DOI: 10.1534/genetics.103.025213

Quantitative Trait Loci Associated With Adventitious Shoot Formation in Tissue Culture and the Program of Shoot Development in Arabidopsis

Sonia Lall,* Dan Nettleton,† Rhonda DeCook,† Ping Che* and Stephen H. Howell*,[1]

**Plant Sciences Institute and* †*Department of Statistics, Iowa State University, Ames, Iowa 50010*

Manuscript received December 1, 2003
Accepted for publication April 19, 2004

ABSTRACT

Arabidopsis ecotypes, Columbia (Col) and Landsberg *erecta* (L*er*), differ in their capacity to regenerate shoots in culture, as do many other cultivars and varieties of the same plant species. Recombinant inbred (RI) lines derived from a cross of Col × L*er* were scored for shoot regeneration, and the Arabidopsis genome was scanned using composite interval mapping for loci associated with shoot regeneration. Three QTL were identified—a major one on chromosome 5 in which the Col parent contributed the superior allele and two minor QTL on chromosomes 1 and 4 in which the L*er* parent contributed the superior alleles. The RI lines were binned into genotypic pools to isolate the effects of the major QTL on chromosome 5 while holding the minor QTL constant. To identify genes with expression levels that are associated with the allelic state of the major QTL on chromosome 5, oligonucleotide array expression patterns for genes in the LLC pool (L*er* alleles at the minor QTL and a Col allele at the major QTL) were compared to those in the LLL pool (L*er* alleles at all QTL). The genes that were significantly differentially expressed between the two pools included several encoding transcription factors and signaling or transposon-related proteins.

ADVENTITIOUS shoot formation in tissue culture is the means by which many plant species are commercially micropropagated, particularly ornamental plants in the families Araceae, Begoniaceae, Gesneriaceae, and Liliaceae (Preil 2003). It is also the means by which many transgenic plants are produced, *i.e.*, by regenerating shoots from transformed cells or tissues. A common problem encountered in micropropagation and plant transformation is that within a plant species, different varieties and cultivars vary widely in their capacity to regenerate. The unpredictable responses of different varieties to standard shoot induction conditions provoked Koornneef *et al.* (1993) to state that "this has left many researchers with the feeling that a successful tissue culture is more an art than a science."

Nonetheless, to uncover the genetic basis for differences in shoot regeneration, investigators have used qualitative or quantitative genetic approaches in *Brassica oleracea* (Buiatti *et al.* 1974), *Lycopersicon esculentum* and/or *peruvianum* (Frankenberger *et al.* 1981; Koornneef *et al.* 1987, 1993), *Zea mays* (Armstrong *et al.* 1992), *Solanum chacoense* (Birhman *et al.* 1994), *Triticum aestivum* (Fennell *et al.* 1996), *Hordeum vulgare* (Komatsuda *et al.* 1993), *Cucumis sativus* (Naldoska-Orezyk and Malepszy 1989), *Helianthus annuus* (Sarrafi *et al.* 1996), and *Arabidopsis thaliana* (Schiantarelli *et al.* 2001).

In addition, individual genes have been identified in Arabidopsis that promote adventitious shoot formation in culture. The first such gene discovered was *CYTOKININ INDEPENDENT1* (*CKI1*), a gene that when overexpressed confers cytokinin-independent shoot formation from callus in Arabidopsis tissue culture (Kakimoto 1996). *CKI1* encodes a histidine kinase related to sensory receptors in two-component signaling pathways (Stock *et al.* 2000; West and Stock 2001; Lohrmann and Harter 2002). At the time of its discovery, *CKI1* was proposed to be a cytokinin receptor (Kakimoto 1996). However, CKI1 has not been shown to bind cytokinins at physiological cytokinin levels, and so its role in cytokinin signaling is unclear.

Overexpression of *ARABIDOPSIS RESPONSE REGULATOR2* (*ARR2*) also stimulates shoot formation in culture (Hwang and Sheen 2001). ARR2 is a B-type response regulator that is thought to be a nuclear activator of cytokinin-responsive genes (Sakai *et al.* 2000; Hwang and Sheen 2001). The stimulatory effects of *ARR2* overexpression on shoot formation are surprising given the assumption that the activation and not merely the presence of the response regulator is required for signaling. In a protoplast system, transfections with *ARR2* constructs basally activated the promoter of a target gene in the absence of cytokinin; however, cytokinin addition resulted in much higher activation (Hwang and Sheen 2001). Regulation of ARR2 by a dissociable repressor might explain how *ARR2* overexpression promotes shoot development—by overwhelming the capacity of a repressor that normally keeps ARR2 in an inactive state.

[1]*Corresponding author:* Plant Sciences Institute, 1073 Roy J. Carver Co-Laboratory, Iowa State University, Ames, IA 50011.
E-mail: shh@iastate.edu

Another gene that stimulates shoot development when overexpressed is *ENHANCER OF SHOOT REGENERATION1* (*ESR1*; Banno *et al.* 2001). *ESR1* was discovered by screening a cDNA library for genes that enhance shoot formation in the absence of cytokinins. *ESR1* encodes an AP2-domain-containing transcription factor, and when *ESR1* expression was placed under the control of an estradiol-inducible promoter, shoot formation was enhanced by the combination of inducer and cytokinin (Banno *et al.* 2001). Overexpression of either *CUP-SHAPED COTYLEDON1* or *CUP-SHAPED COTYLEDON2* (*CUC1* or *CUC2*; driven by the 35S promoter) also promotes shoot formation from callus in the presence of cytokinin (Daimon *et al.* 2003). Hence, *CUC1* and *-2* do not bypass the requirement for cytokinins, but their overexpression enhances shoot development in the presence of cytokinins.

In this study, we analyzed shoot regeneration capacity in recombinant inbred lines of Arabidopsis to identify loci associated with shoot regeneration differences between the Columbia and Landsberg *erecta* ecotypes. Schiantarelli *et al.* (2001) conducted a similar study in Arabidopsis in which they categorized shoot regeneration in recombinant inbred (RI) lines as being similar to one or the other parent or as an intermediate response. In this study, we treated shoot regeneration as a quantitative trait and used metric data to locate controlling loci by composite interval mapping. In addition, we examined the impact of a major shoot regeneration quantitative trait locus (QTL) on the expression pattern of genes during shoot development in culture.

MATERIALS AND METHODS

Plant materials and tissue culture procedures: *Arabidopsis thaliana* (L.) Heynh ecotypes Columbia (Col), Landsberg *erecta* (L*er*), and the first 100 Lister and Dean RI lines (Lister and Dean 1993) were used in this study. Lister and Dean RI lines were obtained from the Arabidopsis Biological Resource Center (ABRC; stock no. CS1899).

