DOI: 10.1534/genetics.107.083899

Quantitative Trait Loci Mapping in Five New Large Recombinant Inbred Line Populations of *Arabidopsis thaliana* Genotyped With Consensus Single-Nucleotide Polymorphism Markers

Matthieu Simon,* Olivier Loudet,* Stéphanie Durand,* Aurélie Bérard,† Dominique Brunel,† François-Xavier Sennesal,* Mylène Durand-Tardif,* Georges Pelletier* and Christine Camilleri*,[1]

**Station de Génétique et d'Amélioration des Plantes UR254, INRA, F-78000 Versailles, France and* †*Etude du Polymorphisme des Génomes Végétaux UR1279, INRA, F-91000 Evry, France*

Manuscript received October 29, 2007
Accepted for publication January 30, 2008

ABSTRACT

Quantitative approaches conducted in a single mapping population are limited by the extent of genetic variation distinguishing the parental genotypes. To overcome this limitation and allow a more complete dissection of the genetic architecture of complex traits, we built an integrated set of 15 new large *Arabidopsis thaliana* recombinant inbred line (RIL) populations optimized for quantitative trait loci (QTL) mapping, having Columbia as a common parent crossed to distant accessions. Here we present 5 of these populations that were validated by investigating three traits: flowering time, rosette size, and seed production as an estimate of fitness. The large number of RILs in each population (between 319 and 377 lines) and the high density of evenly spaced genetic markers scored ensure high power and precision in QTL mapping even under a minimal phenotyping framework. Moreover, the use of common markers across the different maps allows a direct comparison of the QTL detected within the different RIL sets. In addition, we show that following a selective phenotyping strategy by performing QTL analyses on genotypically chosen subsets of 164 RILs (core populations) does not impair the power of detection of QTL with phenotypic contributions >7%.

UNDERSTANDING the genetic networks underlying agronomic trait variation will provide new targets for plant breeders. However, as they are generally under the control of many genes, those characters are quantitatively variable and their study requires specific strategies and techniques. In the model plant species *Arabidopsis thaliana*, studies are now being performed exploiting natural variation as a powerful alternative to classical mutant genetics (Koornneef *et al.* 2004), in particular to identify genes underlying important quantitative trait variation. The major outputs of plant genomics will depend on the development and release of common resources and tools, such as those necessary to help in the identification and the cloning of quantitative trait loci (QTL), a challenging objective to dissect the genetic architecture of complex traits. The use of recombinant inbred lines (RILs) for this purpose is very powerful as each line is nearly homozygous and then can be propagated as genetically identical individuals, allowing genotyping and phenotyping of many traits under various environmental conditions to be performed on the same material.

As the classically used accessions (Col-0, L*er*, Ws) represent only a very limited amount of the variation present in the species, the genetic bases for crossing need to be extended. Although more sampling is required from specific geographic regions, large collections of Arabidopsis accessions have now been obtained from most of the species distribution range and their diversity has been surveyed (Alonso-Blanco and Koornneef 2000). The generation of RIL populations from exotic accessions will allow us to reveal more variation, as was already shown due to frequently used RIL sets generated using parents such as Cvi-0 (Alonso-Blanco *et al.* 1998b) or Shahdara (Loudet *et al.* 2002). RIL sets are currently produced by different groups from crosses between a variety of accessions (listed at http://www.inra.fr/vast/RILs.htm), representing an invaluable resource for the community (Weigel and Nordborg 2005), especially when those accessions show a wide range of genetic backgrounds (Tonsor *et al.* 2005). RILs obtained from divergent parental accessions have already led to the molecular identification of QTL for a number of important complex traits (Koornneef *et al.* 2004).

The accuracy of QTL mapping relies to some extent on the density of the genetic maps and then requires high numbers of genetic markers. Single-nucleotide polymorphisms (SNPs), which are suitable for high-throughput genotyping methods, turn out to be markers of choice to extensively map large sets of individuals. In

[1]*Corresponding author:* Station de Génétique et d'Amélioration des Plantes UR254, INRA, Rte. de St. Cyr, F-78000 Versailles, France. E-mail: camiller@versailles.inra.fr

Arabidopsis, the availability of the whole genomic sequence of the accession Columbia, followed by the sequencing of thousands of fragments located throughout the genome in many other accessions (Nordborg *et al.* 2005), led to the identification of numerous genome-wide SNP sets. The use of such markers allows the construction of consensus genetic maps between different RIL populations. These maps, which rely on common markers anchored to the reference genomic sequence, hence permit a better comparison of the localization of the QTL mapped with different RIL sets.

