

# Global eQTL Mapping Reveals the Complex Genetic Architecture of Transcript-Level Variation in Arabidopsis

Marilyn A. L. West,\* Kyunga Kim,<sup>†</sup> Daniel J. Kliebenstein,\* Hans van Leeuwen,\*  
Richard W. Michelmore,\*<sup>‡</sup> R. W. Doerge<sup>†</sup> and Dina A. St. Clair\*<sup>1</sup>

\*Department of Plant Sciences, University of California, Davis, California 95616-8780, <sup>†</sup>Department of Statistics, Purdue University, West Lafayette, Indiana 47907-2067 and <sup>‡</sup>The Genome Center, University of California, Davis, California 95616

Manuscript received August 17, 2006  
Accepted for publication December 1, 2006

## ABSTRACT

The genetic architecture of transcript-level variation is largely unknown. The genetic determinants of transcript-level variation were characterized in a recombinant inbred line (RIL) population ( $n = 211$ ) of *Arabidopsis thaliana* using whole-genome microarray analysis and expression quantitative trait loci (eQTL) mapping of transcript levels as expression traits (e-traits). Genetic control of transcription is highly complex: one-third of the quantitatively controlled transcripts/e-traits were regulated by *cis*-eQTL, and many *trans*-eQTL mapped to hotspots that regulated hundreds to thousands of e-traits. Several thousand eQTL of large phenotypic effect were detected, but almost all (93%) of the 36,871 eQTL were associated with small phenotypic effects ( $R^2 < 0.3$ ). Many transcripts/e-traits were controlled by multiple eQTL with opposite allelic effects and exhibited higher heritability in the RILs than their parents, suggesting nonadditive genetic variation. To our knowledge, this is the first large-scale global eQTL study in a relatively large plant mapping population. It reveals that the genetic control of transcript level is highly variable and multifaceted and that this complexity may be a general characteristic of eukaryotes.

TRANSCRIPT levels, when assessed in an experimental or mapping population, can be considered as quantitative traits and their variation used to map expression quantitative trait loci (eQTL) (JANSEN and NAP 2001; DOERGE 2002; SCHADT *et al.* 2003). Recent studies in a limited number of organisms indicate that the levels of transcripts are heritable and can be under multigenic control (BREM *et al.* 2002; SCHADT *et al.* 2003; YVERT *et al.* 2003; KIRST *et al.* 2004; BREM and KRUGLYAK 2005; BYSTRYKH *et al.* 2005; CHESLER *et al.* 2005; HUBNER *et al.* 2005; DECOOK *et al.* 2006). Global eQTL analyses in yeast, mice, and humans have detected significant levels of *cis* polymorphism controlling individual genes, as well as evidence for clustered *trans*-eQTL that simultaneously regulate a large fraction of the transcriptome (BREM *et al.* 2002; SCHADT *et al.* 2003; MORLEY *et al.* 2004). In yeast, the complex inheritance of transcript levels (BREM and KRUGLYAK 2005) has been revealed by detecting significant levels of non-additive genetic variance, epistatic interactions, and transgressive segregation. However, comparable studies in plants have yet to be reported. Furthermore, the relationship between transcript-level variation and downstream phenotypic trait variation is not well understood (MACKAY 2001; GIBSON and WEIR 2005).

While the identification of *cis*-eQTL has led to successful QTL cloning in plants (KLIEBENSTEIN *et al.* 2001; LAMBRIX *et al.* 2001; KROYMANN *et al.* 2003; CAICEDO *et al.* 2004; WERNER *et al.* 2005; ZHANG *et al.* 2006), genomewide eQTL analyses in large mapping populations of plants have been lacking. To our knowledge, the study that we describe here is among the first large global eQTL mapping studies in plants. Our overarching goal was to assess and report on the genetic architecture of transcript-level variation in a higher plant species. To address this goal, we employed a sample of 211 recombinant inbred lines (RILs) from a population derived from a cross between two inbred *Arabidopsis thaliana* accessions, Bayreuth-0 (Bay-0) and Shahdara (Sha), that were grown in a biologically replicated experiment to quantify genomewide transcript levels using Affymetrix whole-genome microarrays. The 211 RILs represent a sample of individuals from the population of all possible individuals who represent this particular cross. Transcript-level variation in the 211 RILs was measured and used as quantitative expression traits (e-traits) to map eQTL using a framework map of single feature polymorphism (SFP) markers (WEST *et al.* 2006) for the 211 RILs. We report on the genomic distributions for eQTL, as well as on the phenotypic effects and numbers of *cis*- and *trans*-eQTL. We also estimate the broad-sense heritability of transcript levels/e-traits in the RILs and parent lines and assess transgressive segregation and nonadditive genetic variation.

<sup>1</sup>Corresponding author: Department of Plant Sciences, University of California, Mail Stop 3, 1 Shields Ave., Davis, CA 95616-8780.  
E-mail: dastclair@ucdavis.edu

## MATERIALS AND METHODS

**Plant material and experimental conditions:** Seeds for *A. thaliana* accessions Bay-0 and Sha and the F<sub>8</sub> generation of a Bay-0 × Sha RIL population (LOUDET *et al.* 2002) were obtained from The Arabidopsis Information Resource (TAIR; stock no. CS57920; <http://www.arabidopsis.org>). The advanced generation RILs were created by single-seed descent from the F<sub>2</sub> generation (LOUDET *et al.* 2002). The RIL (F<sub>8</sub>) plants and parental accessions were grown in a single growth chamber at the University of California at Davis and allowed to self-pollinate; seed was harvested from individual plants to produce sufficient seed for each homozygous F<sub>9</sub> line for our replicated experiments.

For the Bay-0 × Sha RIL microarray experiment, five plants/biological replicate for each of 211 RILs, plus parental controls, Bay-0 and Sha, were grown as previously described (WEST *et al.* 2006). At 6 weeks postgermination, the plants were treated with 0.02% Silwet L77, a surfactant (Lehle Seeds, Round Rock, TX), and harvested 28 hr post-treatment, as previously described (WEST *et al.* 2006).

**Microarray hybridization and quality control:** Total RNA was extracted, converted to cDNA, and labeled with biotin as recommended by the manufacturer (Affymetrix, Santa Clara, CA; <http://www.affymetrix.com>). Affymetrix ATH1 GeneChip microarrays were hybridized, washed, scanned, and checked for quality as reported previously (WEST *et al.* 2006). The RIL haplotypes obtained using SFP markers derived from the same GeneChips were consistent with the haplotypes determined previously by microsatellite analysis of genomic DNA, indicating that each GeneChip data set was derived from the designated RIL (LOUDET *et al.* 2002; WEST *et al.* 2006).

