The evolutionary forces shaping cis and trans regulation of gene expression within a population of outcrossing plants.

Emily B. Josephs[1], Young Wha Lee[2], Corlett W. Wood[3], Daniel J. Schoen[4], Stephen I. Wright[5], John R. Stinchcombe[5]

[1] Dept. of Plant Biology, Michigan State University, East Lansing, MI, USA.
[2] Indigo Ag, Boston, MA, USA
[3] Dept. of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, USA
[4] Dept. of Biology, McGill University, Montreal, QC, Canada
[5] Dept. of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

Corresponding Author: Emily Josephs, josep993@msu.edu

Abstract

Understanding the persistence of genetic variation within populations has long been a goal of evolutionary biology. One promising route towards achieving this goal is using population genetic approaches to describe how selection acts on the loci associated with trait variation. Gene expression provides a model trait for addressing the challenge of the maintenance of variation because it can be measured genome-wide without information about how gene expression affects traits. Previous work has shown that loci affecting the expression of nearby genes (local or cis-eQTLs) are under negative selection, but we lack a clear understanding of the selective forces acting on variants that affect the expression of genes in trans. Here, we identify loci that affect gene expression in trans using genomic and transcriptomic data from one population of the obligately outcrossing plant, *Capsella grandiflora*. The allele frequencies of trans-eQTLs are consistent with stronger negative selection acting on trans-eQTLs than cis-eQTLs, and even more strongly on trans-eQTLs associated with the expression of multiple genes. However, despite this general pattern, we still observe the presence of a trans-eQTL at intermediate frequency that affects the expression of a large number of genes in the same coexpression module. Overall, our work highlights the different selective pressures shaping variation in cis- and trans- gene regulation.

Introduction

Understanding why genetic variation persists in populations has long been a goal of evolutionary biology [1]. Variation within populations may be 1) neutral and maintained by mutation-drift balance, 2) deleterious and maintained by mutation-selection balance, or 3) conditionally beneficial and maintained by balancing selection [2]. The availability of large genomic and phenotypic datasets offers the potential to evaluate the relative importance of these three hypotheses by identifying the genetic loci that are associated with a trait and using population genetic approaches to determine how selection acts on these loci [3,4]. In particular, the allele frequencies of eQTLs can provide information about selection, since negative selection against deleterious mutations is expected to keep alleles at lower frequencies than they will be under neutrality or balancing selection.

Gene expression has emerged as a powerful model trait for addressing the challenge of the maintenance of variation [5]. Gene expression is a crucial aspect of the genotype to phenotype map and

expression studies provide a large set of traits that can be easily measured without prior information about how these traits might relate to fitness [6]. Examining a large set of gene expression traits can reveal the evolutionary forces acting on traits in general, rather than a few or a handful of pre-defined traits chosen for specific reasons. The genetic variation that shapes expression can be partitioned into two categories: cis-regulatory variants that only affect the allele they are linked to and trans-regulatory variants, that affect both alleles equally and can be located near or far from the gene they regulate [7,8]. Previous work has mapped the genetic variants that affect expression (eQTLs) of nearby genes and shown that local eQTLs and cis-eQTLs are generally under negative selection [9–12]. However, trans-eQTLs may be under different selection pressures than cis-eQTLs. Since trans-regulatory variation can affecs the expression of multiple genes, trans-regulatory elements may have greater pleiotropic effects on phenotypes and be subject to stronger negative selection than cis-regulatory variants [13]. This prediction is supported by evidence of greater trans-regulatory variation within species compared to between species [7,14], reduced population frequencies of distant eQTLs compared to local eQTLs [15], and greater effect sizes of standing cis-regulatory variation than trans regulation [5,16,17], although these effect size differences may also be caused by differences in mutational input [18].

Despite the expectation that purifying selection will reduce trans-acting regulatory variation within species, there is evidence that trans-regulatory variation is common. Linkage mapping from crossing experiments and population-based association mapping have often found trans-regulatory hotspots, where genetic variation at a locus affects expression of numerous genes [11,16,19–23](but see [17]). Segregating trans-variation is more likely to be tissue-specific than cis-regulatory variation in humans [24] and, in *Arabidopsis thaliana*, trans-eQTLs are particularly important for expression changes in response to drought [23,25]. These findings suggest that trans-eQTLs contribute to standing variation, especially in specific tissues and environments.

Here we both map trans eQTLs for single genes and look for loci associated with the expression of many genes [26–28], in a single population of the plant *Capsella grandiflora*, an obligately outcrossing member of the Brassicaceae family with large effective population size and high levels of genetic sequence diversity [29,30]. To look for eQTLs affecting the expression of multiple genes, we use coexpression networks to summarize expression across many genes and test for associations between genetic variants and the expression of network modules. Coexpression networks are a powerful way to find patterns in large transcriptomic datasets [17,31–35]. For example, coexpression networks made across conditions, tissues, and developmental time can successfully identify specialized metabolic pathways [32] and coexpression modules made with a diverse panel of mouse lines correlate with phenotype [36]. In addition, changes in coexpression module expression have been linked to adaptation [37] and changing ecological conditions [35]. We detect a large number of putative cis and trans eQTLs and show that, based on allele frequencies, trans eQTLs are generally under stronger negative selection than cis eQTLs. We use coexpression networks to summarize expression levels across many genes and detect four eQTL for coexpression module expression. Overall, our results suggest that negative selection acts on trans eQTLs more strongly than cis, but there are some trans-eQTLs affecting large numbers of genes at appreciable frequencies in the population.

Results

Linking allele frequencies and selection in cis and trans eQTLs.

We tested for associations between leaf expression at all genes with 1,873,867 tag SNPs in 146 individuals. We identified 6,231 associations (FDR < 0.1) between 5,468 unique snps and 2,341 genes **(Figure 1)**. These eQTLs were associated with the expression of between 1 and 93 genes. We separated associations into 2,472 trans eQTLs that were more than 5kb away from the gene they regulated and 3759 local (putatively cis) eQTLs that were less than 5kb away from the gene they regulated. We will refer to these eQTLs as cis eQTLs for clarity, while noting that some of them may be caused by trans-eQTLs located near the genes they regulate. 3,300 of these cis-eQTLs were detected previously in a subset of 99 individuals at $p<0.05$, and 2,636 at an FDR of 0.1 (the cutoff used in that study) [9]. Trans-eQTLs had larger effect sizes than cis-eQTLs ($p < 0.001$, mean local effect size = 0.91, mean trans effect size = 1.01).

We used the minor allele frequency ('MAF') to infer the relative strength of selection acting on different types of eQTLs. Trans-eQTLs had lower MAFs than than cis-eQTLs (**Figure 2,** $p < 0.001$, mean trans MAF = 0.214, mean cis MAF = 0.267). This result was robust to the cutoff distance used to define cis- and trans-eQTLs: trans-eQTLs had lower MAFs than cis-eQTLs for distance cutoffs of 1 kb, 2.5 kb, and 10 kb ($p < 0.001$). This difference in allele frequencies is consistent with stronger negative selection on trans-eQTLs than cis-eQTLs. However, since trans-eQTLs have larger effects on expression, there may be more power to detect trans-eQTLs at lower frequencies, causing the observed pattern. We found that when we restricted our analysis to eQTLs in the top quartile of effects (684 cis-eQTL and 692 trans eQTL), trans-eQTLs were still present at lower minor allele frequencies than cis-eQTLs, consistent with allele frequency differences resulting from stronger negative selection on trans-eQTLs than cis-eQTLs ($p < 0.001$, mean cis MAF = 0.115, mean trans MAF = 0.098).

We also investigated the MAFs of eQTLs associated with the expression of many genes compared to the MAFs of eQTLs associated with only one gene. We restricted this analysis to trans eQTLs, to ensure that the differences between cis and trans eQTLs described above did not confound the results. Within trans eQTLs, there were 1,635 eQTLs associated with one gene and 376 eQTLs associated with the expression of more than one gene at an FDR < 0.1. The MAF of eQTLs with one association was significantly higher than the MAF of eQTLs with more than one association ($p < 0.001$, one gene MAF = 0.214, more than one gene MAF = 0.196). This result is consistent with negative selection acting more strongly on eQTLs with many associations than eQTLs with one one association.

