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] and John R. Stinchcombe[5]

[1]Department of Plant Biology, Michigan State University, East Lansing, MI

[2]Indigo Ag, Charlestown, MA

[3]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA

[4]Department of Biology, McGill University, Montreal, QC, Canada

[5]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

***Corresponding author:** E-mail: josep993@msu.edu.

Associate editor: Michael Purugganan

Abstract

Understanding the persistence of genetic variation within populations has long been a goal of evolutionary biology. One promising route toward 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 stronger negative selection acting 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*-regulation.

Key words: **eQTLs, *trans*-regulatory variation, coexpression networks, population genetics.**

Letter

Understanding why genetic variation persists in populations has long been a goal of evolutionary biology (Mitchell-Olds et al. 2007). 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 (Johnson and Barton 2005). The availability of large genomic and phenotypic data sets 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 (Josephs, Stinchcombe, et al. 2017; Sella and Barton 2019). In particular, the allele frequencies of these loci can provide information about selection, since negative selection against deleterious mutations is expected to keep alleles at lower frequencies than neutrality or balancing selection.

Gene expression has emerged as a powerful model trait for addressing the challenge of the maintenance of variation (Kliebenstein 2009). 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 (Rockman and Kruglyak 2006). 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 predefined 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 (Wittkopp et al. 2004; Emerson and Li 2010). 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 (Battle et al. 2014; Josephs et al. 2015; Glassberg et al. 2019; Hernandez et al. 2019). However, *trans*-eQTLs may be under different selection pressures than *cis*-eQTLs. As *trans*-regulatory variation can affect 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 (McGuigan et al. 2014). This prediction is supported by evidence of greater *trans*-regulatory variation within species compared to between species (Wittkopp et al. 2004, 2008), reduced population frequencies of distant eQTLs compared with local eQTLs (Zhang et al. 2011), and greater effect sizes of standing *cis*-regulatory variants than *trans*-regulatory variants (Kliebenstein 2009; Liu et al. 2017; Mähler et al. 2017), although these effect

size differences may also be caused by differences in mutational input (Metzger et al. 2016).

Despite the expectation that negative 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 (Keurentjes et al. 2007; West et al. 2007; Rockman et al. 2010; Lowry et al. 2013; Battle et al. 2014; Liu et al. 2017; Albert et al. 2018, but see Mähler et al. 2017). Segregating *trans*-variation is more likely to be tissue-specific than *cis*-regulatory variation in humans (GTEx Consortium et al. 2017) and, in *Arabidopsis thaliana*, *trans*-eQTLs are particularly important for expression changes in response to drought (Lowry et al. 2013; Clauw et al. 2016). 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 (Kliebenstein et al. 2006; Hore et al. 2016; Brynedal et al. 2017), 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 (Slotte et al. 2010; Williamson et al. 2014). 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 data sets (Saha et al. 2017; Josephs, Stinchcombe, et al. 2017; Mähler et al. 2017; Wisecaver et al. 2017; Palakurty et al. 2018; Mack et al. 2019). For example, coexpression networks made across conditions, tissues, and developmental time can successfully identify specialized metabolic pathways (Wisecaver et al. 2017) and coexpression modules made with a diverse panel of mouse lines correlate with phenotype (Mack et al. 2018). In addition, changes in coexpression module expression have been linked to adaptation (Campbell-Staton et al. 2017) and changing ecological conditions (Palakurty et al. 2018). 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 eQTLs for coexpression module expression. Overall, our results suggest that negative selection acts on *trans*-eQTLs more strongly than *cis*-eQTLs, 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 145 individuals. We identified 6,231 associations (FDR <0.1) between 5,468 unique SNPs and 2,341 genes (fig. 1). Each eQTL was associated with the expression of between 1 and 93 genes. We separated associations into 2,472 *trans*-eQTLs that were >5 kb away from the gene they regulated and 3,759 local (putatively *cis*) eQTLs that were <5 kb 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) (Josephs et al. 2015). *Trans*-eQTLs had larger effect sizes than *cis*-eQTLs ($P < 0.001$, mean local effect size = 0.91, mean *trans*-effect size = 1.01).

