TABLE OF CONTENTS:

* Table S1 and S2 can be found in separate files.

I: SUPPLEMENTAL METHODS:

1) Analytical Pipeline.

The R scripts to conduct this analysis were posted on github (https://github.com/jtlovell/cis-trans-test). The steps that make up each analysis are described briefly below and annotated more completely in the scripts posted on github.

2) Experimental design details.

The parents and F_1 hybrids were planted in the field on October 12, 2012 as part of a larger experiment. The field experimental site was located in a prairie field (30.182° N, 97.879° W) at the south end of the Ladybird Johnson Wildflower Center (Austin, TX). Prior to planting, the

field was covered with Sunbelt Weed Fabric (Peaceful Valley Farms, Grass Valley, CA). For planting, we cut holes in the cloth and used a mechanical auger to drill holes in the soil. Plants were transferred from 3.79L pots into the holes. Plants were arrayed into rows that were two plants wide, with 1.2m wide walking paths between rows to allow for easy access to plants for sampling. Plants were spaced 40 cm apart from each other along rows. The full design included 1000 fully randomized plants, of which the parents (N = 100 FIL2, 100 HAL2) and F_1s (N = 30) were a subset. Due to exceptionally low rainfall in the fall of 2012, we watered plants as needed through November and early December to ensure establishment. Irrigation was ceased once plants entered winter dormancy. The experimental plants emerged from dormancy in the spring of 2013 and were grown without manipulation until early July. Plants grew vigorously through April and May, as rainfall was abundant in those months (LBJ Wildflower Center Weather Station Data: April = 80.0 mm; May = 93.5 mm; Fig. 1). However, this period was followed by a prolonged hot and dry period, where only 5.08 mm of rainfall fell between 26-May and 7-July.

3) Assessment of experimental conditions:
Daily maximum air temperature and total precipitation were requested and downloaded from the National Oceanic and Atmospheric Administration (NOAA) web portal (https://www.ncdc.noaa.gov/cdo-web/) the most proximate weather stations to the collections sites of HAL2 and FIL2. Data for each year over the span of the experiment were parsed from these data files. Kernel density distributions for each climate variable were calculated using the R base function *density* and the percentile distribution of the 2013 drought were compared for each dataset. Local weather station data was available for 2013 at the site of the experiment. These data augmented our broader scale climate data.

4) Processing of counts data:
Libraries were compared by the relative expression of HAL2- and FIL2- allele-specific-expression (ASE). Lines with studentized residuals from the linear model (HAL2 ASE ~ FIL2 ASE) within each genotype (HAL2, FIL2, and F_1) that gave bonferoni-adjusted *p*-values (*t* test) < 0.0001 were excluded from all further analyses. This was accomplished using the function *outlierTest* in the R package *car* (Fox and Weisberg 2010). Genes with mean counts across all samples < 5 were also excluded from all further analyses. We then assessed ASE of the genes with >5 total counts and culled by two criteria. For a gene to be incorporated into our modeling of cis- and trans- acting expression regulation, genes had to have >5 mean FIL2 ASE counts in the FIL2 genotype and >5 mean HAL2 ASE counts in the HAL2 genotype. Furthermore, the mean HAL2 ASE counts in FIL2 genotypes and the mean FIL2 ASE counts in HAL2 genotypes had to be >1% of the FIL2 ASE counts in the FIL2 genotype and the HAL2 ASE counts in the HAL2 genotype respectively. These criteria assured that all genes analyzed had sufficiently precise and quantifiable ASE.

5) Estimation of mapping bias: Many analysis of ASE are sensitive to mapping bias. We corrected for these effects by treating ASE from each plant and allele as its own library in the DESeq2 pipeline. As such, the FIL2 and HAL2 ASE are each normalized with regards to the total number of counts assigned to each bin. Despite these efforts, extreme mapping bias may still introduce statistical artifacts into our analysis. To test mapping bias, we compared the mean number of mapped reads for each allele, for each gene. If the linear correlation falls near a slope of 1, this is an indication that mapping bias is not extreme. Indeed, this is the case (Fig. S6). We find very little evidence of mapping bias.

6) Model Specification:

Our model fit treatment (drought, recovery), genotype (FIL2, HAL2) and their interaction while controlling for sampling scheme. Assessment of differential expression was conducted by fitting two distinct models in *DESeq2*, one with total counts as the response variable and another with allele-specific counts (ASE). In each model, we corrected for the time of sampling by categorizing tissue harvest as early (in the first 50% of sampling) or late (in the last 50% of sampling). This categorical time variable was retained as the first term in all DESeq2 models to account for circadian variation in the linear models.

