Natural variation for gene expression responses to abiotic stress in maize

Amanda J. Waters¹, Irina Makarevitch², Jaclyn Noshay¹, Liana T. Burghardt¹, Candice N. Hirsch³, Cory D. Hirsch⁴ and Nathan M. Springer¹*¹

¹Department of Plant Biology University of Minnesota, Microbial and Plant Genomics Institute, Saint Paul, MN 55108, USA, ²Department of Biology, Hamline University, Saint Paul, MN 55114, USA, ³Department of Agronomy and Plant Genetics, Microbial and Plant Genomics Institute, St. Paul, MN 55108, USA, and ⁴Department of Plant Pathology, Microbial and Plant Genomics Institute, St. Paul, MN 55108, USA

Received 12 August 2016; revised 26 October 2016; accepted 1 November 2016; published online 7 November 2016.

*For correspondence (e-mail springer@umn.edu).

SUMMARY

Plants respond to abiotic stress through a variety of physiological, biochemical, and transcriptional mechanisms. Many genes exhibit altered levels of expression in response to abiotic stress, which requires 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, PH207 and B37 under control, cold and heat conditions. Large numbers of genes that responded differentially to stress between parental inbred lines were identified. RNA sequencing was also performed on similar tissues of the F₁ hybrids produced by crossing B73 and each of the three other inbred lines. By evaluating allele-specific transcript abundance in the F₁ hybrids, we were able to measure the abundance of cis- and trans-regulatory variation between genotypes for both steady-state and stress-responsive expression differences. Although examples of trans-regulatory variation were observed, cis-regulatory variation was more common for both steady-state and stress-responsive expression differences. The genes with cis-allelic variation for response to cold or heat stress provided an opportunity to study the basis for regulatory diversity.

Keywords: regulatory variation, gene expression, abiotic stress, allele-specific expression, Zea mays.

INTRODUCTION

Plants experience a myriad of abiotic stresses over the course of one life cycle including extreme temperatures, drought, and nutrient deficiencies. Because plants are sessile, they utilize multiple mechanisms to respond to, and survive, in these diverse environments. These stresses can have negative effects on plant health, survival, and in the case of crops, a significant loss of yield. For example, temperature stress has been shown to negatively affect germination, stunt growth, increase leaf chlorosis and necrosis, and reduce yield (Yadav, 2010; Hasanuzzaman et al., 2013). Constitutive responses to adverse environments can also limit growth and fitness (Gilmour, 2000; Dubouzet et al., 2003). Therefore, plants have evolved physiological, biochemical, and transcriptional processes to sense and respond to environmental stresses (Mickelbart et al., 2015). While there are important biochemical and physiological responses to abiotic stress (Ashraf and Hafeez, 2004; Crafts-Brandner and Salvucci, 2002; reviewed in Hasanuzzaman et al., 2013), underlying regulation of gene expression plays an important role for tolerance to many different stress responses (Wang et al., 2004; Miura and Furumoto, 2013). Many QTL studies have highlighted the importance of altered gene expression levels or patterns in driving functional variation among alleles. We were interested in understanding the mechanisms that drive allelic variation for response to environmental conditions in order to understand the sources of functional allelic variation that might contribute to local adaptation. Expression of genes coding for key components in stress response pathways is essential for tolerance to abiotic stress. Several transcription factors (TFs) directly affect responses to temperature stress by 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) plays an important role in tolerance to various abiotic stresses including tolerance to cold, drought, and heat (Jaglo-Ottosen, 1998; Schramm et al., 2008; Sakuma et al., 2006). CBF/DREBs play similar roles in many flowering plants including maize, rice, and soybean (Dubouzet et al., 2003; Chen et al., 2007; Liu et al., 2013). There is also evidence from Arabidopsis thaliana studies that expression of cold-responsive (COR) genes and heat shock protein (Hsp) families are important for cold and heat tolerance, respectively (Yadav, 2010; Hasanuzzaman et al., 2013).

Multiple studies have described cis-regulatory elements upstream of known stress-induced genes, including dehydration response element (DRE), abscisic acid (ABA) responsive promoter elements (ABREs), and heat shock elements (HSEs, reviewed in Raghavendra et al., 2010; Yamaguchi-Shinozaki and Shinozaki, 1994; Guiltinan et al., 1990). These elements provide the necessary binding site for TFs to confer stress-responsive gene expression (Yamaguchi-Shinozaki and Shinozaki, 1993; Hovarth et al., 1993; Schoffl et al., 1998; Mittal et al., 2011; Raghavendra et al., 2010). The cold and drought-responsive genes DREB1A and DREB2A also bind to cis-DREs upstream of responsive genes, which resulted in expression of downstream genes. These two DREB genes increase in expression independently under different stresses and induce the up-regulation of different downstream genes. This situation suggests that the DREB genes function and are induced independently (Liu, 1998). Two recent studies have explored sources of regulatory variation for gene expression responses to drought conditions on a genome-wide level. One study, in Arabidopsis thaliana, monitored both steady-state and drought-responsive cis- and trans-regulatory variations (Cubillos et al., 2014). A study in switchgrass also detected signatures of steady-state and stress-responsive regulatory variations in response to drought (Lovell et al., 2016). Both studies found a bias toward cis-regulatory variation when contrasting accessions, as well as treatment and control plants (Cubillos et al., 2014; Lovell et al., 2016).

