Light-Response Quantitative Trait Loci Identified with Composite Interval and eXtreme Array Mapping in Arabidopsis thaliana

David J. Wolyn,*‡ Justin O. Borevitz,† Olivier Loudet,‡ Chris Schwartz,‡ Julin Maloof,‡ Joseph R. Ecker,† Charles C. Berry‡ and Joanne Chory‡,**

*Department of Plant Agriculture, University of Guelph, Guelph, Ontario N1G 2W1, Canada, †Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, ‡Department of Family/Preventative Medicine, University of California, San Diego, California 92093, §Section of Plant Biology, University of California, Davis, California 95616 and **Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92093

Manuscript received November 21, 2003
Accepted for publication February 11, 2004

ABSTRACT

Genetic analysis of natural variation in ecotypes of Arabidopsis thaliana can facilitate the discovery of new genes or of allelic variants of previously identified genes controlling physiological processes in plants. We mapped quantitative trait loci (QTL) for light response in recombinant inbred lines (RILs) derived from the Columbia and Kashmir accessions via two methods: composite interval mapping and eXtreme array mapping (XAM). After measuring seedling hypocotyl lengths in blue, red, far-red, and white light, and in darkness, eight QTL were identified by composite interval mapping and five localized near photoreceptor loci. Two QTL in blue light were associated with CRY1 and CRY2, two in red light were near PHYB and PHYC, and one in far-red light localized near PHYA. The RED2 and RED5 QTL were verified in segregating lines. XAM was tested for the identification of QTL in red light with pools of RILs selected for extreme phenotypes. Thousands of single feature polymorphisms detected by differential DNA hybridized to high-density oligo-nucleotide arrays were used to estimate allele frequency differences between the pools. The RED2 QTL was identified clearly; differences exceeded a threshold of significance determined by simulations. The sensitivities of XAM to population type and size and genetic models were also determined by simulation analysis.

Light is important for plant growth and morphogenesis, affecting germination, seedling development, shade avoidance, flowering, and photosynthesis. Seedling development has been studied extensively in Arabidopsis thaliana as a model for elucidating mechanisms of light signal transduction and has revealed a network of genes encoding photoreceptors, transcription factors, and other interacting proteins (Quail 2002; Kevei and Nagy 2003; Wang and Deng 2003).

Seeds germinated in light or darkness have distinct phenotypes. In darkness, seedlings are etiolated with long hypocotyls, have unexpanded cotyledons, and do not produce chlorophyll. In the presence of light, deetiolated seedlings have short hypocotyls, expanded cotyledons, and are green. Hypocotyl elongation is related inversely to light intensity and is a quantitative measure of light sensitivity for A. thaliana genotypes (Maloof et al. 2001). Mutagenesis and screens for etiolated phenotypes in various wavelengths of light have revealed major regulators of light signal transduction, including photoreceptors such as PHYA (Dehesh et al. 1993; Parks and Quail 1993), PHYB (Somers et al. 1991; Reed et al. 1993; Wester et al. 1994), and CRY1 (Ahmad and Cashmore 1993), identified in far-red, red, and blue light, respectively. Classical mutant screens, however, are limiting because redundant genes or those with small effects are difficult to identify.

Surveys of Arabidopsis accessions for hypocotyl length under varying wavelengths of light indicated continuous or quantitative variation (Maloof et al. 2001; Botto and Smith 2002), suggesting these genetic resources could be valuable for discovering new genes or allelic variants of known loci. Genetic analysis of crosses between wild accessions is likely to reveal quantitative or complex patterns of inheritance and potentially identify loci under natural selection (Mackay 2001).

Quantitative trait loci (QTL) can be detected in populations of varying structures, such as backcross, F2, and recombinant inbred lines (RILs), and through a number of statistical methods, including composite interval mapping (Zeng 1994) and multiple QTL model mapping (Jansen and Stam 1994). Borevitz et al. (2002) identified QTL under different wavelengths of light and hormone treatments in a RIL population derived from a Ler × Cvi cross. Several light-specific loci were mapped...
to genomic regions with no previously discovered signaling genes or candidate genes based on predicted function. One QTL region, *LIGHT2*, with significant effects in both red and white light, included *PHYB*. Another QTL, *HYP2*, with significant effects in blue, far-red, and dark, localized near *ERECTA*. In another study with the *Ler* × *Cvi* population, three QTL for hypocotyl elongation and three others for cotyledon unfolding were identified when seedlings were grown under hourly pulses of far-red light (Bottro *et al.* 2003). A QTL for flowering (El-Assal *et al.* 2001) colocalized with one for cotyledon unfolding (Bottro *et al.* 2003); both were cloned and identified as the *CRI2* locus. Two QTL controlling cotyledon opening under short pulses of far-red light were also identified in a RIL population derived from *Ler* and Col parents (Yanovsky *et al.* 1997).