Another crucial parameter for QTL mapping is the number of RILs that can be studied, which is usually constrained by two factors: the size of available RIL sets and the extent of the phenotyping effort that can be provided. Overall, the more RILs the better (Charcosset and Gallais 1996; Borevitz and Chory 2004), especially when the genetic architecture of the trait becomes more complex and smaller-effect QTL need to be detected. Equally important is the information that—for a given phenotyping investment—studying more RILs is always more powerful than performing more repetitions (Keurentjes *et al.* 2006). In a context where phenotyping remains much more limiting than genotyping, it seems appropriate to generate large RIL populations and adjust the phenotyping framework to the desired level by first limiting the number of repetitions and, only if necessary, the number of RILs observed, for instance, following a selective phenotyping strategy to try to keep the most informative individuals (Xu *et al.* 2005). Of course, the efficiency of this strategy depends on the genetic architecture of the trait, especially the contribution of individual loci.

The aims of this work were (1) to create a powerful permanent resource to facilitate the identification of QTL for a variety of traits, by generating a new series of large RIL populations with a dense consensus genetic map, (2) to validate this resource by mapping QTL for three complex traits (flowering time, rosette size, and seed production) in five RIL sets, and (3) to compare QTL detection in the entire populations and in reduced subsets of RILs (core populations). This study emphasizes the interest of extensively exploiting natural diversity as a source of new alleles for the analysis of complex pathways. Indeed, the excess of rare polymorphisms found when sequencing numerous Arabidopsis accessions (McKhann *et al.* 2004; Nordborg *et al.* 2005) motivated the development of new RIL sets to allow a more complete survey of the diversity present in the species as many of the allelic variations of importance can be found only in a single accession (Clark *et al.* 2007).

MATERIALS AND METHODS

Generation of the RIL populations: The parental accessions were originally obtained from the Nottingham Arabidopsis Stock Center: Blh-1 (N1030), Bur-0 (N1028), Col-0 (N1092), Ct-1 (N1094), Cvi-0 (N902), and Shahdara (Sha, N929). A single plant from each accession was used for crossing after two successive generations of selfing. Five crosses were performed with Blh-1, Bur-0, Ct-1, Cvi-0, and Sha as the female parents and the reference accession Col-0 as the male parent. F_1 plants resulting from each cross were confirmed to be heterozygous with two microsatellite markers showing polymorphism between the parents, and one F_1 plant from each cross was selfed. For each cross, ~500 F_2 seeds were sown individually and each F_2 plant was allowed to self. Four additional cycles of single-seed descent (SSD) were performed to obtain F_7 seeds; to randomly choose each plant to be self-fertilized, 15 seeds were sown per line at each SSD cycle and the pots were then thinned in an unbiased manner to a single plant. The final RIL sets are described and can be ordered at http://dbsgap.versailles.inra.fr/vnat.

Genotyping: For each line, genomic DNA was prepared from a bulk of ~50 F_7 seedlings representing the genotype of the corresponding F_6 plant. The seedlings were grown *in vitro* in 1 ml sterile water in small petri dishes in a culture chamber. After 9 days they were harvested in 1-ml 96-well plates containing metal beads, lyophilized, ground in a vibrator, and then suspended in 200 μl of extraction SDS buffer (200 mm Tris-HCl pH 7.5, 250 mm NaCl, 25 mm EDTA, 0.5% SDS). After centrifugation, the supernatant was precipitated with isopropanol and the pellet was washed in 75% ethanol and resuspended in 100 μl sterile water.

SNP markers were selected that were evenly distributed along the chromosomes and as much as possible polymorphic between Col-0 on the one hand and all (or most) of the female parents on the other hand, *i.e.*, SNPs for which Col-0 has a rare allele, if not a singleton. Most of them were chosen from the Nordborg laboratory's public sequencing data (Nordborg *et al.* 2005) for the accessions Bur-0, Ct-1, Cvi-0, and Sha and were subsequently checked in Blh-1 for which data were not available. When no appropriate SNP was found in the database, we sequenced DNA fragments amplified on the parental accessions at the desired position in the genome to identify suitable SNPs. Ninety-five SNP markers were multiplexed in two sets of 48 and 47 markers, respectively, and genotyped using the SNPlex technology (Applied Biosystems, Foster City, CA) according to the supplier's protocols. Three markers with a too high proportion of missing data were discarded. Two additional SNPs were genotyped with either the TaqMan (Applied Biosystems) or the Amplifluor (Serological Corporation) technologies. The 94 SNP markers finally used are listed in supplemental Table A. To avoid large gaps on the maps, some markers showing no polymorphism in one cross were replaced for the corresponding RIL population by microsatellite or indel PCR-based markers (described in supplemental Table B). The microsatellite and indel markers were amplified by PCR and the length polymorphisms were revealed by agarose gel electrophoresis as described by Loudet *et al.* (2002). The physical positions of the markers are from the TAIR 7.0 genome sequence (April 23, 2007, http://www.arabidopsis.org).