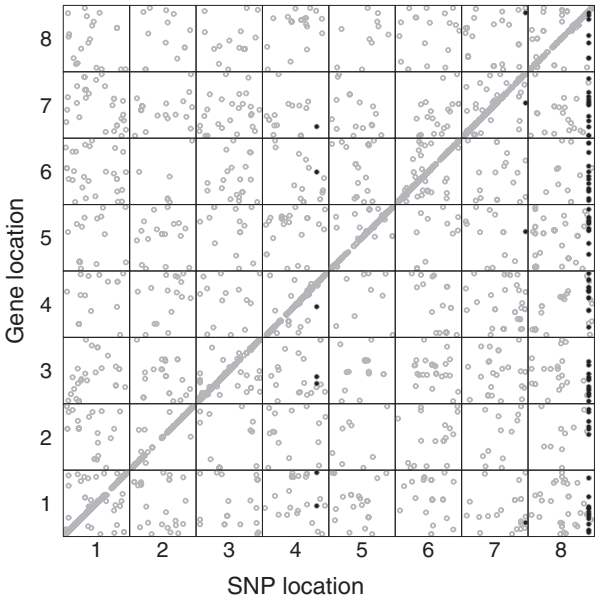

FIG. 1. All-by-all eQTL locations. Each point represents a SNP, whose location is described on the *x* axis, that is associated with the expression of a gene whose location is described on the *y* axis (FDR<0.1). Black points are SNPs that were also identified in the coexpression eQTL analysis.

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 *cis*-eQTLs (fig. 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, 2.5, and 10 kb ($P < 0.001$). This difference in allele frequencies is consistent with stronger negative selection on *trans*-eQTLs than *cis*-eQTLs. However, as *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 effect sizes (684 *cis*-eQTLs and 692 *trans*-eQTLs), *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 with the MAFs of

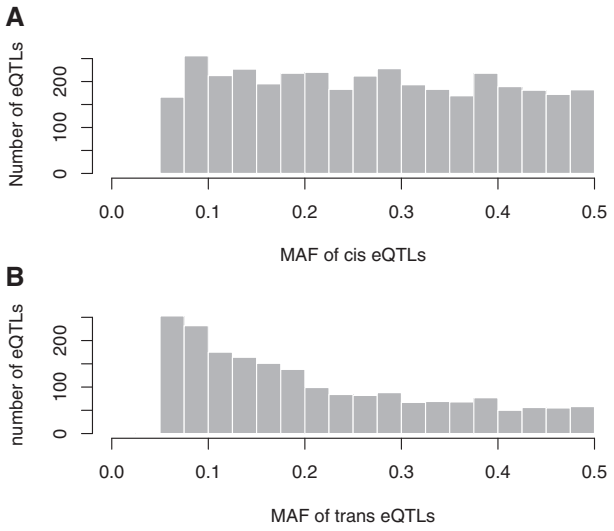

FIG. 2. Distribution of minor allele frequencies (MAF) of *cis*- and *trans*-eQTLs.

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 the expression of 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 association.

Coexpression Module GWAS Identifies SNPs Affecting Expression Modules

We identified 16 coexpression modules ranging in size from 69 to 6,392 genes (supplementary fig. S1, Supplementary Material online). 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 ($P > 0.2$, supplementary fig. S2, Supplementary Material online). 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 with other individuals (supplementary fig. S3, Supplementary Material online). Because the skewed distribution of module expression values could lead to false-positives during association mapping, we quantile-normalized module expression 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 (fig. 3 and 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 (fig. 3*B*), suggesting that there is a longer block of loci in linkage disequilibrium associated with module expression.

Table 1. Information about Significant Coexpression eQTLs (FDR<0.1).

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

NOTE.—CNS, conserved noncoding sequence; MAF, minor allele frequency.

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 *A. 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 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 (fig. 4, $P < 0.05$ after Bonferroni correction for 16 tests). None of the four coexpression eQTLs detected was significantly associated with any phenotypes.