Maize (Zea mays) exhibits substantial genomic and transcriptomic diversity between inbred lines (Chia et al., 2012; Hirsch et al., 2014). Both cis- and trans-regulatory variations contribute to this diversity and have been shown to play important roles in processes including domestication and heterosis (Stupar and Springer, 2006, Holloway et al., 2011; Li et al., 2013; Swanson-Wagner et al., 2009; Lemmon et al., 2014; Hirsch et al., 2014). The sources and prevalence of cis- and trans-regulatory variations and their role in stress response remain poorly characterized in most species including maize. One recent study suggested transposable element (TE) polymorphism may be one source of cis-regulatory variation for the allele-specific response to abiotic stress (Makarevitch et al., 2015).

In this study we utilized the extensive natural variation of maize to study variation in the abiotic stress response among different genotypes. By monitoring allele-specific gene expression in heterozygous plants it was possible to identify examples of cis-regulatory variation for responsiveness to cold or heat stress. In particular, first we confirmed that there is genetic variation for transcriptional responses to stress between our parents. Next we showed that while cis-regulatory variation is more common than trans-regulatory there is ample evidence for both cis- and trans-regulatory variations both at the steady-state and in response to stress. Our results provide a comprehensive picture of the genetic variation that contributes to the evolution of stress responses to the abiotic environment.

RESULTS

RNA-seq was used to survey natural and allelic variation in global gene expression in 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 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 used to examine the relative expression of alleles in hybrids. The transcriptome data were 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 provided three triplet combinations of two parents and their $F_1$ off-spring for all conditions with data from the same experiment. For all seven genotypes, RNA-seq was performed for three biological replicates of 14-day-old seedlings subjected to control, cold or heat treatments (Table S1). Gene expression levels were estimated from the RNA-seq data by aligning the reads to single nucleotide polymorphism (SNP)-corrected reference genomes and counting uniquely aligning reads (Dataset S1). For the hybrid genotypes, the allele-specific expression (ASE) ratios were determined using SNPs within the transcripts to count reads derived from each allele (Dataset S1). This dataset provides data for 25 433-25 641 expressed genes in each genotype by treatment combination and approximately 74% of these genes had ASE data in the hybrid.

Transcriptome changes induced by stress highlight roles of known regulatory pathways

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 could arise...
during subsequent growth (Makarevitch et al., 2015). There were no apparent morphological differences immediately following the stress treatment, but leaf necrosis and reduced growth of stressed plants is apparent several days after the stress treatment. In order to confirm that the cold and heat stress treatments were triggering cold or heat responses we investigated the gene ontology (GO) terms over-represented in up-regulated genes following heat or cold treatment (see Experimental procedures for details). As expected, we found that 72% and 52% of the 25 most significant over-represented GO categories were associated with response to stress in cold and heat conditions, respectively (Tables S2 and S3).

The importance of transcription factors (TFs) for stress response in Arabidopsis thaliana, maize, rice, and wheat is well documented (Abe et al., 1997; Qin et al., 2004; Wang et al., 2004; Lata and Prasad, 2011; Mizoi et al., 2012). We assessed whether maize TF families were enriched for up-regulated genes in response to heat or cold stress. In total, 20 of the 57 annotated maize TF families (Yilmaz et al., 2009) were enriched for up-regulated genes in cold and/or heat stress relative to control conditions in at least three of the inbred genotypes (Figure S1a). Eight families of TFs enriched for genes that were up-regulated in response to heat stress included the heat shock (HS) and the myeloblastosis (MYB) TF families (Figure S1a). Many (15/29) of the HSF TFs showed a significant (FDR corrected P-value < 0.01) increase of expression in B73 plants exposed to heat stress (Figure S1b). There were 18 families of TFs enriched for up-regulation in response to cold stress, including the APETAL2/Ethylene Responsive Factor (AP2/EREB), which contains the maize orthologs of Arabidopsis DREB/CFB 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 S1c).