Coexpression module GWAS identifies SNPs affecting expression modules

We identified 16 coexpression modules ranging in size from 69 to 6392 genes **(Figure S1)**. We summarized expression level across modules, which we will refer to as 'module expression', for each individual using eigengenes. Module expression is not correlated with collection timing **($p > 0.2$, Figure S2).** Module expression values show varying distributions: some modules had normal distributions, some were bimodal, and some showed strong skews where a few individuals had very high module expression compared to other individuals **(Figure S3)**. Because the skewed distribution of module expression values could lead to false positives during association mapping, we quantile-normalized expression level for association mapping.

Genome-wide association mapping for module expression identified four SNPs associated with the expression of two modules (FDR < 0.1, **Table 1)**. We refer to these SNPs as 'coexpression-eQTLs'. All four of the coexpression-eQTLs were also identified as eQTLs in the all-by-all analysis. Two coexpression-eQTLs are associated with expression of the 'lightyellow' module and two with the expression of the 'white' module. **(Figure 3, Table 1)**. Both eQTLs for the 'white' module were located near each other (1.2 kb apart). We mapped the association between all SNPs (not just tagging SNPs) and 'white' module expression in this

region and found additional significant associations (**Figure 3B)**, suggesting that there is a longer block of loci in linkage disequilibrium associated with module expression.

We further investigated the gene containing the one coding coexpression eQTL, which is associated with expression of the 'lightyellow' module. This eQTL was in a 4-fold degenerate site of the gene Carubv10025970m. Its closest ortholog in *Arabidopsis thaliana*, AT5G65683.1 or WAV3 HOMOLOG 2 is a member of the WAVY GROWTH 3 E3 ligase family which is involved in root gravitropism but also shows expression in Arabidopsis young leaves. This coexpression eQTL is also associated with the expression of 93 genes in the all-by-all analysis. All but one of these 93 genes was in the "light yellow" module. The minor allele frequency of this eQTL is 0.317.

Relating coexpression modules to traits.

We also conducted GWAS on phenotypic traits (days to bolting, days to flower, leaf nitrogen content, leaf carbon content and leaf shape traits) following the same procedures described above for coexpression modules. No associations were significant at a FDR < 0.1 or even at an FDR < 0.25. Module expression was correlated with a number of trait measurements. There were four modules whose expression was correlated with days to bolt (**Figure 4**, $p < 0.05$ after Bonferroni correction for 16 tests). None of the four coexpression eQTLs detected were significantly associated with any phenotypes.

Population genetic signatures of selection on eQTLs.

We compared signatures of selection around local and trans eQTLs identified by the all-by-all analysis and found that trans-eQTLs were in windows that had lower Tajima's D on average (Tajima's D = -0.596) than cis-eQTLs (Tajima's D = -0.455, $p < 0.0001$), consistent with the regions around trans eQTLs being under stronger negative selection than regions around cis-eQTLs, although this could also result from differences in the genomic landscapes of trans-eQTLs compared to cis-eQTLs.

While we have evidence that local cis-regulatory eQTLs are in general under negative selection in this population [9], and our present analysis suggests that all-by-all trans-eQTL are subject to stronger puriifying selection, we were curious if we could detect evidence of recent positive or balancing selection on eQTLs detected in the coexpression eQTL analysis as well as in the all-by-all analysis. We measured π and Tajima's D at putatively neutral sites across the genome in 500 bp windows and used SweeD to test for evidence of selective sweeps in 50 SNP windows. None of the coexpression eQTLs were located in windows that were outliers (top 2.5% of windows) for π, Tajima's D, or sweep likelihood (**Fig. S4, Fig. S5, Fig. S6**).

Discussion

In this study, we have mapped the genetic basis of genome-wide expression variation within a single population of an outcrossing plant. The allele frequencies of trans-eQTLs suggest that the variants that affect trans-regulation are under stronger negative selection than cis-eQTLs, and that trans-eQTLs associated with the expression of multiple genes are under stronger negative selection than trans-eQTLs associated with the expression of only one gene. In addition, windows containing trans eQTLs have lower Tajima's D than window containing cis-eQTLs, also consistent with stronger negative selection acting on trans-eQTLs. However, despite the general pattern of negative selection acting on trans-eQTLs, we detected four eQTLs associated with the expression of coexpression network modules, one of which is independently associated

with the expression of 93 genes and present at an intermediate frequency. Overall, this work suggests that trans-eQTLs are under different selective pressures than cis-eQTLs, but that within trans-eQTLs there is also a great deal of variation in selection.

Our results are consistent with previous work showing that distant eQTLs are at lower minor allele frequencies than local eQTLs in *Arabidopsis thaliana[15]*. However, the *A. thaliana* result comes from association mapping done in a diverse panel of lines from across the species range where allele frequencies could be shaped by negative selection or local adaptation, so our results reflect a clearer indication that negative selection acts more strongly on trans-eQTLs than cis-eQTLs. In addition, the pattern that trans-eQTLs that are associated with the expression of multiple genes are at lower minor allele frequency than trans-eQTLs with only one association is consistent with evidence that negative selection acts more strongly on pleiotropic loci in *C. grandiflora [34]* and in other species [13,38].

One important aspect of our use of coexpression modules in the eQTL analysis is that we used "genotype networks' generated from expression data measured in the same tissue type at the same time in a set of genetically distinct individuals. Therefore, the coexpression modules we observed were shaped by genetic perturbations, not tissue or developmental differences. While coexpression measured across multiple timepoints ('developmental networks') has been linked to functional relationships [39,40], coexpression modules generated from genetically distinct individuals have different properties than those generated from different tissue types [17,41]. In some cases, this difference is helpful: analyses combining GWAS and coexpression networks have the most power when using coexpression networks made from genetically distinct samples [41]. However, it is important to keep in mind that the expression datasets used will affect coexpression modules.

Mapping eQTLs has furthered our understanding of the nature of genetic variation maintained within natural populations. Analyses combining genomic and transcriptomic data from natural populations are relevant in the context of models using transcriptomic data to build a mechanistic understanding of the evolutionary forces maintaining variation within populations [42–44]. In addition, since gene expression is important for adaptive divergence [45–47], understanding the maintenance of genetic variation for expression is important for understanding how organisms will adapt to new environments.

Materials and Methods

Genomic, transcriptomic, and phenotypic data

All genomic and transcriptomic sequence data was previously published in Josephs et al. (2015) and Josephs et al. (2017b). We briefly describe data generation here. We collected individuals from a single population of *C. grandiflora* individuals located near Monodendri, Greece. We conducted a generation of random crosses in the greenhouse, and then grew 146 individuals descended from these random crosses in a growth chamber with 16 hours of daylight at 22ºC. We measured traits on these individuals and collected RNA from leaf tissue collected 39 days after planting. Leaves were collected and flash frozen at night sequentially in 9 roughly-equally sized bins, which we will refer to as "collection bin". We extracted RNA using Qiagen RNAeasy kits.

We extracted DNA from leaf material using a CTAB procedure. Both RNA and DNA was sequenced at the Genome Quebec facility with Hiseq 2000 with Truseq libraries with 100bp long reads. DNA was mapped to the standard *C. rubella* reference genome [48] with Stampy [49] and RNA was mapped to an

exon-only reference genome using Stampy as well. SNPs were called from the genomic sequence data using GATK Unified Genotyper [50] and expression levels were measured with HTseq and normalized for sequencing depth by dividing by the median expression level for each individual. [51]. We did not detect interactions between GC content, expression level, and lane [9]. We used the ComBat function from the sva R package to adjust expression level for collection batch [52].

In addition to collecting RNA and DNA for sequencing on these 146 individuals, we measured a number of phenotypes. We measured days to bolting and days to flowering daily (measured since planting date). We collected leaves at day 49 after planting, scanned leaves, and measured leaf shape as reported in [53]. Briefly, dissection index was calculated as DI = $(\text{perimeter}^2)/(4\pi*\text{area})$, so that a circle of the same area would have a value of 1.0 and increasing values indicate increasing complexity and alpha shape dissection index is a similar parameter, but for alpha shapes. We measured leaf carbon and nitrogen content in one leaf per individual. Leaves were collected at day 49 after planting, dried, and ground to powder for elemental analysis by the Ecosystems Analysis Lab at the University of Nebraska. We note that both shape and elemental data came from different leaves than the RNAseq data. We estimated Pearson and spearman correlations between module expression and trait values with the cor.test function in R [54].