Population Genetic Signatures of Selection on eQTLs

We compared signatures of selection around eQTLs identified by the all-by-all analysis and found that *trans*-eQTLs were in windows that had lower Tajima's *D* on an 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 with *cis*-eQTLs. Although we have evidence that *cis*-eQTLs are in general under negative selection in this population (Josephs et al. 2015), and our present analysis suggests that in general *trans*-eQTL are subject to stronger negative selection than *cis*-eQTLs, we were curious if we could detect evidence of recent positive or balancing selection on coexpression eQTLs. We measured π and Tajima's *D* at putatively neutral sites

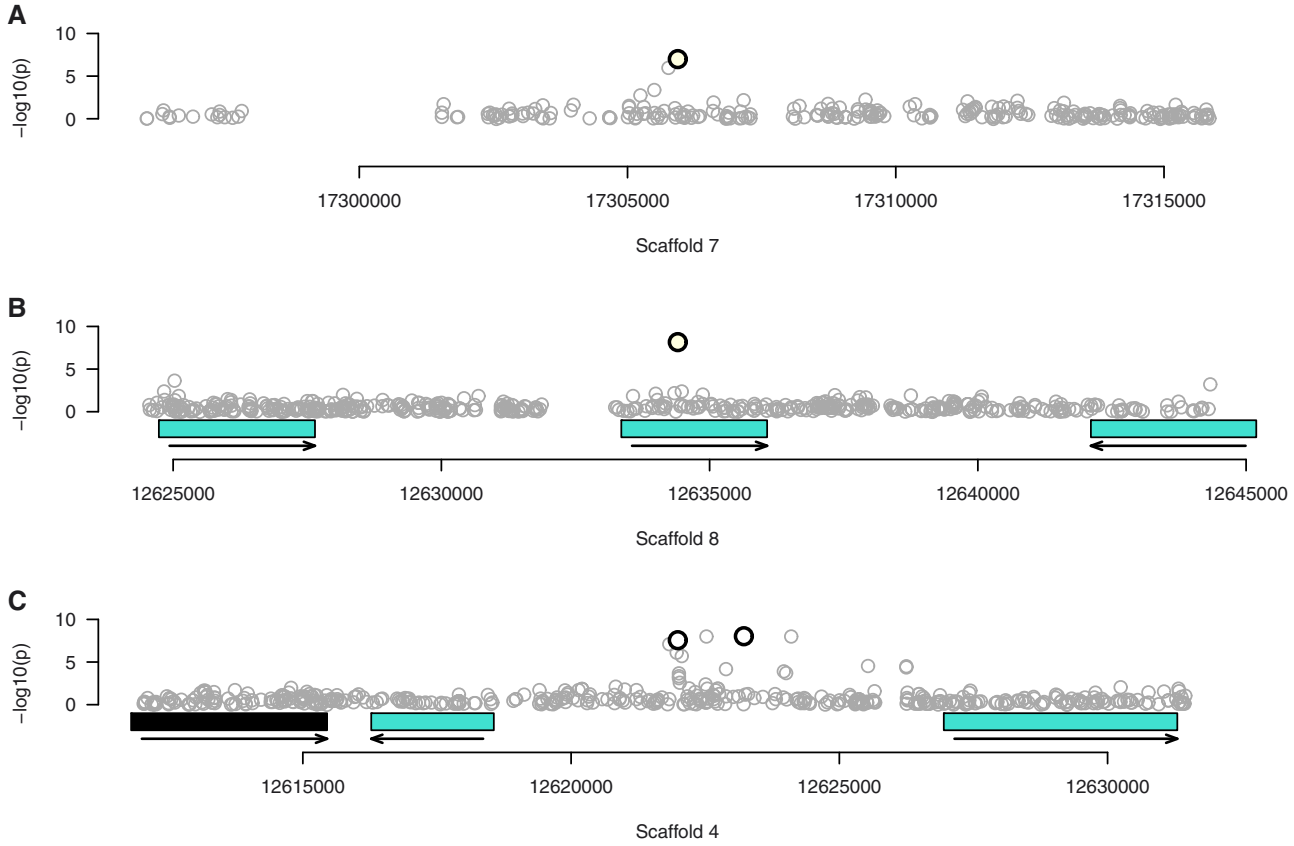

Fig. 3. Physical locations of coexpression QTLs. Coexpression eQTLs are represented by white points with black borders. All other SNPs are plotted in gray. These plots include associations for all SNPs, not just the tagging SNPs, plotted by location, on the *x* axis, and the significance of association with the module 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.

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 was located in windows that were outliers (top 2.5% of windows) for π, Tajima's *D*, or sweep likelihood (supplementary figs. S4–S6, Supplementary Material online).