Prior research has provided evidence that genes responding to abiotic stress often contain proximal cis-elements, including DRE/CRT, ABRE, HSE, NACR, WRKY, MYBR, MYCR, or LTRE motifs, which provide binding sites for DREB/CFB, ABA, HSF, NAC, WRKY, MYB, MYC, or LTR proteins, respectively (Xiao and Lis, 1988; Guiltnan et al., 1990; Yamaguchi-Shinozaki and Shinozaki, 1994, 2006). The genomic distribution of DRE/CRT, ABRE, HSE, NACR, WRKY, MYBR, MYCR, or LTRE motifs in the B73 reference genome was used to assess whether these motifs are enriched near up-regulated genes in response to heat (n = 1627) or cold (n = 1858) stress in B73. The HSEs were enriched near genes that are responsive to heat stress, but not cold-induced genes (Figure S1d). The DRE/CRT, ABRE, MYCR and WRKY binding sites were all enriched in the 1 kb regions upstream of both cold- and heat stress induced genes (Figure S1d). The MYBR and LTRE sites are enriched in distal promoter regions for cold-induced genes. NACR elements were not enriched in any of the regions tested (Figure S1d). These results provide evidence for expected cis/trans-regulatory pathways leading to the observed changes in gene expression in response to cold/heat stress.

**Variation for transcriptional responses to abiotic stress among maize inbreds**

Prior studies have documented numerous examples of cis- and trans-regulatory variation for steady-state transcript abundance in maize (Stupar and Springer, 2006, Holloway et al., 2011; Li et al., 2013; Swanson-Wagner et al., 2009; Lemmon et al., 2014; Hirsch et al., 2014). In this study we were particularly interested in documenting examples of variation for responsiveness to heat or cold stress for specific alleles. Clustering of the gene expression levels revealed significant treatment effects on the transcriptome, but also found evidence that genotypes have slightly different responses (Figure S2a, b), with hybrids often exhibiting intermediate responses in both batches of samples (Figure S2b). Clustering was also performed using genes that are differentially expressed (DE) in at least one inbred genotype in response to cold (n = 7838) or heat (n = 7155) (Figure 1). Visual examination of the clustering suggests

![Figure 1](image-url)
that many genes had consistent responses to heat or cold stress in all inbreds. However, there were also a substantial number of genes with variable responses to cold or heat treatments among genotypes. Both clustering methods revealed large transcriptional changes in response to cold or heat as well as suggesting some variation in those responses among genotypes. Thus, we next used two complementary statistical approaches to assess if there was evidence for genetic variation between parents in transcriptional responses to stress.

For our first approach, we utilized a two-way factorial analysis of variance (ANOVA) on RPM (reads per million mapped reads) levels of each gene. This parses expression variation due to genotype (G), treatment (T) and genotype by treatment (G × T) interactions. We ran this model twice for each pair of inbred parents; once comparing control treatment to heat treatment and once comparing control to cold. (Table 1a; see Experimental procedures for details). Many expressed genes (21–52%) exhibited significant genotype and/or treatment effects after controlling for multiple comparisons. A smaller subset of genes (1–28%) showed evidence of genotype-specific responses to treatment (Table 1a). There were notable differences in the frequency of G × T effects in the batch 1 samples including B73 and Oh43 compared to the genotypes in batch 2. It is not clear whether this is due to biological differences in these genotypes or technical differences among batches of samples. A visualization of the amount of variation ($\eta^2$ values) explained by genotype, treatment and G × T reveal large clusters of genes predominantly affected by either genotype or environment with fewer genes with strong G × T effects (Figure S3). Thus while genotype and treatment explained much of the variation between parents, there is evidence for genetic variation in response to stress.

For our second approach we used a negative binomial generalized linear modeling (GLM) framework to model differential expression of normalized count data. This method also provided an opportunity to model ASE levels of the hybrids enabling an assessment of patterns of gene expression consistent with environmentally sensitive cis- and trans-variation (approach adapted from Lovell et al., 2016). It is important to note that this approach only utilizes reads that can be mapped specifically to one allele in each parental/hybrid contrast and thus uses a different dataset than the one described above (see Experimental procedures for details). We included treatment (control or stress), allele (parental allele source), and generation (parental or F1) as factors in the model as well as all possible interactions between them. As expected, allelic identity and treatment were the most common sources of variation in expression (Table 1b). However, several hundred genes exhibit significant variation due to allele by treatment effects suggesting differential regulation of some alleles in response to heat or cold stress. A similar number of genes exhibit significant allele by generation effects suggesting that these hybrids exhibit differences in expression of certain alleles compared to the parents. An even smaller number of genes exhibit significant generation by treatment interactions that would indicate steady-state trans-responses or three way interactions trans-variation between treatment, allele, and generation, which indicate environmentally sensitive. This suggests relatively rare differences for expression responses to heat or cold stress in hybrids compared to inbred parents. The B73/Oh43 contrast showed a much larger number of significant interaction terms and this could be due to batch effects or increased power for this experiment due to lower variation among replicates. This analysis provides evidence for cis-