Building coexpression networks

We used the program WGCNA [55](version 1.68 running on R version 3.6.2) to identify coexpression modules present within the 145 transcriptomes using the expression level of all genes with median expression greater than 5 reads per gene (n = 18,806). The coexpression analysis groups together genes with similar patterns of pairwise correlation of expression. We were interested in retaining the information embodied in the sign of the gene expression correlations, so we conducted a signed network analysis. We used a soft thresholding value of 12, as suggested by the authors of the WGCNA package for signed networks and a minimum module size of 30. Genes that exhibited similar patterns of connectivity (i.e., genes showing high "topological overlap") were grouped together in the same coexpression modules, based on hierarchical clustering of topological overlap values, in which a dynamic branch-cutting algorithm was used to define initial gene co-expression modules. Module eigengenes (the first principal component of the gene expression values of modules) were calculated, and modules whose eigengenes were correlated at a level greater than 0.8 were merged to arrive at the final set of co-expression modules. The resulting modules were labeled with different colors for ease of referencing[34]. We investigated specific eQTLs in the Joint Genome Institute's Phytozome v12.1 genome browser for *Capsella rubella* v1.0[48] and investigated specific orthologs in *Arabidopsis thaliana* using TAIR[56].

Some of the modules had expression levels that were very skewed, such that a few individuals showed extremely high module expression compared to the rest of the individuals **(Fig S1)**. To reduce potential false-positives in the association mapping study due to skewed expression levels, we quantile normalized module expression levels using the qqnorm function in R [54].

Association mapping

We tested for associations between SNP genotype and individual gene expression, phenotypes, and module expression. For all association mapping analyses, we filtered out SNPs with a minor allele frequency below 0.01 and more than 0.05 missing data, leaving 5,560,798 SNPs. We used Haploview to identify 1,873,867 tag SNPs with minor allele frequency > 0.05 that described the dataset.

We tested for associations between the tag SNPs with the expression of 18,806 genes using the linear model in Matrix eQTL [57]. We quantile normalized gene expression levels using the qqnorm function in R to reduce false positives caused by skewed expression distributions. While all samples came from the same population, we controlled for residual population structure by generating a centered kinship matrix with GEMMA [58] and including the first five principal components of the kinship matrix as covariates.. Since all tag SNPs were tested against all genes, we conducted 35,587,985,444 tests. Matrix eQTL estimates false discovery rates using a Benjamini–Hochberg procedure.

After detecting eQTLs, we compared cis and trans eQTLs. We defined putative cis eQTLs as eQTLs that were less than 5 kB from the transcription start or end site of the gene they were associated with and trans eQTLs as eQTLs that were either on a different chromosome from the gene they were associated with or more than 5 kB away from the transcription start or end site of the chromosome. We compared the minor allele frequencies and effect sizes of eQTLs using the t.test function in R [54].

We did association mapping with GEMMA [58] on module eigengenes (PC1 of expression values of a module), morphological, and life history traits as our phenotypes. We controlled for residual population structure using the standardized kinship matrix and SNPs with minor allele frequency > 0.05 and missing data < 0.05. We used the likelihood ratio p values [59] and calculated the p-value cutoffs corresponding to a false discovery rate of 0.1 for each trait and module expression level using the FDR method of the p.adjust function in R [60].

Population genetic signatures of selection

We used genomic sequence from 188 individuals published in [9]. We downsampled all sites to 320 chromosomes per site and then calculated pi and Tajima's D in 500 bp windows across the genome at non-coding (excluding conserved non-coding sites from [61]), intronic, and 4-fold degenerate sites. We used SweeD to calculate the likelihood of a selective sweep occuring on every 50th SNP (windows were ~600 bp wide on average) using non-conserved intergenic, intronic, and 4-fold degenerate sites for 182 individuals from the focal population [62].

Acknowledgements

We thank Niroshini Epitawalage, Amanda Gorton, Robert Williamson and Jasmina Uzunović for research assistance, as well as Graham Coop, Jeff Ross-Ibarra, and members of the Stinchcombe, Wright, Coop and Ross-Ibarra labs for helpful comments and advice. This work was supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research. This work was funded by a National Science Foundation Graduate Research Fellowship (DGE-1048376) and a National Science Foundation Plant Genome Postdoctoral Fellowship (IOS- 1523733) to EBJ, Natural Sciences and Engineering Research Council of Canada Discovery Grants to DJS, SIW, and JRS and a Value-directed Evolutionary Genomics Initiative grant (Genome Quebec/Genome Canada) to DJS, SIW, and JRS.

Table 1: Information about significant coexpression eQTLs (FDR < 0.1). 'CNS' stands for 'conserved noncoding sequence', 'MAF'' stands for minor allele frequency

Module	SNP	MAF	p	FDR	Site Type
lightyellow	scaffold_7:17305936	0.093	1.03E-07	0.096	intergenic
lightyellow	scaffold_8:12634406	0.317	6.94E-09	0.013	exon
white	scaffold_4:12621988	0.066	2.86E-08	0.027	intergenic
white	scaffold_4:12623220	0.052	9.62E-09	0.018	CNS

Figures

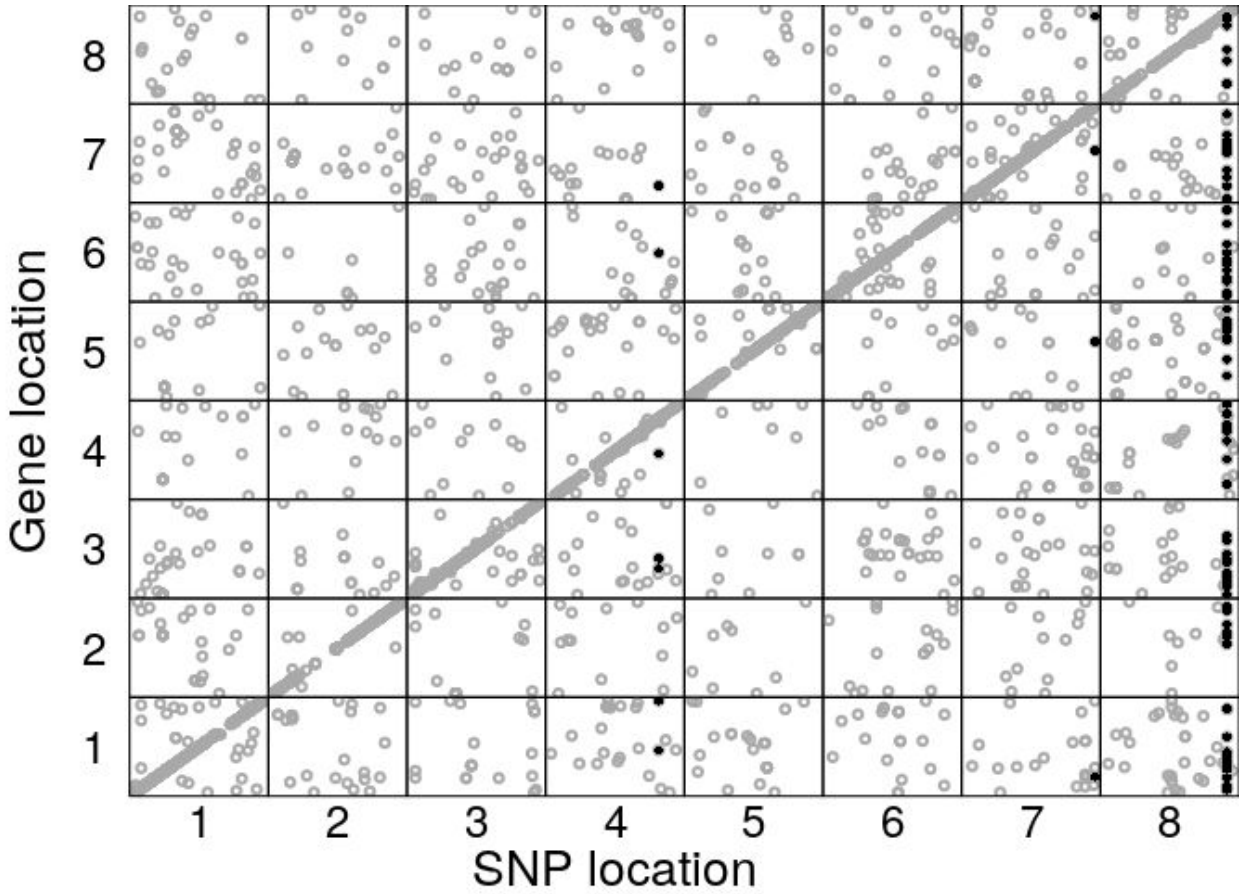

Figure 1: All-by-all eQTL locations. Each point represents a SNP, whose location is described on the X axis, that is associated with the expression of a gene whose location is described on the y axis (FDR < 0.1). Black points are SNPs that were also identified in the coexpression eQTL analysis.