**QTL analysis:** Raw image data from the RIL GeneChips were converted to numeric data via Bioconductor software (<http://www.bioconductor.org>). Due to the nonbiological variation that is present in every microarray experiment, we normalized across all arrays for the purpose of addressing any nonlinear relationships between arrays. Quantile normalization reduces nonbiological variation that is largely due to the technology itself and, when applied at the probe level, it has been shown to outperform other normalization methods that are based on what is referred to as a “base-line array” (BOLSTAD *et al.* 2003). After the quantile normalization, raw intensity values were log<sub>2</sub> transformed. For this analysis of gene e-traits, transcript levels of each gene were averaged over the two microarray replicates per RIL to give a single transcript measurement per RIL.

A genetic map was estimated using JoinMap (Kyazma B.V., Wageningen, The Netherlands); 540 SFP markers were obtained as described previously (WEST *et al.* 2006), and 38 reference SSR markers (LOUDET *et al.* 2002) were scored on the 211 RILs. In previous work (WEST *et al.* 2006), we used a subset of 148 RILs to create a high-density SFP map. When the microarray data for the remaining 63 RILs became available, we reestimated the genetic map on the basis of a total sample of 211 RILs. The resulting map had an average marker density of 0.78 cM. Composite interval mapping (CIM) analysis (ZENG *et al.* 1999) was employed in conjunction with a 5-cM framework map (93 SFP and 2 SSR markers; supplemental Table 1 at <http://www.genetics.org/supplemental/>). The “*zmapqtl*” CIM module of QTL Cartographer Version 1.17 (BASTEN *et al.* 1999) with a walking speed of 1 cM and a window size of 10 cM was employed to analyze each e-trait. To obtain a threshold criterion for declaring statistically significant eQTL, a global permutation threshold (GPT) was obtained by permuting the e-traits while maintaining the genetic information. Because it was computationally not feasible (*i.e.*, at least 1000 permutations for each of the 22,794 e-traits) to compute an empirical

threshold value (CHURCHILL and DOERGE 1994) for each e-trait that we modeled with composite interval mapping, we adopted a GPT approach as follows. For each of 100 randomly selected transcripts or e-traits, the null distribution of the maximum likelihood-ratio test (LRT) statistic was empirically estimated using permutation thresholds based on 1000 permutations (CHURCHILL and DOERGE 1994). Inspection of the means and standard deviations for the 100 null distributions from the permuted transcript values revealed little variation, and the average 95-percentile permutation threshold was 12.0329, with a sample standard deviation of 0.3478. Permutation thresholds were then computed on the basis of each of the 100 representative null distributions. A representative null distribution based on 100,000 maximum LRT statistics (1000 permutations × 100 randomly selected e-traits) was employed for all 22,794 noncontrol transcripts represented on the ATH1 microarray. The GPT was calculated as the 95% upper bound of the representative null distribution and equaled 12.0583.

**Identification of eQTL location:** After conducting composite interval mapping for each of the 22,794 e-traits, it was necessary to summarize the resulting eQTL with a certain level of resolution to the known location of the gene features on the map. For this we utilized the Eql module of QTL Cartographer to count, summarize, and locate potential eQTL (BASTEN *et al.* 1999). Since experimental limitations (*e.g.*, genetic map density and the sample size of the mapping population), statistical issues (*e.g.*, limitations of the statistical method), and genetic factors (*e.g.*, ghost QTL, false-positive QTL, linked QTL, etc.) come into play when locating QTL, we attempted to summarize our eQTL results in a general manner by using an “exclusionary window” that is available in the Eql module to select and count eQTL. However, we first investigated the effect of varying the exclusionary window size (5, 10, and 20 cM) on the eQTL summary statistics. Changing the exclusionary window size had no effect on the number of transcripts/e-traits with detectable eQTL, but increasing it decreased the number of eQTL that were counted for each transcript (supplemental Figure 1 at <http://www.genetics.org/supplemental/>). The average number of eQTL counted per transcript decreased from 3.7 to 2.3 when results were compared between windows of 5 and 20 cM (supplemental Table 2 at <http://www.genetics.org/supplemental/>), but this did not affect the genetic positions of the large *trans*-effect peaks (data not shown). These results suggest that the smaller exclusionary window sizes fail to accommodate closely linked eQTL, ghost QTL (HALEY and KNOTT 1992), and other artifacts that are the result of either experimental limitations or the limitations of composite interval mapping to locate eQTL. After deciding on an exclusionary window of 20 cM to select and count eQTL, the following protocol was used in turn for each of the 22,794 transcripts whose levels were measured and analyzed as e-traits. First, all potential eQTL were sorted by their LRT statistics. The eQTL with the largest LRT statistic was selected, and all other potential eQTL within the exclusionary window (*i.e.*, within 10 cM on either side of the chosen eQTL) were disregarded. The next eQTL was considered and the process iterated until all eQTL for the particular e-trait were exhausted.

Our observation that a 5-cM exclusionary window (Eql module of QTL Cartographer) is most likely affected by additional linked eQTL is supported by the realization that increasing the window size to 20 cM shifted the distribution to more transcripts/e-traits with fewer eQTL (one to three) (supplemental Figure 1 at <http://www.genetics.org/supplemental/>). To investigate this phenomenon further, we examined four genes that were map-based cloned and shown to be controlled by single *cis*-eQTL in Arabidopsis (KLIEBENSTEIN

*et al.* 2001; LAMBRIX *et al.* 2001). These four genes showed large LRT statistics in *cis* to the physical/genetic location of each gene, as expected (supplemental Figure 2 at <http://www.genetics.org/supplemental/>). When a 5-cM exclusionary window was used for identifying eQTL, three to five eQTL were incorrectly associated with each of the four cloned loci instead of a single *cis*-eQTL for each locus. Increasing the window size to 20 cM eliminated these extra eQTL peaks in each test case. Consequently, we used a conservative 20-cM exclusionary window for the global eQTL analysis reported here.

**Estimation of heritability:** We calculated estimates of broad-sense heritability ( $H$ ) for each e-trait as  $H = \sigma_g^2 / \sigma_p^2$ , where  $\sigma_g^2$  is the estimated e-trait (transcript level) genetic variance among different genotypes in this sample of 211 RILs and  $\sigma_p^2$  is the estimated phenotypic variance for an e-trait (FALCONER and MACKAY 1996). Heritability was estimated for all 22,746 noncontrol e-traits in two microarray data sets independently. The first data set included two GeneChips (biological replicates) for each of the 211 RILs, permitting estimation of the heritability for each e-trait within the RIL population. The estimates of  $\sigma_R^2$  (variance due to replicate) and  $\sigma_e^2$  (experimental error variance) were determined for each e-trait in this data set. The second data set included eight GeneChips per parent (Bay-0 and Sha), representing 16 different biological replicates, 4 replicates of each parent grown concurrently with the two RIL replications. The parental data allowed for the estimation of the broad-sense heritability for the levels of each transcript, which were then compared with the estimated heritabilities of e-traits in the 211 RILs.