Seeds were surface sterilized with 30% bleach (Clorox), 0.01% Triton X-100 for 10 min followed by five rinses of sterile water. Seeds were resuspended in 0.1% agarose and dispensed onto petri plates containing PNS medium (Lincoln *et al.* 1990) substituted with 0.5 g liter^{-1} 2-(4-morpholino)-ethane sulfonic acid (MES) at pH 5.7 and 10 g liter^{-1} sucrose. Seeds were stratified in the dark at 4° for 4 days, and seedlings were germinated and grown at 21° in the light, illuminated continuously with cool white fluorescence light at 65–85 μE m^{-2}. Shoots were regenerated from root explants using a protocol derived from Valvekens *et al.* (1988). Essentially, root segments (~5 mm) from 7-day-old seedlings were explanted onto callus induction medium (CIM; B5 medium substituted with 20 g liter^{-1} glucose, 0.5 g liter^{-1} MES, 2.2 μM 2,4-dichlorophenoxy acetic acid, and 0.2 μM kinetin) for 4 days (Gamborg *et al.* 1968). Explants were then transferred to shoot induction medium (SIM; B5 medium with added 20 g liter^{-1} glucose, 0.5 g liter^{-1} MES, 5.0 μM 2-isopentenyl adenine, and 0.9 μM 3-indoleacetic acid) and incubated for 6 days for RNA extraction or 15 days for counting shoots.

Recombinant inbred line analysis: Shoots from the 100 RI lines and parents were scored 15 days after incubation on SIM. One root segment from each of the 100 RI lines and the two parents was randomly assigned to a position within each of six petri plates to obtain six independent measurements of the number of shoots per root explant in each of the 100 RI lines and the two parents. Due to contamination of samples, data from one plate for 7 of the RI lines were lost. Thus results are based on a total of $6 \times 102 - 7 = 605$ shoot counts. The square-root transformation recommended for count data by Anscombe (1948) was used for all analyses to obtain roughly constant variation of trait variance within each line. Untransformed shoot counts were right skewed and exhibited higher variation in lines with higher shoot counts.

Version 1.3 of Windows QTL Cartographer (Basten *et al.* 1999) was used to perform composite interval mapping (Jansen and Stam 1994; Zeng 1994) to scan the Arabidopsis genome for loci conditioning shoot development in the RI lines. A total of 137 markers with an average intermarker distance of ~4.4 cM were utilized for the genome scan. The threshold for genomewise significance at the 0.05 level was estimated from 10,000 data permutations (Churchill and Doerge 1994). Cartographer default settings for composite interval mapping were used. In particular, loci every 2 cM were evaluated for association with the trait using a model that included up to 5 background markers to account for the effects of segregating QTL outside the region of the examined locus. Background markers were selected using stepwise regression and were individually excluded whenever they fell within 10 cM of the locus being evaluated for association with the trait.

In addition to the traditional composite interval mapping scan, several follow-up analyses were conducted as part of a mixed linear model analysis of the shoot-count data. The mixed linear model included random effects for plates and fixed effects for RI lines and parents. Contrasts of line means were used to estimate the effect of each identified QTL, test for interactions among identified QTL, and scan the genome for evidence of epistatic interactions between pairs of loci. A set of 89 markers that provided uniform coverage of the genome was used to scan for epistatic interactions. For each pair of markers, an *F*-statistic for a contrast of line means testing for interaction between the markers was computed. The resulting 3916 *F*-statistics were tested for genomewise significance via a permutation approach. In particular, all 3916 *F*-statistics were recomputed for each of 200 permutations of the line labels relative to their marker information. The maximum of the 3916 observed *F*-statistics was recorded for each data permutation. The significance of the original 3916 *F*-statistics was judged by comparison to the permutation distribution of maximum *F*-statistic values.

RNA extraction and DNA chip analysis: Plant material for RNA extraction was collected at three time points, one at the time of explanting root segments from 1-week-old seedlings (day 0), the second 4 days after incubation on CIM (day 4 CIM), and the third 6 days after transferring the segments to SIM (day 6 SIM). Eight LLC lines and 11 LLL lines were cultured separately, and an equal amount of tissue was pooled together for 1 g of tissue/pool for RNA extraction. RNA extraction and hybridization to Affymetrix ATH1 oligonucleotide arrays were carried out as described in Che *et al.* (2002).

Experimental design: The time course was repeated on three separate occasions. Within a given replication both LLL and LLC were cultured on the same day for a given time point. Thus, this design is equivalent to a standard split-plot design with time as the whole-plot factor and genotype as the split-plot factor (see, for example, Kuehl 2000). RNA from each combination of genotype (LLL or LLC), time point (day 0, day 4 CIM, or day 6 SIM), and experimental replication (1, 2, or 3) was hybridized to its own Affymetrix GeneChip.

However, because of RNA degradation in one replicate of the day 0 sample from the LLC pool and poor RNA quality from all day 4 CIM samples, our results are based on data from a total of 11 and not 18 oligonucleotide arrays.

Oligonucleotide array data analysis: For each of 22,810 probe sets on the oligonucleotide arrays, a mixed linear model analysis of logged and normalized Affymetrix MicroArray Suite (MAS) 5.0 signal intensities was conducted to identify transcripts whose expression differed significantly across genotypes LLL and LLC at either day 0 or day 6 SIM. The mixed linear model for each probe set included time, genotype, and time-by-genotype terms as fixed effects along with random-effect terms for replications, replication-by-time-point interactions, and a general error term. At each time point, the difference in average log-scale expression between genotypes was estimated. The standard error of each difference was determined from the fit of the mixed linear model. The ratio of each estimated difference to its standard error was used to form *t*-statistics for testing the null hypothesis of no expression difference across genotype at each time point. A *P*-value was obtained for each time point by comparing each time-specific *t*-statistic to a *t*-distribution with 3 d.f.

For each time point, a mixture of a uniform and a β-distribution was fit to the observed distribution of 22,810 *P*-values obtained from the mixed linear model analysis. The estimated parameters from the fit of the mixture model were used to estimate the posterior probability of differential expression for each gene and to estimate the total number of genes that were differentially expressed across genotypes as described by Allison *et al.* (2002). The method of Fernando *et al.* (2004) was used to estimate the proportion of false-positive results among all genes with *P*-values ≤ 0.01, 0.001, and 0.0001. In addition to *P*-value analysis, fold-change estimates and 95% confidence intervals associated with the fold-change estimates were computed as part of the mixed linear model analysis.

Data normalization: MAS 5.0 signal intensities for each oligonucleotide array were logged and mean centered so that expression measures would be comparable across slides. Mean centering simply involves subtracting the average of the log MAS 5.0 signal intensities on an oligonucleotide array from each log MAS 5.0 signal intensity on the array. Thus, negative (positive) values indicate that a particular transcript was expressed below (above) the average for a particular array. We chose this relatively simple normalization strategy rather than using robust multiarray average (Irizarry *et al.* 2003) or the model-based expression index proposed by Li and Wong (2001) because these more complex methods introduce dependencies among measures of gene expression from experimental units that are otherwise independent. The *P*-values that we computed as part of our mixed linear model analysis are valid only under the assumption that measures of expression on experimental units from different replications are independent. The estimates of the number of differentially expressed genes and the proportion of false-positive results associated with any threshold for significance require valid *P*-values whose distribution is uniform when the null hypothesis is true.