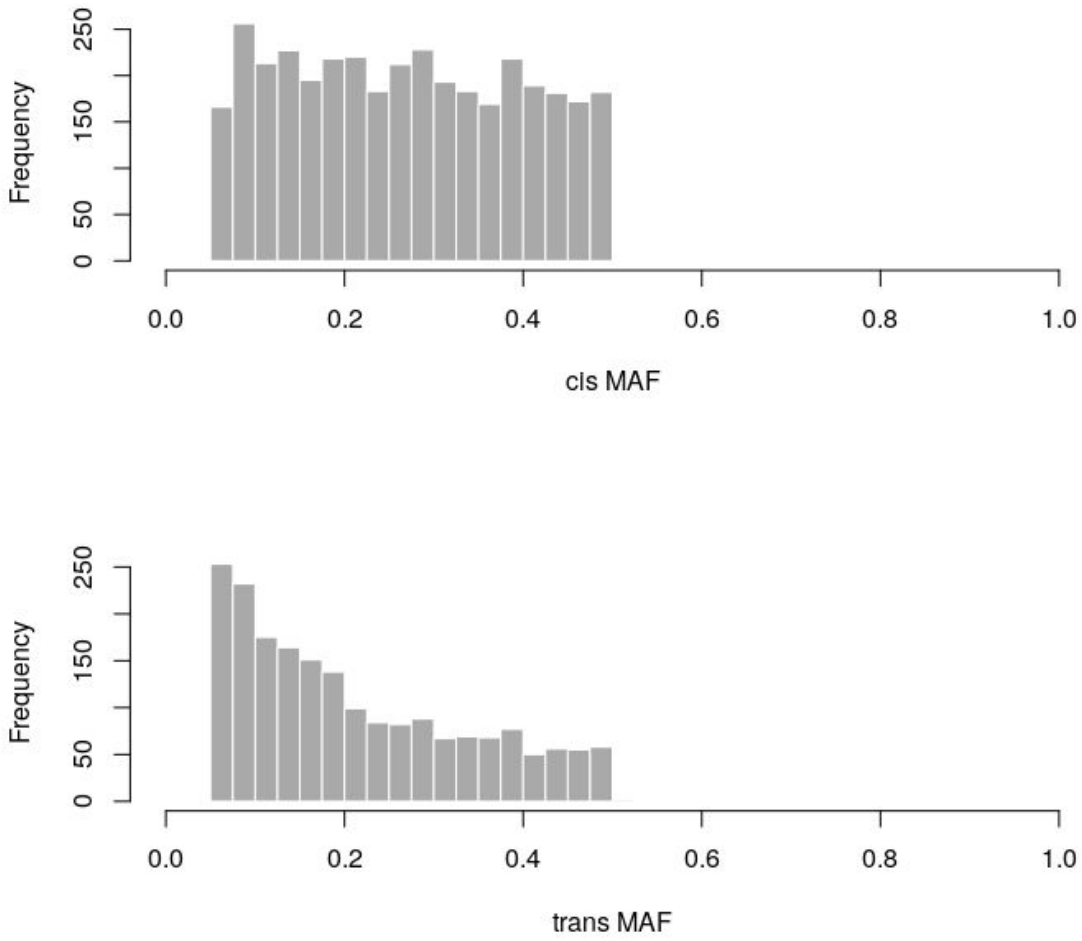

Figure 2 Distribution of minor allele frequencies (MAF) of cis and trans eQTLs.

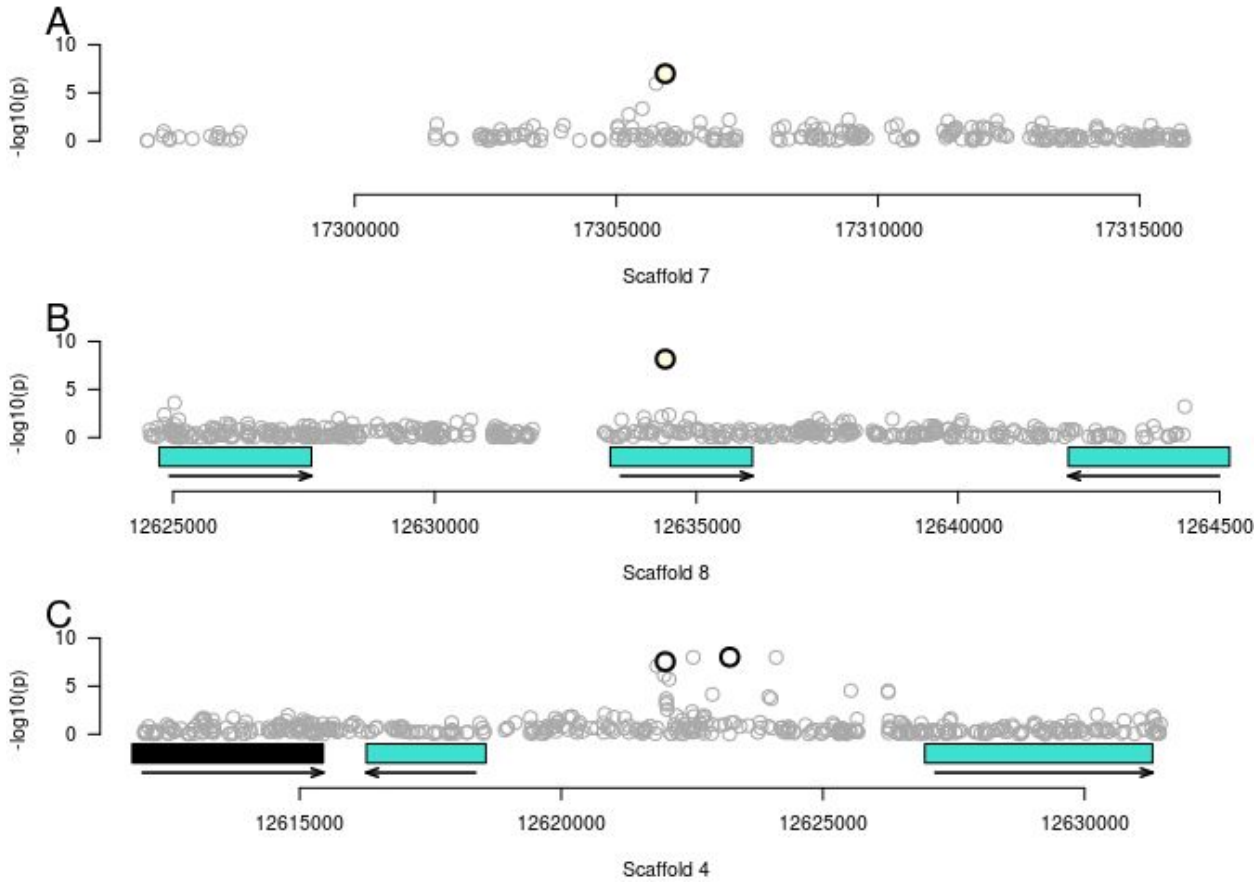

Figure 3: Physical locations of coexpression QTLs. Coexpression eQTLs are represented by white points with black borders. All other SNPs are plotted in gray. These plots include associations for all snps, not just the tagging snps, plotted by location, on the x axis, and the significance of association with the module indicated by color on the y axis. The locations of nearby genes are shown by rectangles, colored by the module the gene is in. The direction of transcription of each gene is shown by a black arrow. A) The first coexpression eQTL on Scaffold 7, B) the second coexpression eQTl on scaffold 8, and C) the third and fourth coexpression eQTLs on Scaffold 4.

