolff et al., 2011; Zhang et al., 2011). In each of these studies, allele-specific expression levels were monitored for a single cross of two parents in *Arabidopsis*, maize, or rice. This allowed for the analysis of imprinting in 50-58% of

genes expressed in endosperm tissue. In each species there is evidence for several hundred imprinted genes with similar numbers of maternally expressed genes (MEGs) and paternally expressed genes (PEGs), but comparisons among flowering plants (Kohler et al., 2012; Luo et al., 2011; Waters et al., 2011; Jiang and Kohler, 2012) have revealed limited overlap in the genes that are imprinted among species.

There has been considerable speculation on the mechanisms that might lead to the origin of imprinted expression as well as the evolutionary mechanisms that would lead to the maintenance of imprinting (Kohler et al., 2012; Haig, 2004; Kohler and Weinhofer-Molisch; Berger et al, 2012). Recent studies suggest that imprinting may arise due to programmed release of silencing marks in specific nuclei of the male and female gametophytes (Feng et al., 2010; Bauer and Fischer, 2011). Plant gametophytes are multi-nucleate structures. The male gametophyte includes a vegetative nucleus and two sperm nuclei. The female gametophyte has multiple cells including the haploid egg cell (which is fertilized by a sperm nuclei to generate the embryo) and the diploid central cell (which is fertilized by a sperm cell to generate the endosperm) (Huh et al., 2008). The loss of DNA methylation before fertilization leads to an epigenetic asymmetry in the endosperm because the maternal genomes (from the central cell) have been demethylated while the paternal genome (from a sperm nucleus) retains normal levels of methylation. Programmed DNA demethylation might result in the generation of siRNAs that could reinforce transposon silencing in adjacent cell types (egg and sperm cells) that contribute genetic material to the next generation (Rodrigues et al., 2013). It has been hypothesized that this process, while targeted to transposons, could inadvertently influence nearby genes, resulting in imprinted expression (Kohler and Weinhofer-Molisch). In support of this idea, several well-characterized imprinted genes contain transposon sequences in adjacent regions (Gehring et al., 2009; Zemach et al., 2010; Kinoshita et al., 2007; Lippman et al., 2004). The potential for transposons to contribute to the origin of imprinted expression of nearby genes may result in examples of imprinting that do not provide a selective advantage and would not be expected to persist over evolutionary time. Because imprinting at such loci would be of limited functional relevance and

dependent on the presence of a transposable element, this could lead to substantial allelic variation for imprinting within a species. Indeed, several of the first characterized examples of imprinting in maize exhibit allelic variation such that certain alleles are imprinted while others are not (Kermicle and Alleman, 1990; Chaudhuri and Messing, 1994).

Regardless of the mechanisms that give rise to imprinted expression, parent-of-origin expression could in some instances provide a selective advantage. The kinship theory (Haig 2004) suggests that maternally expressed genes should restrict growth or limit the flow of resources to offspring while paternally expressed genes might function to promote offspring growth. There are examples of imprinted genes that appear to exhibit these functions (Costa et al., 2012), but there is no clear evidence for these predicted functions in the annotations of the full set of previously identified MEGs or PEGs (Pignatta and Gehring, 2012). Genes that are subject to parental conflict might be expected to exhibit signatures of positive selection (Berger et al., 2012; Hutter et al., 2010). For some imprinted genes, such as the *Arabidopsis* locus MEDEA, potential evidence of positive selection has been found in some cases (Spillane et al., 2007), but not others (Haun et al., 2007).

The presence of imprinting for a particular gene is often assumed to have functional relevance. While this may be the case for a subset of genes, the potential for inadvertent acquisition of imprinting as a result of nearby transposon influences could result in numerous examples of imprinting that have limited functional relevance and thus show intra- or inter-specific variation in imprinting. To distinguish between these possibilities and evaluate the functional importance of imprinting, we analyzed imprinting in multiple diverse genotypes of maize. Reciprocal crosses among four genotypes provided the ability to survey imprinting at over 95% of the genes expressed in endosperm tissue. We find that only a subset of imprinted genes shows conserved imprinting in maize and rice and that these genes show evidence of distinct selective pressures. Comparison of imprinting in different haplotypes within maize reveals allelic

variation for imprinting, further suggesting that imprinting may have limited functional consequence for many maize genes.

RESULTS

Deep sequencing of RNA isolated from 14 day after pollination (DAP) endosperm tissue of five reciprocal hybrid pairs was performed to identify imprinted genes. This intermediate stage of endosperm development was selected because it is before major starch accumulation but after endosperm cellularization and because it reduces the potential for observation of stable transcripts contributed by the gametes. However, the analysis of one stage of endosperm development does not allow for assessment of transiently imprinted genes (Springer and Gutierrez-Marcos, 2008) or imprinting in embryo tissue (Jahnke and Scholten, 2009). The five reciprocal hybrids we assayed included one previously analyzed dataset for the cross of inbred lines B73 x Mo17 (Waters et al., 2011) as well as four new reciprocal hybrids generated by crossing inbred lines Ki11 and Oh43 with both B73 and Mo17 (Table 1). These additional genotypes were selected because whole-genome resequencing provided detailed SNP information (Chia et al., 2012) and because they represent diverse genotypes (Yu et al., 2008).

A large number of reads (180-210 million) were recovered for each of the 10 genotypes and were analyzed to study gene and allelic expression patterns (see Methods, Figure 1 for details). The number of reads that mapped to each allele was summed across all SNPs for a transcript. Only transcripts that had at least 10 reads that could be assigned to a particular allele in each direction of the reciprocal cross were analyzed, resulting in allelic expression data for between 5,851 and 13,478 genes in each cross (Table 2: Figure 1). In total 18,284 genes (95% of genes expressed in 14 DAP endosperm) had allele-specific expression data in at least one of the five reciprocal hybrid pairs (Table 2; Figures 2A and 3). Maternally expressed imprinted genes (MEGs) preferentially express the maternal allele in both directions of a cross while paternally expressed imprinted genes (PEGs) express low levels of the maternal allele in both directions of the cross.

Genes that exhibit consistent bias for the allele from one genotype, independent of parent of origin, reflect cis-regulatory allelic variation.

Comprehensive discovery of maize imprinted genes

A combination of statistical significance and proportion filters was implemented to identify and classify MEGs and PEGs (Figures 1 and 2). We assigned different levels of imprinting to parentally biased genes to compare imprinting strength within and between species in a more nuanced manner. Moderate MEGs / PEGs were defined as having significant allelic bias (χ^2<0.05) and >80% of transcripts from the maternal allele (MEGs) or >60% of the transcripts from the paternal allele (PEGs) (red shaded areas in Figure 2A and 3) in both directions of a reciprocal cross. Strong MEGs and PEGs were defined as having significant allelic bias (χ^2<0.01) and >90% of transcripts from the maternal allele (MEGs) or paternal allele (PEGs) (blue area in figure 2A and 3). Complete MEGs or PEGs have >99% of the transcripts derived from the maternal or paternal allele, respectively. Genes with strong allelic bias (at least 95% reads from one allele) in one direction of the cross but not in the reciprocal hybrid are potentially represent allelic variation for imprinting (green box in Figure 2A and 3).

The number of genes classified as moderate, strong or complete MEGs and PEGs varied for each genotype (Table 1) in large part due to differences in the number of genes with polymorphisms. For the subsequent analyses, only the genes that were classified as strong or complete MEGs/PEGs were used. There are a total of 108 non-redundant strong PEGs, including 31 (28%) examples that were classified as complete PEGs in at least one genotype (Table 1). Additional filtering criteria were applied to MEGs to remove genes that might exhibit maternal bias due to contamination of maternally derived tissues. RNA-seq data from a B73 expression atlas (Sekhon et al., 2013) was used to identify MEGs that may be the result of maternal contamination, resulting in a filtered list of 69 non-redundant strong MEGs, with a larger number (Sekhon et al., 2013; 54%) showing complete imprinting than seen in PEGs (Table 1).

Quantitative SNP assays designed using the Sequenom MassArray platform were used to validate imprinting for 13 MEGs and 13 PEGs (Table 3). These assays are based on a single SNP for each gene and could only be used to assess imprinting in the crosses that were polymorphic for the targeted SNP. The analysis of allele-specific expression in a different 14 DAP endosperm sample for the same set of five reciprocal crosses confirmed imprinting in the majority of samples for both MEGs (23/24) and PEGs (28/28). The one allele that was not validated showed imprinted expression in one direction of the cross but bi-allelic expression in the reciprocal hybrid. The same quantitative SNP assays were also used to assess whether imprinting for these genes was also detected in several other genotypes (NC358, Ms71, and M162W) that were reciprocally crossed with B73 and Mo17. Most of these genes were imprinted in each of the other genotypes that were tested, with the exception of one locus (GRMZM2G020302) (Table 3). Finally, the quantitative SNP assays were also used to assess whether imprinted expression was maintained at earlier and later stages of endosperm development. Imprinting was consistently observed for 26/26 MEGs and 25/26 PEGs at 12 DAP, 14 DAP, 16 DAP, and 20 DAP samples of B73xMo17, B73xNC358 and Mo17xNC358 (Table 3). These data confirm that our RNA-seq data and subsequent imprinting analysis pipeline are highly reproducible.

The expression of imprinted genes is endosperm specific

Several plant imprinted genes have expression that is restricted to the endosperm (Berger et al., 2012). This endosperm-specific expression could be because these genes have specific functions in the endosperm, or because it is beneficial to silence these genes in somatic tissues. Only a subset of MEGs and PEGs exhibit preferential expression in endosperm relative to other tissues in maize (Figure 4 A-B). The majority of MEGs (68%) are preferentially expressed in endosperm while only 26% of PEGs are preferentially expressed in endosperm (Figure 4 A-B). Many MEGs exhibit increasing levels of expression during endosperm development, suggesting that these genes are actively transcribed in endosperm tissue as opposed to being stable, maternally inherited

transcripts. There was no evidence that MEGs or PEGs exhibit unusually high or low expression levels, instead MEGs and PEGs exhibit a range of expression levels in endosperm tissue (Figure 5 A).

PEGs are often targets of H3K27 methylation

We previously (Makarevitch et al., 2013) documented genome-wide H3K27me3 levels for five tissues of maize, including endosperm. We assessed the presence of H3K27me3 for the MEGs and PEGs identified in this study. Consistent with previous work (Wolff et al., 2011; Makarevitch et al., 2013), we find that PEGs are more likely to be targets for histone methylation than MEGs. Only 5 of 69 of the MEGs exhibit H3K27me3 in endosperm tissue, in contrast to 87 of the 108 PEGs. The 87 PEGs that are marked with H3K27me3 in endosperm tissue include 64 genes with expression in vegetative tissues and 23 genes with preferential expression in endosperm. Only 8% of the 64 PEGs that are expressed in vegetative tissues exhibit H3K27me3 in the four vegetative tissues analyzed, while 65% of the 23 PEGs with preferential expression in endosperm are marked by H3K27me3 in at least three of the four vegetative tissues that were analyzed.

PEGs exhibit higher levels of sequence conservation

MEGs and PEGs also differ in their conservation between species and their annotation. The frequency of PEGs with syntenic orthologs in rice (Schnable et al., 2012) was much higher (83%) than MEGs (46%). Similarly, the proportion of PEGs with high sequence similarity (E<1E-50) to an Arabidopsis gene (61%) was higher than the proportion of MEGs (36%). Overrepresentation of functional categories of GO annotations were investigated for the 66 PEGs and 23 MEGs that had high sequence similarity to Arabidopsis (E-score <1E-50) using BinGO (Maere et al., 2005). PEGs exhibit significant ($p<0.05$) enrichment for GO terms including flower development, chromatin modification, regulation of biological process and DNA binding. MEGs

exhibit significant ($p<0.05$) enrichment for terms including DNA binding, transcription factor activity, developmental process and responses to stimulus.

Allelic variation for imprinting

Several of the earliest examples of imprinted loci exhibited imprinting for alleles from some genotypes but not others (Kermicle and Alleman, 1990; Chaudhuri and Messing). Our analysis of multiple maize genotypes provides an opportunity to comprehensively assess allelic variation in imprinting (Figure 6A). In general, when data were available for multiple crosses, many (88%) genes that exhibit imprinting in one cross were also imprinted in the other crosses, but there are examples in which genes imprinted in one cross display allelic variation for imprinting in another cross (Figure 6A-B). We identified 17 genes (8 PEGs and 9 MEGs) that showed consistent patterns of allelic variation in imprinting (Figure 6B and 7). Gene GRMZM2G384780, for example, shows complete maternal allele expression in the Mo17/Oh43 cross, the B73 allele is not silent when it is inherited paternally (Figure 6C). Similar variation was observed for other MEGs (Figure 5B) and PEGs (Figure 6D and 5C). The PEG GRMZM2G106222 shows expression of the maternal allele only when Oh43 is the maternal parent (Figure 6D). A quantitative SNP assay was used to confirm the allele-specific imprinting for this gene (Figure 6D). Overall, these data suggest standing allelic variation for imprinting is about 12% (17/144, the total is the number of genes with data in at least two sets of reciprocal crosses) of the imprinted genes even though we only assayed at most four haplotypes for each locus.

Conservation of imprinting between species

If imprinting plays a similar functional role in all flowering plant species regardless of differences in endosperm growth or development, then it might be expected that there would be strong conservation for the targets of imprinting. Previous work has found only 5-10 examples of conserved imprinting between species (Pignatta and Gehring, 2012; Luo et al., 2011; Waters et al., 2011), but has had limited comparative

power due to the use of only a single cross in which not all genes may show polymorphism. The availability of a comprehensive list of MEGs and PEGs analyzed in multiple crosses in maize and information on syntenic gene relationships in rice allowed us to investigate the conservation of imprinting in monocots in more detail. There are 58 maize PEGs and 27 maize MEGs that have syntenic orthologs in rice that were assessed for imprinting by Luo et al. 2011 in a cross between two haplotypes. Of these, 9 PEGs and 3 MEGs show imprinting for both the maize and rice syntenic orthologs (Figure 8C) and an additional 2 PEGs and 1 MEG that have imprinting for a closely related rice gene not located at a syntenic genomic position (Figure 8A). This is a relatively low level of conservation but is significantly higher than expected by chance (χ^2, $p<0.001$). There are also 3 moderate MEGs and 8 moderate PEGs that show imprinting of their corresponding syntenic rice gene (Figure 8D), and a low, but statistically significant (χ^2, $p<0.001$) level of conservation for imprinting of related sequences (not necessarily syntenic) in maize and *Arabidopsis* (Figure 8B). Genes with conserved imprinting in maize and rice include a variety of annotations that include many genes with putative roles in transcriptional regulation or signaling pathways (Figure 8C-D). Two of these, encoding an ARID/BRIGHT DNA binding domain protein and a flavin-binding monooxygenase protein, also show imprinting for related sequences in *Arabidopsis*. A recent report (Bernardi et al., 2012) provides evidence that one of these genes with conserved imprinting (GRMZM2G091819) plays an important role in influencing endosperm development in maize. Many of the genes with conserved imprinting have not been analyzed functionally but could play important roles in controlling endosperm development.

Finally, we analyzed the conservation of imprinting between paralogs from the recent whole-genome duplication event in maize. Following an allopolyploid whole-genome duplication event 5-12 million years ago (Gaut and Doebley, 2997) subsequent rearrangements and fractionation have resulted in varying patterns of retention and loss (Schnable et al., 2012; Woodhouse et al., 2010) of syntenic paralogs. A larger proportion of PEGs (73 genes, 68%) than MEGs (31 genes, 45%) are found in one of the two

syntenic blocks assigned to subgenomes (Fisher's exact test two-tailed p-value = 0.005). Of the 7 MEGs with retained duplicates in both subgenomes, two of the duplicates exhibit moderate imprinting, two are not imprinted but are expressed in the endosperm, and three are not expressed in the endosperm. Among the 18 PEGs with retained duplicates, 10 are imprinted, 7 are expressed in the endosperm but not imprinted, and 1 is not expressed in the endosperm.

DISCUSSION

Our analysis of allele-specific expression in multiple crosses of maize is the most comprehensive study of imprinting in any plant species to date. Over 95% of the genes that are expressed in endosperm could be tested for imprinting due to the presence of polymorphisms in at least one of the crosses. Several hundred genes show consistent parent-of-origin effects in at least one of the crosses. The availability of a relatively complete set of imprinted genes for maize provides an opportunity to examine the conservation of imprinting within and between species.

Imprinted genes are often treated as a single class in the literature. However, there are differences between MEGs and PEGs that suggest maternally and paternally biased expression might reflect distinct processes. MEGs are much more likely to exhibit endosperm-specific expression than PEGs, more likely to lack predicted function, and much less frequently associated with H3K27me3. There are also differences in the conservation of MEGs and PEGs between species. The maize PEGs have fewer recent duplications, and are more likely to have a retained a syntenic ortholog in rice and highly similar sequence in Arabidopsis. In addition, there are more examples of conserved imprinting in maize and rice, or between paralogs, for the PEGs.

The initial discovery of imprinting was based on studies of the *R* locus in maize (Kermicle 1970), which exhibits allelic variation for imprinting (Kermicle and Allen, 1990). There are several other examples of potential allelic variation for imprinting in maize (Jahnke and Scholten, 2009), but there have been few studies that assess imprinting for multiple alleles within a species. Our data reveal that over 10% of the

genes with strong imprinting show allelic variation among the four maize haplotypes surveyed. This rate would undoubtedly increase if additional haplotypes were tested: we found at least one example of a gene for which the four alleles tested by RNAseq were all imprinted but the allele in at least one additional genotype tested by a gene-specific assay was not (Table 3). This allelic variation in imprinting may reflect differences in transposon content near maize genes, although we have not yet been able to assess this. Studies of haplotype structure variation in maize (Messing and Dooner, 2006) provide evidence for substantial allelic variation in the type of repetitive elements surrounding genes. Further study of the specific haplotypes present at alleles that vary for imprinting may shed further light on the genetic or epigenetic changes that contribute to imprinted expression.

We investigated the conservation of imprinting between two monocots with persistent endosperm, maize and rice. In total we identified 88 imprinted maize genes with a syntenic rice gene evaluated by Luo et al 2011, but only 12 exhibit conserved imprinting. While higher than expected by chance alone, this limited number suggests that conservation of imprinting over longer periods of evolutionary time is not common. It is important to note, however, that there are likely more than 12 examples of conserved imprinting in maize and rice because a number of loci, such as the rice ortholog of the maize MEG *Mez1* (Haun et al., 2007), could not be tested in rice due to a lack of polymorphisms (Luo et al., 2011). The limited conservation of imprinting among species also may help to guide future functional studies. Genes with conserved imprinting among species might play important functional roles in regulating seed development and growth and would be useful targets for reverse-genetic analysis. Future experiments to test the functional role of these genes in seed development are necessary. In contrast, the genes with imprinting only in certain species may reflect unique reproductive strategies in those species or represent imprinting that is not functionally relevant but is simply the result of inadvertent imprinting due to allelic differences such as transposon or epigenetic variation.

MATERIALS AND METHODS

Plant materials and RNAseq: B73, Mo17, Ki11 and Oh43 plants were grown in Saint Paul at the University of Minnesota Agricultural experiment station during the summer of 2011. Reciprocal crosses and self-pollinations for all genotypes were performed between August 4^{th} – 15^{th} and several ears representing each cross were harvested 14 days after pollination (DAP). The endosperm and embryo tissue were dissected by cutting away the cap of each kernel, removing the endosperm and embryo from the pericarp, and then dissecting apart the embryo from the endosperm tissue. At least two ears for each genotype and were pooled together and frozen in liquid nitrogen. RNA was isolated by SDS-Trizol protocol and subsequently purified by LiCl precipitations. These RNA samples were submitted to the University of Minnesota Biomedical Genomics Center for sequencing using the Illumina HiSeq-2500 platform.

These reads were aligned to the 39,540 genes in the filtered gene set (version 5b.60) using the Tophat aligner (Trapnell et al., 2010) and used to generate relative values for gene expression (FPKM – fragments per kilobase per million reads). The sequence reads were run through an allele specific expression pipeline that aligns reads using Tophat aligner, incorporating SNPs from the HapMap2 project (Chia et al., 2012) and allowing a maximum of two mismatches per read to assess allele specific expression rates. In order to eliminate potential false-positive SNPs each SNP had to be supported by at least 1% of the reads at that position in the pair of reciprocal hybrids. After filtering there were 28,195 - 142,033 SNPs that were used to assess allele-specific expression in each pair of reciprocal hybrids (Supplemental Table 1). The number of reads containing the B73, Mo17, Oh43, or Ki11 allele was summed for all SNPs within the same gene. The alignments were then analyzed to assess the number of reads that align to each of the parental alleles. Genes that have at least 10 allelic reads for each direction of the cross were used to perform chi-square tests (relative to an expected 2 maternal : 1 paternal) ratio. RNAseq reads are deposited at the NCBI SRA under project SRP031872.

Allelic variation detection: Allele specific read counts or Sequenom data for each set of

reciprocal crosses were analyzed to discover genes that exhibit allelic variation of imprinting. Genes with at least 20 RNA-seq reads were run through a pipeline that pulls the maize gene ID for genes that showed allelic variation of imprinting in at least two sets of reciprocal crosses, and identifies which alleles are not imprinted at the locus.

Quantitative SNP assays: Quantitative SNP assays (Sequenom MassArray) were used to validate imprinted genes and assess imprinting across additional genotypes or over a time course of seed development. cDNA sequences of candidate genes were used to create primers that amplify an informative SNP, which differentiates between two inbred lines. Genomic DNA from parental lines: B73, Mo17, Ki11, Oh43, Ms71, M162W, and NC358; endosperm cDNA from reciprocal crosses : B73xOh43, Mo17xOh43, B73xKi11, Mo17xKi11, B73xMs71, Mo17xMs71, B73xM162W, Mo17xM162W harvested 14 DAP; and endosperm cDNA from reciprocal crosses: B73xMo17, B73xNC358, and Mo17xNC358 harvested 8 DAP, 14 DAP, 16 DAP, 18 DAP, and 20 DAP were submitted to the University of Minnesota BioMedical Genomics Center. The imprinting status was then assessed across all genotypes and time points using standard Sequenom assay conditions (Stupar and Springer, 2006).

Annotation and comparative genomics of imprinted genes: Maize syntenic orthologs in rice and retained whole-genome duplicates were identified based on the criteria outlined and database created from Schnable et al. 2012. Potential maize homologs in *Arabidopsis thaliana* were identified by using BLAST to compare the nucleotide sequence of maize genes against the Arabidopsis database. Up to three hits were retained for genes that have at least one BLAST hit with an e-value $< 10^{-20}$. Arabidopsis genes with the highest percent identity and lowest e-value were used for analysis.

REFERENCES

Bartolomei MS and Ferguson-Smith AC. 2011. Mammalian genomic imprinting. Cold Spring Harb Perspect Biol 3(7):10.1101/cshperspect.a002592.

Bauer MJ and Fischer RL. 2011. Genome demethylation and imprinting in the endosperm. Curr Opin Plant Biol 14(2):162-7.

Berger F, Vu TM, Li J, Chen B. 2012. Hypothesis: Selection of imprinted genes is driven by silencing deleterious gene activity in somatic tissues. Cold Spring Harb Symp Quant Biol.

Bernardi J, Lanubile A, Li QB, Kumar D, Kladnik A, Cook SD, Ross JJ, Marocco A, Chourey PS. 2012. Impaired auxin biosynthesis in the defective endosperm18 mutant is due to mutational loss of expression in the ZmYuc1 gene encoding endosperm-specific YUCCA1 protein in maize. Plant Physiol 160(3):1318-28.