The ability to detect significant QTL for a trait of interest is dependent upon the genetic variation between parental lines; important loci for which allelic effects are not significantly different between parents of a population will not be identified. Thus, the survey of populations originating from various parental combinations can increase the probability of detecting new, important QTL. A set of RILs derived from Col-*gl1* and Kas-1 was mapped with 26 molecular markers (Wilson *et al.* 2001) and used to identify three QTL for powdery mildew resistance. Since other Kas and Col genotypes varied significantly for hypocotyl length under different wavelengths of light (Maloof *et al.* 2001), analysis of the Col-*gl1* × Kas-1-derived RILs may be useful for the discovery of new genes important for light signal transduction.

QTL analysis is a lengthy process, often involving construction and genotyping of RILs. The development of methods for high-throughput genotyping of pools of lines, employing DNA hybridization to RNA expression GeneChips (Borevitz *et al.* 2003), could expedite QTL discovery. The successful mapping of a monogenic morphological mutation with this technology in combination with bulk segregant analysis (Borevitz *et al.* 2003) suggests that a similar approach may be useful for the identification of QTL when lines with extreme phenotypes are selected.

In this article we report the identification of QTL important for light signaling in a population derived from Col-*gl1* and Kas-1 accessions. Five of eight detected QTL localized near known photoreceptors. Two red-light QTL were verified in segregating RILs. A method of eXtreme array mapping (XAM) was developed that extends the use of array hybridization and bulk segregant mapping to quantitative traits.

**MATERIALS AND METHODS**

**Plant material:** A set of 128 RILs (F2) derived from Col-*gl1* and Kas-1 genotypes (Wilson *et al.* 2001) and parental lines were obtained from the Arabidopsis Biological Resource Center (Columbus, OH). Two plants for each RIL were grown from the original stocks and vernalized for 4 weeks to control variation in flowering time and to reduce maternal effects. F2 seed was saved separately.

**Plant growth:** Seeds were sterilized in 1.5-ml tubes, in 70% EtOH with 0.1% Triton X-100 for 5–10 min and then in 95% EtOH for 5–10 min. After washing in sterile H2O, seeds were resuspended in 0.05% phytagar and stored in darkness at 4°C overnight. Fifteen seeds for each genotype were spotted onto a 15×100-mm petri dish, as described below, containing 1/2 Murashige and Skoog salts and 0.7% phytagar. Plates for all treatments were stratified for 8 days at 4°C in darkness, exposed to white light (120 μmol·m−2·s−1) at 23°C for 4 hr, incubated overnight in darkness at 23°C, and then grown under the following continuous light treatments or a dark treatment for 7 days: blue (4 μmol·m−2·s−1), far-red (0.5 μmol·m−2·s−1), red (35 μmol·m−2·s−1), and white (10 μmol·m−2·s−1). Blue, far-red, and red treatments were conducted in Percival E30LED chambers (Percival Scientific, Boone, IA) where photosynthetic active radiation was calibrated (330–800 nm) under the appropriate LED lights. White light was provided by three 20-W cool white fluorescent and two 25-W incandescent bulbs in Percival E30B chambers; the R/FR ratio (650–680 nm/710–740 nm) was 1.0. For the dark treatment, plates were wrapped in foil and incubated in darkened Percival E30B chambers.

**Experimental design:** For each of two replicate experiments, the RILs and parents were grown on 13 petri dishes (15×100 mm). A dish was divided radially into 12 sectors and contained 10 random RILs and the two parental genotypes. The RILs and parents were randomized among the sectors on each of the 13 dishes, and the identical randomization pattern was used for all light treatments of an experiment. In the replicate experiment, the RILs were rerandomized to the dishes and again with the parents on each dish. Seed was used from an alternate source plant to control maternal effects, and light treatments were assigned to different growth chambers to control chamber effects.

**Statistical analysis:** Analyses of variance were conducted separately for parents and RILs using the Proc Mixed procedure in SAS (Statistical Analysis System, Cary, NC). Parental ecotypes or RILs were analyzed as fixed effects while all remaining sources of variation and interaction terms were considered random.

**Hypocotyl measurements:** Hypocotyls were transferred to acetate sheets and scanned on a flatbed scanner. Length was measured with Scion Image for Windows.

**Mapping analysis:** The 128 RILs were genotyped previously with 26 markers by Wilson *et al.* (2001). Twenty-nine simple sequence length polymorphism (SSLP) markers were added to the map (Figure 1) with Mapmaker 3.0 (Lander *et al.* 1987). The “group” command was used to verify linkage groups (minimum LOD 3; maximum distance 50); “compare” was used to establish order of markers on the linkage map, which was confirmed with the order of markers on the physical map; and “map” was used to estimate map distances (in centimorgans) with the Kosambi map function. Mapping data, genotypes for the RILs at each marker, can be found at http://arabidopsis.info and http://naturalvariation.org/xam.