In the first model, we quantified differential expression in the context of genotype, environment and GxE between the parental genotypes, FIL2 and HAL2 (Fig. 2a). Genes were retained for the analysis if mean total counts were > 5, resulting in a total of 22,322 genes analyzed. For each gene we fit the standard negative binomial linear model DESeq2 pipeline (Love et al. 2013; 2014) with:

MODEL 1:

$$log_2\, q_{ij} = \beta_0 + \beta_D D_i + \beta_T T_i + \beta_A A_i + \beta_{TA} T_i{*}A_i$$

where for individual i and gene j, $D_i = 1$ if the sample was collected late in the day and 0 it was collected early in the day, $T_i = 1$ if the sample is a recovery treatment sample and 0 if it is a drought treatment sample, and $A_i = 1$ if the sample genotype allele is HAL2 and 0 if it is FIL2.

This analysis permitted the estimation of conserved genes that have similar responses across all individuals and environments (conserved), genes with expression that differs between the environments and/or genotypes, or genes that are differentially affected by the environment, depending on the genotype (GxE).

We quantified heterosis by including total counts of F_1 genotypes in MODEL 1. As such this model is a simple expansion of Model 1, where A has three levels (HAL2, FIL2 and F_1). From this model, we conducted specific contrasts between the three genotypes and binned each gene into one of the seven heterosis categories (Table 1) via a custom R script.

To quantify the degree of cis- and trans-acting expression modifiers, we again fit the standard DESeq2 model to allele-specific expression of HAL2, FIL2 and the F_1 hybrid, but with:

MODEL 2:

$$log_2\, q_{ij} = \beta_0 + \beta_D D_i + \beta_T T_i + \beta_A A_i + \beta_G G_i + \beta_{TA} T_i{*}A_i + \beta_{TG} T_i{*}G_i + \beta_{AG} A_i{*}G_i + \beta_{TAG} T_i{*}A_i{*}G_i$$

where for individual *i* and gene *j*, $D_i = 1$ if the sample was collected late in the day and 0 it was collected early in the day, $T_i = 1$ if the sample is from the recovery treatment and 0 if it is from the drought treatment, $A_i = 1$ if the count is for the HAL2 allele and 0 if the count is for the FIL2 allele and $G_i = 1$ if the sample from the F_0 generation and 0 if it is from the F_1 generation.

Under the MODEL 2 specification, β_A estimates the allele contrast within the F_1 generation, which corresponds to the classic cis-test notion of allelic imbalance: $A(F_1)/B(F_1)$. The interaction term β_{AG} can be interpreted as a difference of differences -- the difference between allele contrasts between generations -- which corresponds to the classic trans-test notion: $[A(F_1)/B(F_1)]/[A(F_0)/B(F_0)]$. We also fit MODEL 2 within each environment, to present the similarity of our analysis to traditional cis-trans tests (Figs. S4-5).

It is important to note that the estimation of cis/trans expression regulation is not identical to measuring differential expression. The test of allelic imbalance (Fig. 2a) is a specific contrast:

allele = FIL2 & generation = F_0 (FIL2-F_0) vs. HAL2-F_0.

However, the cis-test is coded so that we contrast:

FIL2-$F_{1\text{-ASE}}$ vs. HAL2-$F_{1\text{-ASE}}$,

and the trans test assesses the allele*generation interaction by contrasting the dependency of allelic imbalance on generation so that:

slope(FIL2-$F_{1\text{-ASE}}$ vs. HAL2-$F_{1\text{-ASE}}$) vs. slope(FIL2-$F_{0\text{-ASE}}$ vs. HAL2-$F_{0\text{-ASE}}$).

Finally, it is important to define the cases where this approach lacks power to detect cis, or trans effects. The most common pattern occurs where significant differential expression exists between the parental genotypes and differential expression exists in the same direction, but is no longer significant in the F_1 generation. In this case, the cis and trans effects are antagonistic, causing our model to detect differential expression between the parents. However, we would be unable to attribute this divergence to either cis, or trans effects. In fact, this pattern is quite common among genes, indicating that both cis and trans acting gene expression regulation exists, but neither has a strong enough effect to be detected via linear modeling (Fig. S4).

7) Multiple testing correction methodology:

Gene-by-gene statistical analysis such as the linear modeling presented here conducted many independent tests, and therefore required multiple testing corrections to maintain an acceptable experiment-wise false discovery rate. Here, we used the R package *qvalue* (Storey 2002) to transform *p*-values into *q*-values that account for the distribution of *p*-values and the number of tests.

For each set of *p*-values, we removed any NA's, then calculated the *q*-value distribution as well as the estimated proportion of true null hypotheses (π_0), which was estimated with the "bootstrap" method (Storey et al. 2004).

8) Promoter motif enrichment analytical pipeline:

A set of 485 annotated and published plant promoter motifs were downloaded from the newPLACE (Higo et al. 1999) database (http://www.dna.affrc.go.jp/PLACE/) on 9-December 2015. These motif sequences were converted into probability matrices via a custom algorithm and read into the R package PWMEnrich (Frith et al. 2004).