Chaudhuri S and Messing J. 1994. Allele-specific parental imprinting of dzr1, a posttranscriptional regulator of zein accumulation. Proc Natl Acad Sci U S A 91(11):4867-71.

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, et al. 2012. Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44(7):803-7.

Costa LM, Yuan J, Rouster J, Paul W, Dickinson H, Gutierrez-Marcos JF. 2012. Maternal control of nutrient allocation in plant seeds by genomic imprinting. Curr Biol 22(2):160-5.

Das R, Hampton DD, Jirtle RL. 2009. Imprinting evolution and human health. Mamm Genome 20(9-10):563-72.

Feng S, Jacobsen SE, Reik W. 2010. Epigenetic reprogramming in plant and animal development. Science 330(6004):622-7.

Gaut BS and Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci U S A 94(13):6809-14.

Gehring M, Bubb KL, Henikoff S. 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324(5933):1447-51.

Gehring M, Missirian V, Henikoff S. 2011. Genomic analysis of parent-of-origin allelic expression in arabidopsis thaliana seeds. PLoS One 6(8):e23687.

Gutierrez-Marcos JF, Constancia M, Burton GJ. 2012. Maternal to offspring resource allocation in plants and mammals. Placenta 33 Suppl 2:e3-10.

Haig D. 2004. Genomic imprinting and kinship: How good is the evidence? Annu Rev Genet 38:553-85.

Haun WJ, Laoueille-Duprat S, O'connell MJ, Spillane C, Grossniklaus U, Phillips AR, Kaeppler SM, Springer NM. 2007. Genomic imprinting, methylation and molecular evolution of maize enhancer of zeste (mez) homologs. Plant J 49(2):325-37.

Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, et al. 2011. Regulation of imprinted gene expression in arabidopsis endosperm. Proc Natl Acad Sci U S A 108(5):1755-62.

Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, et al. 2012. Comparative population genomics of maize domestication and improvement. Nat Genet 44(7):808-11.

Huh JH, Bauer MJ, Hsieh TF, Fischer RL. 2008. Cellular programming of plant gene imprinting. Cell 132(5):735-44.

Hutter B, Bieg M, Helms V, Paulsen M. 2010. Divergence of imprinted genes during mammalian evolution. BMC Evol Biol 10:116,2148-10-116.

Jahnke, S., Scholten, S. 2009. Epigenetic resetting of a gene imprinted in plant embryos. Current Biology, 19, 1677-1681

Jiang H and Kohler C. 2012. Evolution, function, and regulation of genomic imprinting in plant seed development. J Exp Bot 63(13):4713-22.

Kermicle JL and Alleman M. 1990. Gametic imprinting in maize in relation to the angiosperm life cycle. Dev Suppl :9-14.

Kermicle JL. 1970. Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. Genetics 66(1):69-85.

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJ, Koornneef M, Kakutani T. 2007. Control of FWA gene silencing in arabidopsis thaliana by SINE-related direct repeats. Plant J 49(1):38-45.

Kohler C and Weinhofer-Molisch I. 2010. Mechanisms and evolution of genomic imprinting in plants. Heredity (Edinb) 105(1):57-63.

Kohler C, Wolff P, Spillane C. 2012. Epigenetic mechanisms underlying genomic imprinting in plants. Annu Rev Plant Biol 63:331-52.

Li J and Berger F. 2012. Endosperm: Food for humankind and fodder for scientific discoveries. New Phytol 195(2):290-305.

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. Nature 430(6998):471-6.

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A. 2011. A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. PLoS Genet 7(6):e1002125.

Maere S., Heymans K., and Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21: 3448-3449.

Makarevitch I, Eichten SR, Briskine R, Waters AJ, Danilevskaya ON, Meeley RB, Myers CL, Vaughn MW, Springer NM. 2013. Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. Plant Cell .

Messing J and Dooner HK. 2006. Organization and variability of the maize genome. Curr Opin Plant Biol 9(2):157-63.

Miyake T, Takebayashi N, Wolf DE. 2009. Possible diversifying selection in the imprinted gene, MEDEA, in arabidopsis. Mol Biol Evol 26(4):843-57.

Pignatta D and Gehring M. 2012. Imprinting meets genomics: New insights and new challenges. Curr Opin Plant Biol 15(5):530-5.

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D. 2013. Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proc Natl Acad Sci

Schnable JC, Freeling M, Lyons E. 2012. Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol Evol 4(3):265-77.

Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, de Leon N, Kaeppler SM. 2013. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS One 8(4):e61005.

Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo JM, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive darwinian selection at the imprinted MEDEA locus in plants. Nature. 488:349-52.

Springer NM, Gutierrez-Marcos JF. 2008. Imprinting in maize. In The Maize handbook. Eds. Hake S, Bennetzen J) Springer-Verlag. New York.

Stupar RM and Springer NM. 2006. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics 173(4):2199-210.

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J., and Pachter L. 2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28: 511-515.

Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh CT, Xu W, Schnable PS, Vaughn MW, Gehring M, Springer NM. 2011. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell 23(12):4221-33.

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MT, Spillane C, Nordborg M, Rehmsmeier M, Kohler C. 2011. High-resolution analysis of parent-of-origin allelic expression in the arabidopsis endosperm. PLoS Genet 7(6):e1002126.

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol 8(6):e1000409.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178(1):539-51.

Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D. 2010. Local DNA hypomethylation activates genes in rice endosperm. PNAS 107(43):18729-34.

Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y, et al. 2011. Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. PNAS. 108(50):20042-7.

Table 1. Discovery of maize imprinted genes

	BM	BK	MK	BO	MO	All	NR
# genes with >= 10 reads	11,856	10,531	5,851	13,478	9,434	2,087	18,284
Maternal bias	81	58	77	180	134	6	394
Moderate MEGs	75	42	22	118	44	4	198
Strong MEGs	31	28	9	39	25	4	69
Complete MEGs	13	9	3	16	12	3	37
Paternal bias	432	563	403	724	487	24	1,750
Moderate PEGs	171	192	74	191	120	18	367
Strong PEGs	56	55	24	76	45	6	108
Complete PEGs	8	17	3	15	5	0	31

Column headers represent reciprocal crosses between genotypes. B= B73, M=Mo17, K=Ki11, and O=Oh43.

Table 2. Data workflow for discovery of imprinted genes

	B73xMo17	Mo17xB73	B73xKi11	Ki11xB73	Mo17xKi11	Ki11xMo17	B73xOh43	Oh43xB73	Mo17xOh43	Oh43xMo17
Number of raw RNA-seq reads	210,874,775	201,851,318	198,126,556	192,953,028	197,192,383	198,328,755	197,126,727	190,605,935	203,700,260	179,299,550
Per gene read counts	60,432,048	61,450,235	48,761,575	138,160,523	87,590,543	105,458,847	206,600,366	247,736,952	145,469,649	161,979,127
RPKM >1	10,818	11,031	9,545	9,682	5,931	5,690	10,148	10,136	8,036	8,713
# SNPs used	115,813		69,891		28,195		142,033		73,895	
# Genes with SNPs	17,262		15,355		8,329		19,504		13,823	
# Reads aligned to genes	17,134,491	15,681,619	15,620,862	22,676,835	14,936,746	13,409,826	99,084,549	84,317,265	40,865,059	35,523,692
#Genes in FGS	14,405		15,355		7,092		15,648		11,443	
# Reads aligned to FGS	14,427,498	13,552,279	12,930,459	19,086,835	12,830,641	11,160,346	76,529,348	67,101,931	35,638,015	27,879,407
Number of genes with >= 10 reads	12,201	12,381	10,693	11,326	6,298	6,044	13,988	13,873	10,344	9,641

Table 3. Sequenom Validation of Imprinted Genes

Gene ID	Imprinting Level	B73x Mo17	B73x Oh43	Mo17x Oh43	B73x Ki11	Mo17x Ki11	B73x NC358	Mo17x NC358	B73x M162W	Mo17x M162W	B73x Ms71	Mo17x Ms71	Validate in time course
GRMZM2G009465	cMEG	MEG	nd	MEG	nd	nd	nd	MEG	nd	MEG	nd	MEG	Yes
GRMZM2G014119	cMEG	MEG	nd	nd	nd	MEG	MEG	nd	nd	nd	nd	MEG	Yes
GRMZM2G063498	sMEG	MEG	MEG	nd	nd	nd	MEG	nd	nd	nd	nd	nd	Yes
GRMZM2G073700	sMEG	MEG	nd	nd	MEG	nd	nd	MEG	nd	nd	MEG	nd	Yes
GRMZM2G150134	cMEG	MEG	MEG	nd	MEG	nd	MEG	nd	nd	MEG	MEG	nd	Yes
GRMZM2G160687	cMEG[1]	MEG	MEG	nd	nd	nd	MEG	nd	ASI	nd	MEG	nd	Yes
GRMZM2G169695	cMEG	MEG	nd	nd	nd	nd	nd	MEG	nd	nd	nd	MEG	Yes
GRMZM2G178176	MEG[2]	MEG	nd	nd	nd	nd	MEG	nd	nd	nd	MEG	nd	Yes
GRMZM2G345700	MEG[2]	MEG	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Yes
GRMZM2G354579	cMEG	MEG	nd	nd	MEG	nd	MEG	nd	nd	nd	no	MEG	Yes
GRMZM2G370991	cMEG	MEG	MEG	nd	nd	nd	nd	MEG	MEG	nd	MEG	nd	Yes
GRMZM2G374088	mMEG	MEG	nd	MEG	nd	MEG	nd	MEG	nd	nd	nd	ASI	Yes
GRMZM5G802403	cMEG	MEG	nd	nd	nd	nd	nd	MEG	nd	nd	nd	nd	Yes
GRMZM2G000404	cPEG	PEG	nd	PEG	PEG	nd	PEG	nd	nd	PEG	nd	PEG	Yes
GRMZM2G002100	sPEG	PEG	nd	PEG	nd	PEG	nd	nd	nd	PEG	nd	PEG	Yes
GRMZM2G006732	sPEG	PEG	PEG	nd	PEG	nd	PEG	nd	nd	PEG	PEG	nd	Yes
GRMZM2G020302	sPEG	PEG	nd	PEG	nd	PEG	nd	ASI	nd	PEG	nd	PEG	No
GRMZM2G040954	sPEG	PEG	nd	nd	nd	PEG	nd	PEG	nd	PEG	nd	nd	Yes
GRMZM2G047104	cPEG	PEG	nd	PEG	nd	PEG	nd	PEG	nd	PEG	PEG	nd	Yes
GRMZM2G093947	cPEG	PEG	PEG	nd	PEG	nd	nd	PEG	PEG	nd	nd	PEG	Yes
GRMZM2G149903	cPEG	PEG	PEG	nd	PEG	nd	PEG	nd	PEG	nd	PEG	nd	Yes
GRMZM2G164314	sPEG	PEG	PEG	nd	PEG	nd	nd	PEG	nd	PEG	PEG	nd	Yes
GRMZM2G171410	mPEG	PEG	nd	PEG	nd	PEG	nd	PEG	PEG	nd	nd	PEG	Yes
GRMZM2G365731	cPEG	PEG	nd	PEG	nd	PEG	nd	nd	nd	PEG	PEG	nd	Yes
GRMZM2G369203	cPEG	PEG	nd	nd	PEG	nd	nd	PEG	nd	nd	nd	PEG	Yes
GRMZM2G440949	sPEG	PEG	PEG	nd	PEG	nd	PEG	nd	PEG	nd	PEG	nd	Yes

No data (nd) and allele specific imprinting (ASI)

[1] Complete MEG that was filtered out as potential contaminant by Sekhon et al. 2011) expression data, but validated in sequenom

[2] Did not meet the minimum read requirement in one direction of reciprocal crosses, but validated in sequenom

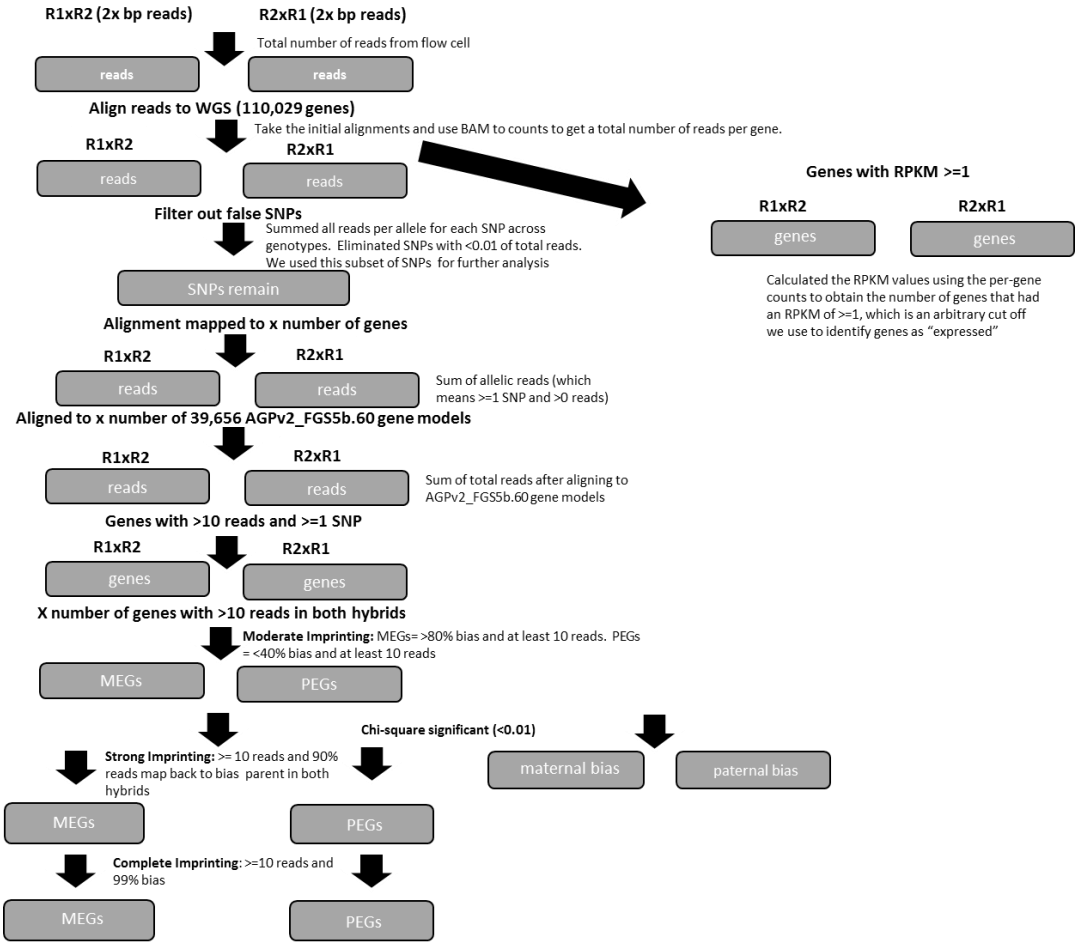


Figure 1- Analysis pipeline for RNA-seq based discovery of imprinted genes. Raw reads from reciprocal hybrids were aligned to the working gene set of version two of the B73 reference genome using TopHat aligner. Two different alignment iterations were completed; allele specific expression analysis (the flowchart) and gene specific alignments (calculate RPKM values). SNPs from resequence data were used to asses allele specific expression rates (Chia et al, 2011). We required at least 1% of the allelic reads for any given SNP came from both parents. False SNPs have all reads map to one parental allele in both reciprocal crosses. Once false SNPs were filtered out the allelic reads covering the remaining SNPs (variant reads) were summed for each gene. Genes in the filtered gene set (AGPv2_FGSv5b) were used for further analysis. At least 10 variant reads are required in both directions of the cross to assess parental bias expression. A chi-square significance test was performed on each gene. Maternally and paternally biased genes have a significant parental bias (<0.01). Additional cutoffs were used to identify and categorize MEGs and PEGs into three categories: moderate, strong, and complete. Moderate MEGs and PEGs have a significant parental bias expression (<0.05) and >80% maternal bias or >60% paternal bias, respectively. Strong MEGs and PEGs have a significant parental allelic bias (<0.01) and >90% maternal or paternal bias, respectively. Complete MEGs and PEGs have >99% maternal or paternal bias, respectively.

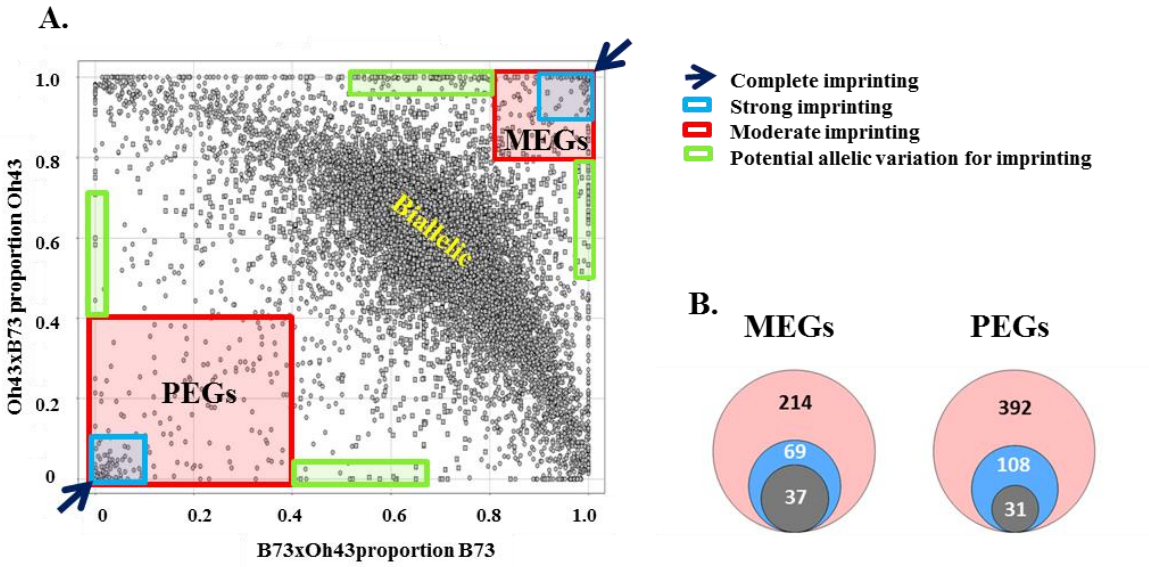


Figure 2. Discovery of imprinted genes in maize. A) Allele-specific expression analysis for the reciprocal F1 genotypes generated by crossing B73 and Oh43. The proportion of maternal transcripts in both reciprocal hybrids is plotted for the 13,478 genes that had at least 10 allelic reads in both directions of the cross of B73 and Oh43. Similar plots for the other reciprocal hybrids are shown in Figure S2. Circle symbols represent genes that are significantly (χ^2 <0.05) different from the expected 2:1 maternal to paternal ratio whereas square symbols are genes that do not significantly differ from expected (χ^2 >0.05). The pink shaded areas indicate moderate MEGs and PEGs, (χ^2 <0.05 and at least 80% maternal bias or 60% paternal bias, respectively). The blue shaded areas indicate strong imprinting have significant allelic bias (χ^2 <0.01) and exhibit at least 90% parental bias for both MEGs and PEGs. The arrows heads indicate genes with complete imprinting (at least 99% parental bias for MEGs and PEGs and χ^2 <0.01)). The green shaded areas indicate genes with potential allelic variation for imprinting (are strongly imprinted in one direction of the cross and biallelic in the reciprocal cross). B) The proportion of moderate (pink), strong (blue) and complete (gray) imprinting for all non-redundant MEGs and PEGs that were detected in at least one of the five reciprocal crosses.

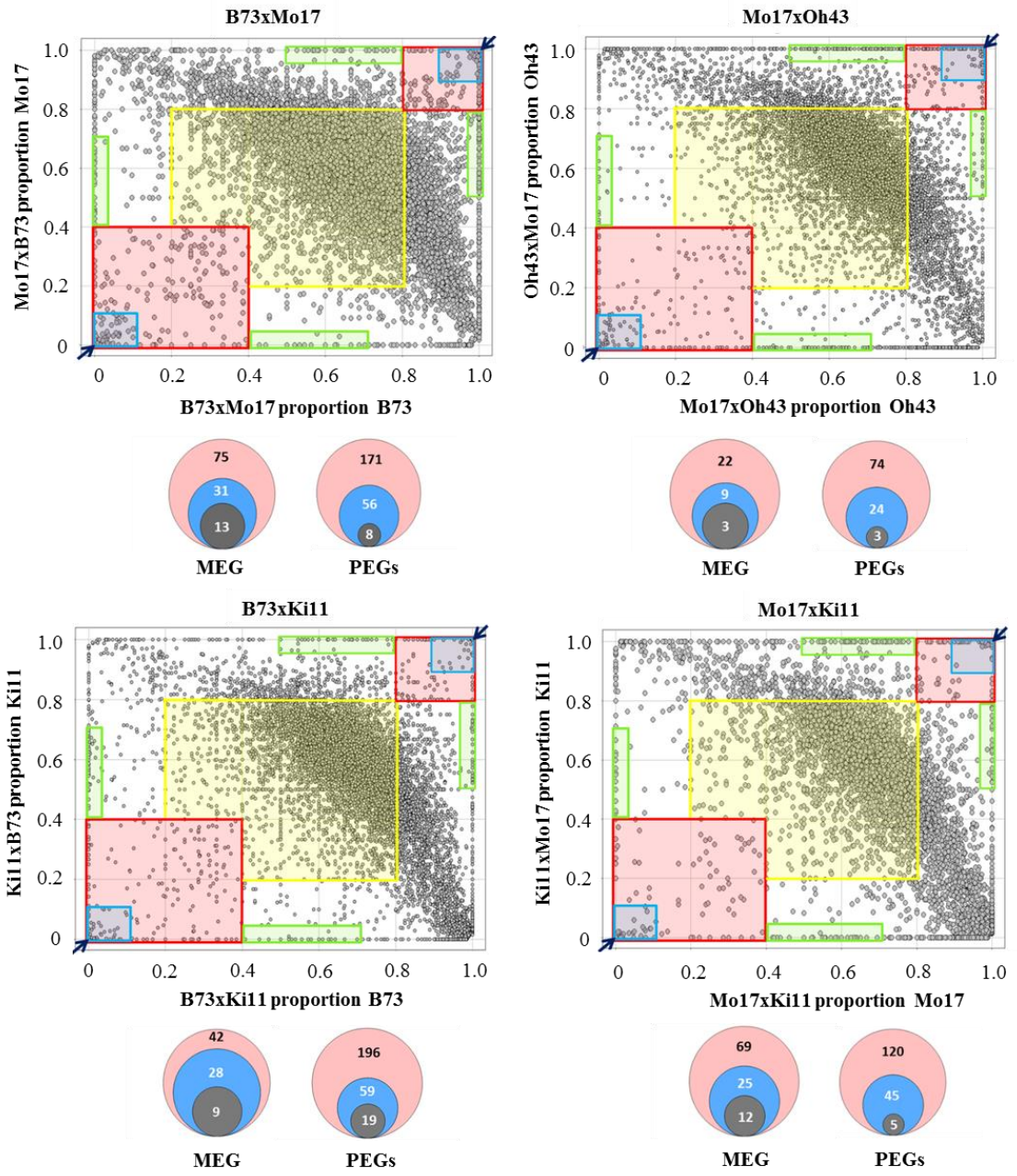


Figure 3. Proportion of maternal transcripts and number of moderate, strong and complete MEGs and PEGs for B73xMo17, Mo17xOh43, B73xKi11, and Mo17xKi11. Moderate MEGs and PEGs (>80% maternal or >60% paternal bias, respectively) are within the red boxes. Strong MEGs and PEGs (>90% parental bias) are within the blue boxes. Complete MEGs and PEGs (>99% parental bias) are represented by the arrows. Biallelic genes are within the yellow squares. The green boxes represent allele specific imprinted genes. The number of moderate, strong, and complete MEGs and PEGs for each genotype are represented by the red, blue, and grey circles, respectively.