Genetic mapping and analysis of transchromosomal linkage disequilibrium: The genetic map was established using MAPMAKER 3.0 (Lander *et al.* 1987); marker distances were estimated with the Kosambi mapping function. The significance of potential segregation distortion of the parental alleles was tested for each marker by a chi-square test. Pairwise linkage disequilibrium (LD) between markers across the genome was detected according to the *GGT32* (2006 edition) LD–heatplot function using "$-10\log(p)$" as the LD estimate (http://www.dpw.wau.nl/pv/pub/ggt/).

Plant growth conditions and phenotyping: Seeds were sown and plants were grown individually in 7 × 7-cm^2 pots in a

greenhouse. Temperatures were 20° during the day and 15° during the night. Long-day growing conditions were maintained (16-hr day and 8-hr night) with a complement of artificial light (105 μE/m²/sec) when necessary. Three traits (flowering time, rosette diameter, and total seed production) were measured on each individual F_7 line of the five populations in a single large experiment after 3 weeks of seed cold treatment at 4°. Nordborg and Bergelson (1999) found that a long seed cold treatment accelerates the germination and decreases the time until flowering in most of the accessions they studied, and such a treatment must have a vernalization effect. All the lines were grown together in the same greenhouse between September and December 2005 with all RILs of each family grouped in a block with no repetition. Three repeats of each parental accession were randomly placed among the lines of the corresponding populations. To avoid border effects, the display was entirely surrounded by Col-0 plants that were not analyzed. Flowering of each plant was checked every day and the number of days between the end of the seed cold treatment and the opening of the first flower was used as an estimate of flowering time. Rosette diameter was measured the day the plant flowered. Watering was arbitrarily suspended 20 days after flowering for each plant, and the whole-seed production was collected when dry and weighted to estimate fitness (seed weight is an appropriate measure of fitness for a selfing species like *A. thaliana*). To avoid losing seeds due to dehiscence of siliques, plants were harvested in several stages during the silique ripening.

Statistical analysis and QTL mapping: QTL analyses were performed using the Unix version of QTL CARTOGRAPHER 1.14 (Basten *et al.* 2000), using essentially standard methods for interval mapping (IM) and composite-interval mapping (CIM) as described by Loudet *et al.* (2003). First, IM (Lander and Botstein 1989) was carried out to determine putative QTL involved in the variation of the trait, and then CIM model 6 of QTL CARTOGRAPHER was performed on the same data: the closest marker to each local LOD score peak (putative QTL) was used as a cofactor to control the genetic background while testing at another genomic position. When a cofactor was also a flanking marker of the tested region, it was excluded from the model. The number of cofactors involved in our models varied between 1 and 5. The walking speed chosen for QTL analyses was 0.1 cM. The global LOD significance threshold (2.2 LOD) was estimated from several permutation test analyses, as suggested by Churchill and Doerge (1994). QTL colocalization was considered only when different QTL peaked in a window of ≤3 cM, which was *a priori* chosen because it represents a very conservative support interval. Additive effects ("2*a*") of detected QTL were estimated from CIM results, as representing the mean effect of the replacement of the non-Col alleles by Col alleles at the locus. The contribution of each identified QTL to the total phenotypic variation (R^2) was estimated by variance component analysis, using phenotypic values for each RIL. The model used the genotype at the closest marker to the corresponding detected QTL as random factors in ANOVA, run using the *aov*() function of the S-PLUS version 3.4 statistical package (Statistical Sciences, Seattle). Only homozygous genotypes were included in the ANOVA analysis. In addition, QTL × QTL interactions, *i.e.*, pairwise epistatic relationships between significant QTL, were searched for in the ANOVA analysis via the corresponding marker × marker interactions, and their contribution to the total phenotypic variation (R^2) was estimated with the ANOVA model including all significant additive and digenic epistatic effects. The threshold used to evaluate the significance of epistatic interactions was $P < 0.01$.