Promoter sequences from all *P. hallii* gene models were downloaded from the Phytozome server (http://phytozome.jgi.doe.gov/pz/portal.html) on 10-December 2015. We developed the background distribution of motifs on a random subset of 2000 promoter sequences using the function makePWMLognBackground. This background distribution was used to generate *p*-values for all motif analyses. We conduced promoter motif enrichment analyses on 7 subsets of genes. For each set of genes, we conducted motif enrichment analyses in the PWMEnrich package using the following parameters: scoring via the "affinity" method and background correction through the *logn* method. *P*-values were transformed to *q*-values using the R package qvalue (Storey 2002).

II: COMPARISON OF CIS-TRANS TEST METHODOLOGIES

1) Comparison to traditional binomial tests:

We implemented the methods of Wittkopp et al. (2004), who utilized a Mann-Whitney t-test to quantify the effects of cis and trans variation using the R base function *wilcox.test*. More recently, the cis-trans test has also been conducted using exact binomial tests, which we implemented using the R base function *binom.test*. *P*-value ranks for these two tests and our negative binomial linear model are presented in Figure S4.

2) Comparison of modeling in DESeq2 and LIMMA:

An important consideration is that the HAL2 and FIL2 alleles in the F1 plants are not independent. While normalization in DESeq2 accounts for differential library size, other factors, such as the specific attributes of the plant, soil characteristics, pests etc. may cause correlations between the alleles in the F_1 plants. As such, it may be appropriate to account for such correlations using a mixed effects model where the ID of each plant (Plant ID) is a blocking factor. Of particular concern would be any case where the estimated effects of a gene are biased in the DESeq2 approach, but are not significant when plant ID is accounted for.

LIMMA is a differential expression analysis package originally designed for microarray experiments. We originally opted for DESeq2 over LIMMA for several reasons: 1) DESeq2 is designed explicitly for RNA-seq datasets (Love et al. 2014), 2) it good FDR control when many biological replicates are present (Soneson and Delorenzi 2013), 3) there is precedent for the analysis of cis and trans factors in DESeq2 (Bader et al. 2015). However, DESeq2 does not offer mixed effects modeling, a drawback that makes it impossible to fit plant ID as a term in the model. LIMMA does permit fitting a model with a random a blocking factor.
However, LIMMA has several drawbacks that make us hesitant to exploit this approach. First, LIMMA fits a common correlation among all levels of random effects. Since our blocking factors occur across diverse sampling times and experimental treatments, it is not clear if this approach is appropriate. Second, LIMMA only permits a single blocking factor, whereas here, we would prefer to fit both time of sampling and unique plant ID as random effects. Finally, the blocking effect in LIMMA was originally designed for technical replicates (e.g. multiple arrays). It is not clear how this approach may perform using a blocking factor, like plant ID with 88 levels.

We first compared identical models in DESeq2 and LIMMA (Fig. S9). Clearly the results are similar, but DESeq2 offers significantly greater power for the estimation of all model contrasts.

3) Comparison of modeling with and without Plant ID as a blocking factor:

To test for the significance of the effect of blocking by plant ID, we implemented our model in LIMMA, accounting for the correlation due to the unique plant, and using plant ID as a blocking factor. We then merged these results with the DESeq2 results and plotted the resultant data. It is clear that the LIMMA approach dramatically improves power to detect both cis and trans effects (Fig S10); however, it is not clear if this more complex model accurately fits the data.

To assess this, we looked at the residual variance of the two models in LIMMA. Combined, the residual variance was reduced by 1.6% (median) in the model with plantID as a blocking effect. We conclude that the difference between models is very modest difference. Due to its increased power and simplicity, we opted for the DESeq2 fixed model over either of the LIMMA models.

III: SUPPLEMENTARY FIGURES

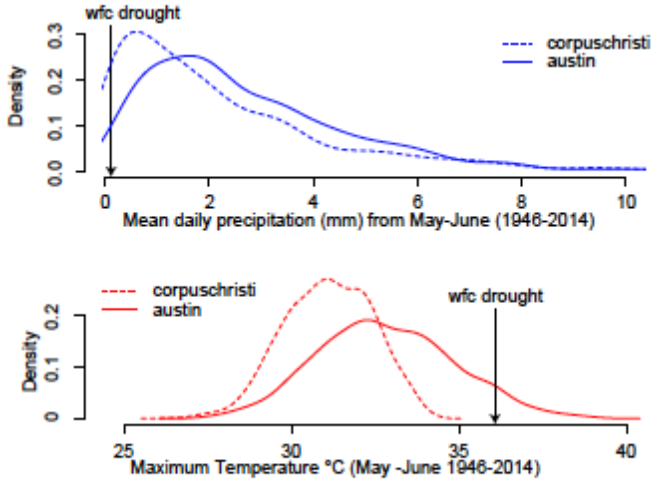


Figure S1. Climatic context of 2013 drought. The native habitats of HAL2 and FIL2 are separated by 280km, but exhibit markedly different historical climates during the study period, as indicated by the amount of precipitation in the warmest quarter. The density distributions of mean daily precipitation (a) and mean maximum temperature (b) for each year from 1946-2014 are plotted for each site. The point estimate of the 2013 drought intensity relative to each site is plotted as a vertical arrow