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 *A. thaliana* (Zhang et al. 2011). 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* (Josephs, Wright, et al. 2017) and in other species (McGuigan et al. 2014; Pickrell et al. 2016).

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. Although coexpression measured across multiple timepoints (developmental networks) has been linked to

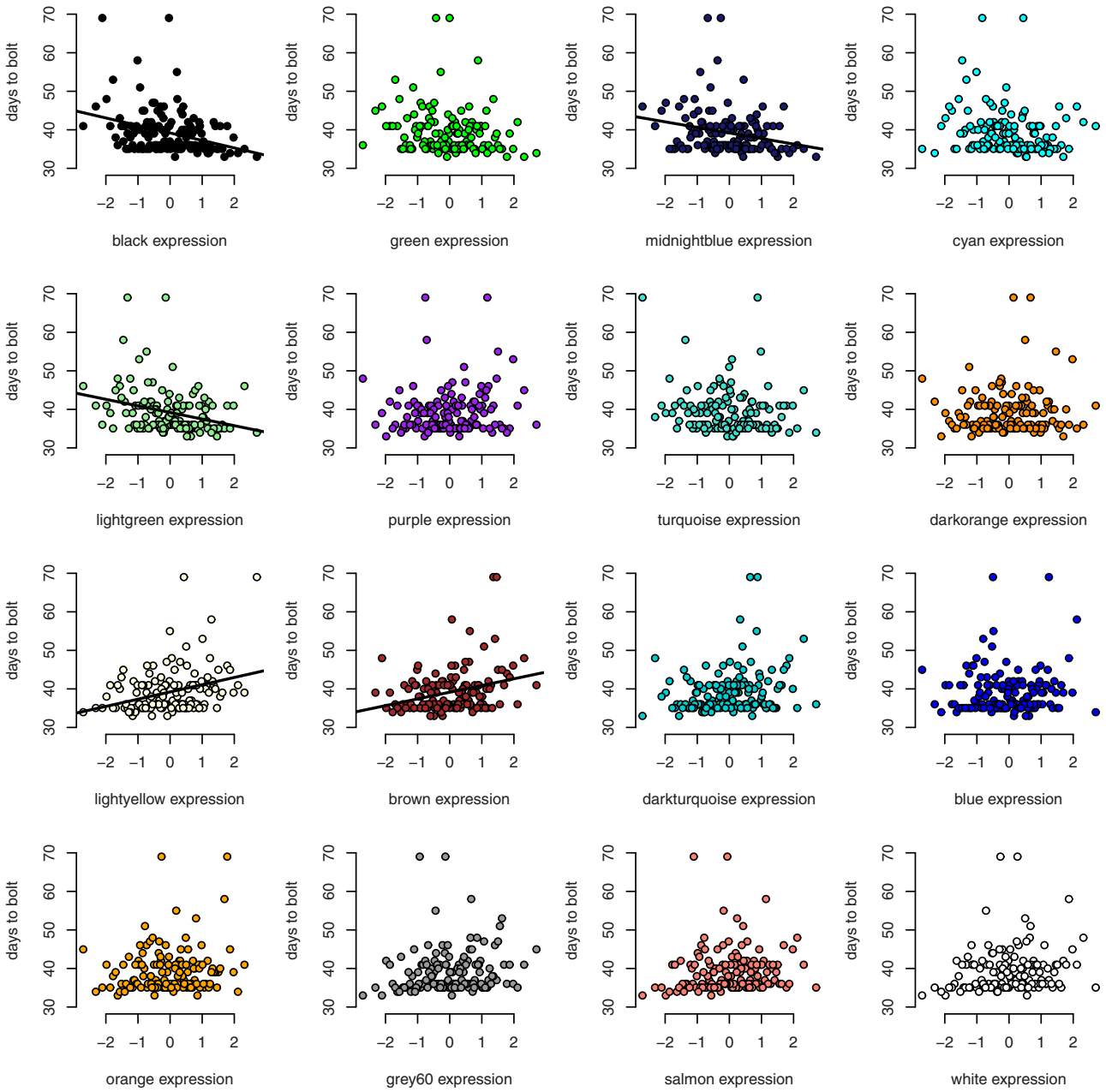

FIG. 4. Correlations between module expression level and traits. The correlation between expression of each module (*x* axis) and days to bolt (*y* axis). If the correlation was significant ($P < 0.05$ after Bonferroni correction), then there is a black trend line drawn.

functional relationships (Eisen et al. 1998; Stuart et al. 2003), coexpression modules generated from genetically distinct individuals have different properties than those generated from different tissue types (Mähler et al. 2017; Schaefer et al. 2018). 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 (Schaefer et al. 2018). However, it is important to keep in mind that the expression data sets used will affect coexpression modules.