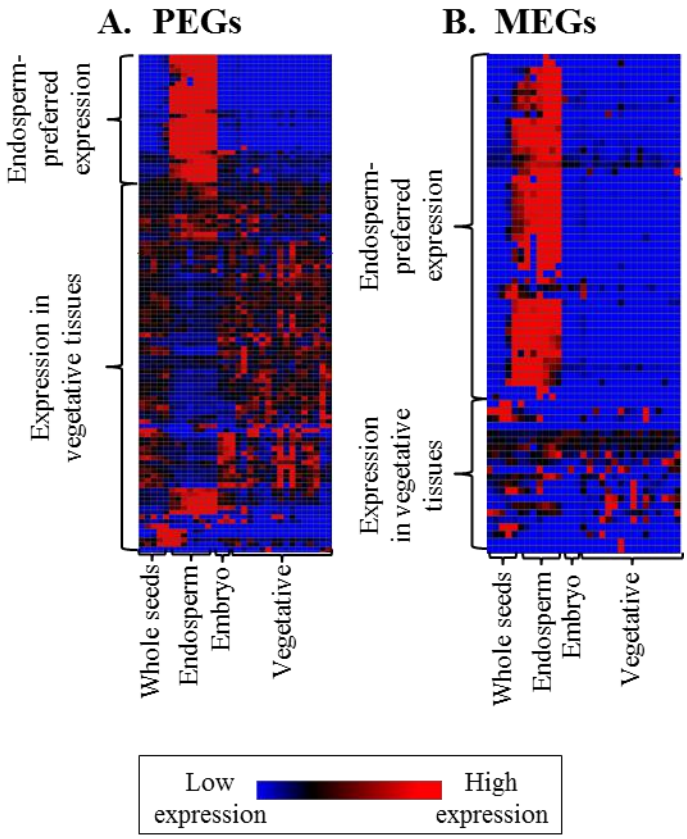


Figure 4. **A subset of imprinted genes show endosperm preferred expression while other imprinted genes are expressed in vegetative tissues.** The gene expression patterns for the PEGs (A) and MEGs (B) were obtained from the maize gene expression atlas (Sekhon et al., 2013). The normalized values (per gene) were used for hierarchical clustering (Ward's method) and the heat map indicates relative levels of expression (red – high; black – intermediate; blue – low). The genes with preferential expression in endosperm are indicated to the left of each heat map. The whole seed samples are 2, 4, 6, 10 and 14 DAP (left to right). The endosperm samples are 12, 14, 16, 18, 20, 22 and 24 DAP. The embryo samples are 16, 18 and 22 DAP. The vegetative samples are 18 DAP pericarp, anthers, pre-pollination cob, silks, leaves, stem, immature tassel, immature leaves, immature cob, meiotic tassel, first internode, shoot tip, and three leaf stages.

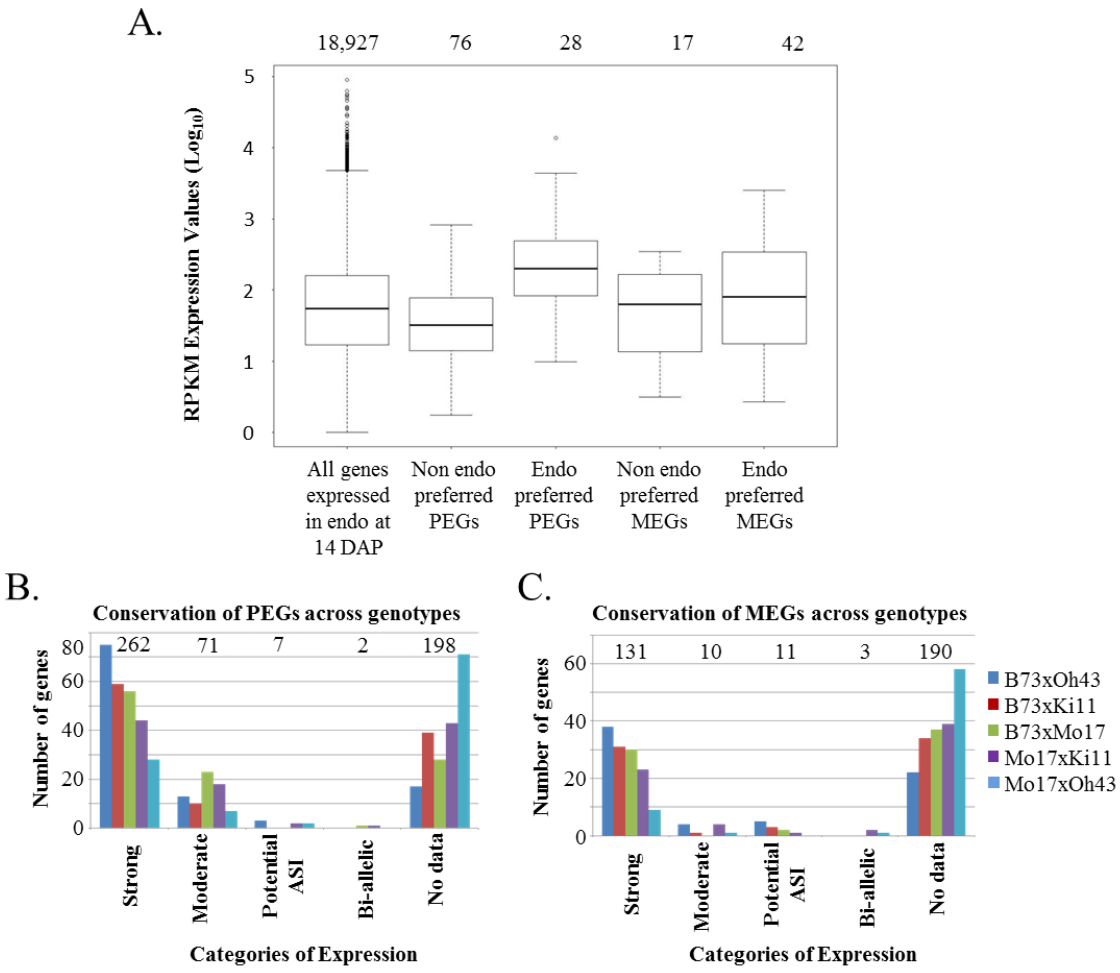


Figure 5. Comparison of tissue specific expression and conservation of imprinting among maize hapoltypes. (A) RPKM expression values are log transformed data from Sekhon et al., 2013. All genes with RPKM values of 0 were removed from this analysis. Gene expression was compared across all genes that are expressed in endosperm tissue at 14 DAP to MEGs and PEGs that are either preferentially expressed in endosperm or show equivalent expression in other tissue types. The values listed above box plots are the number of genes in each group. PEGs that are endosperm preferred have a higher average expression 2.3) relative to all other groups (1.5-1.8). No statistical difference in expression was observed for either group of MEGs or PEGs that are not preferentially expressed in endosperm. B) The conservation level of imprinting was assessed within maize genotypes. Very few genes identified as MEGs and PEGs in one pair of crosses is biallelic in another set of crosses (0.009 and 0.004, respectively). Each color represents a pair of reciprocal F1 hybrid crosses. Each non-redundant imprinted gene could be placed in one of five categories for each pair of reciprocal crosses: Strong imprinting (>90% parental bias), moderate imprinting (>80% maternal bias (MEGs) or >60% paternal bias (PEGs)), potentially allele specific imprinting (ASI, strong imprinting in one direction and biallelic expression in the other direction), or no data (no SNPs or <10 reads). The values above the chart show the number of genes that fit in each category. Over 97% of PEGs and 91% of MEGs are at least moderately imprinted in all genotypes with data. A lack of polymorphisms or reads (no data call) is more prevalent for MEGs (0.55) than PEGs (0.37).

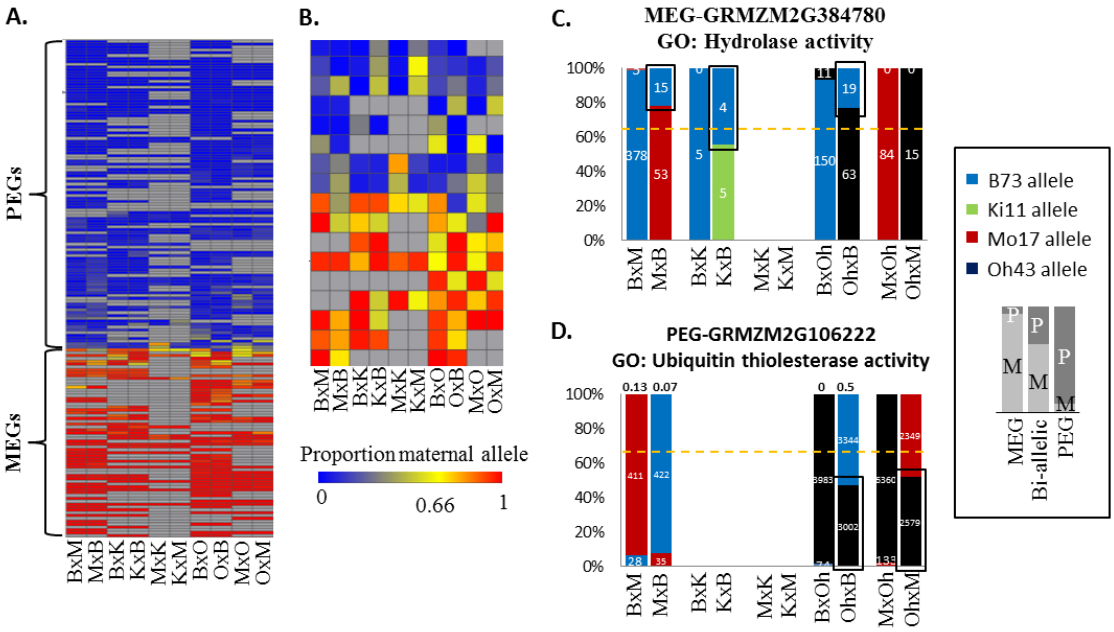
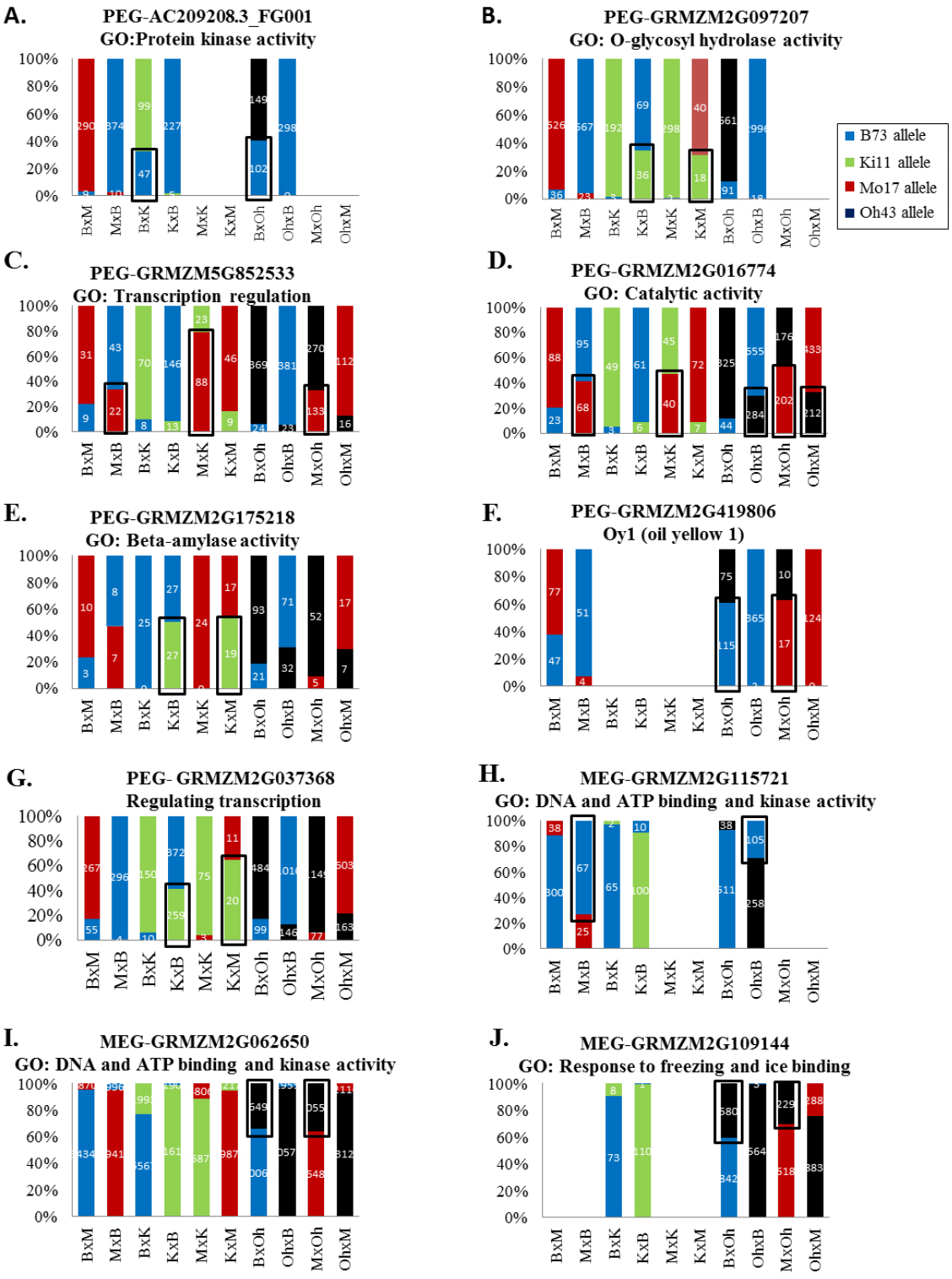


Figure 6. **Conservation of imprinting among maize haplotypes**. A) The proportion of expression from the maternal allele using a heatmap (blue = 0; red = 1; yellow = 0.66; gray = missing data) is shown for all ten genotypes for each of the non-redundant imprinted genes. B) A similar heatmap is shown for the seventeen genes with allelic variation for imprinting. C) The expression patterns for one of the allele-specific imprinted MEGs (GRMZM2G384780) is shown. For this gene, the B73 allele is not silenced when paternally inherited but alleles from the other haplotypes are silenced when inherited from the paternal parent. For each bar, the upper portion represents the proportion of paternal expression and the lower portion represents the proportion of maternal expression (see gray bars in legend for expectations for MEGs, biallelic and PEGs). The colors represent the four alleles assessed (see legend), and the values listed inside the bars are the number of maternal (M, bottom) or paternal (P, top) reads. The orange dashed line across the plot represents the expected biallelic ratio of 66% maternal reads. Black boxes highlight the non-imprinted allele. D) Allele-specific imprinting pattern for the PEG GRMZM2G106222, which exhibits a failure to silence the Oh43 when it is maternally inherited. This gene was also validated by a quantitative SNP assay. SNPs were available to distinguish B73-Mo17 and B73-Oh43 alleles. The values listed above the bars are the proportion of the maternal allele determined from the quantitative SNP assay.


A.
PEG-AC209208.3_FG001
GO:Protein kinase activity
B.
PEG-GRMZM2G097207
GO: O-glycosyl hydrolase activity
B73 allele
Ki11 allele
Mo17 allele
Oh43 allele
C.
PEG-GRMZM5G852533
GO: Transcription regulation
D.
PEG-GRMZM2G016774
GO: Catalytic activity
E.
PEG-GRMZM2G175218
GO: Beta-amylase activity
F.
PEG-GRMZM2G419806
Oy1 (oil yellow 1)
G.
PEG- GRMZM2G037368
Regulating transcription
H.
MEG-GRMZM2G115721
GO: DNA and ATP binding and kinase activity
I.
MEG-GRMZM2G062650
GO: DNA and ATP binding and kinase activity
J.
MEG-GRMZM2G109144
GO: Response to freezing and ice binding
100%
80%
60%
40%
20%
0%
BxM
MxB
BxK
KxB
MxK
KxM
BxOh
OhxB
MxOh
OhxM

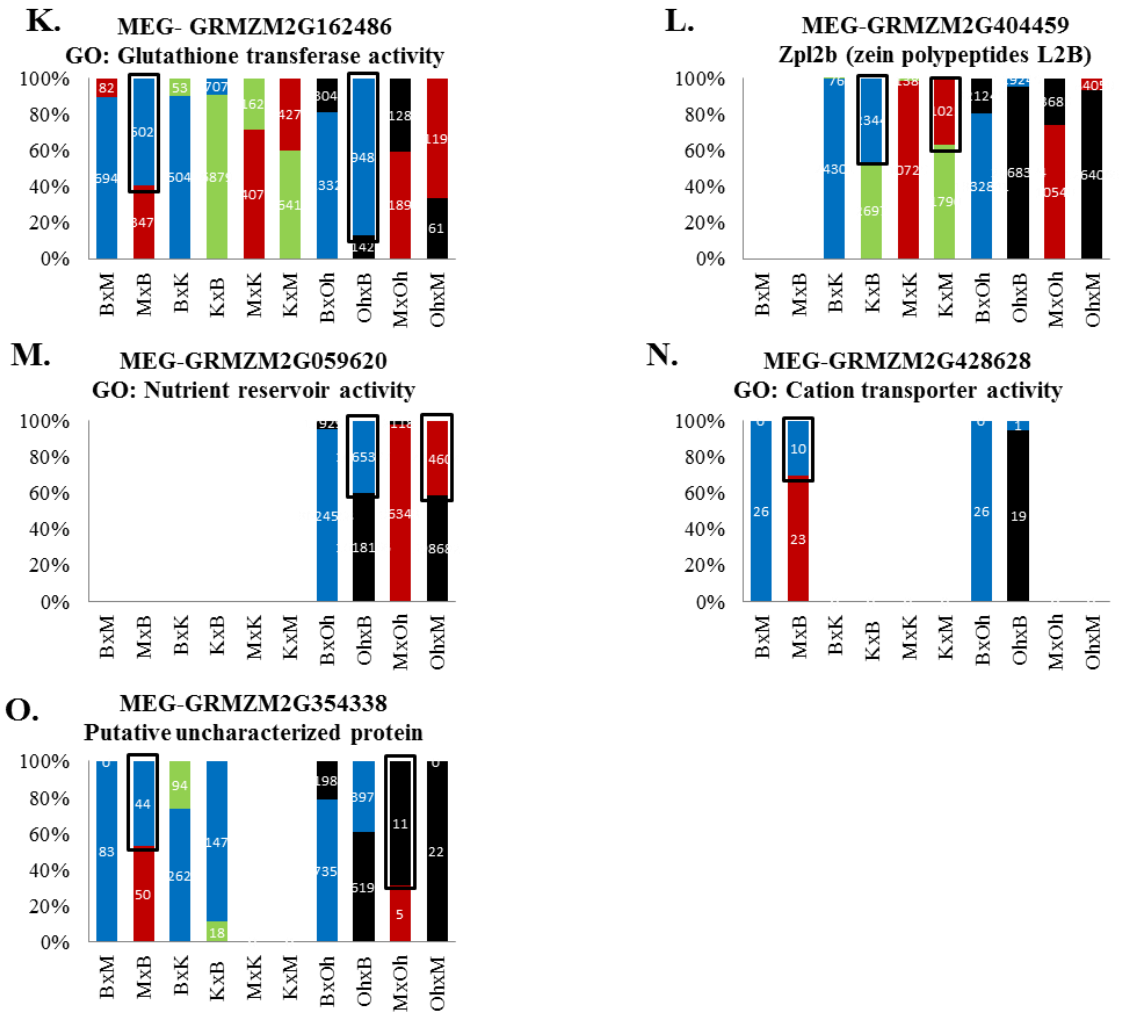
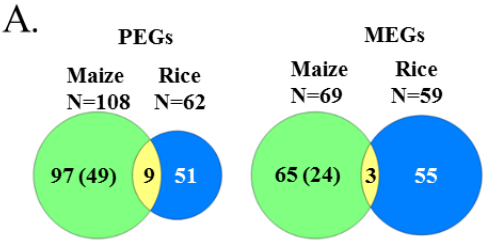
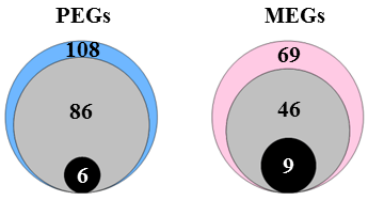


Figure 7. PEGs and MEGs exhibit allelic variation of imprinting. The number of maternal reads (top) and paternal reads (bottom) are shown for each of the seven PEGs and eight MEGs that exhibit allelic variation of imprinting. Each color represents a different allele and the black boxes highlight the allele(s) that fails to silence. A variety of patterns were observed in terms of number of alleles and which alleles fail to silence. Many of these PEGs fail to silence certain alleles when they are maternally inherited (**A-E** and **G**) , only 1 example (**F**) the allele fails to silence when it is paternally inherited. In addition, one allele fails to silence in a majority of these PEGs (**A-C, E and G**), whereas in **D and F** multiple alleles fail to silence when maternally or paternally inherited, respectively. B73 fails to silence in **A** and **F**, Mo17 fails to silence in **C**, **D**, and **F**, Ki11 fails to silence in **B**, **E** and **G,** and Oh43 fails to silence in **F**. Characterized gene oil yellow exhibits allelic variation of imprinting, in that both B73 and Mo17 fail to silence when Oh43 is inherited paternally. Similar patterns were observed for MEGs that exhibit variation of imprinting. The B73 allele failed to silence when it paternally inherited in **H, K,** and **N** and when maternally inherited in **M** and **O.** The Oh43 allele failed to silence when paternally inherited in **I** and **J**, and when maternally inherited in **O**. Failure to silence Mo17 was only observed when Mo17 was inherited maternally and additional alleles failed to silence when Ki11 and Oh43 are maternally inherited (**L** and **M**, respectively).

A.



B.



C.

Gene ID	Type	Rice ID	Annotation	Arabidopsis
GRMZM2G447406	PEG	Os01g70060	Protein of unknown function	
GRMZM2G028366	PEG	Os02g12840	DEAD-box ATP-dependent RNA helicase	
GRMZM5G871520	PEG	Os02g57080	Serine/threonine-protein kinase	
GRMZM5G845175	PEG	Os04g32880	SNF4-like protein	
GRMZM2G042870	PEG	Os06g40490	Glycosyl hydrolase activity	
GRMZM2G000404	PEG	Os09g28940	Ubiquitin sepcific domain	
GRMZM2G365731	PEG	Os10g30944	DNA binding (ARID/BRIGHT)	Imprinted
GRMZM2G108309	PEG	Os10g39780	Protein phosphatase	
GRMZM2G091819	PEG	Os12g08780	Flavin binding monooxygenase (YUC11)	Imprinted
GRMZM2G073700	MEG	Os06g11730	RNA binding	
GRMZM2G118205	MEG	Os08g04290	WD40 repeat containing protein (FIE)	
GRMZM2G003909	MEG	Os10g05750	Allergen and extensin family protein (POEI3)	

D.