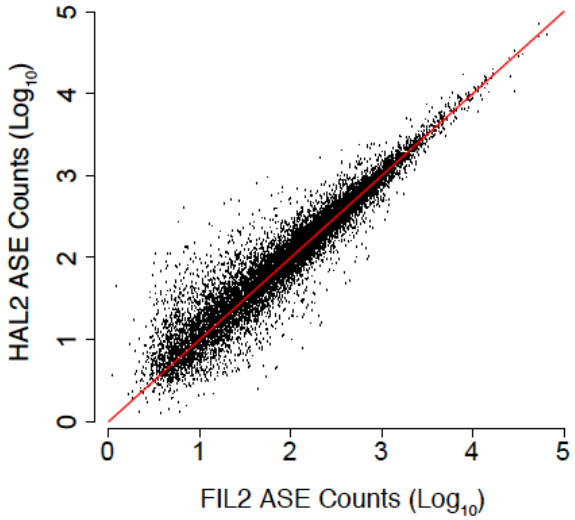


Figure S2. Qualification of mapping bias. Mean number of mapped reads to FIL2 and HAL2 alleles are plotted on $\log_{10}$ scales. The red line indicates a 1:1 mapping ratio (no mapping bias).

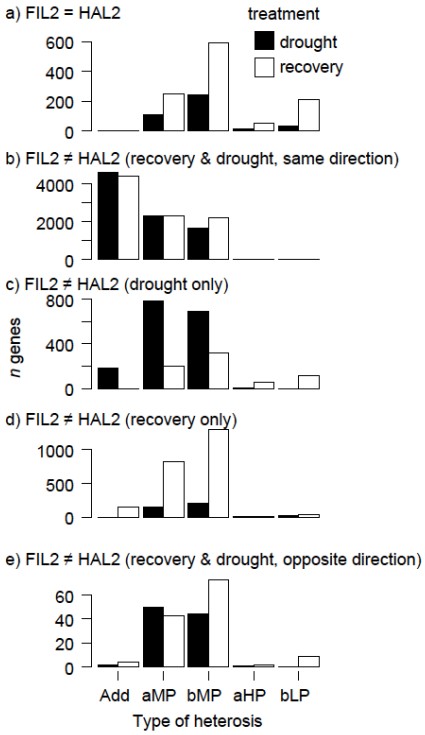


Figure S3: Distribution of heterosis effects across differential expression and GxE categories. Panel a presents heterosis categorization of conserved genes, while panels b-e depict the heterosis categorizations in the four color categories in Fig. 2b-c. The total number of genes in each category are presented and colored by the treatment: black bars indicate the number of genes in each heterosis bin from the drought treatment, while white filled bars represent the expression categorization from the recovery treatment.

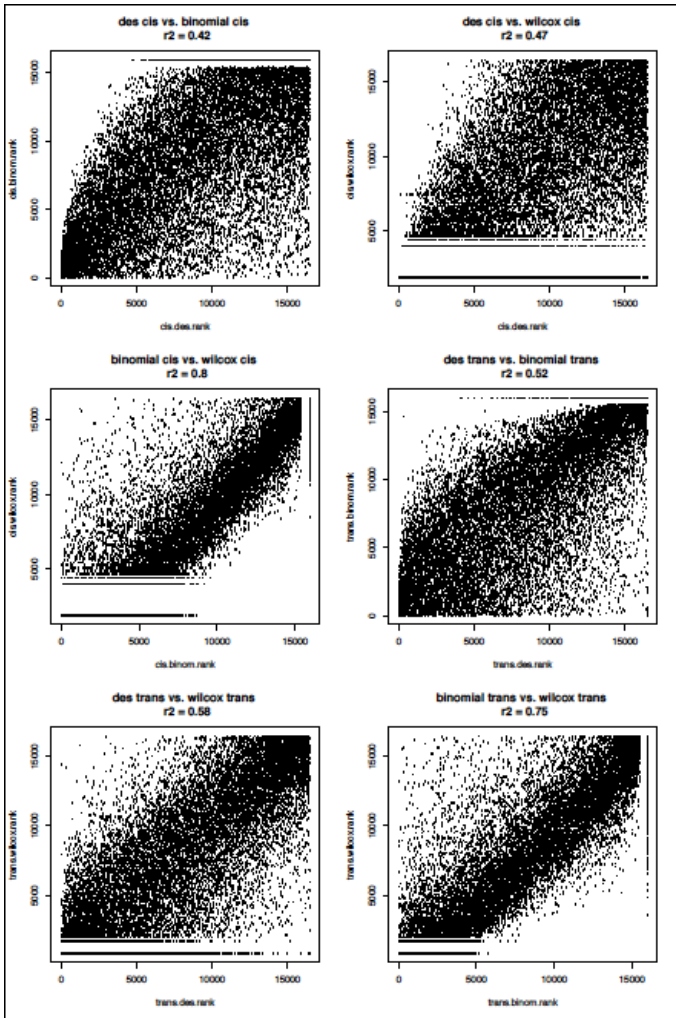


Figure S4. Comparison between linear modeling and traditional cis-trans test analyses. Comparison of the inference of cis (a-c) and trans (d-f) between DESeq2 models ("des cis", "des trans"), mann-whitney test ("Wilcox cis", "Wilcox trans") and exact binomial tests ("binomial cis", "binomial trans"). R^2 values from a linear model implemented in the R base function "lm" accompany the title of each comparison. All tests are completed using only the dry treatment data.