RESULTS

Recombinant inbred lines: Col and L*er* ecotypes in Arabidopsis differ in their capacity to regenerate shoots in tissue culture (Figure 1, A and B). We used a two-step procedure to induce shoot formation by preincubating root explants for 4 days on CIM and then transferring the explants to SIM for further incubation until shoots form (Valvekens *et al.* 1988). Shoot formation was quantified by scoring the number of shoots per root explant (shoot counts) at a given time following transfer to SIM, usually 15 days.

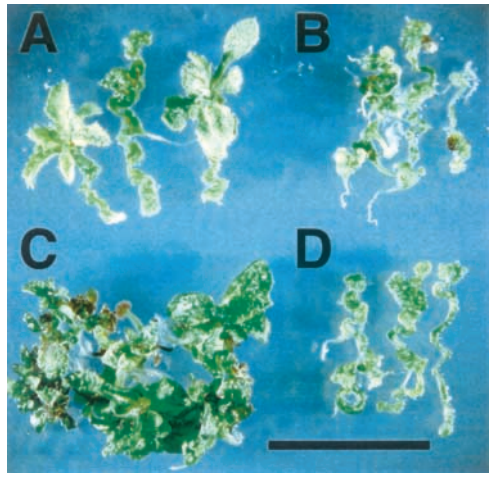

Figure 1.—Shoot regeneration from root explants of Arabidopsis seedlings. (A) Col ecotype, (B) L*er* ecotype, (C) RI line 303, and (D) RI line 370. Photographed after 15 days on SIM. Bar, 1 cm.

Explants from the Col ecotype produced 1.5 ± 0.6 shoots/explant and the L*er* ecotype produced 0.50 ± 0.22 shoots/explant. Recombinant inbred lines generated from the cross of Col and L*er* parents showed a spectrum of shoot counts that varied from lines that were more extreme than those of either parent (highest, 9.2 ± 1.2 shoots/explant; lowest, 0 shoots/explant) to others with more intermediate phenotypes. Two RI lines are shown, CL303, which exhibits more robust shoot production than the Columbia parent (Figure 1C), and CL370, which is less efficient in shoot formation than the Landsberg parent (Figure 1D). This phenomenon in which the RI lines exhibit greater variation than either parent is called transgressive segregation (Darlington and Mather 1949) and can be attributed to the complementary action of genes from the two parents (Rick and Smith 1953).

QTL analysis: Explants from 100 RI lines were scored for number of shoots and a composite interval mapping approach was used to scan the Arabidopsis genome for loci associated with shoot regeneration. Three QTL were identified with likelihood-ratio statistics that approached or exceeded a 0.05 permutation threshold for significance (Figure 2). A major QTL was located on chromosome 5 centering on the marker Cor78 at 106.7 cM. In this interval, the Col ecotype contributed the superior allele. Two minor loci were located on chromosome 1 (near marker ATTSO477 at 12.5 cM) and on chromosome 4 (near marker mi32 at 60.9 cM). The superior alleles at the minor loci were both contributed by the L*er* parent.

The RI lines were binned into eight genotype groups

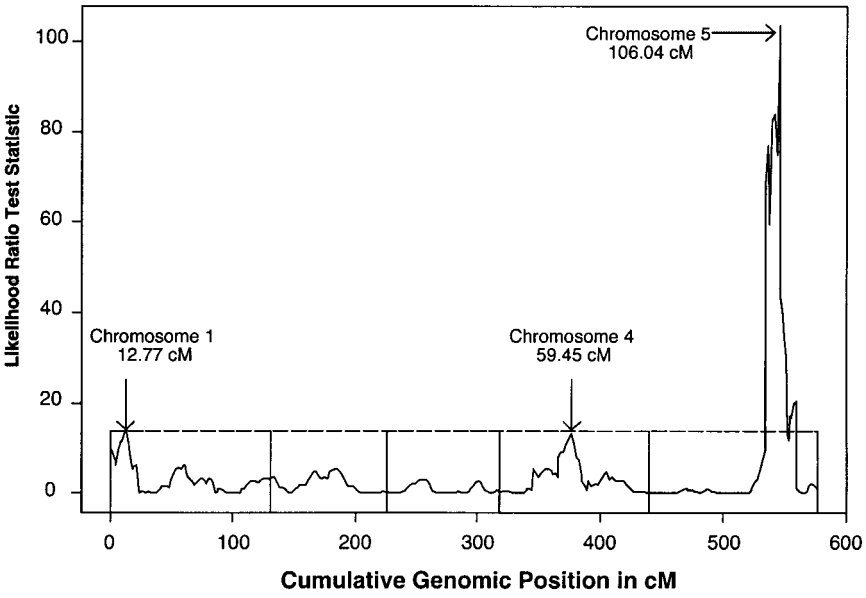

FIGURE 2.—Genome scan for loci conditioning shoot formation. Composite interval mapping was used to identify loci approaching or exceeding the permutation threshold for significance at the 0.05 level.

according to the combination of parental alleles at the three loci (sufficient marker information was available to group 98 of 100 lines). The distributions of shoot counts (on a square-root scale to reduce skewness) are shown in a boxplot (Figure 3). Genotype group identifiers, such as CLC, for example, were used to designate lines in which Columbia alleles were present at the locus on chromosome 1, Landsberg alleles on chromosome 4, and Columbia alleles on chromosome 5. In the boxplot, the effects of any QTL can be discerned by comparing the groups pairwise. For example, the strong effects of the major controlling locus on chromosome 5 can be seen by comparing CCC to CCL, CLC to CLL, LCC to LCL, and so forth. The boxplot clearly shows the transgressive segregation effect; that is, the RI lines with the parental genotype (CCC and LLL) are less extreme than the RI lines with nonparental genotype. There is also visual evidence from the boxplot of epistatic interactions between the loci. For example, it appears that the effect of the major QTL on chromosome 5 is reduced when the Col allele is present at the minor locus on chromosome 4.