Primers for MSAT markers added to the map can be found at http://www.inra.fr/qtlat/ while those for all remaining markers can be found at The Arabidopsis Information Resource (http://www.arabidopsis.org/), except those for the Athch2 marker (forward, GGATCCAAGTGCTGATGAC; reverse, CCTTC GTTCTCAAATAGGAGAAC). PCR was conducted for 40 cycles; each cycle included a 30-sec denaturation at 94°C, a 30-sec annealing at 50°C, and a 30-sec elongation at 72°C. The 40 cycles were preceded by a 1-min denaturation at 94°C and followed by 7-min extension at 72°C.
QTL analysis: For each light treatment, RILs with five or fewer observations in either replicate experiment were eliminated from the analysis due to poor germination. RILs on one dish, across all light treatments, were eliminated due to dramatically poorer growth than those on the other dishes in the first experiment. In addition, 18 lines identified as having genotypes different from those reported by Wilson et al. (2001) for at least one of six markers tested were eliminated from the analysis. Ultimately, 102 of the original 128 lines and 9656 individual hypocotyl measurements were included in the study.

Interval mapping was conducted with QTL Cartographer (http://statgen.ncsu.edu/qtlcart/cartographer.html) using the Zmapqtl (model 3) function to identify putative QTL. Forward/backward stepwise multiple regression was performed in SRmap to identify background markers (P = 0.05). Composite interval mapping (model 6) was then completed by choosing one marker at each significant QTL (P ≤ 0.05) to optimize LOD scores and minimize QTL interval. Mapping was conducted with a walking speed of 0.5 and a window size of 3 cM.

Thresholds for each light treatment were calculated with 5000 permutations (Churchill and Doerge 1994). LOD critical values ranged from 2.5 to 2.7 (P = 0.05) across environments. Additive effect estimates and percentages of variance explained by the QTL were generated with Eqtl, testing hypothesis 10 and using model 6 from Zmapqtl.

To test for significant genotype × environment interactions, JZmapqtl in QTL Cartographer (model 14) was used (Jiang and Zeng 1995). Background markers were included as those identified in the above analyses except a single marker was chosen for closely linked QTL found under different light conditions. Critical LOD values (P = 0.05) were estimated by 2000 permutations to be 6.2 for the G × E likelihood. The G × E likelihood at a given marker was significant when it exceeded the critical value.

Epistasis: Pairwise marker interactions were determined using BQTL (http://hacuna.ucsd.edu/bqtl/; Borevitz et al. 2002). A total of 39,621 tests were performed for 55 markers plus 226 pseudomarkers with a 2-cM walking speed for each light treatment. LOD critical values ranged from 4.2 to 4.5 for P = 0.05 across environments and were determined by 2200–5000 permutations (each permutation consisting of 39,621 tests). These numbers of permutations were determined sufficient since the 95% confidence intervals of the critical values did not include the calculated test statistics.

Confirmation of QTL: RILs 25 and 271 were identified as segregating for the markers PLS7 and MSAT5.22, corresponding to the RED2 and RED5 QTL, respectively. For each line, seed was sterilized, plated, stratified, and cultured for the red light treatment, as described above, except 300 seeds for each line were spaced 10 mm apart on 15–150-mm dishes. Seedlings were harvested, scanned on a flatbed scanner, and then frozen for DNA extraction and genotyping.

DNA sequencing: Overlapping fragments for Kas-1 alleles of the PHYA, PHYB, and PHYC genes were amplified using PCR and sequenced. For PHYB, 2.2 kb of DNA 5′ to the start codon, all introns and exons, and 100 bp 3′ to the stop codon were included in six amplified fragments. Nine fragments were amplified for each of the PHYA and PHYC genes and included 37 or 15 bp 5′ to the start codon and 173 or 147 bp 3′ to the stop codon, respectively, plus all introns and exons. Col and Kas-1 alleles were compared for mutations in protein coding sequences.

XAM: Fifteen RILs with the tallest and 15 RILs with the shortest hypocotyl lengths in red light were chosen. Single, 3-week-old leaves from each RIL were pooled within the short- and long-hypocotyl groups and genomic DNA was extracted using the Qiagen (Chatsworth, CA) DNA Easy kit. DNA was also extracted separately from two independently grown plants for each of the Col-0 and Kas-1 parents. Approximately 300 ng of genomic DNA was labeled overnight at 25°C using 3 vol of the Invitrogen (San Diego) Bioprime random labeling kit. The labeled DNA was ethanol precipitated, resuspended in 100 μl H2O, and then added to the hybridization cocktail. ATH1 GeneChips (Affymetrix) were hybridized and processed as recommended by the manufacturer for RNA probes. Six
ATH1 GeneChip arrays were used in this analysis: two were hybridized with Col-gl1 DNA, two with Kas-1 DNA, one with DNA pooled from the 15 RILs with the shortest hypocotyls in red light, and one with DNA pooled from the 15 RILs with the tallest hypocotyls in red light.