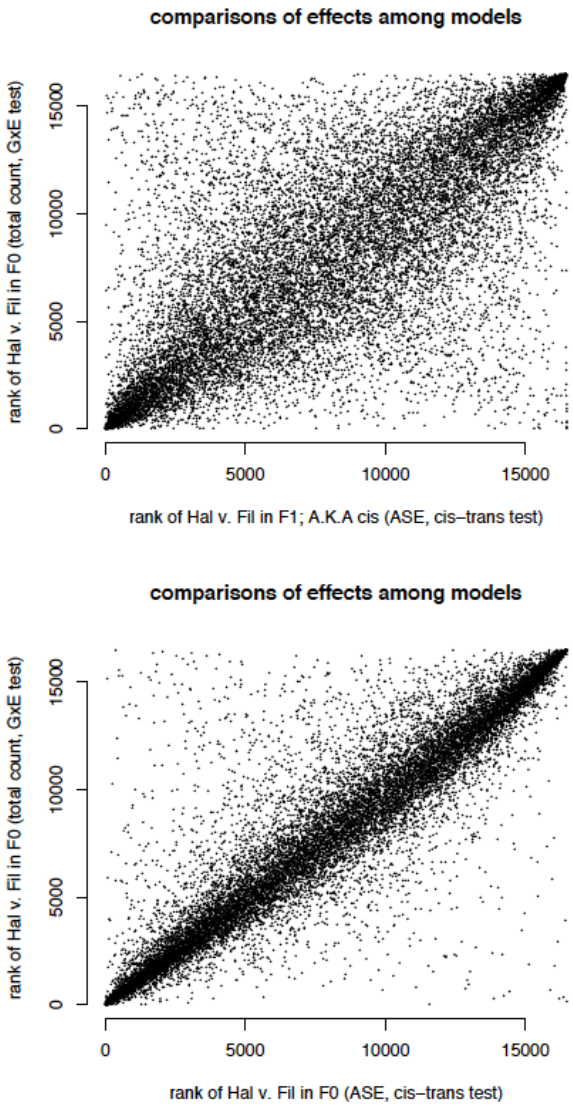


Figure S5. Comparison total and ASE transcript counts in the F_0 and F_1 generations. The $\log_2$ fold changes derived from total count comparisons between F_0-FIL2 and F_0-HAL2 are compared to LFC attributable F_1-FIL2 vs. F_1-HAL2 ASE (cis) (panel a) and F_0-FIL2 vs. F_0-HAL2 ASE (panel b).