The three loci identified with the composite interval mapping scan were further analyzed by examining contrasts of line means estimated via a mixed linear model with fixed effects for lines and random effects for the plates in which root segments were cultured. Table 1 shows the results for several contrasts of interest, which indicate significant differences among the 102 line means, significant differences among the eight QTL genotype groups depicted in Figure 3, significant main effects of each marker associated with the identified QTL, as well as significant interactions among these markers. By comparing the sum of squares associated

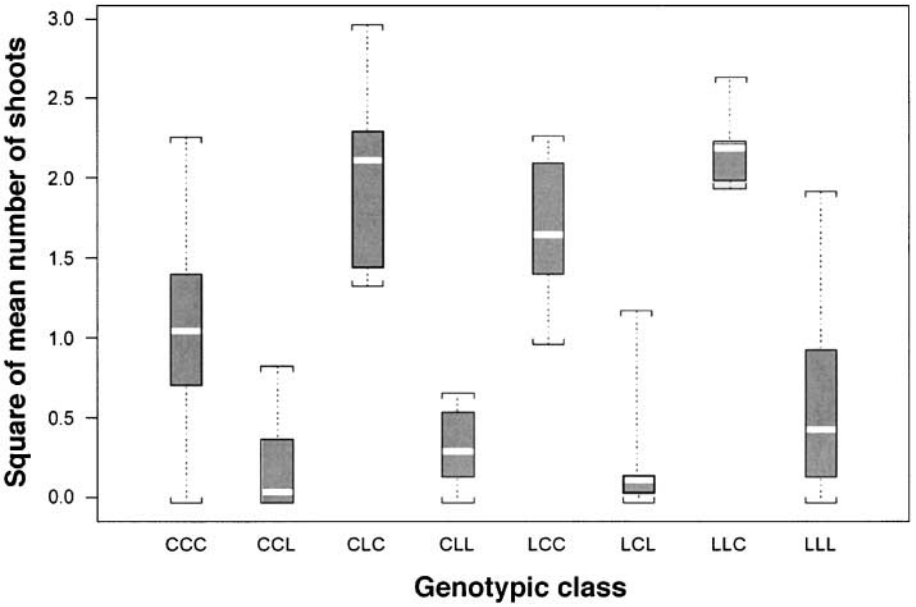

FIGURE 3.—Boxplot showing the distributions of shoot-count means (on the square-root scale) for lines falling into each of the eight QTL genotype groups defined by the QTL sites shown in Figure 2 on chromosomes 1, 4, and 5. Each boxplot (reading downward in each column) displays the maximum, 75th percentile, median, 25th percentile, and minimum for each group. (The minimum and the 25th percentile are the same for the CCL group; the median and the 75th percentile are the same for the LCL group.) The influence of the different parental alleles at each QTL locus can be seen by making various distribution comparisons among the groups. The RI lines from the genotypic groups on the right (LLC and LLL) were used for the gene expression analysis.

TABLE 1

Contrasts among line means

Contrast	Estimate	d.f.	Sum of squares	Mean squares	*F*-statistic	*P*-value
Lines		101	260.175	2.576	15.117	<0.0001
QTL groups		7	172.524	24.646	144.638	<0.0001
M1	−0.1193	1	2.160	2.160	12.676	0.0004
M4	−0.3750	1	21.056	21.056	123.569	<0.0001
M5	0.8987	1	121.493	121.493	712.990	<0.0001
M1 × M4		1	0.583	0.583	3.422	0.0649
M1 × M5		1	0.944	0.944	5.539	0.0190
M4 × M5		1	8.509	8.509	49.935	<0.0001
M1 × M4 × M5		1	13.122	13.122	77.006	<0.0001

The first row provides a test for differences among the 102 line means. The second row provides a test for differences among the eight QTL genotype groups depicted in Figure 3. The remaining rows correspond to tests for the main effects of each of the markers associated with identified QTL as well as all possible interactions. The labels M1, M4, and M5 correspond to the markers ATTSO477, mi32, and Cor78 that lie near the QTL identified on chromosomes 1, 4, and 5. The negative estimates for M1 and M4 indicate that the L*er* allele is associated with enhanced shoot production relative to the Col allele for the QTL on chromosomes 1 and 4. The positive value for M5 indicates that the Columbia allele is superior for shoot production at the QTL on chromosome 5. All analyses were done on the square-root scale, so the estimated QTL effects must be viewed with this in mind.

with the contrast for differences among the eight QTL genotype groups to the sum of squares for differences among the 102 line means, we can see that approximately 66% (172.5/260.2) of the variation in line means was explained by the three identified loci. The contrasts for the marker main effects and interactions are not orthogonal due to imbalance in the number of lines that fall into each category. The contrasts are, however, close to orthogonal so it is possible to gain some insight into the contribution of each marker main effect or interaction by comparing its sum of squares to the sum of squares for differences among QTL groups. For example, most of the sum of squares for differences among the eight QTL groups can be attributed to the large main effect of the QTL on chromosome 5 (70% ≈ 121.5/172.5).

The tests for interactions among markers in Table 1 indicate evidence of epistasis among the identified QTL. However, the genome scan for epistasis (described in MATERIALS AND METHODS) revealed no significant interactions when controlling the genomewise type I error rate at 0.05. Thus the evidence for epistasis among the identified QTL is not as strong as the evidence for QTL main effects, which (as discussed above) were significant at the 0.05 level after adjusting for the multiple testing associated with a genome scan.

Effect of the major shoot regeneration QTL on gene expression: We were interested in determining the effect of the major chromosome 5 QTL on the program of gene expression during shoot development. CHE *et al.* (2002) profiled the expression of ~8000 genes during the development of shoots from root explants in the Col ecotype. In addition, P. CHE, S. LALL and S. HOWELL (unpublished observations) compared the differences in expression of ~22,000 genes during shoot, callus, and root development in the same ecotype. To evaluate the effects of the major chromosome 5 QTL on gene expression during shoot development, RI lines were pooled in two groups, LLC and LLL, such that the major QTL (chromosome 5) was represented by Col alleles (in the LLC group) or L*er* alleles (in the LLL group) and that the minor QTL (chromosomes 1 and 4) were fixed in both groups. The genomes of 8 RI lines in the LLC group and the 11 lines in the LLL group were scanned to determine the actual allele frequency at various genome positions in the pools (Figure 4). The scan clearly shows that in both LLC and LLL groups, the L*er* allele is fixed at the minor QTL positions at the top of chromosome 1 and in the middle of chromosome 4. Likewise, Col alleles are represented at the major QTL position at the bottom of chromosome 5 in the LLC group, and L*er* alleles are represented at the same position in the LLL group. However, because each group was composed of a small number of lines, Col and L*er* alleles were not equally represented at all other unselected genome locations. For example, Col alleles were overrepresented at the bottom of chromosome 1 in the LLL group and similarly Col alleles were overrepresented at the top of chromosome 3 in the LLC group. However, these regions of the genome do not have significant effects on the efficiency differences in shoot regeneration between the two ecotypes, so allele bias in these groups may not have an impact on the gene expression program. Nonetheless, the possible influence from other regions of the genome must be borne in mind in interpreting the effects of the major chromosome 5 QTL on the gene expression program.