After the GeneChips were scanned, .cel files were spatially corrected (Borevitz et al. 2003) and quartile normalized (Iri- zarry et al. 2003). Modified t-tests were performed on 202,806 individual nonredundant features and linear clustering identified potential deletions (Borevitz et al. 2003). Conservatively, 8000 single feature polymorphisms (SFPs) were identified using a previously determined permutation threshold <5% false discovery rate (Borevitz et al. 2003). Potential deletions were identified by a linear clustering method (Borevitz et al. 2003) and are available (http://naturalvariation.org/xam). SFPs were scaled according to the differences in mean hybridization intensities between the parents such that 0.5 indicated homozygous Col-gl1, 0 indicated heterozygous, and −0.5 indicated homozygous Kas-1 genotypes. When all the lines of one pool had only the Col-gl1 allele at a locus and those of the other pool had only Kas-1 allele, the difference between the short and tall pool would be +1 or −1, when Col-gl1 or Kas-1 alleles were responsible for the short hypocotyl phenotype, respectively. To reduce the noise in the estimate of allele frequency difference at a given location, a loess smooth (Cleveland et al. 1992) was applied using a span of 0.25. Thus, the estimate was improved by considering the score of the neighboring feature. The genome-wide median was 13 SFPs/cM with fewer near the centromere and more at chromosome ends; thus, variance due to SFP genotyping was minimal compared to variance from the QTL and population. All data and R scripts used in analyses are available at http://naturalvariation.org/xam.

To determine significance thresholds and confidence intervals for XAM, simulations were performed. Different populations were tested, including 120 RILs, 200 F2’s, or 1000 F2’s. Phenotypes were applied according to the different genetic models (see below). RIL phenotypes were corrected for multiple measurements per line (n = 10). The extreme 10 or 30% of the lines were selected and the precise composite genotype of each simulated pool was calculated. Noise was then added according to the Kas-1/Col-gl1 SFP distribution. For each of 1000 simulations, the position and maximum allele-frequency difference between extreme pools was recorded. These simulations accounted for variation in recombination events in the populations, variation in QTL, phenotypes, and variation in chip genotyping. QTL effects were classified as “major,” explaining 50% of the variance in RIL means, or “moderate,” explaining 20% of the variance in RIL means.

A total of eight genetic models were investigated. Five models tested single QTL with different positions, effects, and dominance. Three genetic models simulated the effects of two QTL: two moderate unlinked QTL, two major QTL linked in repulsion, and two major unlinked epistatic QTL. A major QTL (RED2-like) that explains 50% variance in RIL means was simulated at 40 or 2 cM on chromosome two. In the epistatic model (8), both QTL explained 50% of the variance in RIL means.

For each model, population, and selection method, 1000 simulations were performed. The magnitude and position (in centimorgans) of the maximum difference between the pools was recorded for each chromosome. Using unlinked chromosomes one, three, and five, 95% maximum and minimum horizontal thresholds were calculated and represented the 950th maximum and minimum values for 1000 sorted simulations. The simulation was said to have failed if the maximum or minimum difference on the linked chromosome did not exceed the threshold (Table 2). For the linked chromosome (two and sometimes four), the distribution of maximum positions was an estimate of the precision of XAM. The 95% maximum interval was defined as the width of the chromosomal region that contained the maximum value in 950 of 1000 simulations. All data and analysis scripts are available at http://naturalvariation.org/xam.

RESULTS
Analysis of variance: An experiment was designed to estimate hypocotyl lengths of RILs under varying light conditions. RILs were grown in two repeated experiments and, within each experiment, subgroups of RILs were assigned to individual petri dishes. The parental lines were plated on each dish in each experiment, along with
the selected RILs, to determine and possibly control for dish-to-dish variability.

The analysis of variance for the parental Col-gl1 and Kas-1 lines indicated that the effects of replicate experiments, plates (experiments), and genotype × experiment interaction were not significant ($P > 0.05$) for hypocotyl length in all light treatments (data not shown). A line × plate (experiment) interaction was significant only in red light, suggesting that randomizing dish location twice daily within a chamber controlled random environmental error and improved experimental precision. The lack of significant genotype × experiment interaction indicated that the genotypes responded similarly in experiments conducted at different times and that data are repeatable.

For RILs, the effect of replicate experiments was not significant while those for RIL and RIL × experiment interaction were significant for all treatments. Despite this interaction, hypocotyl lengths of RILs were correlated between experiments for each light treatment ($r = 0.57$ to $0.74$; $P < 0.0001$) and RIL means over experiments were used in QTL analyses as the most precise estimates of line performance.

**Genetic variation in RILs and parents:** For each light treatment, distributions of RIL hypocotyl lengths were continuous, consistent with quantitative genetic variation (Figure 2). Mean hypocotyl length for Col-gl1 was less than that for Kas-1 in all light conditions except darkness, but differences were significant ($P < 0.05$) only in blue and white light. Transgression was considered significant if RIL hypocotyl lengths were two standard errors greater or less than that of the tall or short parent, respectively, for each light condition. For short hypocotyls in all environments and for long hypocotyls in darkness and red light, transgression was significant.