RESULTS

Generation of the RIL populations: We built five large RIL populations originating from crosses between the accession Columbia-0 (Col-0) as the male parent and five distant accessions (McKhann *et al.* 2004; Ostrowski *et al.* 2006) as the females. The accessions crossed to Col-0 were rationally chosen from a core collection that was previously defined to maximize the genetic and phenotypic diversity in a reduced number of accessions (McKhann *et al.* 2004). These RIL populations, Blh-1 × Col-0, Bur-0 × Col-0, Ct-1 × Col-0, Cvi-0 × Col-0, and Sha × Col-0 (hereafter referred to as BlhCol, BurCol, CtCol, CviCol, and ShaCol), consist of 319, 347, 377, 367, and 349 genotyped lines, respectively.

Genotyping: We developed a set of 95 consensus SNP markers evenly distributed throughout the genome, separated by an average distance of 1.25 Mb, that we used to individually genotype the RILs at the F_6 generation. Although these markers were chosen to be polymorphic in most of the crosses, some were not informative in all crosses (Table 1), and thus the final number of useful markers depends on the RIL population. When a SNP marker showed no polymorphism in one cross, leading to a large gap, it was replaced by a PCR-based marker (microsatellite or indel). Depending on the population, the final maps total 75–90 markers, including 52 anchoring markers scored in the five RIL sets and 28 scored in four of them (Table 1). The overall percentage of heterozygous loci for each RIL set is displayed in Table 2. It ranges from 2.9 to 3.7%, which is very close to the theoretical value of 3.1% expected at the F_6 generation. The percentage of heterozygosity was also close to the theoretical expectation when checked marker by marker (data not shown). The level of missing data is very low: 0.1–0.5% depending on the population (Table 2). The average frequencies of parental alleles at the population level are globally close to the Mendelian 1:1 ratio expected for RILs (Table 2). The segregation of parental alleles at each marker locus was also examined: in each population we found regions of the genome with a significant segregation distortion at $P < 0.01$, *i.e.*, regions where the observed genotypic frequencies departed from the 1:1 ratio predicted if no selection bias occurred during the generation of the RILs (Table 1). The highest ratios were found in the CviCol set, with values up to 2.3:1 and 2.4:1, respectively, at markers c3_02968 and c5_02900. Some of the distorted regions appeared cross-specific, while others were found in two or more crosses (Table 1).

Linkage disequilibrium between physically unlinked loci: In two populations (CviCol and ShaCol), LD analyses revealed significant LD between markers on different chromosomes. Two-dimensional LD heat plots from these populations are presented in Figure 1. In the CviCol population, we found two pairs of loci in significant LD, the first one linking a locus at ~27 Mb