***Cis- vs. trans-eQTL identification:*** The SFP markers (WEST *et al.* 2006), which are derived from genes with known physical positions, were used to anchor the genetic map to the genome. The CIM eQTL results for each transcript were examined to investigate if there was an eQTL within 3.5 cM of each gene's physical location. Expanding this exclusionary window to 5 cM had minimal effect on the numbers of *cis*-eQTL identified (data not shown; further discussion below). To test the impact of *cis*-eQTL on the relationship between the parental and RIL heritability estimates, we limited this analysis to only those 4933 transcripts represented by a unique probe set to ensure that the probe set was specific to only one genomic location.

**Transgressive segregation and eQTL of opposite allelic effect:** An ad-hoc approach was devised to test for transgressive segregation (BREM and KRUGLYAK 2005). For each e-trait, we counted the number of RILs whose average transcript-level accumulation was 1 pooled standard deviation above or below the minimum or maximum parental mean value. The Bay-0 and Sha accessions may be fixed for groups of alleles with opposite effects at two eQTL (*e.g.*, Bay-0 as + -/+ - and Sha as - +/+ +), which would either cancel out each other or generate progeny with a greater phenotypic distribution than the parents due to the segregation of recombinant individuals with + +/+ + and - -/- - genotypes for the two eQTL (LYNCH and WALSH 1998). For the 10,084 e-traits with two or more significant eQTL, the additive effect of each eQTL was queried to determine if the transcript was associated with eQTL having opposite allelic effects. An e-trait was defined as exhibiting an opposite allelic effect if it had at least one eQTL with a positive Bay-0 allelic effect and at least one eQTL with a positive Sha allelic effect. If all of the eQTL for an e-trait had only positive Bay-0 or only positive Sha allelic effects, the transcript was classified as not showing transgressive segregation.

**eQTL hotspot significance threshold:** To determine whether a genetic location associated with multiple eQTL was a significant cluster or hotspot, we estimated a significance threshold using permutation. The positions of the 36,871 eQTL at the marker intervals were permuted across the genome 1000 times,

and the maximal number of eQTL per genetic position per permutation was obtained. Using the distribution of the maximum number of eQTL, the criterion for declaration of a significant eQTL hotspot was 133 eQTL/genetic position at  $\alpha = 0.05$ .

**Gene ontology analysis of *trans*-eQTL hotspots:** The gene ontology (GO) analysis was performed using GOSlim terms (a high-level GO term for functional categorization). Analysis was performed using the GO annotation (BERARDINI *et al.* 2004) from the TAIR website (download 20051119; <http://www.arabidopsis.org/>). The frequency of each classification was obtained for the completely sequenced genome and for the gene lists for each of the 17 *trans*-eQTL hotspots (see RESULTS). Each gene list was tested for significant deviation ( $P \leq 0.001$ ) from the expected frequencies for the complete genome using  $\chi^2$  analysis (CHEN *et al.* 2005).

**Accession numbers:** The microarray data set used in this study has been deposited at EBI ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) under nos. E-TABM-126 and E-TABM-224.

## RESULTS

**Numbers of eQTL detected:** The majority of genes within the Arabidopsis genome exhibited heritable transcriptional variation that is controlled by eQTL in a Bay-0  $\times$  Sha RIL population (Table 1). Of the 22,746 Arabidopsis transcripts represented on the Affymetrix ATH1 microarray and whose levels were measured as e-traits in the 211 RILs, 69% were associated with one or more significant eQTL (LRT statistics  $> 12.0583$ , the 95% global permutation threshold; see MATERIALS AND METHODS). A total of 36,871 distinct eQTL were detected and counted using the Eql module of QTL Cartographer and an exclusionary window size of 20 cM (for details, see MATERIALS AND METHODS). The number of eQTL detected per transcript/e-trait varied from 0 to 11. The transcripts/e-traits associated with eQTL had an average of 2.34 and a median of two eQTL/transcript. Additionally, eQTL were identified for non-nuclear genes: 69% of the 155 organelle-encoded transcripts identified nuclear-encoded eQTL (Table 1).

**eQTL genomic distribution:** One-third of the e-traits with detectable eQTL had a *cis*-eQTL (*i.e.*, coincident with the physical position of the gene whose transcript level varied) (Table 1). This proportion was strikingly uniform across the five Arabidopsis chromosomes, as evidenced by the diagonal of significant LRT statistics across each chromosome (Figure 1A). Furthermore, there were approximately equal contributions of parental alleles that increased transcript abundance (Figure 1A).

The remaining (*trans*-)eQTL were nonuniformly distributed across the chromosomes, with the majority of eQTL localized on chromosome II (Figure 1, Table 1). eQTL were clustered into *trans*-eQTL hotspots that are visible as vertical bands of eQTL (Figure 1A) and as peaks of transcript numbers with common eQTL map positions (Figure 1B). Seventeen significant *trans*-eQTL hotspots were detected across the genome, with a minimum of one/chromosome. Of these 17 hotspots, 10

**TABLE 1**  
**Global eQTL description**

Chromosome	Genes <sup>a</sup>	Genes with an eQTL <sup>b</sup>		Genes with <i>cis</i> -eQTL <sup>c</sup>		Genes with an eQTL on this chromosome <sup>d</sup>		No. of eQTL <sup>e</sup>	
		Total	%	Total	%	Total	%	Total	%
I	5,862	4,114	70	1333	32	5,101	32	6,404	17
II	3,752	2,473	66	794	32	11,174	71	16,699	45
III	4,490	3,135	70	1024	33	3,096	20	3,664	10
IV	3,379	2,308	68	787	34	3,004	19	3,591	10
V	5,108	3,634	71	1189	33	4,988	32	6,513	18
Non-nuclear	155	107	69	NA		NA		NA	
Total	22,794	15,771	69	5127	32	15,771	100	36,871	100

The Affymetrix ATH1 microarray detects transcripts for genes on all five Arabidopsis chromosomes (I–V) as well as additional non-nuclear-encoded genes. The control probe sets are not included.

<sup>a</sup>The number of genes per chromosome or group whose transcripts are detected on the Affymetrix ATH1 microarray. Total is the sum for the entire genome.

<sup>b</sup>The number of genes from each chromosome or group whose transcript accumulation was controlled by at least one eQTL and the percentage of genes from each chromosome whose transcript accumulation was controlled by at least one eQTL.

<sup>c</sup>The number of genes per chromosome whose transcripts had a significant eQTL in the *cis* position, defined as within 3.5 cM of the gene's genetic position, and the percentage of genes determined by only those genes with at least one detectable eQTL.

<sup>d</sup>Genes with an eQTL on this chromosome lists the number of genes within the genome whose transcripts were significantly controlled by at least one eQTL on this specific chromosome.

<sup>e</sup>The number of significant eQTL detected on this specific chromosome.

contained eQTL associated with >200 different transcripts; the largest hotspot was associated with 2528 transcripts (Figure 1B; Table 2). Chromosome II had the greatest number of hotspots (5), as well as the hotspots associated with the greatest number of transcripts (Figure 1; Table 2). A GO analysis of the transcripts associated with the 17 *trans*-eQTL hotspots did not reveal any significant over- or underrepresentation of any functional or biological process GO categories (data not shown). In addition, our previous network eQTL analysis in this RIL population using 18 *a priori*-defined gene networks indicated that the chromosome II hotspots did not control the majority of these networks (KLIEBENSTEIN *et al.* 2006a).