**QTL identification:** Composite interval mapping with background markers was conducted to identify QTL important for light signaling. Eight were significant among the five light treatments (Figure 3, Table 1) and were named according to the chromosomes on which

---

**Figure 3.—QTL map for light response in Col-gl1 × Kas-1 recombinant inbred lines derived from composite interval mapping.** The QTL and their effects, expressed as $R^2$, are represented by horizontal lines. The grouping of closely linked QTL from different light conditions are indicated by a named QTL. Vertical lines represent 2-LOD support intervals for individual QTL: B, blue; D, dark; FR, far-red; R, red; W, white. Where QTL are closely linked, the LOD intervals diagrammed from left to right are associated with QTL diagrammed top to bottom on the chromosome. Approximate positions of photoreceptor loci are indicated. Horizontal lines on chromosomes correspond to mapped molecular markers.
they were located and light conditions under which they were significant. BLUE1 was identified in both blue and white light and maps near CRY2, a blue-light photoreceptor (Guo et al. 1999), with activity also in white light (Mas et al. 2000). FR1 was identified near the PHYA gene; significant QTL in both white and blue light were also identified in this region. Since PHYA is known to have activity in blue (Whitelam et al. 1993; Ahmad and Cashmore 1997; Neff and Chory 1998) and white (Reed et al. 1994) light, in addition to far-red (Parks and Quail 1993), the QTL for the three light environments were grouped as the FR1 locus. The RED2 QTL was significant only in red light and was linked to PHYB. HYP2 was significant in all light conditions except red. BLUE4 was significant in both blue and white light and mapped near CRY1. QTL in red, white, and blue light were colocalized at the distal end of the long-arm of chromosome four and were collectively designated LIGHT4. RED5 was linked to the red-light photoreceptor, PHYC (Monte et al. 2003), on chromosome five. FR5 was significant only in far-red light and was not near recognizable candidate genes. All photoreceptor genes except PHYB were within 2-LOD support intervals of their respective QTL (Figure 3).

All QTL were highly significant ($P < 0.01$; Table 1). Additive effects ($2a$) for hypocotyl length ranged from 0.77 to 2.59 mm although most were $>1.0$ mm; RED2 had the greatest effect ($2a = 2.59$ mm). Percentage of variance explained by the QTL ranged from 7.5 to 48.2% for the red-light effects of LIGHT4 and RED2, respectively. Kas-1 alleles at all QTL contributed to long hypocotyls while those from Col-gl1 contributed to short hypocotyls. A significant $G \times E$ interaction was detected only for RED2, a red-light-specific QTL. Although all other QTL were unique to subsets of the environments tested, their $G \times E$ likelihood values did not exceed the threshold of significance, as determined by JZmapQTL. Two-way interactions of main effect QTL were not significant, nor were pairwise interactions of all 2-cM intervals for each light environment, as assessed by permutations.

**Confirmation of QTL:** RILs 25 and 271 segregated for markers PLS7 or MSAT 5.22, corresponding to the RED2 and RED5 QTL, respectively (Figure 4). Hypocotyl lengths for the Col-gl1 and Kas-1 homozygotes and heterozygotes at PLS7 (RED2) differed significantly; the heterozygote was intermediate to the homozygotes, indicating additivity. At RED5, the Col-gl1 and Kas-1 homozygotes for the MSAT 5.22 marker differed significantly; the hypocotyl length for the heterozygote was intermediate to but did not differ from them. The chromosome regions segregating in the two RILs included at least 10 cM on the genetic map (data not shown). Consequently, a large number of candidate genes, including PHYB and PHYC, could explain the two QTL.

