eQTL mapping in the Bay x Sha recombinant population.

Expression levels for 24,576 traits (Gene-specific Sequence Tags: GSTs, CATMA array version 2) was measured in RNA extracted from developing seeds (harvested 10 days after pollination) of 157 RILs from the Bay-0 x Sha population (hereafter BaySha)[1]. Display, sample treatment, microarray analysis and data normalisation were performed as indicated in Methods within the main text. We used this dataset to develop and test the efficiency of the novel R/eqtl package as a tool for identifying expression quantitative trait loci (eQTLs) in recombinant populations.

Linkage analyses on expression traits allowed us to identify 19,021 significant eQTLs ($FDR = 5\%, \text{LOD} > 2.47$ after 1,000 permutation tests on 100 randomly chosen expression traits; Figure S6a; Table S2c). For each of them, the variance explained, the supporting interval, the additive effect and the classification as potentially cis- or trans-acting eQTL (see main text for details) were determined. The accurate results provided by R/eqtl in the BaySha population allowed us to perform further mapping and crosses comparisons with the microarray data generated in the CviCol and BurCol recombinant populations.

The number of eQTLs identified in BaySha exceeds those ones found for the CviCol and BurCol recombinant populations by 4.4 and 2.9 fold, respectively. The majority of the eQTLs were mapped as distant (91.7%), and only a minority (8.3%) was found to act locally (Figure S6a-b). The significant over-representation of distant association and the total lower number of local-eQTLs detected in the BaySha population explains the greater difference between this set and the two crosses involving the Col-0 accession. Nevertheless, in agreement with CviCol and BurCol findings, we observed a higher average variance explained by local eQTLs compared to distant-eQTLs (13.2% vs. 11.3% respectively, $P < 2.2\text{e-16}$, Figure S6b, Figure S7) with more local-eQTLs having larger effects and significance (Figure S6b). Most of the traits were associated with a unique genomic region (Figure S8), in agreement with CviCol and BurCol findings.

The distribution of the distant-eQTLs along the genome was not uniform and most of them gathered in specific regions of the genome (Figure S9). In order to map intervals significantly over-represented with distant-eQTLs (hotspots), we performed a permutation test after dividing the genome into 1Mb intervals (see Methods in main text). Only three significant hotspots were detected localised in the left arm of chromosome 4 and 5. Interestingly, two well-characterised pleiotropic QTLs known to segregate in this cross were found within these regions. Below the strongest hotspot (gathering ~12,000 eQTLs on chromosome 4) lies the flowering time gene FRIGIDA ($FRI$) as an obvious effector. Similarly, the second major hotspot, linked to the expression of over 2,000 genes, localised on chromosome 5 in an
interval neighbouring the FLC locus. Variation at FRI and FLC is known to control most of the variation for flowering time in this cross [1]. Additionally, these hotspots are not independent, with most of the genes controlled by a distant-eQTL at FLC also controlled by a distant-eQTL at FRI, in most instances with opposite allelic effects. In other words, this means that hundreds of transcripts have a genetic architecture (= detected QTLs and their parameters) which is very close to the one obtained for the flowering precocity trait itself. Although FLC has been hypothesised to directly regulate many genes outside of the flowering time pathway [2], the observed pattern fits better with an indirect effect of flowering time on the timing of gene expression in the developing seeds after pollination.

Previously, linkage mapping of rosette transcripts for this line detected 36,871 eQTLs, which is almost twice as much as detected in this study. Contrary to our results, many of the eQTLs gathered in a distant-hotspot localised on chromosome 2 and no evidence for a large number of transcripts regulated by either FRI or FLC was reported [3]. Discrepancies between studies could be explained by the different stages at which very different tissues were sampled, where different trans-factors could lead the expression of thousands of transcripts.

The BaySha developing seeds eQTLs results are stored in the QTLStore portal.

References

Supporting Figures

Figure S1. Venn diagram depicting the overlap between genes with differential expression in parental accessions pairs (Cvi vs. Col and Bur vs. Col).

Figure S2. Histograms of the explained phenotypic variance ($R^2$; %) for the eQTLs in the a.CviCol and b. BurCol populations.

Figure S3. Number of eQTLs per trait. The percentage and number of traits explained by 1 to 5 eQTLs are indicated along the y-axis and on top of each bar, respectively. a. CviCol b. BurCol at a FDR of 5%.

Figure S4. Venn diagram depicting the overlap between probes with local eQTLs in the CviCol and BurCol populations.

Figure S5. Histogram of the number of probes with a significant eQTL for different numbers of hidden factors tested with VBQTL in CviCol.

Figure S6. Genetic landscape for transcript accumulation variation in BaySha. a. eQTL heatmap for BaySha population significant at a 5% FDR. Each horizontal bar represents an eQTL mapped on the x-axis and controlling the accumulation of a transcript expressed from the locus indicated on the y-axis. The colour of the bar indicates the direction and strength of the eQTL additive effect, and its length along the x axis encompasses the eQTL support interval. Local eQTLs form the diagonal, while distant eQTLs fall elsewhere in the map. b. Bar plot indicating the proportion of local and distant eQTLs for increasing LOD value intervals.

Figure S7. Histogram of the explained phenotypic variance ($R^2$) for the eQTLs in the BaySha population

Figure S8. Number of eQTLs per trait in BaySha. The percentage and number of traits explained by 1 to 5 eQTLs are indicated along the y-axis and on top of each bar, respectively.

Figure S9. Distribution of distant-eQTLs along the genome in BaySha. The number of eQTLs (y-axis) is plotted against the physical position of the 1Mb-window where they peak (x-axis). Intervals with an excess of eQTLs relative to the threshold estimated by permutation (red dashed line) were classified as hotspots.
Figure S1
Figure S2
Figure S3

(a) Distribution of eQTLs for CviCol with percentages:
- 3464 eQTLs
- 342 eQTLs
- 42 eQTLs
- 708 eQTLs
- 4 eQTLs

(b) Distribution of eQTLs for BurCol with percentages:
- 4905 eQTLs
- 708 eQTLs
- 61 eQTLs
- 5 eQTLs
- 2 eQTLs
Figure S5
a

![Graph showing physical eQTL location (Chromosome) on the x-axis and additive effect on the y-axis. The graph has a color gradient indicating Bay and Sha.](image)

b

![Bar chart showing percentage of local and distant effect across different LOD values.](image)

Figure S6
Figure S7 Frequency

Explained Phenotypic Variance (%)
Figure S8
Figure S9