A strong directional bias was evident for each of the *trans*-eQTL hotspots (Figure 1A; Table 2), such that the same parental allele upregulated most of the transcripts associated with a hotspot. At 8 *trans*-eQTL hotspots, the Sha allele increased transcript levels (positive effect); the other 9 hotspots were associated with the opposite allelic effect (*i.e.*, the Sha allele decreased accumulation). A global average of 54.7% of the 36,871 eQTL had a positive Sha allelic effect and 45.3% had a positive Bay-0 allelic effect. In contrast, the effects of each *trans*-eQTL hotspot significantly deviated from these global averages: for 15 of the 17 hotspots, >80% of the transcripts associated with each *trans*-eQTL hotspot, were affected positively by the same parental allele (Table 2).

**Proportion of phenotypic variation per eQTL:** Most individual eQTL accounted for only a small proportion of the associated transcript/e-trait's estimated phenotypic variation ( $R^2$ ): 89% of eQTL accounted for <20%

of the  $R^2$  for each transcript (Figure 2). The *cis*-eQTL typically controlled more of an e-trait's phenotypic variation than did the *trans*-eQTL (Figure 2). This difference in the  $R^2$  distributions between *cis*- and *trans*-eQTL was not due to the presence of SFPs because >91% of all genes with a *cis*-eQTL did not have a SFP (supplemental Table 3 at <http://www.genetics.org/supplemental/>). Furthermore, the genes with a SFP and a *cis*-eQTL had an  $R^2$  distribution similar to genes with a *cis*-eQTL that lacked a SFP (data not shown).

**eQTL allelic effects and transcript heritabilities:** The opposing directionality of allelic effects in the *trans*-eQTL hotspots suggests the potential for nonadditive genetic variation in transcript levels between Bay-0 and Sha. The presence of nonadditive variation is also supported by the observation that while the parental microarray data showed a significant difference in transcript levels for only 3351 of the 22,746 transcripts (false discovery rate of 0.05) (data not shown), the RIL microarray data allowed detection of eQTL for 69% of the e-traits, indicating that the levels of those transcripts are most likely genetically controlled. To further test for nonadditive genetic variation in this population, we estimated and compared the broad-sense heritability ( $H$ ) of 22,746 transcripts/e-traits with data from parental microarrays and from replicated microarrays for the 211 RILs. The vast majority of the e-traits showed greater heritability in the RILs than in the parents (Figure 3A), suggesting nonadditive genetic variation in the RILs. Therefore, for the majority of e-traits, the observed variation of transcript levels in the parents was not predictive of the variation present in their progeny.

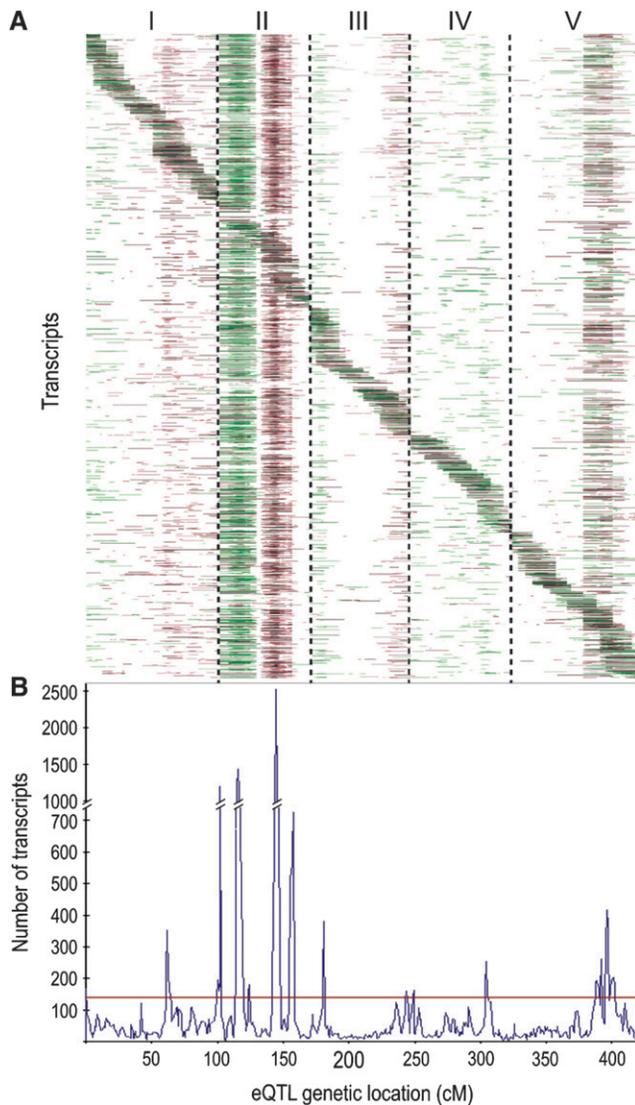


FIGURE 1.—Genomic architecture of eQTL across five Arabidopsis chromosomes. (A) Heat map of LRT statistics obtained by CIM eQTL analysis for 22,591 nuclear-encoded transcripts (y-axis) plotted against 464 markers (x-axis) across five chromosomes. Colors indicate chromosomal regions where LRT statistics were significantly greater than the global permutation threshold ( $GPT > 12.0583$ ) at  $P < 0.05$ . Red indicates a positive effect of the presence of the Sha allele, and green indicates a positive effect of the Bay-0 allele. Vertical dotted lines separate the five chromosomes (I–V). (B) Numbers of transcripts/e-traits for which eQTL are detected. The number of transcripts is indicated on the y-axis, plotted against the genetic location of the eQTL in centimorgans on the x-axis. The permuted threshold ( $P = 0.05$ ) for detection of a significant *trans*-eQTL hotspot is 133 transcripts, indicated by the red horizontal line.

**Transgressive segregation:** One potential source of nonadditive genetic variation is transgressive segregation whereby the progeny have average e-trait values that exceed the range of the parent values. Comparing transcript accumulation for 22,746 transcripts between the parents and RILs identified 14,258 transcripts for which at least 10% of the RILs were 1 standard deviation

beyond the parental range (data not shown). Since transgressive segregation can be due to the presence of eQTL with opposing parental allelic effects, our investigation revealed that of the 10,084 e-traits associated with two or more eQTL, 6911 exhibited eQTL with opposing additive effects, suggesting that such eQTL are a significant component of the nonadditive genetic variation within the 211 Bay-0  $\times$  Sha RILs. Of these 6911 e-traits, 4871 show evidence of transgressive segregation as described above, indicating that 1 standard deviation above and below the parental range is a conservative assessment of nonadditive genetic variation. Our statistical analyses and these summaries do not, however, indicate whether epistasis is a component of the nonadditive genetic variation that is prevalent in this population. Due to the complexity of the required statistical analyses involving both a large number of tests and a large number of eQTL, epistasis will be investigated in a subsequent effort.