TABLE 1

Consensus genetic map of the five RIL populations

Marker	Type	Physical position (Mb)	Genetic position (cM)				
			BlhCol	BurCol	CtCol	CviCol	ShaCol
c1_00593		0.6	0.0	0.0	0.0 (−)	0.0 (−)	0.0
c1_02212		2.2		5.4 (+)	7.3	5.8 (−)	4.5
c1_02992		3.0	8.1	6.3 (+)	10.2	10.7 (−)	7.2
c1_04176		4.2	10.7	8.3 (+)	15.4	17.7 (−)	11.7
c1_05593		5.6	13.4	12.4 (+)	21.8	24.4 (−)	15.2
c1_08385		8.4	25.7	17.8 (+)	31.6	37.3	27.9
c1_09621	M	9.6	29.2				
c1_09782		9.8		23.7 (+)	36.7	41.6	31.9
c1_11160		11.2		28.9 (+)	45.0	48.9	
c1_11723	M	11.7	39.3				
c1_12295		12.3		34.6 (+)	48.1	54.2	43.2
c1_13869		13.9	50.1	42.0 (+)	55.4	64.5	49.1
c1_13926		13.9				64.5	
c1_15634		15.6	50.3	42.2 (+)		65.4	49.4
c1_15927	M	15.9	50.5				
c1_16875		16.9	55.5 (+)	50.8 (+)	61.3	71.7	51.2
c1_18433		18.4	63.1 (+)	62.5	69.2	82.0	59.3
c1_19478		19.5			72.0	85.6	61.5 (−)
c1_20384		20.4	69.2 (+)	68.0	73.9	89.0	65.7 (−)
c1_22181		22.2	72.9 (+)	70.5 (−)		92.3	70.3 (−)
c1_23381		23.4	77.5 (+)	74.3 (−)	82.9	97.0	73.2 (−)
c1_24795		24.8	83.7 (+)	81.8 (−)	89.3	105.9	79.7 (−)
c1_25698		25.7	88.7 (+)	84.6 (−)	92.9	111.7	83.1 (−)
c1_26993		27.0		90.3	102.2	122.6 (−)	87.7
c1_28454		28.5	97.0 (+)	94.4	104.9	129.4 (−)	91.5
c1_28667		28.7		94.7	105.6	130.4	
c1_29898		29.9	104.0 (+)	99.4	112.7	137.9	96.4
c2_00593		0.6	0.0	0.0	0.0	0.0	0.0
c2_02365		2.4	11.9	5.2 (+)	7.6	13.8	7.2 (+)
c2_03041		3.0			7.9	16.9	7.7 (+)
c2_04263		4.3	12.6	5.6 (+)		17.7	7.7 (+)
c2_05588		5.6		6.6 (+)	9.8		8.3 (+)
c2_06280	I	6.3				25.9	
c2_06655		6.7	17.4	9.3 (+)	15.0		12.2 (+)
c2_07650		7.7	23.3	13.0	22.4	34.6	19.3 (+)
c2_10250		10.3	33.6	24.2	37.9	44.4	30.7 (+)
c2_11457		11.5	37.1	28.5	42.2	48.8	35.1 (+)
c2_12435		12.4	43.6	33.7	47.7	54.7	41.5 (+)
c2_13472		13.5	47.7	38.3		61.6	45.0 (+)
c2_15252		15.3	52.3	43.8	60.9	68.5	48.5 (+)
c2_16837		16.8	58.5	49.7	65.5	77.2	54.4 (+)
c2_17606		17.6		51.2	66.9	79.9	56.7 (+)
c2_18753		18.8	66.7	55.8	70.8	86.7	61.4 (+)
c3_00580		0.6			0.0	0.0 (+)	0.0 (−)
c3_00885		0.9	0.0	0.0		1.1 (+)	0.9 (−)
c3_01901		1.9	3.0	3.7		4.5 (+)	3.8 (−)
c3_02133	M	2.1			6.0		
c3_02968		3.0	6.8	5.7		8.3 (+)	5.9 (−)
c3_04141		4.1	10.4	8.0	11.8	14.6 (+)	5.9 (−)
c3_05141		5.1	13.8	11.1	15.8	18.7 (+)	5.9 (−)
c3_06631		6.6		14.5	19.6	24.0 (+)	9.4 (−)
c3_07673	M	7.7	21.9				
c3_08042		8.0		19.3	25.9	30.2 (+)	12.9 (−)
c3_09056	M	9.1	32.4				
c3_09748		9.7		28.7	35.0	39.9 (+)	24.7
c3_10996		11.0		34.2		48.4 (+)	

(*continued*)

TABLE 1

(Continued)