| Table 1 Number of genes with significant effects for total expression levels or allelic expression levels |
|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
|                                                  | Control and cold treatment                       | Control and heat treatment                       |
|                                                  | B73/Mo17 | B73/PH207 | B73/Oh43   | B73/Mo17 | B73/PH207 | B73/Oh43   |
| (a) ANOVA (RPMs)                                 |          |          |           |          |          |           |
| Genotype                                        | 8691     | 7851     | 17 135    | 9030     | 7834     | 16 049     |
| Treatment                                       | 8291     | 6439     | 16 119    | 9461     | 9543     | 20 461     |
| Genotype by treatment                           | 584      | 308      | 8075      | 918      | 472      | 10 950     |
| Total genes with data                           | 39 656   | 39 656   | 39 656    | 39 656   | 39 656   | 39 656     |
| (b) GLM (ASE counts)                            |          |          |           |          |          |           |
| Allele                                          | 5758     | 5630     | 5032      | 5482     | 5679     | 4185       |
| Treatment                                       | 6482     | 6554     | 5821      | 5835     | 5982     | 4957       |
| Generation                                      | 596      | 386      | 709       | 244      | 413      | 682        |
| Allele by generation                            | 275      | 297      | 2273      | 302      | 316      | 1097       |
| Allele by treatment                             | 273      | 507      | 1930      | 236      | 202      | 1594       |
| Treatment by generation                         | 1        | 3        | 1176      | 52       | 5        | 1181       |
| Treatment by Generation by Allele               | 1        | 0        | 276       | 2        | 0        | 880        |
| Total genes with data                           | 15 997   | 15 360   | 15 609    | 15 837   | 15 283   | 15 458     |

Significance for both tests: FDR < 0.01.
regulatory variation both for steady-state expression and for responsiveness to stress.

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

Previous research has documented high levels of cis-regulatory variation, particularly for genes with strong expression variation between maize 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 lines within each environmental treatment. Biased allelic expression in the F₁ provided evidence of cis-regulatory variation, while balanced expression of the two alleles suggested trans-regulatory variation, despite differences in expression in the parental genotypes (Figure 2). Genes were classified as having cis- or trans-regulatory variation by performing a chi-square analyses of the observed read counts relative to either the cis- or trans-expectation and identifying the genes with significant ($\chi^2 < 0.01$) from one expected value but not the other (Table S4). 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 3 and S4). Cis-regulatory variation is quite prevalent in all conditions with fewer examples of trans-regulatory variation although higher rates of trans-regulatory variation are observed for the B73-Oh43 comparison from batch 1 (Table S4). 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 S5a). In contrast, it was more common for genes classified as having trans-regulatory variation in control conditions to exhibit cis-regulatory variation in heat or stress conditions (Figure S5b).

**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 an abiotic stress. The analysis of allelic variation for response to heat or cold stress may help uncover the mechanisms that underlie gain or loss of stress-induced gene expression. In order to identify examples of genes with regulatory variation in response to heat or cold stress we found genes with ASE data with equivalent levels of expression in the two inbred parents in control conditions, but exhibiting significant differences in expression following heat or cold stress (Figure 4). These include genes with both up- and down-regulation in response to the stress, as well as genes that exhibited low or no expression in control conditions. These could include examples of cis- or trans-regulatory variation for responsiveness. If a gene has purely trans-responsive

---

**Figure 2.** Genes with cis- or trans-regulatory variation between genotypes within a condition were observed. (a) In this example the allele of inbred 1 is more highly expressed than the allele of inbred 2, which is indicated by the thickness of the arrow. 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. (b, c) The reads per million mapped reads (RPMs) for a trio of samples across all treatments are shown for an example of cis-regulatory (b) and trans-regulatory (c) 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.
regulatory variation, then the two alleles in the stress \( F_1 \) hybrid should not be statistically different than the proportion observed in the \( F_1 \) control. Whereas, a biased expression of alleles in the \( F_1 \) stress sample suggests responsive cis-regulatory variation (Figure 4b). For example, by contrasting ASE in the \( F_1 \) hybrids for gene GRMZM2G140082 we observed that both Mo17 and PH207 alleles showed biased expression compared with the B73 allele, whereas the Oh43 allele showed equivalent expression in the cold condition (Figure 4c). The observed bias expression of the Mo17 and PH207 alleles in the hybrids was in the direction we would predict given the parental expression values (Figure 4d). An example of trans-regulatory variation for stress response (Figure 4e) was observed for gene GRMZM2G077463, which showed equivalent parental expression for B73 and Oh43, but higher expression of Mo17 and PH207 when exposed to heat (Figure 4f). Allelic ratios of B73 to Mo17 and B73 to PH207 were equivalent in the respective hybrids under control and stress conditions (Figure 4g). Hundreds of genes exhibited variation in stress-induced expression between two parents (Figure 5). Overall, more genes had altered allelic proportions in the