**Transcript heritabilities in RILs vs. parents:** While the majority of e-trait estimated heritabilities showed little correlation between the parents and the RILs, some did exhibit a positive relationship (Figure 3, B and C). We found that the transcripts/e-traits associated with a *cis*-eQTL are most likely to exhibit a linear heritability relationship in the parents and RILs, because *cis*-eQTL tend to control more of the e-trait's variation ( $R^2$ , Figure 2). In fact, the 5127 e-traits with a *cis*-eQTL exhibited a stronger linear relationship of heritability between the parents and the RILs (Figure 3C) than did the 10,644 e-traits with only *trans*-eQTL (*i.e.*, no *cis*-linkages) (Figure 3D). Even though  $R^2$  values are not additive, if used as a relative measure of phenotypic variation for a specific transcript, the totaled  $R^2$  for transcripts with a *cis*-eQTL explains more of the relative phenotypic variation than the corresponding totaled  $R^2$  for transcripts with only *trans*-eQTL (supplemental Figure 3 at <http://www.genetics.org/supplemental/>). A small proportion of e-traits had high heritability ( $H > 0.7$ ) in the RILs, but no significant eQTL, suggesting that these effects are too small to detect in a sample of 211 individuals.

## DISCUSSION

The genetic architecture of transcript-level variation in the *A. thaliana* Bay-0  $\times$  Sha RIL population was surprisingly highly variable and complex. Single eQTL were detected for a large proportion (69%) of the 22,746 transcripts/e-traits (Table 1). Many e-traits were also associated with multiple eQTL. We realize that this experiment and these data represent only a single sample of the life-history variation for this species; however, the expression levels of the majority of transcripts showed quantitative variation. These transcript levels are probably quantitatively controlled, given the

**TABLE 2**  
**Assessment of directional bias of allelic effects at *trans*-eQTL hotspots**

Chromosome	Marker position (cM)	No. of eQTL <sup>a</sup>	Negative <sup>b</sup>	Positive <sup>c</sup>	% Bay-0 <sup>d</sup>	% Sha <sup>e</sup>	<i>P</i> -value <sup>f</sup>
I	0.01	164	142	22	86.6	13.4	$2.37 \times 10^{-26}$
I	61.51	353	16	337	4.5	95.5	$2.00 \times 10^{-53}$
I	99.88	193	51	142	26.4	73.6	$1.38 \times 10^{-7}$
I	101.88	186	175	11	94.1	5.9	$9.52 \times 10^{-41}$
II	0.01	1202	1028	174	85.5	14.5	$1.06 \times 10^{-172}$
II	13.59	1439	1341	98	93.2	6.8	$1.35 \times 10^{-291}$
II	22.97	182	164	18	90.1	9.9	$6.16 \times 10^{-34}$
II	42.44	2528	5	2523	0.2	99.8	0
II	55.87	710	10	700	1.4	98.6	$4.64 \times 10^{-122}$
III	9.14	381	367	14	96.3	3.7	$4.68 \times 10^{-89}$
III	72.38	162	26	136	16.0	84.0	$7.48 \times 10^{-14}$
III	77.48	163	30	133	18.4	81.6	$5.27 \times 10^{-12}$
IV	55.40	256	255	1	99.6	0.4	$3.08 \times 10^{-68}$
V	63.54	195	123	72	63.1	36.9	$6.14 \times 10^{-7}$
V	66.48	261	245	16	93.9	6.1	$5.58 \times 10^{-56}$
V	70.70	417	69	348	16.5	83.5	$4.12 \times 10^{-32}$
V	75.51	202	37	165	18.3	81.7	$1.32 \times 10^{-14}$

All eQTL with their LRT statistic maxima mapping to a marker at a *trans*-eQTL hotspot were analyzed for additive-effect estimates contributed by each parental allele. Permutation analysis established 133 eQTL as the significance threshold ( $\alpha = 0.05$ ) for declaring a significant *trans*-eQTL hotspot (see MATERIALS AND METHODS). Within each hotspot, the number of eQTL with a negative additive-effect estimate (*i.e.*, Bay-0 allele increased transcript level) was compared with the number of eQTL with a positive additive-effect estimate (*i.e.*, Sha allele increased transcript level).

<sup>a</sup> Number of eQTL with LRT statistic maxima located at the marker.

<sup>b</sup> Number of eQTL with a negative additive-effect estimate.

<sup>c</sup> Number of eQTL with a positive additive-effect estimate.

<sup>d</sup> Percentage of transcripts positively influenced by the Bay-0 allele.

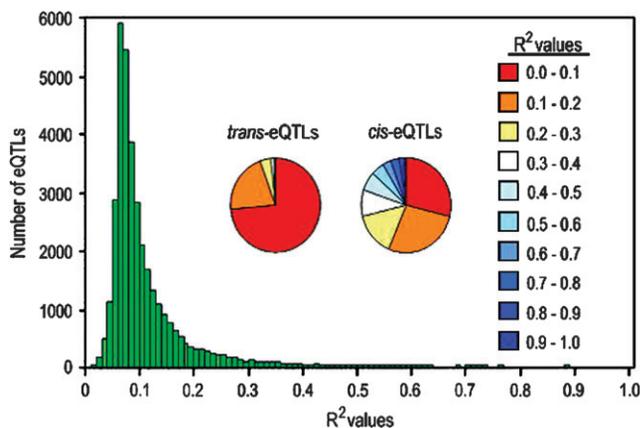
<sup>e</sup> Percentage of transcripts positively influenced by the Sha allele.

<sup>f</sup> *P*-values for  $\chi^2$  test for significant deviation from expected 50% contribution from each allele.

fact that this RIL population, derived from two inbred accessions, represents only a fraction of the variation present in this diverse species (NORDBORG *et al.* 2005; KLIEBENSTEIN *et al.* 2006b). The complexity of this

control also implies that transcriptional regulation in other higher plant species is likely to be as multifaceted. Our findings in *Arabidopsis*, in conjunction with studies in yeast (YVERT *et al.* 2003; BREM and KRUGLYAK 2005), mice (SCHADT *et al.* 2003; CHESLER *et al.* 2005), and humans (MORLEY *et al.* 2004), support the working hypothesis that global complexity of transcriptional regulation may be a general feature of eukaryotes.