Marker	Type	Physical position (Mb)	Genetic position (cM)				
			BlhCol	BurCol	CtCol	CviCol	ShaCol
c3_11192		11.2				49.8 (+)	35.5
c3_11208	M	11.2	43.0				
c3_12647		12.6	46.4 (+)	40.9	48.6	55.7	38.7
c3_14097		14.1	46.6 (+)	41.2	48.6		38.7
c3_15117		15.1	47.7 (+)	42.0	48.9	57.6	
c3_15714	M	15.7					41.0
c3_16677		16.7		50.0	57.3 (−)	68.4	
c3_17283	M	17.3					48.9
c3_17651	M	17.7	55.1				
c3_18180		18.2		54.2	63.6 (−)	75.6	51.8
c3_20729		20.7	67.5	63.4	70.9 (−)	85.9	63.4 (+)
c3_22147		22.1	70.2	65.5	74.2	90.3	67.7 (+)
c4_00012		0.0	0.0			0.0	0.0
c4_00269	I	0.3			0.0		
c4_00641		0.6		0.0	4.1	6.3	7.0
c4_02133		2.1	11.3	6.9	12.4	16.2	15.6
c4_03833		3.8	11.5	7.0	12.5	16.4	15.6
c4_04877		4.9	11.7	7.3	12.5	18.9	15.9
c4_05629	M	5.6		13.3			
c4_05850		5.9	19.2		20.1	28.4	22.2
c4_06923		6.9	26.7	22.6	28.1	35.8	27.7
c4_07549	M	7.5	31.3	25.6	31.3	39.3	30.8
c4_07740		7.7		26.4	31.6	39.8	
c4_08930		8.9	38.6	30.9 (+)	36.7	43.7	35.5
c4_10609		10.6		35.7 (+)	46.0	51.1	42.4
c4_11878		11.9	51.6	38.7 (+)	52.2	56.6	47.9
c4_13171		13.2	56.1 (+)	40.9 (+)	56.8	62.5	55.1 (−)
c4_14819		14.8	60.9	44.2 (+)	62.6	68.6	59.4
c4_15765		15.8		46.4 (+)	64.9	71.9	60.8
c4_17684		17.7	68.8	52.7 (+)	73.7	80.9	
c4_18056		18.1					68.3
c5_00576		0.6	0.0	0.0 (−)	0.0 (−)	0.0 (+)	0.0
c5_01587		1.6	7.0	4.6 (−)	4.5	5.4 (+)	6.4
c5_02900		2.9	11.0	6.1 (−)	8.9 (−)	12.1 (+)	9.0 (−)
c5_04011		4.0	15.6	9.1 (−)	13.3 (−)	17.9 (+)	11.8
c5_05319		5.3	19.1	11.5 (−)	17.4	22.6 (+)	17.2
c5_06820		6.8	25.5	18.9 (−)	23.9	28.4 (+)	21.7
c5_07442		7.4		20.4 (−)	27.6	31.5 (+)	24.9
c5_08563		8.6	33.3	25.1	32.1	35.2 (+)	
c5_09082	M	9.1					36.1
c5_10428		10.4	45.7	33.7	46.1	48.1	41.7
c5_12699		12.7		35.4	47.4		
c5_13614		13.6	49.0	36.4	48.2	53.1	43.1
c5_14766		14.8	54.8	42.0	53.0	58.3	
c5_16368		16.4			60.5		53.0
c5_17570		17.6	64.4	53.3	64.5	70.0	
c5_19316		19.3	73.2	62.6	68.8	77.5	62.1 (+)
c5_20318		20.3	77.1	64.2	71.4	81.4	63.6 (+)
c5_21319		21.3	82.8	68.5	75.4	84.9	66.1 (+)
c5_22415		22.4	85.9	73.2	80.3	91.4	72.5 (+)
c5_23116		23.1	87.5	75.4	82.5	96.3	74.7 (+)
c5_24997		25.0	95.1	80.4	88.1	106.3	80.9 (+)
c5_26671		26.7	101.0	89.6	98.1 (−)	112.6	87.3 (+)

Each marker name includes the chromosome number followed by the physical position in kilobases of the marker on the chromosome (Col-0 genomic sequence). "M" indicates a microsatellite marker; "I" indicates an indel marker. Other markers are SNPs. Markers showing a significant segregation distortion between the parental alleles are indicated by (+) when there is an excess of Col-0 allele and by (−) when there is a lack of Col-0 allele relative to the other parent allele.

TABLE 2

Overall genotyping data for the five RIL populations

	BlhCol	BurCol	CtCol	CviCol	ShaCol
No. of genotyped lines	319	347	377	367	349
No. of markers scored	75	87	85	90	86
% of Col-0 allele	52.7	51.3	47.1	53.2	50.8
% of heterozygosity	3.1	2.9	3.2	3.7	3.3
% of missing data	0.5	0.5	0.1	0.2	0.4

on chromosome 1 to a locus at ~3 Mb on chromosome 5 and the second one between a locus at ~5 Mb on chromosome 1 and a locus at ~3 Mb on chromosome 3. In the ShaCol population, a locus at ~13 Mb on chromosome 4 is in significant LD with a locus at ~26 Mb on chromosome 5. In all three cases, the observed LD is entirely explained by the fact that one of the four homozygous allelic combinations expected from the segregation of two independent loci is absent (or extremely rare) among the RILs (either a Col-0/Cvi-0 or a Col-0/Sha allelic combination; data not shown).