---

**Figure 3.** Classification of differentially expressed genes into regulatory variation categories. Genes that are differentially expressed (DE) between inbreds were identified for each condition. The gap in the middle of each figure is the result of the requirement that genes have differential expression in the parents. Each scatterplot shows the entire set of DE genes between B73 and Mo17 for control (left), cold (middle), and heat (right). 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-regulatory (blue) or trans-regulatory (red) 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 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 4.** Examples of responsiveness cis- and trans-regulatory variation. (a) Responsiveness regulatory variation candidates were identified by assessing genes that have similar expression in the control (black circles) and variable response to stress (blue and red dashed circles) between two inbreds (thickness of arrows). (b) 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). (c) The expression (RPMs) for all genotypes in control (black) and cold (blue) conditions. Error bars are standard error for all replicates for each batch, except for B73 which is the standard error between batch 1 and batch 2. (d) Is the proportion of B73 in the \( F_1 \) hybrids for a gene that is biased toward Mo17 and PH207 alleles under cold conditions is shown. (e) Conversely, if a gene exhibits trans-regulatory responsiveness variation then both parental alleles would be equivalently expressed in both control and stress conditions in the hybrid. (f, g) (f) The expression levels and (f) proportion of B73 allele in the hybrids for a gene that is biased toward Mo17 when contrasting B73 and Mo17 (PH207 was not DE between control and heat) under heat conditions.

---

**© 2016 The Authors**
Figure 5. Genes with stress-induced differential expression more often exhibit cis-regulatory variation.

(a) Categorization of cis- and trans-regulatory variation for response to cold and heat stress (gene number listed under the pie charts indicates the number of genes that could be classified). Each gene was categorized into one of three categories. A majority of genes exhibit biased allelic proportions between the control and stress $F_1$ samples (blue and grey). Of the genes that show allelic bias a majority showed a bias in the direction we would predict given parental expression values (blue). This pattern is consistent with cis-regulatory variation. A small proportion of genes (4–16%) showed an allelic bias toward the opposite parental allele (grey). The rest of the genes exhibit trans-regulatory variation patterns (red) in which the proportion of alleles were not statistically different ($\chi^2 > 0.01$) between control and stress conditions in the $F_1$.

(b) An ABRE (coloured oval) located upstream of gene GRMZM2G119258 has a SNP in Mo17, Oh43 and PH207 relative to B73 and B37 (Bukowski et al., 2015).

(c) The relative fold change in gene expression in cold relative to control is shown for GRMZM2G119258 in each of the five genotypes.

(d) The proportion of genes that have a retained duplicate are shown the filtered gene set (FGS, black), classified as cis-regulatory (blue), or trans-regulatory (red) variation. Maize paralogs (as reported in Schnable et al., 2011) 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 were utilized for this analysis.

(e) The proportion of paralogs in (d) 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-regulatory (blue), and trans-regulatory variation (red).

$F_1$ between control and stress treatments, which would suggest cis-regulatory variation (Figure 5a). Of the genes that exhibited allelic bias 84–96% were in the direction we would predict based on parental expression.

Allelic variation for stress-induced expression can be the result of a loss or gain of expression activation in response to stress. Variation in important cis-regulatory elements upstream of genes could provide a potential mechanism for loss of expression activation. SNPs within important cis-binding elements (ABRE, DRE/CRT, and HSE) were identified within the promoter region for two groups of genes. The first group of genes exhibited increased expression following a stress treatment in B73, but no difference in expression for the other genotype in the trio, whereas the second group showed an increase in both alleles. If SNPs in binding motifs are associated with loss of stress-induced expression, then we would predict a greater rate of SNPs in binding regions upstream of the genes with variable response than those with consistent responses for both alleles. Genes with allelic variation for response to stress showed a higher rate of SNPs in nearby cis-regulatory elements than genes where both alleles are up-regulated for most of the stress by motif combinations (Table S5). One example, GRMZM2G119258, was categorized as a gene exhibiting cis-regulatory variation for response to cold stress between B73 and Oh43. B73 had an intact ABRE element upstream of the gene, while Oh43 and Mo17 contained a SNP (Bukowski et al., 2015) in that element. The expression level and genotype for PH207 and B37 (RNA-seq data from Makarevitch et al., 2015) were also determined (Figure 5b). If the SNP in this motif is associated with expression differences, then genotypes with the B73-like haplotype should increase in expression under stress conditions, and non-B73 haplotypes should not increase. Analysis of expression showed that B73 and B37 are DE following cold stress while Mo17, Oh43, and PH207 do not respond to a cold treatment (Figure 5b).