The transcript levels of most genes in this *Arabidopsis* RIL population were controlled by multiple eQTL with small  $R^2$  values (Figure 2). The RIL sample size, the observable number of recombinants, and the statistical analysis that we employed can all be considered limiting factors of this study that affect our ability to detect eQTL of small phenotypic effect (BEAVIS 1998; MACKAY 2001; DOERGE 2002; KIM *et al.* 2005). The genetic variance associated with an eQTL (*i.e.*, the proportion of the phenotypic variance for which it accounts) is larger in a RIL population than in  $F_2$  or backcross populations (BEAVIS 1998; FLINT *et al.* 2005). Therefore, our use of a RIL mapping population structure in conjunction with 211 biologically replicated RILs greatly facilitated the detection of smaller-effect eQTL due to the relatively large number of observed recombination events that, in turn, allowed more precise estimation of phenotypic



**FIGURE 2.**—Distribution of the percentage of phenotypic effect ( $R^2$ ) for all eQTL. A histogram of the distribution of  $R^2$  values for all 36,871 eQTL is shown separated into 0.01 bins; the maximum  $R^2$  was 0.97. The two pie graphs illustrate the  $R^2$  distributions for eQTL that are *cis* (5127 total *cis*-eQTL) or *trans* (31,777 total *trans*-eQTL) to the gene's physical position. The color scale to the right indicates the  $R^2$  bins for the pie graphs.

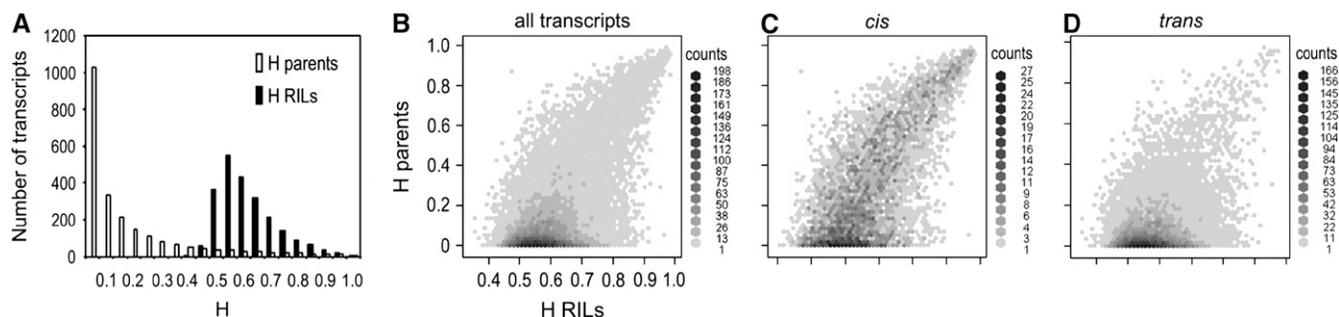


FIGURE 3.—Transcript-level/e-trait heritabilities in RILs *vs.* parents. (A) Histograms of estimated broad-sense heritability ( $H$ ) values in RILs and parents (Bay-0 and Sha). Solid bars show the histogram of heritability values for all 22,746 transcripts/e-traits as estimated with the RIL microarray data. Open bars show the histogram of  $H$  for all 22,746 transcripts/e-traits as estimated with the parental (Bay-0 and Sha) microarray data. (B–D) Relationship between estimated transcript/e-trait  $H$  in the RILs and in the parents.  $H$ -values for all 22,746 transcripts are plotted as a hexbin graph for both the RIL ( $x$ -axis) and the parental ( $y$ -axis)  $H$  estimates. The shading scale to the right of each figure indicates data-point density per bin. The  $x$ - and  $y$ -axes for B–D are identical. (B) All transcripts. (C) Transcripts for which a *cis*-eQTL was mapped. (D) Transcripts that mapped to only *trans*-eQTL (no *cis*).

variances. Nevertheless, it is likely that additional small-effect eQTL remain undetected in this population. The observed skewed (*e.g.*, long right-hand tail)  $R^2$  histogram (Figure 2) suggests that an increase in population size ( $n > 211$ ) and additional biological replication per line would most likely increase detection of eQTL with  $R^2$  values  $< 0.1$  (KIM *et al.* 2005).

This study revealed that small-effect eQTL were prevalent in the 211 RILs, with  $< 6\%$  of the 36,871 eQTL accounting for a large proportion of the estimated phenotypic variation ( $R^2 > 0.3$ ). These results contrast to studies that were based on smaller population sizes in which large phenotypic effect *cis*-eQTL were prevalent (SCHADT *et al.* 2003; DECOOK *et al.* 2006). Studies based on small sizes in combination with population structures that do not maximize recombination events reduce QTL resolution and lead to sampling bias when detecting eQTL/QTL of larger phenotypic effect (BEAVIS 1998; BOGDAN and DOERGE 2005; DE KONING and HALEY 2005; GIBSON and WEIR 2005).

Our observation that *cis*- and *trans*-eQTL are associated with different distributions of  $R^2$  (Figure 2) was also found in *Drosophila* (MEIKLEJOHN *et al.* 2003; WAYNE *et al.* 2004; HUGHES *et al.* 2006). A potential explanation for this observation is that the transcript abundance of most genes is regulated by multiple signal transduction networks at multiple levels (*e.g.*, transcription and transcript stability); thus a polymorphism in any one of these regulatory levels may be functionally limited to only a small change in transcript accumulation for those genes controlled in *trans* by that polymorphism. However, induced mutations in transcription factors can generate large expression differences in genes regulated in *trans* by the transcription factor. Alternatively, the polymorphism underlying a *trans*-eQTL hotspot has greater potential to be pleiotropic than a *cis*-eQTL with no corresponding hotspot. Large-effect mutations in pleiotropic genes are more likely to be deleterious than

mutations in less interconnected genes (WRIGHT 1977; TURELLI 1988; WAGNER 2000; JEONG *et al.* 2001; FRASER *et al.* 2002; YU *et al.* 2004). There may be an evolutionary fitness limitation on the potential genetic effect of polymorphisms that generate *trans*-eQTL hotspots. Additional populations and species need to be evaluated to determine if *cis*-eQTL effects are larger than *trans*-eQTL effects in general.

The majority of transcripts/e-traits exhibited higher estimated broad-sense heritabilities ( $H$ ) within the RILs than in the parents (Figure 3), indicating that the transcript-level variation present in these two homozygous inbred *Arabidopsis* accessions does not accurately predict transcript variation in their RIL progeny. Furthermore, our results indicate that if parental transcript variation is used to select a restricted set of genes for subsequent study in the progeny, many genes would be overlooked. The difference in estimated transcript-level heritabilities for parents and RILs also suggests that the majority of the genetic variance influencing the levels of most transcripts is nonadditive (*i.e.*, includes dominance and epistatic interaction components; LYNCH and WALSH 1998). Indeed,  $\sim 70\%$  of the transcripts with two or more eQTL showed opposite allelic effects. Since the parents and RILs were grown together in the same environment, it is unlikely that genotype  $\times$  environment interactions were a major source of differences in phenotypic variance estimates between parents and RILs, supporting the conclusion that nonadditive genetic variance was the main contributor to differences in e-trait  $H$  between parents and RILs. Our results are consistent with a yeast global transcriptome study for which transgressive segregation and nonadditive genetic variation were also prevalent (BREM and KRUGLYAK 2005). The predominance of nonadditive genetic variation in our study also suggests that surveys of transcript-level variation in parental accessions, at least within *Arabidopsis*, will significantly underestimate the actual

transcript-level variation in progeny populations derived from intercrossing these accessions. However, the underestimation will primarily affect *trans*-eQTL as the *cis*-eQTL typically exhibit larger  $R^2$  values (Figure 2) and a positive correlation of transcript-level heritability between the parents and their RILs (Figure 3C). This latter observation supports our previous suggestion that correlation between transcript-level variation and sequence polymorphism within seven Arabidopsis accessions was due to *cis* control of the majority of the detected transcript-level variation (KLIEBENSTEIN *et al.* 2006b). Interestingly, a small number of transcripts (41) exhibited high heritability, but no detectable eQTL. This could be due to control by multiple loci with small additive effects or control by loci that interact non-additively (BREM and KRUGLYAK 2005). Neither of these issues is investigated here, which serves as motivation for further work in this area.