Gene ID	Type	Rice ID	Annotation
GRMZM5G897988	PEG	Os01g63250	Ataxin/ Josephin
AC197717.3_FG002	PEG	Os03g11550	G-protein coupled receptor
GRMZM2G146819	PEG	Os03g38970	Tyrosine phosphatase activity
GRMZM2G041312	PEG	Os04g32880	AMP-activated protein kinase
GRMZM2G075582	PEG	Os06g06870	Zinc ion binding
GRMZM2G104866	PEG	Os06g42990	Regulation of transcription and DNA binding
GRMZM2G472052	PEG	Os07g12490	RNA binding
GRMZM2G087212	PEG	Os11g07910	Transporter activity
GRMZM2G102356	MEG	Os04g39150	Defense response (Prp2)
GRMZM2G178435	MEG	Os02g43460	DNA and ATP binding
GRMZM2G323353	MEG	Os01g10520	Expressed protein

Figure 8. Limited, but significant, conservation of imprinting between maize, rice and *Arabidopsis*. Syntenic regions in rice were identified and the level of conservation was assessed (A). The values in parentheses are the number of maize imprinted genes that have a syntenic ortholog in rice and were assessed by Luo et al (2011). Assessing conservation of syntenic genes identified 9 PEGs and 3 MEGs that are conservatively imprinted between maize and rice. (B) The level of conservation between maize and *Arabidopsis thaliana* was assessed. The protein sequence of each maize imprinted gene was used to find potential orthologs in *Arabidopsis* (BLASTp). The blue outer circle is the number of non-redundant maize PEGs and the pink outer circle is the number of non-redundant maize MEGs. The grey circles are the number of PEGs (0.80) or MEGs (0.67) that have a potential ortholog in *Arabidopsis* (BLAST e-value <1e-20) for which the Arabidopsis ortholog was assessed by either Gehring et al (2011) or Heish et al (2011). The black circles are the number of MEGs or PEGs that are conservatively imprinted between maize and *Arabidopsis*. Approximately, 7% of PEGs compared to 20% of MEGs, for which an ortholog was assessed are conserved between maize and *Arabidopsis*. C) The maize gene ID, rice gene ID, annotation and imprinting status in *Arabidopsis* for PEGs and MEGs that are conservatively imprinted in maize and syntenic blocks in rice. (D) Maize gene ID, rice gene ID, and annotation for the 3 moderate MEGs and 8 moderate PEGs that show conservation of imprinting between maize and syntenic blocks in rice.

Chapter IV: **context statement**

SUMMARY

Plants respond to abiotic stress through a variety of physiological, biochemical, and transcriptional mechanisms. A large number of genes exhibit altered levels of expression in response to abiotic stress, which requires a concerted action of both cis- and trans-regulatory features. In order to study the variability in transcriptome response to abiotic stress, RNA sequencing was performed using 14 day old maize seedlings of inbreds B73, Mo17, Oh43, and PH207 as well as F1 hybrids B73xMo17, Oh43xB73, and B73xPH207 under control, cold, and heat conditions. We identified a large number of differentially expressed genes between parental inbred lines for each condition, as well as many examples of stress-responsive gene expression within each of the inbred lines. By evaluating allele specific transcript abundance in the F1 hybrids we were also able to study natural variation for cis- and trans-regulatory variation between genotypes as well as for stress responsive expression. Although examples of trans- regulatory variation were observed, a large proportion of genes exhibit cis- regulatory variation for both genotype and stress-responsive expression. These genes with allelic variation for cis-regulation in response to cold or heat stress provide an opportunity to study the basis for regulatory diversity. The combined analysis of promoter haplotype diversity (especially at predicted transcription factor binding sites) and stress-responsive expression at these genes in a diverse panel of maize lines revealed potential sources for regulatory variation for maize responses to abiotic stress. An improved understanding of the molecular basis for diversity of gene expression responses to abiotic stress can be used to inform strategies for engineering more stress-tolerant genotypes.

Additional authors contributed to this work. Nathan Springer, Irina Makarevitch, and I designed the experiment. Candice Hirsch and Cory Hirsch provided computational and statsical assistance. Irina Makarevitch, Peter Hermanson, and I collected tissue and prepared RNA for sequencing. Jaclyn Noshay provided the CBF/DREB information.

Nathan Springer provided assistance with development of analytical criteria, comments on the manuscript and figures, and ran the enrichment for transcription factors. I analyzed the RNAseq expression data, performed the stastical tests, ran the motif enrichment, made the figures and tables, and wrote the manuscript.

CHAPTER IV
Regulatory variation for gene expression response to abiotic stress

INTRODUCTION

As sessile organisms, plants utilize multiple mechanisms to respond to, and survive in, many different environments. Plants may experience a myriad of abiotic stresses over the course of one life cycle including extreme temperatures, drought, and nutrient deficiencies. These stresses can have negative effects on plant health, survival, and in the case of crops a significant loss of yield. However, constitutive responses to adverse environments can limit growth and fitness (Gilmour et al., 2000; Dubouzet et al., 2003). Therefore, plants have developed physiological, biochemical, and transcriptional processes to sense and respond to environmental stresses. Temperature stress, in particular, has been shown to negatively affect germination, stunt growth, increase leaf chlorosis and necrosis, and reduce yield (Yadav 2009; Hasanuzzaman et al., 2013). While there are important biochemical and physiological responses to abiotic stress (Ashraf and Hafeex, 2004; Crafts-Brandner and Salvucci, 2002; reviewed in Hasanuzzaman et al., 2013) the regulation of gene expression has also been shown to play significant roles in many different stress responses (Miura and Furumoto, 2013; Wang et al., 2004).

Differential gene expression of key components in stress response pathways is essential for tolerance to abiotic stress. Expression of several transcription factors (TFs) has been documented to have direct roles in temperature stress response as well as inducing transcription of important stress response genes (Abe et al, 1997; Hasanuzzaman et al., 2013; reviewed in Qu et al., 2013). In *Arabidopsis thaliana* a set of TFs called dehydration-responsive element binding (DREB) factors or C-repeat binding factors (CBF) play important roles in tolerance to various abiotic stresses including tolerance to cold, drought, and heat (Jaglo-Ottosen, 1998; Schramm et al., 2007; Sakuma et al., 2006). CBF/DREBs play similar roles in many flowering plants including maize, rice, and soybean (Liu et al., 2013; Dubouzet et al., 2003; Chen et al., 2007). There is also evidence that expression of the cold-responsive (COR) group of genes discovered in

Arabidopsis thaliana and the heat shock protein (Hsp) families are important for cold and heat tolerance, respectively (Hasanuzzaman et al., 2013; Yadav 2009).

One key abiotic stress response mechanism is altering gene expression, which likely involves both cis- and trans-regulatory mechanisms. Examples of both cis- and trans- regulatory elements have been described for responses to abiotic stress. Cis-regulatory elements 5' of COR genes, including DRE, as well as ABA responsive promoter elements (ABREs), and heat shock elements (HSEs) confer stress responsive gene expression (Yamaguchi-Shinozaki and Shinozaki., 1994, Baker et al., 1994, Yamaguchi-Shinozaki and Shinozaki, 1993, Hovarth et al., 1993, Mittal et al., 2011, reviewed in Raghavenadra et al., 2010). These elements provide the sequence necessary for trans- acting TFs, including DREBs, to bind and transcribe genes (Liu et al., 1998). There is evidence for contributions of both cis- and trans-regulatory variation for *Arabidopsis thaliana* responses to drought conditions with greater contributions of cis-regulatory variation both between accessions and between treatments (Cubillos et al., 2014).

Maize (*Zea mays*) exhibits a vast amount of transcriptomic diversity between inbred lines (Chia et al., 2012, Hirsch et al., 2014). This variation has been used to document cases of cis- and trans- regulatory variation between inbred lines and tissues as well as associations with important processes like domestication and heterosis (Stupar and Springer 2006, Li et al., 2013, Swanson-Wagner et al., 2009, Lemmon et al., 2014). One source of cis- regulatory variation for allelic specific response to abiotic stress in maize is polymorphisms of certain transposable element (TE) families proximal to the transcription start site of genes (Makarevitch et al., 2015).

In this study we utilized the extensive natural variation of maize to study the differences in gene expression responses to abiotic stress among different genotypes. By monitoring allele-specific gene expression in heterozygous plants it is possible to identify examples of cis-regulatory variation for the ability to respond to cold or heat stress. This provides an opportunity to study the evolution of gene expression responses to stress and the cis-factors that are necessary for responses to abiotic stress.

RESULTS

RNAseq was used to survey the extent of natural variation and allelic response to cold and heat stress in several maize inbred and hybrid genotypes. The inbred parents (B73, Mo17, PH207 and Oh43) were selected because they have genetic/genomic resources including whole genome assemblies (B73 and PH207) or deep re-sequencing data (Mo17 and Oh43). These inbred lines also represent four different population groups of maize (Nelson et al., 2008) and provide transcriptome diversity (Stupar et al., 2008). The F1 hybrids B73xMo17, B73xPH207, and Oh43xB73 were also used to examine the relative expression of both alleles. The transcriptome data was produced in two separate experiments. The first experiment included B73, Oh43 and Oh43xB73. The second experiment included B73, PH207, Mo17, B73xMo17 and B73xPH207. In total, this provides three triplet combinations of two parents and their F1 off-spring for all conditions with data from the same experiment. For all seven genotypes RNAseq was performed for three biological replicates of 14-day seedlings subjected to control, cold, or heat treatments (Table 1). Stressed samples were collected immediately following the treatment to reduce gene expression differences due to morphological changes and developmental differences between control and stressed plants that arise during subsequent growth. Gene expression levels were estimated from the RNAseq data by aligning the reads to SNP-corrected reference genomes and counting uniquely aligned reads. For the hybrid genotypes, the allele-specific expression (ASE) ratio was also determined by using SNPs within the transcripts to count reads derived from each allele. This dataset provides data for 25,433-25,641 expressed genes in each genotype by treatment combination and approximately 74% of these genes have ASE data.

Variation for transcriptional responses to abiotic stress among maize inbreds

In order to confirm that the cold and heat stress treatments were triggering cold or heat responses we investigated the GO terms that are over-represented in up-regulated genes for each of the treatments (see Methods for details). As expected, we found that

many of the most significant over-represented GO categories were associated with response to stress in cold and heat conditions, respectively (Tables 2-3). Transcription factors (TFs) have been shown to be important for stress responses (Abe et al., 1997, Wang et al., 2004, Mizoi et al., 2012). We therefore assessed whether maize TF families were enriched for up-regulated genes in response to heat or cold stress. There are 20 TF families (of 57 maize TF families, Yilmaz et al., 2008) that are enriched for up-regulated genes in cold and/or heat stress relative to control conditions (Figure 1A). There are 8 families of TFs that are enriched for genes that are up-regulated in response to heat stress including the heat shock (HS) and the myeloblastosis (MYB) TF families (Figure 1A). Many (15/29) of the HSF TFs in maize (Yilmaz et al., 2008) showed a significant (FDR corrected p-value <0.01) increase in expression when plants were exposed to heat stress (Figure 1B). There are 18 families of TFs that are enriched for up-regulation in response to cold stress including the APETALA2/Ethylene Responsive Factor (AP2/EREB), which contains the maize orthologs of DREB/CBF genes. Many (8/13) of the maize CBF/DREB orthologs (Liu et al., 2013) exhibit up-regulation in response to cold or heat stress (Figure 1C).

Prior research has provided evidence that genes which respond to abiotic stress often contain DRE/CRT, ABRE, or HSE motifs provide binding sites for DREB/CBF, ABA, or HSF genes, respectively (Yamaguchi-Shinozaki and Shinozaki, 1994, Guiltinan et al., 1990, Xiao and Lis, 1988). The genomic distribution of DRE/CRT, ABRE, and HSE motifs in the B73 reference genome was used to assess whether these motifs are enriched near genes that are up-regulated in response to heat or cold stress in B73. Genes that are up-regulated in response to cold stress are enriched for ABRE elements in the 1kb 5' of the gene while these genes show enrichment of DRE/CTF motifs in the 1kb 5' and 3' of the gene as well as within the gene bodies and no enrichment was found for HSEs (Figure 1D). Genes that are up- regulated in response to heat show enrichment in the 1kb 5' and 3' as well as within the gene body for both DRE/CRTs and HSEs while ABREs are enriched in the 1kb 5' upstream.

A major goal of this study was to document genotype x environment effects on gene expression. Several complementary approaches were used to assess the patterns of gene expression in response to genotype and environment. Clustering of the gene expression levels reveals significant environmental effects on the transcriptome but also finds evidence that genotypes have slightly different responses, with hybrids often exhibit intermediate responses (Figure 2). Clustering was performed using genes that are differentially expressed (DE) in at least one inbred genotype in response to heat (n=7,838) or cold (n=7,155) (Figure 3). Many genes have consistent responses to heat or cold stress in all inbreds. However, there are also a substantial number of genes that have variable responses to the cold or heat treatments among genotypes. Together, the clustering of all expressed genes or DE genes provides evidence for many changes in gene expression as a result of the cold or heat stress and reveals some variation for response among different genotypes.

A two-way factorial ANOVA was used to determine the frequency of significant genotype, environment, and genotype x environment (GxE) effects on gene expression levels in the two inbred parents in each of the three stresses for each triplet of samples (Table 4). The ANOVA finds that 21%-52% of genes exhibit significant effects from genotype and/or environment. There are also 1-28% of genes with significant GxE effects on gene expression in the genotypes and environments used for this analysis (Table 4). A visualization of the amount of variation (η^2 values) that is explained by genotype, environment and GxE reveal large clusters of genes predominantly affected by either genotype or environment with fewer genes with strong GxE effects (Figure 4). Given our interest in studying allelic variation for abiotic stress response the subsequent analyses are focused on genes with significant genotype or GxE effects that have allele-specific expression data.

High-levels of cis-regulatory variation contribute to gene expression differences among genotypes

Previous research on maize has documented high levels of cis-regulatory variation, particularly for genes with strong expression variation between inbred lines (Stupar and Springer 2006; Holloway et al., 2011; Li et al., 2013). In this study we had the opportunity to evaluate whether the contributions of cis- and trans-regulatory variation are similar in different environmental conditions. The relative contribution of cis- and trans-regulatory variation was estimated for each trio of parents and the hybrid within each environmental treatment. Biased allelic expression in the F1 provides evidence for cis-regulatory variation while balanced expression of the two alleles, despite differences in expression in the parental genotypes, suggests trans-regulatory variation (see examples in Figure 5A-C). Genes were classified as having cis- or trans-regulatory variation by performing chi-square analyses of the observed read counts relative to either the cis- or trans-expectation and identifying the genes with significant ($X<0.01$) from one expected value but not the other (Table 5). The observed allelic proportions in the F1 hybrids were compared to the expected allelic proportions predicted from the relative expression level of the parents for each gene in each treatment (see methods for details) in order to visualize the relative contributions of cis- and trans-regulatory variation (Figures 6). Cis-regulatory variation is quite prevalent in all conditions with fewer examples of trans-regulatory variation. Genes that are classified as subject to cis-regulatory variation in the control experiments frequently exhibit cis-regulatory variation in both of the stress treatment conditions with relatively few examples of the same genes switching to trans-regulatory variation (Figure 7A). In contrast, it was more common for genes that are classified as having trans-regulatory variation in control conditions to exhibit cis-regulatory variation in heat or stress conditions (Figure 7B).

Allelic variation for responsiveness to abiotic stress

Given the high levels of transcriptional variation in response to abiotic stress among maize inbreds and the frequent cis-regulatory variation within a treatment we

expected that there would also be examples of cis-regulatory variation for gene expression responses to abiotic stress. By studying the allelic variation for the ability to respond to heat or cold stress it should be possible to begin to understand the mechanisms of evolution of gene expression responses. In order to identify examples of genes with regulatory variation in response to heat or cold stress we found genes (with ASE data) that have equivalent levels of expression in the two inbred parents in control conditions but exhibit significant differences in expression following heat or cold stress . These include genes with both up- and down-regulation in response to the stress as well as genes that are silent in control conditions. These could include examples of cis- or trans-regulatory variation for responsiveness. If a gene has purely trans- responsive regulatory variation then the two alleles in the stress F1 hybrid should not be statistically different than the proportion observed in the F1 control. Whereas, a biased expression of alleles in the F1 stress sample suggests responsive cis- regulatory variation (Figure 8A). Genes can exhibit allelic bias either toward the allele we would expect based on parental expression or in the opposite direction than expected. For example, contrasting allele specific expression in the F1 hybrids for gene GRMZM2G140082 we observed that both Mo17 and PH207 alleles showed biased expression when compared to B73, whereas the Oh43 allele showed equivalent expression in the cold condition. The observed bias expression of Mo17 and PH207 alleles in the hybrids was in the direction we would predict given parental expression values (Figure 8B). Another example of allelic bias was observed for gene GRMZM2G007957, which showed equivalent parental expression for B73, Mo17, and PH207, but higher expression of Oh43 when exposed to heat. The observed proportion of parental alleles was reflected of allelic ratios observed in the hybrids (Figure 8C). Hundreds of genes exhibit variation in stress induced expression between two parents (Figure 9). Overall, more genes have altered allelic proportions in the F1 between control and stress treatments, which would suggest cis- regulatory variation (Figure 9). Of the genes that exhibit allelic bias 84-96% were in the direction we would predict based on parental expression.

We were interested in contrasting patterns of expression, rate of retention of retained syntenic duplicates, and annotation of genes that exhibit cis- or trans- regulatory variation for stress response. Previous work has shown that genes that exhibit steady-state changes and cis- regulatory variation patterns tend to have a higher fold change between inbred (Stupar and Springer 2006). Comparing the fold change between stress and control treatments for genes that have cis- or trans- regulatory variation patterns allowed us to test whether this pattern holds true for stress induced expression. A non-redundant list of cis- and trans- regulatory variation was created by combining all genes categorized as cis- or trans- from all three trio comparisons. In contrast to steady-state expression patterns, we do not observe a bias of genes that have strong expression variation also exhibiting cis- regulatory variation for any of the comparisons made in this experiment (Figure 10). Interestingly, in the contrast of B73 and Oh43 we observed a higher mean and larger variation for genes that exhibit trans- regulatory variation (Figure 10).

We also assessed the rate of retention and stress induced expression variation for maize paralogs. Maize underwent a whole genome duplicate event approximately 6-12 million years ago (Gaut and Doebley 1997). Subsequent fractionation and rearrangements have resulted in different levels of loss of paralogs across the genome. Genes that exhibit cis- or trans- regulatory variation retain paralogs at a similar rate (44-49%) to that of other expressed maize genes (44%, Figure 11A). We were interested in how often the paralogs of genes with cis- or trans- regulatory variation were also differentially expressed. No difference in the rate of paralogs for genes with cis- or trans-regulatory variation and the reference gene set was observed (Figure 11B). Many transcription factor families were enriched in our stress-induced expression gene sets, so we were interested in which TFs that show stress responsive cis- or trans- regulatory variation. Although some TF families overlap between cis- and trans- genes (EREB and Orphan), a subset of important abiotic stress response TF families (MYB and NAC) were only present in the cis- regulatory variation category (Table 6).

DISCUSSION

Differential expression of alleles is a major component of phenotypic variance and evolution. Differential regulation of these alleles could be the result of changes the proximal regulatory regions (cis-), factors that can influence many genes by providing binding sites or distally (trans-), or a combination. Trans- acting factors, like TFs, can influence the expression of many downstream genes. This study aimed to assess the regulatory variation for steady-state expression levels as well as expression responses to heat or cold stress in several maize genotypes. Thousands of genes exhibit differences in steady-state expression within a condition between two inbreds. We assessed the contribution of cis- and trans- regulatory variation for this subset of genes. Hundreds of genes exhibit expression variation for stress response between two inbreds. Through assessing the allele-specific expression levels of these genes in the F1 hybrids we can monitor the contribution of cis- or trans- regulatory variation for stress response.

Cis- regulatory variation is prevalent for both steady-state and stress-induced expression variation

Regulatory variation for steady-state changes in expression has been documented in maize (Stupar and Springer 2006; Li et al., 2013; Swanson-Wagner et al., 2009; Lemmon et al., 2014). Here we show that thousands of genes have variable levels of expression between inbreds in any of our conditions. Cis- regulatory variation is more prevalent and often stable across treatments. There are more examples of trans-regulatory variation that shift to other regulatory types in other conditions, but overall these calls are also fairly stable.

Genes that showed variable stress-induced expression among different genotypes were identified to study regulatory variation for stress response. As noted for steady-state expression variation, there is a bias toward cis- regulatory variation. For responsiveness variation the genes with cis- or trans- regulatory variation show similar fold change expression patterns, which is not true for steady-state expression variation. Although trans- regulatory elements are important components of stress response

pathways, the goal and focus of future experiments as a result of this work would be to understand how alleles could lose or gain stress-induced expression through cis-variation.

Using retained syntenic duplicates to infer ancestral state of stress response

Expression variation in response to stress conditions can be the result of an allele gaining or losing stress-induced expression. Assessing whether the responsiveness variation represents a loss or gain of responsive expression can be difficult. One method would be to assess responsive expression of orthologous genes in closely related species. If the orthologs are DE in response to stress in other species then alleles that do not exhibit altered expression would likely represent a loss of stress-induced expression as the ancestral state is likely to be responsive. Alternatively, if the orthologs in related species do not have altered expression in response to the same stress then we would infer that the allelic variation in maize is due to the gain of expression response for some maize alleles. Using gene expression responses in related species can be complicated by the fact that related species have different morphologies and stress sensitivities. An alternative method for inferring loss or gain of responsive expression is to utilize the expression responses for retained syntenic duplicates. A whole genome duplication event in maize resulted in two subgenomes (Gaut and Doebley 1997; Schnable et al., 2011). A subset of maize genes have retained both syntenic paralogs (Schnable et al., 2011), which can be used to assess the likely ancestral state for expression responses. If one gene exhibits allelic variation with some alleles responding to a stress and other alleles that do not respond we can look at the responses of the retained syntenic duplicate. If the retained duplicate responds then we infer that some alleles have likely lost the expression response. Alternatively, if the retained duplicate does not respond to the abiotic stress we would infer that some alleles for the gene with a variable response have gained novel responsiveness. In our experiment 49% of genes with cis- regulatory variation for stress response had a retained syntenic duplicate, of which, 88% of the paralogs were also DE in the same direction (up- or down-) under the same treatment suggesting that alleles

have lost stress induced expression (Figure 11). Cis- regulatory variation genes for which the paralog does not respond in any of the other genotypes could be examples of alleles gaining stress induced expression.

Potential mechanism for loss of stress-induced expression

Genes with differential expression responses to stress among two inbreds that are attributed to cis- regulatory variation can be used to study regulatory regions or elements that contribute to gene expression responses to abiotic stress. Cis-acting sequences often act by providing binding sites for transcription factors that can regulate expression. We identified several cis-acting regulatory regions that are enriched in genes that respond to heat and/or cold stress in B73 including ABRE, DRE/CRT, and HSE sites. Mutations in these binding regions such as these could result in the loss of stress-induced expression for some inbreds. For gene GRMZM2G119258, B73 is the only genotype tested that has stress-induced expression and also has an intact ABRE element upstream of the gene, whereas, Mo17 and Oh43 have a SNP within the element. In order to test whether a SNP in this element resulted in the loss of stress-induced expression, one could cluster genes into haplotype groups based on the nucleotide at this SNP and assess expression of these lines under control and stress conditions. Screening and assessing differential expression across a panel of diverse lines for polymorphisms in cis- regulatory elements would shed light on important cis- elements as well as frequency of responsiveness for these genes.

Potential mechanism for gain of stress-induced expression

Alternatively, alleles could gain stress-induced expression. One novel source of variation for differential expression is the insertion of TEs nearby genes. Certain families of TEs are associated with stress-induced expression in maize in addition to expression and phenotypic diversity across many plant species (Makarevitch et al., 2015; Lisch et al., 2013). Different TE families are associated with induced expression under different stress treatments (Makarevitch et al., 2015). The insertion of particular TE families putatively provides novel cis- binding regions, so alleles with the TE are more highly

expressed under stress conditions compared to those without the insertion (Makarevitch et al., 2015).