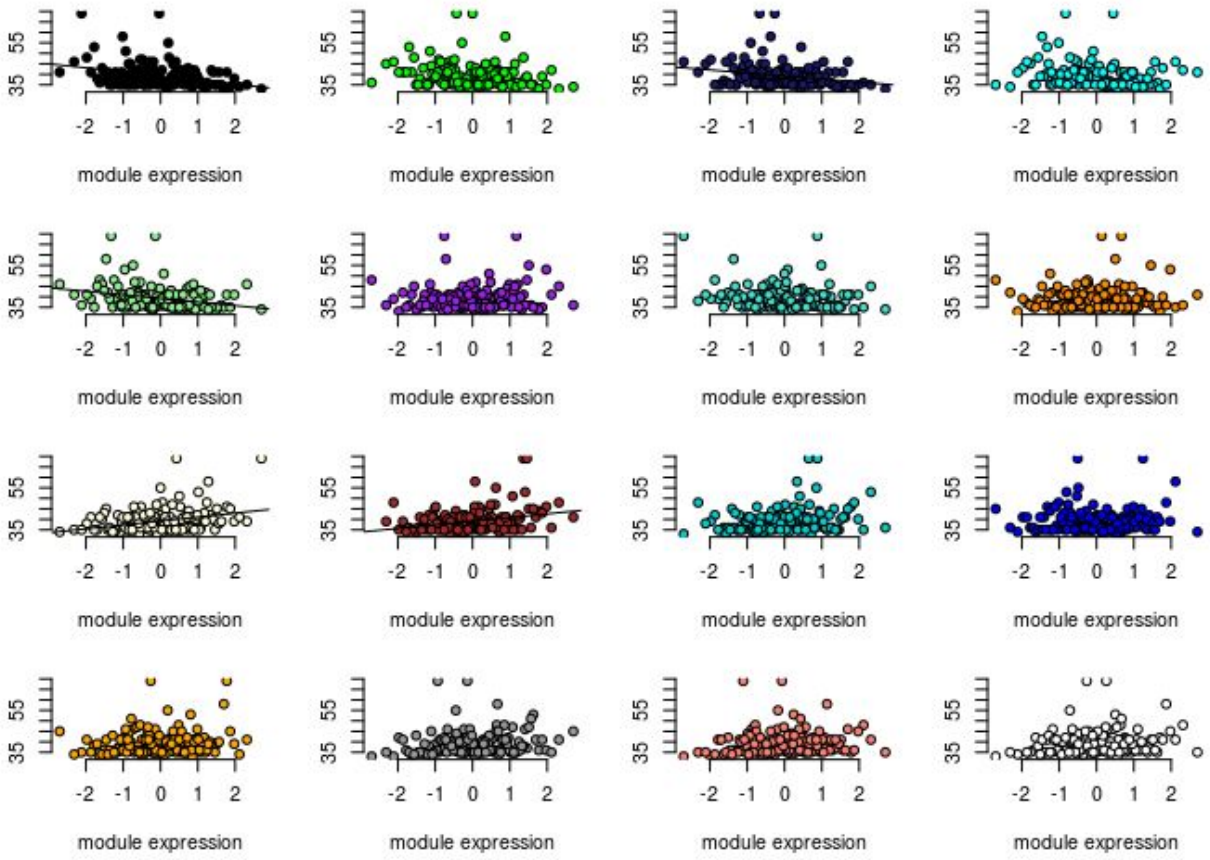

Figure 4: Correlations between module expression level and traits. The correlation between expression of each module (x axis) and days to flower (y axis). If the correlation was significant ($p < 0.05$ after Bonferonni correction), there is a black trend line drawn.

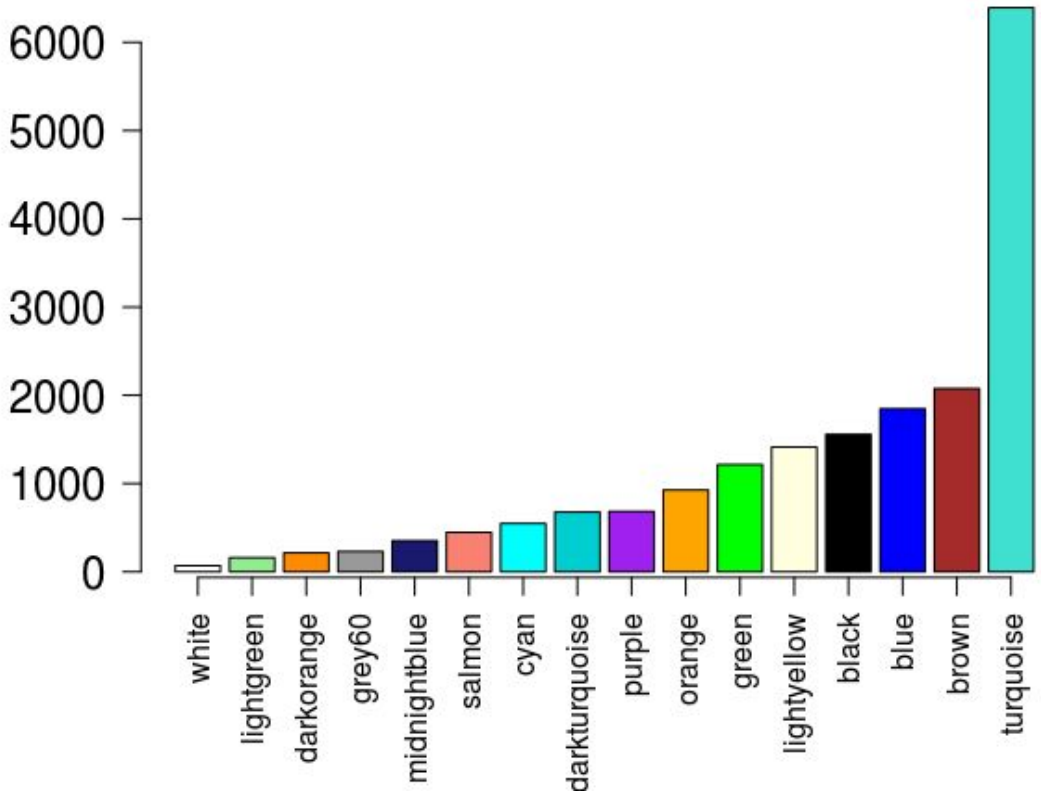

Figure S1: The number of genes in each module.