Mapping eQTLs have 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 (Boyle et al. 2017; Wray et al. 2018; Liu et al. 2019). In addition, as gene expression is important for adaptive divergence (Shapiro et al. 2004; Whitehead and Crawford 2006; Fraser 2013), 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 were previously published in Josephs et al. (2015) and Josephs, Wright,

et al. (2017). We briefly describe data generation here. We collected individuals from a single population of *C. grandiflora* located near Monodendri, Greece. We conducted a generation of random crosses in the greenhouse, and then grew 145 individuals descended from these random crosses in a growth chamber with 16 h of daylight at 22 °C. We measured traits on these individuals and collected RNA from young leaf tissue 39 days after planting. Leaves were collected and flash frozen at night sequentially in nine roughly equally sized bins, which we will refer to as "collection bins." We extracted RNA with RNA extraction kits (Sigma). We extracted DNA from leaf material using a CTAB procedure. Both RNA and DNA were sequenced at the Genome Quebec facility with Hiseq 2000 with Truseq libraries with 100-bp long reads. DNA was mapped to the standard *C. rubella* reference genome (Slotte et al. 2013) with Stampy (Lunter and Goodson 2011) and RNA was mapped to an exon-only reference genome using Stampy. SNPs were called from the genomic sequence data using GATK Unified Genotyper (Van der Auwera et al. 2013) and expression levels were measured with HTseq and normalized for sequencing depth by dividing by the median expression level for each individual. (Anders et al. 2015). We did not detect interactions between GC content, expression level, and lane (Josephs et al. 2015). We used the ComBat function from the sva R package to adjust expression level for collection batch (Leek et al. 2012).

In addition to collecting RNA and DNA for sequencing on these 145 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 Sicard et al. (2014). Briefly, dissection index was calculated as DI = (perimeter^2)/(4$\pi\times$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. The 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 (R Core Team 2018).

Building Coexpression Networks

We used the program WGCNA (Langfelder and Horvath 2008) (version 1.68 running on R version 3.6.2) to identify coexpression modules present within the transcriptomes using the expression level of all genes with median expression >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 coexpression 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 >0.8 were merged to arrive at the final set of coexpression modules. The resulting modules were labeled with different colors for ease of referencing. We investigated specific eQTLs in the Joint Genome Institute's Phytozome v12.1 genome browser for *C. rubella* v1.0 (Slotte et al. 2013) and investigated specific orthologs in *A. thaliana* using TAIR (Berardini et al. 2015).

Some of the modules had expression levels that were very skewed, such that a few individuals showed extremely high module expression compared with the rest of the individuals (supplementary fig. S1, Supplementary Material online). 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 (R Core Team 2018).

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 <0.01 and >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 data set .

We tested for associations between the tag SNPs and the expression of 18,806 genes using the linear model in Matrix eQTL (Shabalin 2012). We quantile-normalized gene expression levels using the qqnorm function in R to reduce false-positives caused by skewed expression distributions. Although all samples came from the same population, we controlled for residual population structure by generating a centered kinship matrix with GEMMA (Zhou and Stephens 2012) and including the first five principal components of the kinship matrix as covariates. As 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 <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 (R Core Team 2018).

We did association mapping with GEMMA (Zhou and Stephens 2012) 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 (Xing et al. 2012) 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 (Benjamini and Hochberg 1995).

Population Genetic Signatures of Selection

We used genomic sequence from 188 individuals published in Josephs et al. (2015) which include the 145 individuals used in the eQTL analysis. 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 noncoding (excluding conserved noncoding sites from Haudry et al. [2013]), intronic, and 4-fold degenerate sites. We used SweeD to calculate the likelihood of a selective sweep occurring on every 50th SNP (windows were ~600 bp wide on an average) using nonconserved intergenic, intronic, and 4-fold degenerate sites for 182 individuals from the focal population (Pavlidis et al. 2013).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online. All code for analysis is available at https://github.com/emjosephs/capsella-eQTL and archived on Zenodo: 10.5281/zenodo.3774082.