Characterizing variation in regulatory regions could shed light on important stress response pathways and has the potential to help engineer more stress tolerant genotypes. Here we characterized expression and regulatory variation for a diverse set of inbred lines in heat and cold conditions as well as propose methods for assessing loss or gain of stress-induced expression. Studying regulatory mechanisms on a large scale is quite complicated, but improvements in sequencing technology, annotation of genomes, and novel analytic tools are providing the platform to begin understanding complex regulatory mechanisms.

MATERIALS AND METHODS

Stress conditions, tissue collection, and RNA extraction: Inbred and hybrid maize seedlings were grown under 16 hours of light at 24°C in growth chambers for 14 days. Plants were cold stressed, heat stressed, or remained in control conditions. Plants that were cold stressed were placed in a cold room (50°C) for 16 hours. Heat stress plants were placed in an incubator (50°C) for 4 hours. Above ground seedling tissue was collected immediately following each stress treatment and 6 plants were pooled per sample. Biological replicates were grown five days apart for all treatments. After tissue collection each pooled sample was ground using liquid nitrogen. Total RNA was extracted in Trizol (Life Technologies, NY, USA) and then purified with LiCl.

Sequencing and Data Processing: RNA-seq libraries were prepared by the University of Minnesota Genomics Center in accordance with TruSeq library creation protocol (Illumina, San Diego, CA). Samples were sequenced using the Illumina HiSeq-2000 platform developing 15-3 0 million reads per sample. All replicates and conditions of B73 were aligned to the maize reference genome (B73 AGPv2). To improve mapping quality of the non-reference inbreds SNP corrected references were created for Mo17, PH207, and Oh43. Each sample of Mo17, PH207, and Oh43 was aligned to their

respective corrected reference. All alignments were made using Tophat 2 (Kim et al., 2013) and bowtie2 (Langmead and Salzberg 2012). To keep alignments consistent across all samples only reads that mapped without any mismatches were retained and used for counts and rpm calculations (Tophat2 option -N 0). Bamtools was used to further filter out reads that mapped with indels or to multiple places in the genome (Barnett et al., 2011). Hybrid samples were aligned to both parental references using the same criteria (e.g. B73xOh43 was aligned to both B73 and Oh43 corrected references). To obtain ASE counts, reads from the two alignments for each hybrid were compared and reads that mapped uniquely to one parent were retained. All read counts from inbred parents and the ASE counts from the hybrids were run through HTSeq (Anders et al., 2015) to get per gene counts and per gene ASE counts, respectively. Reads per million (rpm), normalized expression values, were calculated for all inbred parents by using the count files from HTSeq. Hybrid rpms were calculated by summing the parent specific reads and the reads that mapped equally to both parents.

GO analysis: Genes that were up- regulated in response to cold or heat for each genotype were independently run through BinGO plugin from Cytoscape to assess enrichment of biological processes (Maere et al., 2005). The top Arabidopsis BLAST hit for each gene was used to run the enrichment. The entire list of the enriched biological processes for each genotype was combined and a non-redundant (NR) list was created. The corrected p-value for each inbred was pulled for the top 25 most significant NR GO categories.

Transcription factor and motif enrichment: The entire set of characterized maize transcription factors (TFs, Yilmaz et al., 2008) were used test whether certain TF families were associated with stress-induced differentially expressed genes. Enrichment of TFs were calculated by contrasting the number of genes that are up- regulated in at least 3 genotypes under cold or heat treatments was contrasted to the expected number (ratio of genes that are up- regulated in at least three genotypes in the entire filtered gene set). A

chi-square test was used to see which families were statistically enriched (X^2 <0.01) in our up- regulated gene sets.

The Find Individual Motif Occurrences (FIMO, Grant et al., 2011) tool, which is part of the MEME suite was used to count the number of DRE/CRT, ABRE, and HSE motifs throughout the B73 reference genome. The number of motifs per bin upstream, downstream, and within each was calculated by summing the number of motifs that overlap the specified bin using the BEDTools (Quinlan and Hall 2010) window tool. Each motif was counted per bin 5-2.5 kb, 2.5-1kb, and 1kb 5' and 3' and within the gene for up- regulated genes in response to cold or heat and across the entire filtered gene set. Enrichment of motifs was calculated for each bin by contrasting the proportion of up regulated genes that have at least one motif in a particular bin to the proportion of genes that have at least one motif in that bin in the genome. A chi-square test was run to see if the observed number of motifs per bin was statistically different from expected.

Differential expression: Raw read counts were run through the DESeq package in R (Anders and Huber 2010) to run statistical tests for differential expression between control and stress (stress induced DE) for each parent comparing control to cold or heat as well as between inbred genotypes within each condition (genotype DE). Differentially expressed genes were identified by having an adjusted p-value (FDR) less than 0.01 as calculated by DESeq, an rpm greater than 0.5 for both parents (genotype DE) or in control and stress (stress induced DE), and at least a 2 fold difference in expression between genotypes (genotype DE) or between control and stress (stress induced DE). Genes that turn "on" or "off" in response to stress have an FDR less than 0.01 and exhibit a tenfold difference in expression.

Identifying cis- and trans- regulatory variation

Genotypic regulatory variation: Each gene that is differentially expressed between two inbred parents was run through independent chi-square tests to test for cis- and trans-regulatory variation. Genes that exhibit cis- regulatory variation will show a biased

allelic proportion in the F_1 where the expected value is the total number of allele specific reads for a given condition multiplied by the proportion of the parental alleles and the observed values are the number of allele specific read counts for both alleles in the F_1. The test for trans- regulatory variation used the total number of allele specific reads in the hybrid divided by two for the expected value and the same observed values as the test for cis- regulatory variation. Genes that are not statistically different than expected for cis- or trans- regulatory variation patterns ($X > 0.01$) were categorized as examples of cis- or trans- regulatory variation.

Stress-responsive regulatory variation: Genes that exhibited variation in response to stress between two inbreds were used to assess the contribution of cis- and trans- regulatory variation. A chi-square test was applied to these genes sets to test for trans- regulatory variation patterns. If the proportion of alleles in the control F_1 sample was not statistically different from the proportion in the stress F_1 then those genes would be classified as trans- regulatory variation. Genes whose allelic proportion differed between the two F_1 samples would be classified as cis- regulatory variation examples. Genes with allelic bias were further binned into genes whose allelic bias is in the direction we would predict given parental expression values or toward the opposite allele.

REFERENCES

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., & Shinozaki, K. (1997). Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. The Plant Cell, 9(10), 1859–68. http://doi.org/10.1105/tpc.9.10.1859

Anders, S., Pyl, P. T., & Huber, W. (2014). HTSeq - A Python framework to work with high-throughput sequencing data. Bioinformatics (Oxford, England), 31(2), 166–9. http://doi.org/10.1093/bioinformatics/btu638

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology, 11(10), R106. http://doi.org/10.1186/gb-2010-11-10-r106

Ashraf, M., & Hafeez, M. (2004). Thermotolerance of pearl millet and maize at early growth stages : growth and nutrient relations (Vol. 48, pp. 81–86).

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics (Oxford, England), 27(12), 1691–2. http://doi.org/10.1093/bioinformatics/btr174

Chen, M., Wang, Q.-Y., Cheng, X.-G., Xu, Z.-S., Li, L.-C., Ye, X.-G., et al Ma, Y.-Z. (2007). GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. Biochemical and Biophysical Research Communications, 353(2), 299–305. http://doi.org/10.1016/j.bbrc.2006.12.027

Chia, J.-M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J., et al Ware, D. (2012). Maize HapMap2 identifies extant variation from a genome in flux. Nature Genetics, 44(7), 803–7. http://doi.org/10.1038/ng.2313

Crafts-Brandner, S. J., & Salvucci, M. E. (2002). Sensitivity of photosynthesis in a C4 plant, maize, to heat stress. Plant Physiology, 129(4), 1773–80. http://doi.org/10.1104/pp.002170

Cubillos, F. A., Stegle, O., Grondin, C., Canut, M., Tisné, S., Gy, I., & Loudet, O. (2014). Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis. The Plant Cell, 26(11), 4298–310. http://doi.org/10.1105/tpc.114.130310

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al Yamaguchi-Shinozaki, K. (2003). OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene

expression. The Plant Journal, 33(4), 751–763. http://doi.org/10.1046/j.1365-313X.2003.01661.x

Gaut, B. S., & Doebley, J. F. (1997). DNA sequence evidence for the segmental allotetraploid origin of maize. Proceedings of the National Academy of Sciences, 94(13), 6809–6814. http://doi.org/10.1073/pnas.94.13.6809

Gilmour, S. J. (2000). Overexpression of the Arabidopsis CBF3 Transcriptional Activator Mimics Multiple Biochemical Changes Associated with Cold Acclimation. PLANT PHYSIOLOGY, 124(4), 1854–1865. http://doi.org/10.1104/pp.124.4.1854

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics (Oxford, England), 27(7), 1017–8. http://doi.org/10.1093/bioinformatics/btr064

Guiltinan, M., Marcotte, W., & Quatrano, R. (1990). A plant leucine zipper protein that recognizes an abscisic acid response element. Science, 250(4978), 267–271. http://doi.org/10.1126/science.2145628

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., & Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. International Journal of Molecular Sciences, 14(5), 9643–84. http://doi.org/10.3390/ijms14059643

Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., et al Buell, C. R. (2014). Insights into the maize pan-genome and pan-transcriptome. The Plant Cell, 26(1), 121–35. http://doi.org/10.1105/tpc.113.119982

Holloway, B., Luck, S., Beatty, M., Rafalski, J.-A., & Li, B. (2011). Genome-wide expression quantitative trait loci (eQTL) analysis in maize. BMC Genomics, 12(1), 336. http://doi.org/10.1186/1471-2164-12-336

Horvath, D. (1993). Regulation of Arabidopsis thaliana L. Heyn) cor78 in response to low temperature. PLANT PHYSIOLOGY, 103(4), 1047–1053. http://doi.org/10.1104/pp.103.4.1047

Jaglo-Ottosen, K. R. (1998). Arabidopsis CBF1 Overexpression Induces COR Genes and Enhances Freezing Tolerance. Science, 280(5360), 104–106. http://doi.org/10.1126/science.280.5360.104

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology, 14(4), R36. http://doi.org/10.1186/gb-2013-14-4-r36

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–9. http://doi.org/10.1038/nmeth.1923

Lemmon, Z. H., Bukowski, R., Sun, Q., & Doebley, J. F. (2014). The role of cis regulatory evolution in maize domestication. PLoS Genetics, 10(11), e1004745. http://doi.org/10.1371/journal.pgen.1004745

Li, L., Petsch, K., Shimizu, R., Liu, S., Xu, W. W., Ying, K., et al Muehlbauer, G. J. (2013). Mendelian and non-Mendelian regulation of gene expression in maize. PLoS Genetics, 9(1), e1003202. http://doi.org/10.1371/journal.pgen.1003202

Liu, S., Wang, X., Wang, H., Xin, H., Yang, X., Yan, J., et al Qin, F. (2013). Genome-Wide Analysis of ZmDREB Genes and Their Association with Natural Variation in Drought Tolerance at Seedling Stage of Zea mays L. PLoS Genetics, 9(9), e1003790. http://doi.org/10.1371/journal.pgen.1003790

Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics (Oxford, England), 21(16), 3448–9. http://doi.org/10.1093/bioinformatics/bti551

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., & Springer, N. M. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genetics, 11(1), e1004915. http://doi.org/10.1371/journal.pgen.1004915

Mittal, D., Enoki, Y., Lavania, D., Singh, A., Sakurai, H., & Grover, A. (2011). Binding affinities and interactions among different heat shock element types and heat shock factors in rice (Oryza sativa L.). The FEBS Journal, 278(17), 3076–85. http://doi.org/10.1111/j.1742-4658.2011.08229.x

Miura, K., & Furumoto, T. (2013). Cold signaling and cold response in plants. International Journal of Molecular Sciences, 14(3), 5312–37. http://doi.org/10.3390/ijms14035312

Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta, 1819(2), 86–96. http://doi.org/10.1016/j.bbagrm.2011.08.004

Nelson, P. T., Coles, N. D., Holland, J. B., Bubeck, D. M., Smith, S., & Goodman, M. M. (2008). Molecular characterization of maize inbreds with expired U.S. plant variety protection. Crop Science, 48(5), 1673–1685. http://doi.org/10.2135/cropsci2008.02.0092

Qu, A.-L., Ding, Y.-F., Jiang, Q., & Zhu, C. (2013). Molecular mechanisms of the plant heat stress response. Biochemical and Biophysical Research Communications, 432(2), 203–7. http://doi.org/10.1016/j.bbrc.2013.01.104

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England), 26(6), 841–2. http://doi.org/10.1093/bioinformatics/btq033

Raghavendra, A. S., Gonugunta, V. K., Christmann, A., & Grill, E. (2010). ABA perception and signalling. Trends in Plant Science, 15(7), 395–401. http://doi.org/10.1016/j.tplants.2010.04.006

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2006). Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. Proceedings of the National Academy of Sciences of the United States of America, 103(49), 18822–7. http://doi.org/10.1073/pnas.0605639103

Schramm, F., Larkindale, J., Kiehlmann, E., Ganguli, A., Englich, G., Vierling, E., & von Koskull-Döring, P. (2008). A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of Arabidopsis. The Plant Journal : For Cell and Molecular Biology, 53(2), 264–74. http://doi.org/10.1111/j.1365-313X.2007.03334.x

Schnable, J. C., Springer, N. M., & Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences of the United States of America, 108(10), 4069–74. http://doi.org/10.1073/pnas.1101368108

Stupar, R. M., Gardiner, J. M., Oldre, A. G., Haun, W. J., Chandler, V. L., & Springer, N. M. (2008). Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biology, 8(1), 33. http://doi.org/10.1186/1471-2229-8-33

Stupar, R. M., & Springer, N. M. (2006). Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics, 173(4), 2199–210. http://doi.org/10.1534/genetics.106.060699

Swanson-Wagner, R. A., DeCook, R., Jia, Y., Bancroft, T., Ji, T., Zhao, X., et al Schnable, P. S. (2009). Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids. Science (New York, N.Y.), 326(5956), 1118–20. http://doi.org/10.1126/science.1178294

Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in Plant Science, 9(5), 244–52. http://doi.org/10.1016/j.tplants.2004.03.006

Xiao, H., & Lis, J. (1988). Germline transformation used to define key features of heat-shock response elements. Science, 239(4844), 1139–1142. http://doi.org/10.1126/science.3125608

Yadav, S. K. (2010). Cold stress tolerance mechanisms in plants. A review. Agronomy for Sustainable Development, 30(3), 515–527. http://doi.org/10.1051/agro/2009050

Yamaguchi-Shinozaki, K., & Shinozaki, K. (1993). Characterization of the expression of a desiccation-responsive rd29 gene of Arabidopsis thaliana and analysis of its promoter in transgenic plants. Molecular & General Genetics : MGG, 236(2-3), 331–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8437577

Yamaguchi-Shinozaki, K., & Shinozaki, K. (1994). A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. The Plant Cell, 6(2), 251–64. http://doi.org/10.1105/tpc.6.2.251

Yilmaz, A., Nishiyama, M. Y., Fuentes, B. G., Souza, G. M., Janies, D., Gray, J., & Grotewold, E. (2009). GRASSIUS: a platform for comparative regulatory genomics across the grasses. Plant Physiology, 149(1), 171–80. http://doi.org/10.1104/pp.108.128579

Table 1. Sequencing depth for samples used in this study

Sample	Batch	Treatment	Replicate	Total Read Pairs
B73	1	Control	1	13748589
B73	1	Control	2	33557107
B73	1	Control	3	40175033
B73	1	Cold	1	11271326
B73	1	Cold	2	32590508
B73	1	Cold	3	44422177
B73	1	Heat	1	8627575
B73	1	Heat	2	40804694
B73	1	Heat	3	35587015
B73	2	Control	1	20418064
B73	2	Control	2	22461858
B73	2	Control	3	15763628
B73	2	Cold	1	22931175
B73	2	Cold	2	28224749
B73	2	Cold	3	13983222
B73	2	Heat	1	22278888
B73	2	Heat	2	18177209
B73	2	Heat	3	20560891
Mo17	2	Control	1	22078493
Mo17	2	Control	2	18600605
Mo17	2	Control	3	20938964
Mo17	2	Cold	1	29222649
Mo17	2	Cold	2	16926075
Mo17	2	Cold	3	18356387
Mo17	2	Heat	1	16339495
Mo17	2	Heat	2	12766970
Mo17	2	Heat	3	22386299
PH207	2	Control	1	16535313
PH207	2	Control	2	20237611
PH207	2	Control	3	22673151
PH207	2	Cold	1	24284778
PH207	2	Cold	2	27650600
PH207	2	Cold	3	24528419
PH207	2	Heat	1	16270762
PH207	2	Heat	2	27072347

Sample	Batch	Treatment	Replicate	Total Read Pairs
PH207	2	Heat	3	22115560
Oh43	1	Control	1	10466243
Oh43	1	Control	2	27403769
Oh43	1	Control	3	28129362
Oh43	1	Cold	1	9925157
Oh43	1	Cold	2	32210203
Oh43	1	Cold	3	32586844
Oh43	1	Heat	1	10466243
Oh43	1	Heat	2	30385623
Oh43	1	Heat	3	32486290
B73xMo17	2	Control	1	45003370
B73xMo17	2	Control	2	42462573
B73xMo17	2	Control	3	44475994
B73xMo17	2	Cold	1	57871903
B73xMo17	2	Cold	2	51889655
B73xMo17	2	Cold	3	40771421
B73xMo17	2	Heat	1	46364994
B73xMo17	2	Heat	2	42139533
B73xMo17	2	Heat	3	47862136
B73xPH207	2	Control	1	46111079
B73xPH207	2	Control	2	38246329
B73xPH207	2	Control	3	46963457
B73xPH207	2	Cold	1	44393869
B73xPH207	2	Cold	2	45502576
B73xPH207	2	Cold	3	51216447
B73xPH207	2	Heat	1	51850349
B73xPH207	2	Heat	2	46996865
B73xPH207	2	Heat	3	33552577
Oh43xB73	1	Control	1	14741105
Oh43xB73	1	Control	2	41580255
Oh43xB73	1	Control	3	32222079
Oh43xB73	1	Cold	1	10807124
Oh43xB73	1	Cold	2	44318286
Oh43xB73	1	Cold	3	47143405
Oh43xB73	1	Heat	1	10602433
Oh43xB73	1	Heat	2	47209241
Oh43xB73	1	Heat	3	38456627

Table 2. GO enrichments for up- regulated genes in response to cold treatment

GO description	B73 batch1	B73 batch2	Mo17	Oh43	PH207
response to stimulus	8.35E-37	1.00E-17	8.98E-19	6.51E-31	9.37E-15
response to chemical stimulus	1.10E-30	7.22E-19	1.27E-13	6.70E-24	8.55E-12
multicellular organismal development	1.67E-13	1.93E-13	1.81E-28	2.28E-30	1.07E-25
multicellular organismal process	1.90E-13	4.15E-13	5.24E-28	2.44E-30	1.48E-27
developmental process	3.64E-13	5.42E-12	1.94E-27	3.22E-30	1.39E-23
regulation of cellular metabolic process	1.63E-26	2.34E-16	6.03E-16	7.80E-13	4.83E-17
biological regulation	1.77E-24	3.38E-19	4.02E-23	1.51E-15	1.04E-25
regulation of cellular process	1.90E-25	7.74E-17	7.80E-19	4.76E-14	1.13E-19
cellular process	2.59E-08	1.04E-03	2.28E-14	3.89E-25	6.84E-13
regulation of nitrogen compound metabolic process	1.51E-24	9.60E-16	8.90E-16	4.61E-13	2.92E-16
regulation of biosynthetic process	1.77E-24	1.17E-15	3.72E-14	3.71E-13	1.32E-13
regulation of cellular biosynthetic process	1.77E-24	1.17E-15	3.72E-14	3.71E-13	1.32E-13
anatomical structure development	3.17E-08	2.10E-09	2.32E-22	8.78E-24	1.55E-21
regulation of biological process	2.02E-23	2.70E-18	8.02E-22	4.28E-13	1.03E-23
regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1.20E-23	1.17E-15	2.19E-15	1.69E-12	4.82E-15
regulation of macromolecule biosynthetic process	1.39E-23	1.03E-15	5.44E-14	1.43E-12	6.34E-13
regulation of transcription	1.62E-23	1.17E-15	6.00E-13	2.05E-11	6.84E-13
regulation of primary metabolic process	2.02E-23	4.60E-16	3.79E-16	1.43E-12	5.26E-17
regulation of metabolic process	2.43E-23	2.51E-15	1.81E-16	7.03E-12	4.09E-19
response to organic substance	8.23E-23	7.22E-19	3.72E-14	4.67E-15	8.76E-14
response to abiotic stimulus	3.79E-21	1.44E-14	5.76E-12	7.00E-22	1.67E-10
regulation of macromolecule metabolic process	1.09E-20	2.03E-14	4.05E-15	8.41E-12	1.56E-16

GO description	B73 batch1	B73 batch2	Mo17	Oh43	PH207
regulation of gene expression	2.28E-20	1.08E-13	3.72E-13	6.32E-11	3.44E-14
response to stress	1.50E-19	7.66E-08	4.66E-08	2.06E-17	8.94E-06
nucleic acid metabolic process	n.s.	n.s.	1.51E-19	1.94E-11	1.22E-09

Values shown are FDR corrected p-values

Table 3. Enrichment of GO categories for genes that are up- regulated in response to heat stress

GO description	B73 batch1	B73 batch2	Mo17	Oh43	PH207
response to stimulus	2.55E-32	9.00E-16	1.96E-14	3.53E-33	5.35E-14
response to heat	3.76E-30	1.74E-08	4.68E-10	3.57E-27	n.s.
response to chemical stimulus	9.36E-30	8.53E-12	9.97E-14	2.50E-28	3.09E-13
response to stress	6.40E-25	1.08E-07	2.30E-11	2.89E-28	2.20E-06
response to temperature stimulus	3.10E-23	1.04E-06	2.56E-07	3.57E-27	n.s.
response to abiotic stimulus	4.54E-26	1.33E-09	1.13E-10	1.18E-25	2.29E-07
protein folding	5.47E-17	n.s.	3.24E-05	3.75E-15	n.s.
response to organic substance	5.76E-16	1.47E-09	6.75E-11	1.45E-13	1.03E-12
biological regulation	3.91E-05	1.33E-09	4.19E-15	3.40E-06	8.02E-11
cellular process	1.06E-03	n.s.	4.19E-15	8.20E-10	2.82E-03
DNA metabolic process	n.s.	n.s.	4.19E-15	1.28E-04	n.s.
regulation of biological process	5.75E-03	5.43E-07	4.19E-15	3.21E-05	7.07E-08
response to water deprivation	2.53E-10	8.66E-04	1.21E-04	7.29E-15	5.21E-03
multicellular organismal process	n.s.	2.38E-14	3.67E-11	8.83E-04	1.61E-04
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	n.s.	n.s.	2.88E-14	n.s.	n.s.
response to water	6.40E-10	1.28E-03	2.58E-04	3.00E-14	7.38E-03
nucleic acid metabolic process	n.s.	n.s.	6.36E-14	n.s.	n.s.
nitrogen compound metabolic process	n.s.	3.24E-03	7.37E-14	n.s.	1.82E-02
cellular nitrogen compound metabolic process	n.s.	7.52E-03	8.02E-14	n.s.	n.s.
multicellular organismal development	4.26E-02	6.58E-13	1.13E-10	1.33E-03	6.59E-04
anatomical structure development	n.s.	8.01E-13	3.93E-12	7.79E-03	1.77E-04
regulation of metabolic process	2.54E-03	1.85E-04	9.62E-13	5.30E-04	1.49E-05

GO description	B73 batch1	B73 batch2	Mo17	Oh43	PH207
developmental process	n.s.	1.14E-12	1.33E-10	3.11E-04	2.72E-04
regulation of cellular process	4.31E-03	2.23E-05	2.17E-12	1.40E-05	2.95E-08
regulation of cell cycle	n.s.	n.s.	3.06E-12	3.38E-05	n.s.