The RNA samples were used to make cDNAs, which were hybridized to the 22,810 probe sets on Affymetrix oligonucleotide chips, ATH1. Plant material was col-

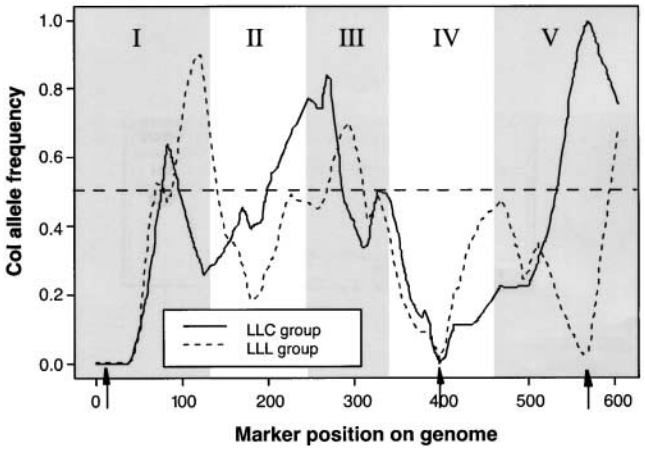

FIGURE 4.—Genome scan of the Col allele frequency in two groups of RI lines (LLC and LLL groups) used to evaluate the effect of the major shoot regeneration QTL on chromosome 5. Genome coordinates are indicated by marker positions on the *x*-axis and chromosome (1–5) locations. Peak positions for the minor QTL on chromosomes 1 and 4 and the major QTL on chromosome 5 are indicated by arrows. The dashed line parallel to the *x*-axis represents equal frequency of Col and L*er* alleles. The allele frequency curves are linear approximations derived from fitting multiple linear regressions over small sections of the *x*-axis.

lected at three time points, day 0 (time of explant), day 4 CIM (4 days after transfer to CIM), and day 6 SIM (6 days after transfer to SIM). The time points were chosen because day 4 CIM represents the time when root explants "acquire competence" to respond to shoot induction signals after transfer to SIM, and day 6 SIM represents the time of "shoot commitment," when root explants continue to form shoots after transfer to basal medium (CARY *et al.* 2002). Shoot commitment is also a time when many regulatory genes are upregulated (genes encoding transcription factors and signal components) during shoot development (CHE *et al.* 2002). Three repeats of each time course were performed, allowing us to treat the gene expression data statistically. However, tissue from day 4 CIM yielded poor RNA; therefore the data from these hybridizations were not used in the analysis.

For each gene, the expression differences between the LLC and the LLL genotype pools at the two time points were determined and *P*-values were computed (see MATERIALS AND METHODS). This analysis strategy assigns small *P*-values to genes in which signal intensities at a given time point differed between LLL and LLC genotypes in a consistent manner across replications. The distribution of *P*-values for testing for differences in expression between LLC and LLL genotypes at each of the two time points was plotted for comparison (Figure 5, A and B). Note that when a gene is not differentially expressed between genotypes at a time point, its *P*-value should be uniformly distributed on the interval 0–1. If no gene is differentially expressed at a given time point, the distribution of *P*-values will have a uniform shape. On the other hand, the presence of differentially expressed genes will lead to an excess of *P*-values near 0. This analysis shows clear evidence of many differentially expressed genes at day 6 SIM (Figure 5B) while few genes appear to be differentially expressed at day 0 (Figure 5A).

The method described by ALLISON *et al.* (2002) was used to fit a mixture of a uniform and a β-distribution to each empirical distribution of *P*-values. The best-fitting curves estimated via maximum-likelihood methods are plotted in each panel (Figure 5, A and B). From the fit of these mixture models the proportion of differentially expressed genes at each time point can be estimated along with posterior probabilities of differential expression for each gene at each time point. Just over 2% of the 22,810 genes are estimated to be differentially expressed at day 0. In contrast, ~37% of the genes appear to be differentially expressed at day 6 SIM. The estimated posterior probabilities of differential expression are all $<1/3$ at the day 0 time point. At day 6 SIM, the estimated posterior probabilities of differential expression for the top 1000 genes range from >78 to ~95%. The method of FERNANDO *et al.* (2004) was used to estimate the proportion of false-positive results (genes incorrectly declared as differentially expressed) among all genes with *P*-values <0.01, 0.001, and 0.0001 at each time point (Table 2). There are many more genes at or below each significance threshold at day 6 SIM compared to day 0. Furthermore the estimated proportion of false positives is relatively low at day 6 SIM compared to day 0.

Using 0.01 as a *P*-value threshold, we identified 845 genes as being significantly differentially expressed between the genotypes at day 6 SIM. Within this group of genes, ~16% are expected to be false positives (Table 2). Sixty-six genes among the 845 were estimated to be expressed at least fourfold higher in the LLC pool than in the LLL pool at day 6 SIM. Most of these genes also had small *P*-values for the test of interaction between time and genotype with the differences between LLC and LLL at day 6 SIM tending to be significantly greater than the differences between LLC and LLL at day 0. Table 3 contains a list of the 35 of these 66 genes that have been annotated [The Arabidopsis Information Resource (TAIR)]. It was surprising to find that 19 of those genes encode transcription factors and signaling or transposable element-related molecules. The transcription factors include a putative chloroplast-located zinc finger protein (At5g42280), a CAATT-binding factor B subunit (At1g54160), a homeodomain leucine zipper protein (ATHB-17; At2g01430), a *NO APICAL MERISTEM* (*NAM*) family member (At5g50820), two MADS-box proteins (AGL6; At2g45650 and At4g36590), and *ARABIDOPSIS RESPONSE REGULATOR18* (*ARR18*; At5g58080). Four genes were associated with the activity of transposable elements (transposases, etc.) Four encode proteases, three of which encode subtilisin-like

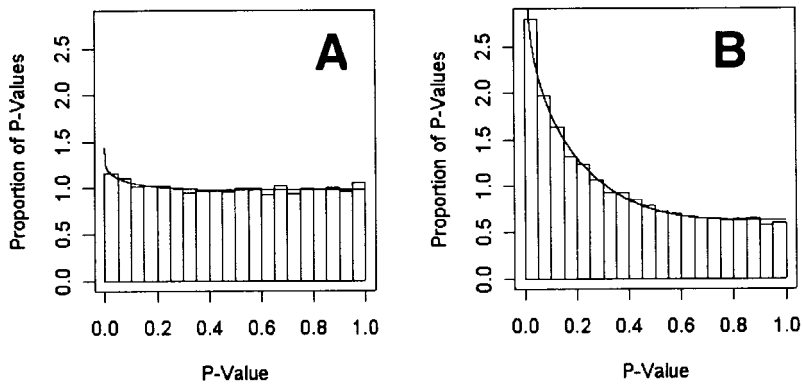

FIGURE 5.—Distributions of 22,810 *P*-values for testing for probe-set-specific expression differences between LLL and LLC at (A) day 0 and (B) day 6 SIM. The curves approximating each distribution of *P*-values were obtained using the mixture model method proposed by ALLISON *et al.* (2002).

serine proteases (At1g01900, At5g59120, and At4g26330) and one (At4g08340) encodes a protein related to desumoylating proteases (At4g08340).