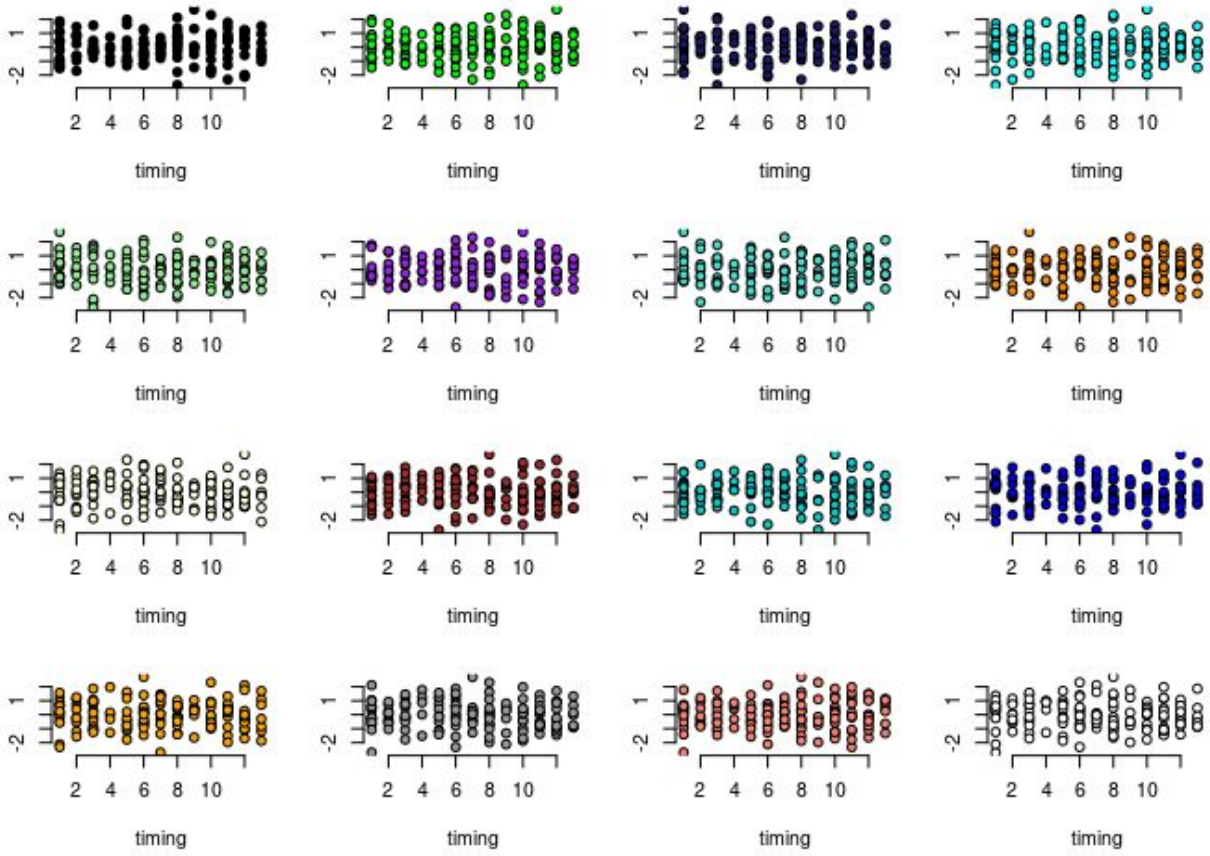

Figure S2. Relationship between collection box (x axis) and module expression (y axis) for 16 coexpression modules. Each point represents one plant.

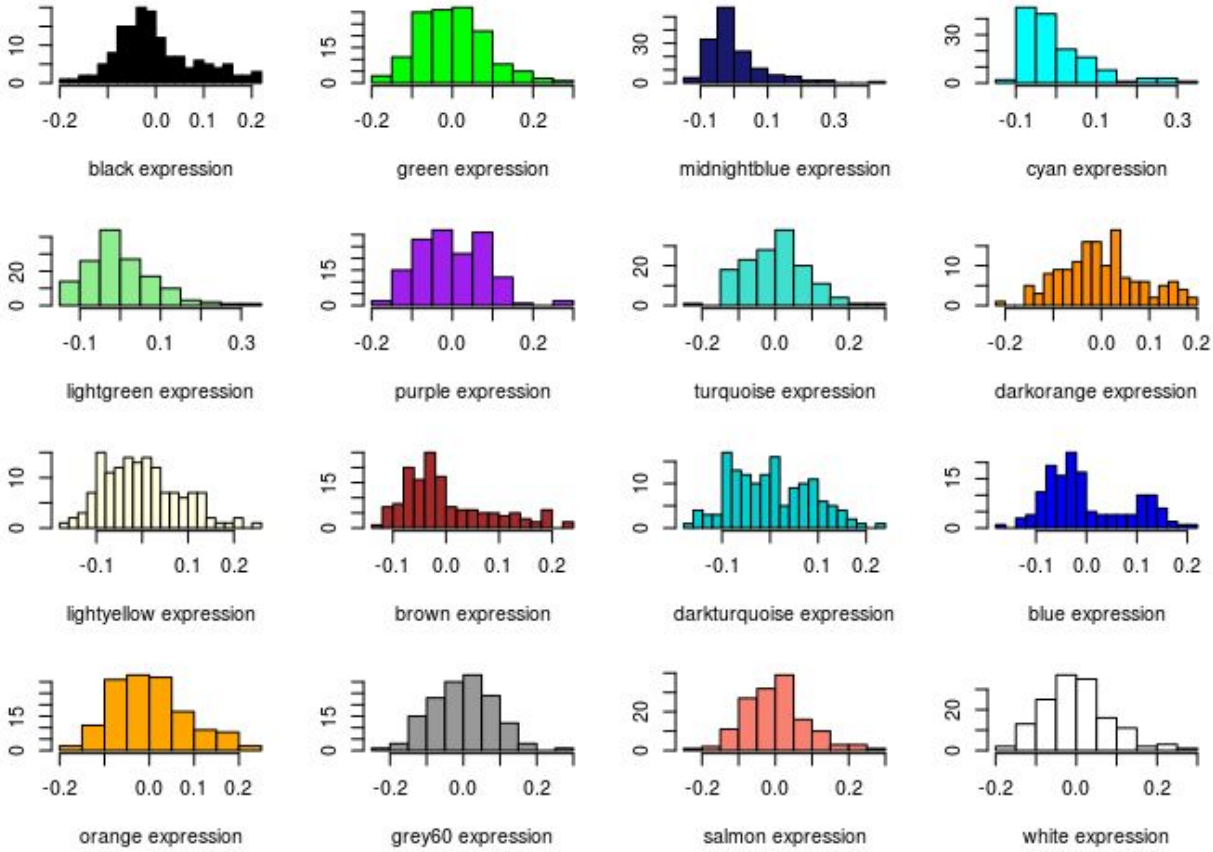

Figure S3: Distribution of un-normalized eigengenes. Each plot shows a histogram of eigengene expression for a specific coexpression module. The module name is labeled below the x axis.