**Candidate gene sequences:** Since five of eight QTL localized near photoreceptors and the light conditions

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Chromosome no.</th>
<th>Marker no. and name</th>
<th>QTL no. and name</th>
<th>cm LOD (mm) variance</th>
<th>cm LOD (mm) variance</th>
<th>cm LOD (mm) variance</th>
<th>cm LOD (mm) variance</th>
<th>cm LOD (mm) variance</th>
<th>cm LOD (mm) variance</th>
<th>cm LOD (mm) variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BLUE1</td>
<td>1 (NGA59)</td>
<td>2.6 4.88 0.92</td>
<td>1.1</td>
<td>2.6 4.88 0.92</td>
<td>1.1</td>
<td>2.6 4.88 0.92</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FR1</td>
<td>3 (NGA63)</td>
<td>10.0 6.67 0.77</td>
<td>16.4</td>
<td>10.0 6.67 0.77</td>
<td>16.4</td>
<td>10.0 6.67 0.77</td>
<td>16.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RED2</td>
<td>8 (PLS7)</td>
<td>56.6 8.31 1.20</td>
<td>25.0</td>
<td>56.6 8.31 1.20</td>
<td>25.0</td>
<td>56.6 8.31 1.20</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HYP2</td>
<td>10 (NCAG1120)</td>
<td>12.5 3.40 0.48</td>
<td>11.1</td>
<td>12.5 3.40 0.48</td>
<td>11.1</td>
<td>12.5 3.40 0.48</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BLUE4</td>
<td>11 (NCAG1107)</td>
<td>61.2 6.67 1.00</td>
<td>18.6</td>
<td>61.2 6.67 1.00</td>
<td>18.6</td>
<td>61.2 6.67 1.00</td>
<td>18.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LIGHT4</td>
<td>16 (MSAT5.22)</td>
<td>5.17 1.34 13.8</td>
<td></td>
<td>5.17 1.34 13.8</td>
<td></td>
<td>5.17 1.34 13.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RED5</td>
<td>11 (LFY)</td>
<td>88.8 6.00 0.80</td>
<td>20.2</td>
<td>88.8 6.00 0.80</td>
<td>20.2</td>
<td>88.8 6.00 0.80</td>
<td>20.2</td>
</tr>
</tbody>
</table>
under which they were significant were also appropriate for the photoreceptor, these genes are obvious candidates to explain the genetic variation. To gain insight for further investigations, PHYB and PHYC alleles of Kas-1 were sequenced and compared to those of Col because associated QTL were confirmed in segregating RILs. PHYA was also sequenced.

PHYB: The Kas-1 PHYB allele contained a 12-bp deletion corresponding to amino acids 15–18 in the Col allele (data not shown). In addition, a glutamic acid at position 19 of Col PHYB was replaced by a lysine in Kas-1.

PHYC: Sequence analysis of the PHYC coding region indicated that Kas-1 and Col PHYC differed by 10 amino acid substitutions. One was in a PAS domain, and three were in the kinase domain (Figure 5).

PHYA: Analysis of the Kas-1 PHYA allele sequence detected no mutations in the protein coding region when compared to that of the Col allele (data not shown).

XAM: To develop rapid procedures for identifying major QTL in new crosses between accessions, methods of selective genotyping using pools of lines with extreme phenotypes (Michelmore et al. 1991; Churchill et al. 1993; Tanksley 1993) were tested. In each pool, allele frequencies at loci near the QTL are expected to be skewed in opposite directions while unlinked loci are expected to have equal contributions of Col-gl1 and Kas-1 alleles. Variance due to random segregation of chromosomes will decrease as the number of plants in each pool increases.

To test XAM and to verify the red-light QTL identified in this study by composite interval mapping, RILs with extreme phenotypes in red light were selected for DNA hybridization to Arabidopsis GeneChips (ATH1). A total of 8000 SFPs were identified when Col-gl1 and Kas-1 DNA was hybridized to these chips and served as markers in this analysis. A total of 542 genes that are potentially deleted in the Kas genome were also identified. Those that map to QTL regions represent candidate genes. The genes, their locations, and annotations are provided in tabular and graphical format (http://naturalvariation.org/xam).

The differences in allele frequencies between the pools were calculated and a loess smooth was applied. The RED2 QTL was clearly identified (Figure 6). A marker at this locus explained 48% of the variance in RIL line means, thus confirming that XAM can easily identify major QTL. XAM also suggested the presence of two additional QTL at the ends of chromosomes four and five. The former may correspond to LIGHT4, which explained 7% of the variance in previous analyses. The significant QTL detected at the end of chromosome five, outside the region of the RED5 QTL, could be explained by several linked QTL on the chromosome, suggested by composite interval mapping (data not shown).

Simulations using XAM: To establish thresholds and confidence intervals and to test the robustness of XAM to other possible genetic models and mapping populations, simulations were performed for each of three mapping populations, with either 10 or 30% selection of extreme phenotypes (Table 2). Single major QTL were readily identified, independent of additive or dominant gene action, and at various chromosomal locations (models 1–3, Table 2). Nearly all simulations for these models in all populations and selection methods succeeded in identifying the single QTL (Table 2). The maximum interval, where 95% of simulated maximum values occurred, decreased as the number of recombination events increased, as with large numbers of F2’s, and/or as heterozygotes were eliminated, as when RILs were assessed.

For simple QTL models in small populations, increasing the number of selected individuals improved XAM precision because recombination events increased with...
increasing numbers of individuals. Large populations are required to precisely map moderate-effect QTL (model 4), which had tolerable failure rates in small populations but were detected always in large populations. In all cases, the maximum interval was centered on the simulated position of the QTL (see supplemental Table 3 available at http://www.genetics.org/supplemental/). As expected, XAM was not useful under a genetic model of overdominance (model 5), as pools of homozygous genotypes for one extreme phenotype were equivalent to pools of heterozygous genotypes for the other extreme phenotype.