Consensus genetic map: The genetic maps obtained for the five RIL populations are presented in Table 1. The five maps are collinear. All the markers mapped genetically according to their physical position on the Col-0 genomic sequence, although some of them, particularly in the centromeric regions, could not be separated as no recombination occurred between them (c2_03041 and c2_04263 in ShaCol, c3_12647 and c3_14097 in CtCol and in ShaCol, c4_03833 and c4_04877 in CtCol, and c4_02133 and c4_03833 in ShaCol). In addition, in ShaCol no recombination event was found on the upper arm of chromosome 3 between c3_02968, c3_04141, and c3_05141. The total lengths of the genetic maps (Table 3) are in the range of 363 cM (BurCol) to 508 cM (CviCol). The average genetic distances between two adjacent markers vary between 4.4 cM (BurCol) and 6 cM (CviCol) and the maximal distances between two consecutive markers range from 11.7 cM (BurCol) to 15.5 cM (CtCol). The mean ratio between physical and genetic distances, which gives the average density of recombination events, is of 227 kb/cM (CviCol) to 317 kb/cM (BurCol) (Table 3). However, disparities of these frequencies between the chromosomes can be seen in the five RIL sets (Table 3) even though no general rule emerges except that recombination seems to occur less frequently on chromosome 3 than on the others in all the populations. In each population, comparing physical to genetic lengths reveals variation in the recombination rate along the chromosomes, including, as expected, for the centromeric regions where nearly no recombination occurs.

Construction of "core populations": From each RIL population, we established a core population that is an optimal subset of 164 lines allowing the user to phenotype only a limited number of lines without losing much QTL detection power. Similar in principle to the selective mapping strategy described by Vision *et al.* (2000), it is here intended to reduce the phenotyping task when studying the whole RIL set is impractical or too expensive, as described by Xu *et al.* (2005). To take more parameters into account, the choice of lines was implemented by hand: to build these core populations, we eliminated the lines with the most missing data and heterozygous loci and kept the lines with more recombination events that were thus the most informative. Moreover, care was taken to maintain the ratio of the parental alleles in the core populations under 1.5:1 for every marker. The core populations are described and are available at http://dbsgap.versailles.inra.fr/vnat.

Phenotyping: All lines of the five populations were simultaneously scored under long-day conditions in a single greenhouse. Three traits were measured: flowering time (FLO), rosette diameter (DIA), and total seed production (FIT). For each trait, we observed an important variation between the RILs and a transgression in both directions; *i.e.*, the range of variation found in the RIL population extended far beyond the variation in phenotypes of the parental accessions, even when the

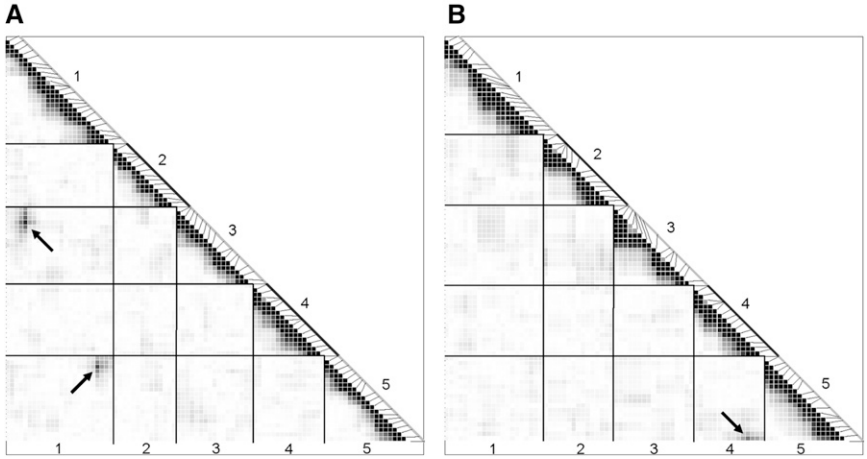

Figure 1.—Transchromosomal linkage disequilibrium (LD) within the Cvi-0 × Col-0 (A) and Sha × Col-0 (B) RIL sets. Heat plots from GGT32 software are shown, where darker spots indicate high LD between pairs of markers. Significant LD between physically unlinked genomic regions are indicated by arrows and represent pairs of markers that do not segregate independently from each other.

TABLE 3

Characteristics of the five genetic maps

	BlhCol	BurCol	CtCol	CviCol	ShaCol
Total map length (cM)	410.7	363.0	429.5	508.4	381.1
Average distance[a] (cM)	5.9	4.4	5.4	6.0	4.7
Maximum distance[a] (cM)	13.0	11.7	15.5	14.0	12.7
Physical/genetic length ratio (kb/cM)	280	317	268	227	302
Chromosome 1 ratio kb/cM	287	301	265	217	310
Chromosome 2 ratio kb/cM	281	336	265	216	305
Chromosome 3 ratio kb/cM	315	338	298	245	327
Chromosome 4 ratio kb/cM	257	336	240	219	264
Chromosome 5 ratio kb/cM	264	298	272	237	309