Values shown are FDR corrected p-values

Table 4. Proportion of genes with significant genotype, condition, or genotype by condition interactions

	B73/Mo17		B73/PH207		B73/Oh43	
	Cold	Heat	Cold	Heat	Cold	Heat
Genotype	0.22	0.23	0.20	0.20	0.43	0.40
Environment	0.21	0.24	0.16	0.24	0.41	0.52
Genotype by Environment	0.01	0.02	0.01	0.01	0.20	0.28

Significant: FDR corrected p-value <0.05

Table 5. Distribution of cis- and trans- regulatory variation classifications for genes DE between genotypes within a condition

	B73/Mo17			B73/PH207			B73/Oh43		
	Control	Cold	Heat	Control	Cold	Heat	Control	Cold	Heat
Extreme allelic bias	303	317	306	261	316	269	277	382	368
cis- regulatory variation	1053	1164	1135	1170	1188	1126	1306	1623	1226
cis- and trans-	1422	1244	1294	873	1097	1199	1710	1366	1126
trans- regulatory variation	395	418	341	211	306	328	1362	681	710
unknown	154	144	133	119	109	121	365	159	239

Table 6. Stress response genes that are characterized transcription factors

Gene ID	Cis- or Trans-	Transcription Factor
GRMZM2G025812	cis	ZmbZIP112
GRMZM2G180979	cis	ZmC3H17
GRMZM2G180430	cis	ZmDBP2
GRMZM2G057386	cis	ZmEREB107
GRMZM2G457562	cis	ZmEREB113
GRMZM2G021573	cis	ZmEREB161
GRMZM2G146028	cis	ZmEREB8
GRMZM2G082264	cis	ZmGLK21
GRMZM2G168002	cis	ZmGLK49
GRMZM2G098986	cis	ZmMADS77
GRMZM2G000818	cis	ZmMYB11
GRMZM2G001875	cis	ZmMYB131
GRMZM2G701063	cis	ZmMYB2
GRMZM2G051793	cis	ZmMYB45
GRMZM2G139760	cis	ZmMYBR61
GRMZM2G079632	cis	ZmNAC4
GRMZM2G054277	cis	ZmNAC63
GRMZM2G136765	cis	ZmOrphan102
GRMZM2G444075	cis	ZmOrphan137
GRMZM2G008482	cis	ZmOrphan184
GRMZM2G462417	cis	ZmZHD18
AC203957.3_FG004	trans	ZmbZIP40
GRMZM2G012717	trans	ZmCOL10
GRMZM2G076896	trans	ZmEREB111
GRMZM2G031983	trans	ZmGATA15
GRMZM2G068672	trans	ZmHB60
GRMZM2G430871	trans	ZmOFP27
GRMZM2G476637	trans	ZmOrphan274

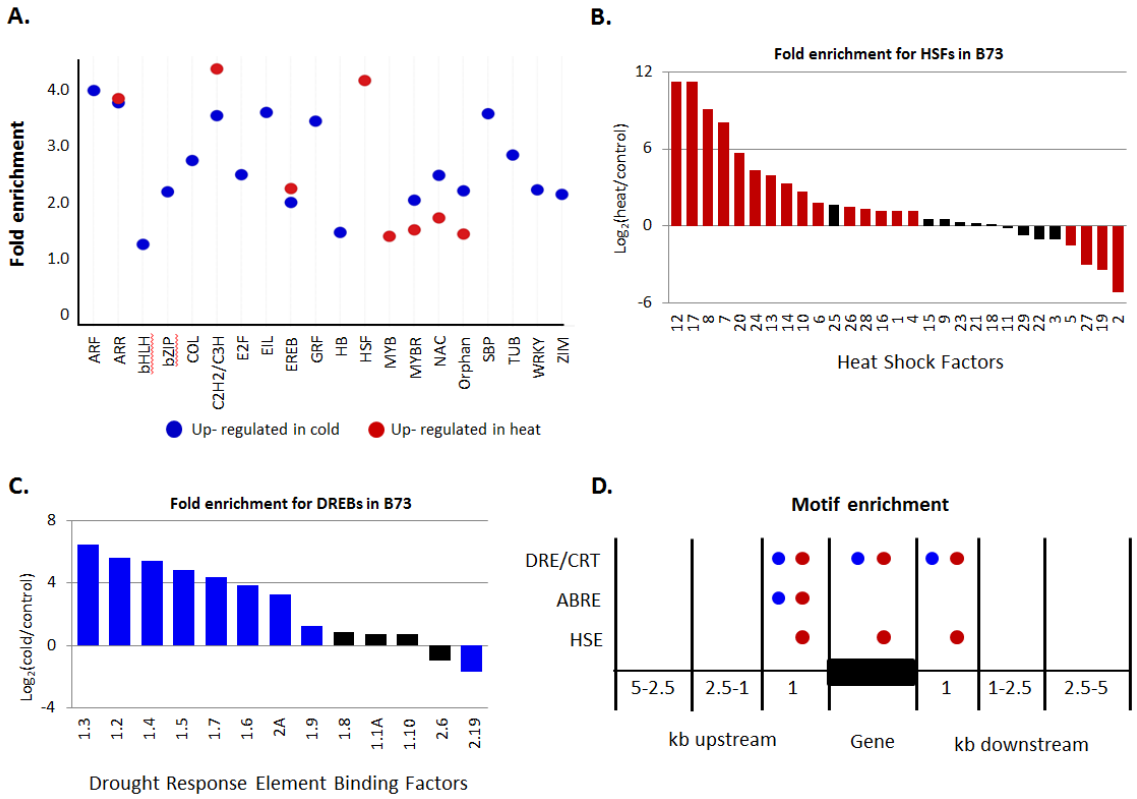


Figure 1. A subset of transcription factors and regulatory motifs are enriched in cold and heat DE gene sets. A. Maize transcription factors (TFs) were obtained from Grassius (Yilmaz et al., 2008) and assessed in the lists of genes that are up-regulated in cold or heat treatments. The observed number of up-regulated genes (in at least 3 genotypes) per TF family was contrasted to the number expected (#TF per family * proportion of FGS genes up-regulated in at least 3 genotypes). The fold enrichment (log2(observed/expected)) for the 20 TF families that have a statistically higher number of up-regulated genes than predicted (X^2 <0.01) for cold (blue) or heat (red) are shown. B. The log2(heat/control) of expression values are shown for all 29 characterized maize heat shock factors (HSF, Yilmaz et al., 2008). Nineteen of the HSFs have statistically significant (FDR p-value <0.01) differential expression between control and heat (red bars). C. Maize orthologs of DREB (TF family EREB, Lui et al., 2013) that had expression data in our experiment are shown as log2(cold/control) expression values. The 8 DREBs that exhibit significant increases in expression in cold treatments are indicated by blue bars. D. The number of drought response element (DRE), ABA response element (ABRE), and heat shock elements (HSE) sequences were counted for genes that are up- regulated in response to cold or heat as well as the filtered gene set (FGS) for windows flanking genes. Enrichment for these motifs in genes responsive to cold or heat stress was assessed by contrasting the frequency of observed motifs in each region, with the frequency of these motifs in the same region in the entire FGS. Motifs that are enriched (X^2 < 0.01) in each region of the cold (blue) or heat (red) responsive genes are shown.

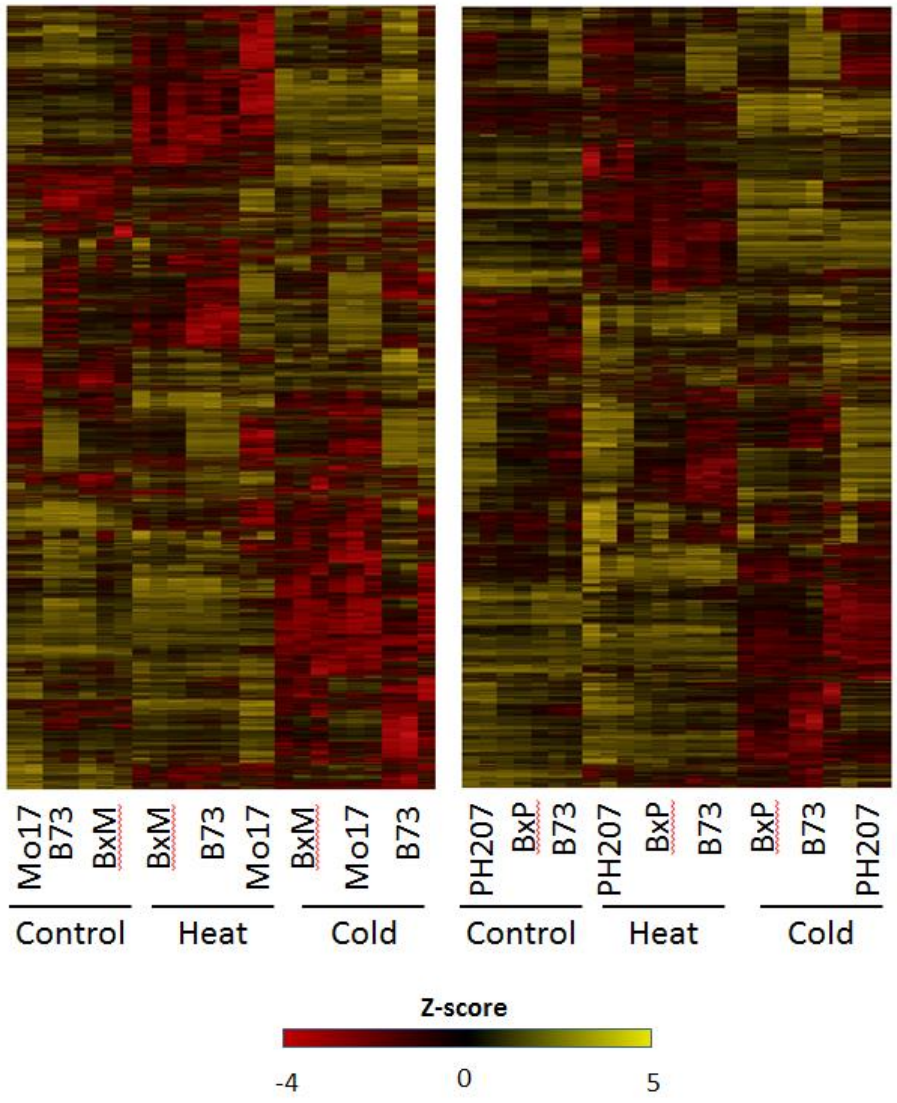


Figure 2- Transcriptional responses in maize seedlings exposed to heat or cold treatment. Genes that are expressed (rpm >0.5) in at least one genotype were used to perform hierarchal clustering (UMPGA method) for the two trios of genotypes used in batch 2. Expression is represented as a z-score normalized rpm and high expression (red), low expression (yellow), or average expression (black) can be differentiated. Although some genes respond similarly across genotypes or treatments many examples of variable expression were observed.

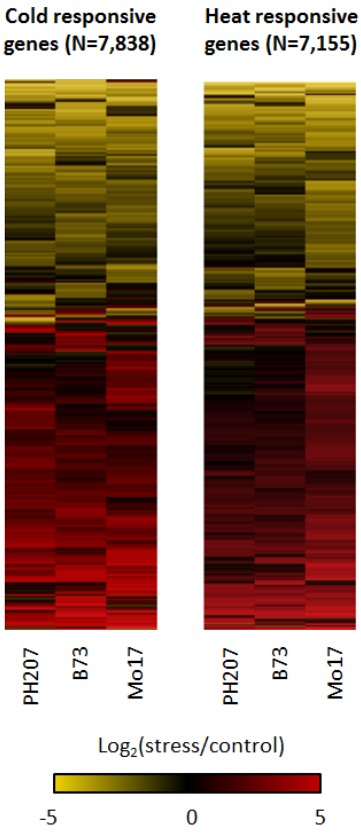


Figure 3. Clustering of cold and heat responsive genes. After exposure to cold or heat stress many genes are differentially expressed (DE) in each inbred. The non-redundant set of cold (left) or heat (right) responsive genes from genotypes in batch 2 were used to perform hierarchal clustering (UPGMA method) using the gene expression response values (log_2 stress rpm divided by control rpm). Genes that are up-regulated (red) or down- regulated (yellow) in response to stress in at least one inbred comparison are visualized. A subset of genes are DE in all three inbred comparisons, but some genes only have strong responses in one or two genotypes and have no change in gene expression (black) in other genotypes.

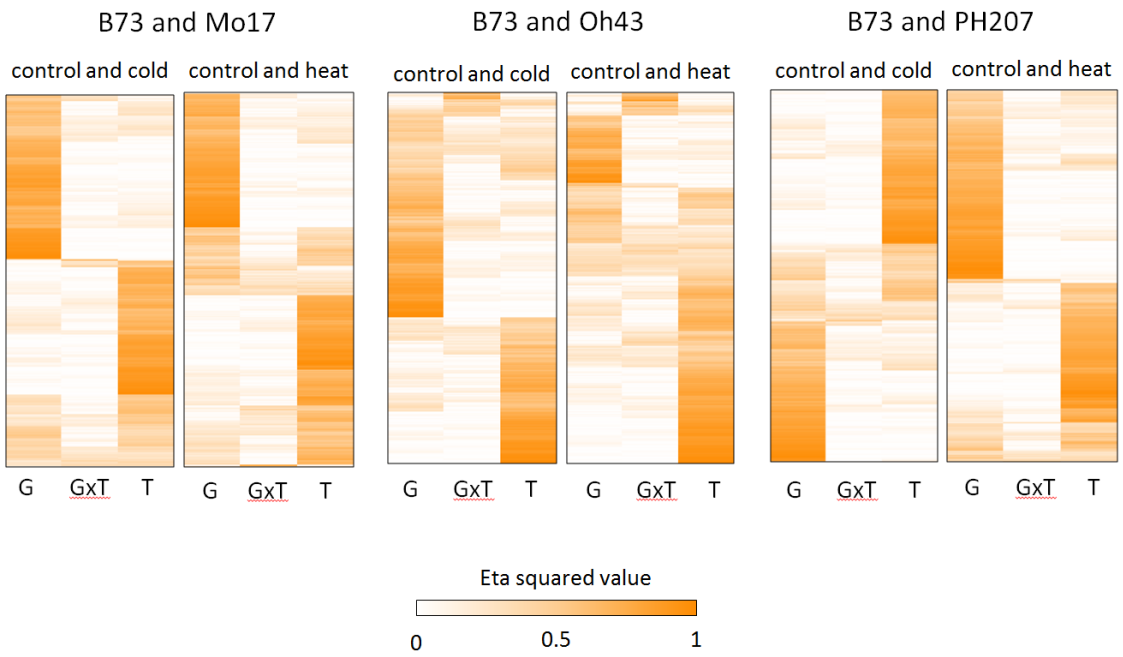


Figure 4. Genotype or treatment explain most of the variation observed in ANOVA analysis. Eta squared values were calculated to determine the amount of variation explained by genotype (G), treatment (T), and genotype by treatment (GxT) factors in our ANOVA. The proportion of variance explained by each factor is shown on a scale from 0-1 (white to orange). Hierarchal clustering (UPGMA method) of genes that have at least one significant factor (FDR corrected p-value <0.05) when contrasting two inbred parents B73 and Mo17 (left) B73 and Oh43 (middle) and B73 and Ph207 (right) in control and cold or heat conditions are shown.

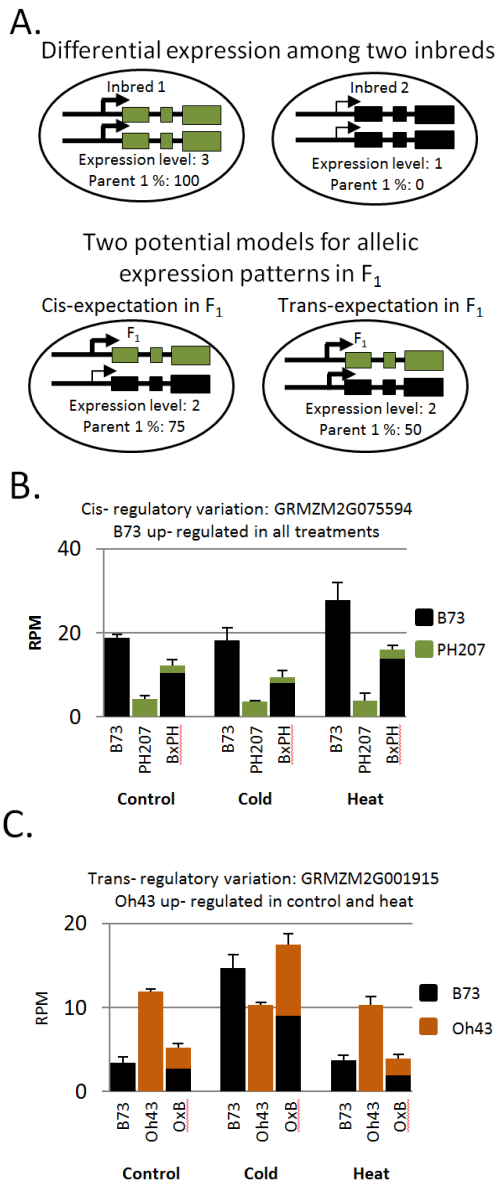


Figure 5. Genes with cis- or trans- regulatory variation between genotypes within a condition were observed A. Genes with cis- and trans- regulatory variation between inbred parents will exhibit different allelic ratios in the F1 hybrids. Cis- and trans- regulatory variation was assessed for genes that are differentially expressed between the two inbred parents. Here the size of the arrows is representative of level of transcription of that gene, where a thicker arrow is indicative of higher expression. In this example the allele of inbred 1 is more highly expressed than the allele of inbred 2. Predictions of allelic proportion in the F1 hybrid are different for cis- and trans- regulatory variation. If a gene exhibits cis- regulatory variation then the allelic ratio in the F1 would be biased toward inbred 1. Whereas, if the gene exhibited trans- regulatory variation both alleles from the parents would be equivalent. The reads per million (RPMs) for a trio of samples across all treatments is shown for an example of cis- (B) and trans- (C) regulatory variation. B. Shows an example of cis- regulatory variation where the B73 (black) allele is preferentially expressed in all conditions as compared to the PH207 allele (green). C. An example of trans- regulatory variation for a gene where the Oh43 allele (orange) is more highly expressed in control and heat treatments. However, in the F1 plants the B73 (black) and Oh43 (orange) alleles exhibit equivalent expression levels.

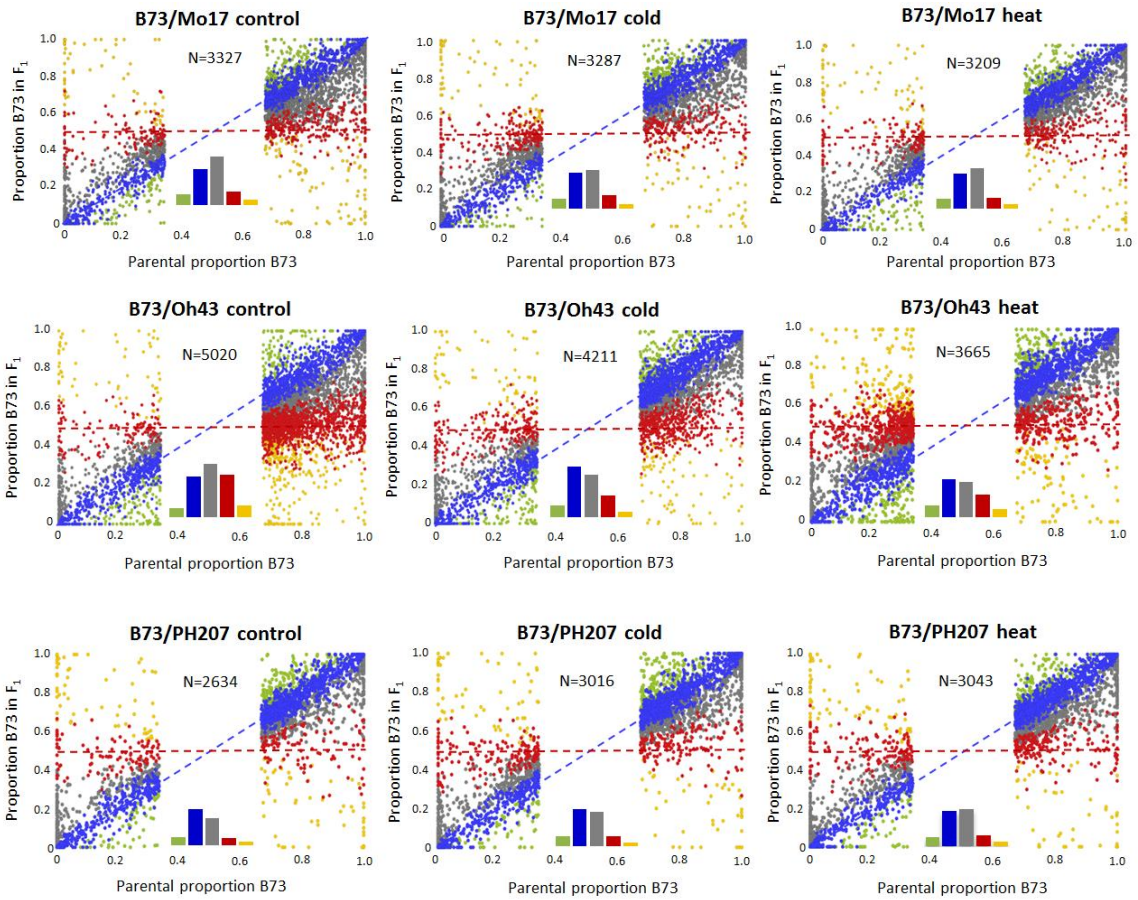


Figure 6. Classification of differentially expressed genes into regulatory variation categories. Genes that are differentially expressed (DE) between inbreds were identified for each condition. Each scatterplot shows the entire set of DE genes between B73 and Mo17 (top), B73 and Oh43 (middle), and B73 and PH207 (bottom) for control (left), cold (middle), and heat (right). Each DE gene was classified into one of 5 regulatory variation categories based on a comparison between the expected allelic proportion calculated from parental expression values (x-axis) and the observed allelic proportion in the F1 hybrid (y- axis). Genes with purely cis-regulatory variation are expected to plot along the diagonal blue dashed line while genes subject to trans-regulatory variation will plot along the horizontal red dashed line. The inset bar charts in each scatterplot show the distribution of genes in five different categories. These include genes that are not statistically distinguishable from cis- (blue) or trans- (red) regulatory variation expectations. In addition, some genes exhibit a greater allelic bias in the expected direction (green) or a bias toward the opposite allele (yellow) than predicted by parental ratios. The remaining genes that did not fit the statistical cutoffs for cis- or trans- regulatory variation and have an observed allelic proportion that falls between expected cis- and trans- expected values (grey).