While *cis*-eQTL accounted for a greater proportion of the phenotypic variance, the majority of eQTL within this RIL population are situated in *trans* to the genes whose transcript levels varied. Furthermore, most *trans*-eQTL clustered into regions defined as *trans*-eQTL hotspots that control the levels of a larger number of transcripts but only a small fraction of the variation for each transcript. *Trans*-eQTL hotspots have been reported in other organisms (BREM *et al.* 2002; SCHATZ *et al.* 2003). Interestingly, in our study the e-traits affected by a *trans*-eQTL hotspot also showed directionality such that most transcripts were either up- or downregulated by the same parental allele at the *trans*-eQTL (Figure 1A, Table 2). However, examination of their genomic positions and GO annotation analysis did not identify any functional categories significantly associated with any of the 17 *trans*-eQTL hotspots (data not shown); thus the biological function of the *trans*-eQTL hotspots remains unknown. Finally, our comparison of the *trans*-eQTL hotspots to known phenotypic QTL locations did not identify any significant enrichment of phenotypic QTL in these regions (KLIEBENSTEIN *et al.* 2006a; data not shown).

Clearly, further genetic dissection of each *trans*-eQTL hotspot is required to identify the genes that influence so many other genes in *trans* and to understand the biological function of such hotspots. One approach to finding candidate causal genes is to identify *cis*-eQTL underlying a *trans*-eQTL hotspot; our analysis identified ~100 *cis*-eQTL/hotspot. However, due to the potential genetic complexity of e-QTL regions, it may not be a straightforward task to identify the specific causal gene or genes. QTL and e-QTL detected as single genetic effects may be composed of multiple, physically linked loci, each associated with a small portion of the phenotypic variation (FLINT *et al.* 2005). Understanding how closely linked loci of small effect interact to control a quantitative phenotype would require a very large mapping population (*i.e.*, many recombinants) and appro-

priate follow-up studies to separate and isolate each of the QTL involved (KIM *et al.* 2005).

An important question that remains to be addressed is how transcript-level variation is related to and controls downstream phenotypic variation (GIBSON and WEIR 2005). Studies in plants have correlated specific phenotypic QTL with transcript-level variation controlled in *cis* (KLIEBENSTEIN *et al.* 2001; LAMBRIX *et al.* 2001; ZHANG *et al.* 2006). Several of these studies identified *cis*-regulated transcript-level variation in signal transduction genes controlling flowering time or circadian rhythm variation (JOHANSON *et al.* 2000; CAICEDO *et al.* 2004; WERNER *et al.* 2005). This *cis* variation in a signal transduction component potentially controls small changes in transcript accumulation for numerous genes in *trans*. As mentioned previously, we hypothesize that *cis*-eQTL variation may underlie some *trans*-eQTL hotspots whereby the genes affected in *trans* are the mechanism through which the *cis*-affected gene influences the downstream phenotype. Testing this hypothesis will require high-throughput analysis of a RIL population for both transcript level and downstream phenotypic variation and subsequent functional analysis of the causal genes.

A fundamental issue in quantitative genetics is how the genotype determines the quantitative trait phenotype (MACKAY 2001). A study of transcriptional variation is merely one component of this question. Determining the actual biological relationship between transcript-level variation and other phenotypes, such as protein and metabolite levels, that in turn influence trait phenotypes farther downstream, is a highly complex task (JANSEN *et al.* 2002). For example, post-transcriptional regulation can diminish the correlation between transcript levels and protein levels. Furthermore, inaccuracy in measurements of global transcript accumulation can reduce the ability to detect the relationship between mRNA levels and downstream phenotypes. Reducing experimental error via biological replication of large segregating populations representing many recombination events (JANSEN and NAP 2001) can improve the ability to detect meaningful biological associations of transcriptomic variation with proteomic and metabolomic variation. The Bay-0  $\times$  Sha RIL population, consisting of hundreds of genetically stable inbred lines, and the global eQTL analysis described here provide a basis from which to study how transcript-level variation may influence downstream phenotypic trait variation in higher plants. Such studies will enable determination of how the genotype determines quantitative trait phenotypes and contribute toward a deeper understanding of the genotype-phenotype relationship in eukaryotes.

We thank Rebecca Walker and Tanya Tang for technical assistance and Steve Edberg for assistance with data management and dissemination. This research was supported by the National Science Foundation 2010 Project, grant no. MCB-0115109 to D.A.S., R.W.M., and R.W.D.