The allelic variation for responsiveness to an abiotic stress could be the result of loss of regulatory information for one allele, or the result of acquisition of novel regulatory information for the other allele. The presence of retained duplicates from a recent whole genome
dissociation event provides an opportunity to assess the relative frequency of gains and losses for responsive expression. Maize underwent a whole genome duplication event approximately 6–12 million years ago (Gaut and Doebley, 1997). Subsequent fractionation has resulted in frequent loss of paralogs or cis-regulatory elements (Freeling et al., 2015). 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. We can look at the responses of the retained syntenic duplicates if a gene exhibits allelic variation with some alleles responding to a stress, and other alleles that do not respond. 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. Genes that exhibited cis- or trans-regulatory variation retain paralogs at a similar rate (44–49%) to that of other expressed maize genes (44%, Figure 5c). For the genes classified as exhibiting cis- or trans-regulatory variation for responses to abiotic stress, we determined whether the retained duplicates also responded to the same stress (Figure 5d). The majority of these genes exhibited stress responses for the retained duplicates, suggesting that most examples of allele variation reflect recent loss of stress-responsive expression.

Previous work has shown that genes exhibiting steady-state changes and cis-regulatory variation patterns tended to have a higher fold change between inbreds than trans-regulatory variation (Stupar and Springer, 2006). Comparing the fold change between stress and control treatments for genes that have responsive cis- or trans-regulatory variation patterns allowed us to test whether this pattern holds true for stress-induced expression. A non-redundant list of genes exhibiting cis- and trans-regulatory variation was created by combining all genes categorized as cis or trans from all three trio comparisons. GO enrichment analysis was performed using the top Arabidopsis thaliana BLAST hit for this set of genes, and the only enriched biological process was response to stimulus. In contrast to expression patterns, we did not observe a bias of genes with strong expression variation also exhibiting cis-regulatory variation for any of the comparisons made in this experiment (Figure S6). Interestingly, in the contrast of B73 and Oh43 from batch 1 data we observed a higher mean and larger variation for genes exhibiting trans-regulatory variation then for the genotypes monitored in batch 2 (Figure S6). This result could reflect a biological difference among these genotypes or technical variation among batches. Many transcription factor families were enriched in our stress-induced expression gene sets, therefore 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 S6).

DISCUSSION

The expression of many genes is influenced by cold (Miura and Furumoto, 2013) or heat stress (Wang et al., 2004) and we found that many of these expected changes in gene expression were observed in our cold or heat stressed samples. Expression variation is a major contributor to phenotypic variation within species and divergence among species. In particular, the developmental timing of gene expression or responsiveness of expression to environmental stimuli can vary among species. Differential regulation of alleles can result from changes at proximal regulatory regions (cis), changes in unlinked loci that encode regulatory molecules such as TFs or regulatory RNAs (trans), or a combination of both. While researchers appreciate the importance of variation at cis-regulatory regions for contributing to differences in both steady-state expression variation and differences in responsiveness to environmental stimuli, we lack a detailed understanding of the sequence changes underlying this variation. In this study we utilized natural variation in maize to assess the regulatory variation for steady-state expression levels, as well as expression responses to heat or cold stress in several maize genotypes. Given the previous evidence for abundant cis- and trans-regulatory variation contributing to differences in steady-state expression levels (Stupar and Springer, 2006; Holloway et al., 2011; Li et al., 2013), we were interested in documenting the contribution of cis- and trans-regulatory variation for gene expression responsiveness to specific treatments such as heat or cold stress. The relative expression for both parental lines, as well as the two alleles in hybrid plants grown in control or stressed conditions were used to identify examples of differential response to environment and then to partition these into cis- or trans-regulatory variation. We found hundreds of genes with variable expression responses to heat or cold stress, with much of differences due to cis-regulatory variation.

These examples of cis-regulatory variation for stress responsiveness can be utilized to learn about the mechanisms leading to expression variation within species. The allelic variation for stress-responsive expression could arise from the loss of responsiveness for an allele or through the acquisition of novel responsiveness. An analysis of the expression response for paralogs from a recent whole genome duplication event suggested that in many cases the variation is likely due to loss of responsiveness for one allele (Figure 4d, e). This variation can occur due to SNPs or deletions in binding sites for TFs that are stress-induced. We were interested in the rare cases of novel
acquisition of stress responsiveness. These examples may shed light on the mechanisms that could lead to novel variation for gene expression responses. Previous research has suggested that transposon polymorphisms may contribute to stress-responsive expression in maize (Makarevitch et al., 2015) and further research into the haplotypes of genes identified in this study will document the role of transposon variation at these loci.