Acknowledgments

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 E.B.J., Natural Sciences and Engineering Research Council of Canada Discovery Grants to D.J.S., S.I.W., and J.R.S., and a Value-directed Evolutionary Genomics Initiative grant (Genome Quebec/Genome Canada) to D.J.S., S.I.W., and J.R.S.

References

Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. 2018. Genetics of *trans*-regulatory variation in gene expression. *Elife* 7. Available from: http://dx.doi.org/10.7554/eLife.35471.

Anders S, Pyl PT, Huber W. 2015. HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.

Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, et al. 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* 24(1):14–24.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Stat Methodol.* 57(1):289–300.

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E. 2015. The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis* 53(8):474–485.

Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169(7):1177–1186.

Brynedal B, Choi J, Raj T, Bjornson R, Stranger BE, Neale BM, Voight BF, Cotsapas C. 2017. Large-scale *trans*-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. *Am J Hum Genet.* 100(4):581–591.

Campbell-Staton SC, Cheviron ZA, Rochette N, Catchen J, Losos JB, Edwards SV. 2017. Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. *Science* 357(6350):495–498.

Clauw P, Coppens F, Korte A, Herman D, Slabbinck B, Dhondt S, Van Daele T, De Milde L, Vermeersch M, Maleux K, et al. 2016. Leaf growth response to mild drought: natural variation in Arabidopsis sheds light on trait architecture. *Plant Cell* 28(10):2417–2434.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A.* 95(25):14863–14868.

Emerson JJ, Li W-H. 2010. The genetic basis of evolutionary change in gene expression levels. *Philos Trans R Soc B.* 365(1552):2581–2590.

Fraser HB. 2013. Gene expression drives local adaptation in humans. *Genome Res.* 23(7):1089–1096.

Glassberg EC, Gao Z, Harpak A, Lan X, Pritchard JK. 2019. Evidence for weak selective constraint on human gene expression. *Genetics* 211(2):757–772.

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, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. 2017. Genetic effects on gene expression across human tissues. *Nature* 550:204–213.

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat Genet.* 45(8):891–898.

Hernandez RD, Uricchio LH, Hartman K, Ye C, Dahl A, Zaitlen N. 2019. Ultrarare variants drive substantial *cis* heritability of human gene expression. *Nat Genet.* 51(9):1349–1355.

Hore V, Viñuela A, Buil A, Knight J, McCarthy MI, Small K, Marchini J. 2016. Tensor decomposition for multiple-tissue gene expression experiments. *Nat Genet.* 48:1094–1100.

Johnson T, Barton N. 2005. Theoretical models of selection and mutation on quantitative traits. *Philos Trans R Soc B.* 360(1459):1411–1425.

Josephs EB, Lee YW, Stinchcombe JR, Wright SI. 2015. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proc Natl Acad Sci U S A.* 112(50):15390–15395.

Josephs EB, Stinchcombe JR, Wright SI. 2017. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytol.* 214(1):21–33.

Josephs EB, Wright SI, Stinchcombe JR, Schoen DJ. 2017. The relationship between selection, network connectivity, and regulatory variation within a population of *Capsella grandiora*. *Genome Biol Evol.* 9(4):1099–1109.

Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC. 2007. Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci U S A.* 104(5):1708–1713.

Kliebenstein D. 2009. Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Annu Rev Plant Biol.* 60(1):93–114.