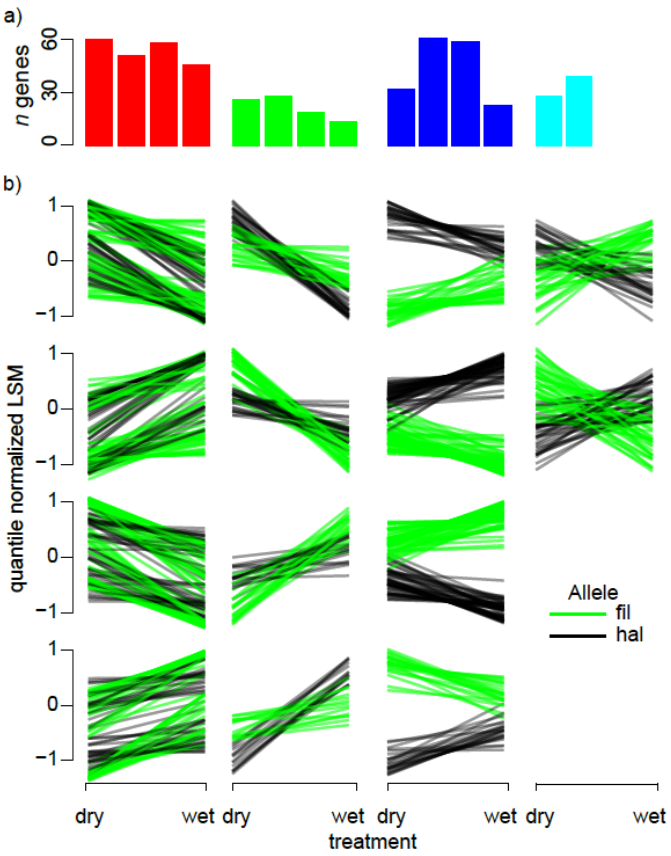


Figure S6. Counts and reaction norms of cis*treatment interactions. The values plotted are the quantile normalized least square means of the expression values for each combination of treatment and cis effect for all genes with significant cis*treatment interactions. Genes with significant treatment * cis interactions are broken out into the 14 possible expression patterns. The alleles are colored as FIL2 (green) and HAL2 (black). Accompanying the line plots are bar plots indicating membership in each expression pattern type. Bars from left to right represent membership of each pattern from top to bottom in each column. Column 1 (red barplot) shows all genes where the slopes of each allele are in the same direction, but the relative ranks of the alleles do not change. Column 2 (green barplot) is reserved for genes where the slopes are in the same direction, but the allelic LSmeans change ranks. Column 3 (blue barplot) genes have opposite slopes but retain the same ranks, while column 4 (cyan barplot) genes have both opposite slopes and different ranks.

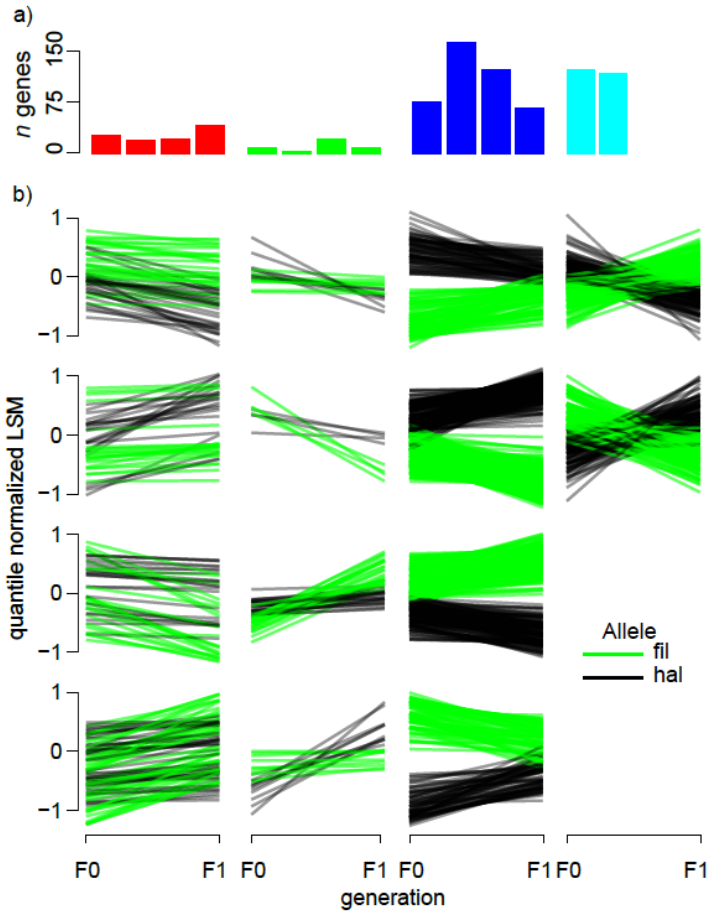


Figure S7. Counts and reaction norms of trans effects. The values plotted are the quantile normalized least square means of the expression values for each combination of allele and generation (trans) effect for all genes with significant trans effects. Genes with significant trans effects are broken out into the 14 possible expression patterns. The alleles are colored as FIL2 (green) and HAL2 (black). Accompanying the line plots are bar plots indicating membership in each expression pattern type. Bars from left to right represent membership of each pattern from top to bottom in each column. Column 1 (red barplot) shows all genes where the slopes of each allele are in the same direction, but the relative ranks of the alleles do not change. Column 2 (green barplot) is reserved for genes where the slopes are in the same direction, but the allelic LSmeans change ranks. Column 3 (blue barplot) genes have opposite slopes but retain the same ranks, while column 4 (cyan barplot) genes have both opposite slopes and different ranks.