The expression patterns during shoot development for some of the above genes were plotted using the data reported here and additional gene expression data collected at 4 days CIM. The gene encoding a zinc finger protein (At5g42280), for example, was expressed similarly in the LLC and LLL pools at 0 time and 4 days CIM (Figure 6A). However, at 6 days SIM the gene was expressed >13-fold higher in the LLC pool than in the LLL pool. A gene encoding a subtilisin-like serine protease (At1g01900) was also expressed similarly in the LLC and LLL pools at 0 time and 4 days CIM (Figure 6B). At 6 days SIM, the gene was expressed >10-fold higher in the LLC pool than in the LLL pool. The expression patterns of these genes in the LLC pool were similar to those in the Col parent (data not shown). Also of note was the significant differential expression of a member of the family of cytokinin oxidases (At3g63440) (Table 3). However, other data from our lab (not shown) show that this gene is highly upregulated during incubation on CIM and its expression simply falls more precipitously in genotypes in the LLL pool than in the LLC pool.

DISCUSSION

We have identified three QTL that contribute to the differences in shoot regeneration efficiency between the Col and L*er* ecotypes in Arabidopsis. The major QTL lies on chromosome 5 and the superior allele is derived from the Col parent. The other two minor QTL lie on chromosomes 1 and 4, and the superior alleles are derived from the L*er* parent. In addition, we have uncovered suggestive evidence of epistatic interaction among the identified loci. For example, the minor QTL on chromosome 4 appears to interact with the major chromosome 5 QTL in that the shoot-promoting effect of the major QTL is enhanced when the L*er* allele is present at the minor QTL locus on chromosome 4. Because the superior alleles and the enhancing epistatic effects are not found in one ecotype, the shoot regeneration phenotype in progeny shows greater extremes in shoot regeneration efficiencies than does either parent. Because of this, more robust shoot-regenerating lines than those of either parent are found in the RI lines, particularly those with an LLC genotype.

SCHIANTARELLI *et al.* (2001) conducted quantitative trait analysis on RI lines from a similar cross of Columbia × Landsberg to identify QTL associated with shoot formation in Arabidopsis. They used MAPMAKER/QTL and categorized shoot regeneration phenotypes in each line by whether they represented one parent or another or were intermediate between the two. In doing so, they found one highly significant determinant on chromosome 1 in one replicate of an experiment involving shoot regeneration from leaf explants. They found 5 other determinants that significantly contributed to the regeneration of normal or abnormal shoots from roots—two were on chromosome 5 but at a different location (position 136–139), more toward the tip of the chromosome than the major QTL that we identified. However, since their findings were reported in tabular form rather than in chromosome scans, it is difficult to say whether the chromosome locations of the determi-

TABLE 2

Estimation of the number of false positives

Significance threshold	No. of genes with *P*-values less than or equal to threshold		Estimated proportion of false positives (%)[c]	
	Day 0[a]	Day 6 SIM[b]	Day 0	Day 6 SIM
0.01	277	845	80	16
0.001	34	95	66	14
0.0001	6	9	37	15

[a] Number of genes that would be declared as significantly differentially expressed between LLL and LLC at day 0 for each of three significance thresholds.

[b] Number of genes that would be declared as significantly differentially expressed between LLL and LLC at day 6 SIM for each of three significance thresholds.

[c] Proportion of genes that are not truly differentially expressed among the genes with *P*-values less than or equal to each threshold.

TABLE 3

Annotated genes differentially expressed between LLC and LLL genotypes

Locus name	LLL *vs.* LLC at day 6 SIM *P*-value[a]	LLC/LLL[b]	Time-by-genotype interaction *P*-value[c]	Assigned function[d]
At3g21340	0.0009	21.5522	0.0017	Leucine-rich repeat protein kinase, putative
At4g04080	0.0001	21.2951	0.0003	NifU-related metallocluster assembly factor
At3g63440	0.0018	17.5570	0.0184	Cytokinin oxidase family
At2g15470	0.0049	16.5874	0.0193	Polygalacturonase, putative
At2g32540	0.0021	16.5407	0.0367	Cellulose synthase related
At5g42280	0.0070	13.3628	0.0186	CHP-rich zinc finger protein, putative
At4g38560	0.0068	13.2675	0.0064	Phospholipase like protein; pEARLI 4
At4g07700	0.0090	12.1400	0.0227	Athila transposon protein related
At1g70130	0.0050	11.0677	0.0745	Receptor lectin kinase, putative
At1g01900	0.0075	10.4057	0.0105	Subtilisin-like serine protease, putative
At1g66930	0.0060	9.4398	0.0168	Receptor-related kinase
At1g54160	0.0074	8.4222	0.0327	CCAAT-binding factor B subunit related
At5g25430	0.0029	7.6552	0.0069	Anion exchange protein family
At4g08340	0.0017	6.7000	0.0140	Ulp1 protease family
At2g01430	0.0027	6.6475	0.0074	Homeodomain-leucine zipper protein ATHB-17
At5g50820	0.0006	6.4802	0.0011	No apical meristem (NAM) protein family
At2g45650	0.0093	6.3730	0.0293	MADS-box protein (AGL6)
At1g60450	0.0023	6.0899	0.0150	Galactinol synthase, putative
At4g04270	0.0045	5.9428	0.0163	Plant transposase (Ptta/En/Spm) family
At4g17380	0.0074	5.5140	0.0357	DNA mismatch repair MutS family
At5g48485	0.0023	5.3021	0.0052	Protease inhibitor/seed storage/lipid transfer protein
At1g31530	0.0016	5.0670	0.0024	Endonuclease/exonuclease/phosphatase family
At2g43050	0.0100	4.9074	0.0406	Pectinesterase family
At5g36655	0.0014	4.8416	0.0017	Plant transposase (Ptta/En/Spm) family
At5g59120	0.0008	4.6048	0.2217	Subtilisin-like serine protease
At5g49290	0.0097	4.4301	0.0272	Leucine-rich repeat protein family
At2g21910	0.0049	4.4035	0.0048	Cytochrome P450, putative
At4g36590	0.0059	4.4013	0.0423	MADS-box protein
At1g33220	0.0005	4.3931	0.0014	Glycosyl hydrolase family 17 (β-1,3-glucanase)
At5g42720	0.0005	4.3263	0.0121	Glycosyl hydrolase family 17
At3g28380	0.0030	4.2623	0.1031	P-glycoprotein, putative;
At1g07180	0.0036	4.2277	0.0033	NADH dehydrogenase family;
At4g26330	0.0075	4.1793	0.0094	Subtilisin-like serine protease
At5g58080	0.0100	4.1689	0.0452	Response regulator ARR18
At2g15750	0.0090	4.0243	0.0441	Non-LTR retroelement reverse transcriptase related

[a] *P*-values for the contrast between LLC and LLL means at 6 days on SIM.
[b] Estimated fold change for expression in LLC relative to LLL at 6 days on SIM.
[c] *P*-values for the contrast comparing the change in expression between LLC and LLL at day 0 to the change in expression between LLC and LLL at 6 days on SIM.
[d] Functional annotation according to TAIR.

nants found in their study were significantly different from ours.

Nearly 1000 genes lie in the 20-cM interval flanking the peak position for the major QTL on chromosome 5. The set of genes in the interval is too large to speculate on candidate genes responsible for the QTL, but fine mapping of the interval is underway.