Kliebenstein DJ, West MAL, van Leeuwen H, Loudet O, Doerge RW, St Clair DA. 2006. Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics* 7(1):308.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(1):559.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28(6):882–883.
Liu H, Luo X, Niu L, Xiao Y, Chen L, Liu J, Wang X, Jin M, Li W, Zhang Q, et al. 2017. Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. *Mol Plant* 10(3):414–426.
Liu X, Li YI, Pritchard JK. 2019. *Trans* effects on gene expression can drive omnigenic inheritance. *Cell* 177(4):1022–1034.e6.
Lowry DB, Logan TL, Santuari L, Hardtke CS, Richards JH, DeRose-Wilson LJ, McKay JK, Sen S, Juenger TE. 2013. Expression quantitative trait locus mapping across water availability environments reveals contrasting associations with genomic features in Arabidopsis. *Plant Cell* 25(9):3266–3279.
Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 21(6):936–939.
Mack KL, Ballinger MA, Phifer-Rixey M, Nachman MW. 2018. Gene regulation underlies environmental adaptation in house mice. *Genome Res.* 28(11):1636–1645.
Mack KL, Phifer-Rixey M, Harr B, Nachman MW. 2019. Gene expression networks across multiple tissues are associated with rates of molecular evolution in wild house mice. *Genes* 10(3):225.
Mähler N, Wang J, Terebieniec BK, Ingvarsson PK, Street NR, Hvidsten TR. 2017. Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet.* 13(4):e1006402.
McGuigan K, Collet JM, Allen SL, Chenoweth SF, Blows MW. 2014. Pleiotropic mutations are subject to strong stabilizing selection. *Genetics* 197(3):1051–1062.
Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ. 2016. Contrasting frequencies and effects of *cis*- and *trans*-regulatory mutations affecting gene expression. *Mol Biol Evol.* 33(5):1131–1146.
Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat Rev Genet.* 8(11):845–856.
Palakurty SX, Stinchcombe JR, Afkhami ME. 2018. Cooperation and coexpression: how coexpression networks shift in response to multiple mutualists. *Mol Ecol.* 27(8):1860–1873.
Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N. 2013. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol.* 30(9):2224–2234.
Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. 2016. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet.* 48(7):709–717.
R Core Team. 2018. R: a language and environment for statistical computing. Available from: https://www.R-project.org/.
Rockman MV, Kruglyak L. 2006. Genetics of global gene expression. *Nat Rev Genet.* 7(11):862–872.
Rockman MV, Skrovanek SS, Kruglyak L. 2010. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* 330(6002):372–376.
Saha A, Kim Y, Gewirtz ADH, Jo B, Gao C, McDowell IC, GTEx Consortium, Engelhardt BE, Battle A. 2017. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* 27(11):1843–1858.
Schaefer R, Michno J-M, Jeffers J, Hoekenga OA, Dilkes BP, Baxter IR, Myers C. 2018. Integrating co-expression networks with GWAS to prioritize causal genes in maize. *Plant Cell* 30(12):2922–2942.
Sella G, Barton NH. 2019. Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annu Rev Genomics Hum Genet.* 20:461–493..
Shabalin AA. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28(10):1353–1358.
Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428(6984):717–723.
Sicard A, Thamm A, Marona C, Lee YW, Wahl V, Stinchcombe JR, Wright SI, Kappel C, Lenhard M. 2014. Repeated evolutionary changes of leaf morphology caused by mutations to a homeobox gene. *Curr Biol.* 24(16):1880–1886.
Slotte T, Foxe JP, Hazzouri KM, Wright SI. 2010. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Mol Biol Evol.* 27(8):1813–1821.
Slotte T, Hazzouri KM, Arvid Ågren J, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Killian Newman L, et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet.* 45(7):831–835.
Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302(5643):249–255.
Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics 43(1110):11.10.1–11.10.33.
West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA. 2007. Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics* 175(3):1441–1450.
Whitehead A, Crawford DL. 2006. Neutral and adaptive variation in gene expression. *Proc Natl Acad Sci U S A.* 103(14):5425–5430.
Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI. 2014. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora*. *PLoS Genet.* 10(9):e1004622.
Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. 2017. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* 29(5):944–959.
Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430(6995):85–88.
Wittkopp PJ, Haerum BK, Clark AG. 2008. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet.* 40(3):346–350.
Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. 2018. Common disease is more complex than implied by the core gene omnigenic model. *Cell* 173(7):1573–1580.
Xing G, Lin C-Y, Wooding SP, Xing C. 2012. Blindly using Wald's test can miss rare disease-causal variants in case-control association studies. *Ann Hum Genet.* 76(2):168–177.
Zhang X, Cal AJ, Borevitz JO. 2011. Genetic architecture of regulatory variation in *Arabidopsis thaliana*. *Genome Res.* 21(5):725–733.
Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 44(7):821–824.