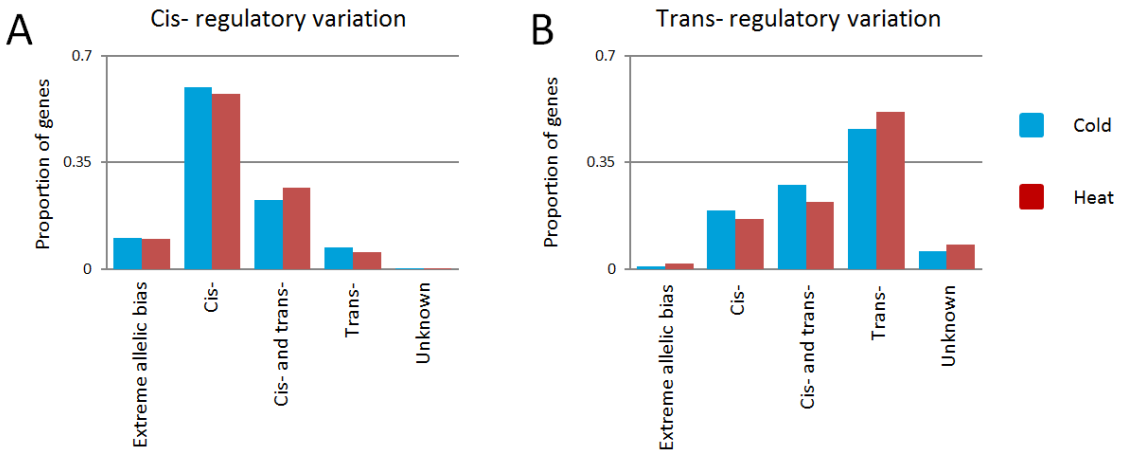


Figure 7. Conservation of cis- and trans- regulatory variation classifications. Genes that are DE in all three conditions for at least one of the three genotype comparisons were used to assess the stability of cis- and trans- regulatory variation categories across stress treatments. The distribution of regulatory variation categories in cold (blue) and heat (red) treatments were assessed for genes that were classified as cis- (A) or trans- (B) in the control treatment. The categories are based off statistical criteria where some genes exhibit more allelic bias than predicted (extreme allelic bias) or in the opposite direction (unknown). Cis- and trans- category are statistically significant from both the cis- and trans- regulatory variation expectations, but fall between the two expectations.

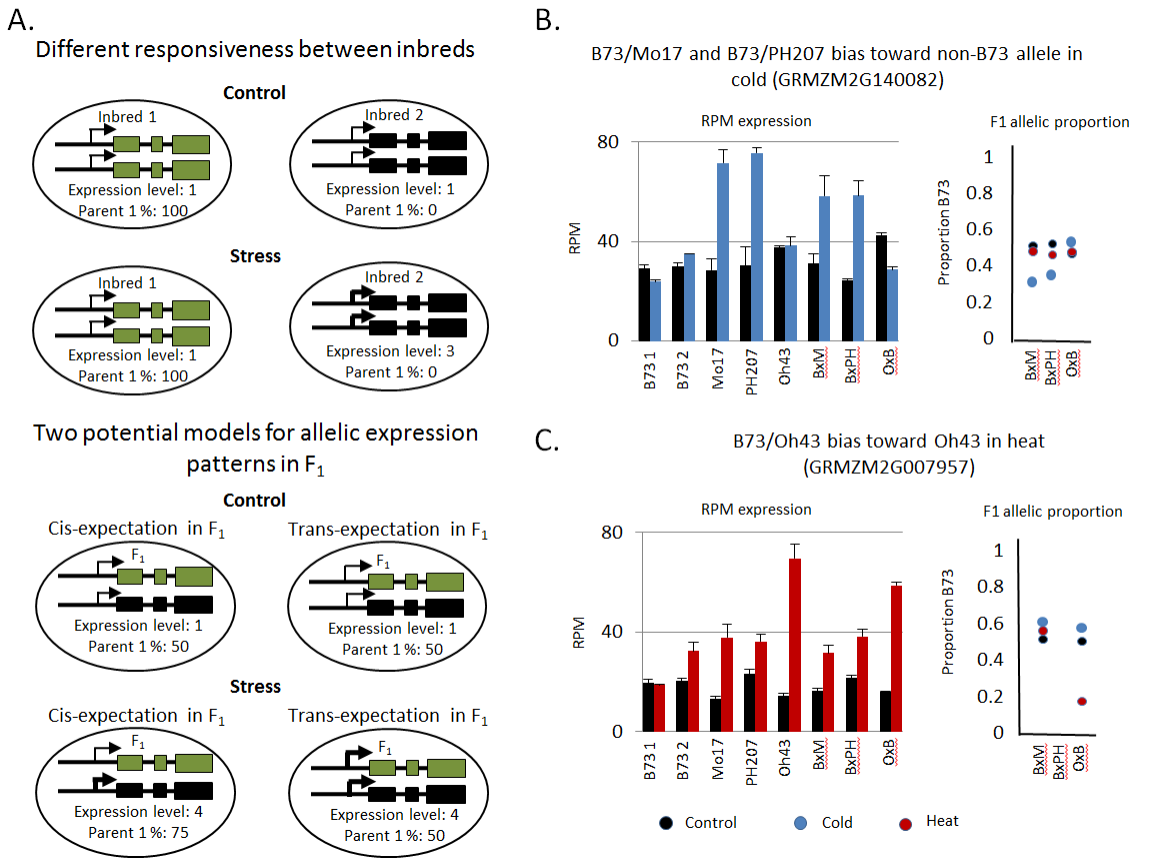


Figure 8. Examples of responsiveness cis- and trans- regulatory variation. Responsiveness regulatory variation candidates were identified by assessing genes that have variable response to stress between two inbreds (thickness of arrows). Genes with cis- or trans- regulatory variation for responsiveness to an environmental stimulus are expected to have distinct patterns of allelic expression levels in F1 plants. Cis-regulatory responsiveness variation would result in approximately equally expression of both alleles in control conditions, but biased allelic expression in stress conditions (thicker arrow). Conversely, if a gene exhibits trans- regulatory responsiveness variation then both parental alleles would be equivalently expressed in both control and stress conditions. B. Shows the expression (RPMs) for all genotypes in control (black) and cold (blue) conditions as well as the proportion of B73 in the F1 hybrids for a gene that is biased toward Mo17 and PH207 alleles under cold conditions. The biased expression of the Mo17 and PH207 alleles under cold treatment in the F1 is consistent with cis- regulatory variation for response to cold. C. Shows the expression levels and proportion of B73 allele for a gene that is bias toward Oh43 when contrasting B73 and Oh43 under heat conditions. The similar ratio of B73 and Mo17 in B73xMo17 and biased expression on Oh43 in the Oh43xB73 hybrid fits the predictions based on parental expression.

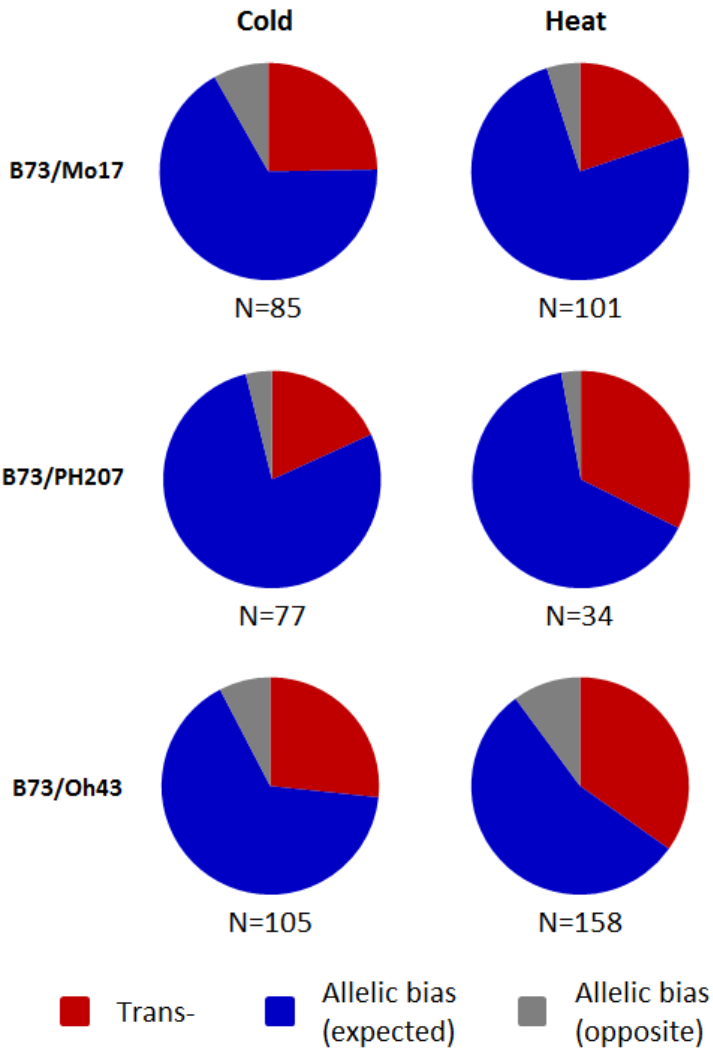


Figure 9. Proportion of genes that fit statistical criteria for cis- and trans- regulatory variation in response to stress. All genes that fit the criteria for allelic variation for response to stress for each trio (gene number listed under the pie charts) were classified into one of three categories. The proportions of genes that fit into each category are shown in the pie charts. A majority of genes exhibit biased allelic proportions (cis- regulatory variation, blue and grey). Of the genes that show allelic bias, a majority exhibited a bias in the direction we would predict given parental expression values (blue). A small proportion of genes showed an allelic bias toward the opposite parental allele (grey). A moderate proportion of genes exhibit trans- regulatory variation patterns (red).

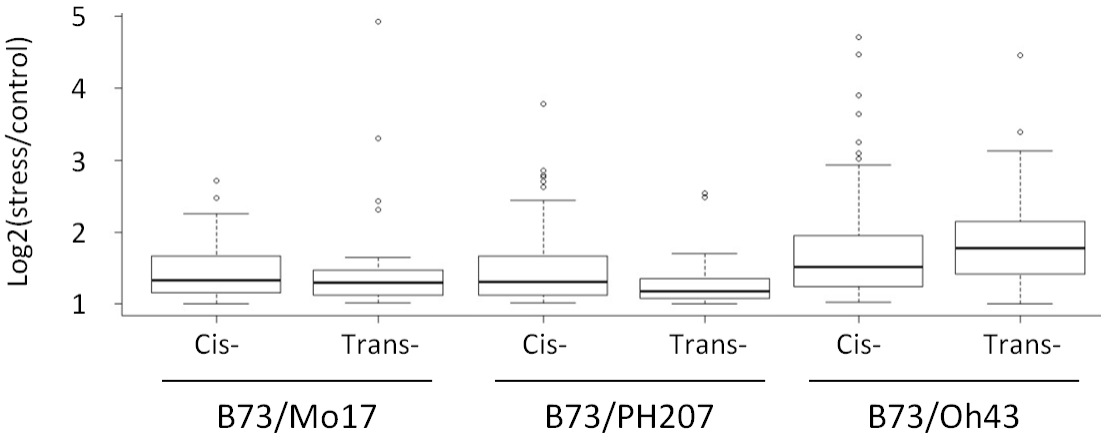


Figure 10- Cis- and trans- regulatory variation genes exhibit similar fold change between stress and control. The entire set of genes that exhibits cis- or trans- regulatory variation for stress response for B73 and Mo17 (left), B73 and PH207 (middle) and B73 and Oh43 (right) was used to contrast fold change (log2) values between the two regulatory variation categories. The absolute log2 of stress rpm divided by control rpm for the inbred that increased in response to stress for each gene in the trio is show. Similar means and variation are observed between cis- and trans- regulatory variation genes in each trio.

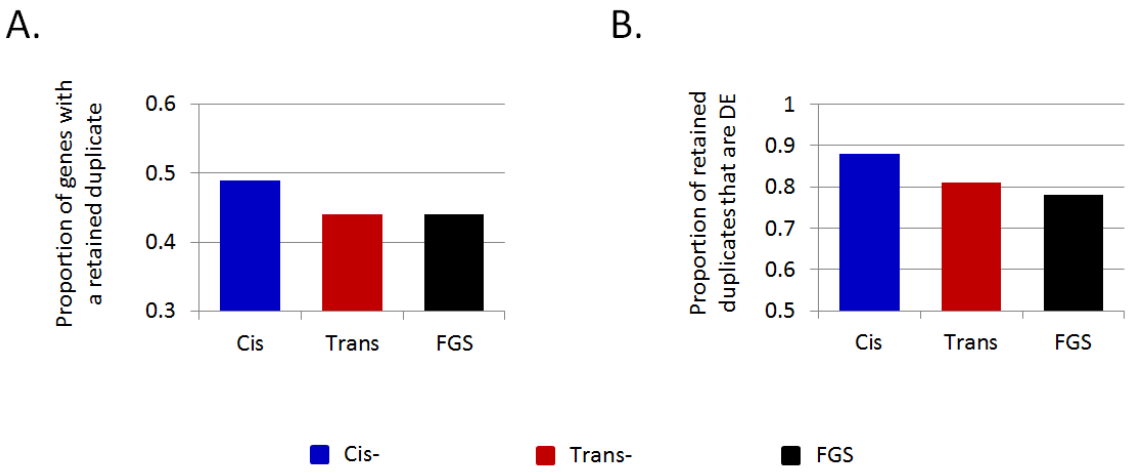


Figure 11. Paralogs of genes with cis- or trans- regulatory variation in response to stress are not DE more than expected. A. Maize paralogs (as reported in Schnable et al., 2009) were pulled for genes that are expressed, have at least 25 allele specific expression reads, and are DE between control and stress in at least one genotype. The proportion of genes that have a retained duplicate are shown for genes in the filtered gene set (FGS, black), classified as cis- (blue), or trans- (red) regulatory variation. B. The proportion of paralogs in (A) that are also DE in the same direction (up- or down-) under the same stress conditions are plotted for the three categories: FGS (black), cis- (blue), and trans- regulatory variation (red).

BIBLIOGRAPHY

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., & Shinozaki, K. 1997. Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. The Plant Cell, 910, 1859–68. http://doi.org/10.1105/tpc.9.10.1859

Albert, F. W., & Kruglyak, L. 2015. The role of regulatory variation in complex traits and disease. Nature Reviews Genetics, 164, 197–212. http://doi.org/10.1038/nrg3891

Anders, S., Pyl, P. T., & Huber, W. 2014. HTSeq - A Python framework to work with high-throughput sequencing data. Bioinformatics Oxford, England, 312, 166–9. http://doi.org/10.1093/bioinformatics/btu638

Anders, S., & Huber, W. 2010. Differential expression analysis for sequence count data. Genome Biology, 1110, R106. http://doi.org/10.1186/gb-2010-11-10-r106

Ashraf, M., & Hafeez, M. 2004. Thermotolerance of pearl millet and maize at early growth stages : growth and nutrient relations Vol. 48, pp. 81–86.

Autran, D., Baroux, C., Raissig, M. T., Lenormand, T., Wittig, M., Grob, S., et al Grossniklaus, U. 2011. Maternal epigenetic pathways control parental contributions to Arabidopsis early embryogenesis. Cell, 1455, 707–19. http://doi.org/10.1016/j.cell.2011.04.014

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics Oxford, England, 2712, 1691–2. http://doi.org/10.1093/bioinformatics/btr174

Baroux, C., Gagliardini, V., Page, D. R., & Grossniklaus, U. 2006. Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis. Genes & Development, 209, 1081–6. http://doi.org/10.1101/gad.378106

Bartolomei, M. S., & Ferguson-Smith, A. C. 2011. Mammalian genomic imprinting. Cold Spring Harbor Perspectives in Biology, 37, a002592–. http://doi.org/10.1101/cshperspect.a002592

Bauer M.J. and Fischer R.L. 2011. Genome demethylation and imprinting in the endosperm. Curr. Opin. Plant Biol. 14: 162-167.

Berger F, Vu TM, Li J, Chen B. 2012. Hypothesis: Selection of imprinted genes is driven by silencing deleterious gene activity in somatic tissues. Cold Spring Harb Symp Quant Biol.

Bernardi J, Lanubile A, Li QB, Kumar D, Kladnik A, Cook SD, Ross JJ, Marocco A, Chourey PS. 2012. Impaired auxin biosynthesis in the defective endosperm18 mutant is due to mutational loss of expression in the ZmYuc1 gene encoding endosperm-specific YUCCA1 protein in maize. Plant Physiol 1603:1318-28.

Boyer, J. S., Byrne, P., Cassman, K. G., Cooper, M., Delmer, D., Greene, T., et al Gaffney, J. 2013. The U.S. drought of 2012 in perspective: A call to action. Global Food Security, 23, 139–143. http://doi.org/10.1016/j.gfs.2013.08.002

Bushell C., Spielman M., and Scott R.J. 2003. The basis of natural and artificial postzygotic hybridization barriers in Arabidopsis species. The Plant Cell. 15:1430-1442.

Calarco J.P., Borges F, Donoghue M.T.A., Van Ex F., Jullien P.E., Lopes T., Gardner R., Berger F., Feijó J.A., Becker J.D., Martienssen R.A. 2012 Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell. 1511: 194-205.

Carroll, S. B. 2005. Evolution at two levels: on genes and form. PLoS Biology, 37, e245. http://doi.org/10.1371/journal.pbio.0030245

Chaudhuri S. and Messing J. 1994. Allele-specific parental imprinting of dzr1, a posttranscriptional regulator of zein accumulation. Proc Natl Acad Sci U S A 9111:4867-71.

Chen, L., Song, Y., Li, S., Zhang, L., Zou, C., & Yu, D. 2012. The role of WRKY transcription factors in plant abiotic stresses. Biochimica et Biophysica Acta, 18192, 120–8. http://doi.org/10.1016/j.bbagrm.2011.09.002

Chen, M., Wang, Q.-Y., Cheng, X.-G., Xu, Z.-S., Li, L.-C., Ye, X.-G., et al Ma, Y.-Z. 2007. GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. Biochemical and Biophysical Research Communications, 3532, 299–305. http://doi.org/10.1016/j.bbrc.2006.12.027

Chia J.M., Song C., Bradbury P.J., Costich D., de Leon N., Doebley J., Elshire R.J., Gaut B., Geller L., Glaubitz J.C., et al. 2012. Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 447:803-7.

Chiesa N., De Crescenzo A., Mishra K., Perone L., Carella M., Palumbo O., Mussa A., Sparago A., Cerrato F., Russo S., Lapi E., Cubellis M.V., Kanduri C., Silengo M.C.,

Riccio A., and Ferrero G.B. 2012. The KCNQ10T1 imprinting control region and non-coding RNA: new properties derived from the study of Beckwidth-Wiedemann syndrome and Silver-Russell syndrome cases. Human Molecular Genetics 21:10-25.

Costa L.M., Yuan J., Rouster W.P., Dickinson H., and Gutierrez-Marcos J.F., Maternal control of nutrient allocation in plant seeds by genomic imprinting. Current Biology 22:160-165.

Crafts-Brandner, S. J., & Salvucci, M. E. 2002. Sensitivity of photosynthesis in a C4 plant, maize, to heat stress. Plant Physiology, 1294, 1773–80. http://doi.org/10.1104/pp.002170

Cubillos, F. A., Stegle, O., Grondin, C., Canut, M., Tisné, S., Gy, I., & Loudet, O. 2014. Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis. The Plant Cell, 2611, 4298–310. http://doi.org/10.1105/tpc.114.130310

Danilevskaya O.N., Hermon P., Hantke S., Muszynski M.G., Kollipara K., and Ananiev E.V. 2003. Duplicated fie genes in maize: expression pattern and imprinting suggest distinct functions. Plant Cell 15: 425-438.

Das R, Hampton DD, Jirtle RL. 2009. Imprinting evolution and human health. Mamm Genome 209-10:563-72.

Deng X, Disteche CM 2010 Genomic Responses to Abnormal Gene Dosage: The X Chromosome Improved on a Common Strategy. PLoS Biol 82.

Dilkes, B. P., & Comai, L. 2004. A differential dosage hypothesis for parental effects in seed development. The Plant Cell, 1612, 3174–80. http://doi.org/10.1105/tpc.104.161230

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al Yamaguchi-Shinozaki, K. 2003. OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. The Plant Journal, 334, 751–763. http://doi.org/10.1046/j.1365-313X.2003.01661.x

Eichten, S. R., Ellis, N. A., Makarevitch, I., Yeh, C.-T., Gent, J. I., Guo, L., et al Springer, N. M. 2012. Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS Genetics, 812, e1003127. http://doi.org/10.1371/journal.pgen.1003127

Eichten S.E., Swanson-Wagner R.A., Schnable J.C., Waters A.J., Hermanson P.J., Liu S., Yeh C-T., Jia Y., Gendler K., Freeling M., Schnable P.S., Vaughn M.W., and Springer

N.M. 2011 Heritable epigenetic variation among maize inbreds. PLoS Genetics. 7(11): e1002372. doi:10.1371/journal.pgen.1002372.

Ewing B. and Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8: 186-194.

Ewing B., Hillier L., Wendl M.C., and Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8: 175-185.

Feil R. and Berger F. 2007. Convergent evolution of genomic imprinting in plants and mammals. Trends Genet. 23: 192-199.

Feng S., Jacobsen S.E., Reik W. 2010. Epigenetic reprogramming in plant and animal development. Science 3306004:622-7.

Fitz Gerald, J. N., Hui, P. S., & Berger, F. 2009. Polycomb group-dependent imprinting of the actin regulator AtFH5 regulates morphogenesis in Arabidopsis thaliana. Development Cambridge, England, 13620, 3399–404. http://doi.org/10.1242/dev.036921

Gaut, B. S., & Doebley, J. F. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proceedings of the National Academy of Sciences, 9413, 6809–6814. http://doi.org/10.1073/pnas.94.13.6809

Gehring M., Missirian V., Henikoff S. 2011. Genomic analysis of parent-of-origin allelic expression in arabidopsis thaliana seeds. PLoS One 68:e23687.

Gehring M, Bubb KL, Henikoff S. 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 3245933:1447-51.

Gehring M., Huh J.H., Hsieh T.F., Penterman J., Choi Y., Harada J.J., Goldberg R.B., and Fischer R.L. 2006. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124: 495-506.

Gehring M., Choi Y., and Fischer R.L. 2004. Imprinting and seed development. Plant Cell 16 Suppl: S203-13.

Gilmour, S. J. 2000. Overexpression of the Arabidopsis CBF3 Transcriptional Activator Mimics Multiple Biochemical Changes Associated with Cold Acclimation. PLANT PHYSIOLOGY, 1244, 1854–1865. http://doi.org/10.1104/pp.124.4.1854

Gilmour, S. J., Hajela, R. K., & Thomashow, M. F. 1988. Cold Acclimation in Arabidopsis thaliana. PLANT PHYSIOLOGY, 873, 745–750. http://doi.org/10.1104/pp.87.3.745

Goodrich J., Puangsomlee P., Martin M., Long D., Meyerowitz E.M., and Coupland G. 1997. A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature 386: 44-51.

Grant, C. E., Bailey, T. L., & Noble, W. S. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics Oxford, England, 277, 1017–8. http://doi.org/10.1093/bioinformatics/btr064

Greb T., Mylne J.S., Crevillen P., Geraldo N., An H., Gendall A.R., and Dean C. 2007. The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. Curr. Biol. 17: 73-78.

Grennan, A. K. 2006. Abiotic stress in rice. An “omic” approach. Plant Physiology, 1404, 1139–41. http://doi.org/10.1104/pp.104.900188

Grimanelli D., Perotti E., Ramirez J., and Leblanc O. 2005. Timing of the maternal-to-zygotic transition during early seed development in maize. Plant Cell 17: 1061-1072.

Grossniklaus U., Vielle-Calzada J.P., Hoeppner M.A., and Gagliano W.B. 1998. Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science 280: 446-450.