Thirty-five genes for which annotation is available were more than fourfold differentially expressed between the LLC and LLL genotypes and, of those, 19 encode transcription factors and signaling or transposon-related proteins. Among the list of differentially expressed genes were 15 genes (annotated and nonannotated) that lie in the region of the major QTL on chromosome 5 of which 8 (6 annotated; Table 3) were more than fourfold differentially expressed. Some of the transcription factor genes are members of gene families that are reported to influence shoot development. For example, At5g50820 is a putative NAC-domain-containing transcription factor related to other factors in the *NAM* gene family of transcription factors. This family derives its name from a gene required for shoot apical meristem formation in petunia (Souer *et al.* 1996). Related NAC-domain-containing factors in Arabidopsis, *CUC1* (At3g15170) and *CUC2* (At5g53950), activate the expression of shoot apical meristem genes

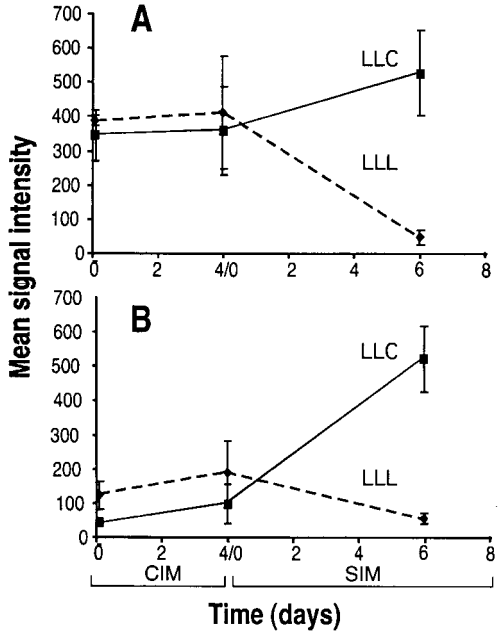

FIGURE 6.—Time courses of expression of genes with significantly different expression levels between the LLC and LLL genotypes at 6 days on SIM. Shown are the means and standard errors of the signal intensity values obtained from three repetitions of the same experiment for (A) a gene encoding a zinc finger protein, At5g42280, and (B) a subtilisin-like serine protease, At1g01900. Gene expression data at the day 4 CIM time point were not presented elsewhere in this study, because the data, in general, showed greater error variation because of difficulties in obtaining RNA samples of consistent quality at this time point. However, the day 4 CIM time point data appeared adequate to provide a better understanding of the expression patterns of these two genes.

and induce adventitious shoot formation (DAIMON *et al.* 2003; HIBARA *et al.* 2003). Two other genes (At2g45650 and At4g36590) that were differentially expressed between the two pools in our study encode MADS-box transcription factors. A MADS-box gene (PkMADS1) in *Paulownia kawakamii* has been implicated as a regulator of adventitious shoot formation from leaf explants (PRAKASH and KUMAR 2002).

Another transcription factor gene that is significantly upregulated in the LLC pool is *ARR18* (At5g58080), which encodes a B-type response regulator (RR). HWANG and SHEEN (2001) demonstrated that overexpression of *ARR2*, which also encodes a B-type RR, enhanced shoot formation in Arabidopsis in a cytokinin-dependent manner. Overexpression of *ARR11*, another B-type RR, leads to ectopic shoot formation in cotyledons and greening of callus in culture even at low cytokinin levels (IMAMURA *et al.* 2003). In addition, overexpression of *ARR1*, yet another B-type RR, also enhances shoot formation and T-DNA knockouts show decreased or no shoot formation although there is greening of callus (SAKAI *et al.* 2001).

Three other genes with significant differential expression between the two LLC and LLL genotypes encode subtilisin-like serine proteases (subtilases). Subtilases are a large gene family in Arabidopsis composed of nearly 60 members, some of which appear to have important developmental functions. Loss-of-function mutations in a gene called *ABNORMAL LEAF SHAPE1* (*ALE1*; At1g62340) encoding a subtilase affect the formation of cuticle on embryos and juvenile plants (TANAKA *et al.* 2001). The mutation has profound developmental effects because the cuticle is required for separation of the endosperm from the embryo and for prevention of organ fusion. *STOMATAL DENSITY AND DISTRIBUTION1* (*SDD1*; At1g04110) of Arabidopsis, which encodes another subtilase, is required for the control of cell lineage that leads to formation of stomatal guard cells (BERGER and ALTMANN 2000). In *sdd1* mutants, the stomatal pattern is disrupted, resulting in stomata clustering and increased stomatal density. VON GROLL *et al.* (2002) proposed that *SDD1* generates an extracellular signal by meristemoids/guard mother cells and demonstrated that the function of *SDD1* is dependent on *too many mouths* activity. Of all the subtilase-encoding genes in Arabidopsis, *SDD1* has the closest phylogenetic relatationship to At1g01900, the most highly differentially expressed subtilase gene in our study.

In conclusion, the finding in this study of a major QTL influencing shoot regeneration in Arabidopsis has prompted a search for the candidate gene(s) controlling the locus. Discovery of controlling gene(s) will be important in understanding the regulation of shoot regeneration in Arabidopsis and determining whether these genes control differences in shoot regeneration in other plant species. However, the identification in this study of genes that are differentially regulated in association with the allelic state of the major shoot regeneration QTL has pointed us to many other genes that may be key regulators in the shoot regeneration process. Expression profiling in our laboratory of other Arabidopsis mutants with defects in shoot regeneration has revealed many of the same genes found in this study (P. CHE, S. LALL and S. HOWELL, unpublished observations).

We thank Andrew Cary, Owen Hoekenga, Todd Vision, and Shyam Uttamchandani for their preliminary work on this problem. This work was supported by the National Science Foundation (IBN-0236060) and by the Plant Sciences Institute at Iowa State University.

LITERATURE CITED

ALLISON, D. B., G. L. GADBURY, M. HEO, J. R. FERNANDEZ, C.-K. LEE *et al.*, 2002 A mixture model approach for the analysis of microarray gene expression data. Comp. Stat. Data Anal. **39:** 1–20.

ANSCOMBE, F. J., 1948 The transformation of Poisson, binomial and negative-binomial data. Biometrika **35:** 246–254.

ARMSTRONG, C. L., J. ROMERO-SEVERSON and T. K. HODGES, 1992 Improved tissue culture response of an elite maize inbred through backcross breeding, and identification of chromosomal regions

important for regeneration by RFLP analysis. Theor. Appl. Genet. **84:** 755–762.

Banno, H., Y. Ikeda, Q. W. Niu and N. H. Chua, 2001 Overexpression of Arabidopsis *ESR1* induces initiation of shoot regeneration. Plant Cell **13:** 2609–2618.

Basten, C. J., B. S. Weir and Z-B. Zeng, 1999 *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC.

Berger, D., and T. Altmann, 2000 A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in Arabidopsis thaliana. Genes Dev. **14:** 1119–1131.