Two-gene models required large populations and sometimes strict selection of extreme phenotypes for QTL to be identified. This was due to added variance in the genetic model that led to incorrect genotypes in the selection pools. In the case of two QTL linked in repulsion (model 7), only recombinant classes showed a phenotype; large populations were required to increase the number of recombination events enabling QTL detection. For epistasis, only 1/16 of the plants distinguished the phenotype/genotype association. In this case an increased selection intensity helped identify QTL location under the random phenotypic noise. The trends for two moderate unlinked QTL (model 6) were similar to those for single genes except that added genetic variance at the unlinked locus reduced efficiency and precision.

DISCUSSION

Eight QTL affecting hypocotyl length in one or more light conditions were identified. Five localized near photoreceptor loci and were significant at wavelengths appropriate for the photoreceptor. Two QTL, \textit{RED2} and \textit{RED5}, were confirmed in segregating RILs. \textit{RED2} was clearly identified using the rapid XAM method.

The experiment performed here using composite interval mapping was similar to that reported previously for a RIL population derived from the Ler and Cvi ecotypes (Borevitz \textit{et al.} 2002), where 12 QTL for various light and hormone treatments were identified. Only two of those loci may overlap with QTL identified here: \textit{LIGHT2} and \textit{HYP2} from the Ler \times Cvi population or \textit{RED2} and \textit{HYP2} from Col-gl1 \times Kas-1, respectively. Analysis of the Ler \times Cvi population for hypocotyl length response to continuous, or pulses of, red or far-red light or darkness indicated 5 QTL (Botto \textit{et al.} 2003); 1 colocalized with the \textit{HYP2} region identified here. The large number of unique QTL in each population suggested that new QTL could be identified upon analysis of additional crosses. With multiple QTL analyses complete, a large number of new loci and allelic variants at candidate genes have been identified for future studies. This undoubtedly will enhance our understanding of light responses in plants.

Interestingly, all Kas-1 QTL alleles contributed to long hypocotyls across all environments, while those for Col-
TABLE 2

XAM simulations for three populations, two selection intensities, and eight genetic models

<table>
<thead>
<tr>
<th>Model of effect</th>
<th>Allelic effect</th>
<th>120 RILs</th>
<th>200 F2</th>
<th>1000 F2</th>
<th>95% maximum interval (cM)</th>
<th>Chromosome no.</th>
<th>Position (cM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major Additive</td>
<td>11 5</td>
<td>16 13</td>
<td>5 6</td>
<td>Two</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Additive</td>
<td>5 3</td>
<td>10 7</td>
<td>4 4</td>
<td>Two</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Dominant</td>
<td>11 5</td>
<td>22 17</td>
<td>5 6</td>
<td>Two</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate Additive</td>
<td>40 20</td>
<td>83 51</td>
<td>15 13</td>
<td>Two</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate Overdominant</td>
<td>83 83</td>
<td>83 83</td>
<td>83 83</td>
<td>Two</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate Additive</td>
<td>55 25</td>
<td>68 58</td>
<td>15 13</td>
<td>Two</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate Additive</td>
<td>40 24</td>
<td>57 50</td>
<td>15 14</td>
<td>Four</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Additive</td>
<td>26 32</td>
<td>41 41</td>
<td>22 24</td>
<td>Two</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Additive</td>
<td>25 25</td>
<td>25 25</td>
<td>23 23</td>
<td>Two</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Additive</td>
<td>50 42</td>
<td>83 83</td>
<td>51 83</td>
<td>Two</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Epistatic</td>
<td>44 32</td>
<td>64 64</td>
<td>44 64</td>
<td>Four</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Threshold: 0.49 0.29 0.27 0.16 0.12 0.08 Unlinked

Failures of 1000 simulations

1 Major Additive 0 0 0 0 0 0 Two 40
2 Major Additive 0 0 0 0 0 0 Two 2
3 Major Additive 0 0 0 0 0 0 Two 40
4 Moderate Additive 56 14 131 66 0 0 Two 40
5 Major Overdominant 949 960 951 945 946 950 Two 40
6 Moderate Additive 96 32 143 84 0 0 Two 40
7 Moderate Additive 85 24 147 61 0 0 Four 30
8 Major Additive 176 289 570 527 14 4 Two 40
7 Major Additive 206 362 641 580 17 21 Two 60
8 Major Epistatic 208 121 597 707 65 222 Two 40
8 Major Epistatic 163 89 580 705 48 191 Four 30

(Top) Maximum allele frequency was recorded on the unlinked chromosome for 1000 simulations and 95% of the simulations identified locations that are within the shown interval. This demonstrates the precision of eXtreme array mapping with different models. (Bottom) The failures in 1000 simulations represent the number of simulations from which the maximum or minimum value on the linked chromosome did not exceed the upper and lower 95% thresholds identified from maximum or minimum values on unlinked chromosomes.

**gl1** contributed to short hypocotyls. These data would not predict transgression in the RIL population; however, it was observed for short hypocotyls in all environments and long hypocotyls in two environments. The discrepancy could be explained by undetected minor QTL in the population or by random environmental variation in the RILs. The exceedingly long hypocotyls of Kas-1 compared to those of other ecotypes (Maloof et al. 2001) and the limited transgressive segregation in this experiment are consistent with allelic effects at all significant QTL in this ecotype contributing to long hypocotyls. For the Ler × Cvi population, alleles with positive and negative effects on hypocotyl length were identified in each parent (Borevitz et al. 2002).