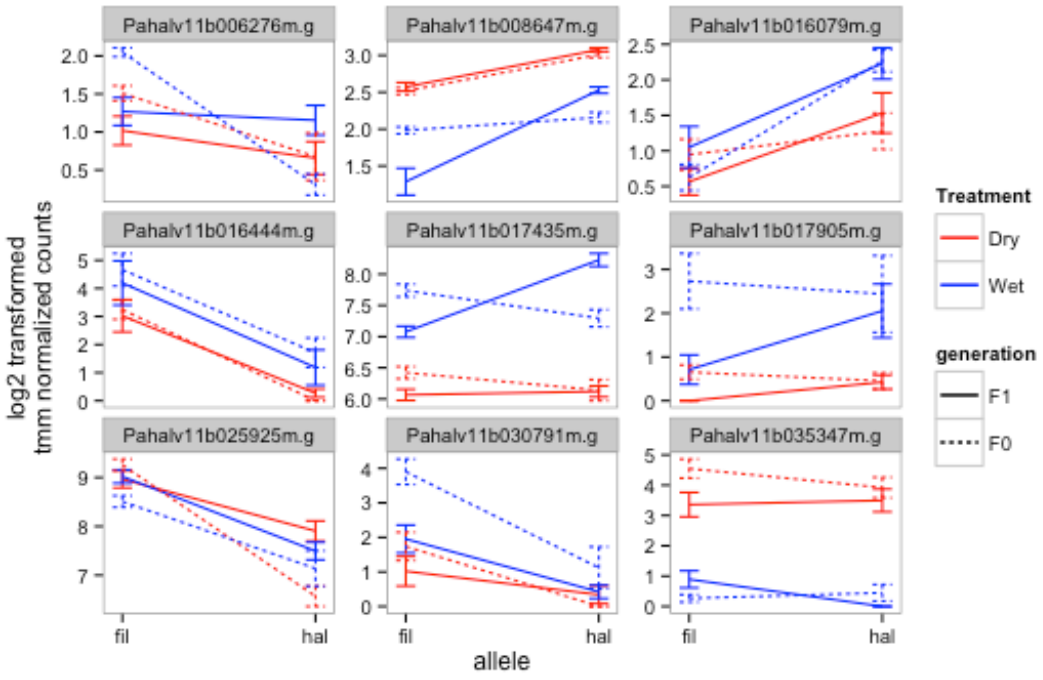


Figure S8. Reaction norms across the 9 genes significantly affected by trans*treatment interactions. Mean log2 transformed, variance stabilized and library size corrected counts are plotted for all genes with significant trans*trt effects. The Arabidopsis ortholog TAIR gene identifiers and pseudonyms for each gene can be found in Table 2.

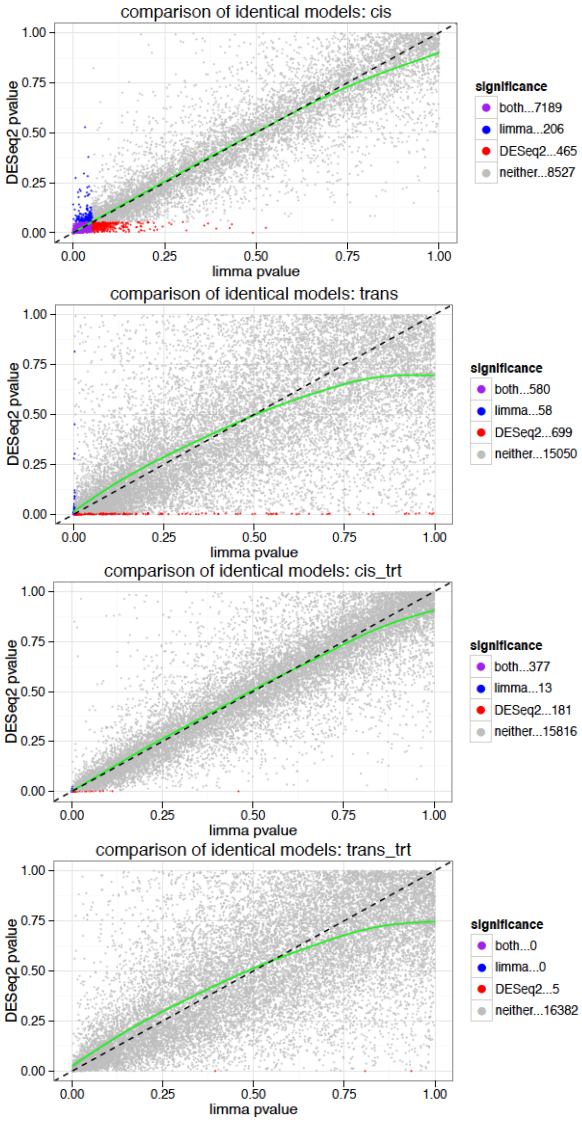
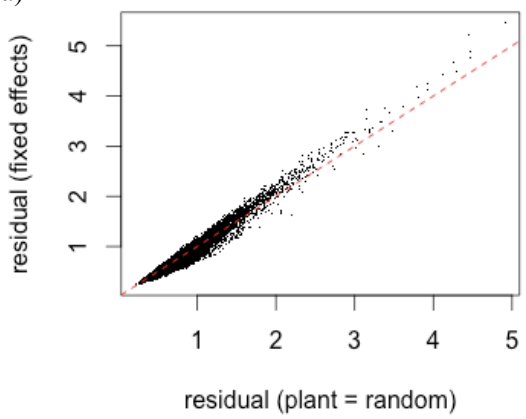
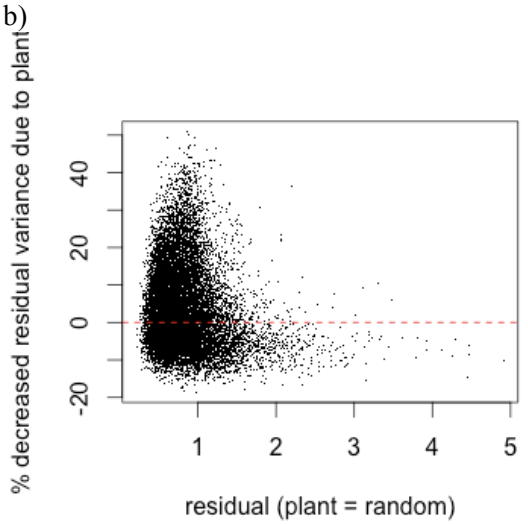