Several recent studies (Cubillos et al., 2014; Lovell et al., 2016) have also studied variation in expression responses in Arabidopsis and switchgrass. While Lovell et al. (2016) focused on parental varieties exhibiting strong phenotypic differences for tolerance to drought stress, both Cubillos et al. (2014) and this study utilized parental varieties without strong phenotypic differences for tolerance to the treatment conditions. There are some differences in the sensitivity of B73, Mo17, Oh43 and PH207 to cold and heat stress in terms of relative growth rates in the days following the stress, but these varieties tolerate our stress regimes and recover. This suggests the abundant molecular differences for stress-responsive expression are not contributing to a long-term, major phenotypic difference. This study was not designed to identify genes that would provide tolerance to cold or heat stress in maize. Instead, our goal was to begin to document regulatory features that contribute to the ability of a gene to respond to cold or heat stress in maize, and to be able to monitor how variation in promoter regions may contribute to changes in responsiveness to stress. Studying regulatory mechanisms on a large scale is quite complicated, but improvements in sequencing technology, annotation of genomes, and novel analytic tools are providing a platform to begin understanding complex regulatory mechanisms. This information may be useful in developing predictive models for gene expression responses and for engineering novel responses to abiotic stress in crops.

**EXPERIMENTAL PROCEDURES**

**Stress conditions, tissue collection, and RNA extraction**

Inbred and hybrid maize seedlings were grown under 16 h 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 (7°C) for 16 h. Heat stressed plants were placed in an incubator (50°C) for 4 h. Above-ground seedling tissue was collected immediately following each stress treatment with six plants pooled per sample and the tissue kept at −80°C. Biological replicates were grown 5 days apart for all treatments. After tissue collection each pooled sample was ground in 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, USA). Samples were sequenced using the Illumina HiSeq-2000 platform outputting 15–30 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 made for Mo17, PH207, and Oh43. The SNP-corrected references were made by taking the B73 reference fasta file and replacing the nucleotides where a SNP was present between the two inbreds. Each sample of Mo17, PH207, and Oh43 was aligned to their respective corrected reference. All alignments were made using Tophat 2 (version 2.1.0, Kim et al., 2013) and bowtie2 (version 2.0.10,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 reads per million mapped reads (RPM) calculations (Tophat2 option -N 0). Bamtools (version 2.3.0) was used to filter out reads that mapped with indels or to multiple places in the genome (Barnett et al., 2011). Inbred and hybrid samples were aligned to both parental references using the same criteria (e.g. B73xOh43 was aligned to both B73 and Oh43 corrected references). Any SNPs that were not fully validated for the inbred samples were removed from subsequent analyses. To obtain ASE counts for hybrids, reads from the two alignments for each hybrid were compared and reads that mapped uniquely to one parent were retained. All uniquely mapping reads from inbred parents and the ASE counts from the hybrids were run through HTSeq (version 0.6.1p1, Anders et al., 2014) to get per gene counts and per gene ASE counts, respectively. RPM values, were calculated for all inbred parents by using the count files from HTSeq. Hybrid RPM values were calculated by summing the parent specific reads and the reads that mapped equally to both parents.

**ANOVA and eta-squared**

A pairwise two-way factorial ANOVA was run for each gene on all biological replicates for a pair of inbreds and conditions (e.g. control and cold) using the aov function in R (Chambers, 1992) to look for significant genotype, treatment, and genotype by treatment effects. RPMs for each sample were used. Statistical significance was determined for each factor by calculating FDR corrected P-values. For a factor to be considered significant the adjusted P-value had to be <0.05. Eta-squared values were used to estimate the effect size of each factor from the ANOVA (G, T, and G x T) for each gene. Eta-squared values were also calculated for each gene with at least one significant factor from the ANOVA. Eta-squared values were calculated in R using the eta-Squared function within the far package in R (Navarro, 2015).