## LITERATURE CITED

- BASTEN, C. J., B. S. WEIR and Z.-B. ZENG, 1999 *QTL Cartographer*, Version 1.13. Department of Statistics, North Carolina State University, Raleigh, NC.
- BEAVIS, W. D., 1998 QTL analyses: power, precision, and accuracy, pp. 145–162 in *Molecular Dissection of Complex Traits*, edited by A. H. PATERSON. CRC Press, Cleveland/Boca Raton, FL.
- BERARDINI, T., S. MUNDODI, R. REISER, E. HUALA, M. GARCIA-HERNANDEZ *et al.*, 2004 Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiol.* **135**: 1–11.
- BOGDAN, M., and R. W. DOERGE, 2005 Biased estimators of quantitative trait locus heritability and location in interval mapping. *Heredity* **95**: 476–484.
- BOLSTAD, B. M., R. A. IRIZARRY, M. ASTRAND and T. P. SPEED, 2003 A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**: 185–193.
- BREM, R. B., and L. KRUGLYAK, 2005 The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. USA* **102**: 1572–1577.
- BREM, R. B., G. YVERT, R. CLINTON and L. KRUGLYAK, 2002 Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**: 752–755.
- BYSTRYKH, L., E. WEERSING, B. DONTJE, S. SUTTON, M. T. PLETCHER *et al.*, 2005 Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics.' *Nat. Genet.* **37**: 225–232.
- CAICEDO, A. L., J. R. STINCHCOMBE, K. M. OLSEN, J. SCHMITT and M. D. PURUGGANAN, 2004 Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proc. Natl. Acad. Sci. USA* **101**: 15670–15675.
- CHEN, W. J., S. H. CHANG, M. E. HUDSON, W.-K. KWAN, J. LI *et al.*, 2005 Contribution of transcriptional regulation to natural variations in Arabidopsis. *Genome Biol.* **6**: R32.
- CHESLER, E. J., L. LU, S. M. SHOU, Y. H. QU, J. GU *et al.*, 2005 Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.* **37**: 233–242.
- CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.
- DECOOK, R., S. LALL, D. NETTLETON and S. H. HOWELL, 2006 Genetic regulation of gene expression during shoot development in Arabidopsis. *Genetics* **172**: 1155–1164.
- DE KONING, D. J., and C. S. HALEY, 2005 Genetical genomics in humans and model organisms. *Trends Genet.* **21**: 377–381.
- DOERGE, R. W., 2002 Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* **3**: 43–52.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*. Longman Group, Essex, UK.
- FLINT, J., W. VALDAR, S. SHIFMAN and R. MOTT, 2005 Strategies for mapping and cloning quantitative trait genes in rodents. *Nat. Rev. Genet.* **6**: 271–286.
- FRASER, H. B., A. E. HIRSH, L. M. STEINMETZ, C. SCHARFE and M. W. FELDMAN, 2002 Evolutionary rate in the protein interaction network. *Science* **296**: 750–752.
- GIBSON, G., and B. WEIR, 2005 The quantitative genetics of transcription. *Trends Genet.* **21**: 616–623.
- HALEY, C. S., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.
- HUBNER, N., C. A. WALLACE, H. ZIMDAHL, E. PETRETTO, H. SCHULZ *et al.*, 2005 Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet.* **37**: 243–253.
- HUGHES, K. A., J. F. AYROLES, M. M. REEDY, J. M. DRNEVICH, K. C. ROWE *et al.*, 2006 Segregating variation in the transcriptome: cis regulation and additivity of effects. *Genetics* **173**: 1347–1355.
- JANSEN, R. C., and J. P. NAP, 2001 Genetical genomics: the added value from segregation. *Trends Genet.* **17**: 388–391.
- JANSEN, R. C., J. P. NAP and L. MLYNAROVA, 2002 Errors in genomics and proteomics. *Nat. Biotechnol.* **20**: 19.
- JEONG, H., S. P. MASON, A. L. BARABASI and Z. N. OLTVAI, 2001 Lethality and centrality in protein networks. *Nature* **411**: 41–42.
- JOHANSON, U., J. WEST, C. LISTER, S. MICHAELS, R. AMASINO *et al.*, 2000 Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science* **290**: 344–347.
- KIM, K., M. A. L. WEST, R. W. MICHELMORE, D. A. ST. CLAIR and R. W. DOERGE, 2005 Old methods for new ideas: genetic dissection of the determinants of gene expression levels, pp. 89–108 in *Genome Exploitation: Data Mining the Genome*, edited by J. P. GUSTAFSON, R. SHOEMAKER and J. W. SNAPE. Springer Science & Business Media, University of Missouri, Columbia, MO.
- KIRST, M., A. A. MYBURG, J. P. G. DE LEON, M. E. KIRST, J. SCOTT *et al.*, 2004 Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiol.* **135**: 2368–2378.
- KLIEBENSTEIN, D., V. LAMBRIX, M. REICHEL, J. GERSHENZON and T. MITCHELL-OLDS, 2001 Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*. *Plant Cell* **13**: 681–693.
- KLIEBENSTEIN, D., M. WEST, H. VAN LEEUWEN, O. LOUDET, R. DOERGE *et al.*, 2006a Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics* **7**: 308.
- KLIEBENSTEIN, D. J., M. A. L. WEST, H. VAN LEEUWEN, K. KYUNGA, R. W. DOERGE *et al.*, 2006b Genomic survey of gene expression diversity in *Arabidopsis thaliana*. *Genetics* **172**: 1179–1189.
- KROYMANN, J., S. DONNERHACKE, D. SCHNABELRAUCH and T. MITCHELL-OLDS, 2003 Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus. *Proc. Natl. Acad. Sci. USA* **100**: 14587–14592.
- LAMBRIX, V., M. REICHEL, T. MITCHELL-OLDS, D. KLIEBENSTEIN and J. GERSHENZON, 2001 The Arabidopsis epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichopusia ni* herbivory. *Plant Cell* **13**: 2793–2807.
- LOUDET, O., S. CHAILLOU, C. CAMILLERI, D. BOUCHEZ and F. DANIEL-VEDELE, 2002 Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis. *Theor. Appl. Genet.* **104**: 1173–1184.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- MACKAY, T. F. C., 2001 The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**: 303–339.
- MEIKLEJOHN, C. D., J. PARSCH, J. M. RANZ and D. L. HARTL, 2003 Rapid evolution of male-biased gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **100**: 9894–9899.
- MORLEY, M., C. M. MOLONY, T. M. WEBER, J. L. DEVLIN, K. G. EWENS *et al.*, 2004 Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**: 743–747.
- NORDBORG, M., T. T. HU, Y. ISHINO, J. JHAVERI, C. TOOMAJIAN *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**: e196.
- SCHADT, E. E., S. A. MONKS, T. A. DRAKE, A. J. LUSIS, N. CHE *et al.*, 2003 Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**: 297–302.
- TURELLI, M., 1988 Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* **42**: 1342–1347.
- WAGNER, A., 2000 The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. *Genetics* **154**: 1389–1401.
- WAYNE, M. L., Y.-J. PAN, S. V. NUZHIDIN and L. M. MCINTYRE, 2004 Additivity and transacting effects on gene expression in male *Drosophila simulans*. *Genetics* **168**: 1413–1420.
- WERNER, J. D., J. O. BOREVITZ, N. WARTHMAN, G. T. TRAINER, J. R. ECKER *et al.*, 2005 Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc. Natl. Acad. Sci. USA* **102**: 2460–2465.
- WEST, M. A. L., H. VAN LEEUWEN, A. KOZIK, D. J. KLIEBENSTEIN, R. W. DOERGE *et al.*, 2006 High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. *Genome Res.* **16**: 787–795.
- WRIGHT, S., 1977 *Evolution and the Genetics of Populations*. University of Chicago Press, Chicago.
- YU, H. Y., D. GREENBAUM, H. X. LU, X. W. ZHU and M. GERSTEIN, 2004 Genomic analysis of essentiality within protein networks. *Trends Genet.* **20**: 227–231.

- YVERT, G., R. B. BREM, J. WHITTLE, J. M. AKEY, E. FOSS *et al.*, 2003 Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35**: 57–64.
- ZENG, Z.-B., C.-H. KAO and C. J. BASTEN, 1999 Estimating the genetic architecture of quantitative traits. *Genet. Res.* **75**: 345–355.

- ZHANG, Z.-Y., J. A. OBER and D. J. KLIEBENSTEIN, 2006 The gene controlling the quantitative trait locus *EPITHIOSPECIFIER MODIFIER1* alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. *Plant Cell* **18**: 1524–1536.

Communicating editor: J. F. DOEBLEY