Figure S9. Comparison between LIMMA and DESeq2 models for the estimation of cis, trans and their interactions with treatment. P-values from each model are plotted on linear scales. The 1:1 line is plotted as the dashed black line. A gam (formula: y ~ s(x, bs = "cs")) smoothing curve is overlaid to demonstrate where the density of points lies. Points are colored by the method that determined significance. If neither model found significance, the points are grey. The number of genes in each category are reported in the legend.

a)



b)



c)

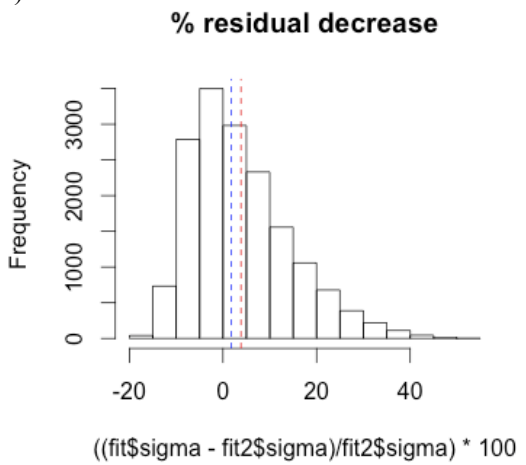


Figure S10: Comparisons of a fixed effects model in LIMMA to one with PlantID as a blocking factor. There is a general increase in the residual variance for the fixed effects model where the residual variance is very high, but the two models perform equally well with models with lower residual variance (a-b). Panel c displays a histogram of the y-axis in panel b, marking the mean

(red) and median (blue) change in the percent decrease in residual variance from the fixed effects model to that with plantID as a blocking factor.

IV: SUPPLEMENTARY TABLE CAPTIONS

Table S1. Significant GO terms for five categories of genes: 1) drought-responsive, 2) up-regulated in drought, 3) down-regulated in drought, 4) cis-by-treatment interaction affects expression, 5) trans regulated. FDR-corrected *P*-values are included along with GO id numbers and descriptions of each GO term. Terms are ordered by significance.

Table S2: Significantly enriched promoter motifs terms for 7 categories of genes. All motifs passing an $\alpha = 0.05$ threshold are reported. The annotations and references, parsed from the newPLACE database, accompany the motifs. The column "TEST" corresponds to the subset of genes analyzed: GxE genes with crossing reaction norms (green points, fig 2a) that were either up regulated ("gxe.halup.fildown") or down regulated in HAL2 in drought ("gxe.haldown.filup"), 3) Genes with very significant trans effects ("all.trans q<0.01"), 4) genes with significant signatures of compensatory evolution ("compensatory"), 5) genes with cis-by-treatment interactions ("cis.trt").

Column names: *ID* = motif ID (newPLACE), *RANK* = q value derived rank of each motif for each test, *RAW SCORE* = the affinity test statistic score, *P VALUE* = background corrected, *p*-value for the affinity test, *TEST* = the group of genes used to make the test (described above), *Q VALUE* = *q*-value calculated from the complete *p*-value distribution, *SEQUENCE* = The motif sequence, *ACCESSION#* = the newPLACE accession numeric identifier, *DESCRIPTION* = the newPLACE description, *ASSOCIATED* = newPLACE associated characteristics, *ORGANISM* = newPLACE organism in which the motif was identified, *REFERNCE* = newPLACE databased references that documented said motif.

V: SUPPLEMENTARY REFERENCES

Bader DM, Wilkening S, Lin G, Tekkedil MM, Dietrich K, Steinmetz LM, Gagneur J. 2015. Negative feedback buffers effects of regulatory variants. *Molecular Systems Biology* **11**: 785–785.

Fox J, Weisberg S. 2010. *An R Companion to Applied Regression*. SAGE.

Frith MC, Fu Y, Yu L, Chen J-F, Hansen U, Weng Z. 2004. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Research* **32**: 1372–1381.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* **27**: 297–300.

Love M, Anders S, Huber W, Love MM. 2013. Package "DESeq2."

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

404 Soneson C, Delorenzi M. 2013. A comparison of methods for differential expression analysis of
405 RNA-seq data. *BMC Bioinformatics* **14**: 91.

406 Storey JD. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical*
407 *Society: Series B (Statistical Methodology)* **64**: 479–498.

408 Storey JD, Taylor JE, Siegmund D. 2004. Strong control, conservative point estimation and
409 simultaneous conservative consistency of false discovery rates: a unified approach. *Journal*
410 *of the Royal Statistical Society: Series B (Statistical Methodology)* **66**: 187–205.

411