Birhman, R. K., G. Laublin and M. Cappadocia, 1994 Genetic control of *in vitro* shoot regeneration from leaf explants in *Solanum chacoense Bitt.* Theor. Appl. Genet. **88:** 535–540.

Buiatti, E. T., S. Baroncelli, A. Bennici, M. Pagliai and R. Tesi, 1974 Genetics of growth and differentiation *in vitro* of *Brassica oleracea var Botrytis.* II. An *in vitro* analysis of a diallel cross. Z. Pflanzenzuchtung **72:** 269–274.

Cary, A. J., P. Che and S. H. Howell, 2002 Developmental events and shoot meristem gene expression patterns during shoot development in *Arabidopsis thaliana.* Plant J. **32:** 867–877.

Che, P., D. J. Gingerich, S. Lall and S. H. Howell, 2002 Global and cytokinin-related gene expression changes during shoot development in Arabidopsis. Plant Cell **14:** 2771–2785.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

Daimon, Y., K. Takabe and M. Tasaka, 2003 The *CUP-SHAPED COTYLEDON* genes promote adventitious shoot formation on calli. Plant Cell Physiol. **44:** 113–121.

Darlington, C. D., and K. Mather, 1949 *The Elements of Genetics.* Allen & Unwin, London.

Fennell, S., N. Bohorova, M. van Ginkel, J. Crossa and D. Hoisington, 1996 Plant regeneration from immature embryos of 48 elite CIMMYT bread wheats. Theor. Appl. Genet. **92:** 163–169.

Fernando, R. L., D. Nettleton, B. R. Southey, J. C. M. Dekkers, M. F. Rothschild *et al.*, 2004 Controlling the proportion of false positives in multiple dependent tests. Genetics **166:** 611–619.

Frankenberger, E. A., P. M. Hasegawa and E. C. Tigchelaar, 1981 Diallel analysis of shoot-forming capacity among selected tomato genotypes. Z. Pflanzenphysiol. **102:** 233–242.

Gamborg, O. J., R. A. Miller and K. Ojima, 1968 Nutrient requirement of suspension cultures of soybean root cells. Exp. Cell Res. **50:** 151–158.

Hibara, K., S. Takada and M. Tasaka, 2003 *CUC1* gene activates the expression of SAM-related genes to induce adventitious shoot formation. Plant J. **36:** 687–696.

Hwang, I., and J. Sheen, 2001 Two-component circuitry in Arabidopsis cytokinin signal transduction. Nature **413:** 383–389.

Imamura, A., T. Kiba, Y. Tajima, T. Yamashino and T. Mizuno, 2003 In vivo and in vitro characterization of the *ARR11* response regulator implicated in the His-to-Asp phosphorelay signal transduction in *Arabidopsis thaliana.* Plant Cell Physiol. **44:** 122–131.

Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis *et al.*, 2003 Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:** 249–264.

Jansen, R. C., and P. Stam, 1994 High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136:** 1447–1455.

Kakimoto, T., 1996 CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. Science **274:** 982–985.

Komatsuda, T., T. Annaka and S. Oka, 1993 Genetic mapping of quantitative trait loci (QTLs) that enhance the shoot differentiation rate in *Hordeum vulgare L.* Theor. Appl. Genet. **86:** 713–720.

Koornneef, M., C. J. Hanhart and L. Martinelli, 1987 A genetic analysis of cell culture traits in tomato. Theor. Appl. Genet. **74:** 633–641.

Koornneef, M., J. Bade, C. J. Hanhart, J. Schel, W. Soppe *et al.*, 1993 Characterization and mapping of a gene controlling shoot regeneration in tomato. Plant J. **3:** 131–141.

Kuehl, R. O., 2000 *Design of Experiments: Statistical Principles of Research Design and Analysis.* Duxbury Press, Belmont, CA.

Li, C., and W. H. Wong, 2001 Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc. Natl. Acad. Sci. USA **98:** 31–36.

Lincoln, C., J. H. Britton and M. Estelle, 1990 Growth and development of the *axr1* mutants of *Arabidopsis.* Plant Cell **2:** 1071–1080.

Lister, C., and C. Dean, 1993 Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana.* Plant J. **4:** 745–750.

Lohrmann, J., and K. Harter, 2002 Plant two-component signaling systems and the role of response regulators. Plant Physiol. **128:** 363–369.

Naldoska-Orezyk, A., and S. Malepszy, 1989 *In vitro* culture of *Cucumis sativus L.* Genes controlling plant regeneration. Theor. Appl. Genet. **78:** 836–840.

Prakash, A. P., and P. P. Kumar, 2002 PkMADS1 is a novel MADS box gene regulating adventitious shoot induction and vegetative shoot development in Paulownia kawakamii. Plant J. **29:** 141–151.

Preil, W., 2003 Micropropagation of ornamental plants, pp. 115–133 in *Plant Tissue Culture: 100 Years since Gottlieb Haberlandt*, edited by M. Laimer and W. Ruecker. Springer-Verlag, Vienna.

Rick, C. M., and P. G. Smith, 1953 Novel variation in tomato species hybrids. Am. Nat. **88:** 359–373.

Sakai, H., T. Aoyama and A. Oka, 2000 Arabidopsis ARR1 and ARR2 response regulators operate as transcriptional activators. Plant J. **24:** 703–711.

Sakai, H., T. Honma, T. Aoyama, S. Sato, T. Kato *et al.*, 2001 ARR1, a transcription factor for genes immediately responsive to cytokinins. Science **294:** 1519–1521.

Sarrafi, A., J. P. Roustan, J. Fallot and G. Alibert, 1996 Genetic analysis of organogenesis in the cotyledons of zygotic embryos of sunflower (*Helianthus annuus L.*). Theor. Appl. Genet. **92:** 225–229.

Schiantarelli, E., A. de la Pena and M. Candela, 2001 Use of recombinant inbred lines (RILs) to identify, locate and map major genes and quantitative trait loci involved with *in vitro* regeneration ability in *Arabidopsis thaliana.* Theor. Appl. Genet. **102:** 335–341.

Souer, E., A. Van Houwelingen, D. Kloos, J. Mol and R. Koes, 1996 The *NO APICAL MERISTEM* gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell **85:** 159–170.

Stock, A. M., V. L. Robinson and P. N. Goudreau, 2000 Two-component signal transduction. Annu. Rev. Biochem. **69:** 183–215.

Tanaka, H., H. Onouchi, M. Kondo, I. Hara-Nishimura, M. Nishimura *et al.*, 2001 A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. Development **128:** 4681–4689.

Valvekens, D., M. Van Montagu and M. V. Lijsebettens, 1988 *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. Proc. Natl. Acad. Sci. USA **85:** 5536–5540.

Von Groll, U., D. Berger and T. Altmann, 2002 The subtilisin-like serine protease SDD1 mediates cell-to-cell signaling during Arabidopsis stomatal development. Plant Cell **14:** 1527–1539.

West, A. H., and A. M. Stock, 2001 Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem. Sci. **26:** 369–376.

Zeng, Z-B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457–1468.

Communicating editor: D. Weigel