**Generalized linear model (GLM)**

Allele-specific counts (see the Sequencing and data processing section for details) for both the inbreds and F1 hybrids in a trio for control and one stress condition were used to assess which factors contribute to the variation in allelic expression. (similar to Lovell et al., 2016). The allele-specific counts for genes that have a mean ASE count >5 for both alleles were fit to a negative binomial model using DESeq2 (Love et al., 2014) with:

\[
\log(y) = b_0 + b_T T + b_A A_i + b_G G_i + b_T A_i \times T + b_T G_i \times T + b_A A_i \times G_i + b_T A_i \times G_i
\]

where for individual \(i\) at gene \(j\), \(T, A_i\) describes treatment where control is the baseline and the treatment (cold or heat) is the alternate set, \(A_i\) is the allele where the baseline is B73 and the alternate is non-B73, \(G_i\) represents the generation where the baseline is the inbreds (F0) and the alternate is the hybrid (F1). In order to test whether any factor or interaction between factors were significant
we applied nested likelihood ratio tests (LRT) and used FDR corrected P-values (<0.01). We tested the significance of all possible individual factors and interactions beginning with the highest order interaction. If the highest order interaction was not significant, we dropped that term and tested each subsequent two-way interaction while controlling for the other two interactions. If none of the two-way interactions was significant, we tested for each individual factor while controlling for the other two individual factors.

**GO analysis**

Genes that were up-regulated in response to cold or heat for each genotype were independently run through the BinGO plug-in 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 analysis using the default statistical options and Arabidopsis thaliana database. 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 available characterized maize transcription factors (TFs, Yilmaz et al., 2009) was used to test whether certain TF families were associated with stress-induced DE genes. Enrichment of TFs were calculated by contrasting the number of up-regulated genes in at least three genotypes under cold or heat treatments to the expected number (ratio of genes that are up-regulated in at least three genotypes in the entire filtered gene set). A chi-squared test was used to determine which families were statistically enriched ($\chi^2 < 0.01$) in our up-regulated gene sets.

The Find Individual Motif Occurrences (FIMO, Grant et al., 2011) tool, which is part of the MEME suite (version 4.4, Bailey and Elkan, 1994) was used to count the number of DRE/CRT, ABRE, HSE, NACR, WRKY, MYBR, MYCR, and LTRE motifs throughout the B73 reference genome. The number of motifs per bin upstream, downstream, and within each gene was calculated by summing the number of motifs overlapping the specified bin using the BEDTools (version 2.25.0, Quinlan and Hall, 2010) window tool. Each motif was counted per bin (5-2.5, 2.5-1, and 1 kb for both 5’ and 3’ ends) 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-squared test was run to see if the observed number of motifs per bin was statistically different from expected.

**Identifying cis- and trans-regulatory variation**

**Genotypic regulatory variation.** Each gene that is DE between two inbred parents was run through independent chi-squared 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 ($\chi^2 > 0.01$) 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-squared test was applied to these gene 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.

**ACCESSION NUMBERS**

The RNA-seq reads used in this study are deposited at the National Center for Biotechnology Information Sequence Read Archive under project numbers PRJNA244661.

**ACKNOWLEDGEMENTS**

The authors would like to acknowledge the Texas Advanced Computing Center (TACC) at the University of Texas at Austin for computational resources and storage space necessary to complete this project. All of the Illumina sequencing was performed at the University of Minnesota Genomics Center (UMGC). The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. In addition, Peter Hermanson and James Satterlee assisted with sample processing and wet lab techniques. This research was funded by a NSF grant awarded to N.M.S. and I.M. (IOS-1444456). The authors declare no conflict of interest.

**AUTHOR CONTRIBUTIONS**

A.J.W., I.M. and N.M.S. designed the research, A.J.W., J.N., and I.M. preformed the research, C.N.H, C.D.H. and N.M.S. designed the research, A.J.W., I.M. and N.M.S. wrote the article.

**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** A subset of transcription factors and regulatory motifs are enriched in cold and heat DE gene sets.
Figure S2. Transcriptional responses in maize seedlings exposed to heat or cold treatment.

Figure S3. Genotype or treatment explain most of the variation observed in ANOVA analysis.

Figure S4. Classification of differentially expressed genes into regulatory variation categories for additional trios.

Figure S5. Conservation of cis- and trans-regulatory variation calls.

Figure S6. cis- and trans-regulatory variation genes exhibit similar fold change between stress and control.

Table S1. Sequencing depth for samples used in this study.

Table S2. GO enrichments for up-regulated genes in response to cold treatment.

Table S3. Enrichment of GO categories for genes that are up-regulated in response to heat stress.

Table S4. Distribution of cis- and trans-regulatory variation calls for genes DE between genotypes within a condition.

Table S5. Proportion of genes that have cis-elements within 5 kb upstream and have SNPs within elements.

Table S6. cis-regulatory effects in response to cold treatment.

Dataset S1. Gene and allele-specific expression counts for all samples used in the study. For each gene the reads per million (RPM) and allele-specific counts are provided for each of the RNA-seq samples.

REFERENCES