The Kas-1 allele of RED2 was insensitive to light and acted additively; PHYB/phyB (null) heterozygotes also showed incomplete dominance (Koornneef et al. 1980), depending upon light intensity. The Cvi allele of LIGHT2 (also near PHYB), however, was insensitive to light but dominant to the Ler allele. RED2 could be PHYB because the Kas-1 allele has a 12-bp deletion and an amino acid substitution in comparison to Col PHYB. The deletion eliminates a four-amino-acid repeat and is present at high frequency among *A. thaliana* accessions (J. N. Maloof, unpublished results). Four observations suggest, however, that RED2 may not be PHYB. Activity for PHYB is observed in transgenic plants with the allele containing the deletion (J. N. Maloof, unpublished results). RED2 was detected only in red light, whereas PHYB is expected to be responsive in both red and white light, as was observed for LIGHT2 in the Ler × Cvi population (Borevitz et al. 2002). The QTL peak for RED2 is centered at molecular marker PLS 7. PHYB, used as an SSLP marker in this analysis, is located at the limits of the RED2 QTL peak (data not shown), 1.7 Mb from PLS 7 on the physical map, and is not included in the 2-LOD support interval of the QTL. Fine mapping is required to determine if RED2 is allelic to LIGHT2 or PHYB.

The RED5 QTL was confirmed in a segregating RIL and localized near PHYC. Recently, PHYC knock-out lines have been shown to have longer hypocotyls than wild type in red light (Monte et al. 2003). Heterozygous
phyC mutants also showed partial dominance, consistent with observations for RED5. These data, in combination with the point mutations leading to amino acid substitutions in the Ks-1 allele of PHYC, make this gene a strong candidate for the QTL.

Five of the eight identified QTL localize near photoreceptors and those associations were appropriate for the light conditions studied: BLUE1/CRY2, BLUE4/CRY1, FR1/PHYA, RED2/PHYB, and RED5/PHYC. The 2-LOD support intervals at these loci, however, were large. No direct evidence suggested that the QTL could be explained by genetic variation at photoreceptor loci; therefore, other genes linked to these loci could be the QTL. If many of these QTL are actually photoreceptor loci, one could hypothesize that point mutations in genes at the beginning of a signal transduction pathway (e.g., those for photoreceptors) may explain much of the natural variation in light response.

In this article we developed a rapid method to quickly identify QTL in new crosses. XAM can estimate the allele frequency differences between pools of lines selected for extreme phenotypes by hybridization of total genomic DNA to GeneChips. We confirmed the method by identifying the RED2 QTL, a major QTL detected in this study by traditional QTL mapping. XAM is an attractive alternative to studying allelic variation in new crosses, thus avoiding the tedious genotyping of individuals. Several crosses can be tested quickly to determine if new or previously identified QTL are segregating. XAM can also be used to confirm associations identified through linkage disequilibrium mapping in crosses between accessions with functionally different haplotypes.

Simulations indicated that complex genetic models require high selection intensities in large populations for QTL identification by XAM. These simulations provide a tool to help determine the population type, the number of plants, and the extreme selection intensity required under other, perhaps more complex, genetic models. Finally, array hybridization on parental lines is useful to identify SFPs and potential deletions in QTL regions that suggest candidate genes.

Future studies aim to extend XAM to fine mapping. Many recombinants must first be identified across the QTL region, probably with PCR-based markers. Progeny are then scored for the QTL phenotype and pooled accordingly. With nonrecombinant chromosomes removed, XAM can precisely predict the actual sites of recombination because SFP markers are abundant across the QTL interval.

The synthesis and genotyping of several RIL populations as resources for the Arabidopsis community have led to QTL identification for a broad spectrum of traits (ALONSO-BLANCO et al. 1998; WILSON et al. 2001; KOBAYASHI and KOYAMA 2002; WEINIG et al. 2002; LOUDET et al. 2003a,b) and allowed for important studies of pleiotropy when many traits are studied in the same population. When new allelic variation is of interest for a specific trait, XAM offers a time-saving and cost-effective method to discover, and ultimately clone, new QTL, advancing our understanding of the natural genetic control of plant processes.

These studies were supported by a grant to J.C. (GM52413) from the National Institutes of Health (NIH). O.L. is supported by Institut National de la Recherche Agronomique and funds from NIH (GM62992) and the Human Frontier Science Program to Detlef Weigel. J.O.B. acknowledges the generous support of the Helen Hay Whitney Foundation, and C.S. is supported by National Research Service Award, F32 GM65032. J.C. is an Investigator of the Howard Hughes Medical Institute.

LITERATURE CITED


Communicating editor: B. Bartel