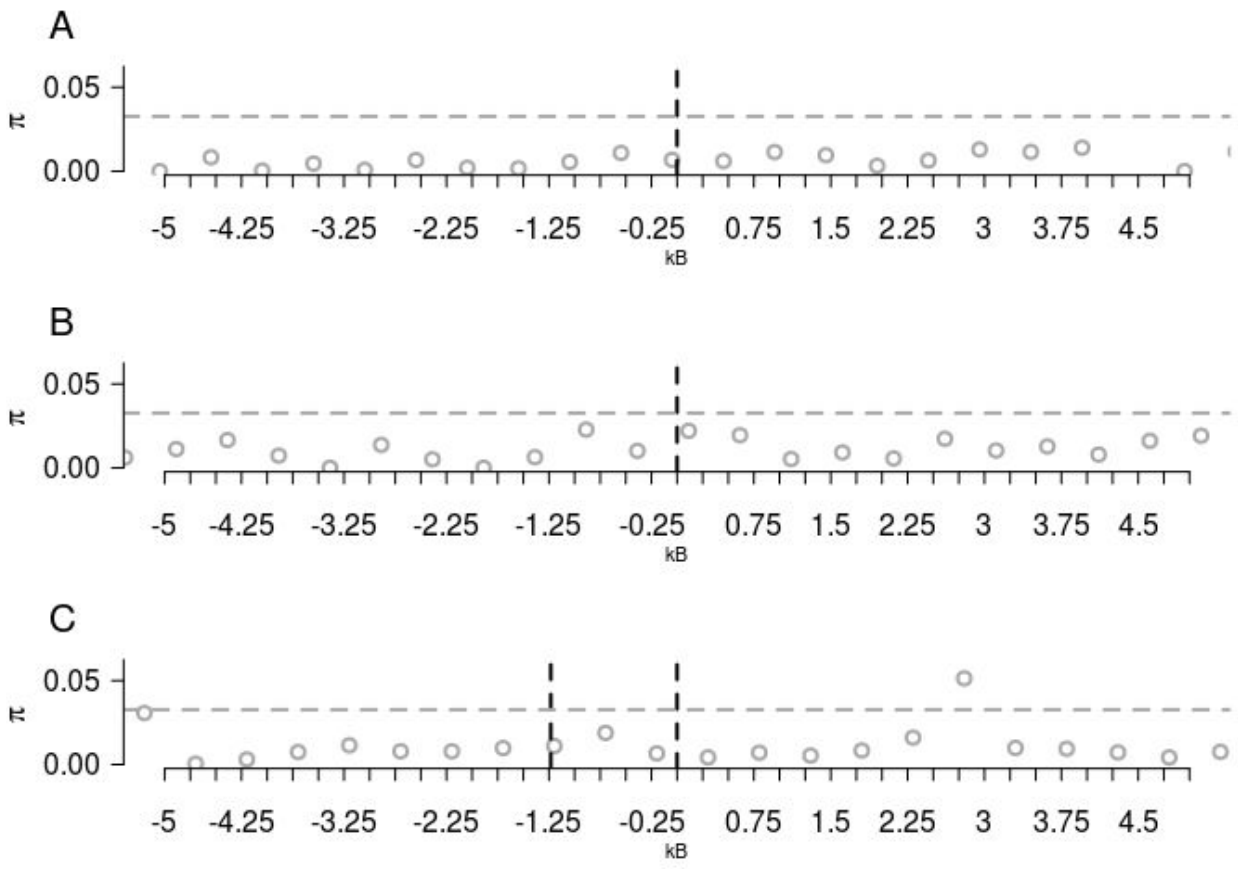

Figure S4: Diversity (**π**) at putatively neutral sites in 500 bp windows around coexpression-eQTLs. Gray horizontal dotted lines show 95% cutoffs of the observed distribution of likelihoods and black vertical lines show the locations of eQTLs. Each panel corresponds to the eQTL shown in **Fig. 3.**

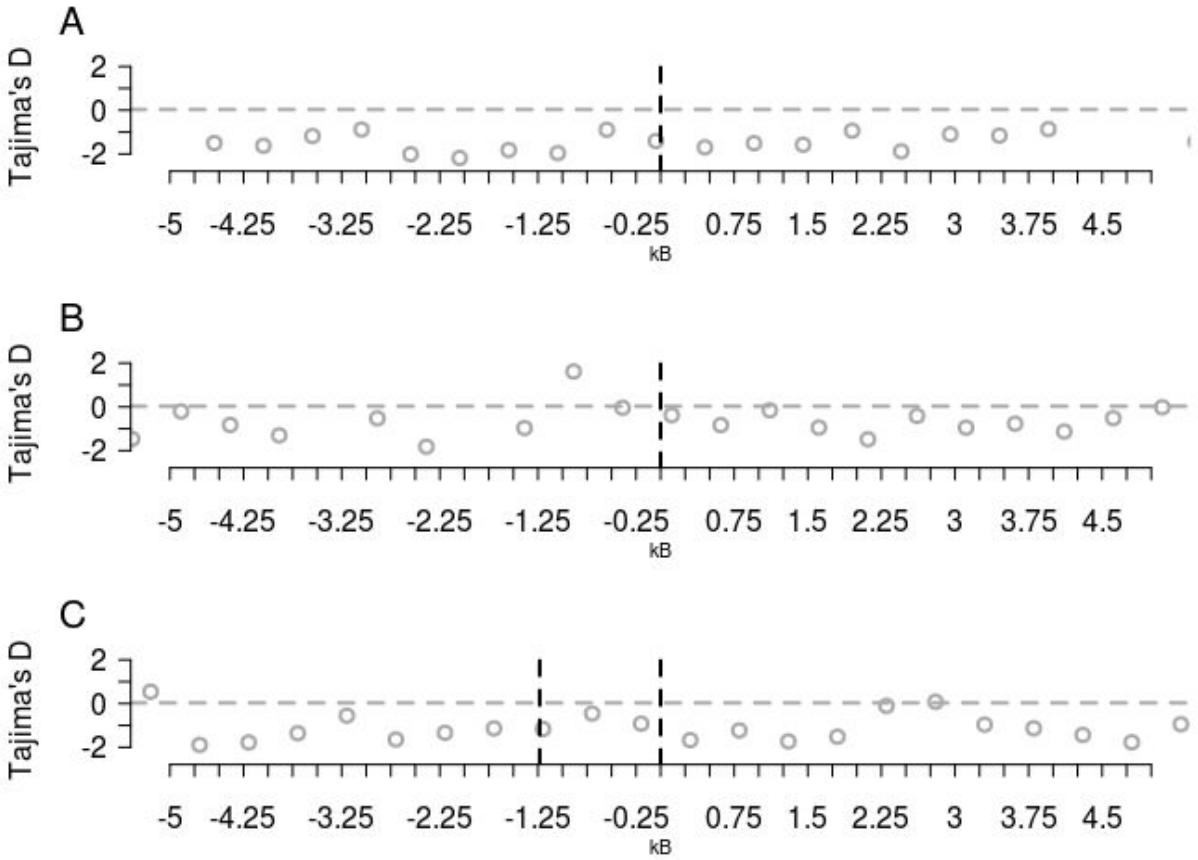

Figure S5: Tajima's D at putatively neutral sites in 500 bp windows around coexpression-eQTLs. Gray horizontal dotted lines show 95% cutoffs of the observed distribution of likelihoods and black vertical lines show the locations of eQTLs.. Each panel corresponds to the eQTL shown in **Fig. 1.**

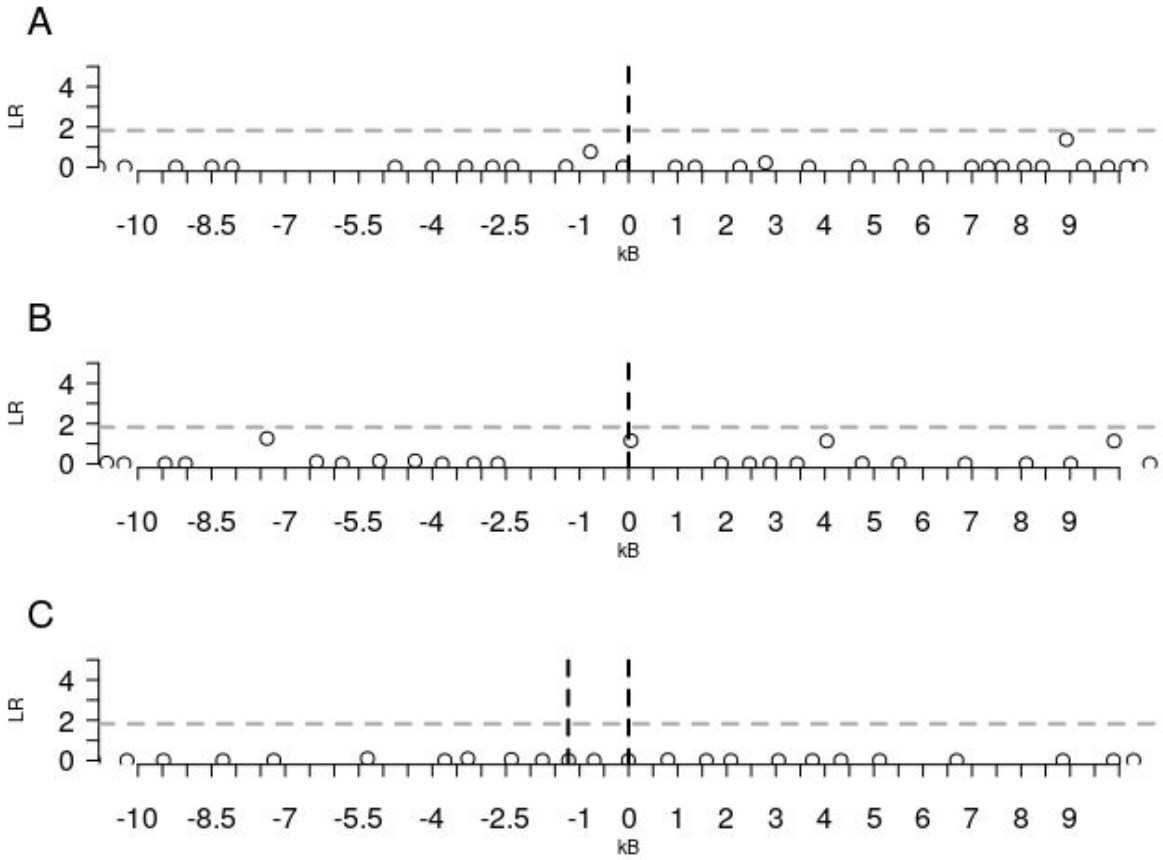

Figure S6: SweeD sweep likelihoods in windows around coexpression-eQTLs. Gray horizontal dotted lines show 95% cutoffs of the observed distribution of likelihoods and black vertical lines show the locations of eQTLs. Each panel corresponds to the eQTL shown in **Fig. 1.**

References

1. Mitchell-Olds T, Willis JH, Goldstein DB. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat Rev Genet. 2007;8: 845–856.