Guiltinan, M., Marcotte, W., & Quatrano, R. 1990. A plant leucine zipper protein that recognizes an abscisic acid response element. Science, 2504978, 267–271. http://doi.org/10.1126/science.2145628

Guo M., Rupe M.A., Danilevskaya O.N., Yang X., and Hu Z. 2003. Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. Plant J. 36: 30-44.

Gutierrez-Marcos J.F., Constancia M., Burton G.J. 2012. Maternal to offspring resource allocation in plants and mammals. Placenta 33 Suppl 2:e3-10.

Gutierrez-Marcos J.F., Costa L.M., Biderre-Petit C., Khbaya B., O'Sullivan D.M., Wormald M., Perez P., and Dickinson H.G. 2004. maternally expressed gene1 Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. Plant Cell 16: 1288-1301.

Gutierrez-Marcos J.F., Pennington P.D., Costa L.M., and Dickinson H.G. 2003. Imprinting in the endosperm: a possible role in preventing wide hybridization. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 358: 1105-1111.

Haig D. 2004. Genomic imprinting and kinship: How good is the evidence? Annu Rev Genet 38:553-85.

Haig D. and Westoby M. 1989. Parent-specific gene-expression and the triploid endosperm. Am. Nat. 134: 147-155.

Hajela, R. K., Horvath, D. P., Gilmour, S. J., & Thomashow, M. F. 1990. Molecular Cloning and Expression of cor Cold-Regulated Genes in Arabidopsis thaliana. Plant Physiology, 933, 1246–52.

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., & Fujita, M. 2013. Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. International Journal of Molecular Sciences, 145, 9643–84. http://doi.org/10.3390/ijms14059643

Haun W.J. and Springer N.M. 2008. Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. Plant J. 56: 903-912.

Haun W.J., Laoueille-Duprat S., O'connell M.J., Spillane C., Grossniklaus U., Phillips A.R., Kaeppler S.M., and Springer N.M. 2007. Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste Mez homologs. Plant J. 49: 325-337.

Hermon P., Srilunchang K.O., Zou J., Dresselhaus T., and Danilevskaya O.N. 2007. Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of the maternal allele. Plant Mol. Biol. 64: 387-395.

Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., et al Buell, C. R. 2014. Insights into the maize pan-genome and pan-transcriptome. The Plant Cell, 261, 121–35. http://doi.org/10.1105/tpc.113.119982

Holloway, B., Luck, S., Beatty, M., Rafalski, J.-A., & Li, B. 2011. Genome-wide expression quantitative trait loci eQTL analysis in maize. BMC Genomics, 121, 336. http://doi.org/10.1186/1471-2164-12-336

Hong, S.-W., & Vierling, E. 2001. Hsp101 is necessary for heat tolerance but dispensable for development and germination in the absence of stress. The Plant Journal, 271, 25–35. http://doi.org/10.1046/j.1365-313x.2001.01066.x

Horvath, D. 1993. Regulation of Arabidopsis thaliana L. Heyn cor78 in response to low temperature. PLANT PHYSIOLOGY, 1034, 1047–1053. http://doi.org/10.1104/pp.103.4.1047

Hsieh T.F., Shin J., Uzawa R., Silva P., Cohen S., Bauer M.J., Hashimoto M., Kirkbride R.C., Harada J.J., Zilberman D., and Fischer R.L. 2011. Regulation of imprinted gene expression in Arabidopsis endosperm. Proc. Natl. Acad. Sci. U. S. A. 108: 1755-1762.

Hsieh T.F., Ibarra C.A., Silva P., Zemach A., Eshed-Williams L., Fischer R.L., and Zilberman D. 2009. Genome-wide demethylation of Arabidopsis endosperm. Science 324: 1451-1454.

Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, et al. 2012. Comparative population genomics of maize domestication and improvement. Nat Genet 447:808-11.

Huh J.H., Bauer M.J., Hsieh T.F., and Fischer R.L. 2008. Cellular programming of plant gene imprinting. Cell 132: 735-744.

Hutter B., Bieg M., Helms V., Paulsen M. 2010. Divergence of imprinted genes during mammalian evolution. BMC Evol Biol 10:116,2148-10-116.

Jaglo-Ottosen, K. R. 1998. Arabidopsis CBF1 Overexpression Induces COR Genes and Enhances Freezing Tolerance. Science, 2805360, 104–106. http://doi.org/10.1126/science.280.5360.104

Jahnke S. and Scholten S. 2009. Epigenetic resetting of a gene imprinted in plant embryos. Curr. Biol. 19: 1677-1681.

Jiang H. and Kohler C. 2012. Evolution, function, and regulation of genomic imprinting in plant seed development. J Exp Bot 6313:4713-22.

Johannes, F., Colot, V., & Jansen, R. C. 2008. Epigenome dynamics: a quantitative genetics perspective. Nature Reviews. Genetics, 911, 883–90. http://doi.org/10.1038/nrg2467

Jullien P.E., Berger F. 2010. Parental Genome Dosage Imbalance Deregulates Imprinting in Arabidopsis. PLoS Genet 63 : e1000885

Jullien P.E. and Berger F. 2009. Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr. Opin. Plant Biol. 12: 637-642.

Jullien P.E., Katz A., Oliva M., Ohad N., and Berger F. 2006. Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis. Curr. Biol. 16: 486-492.

Kermicle J.L. and Alleman M. 1990. Gametic imprinting in maize in relation to the angiosperm life cycle. Dev Suppl :9-14.

Kermicle J.L. 1970. Dependence of the R-Mottled Aleurone Phenotype in Maize on Mode of Sexual Transmission. Genetics 66: 69-85.

Kidwell, M.G. & D. Lisch, 1997. Transposable elements as sources of variation in animals and

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology, 144, R36. http://doi.org/10.1186/gb-2013-14-4-r36

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJ, Koornneef M, Kakutani T. 2007. Control of FWA gene silencing in arabidopsis thaliana by SINE-related direct repeats. Plant J 491:38-45.

Kinoshita T., Miura A., Choi Y., Kinoshita Y., Cao X., Jacobsen S.E., Fischer R.L., and Kakutani T. 2004. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science 303: 521-523.

Kinoshita T., Yadegari R., Harada J.J., Goldberg R.B., and Fischer R.L. 1999. Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. Plant Cell 11: 1945-1952.

Knoll J.H., Cheng S.D., LALANDE M. 1994 Allele specificity of DNA replication timing in the Angelman/ Prader-Willi syndrome imprinted chromosomal region. Nat Genet. 6:41-6

Kohler C., Wolff P., Spillane C. 2012. Epigenetic mechanisms underlying genomic imprinting in plants. Annu Rev Plant Biol 63:331-52.

Kohler C. and Weinhofer-Molisch I. 2010. Mechanisms and evolution of genomic imprinting in plants. Heredity Edinb 1051:57-63.

Kohler C., and Makarevitch G. 2006. Epigenetic mechanisms governing seed development in plants. EMBO Reports. 712: 1223-1227.

Kohler C., Page D.R., Gagliardini V., and Grossniklaus U. 2005. The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nature Genetics. 371: 28-30.

Kohler C., Hennig L., Spillane C., Pien S., Gruissem W., and Grossniklaus U. 2003. The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes Dev. 17: 1540-1553.

Komatsu, S., Kamal, A. H. M., & Hossain, Z. 2014. Wheat proteomics: proteome modulation and abiotic stress acclimation. Frontiers in Plant Science, 5, 684. http://doi.org/10.3389/fpls.2014.00684

Langmead, B., & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods, 94, 357–9. http://doi.org/10.1038/nmeth.1923

Lauria M., Rupe M., Guo M., Kranz E., Pirona R., Viotti A., and Lund G. 2004. Extensive maternal DNA hypomethylation in the endosperm of Zea mays. Plant Cell 16: 510-522.

Lemmon, Z. H., Bukowski, R., Sun, Q., & Doebley, J. F. 2014. The role of cis regulatory evolution in maize domestication. PLoS Genetics, 1011, e1004745. http://doi.org/10.1371/journal.pgen.1004745

Li J. and Berger F. 2012. Endosperm: Food for humankind and fodder for scientific discoveries. New Phytol 1952:290-305.

Li S. and Chou H.H. 2004. LUCY2: an interactive DNA sequence quality trimming and vector removal tool. Bioinformatics 20: 2865-2866.

Li, L., Petsch, K., Shimizu, R., Liu, S., Xu, W. W., Ying, K., et al Muehlbauer, G. J. 2013. Mendelian and non-Mendelian regulation of gene expression in maize. PLoS Genetics, 91, e1003202. http://doi.org/10.1371/journal.pgen.1003202

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. Nature 4306998:471-6.

Liu, Q. 1998. Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low-Temperature-Responsive Gene Expression, Respectively, in Arabidopsis. THE PLANT CELL ONLINE, 108, 1391–1406. http://doi.org/10.1105/tpc.10.8.1391

Liu, S., Wang, X., Wang, H., Xin, H., Yang, X., Yan, J., et al Qin, F. 2013. Genome-Wide Analysis of ZmDREB Genes and Their Association with Natural Variation in Drought Tolerance at Seedling Stage of Zea mays L. PLoS Genetics, 99, e1003790. http://doi.org/10.1371/journal.pgen.1003790

Lu J., Zhang C., Baulcombe D. C., and Chen J. 2012. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. PNAS. 10914: 5529-5534.

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A. 2011. A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. PLoS Genet 76:e1002125.

Luo M., Bilodeau P., Dennis E.S., Peacock W.J., and Chaudhury A. 2000. Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. Proc. Natl. Acad. Sci. U. S. A. 97: 10637-10642.

Maere S., Heymans K., and Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21: 3448-3449.

Makarevich, G., Villar, C. B. R., Erilova, A., & Köhler, C. 2008. Mechanism of PHERES1 imprinting in Arabidopsis. Journal of Cell Science, 121Pt 6, 906–12. http://doi.org/10.1242/jcs.023077

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., Springer, N. M. 2015. Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genetics, 111, e1004915. http://doi.org/10.1371/journal.pgen.1004915

Makarevitch I, Eichten SR, Briskine R, Waters AJ, Danilevskaya ON, Meeley RB, Myers CL, Vaughn MW, Springer NM. 2013. Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. Plant Cell. 25:780-793.

McKeown P.C., Laouielle-Duprat S., Prins P., Wolff P., Schmid M.W., Donoghue M.T., Fort A., Duszynska D., Comte A., Lao N.T., Wennblom T.J., Smant G., Kohler C., Grossniklaus U., and Spillane C. 2011. Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds. BMC Plant. Biol. 11: 113.

Messing J and Dooner HK. 2006. Organization and variability of the maize genome. Curr Opin Plant Biol 92:157-63.

Mickelbart, M. V, Hasegawa, P. M., & Bailey-Serres, J. 2015. Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nature Reviews. Genetics, 164, 237–51. http://doi.org/10.1038/nrg3901

Mittal, D., Enoki, Y., Lavania, D., Singh, A., Sakurai, H., & Grover, A. 2011. Binding affinities and interactions among different heat shock element types and heat shock factors in rice Oryza sativa L.. The FEBS Journal, 27817, 3076–85. http://doi.org/10.1111/j.1742-4658.2011.08229.x

Miura, K., & Furumoto, T. 2013. Cold signaling and cold response in plants. International Journal of Molecular Sciences, 143, 5312–37. http://doi.org/10.3390/ijms14035312

Miyake T, Takebayashi N, Wolf DE. 2009. Possible diversifying selection in the imprinted gene, MEDEA, in arabidopsis. Mol Biol Evol 264:843-57.

Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. 2012. AP2/ERF family transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta, 18192, 86–96. http://doi.org/10.1016/j.bbagrm.2011.08.004

Mosher R.A. and Melnyk C.W. 2010. siRNAs and DNA methylation: seedy epigenetics. Trends Plant Sci. 15: 204-210.

Munné-Bosch, S., & Müller, M. 2013. Hormonal cross-talk in plant development and stress responses. Frontiers in Plant Science, 4, 529. http://doi.org/10.3389/fpls.2013.00529

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. 2012. NAC transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta, 18192, 97–103. http://doi.org/10.1016/j.bbagrm.2011.10.005

Nelson, P. T., Coles, N. D., Holland, J. B., Bubeck, D. M., Smith, S., & Goodman, M. M. 2008. Molecular characterization of maize inbreds with expired U.S. plant variety protection. Crop Science, 485, 1673–1685. http://doi.org/10.2135/cropsci2008.02.0092

O'Brien, J. A., & Benková, E. 2013. Cytokinin cross-talking during biotic and abiotic stress responses. Frontiers in Plant Science, 4, 451. http://doi.org/10.3389/fpls.2013.00451

Ohad N., Yadegari R., Margossian L., Hannon M., Michaeli D., Harada J.J., Goldberg R.B., and Fischer R.L. 1999. Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell 11: 407-416.

Orlando V. 2003. Polycomb, epigenomes, and control of cell identity. Cell 112: 599-606.

Ouyang S., Zhu W., Hamilton J., Lin H., Campbell M., Childs K., Thibaud-Nissen F., Malek R.L., Lee Y., Zheng L., Orvis J., Haas B., Wortman J., and Buell C.R. 2007. The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. 35: D883-7.

Peleg, Z., & Blumwald, E. 2011. Hormone balance and abiotic stress tolerance in crop plants. Current Opinion in Plant Biology, 143, 290–295. http://doi.org/10.1016/j.pbi.2011.02.001

Pignatta, D., Erdmann, R. M., Scheer, E., Picard, C. L., Bell, G. W., & Gehring, M. 2014. Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. eLife, 3, e03198. http://doi.org/10.7554/eLife.03198

Pignatta D., Gehring M. 2012 Imprinting meets genomics: new insights and new challenges. Curr Opin. In Plant Bio. 155: 530-535

Qu, A.-L., Ding, Y.-F., Jiang, Q., & Zhu, C. 2013. Molecular mechanisms of the plant heat stress response. Biochemical and Biophysical Research Communications, 4322, 203–7. http://doi.org/10.1016/j.bbrc.2013.01.104

Quinlan, A. R., & Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics Oxford, England, 266, 841–2. http://doi.org/10.1093/bioinformatics/btq033

Raghavendra, A. S., Gonugunta, V. K., Christmann, A., & Grill, E. 2010. ABA perception and signalling. Trends in Plant Science, 157, 395–401. http://doi.org/10.1016/j.tplants.2010.04.006

Raissig, M. T., Bemer, M., Baroux, C., & Grossniklaus, U. 2013. Genomic imprinting in the Arabidopsis embryo is partly regulated by PRC2. PLoS Genetics, 912, e1003862. http://doi.org/10.1371/journal.pgen.1003862

Raissig M.T., Baroux C., and Grossniklaus U. 2011. Regulation and flexibility of genomic imprinting during seed development. Plant Cell 23: 16-26.

Richards, E. J. 2006. Inherited epigenetic variation--revisiting soft inheritance. Nature Reviews. Genetics, 75, 395–401. http://doi.org/10.1038/nrg1834

Rodrigues J.A., Ruan R., Nishimura T., Sharma M.K., Sharma R., Ronald P.C., Fischer R.L., Zilberman D. 2013. Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proc Natl Acad Sci U S A.

Roudier F., Ahmed I., Berard C., Sarazin A., Mary-Huard T.,Cortijo S., Bouyer D., Caillieux E., Duvernois-Berthet E., Al-Shikhley L., Giraut L., Despres B., Drevensek S., Barneche F., Derozier S., Brunaud V., Aubourg S., Schnittger A., Bowler C., Martin-Magniette M-L., Robin S., Caboche M., and Colot V. 2011. Integrative epigenomic mapping defines four main chromatin states in Arbidopsis. EMBO J. 3010: 1928-1938

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., & Yamaguchi-Shinozaki, K. 2006. Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. Proceedings of the National Academy of Sciences of the United States of America, 10349, 18822–7. http://doi.org/10.1073/pnas.0605639103

Schnable J.C., Freeling M., Lyons E. 2012. Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol Evol 43:265-77.

Schnable, J. C., Springer, N. M., & Freeling, M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences of the United States of America, 10810, 4069–74. http://doi.org/10.1073/pnas.1101368108

Schramm, F., Larkindale, J., Kiehlmann, E., Ganguli, A., Englich, G., Vierling, E., & von Koskull-Döring, P. 2008. A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of Arabidopsis. The Plant Journal : For Cell and Molecular Biology, 532, 264–74. http://doi.org/10.1111/j.1365-313X.2007.03334.x

Sekhon R.S., Briskine R., Hirsch C.N., Myers C.L., Springer N.M., Buell C.R., de Leon N., Kaeppler S.M. 2013. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS One 84:e61005.

Sekhon R.S., Lin H., Childs K.L., Hansey C.N., Buell C.R., de Leon N., and Kaeppler S.M. 2011. Genome-wide atlas of transcription during maize development. Plant J. 66: 553-563.

Selinger D.A. and Chandler V.L. 2001. B-Bolivia, an allele of the maize b1 gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream. Plant Physiol. 125: 1363-1379.

Shi, Y., Mosser, D. D., & Morimoto, R. I. 1998. Molecularchaperones as HSF1-specific transcriptional repressors. Genes & Development, 125, 654–666. http://doi.org/10.1101/gad.12.5.654

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J. D., Feijó, J. A., & Martienssen, R. A. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell, 1363, 461–72. http://doi.org/10.1016/j.cell.2008.12.038

Smyth G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3: Article3.

Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo JM, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive darwinian selection at the imprinted MEDEA locus in plants. Nature. 488:349-52.

Springer N.M. 2009. Small RNAs: how seeds remember to obey their mother. Curr. Biol. 19: R649-51.

Springer N.M., and Gutierrez-Marcos F. 2009. Imprinting in Maize. Maize Handbook-Genetics and Genomics. 2:429-440.

Springer, N. M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., et al Schnable, P. S. 2009. Maize inbreds exhibit high levels of copy number variation CNV and presence/absence variation PAV in genome content. PLoS Genetics, 511, e1000734. http://doi.org/10.1371/journal.pgen.1000734

Stupar, R. M., Gardiner, J. M., Oldre, A. G., Haun, W. J., Chandler, V. L., & Springer, N. M. 2008. Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biology, 81, 33. http://doi.org/10.1186/1471-2229-8-33

Stupar R.M., Hermanson P.J., and Springer N.M. 2007. Nonadditive expression and parent-of-origin effects identified by microarray and allele-specific expression profiling of maize endosperm. Plant Physiol. 145: 411-425.

Stupar, R. M., & Springer, N. M. 2006. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics, 1734, 2199–210. http://doi.org/10.1534/genetics.106.060699

Swanson-Wagner, R. A., Eichten, S. R., Kumari, S., Tiffin, P., Stein, J. C., Ware, D., & Springer, N. M. 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Research, 2012, 1689–99. http://doi.org/10.1101/gr.109165.110

Swanson-Wagner, R. A., DeCook, R., Jia, Y., Bancroft, T., Ji, T., Zhao, X., et al Schnable, P. S. 2009. Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids. Science New York, N.Y., 3265956, 1118–20. http://doi.org/10.1126/science.1178294

Thomashow, M. F. 1999. PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology, 50, 571–599. http://doi.org/10.1146/annurev.arplant.50.1.571

Tiwari, S., Schulz, R., Ikeda, Y., Dytham, L., Bravo, J., Mathers, L., et al Scott, R. J. 2008. MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in Arabidopsis, encodes the conserved C-terminal domain of polyadenylate binding proteins. The Plant Cell, 209, 2387–98. http://doi.org/10.1105/tpc.108.061929

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J., and Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28: 511-515.

Tuteja, N. 2007. Abscisic Acid and abiotic stress signaling. Plant Signaling & Behavior, 23, 135–8.

van Oeveren, J., & Janssen, A. 2009. Single Nucleotide Polymorphisms. In Single Nucleotide Polymorphisms Vol. 578, pp. 73–91. http://doi.org/10.1007/978-1-60327-411-1

Vaughn, M. W., Tanurdzić, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P. D., et al Martienssen, R. A. 2007. Epigenetic natural variation in Arabidopsis thaliana. PLoS Biology, 57, e174. http://doi.org/10.1371/journal.pbio.0050174

Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in Plant Science, 95, 244–52. http://doi.org/10.1016/j.tplants.2004.03.006

Waters, A. J., Bilinski, P., Eichten, S. R., Vaughn, M. W., Ross-Ibarra, J., Gehring, M., & Springer, N. M. 2013. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proceedings of the National Academy of Sciences of the United States of America, 11048, 19639–44. http://doi.org/10.1073/pnas.1309182110

Waters A.J., Makarevitch I., Eichten S.R., Swanson-Wagner R.A., Yeh C-T, Xu W., Schnable P.S., Vaughn M.W., Gehring M., and Springer N.M. 2011. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell. 2312:4221-4233.

Weinhofer I., Hehenberger E., Roszak P., Hennig L., and Kohler C. 2010. H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS Genet. 6: e1001152.

Wolff P., Weinhofer I., Seguin J., Roszak P., Beisel C., Donoghue M.T., Spillane C., Nordborg M., Rehmsmeier M., and Kohler C. 2011. High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. PLoS Genet. 7: e1002126.

Woo H.R., Pontes O., Pikaard C.S., and Richards E.J. 2007. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev. 21: 267-277.

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol 86:e1000409.

Wray, G. A. 2007. The evolutionary significance of cis-regulatory mutations. Nature Reviews. Genetics, 83, 206–16. http://doi.org/10.1038/nrg2063

Wu T.D. and Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873-881.

Xiao, H., & Lis, J. 1988. Germline transformation used to define key features of heat-shock response elements. Science, 2394844, 1139–1142. http://doi.org/10.1126/science.3125608

Yadav, S. K. 2010. Cold stress tolerance mechanisms in plants. A review. Agronomy for Sustainable Development, 303, 515–527. http://doi.org/10.1051/agro/2009050

Yamaguchi-Shinozaki, K., & Shinozaki, K. 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annual Review of Plant Biology, 57, 781–803. http://doi.org/10.1146/annurev.arplant.57.032905.105444

Yamaguchi-Shinozaki, K., & Shinozaki, K. 1994. A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. The Plant Cell, 62, 251–64. http://doi.org/10.1105/tpc.6.2.251

Yamaguchi-Shinozaki, K., & Shinozaki, K. 1993. Characterization of the expression of a desiccation-responsive rd29 gene of Arabidopsis thaliana and analysis of its promoter in transgenic plants. Molecular & General Genetics : MGG, 2362-3, 331–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8437577

Yilmaz, A., Nishiyama, M. Y., Fuentes, B. G., Souza, G. M., Janies, D., Gray, J., & Grotewold, E. 2009. GRASSIUS: a platform for comparative regulatory genomics across the grasses. Plant Physiology, 1491, 171–80. http://doi.org/10.1104/pp.108.128579

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 1781:539-51.

Zemach A., Kim M.Y., Silva P., Rodrigues J.A., Dotson B., Brooks M.D., and Zilberman D. 2010. Local DNA hypomethylation activates genes in rice endosperm. Proc. Natl. Acad. Sci. U. S. A. 107: 18729-18734.

Zhang, M., Xie, S., Dong, X., Zhao, X., Zeng, B., Chen, J., et al Lai, J. 2014. Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. Genome Research, 241, 167–76. http://doi.org/10.1101/gr.155879.113

Zhang M., Zhao H., Xie S., Chen J., Xu Y., Wang K., Zhoa H., Guan H., Hu X., Jiao Y., Song W., and Lai J. 2011 Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. PNAS. 10850: 20042-20047.

Zhang, N., Vierling, E., & Tonsor, S. 2016. Adaptive divergence in transcriptome response to heat and acclimation in Arabidopsis thaliana plants from contrasting climates. bioRxiv. Cold Spring Harbor Labs Journals.