2. Johnson T, Barton N. Theoretical models of selection and mutation on quantitative traits. Philos Trans R Soc Lond B Biol Sci. 2005;360: 1411–1425.

3. Sella G, Barton NH. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. Annu Rev Genomics Hum Genet. 2019. doi:10.1146/annurev-genom-083115-022316

4. Josephs EB, Stinchcombe JR, Wright SI. What can genome‐wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? New Phytol. 2017. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.14410

5. Kliebenstein D. Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. Annu Rev Plant Biol. 2009;60: 93–114.

6. Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet. 2006;7: 862–872.

7. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430: 85–88.

8. Emerson JJ, Li W-H. The genetic basis of evolutionary change in gene expression levels. Philos Trans R Soc Lond B Biol Sci. 2010;365: 2581–2590.

9. Josephs EB, Lee YW, Stinchcombe JR, Wright SI. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. Proc Natl Acad Sci U S A. 2015;112: 15390–15395.

10. Glassberg EC, Gao Z, Harpak A, Lan X, Pritchard JK. Evidence for Weak Selective Constraint on Human Gene Expression. Genetics. 2019;211: 757–772.

11. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. 2014;24: 14–24.

12. Hernandez RD, Uricchio LH, Hartman K, Ye C, Dahl A, Zaitlen N. Ultrarare variants drive substantial cis heritability of human gene expression. Nat Genet. 2019;51: 1349–1355.

13. McGuigan K, Collet JM, Allen SL, Chenoweth SF, Blows MW. Pleiotropic mutations are subject to strong stabilizing selection. Genetics. 2014;197: 1051–1062.

14. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet. 2008;40: 346–350.

15. Zhang X, Cal AJ, Borevitz JO. Genetic architecture of regulatory variation in Arabidopsis thaliana. Genome Res. 2011;21: 725–733.

16. Liu H, Luo X, Niu L, Xiao Y, Chen L, Liu J, et al. Distant eQTLs and non-coding sequences play critical

roles in regulating gene expression and quantitative trait variation in maize. Mol Plant. 2016. doi:10.1016/j.molp.2016.06.016

17. Mähler N, Wang J, Terebieniec BK, Ingvarsson PK, Street NR, Hvidsten TR. Gene co-expression network connectivity is an important determinant of selective constraint. PLoS Genet. 2017;13: e1006402.

18. Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ. Contrasting frequencies and effects of cis- and trans-regulatory mutations affecting gene expression. Mol Biol Evol. 2016. doi:10.1093/molbev/msw011

19. Rockman MV, Skrovanek SS, Kruglyak L. Selection at linked sites shapes heritable phenotypic variation in C. elegans. Science. 2010;330: 372–376.

20. Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. Genetics of trans-regulatory variation in gene expression. Elife. 2018;7. doi:10.7554/eLife.35471

21. West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, et al. Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics. 2007;175: 1441–1450.

22. Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, et al. Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci U S A. 2007;104: 1708–1713.

23. Lowry DB, Logan TL, Santuari L, Hardtke CS, Richards JH, DeRose-Wilson LJ, et al. Expression quantitative trait locus mapping across water availability environments reveals contrasting associations with genomic features in Arabidopsis. Plant Cell. 2013;25: 3266–3279.

24. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550: 204–213.

25. Clauw P, Coppens F, Korte A, Herman D, Slabbinck B, Dhondt S, et al. Leaf Growth Response to Mild Drought: Natural Variation in Arabidopsis Sheds Light on Trait Architecture. Plant Cell. 2016;28: 2417–2434.

26. Kliebenstein DJ, West MAL, van Leeuwen H, Loudet O, Doerge RW, St Clair DA. Identification of QTLs controlling gene expression networks defined a priori. BMC Bioinformatics. 2006;7: 308.

27. Brynedal B, Choi J, Raj T, Bjornson R, Stranger BE, Neale BM, et al. Large-Scale trans-eQTLs Affect Hundreds of Transcripts and Mediate Patterns of Transcriptional Co-regulation. Am J Hum Genet. 2017;100: 581–591.

28. Hore V, Viñuela A, Buil A, Knight J, McCarthy MI, Small K, et al. Tensor decomposition for multiple-tissue gene expression experiments. Nat Genet. 2016. doi:10.1038/ng.3624

29. Slotte T, Foxe JP, Hazzouri KM, Wright SI. Genome-wide evidence for efficient positive and purifying selection in Capsella grandiflora, a plant species with a large effective population size. Mol Biol Evol.

2010;27: 1813–1821.

30. Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, et al. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of Capsella grandiflora. PLoS Genet. 2014;10: e1004622.

31. Mack KL, Phifer-Rixey M, Harr B, Nachman MW. Gene Expression Networks Across Multiple Tissues Are Associated with Rates of Molecular Evolution in Wild House Mice. Genes . 2019;10. doi:10.3390/genes10030225

32. Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. Plant Cell. 2017;29: 944–959.

33. Saha A, Kim Y, Gewirtz ADH, Jo B, Gao C, McDowell IC, et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. bioRxiv. 2016. p. 078741. doi:10.1101/078741

34. Josephs EB, Wright SI, Stinchcombe JR, Schoen DJ. The relationship between selection, network connectivity, and regulatory variation within a population of Capsella grandiora. Genome Biol Evol. 2017. doi:10.1093/gbe/evx068

35. Palakurty SX, Stinchcombe JR, Afkhami ME. Cooperation and coexpression: How coexpression networks shift in response to multiple mutualists. Mol Ecol. 2018;27: 1860–1873.

36. Mack KL, Ballinger MA, Phifer-Rixey M, Nachman MW. Gene regulation underlies environmental adaptation in house mice. Genome Res. 2018;28: 1636–1645.

37. Campbell-Staton SC, Cheviron ZA, Rochette N, Catchen J, Losos JB, Edwards SV. Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. Science. 2017;357: 495–498.

38. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet. 2016;48: 709–717.

39. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95: 14863–14868.

40. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science. 2003;302: 249–255.

41. Schaefer R, Michno J-M, Jeffers J, Hoekenga OA, Dilkes BP, Baxter IR, et al. Integrating co-expression networks with GWAS to prioritize causal genes in maize. Plant Cell. 2018. doi:10.1105/tpc.18.00299

42. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169: 1177–1186.

43. Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. Cell. 2018;173: 1573–1580.

44. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell. 2019;177: 1022–1034.e6.

45. Fraser HB. Gene expression drives local adaptation in humans. Genome Res. 2013;23: 1089–1096.

46. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature. 2004;428: 717–723.

47. Whitehead A, Crawford DL. Neutral and adaptive variation in gene expression. Proc Natl Acad Sci U S A. 2006;103: 5425–5430.

48. Slotte T, Hazzouri KM, Arvid Ågren J, Koenig D, Maumus F, Guo Y-L, et al. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nat Genet. 2013;45: 831–835.

49. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011;21: 936–939.

50. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in Bioinformatics. John Wiley & Sons, Inc.; 2002.

51. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31: 166–169.

52. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28: 882–883.

53. Sicard A, Thamm A, Marona C, Lee YW, Wahl V, Stinchcombe JR, et al. Repeated evolutionary changes of leaf morphology caused by mutations to a homeobox gene. Curr Biol. 2014;24: 1880–1886.

54. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available: https://www.R-project.org/

55. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9: 559.

56. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis. 2015;53: 474–485.

57. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28: 1353–1358.

58. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44: 821–824.

59. Xing G, Lin C-Y, Wooding SP, Xing C. Blindly using Wald's test can miss rare disease-causal variants in case-control association studies. Ann Hum Genet. 2012;76: 168–177.

60. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57: 289–300.

61. Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, et al. An atlas of over 90,000

conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genet. 2013;45: 891–898.

62. Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Mol Biol Evol. 2013;30: 2224–2234.