[a] Between two consecutive markers.

latter was extremely limited (Table 4). We here intentionally used a minimal phenotyping framework to estimate the power of QTL mapping in this context. Studying more repetitions would also have resulted in increasing the experimental variance on phenotypic estimations (Borevitz and Chory 2004) since the display is already quite extensive. However, environmental heterogeneity was not totally avoided as shown by the study of the Col-0 repetitions across the display (Table 4): Col-0 shows significantly different DIA and FIT phenotypic values from the Ct block compared to all other experimental blocks. This is very likely due to the fact that the CtCol set was grown at the southern end of the greenhouse, where the conditions seem to be different from the rest of the greenhouse. One should then not try to directly compare phenotypes obtained in the CtCol set with those of the other populations. However, this does not preclude QTL analyses, which were performed within each RIL set.

We found different significant correlations between traits (Table 5) depending on the cross, including weak positive correlations between FLO and DIA in all populations, rather strong positive correlations between DIA and FIT in CtCol and CviCol, and negative correlations between FLO and FIT in BlhCol, BurCol, and ShaCol.

QTL analysis: The detected QTL explaining the variation among the RILs for the three analyzed traits are summarized in Figure 2. They individually explain from 2 to 28% of the total intrapopulation phenotypic variation of the trait. The detailed results of the QTL analyses are presented in supplemental Table C. The use of numerous markers common to the different maps and anchored on the genomic sequence of Col-0 allowed us to compare the QTL detected in the five populations.

FLO variation is explained by three to six QTL depending on the RIL set, which were individually responsible for 2–25% of the phenotypic variation. Most of them are population specific, but three QTL were found at the same location in two populations: on chromosome (chr)1 at ~23.5 Mb in CtCol and BurCol, for which the Col-0 allele promotes earlier flowering; on chr4 at ~0.5 Mb in BlhCol and BurCol (Col-0 accelerates flowering); and on chr5 at ~3.5 Mb in CviCol (Col-0 accelerates flowering) and ShaCol (Sha accelerates flowering). In addition, one flowering-time QTL

TABLE 4

Phenotypic data of the parental accessions and the five RIL populations for three traits: long-day flowering-time, rosette diameter, and total seed weight

	BlhCol	BurCol	CtCol	CviCol	ShaCol
			Flowering time (day)		
Col-0	27.7	28.0	27.7	29.7	29.3
Other parent	39.0	41.7	27.0	26.7	27.3
RIL average	30.6	36.5	26.3	22.6	28.5
RIL range	22–51	23–66	20–39	18–34	19–47
			Rosette diameter (mm)		
Col-0	131	127	98	118	138
Other parent	149	161	74	59	116
RIL average	123	131.9	76.3	57.6	101.6
RIL range	50–189	25–197	25–157	26–135	25–179
			Seed production (mg)		
Col-0	319	276	141	262	253
Other parent	97	47	123	124	238
RIL average	162	200	135	87	178
RIL range	1–497	1–521	2–341	6–242	1–430

TABLE 5

Correlation coefficients between the three analyzed traits in each RIL population: long day flowering time (FLO), rosette diameter (DIA), and total seed weight (FIT)

	FLO/DIA	DIA/FIT	FLO/FIT
BlhCol	0.38***	NS	−0.36***
BurCol	0.14**	0.29***	−0.46***
CtCol	0.36***	0.69***	NS
CviCol	0.49***	0.68***	0.28***
ShaCol	0.39***	0.20***	−0.27***

NS, not significant. Significant at ***$P < 0.001$ and at **$P < 0.01$.

(chr5, ~26 Mb) was mapped in four of the five populations, with the Col-0 allele delaying flowering in all cases. In the CviCol and CtCol populations, which originate from early-flowering parents, the variation among RILs can be explained in both cases by six QTL with contrasting effects: for three of them the Col-0 allele is responsible for early flowering while for the three others it promotes later flowering. In the BlhCol and BurCol populations, the parental accessions have much more contrasting flowering times than in the other sets. For BlhCol, variation between the RILs is mainly due to a major QTL on chromosome 4 (~0.5 Mb) that explains 17% of the population variation and for which the Col-0 allele accelerates flowering. For BurCol, the variation can be explained by six QTL, five of which have allelic effects in the same direction (the Col-0 allele accelerates flowering).

DIA is explained according to the RIL populations by one to seven QTL with individual contributions (R^2) of 3–16%. Most of these QTL are population specific except one on chr4 at ~0.5 Mb that was mapped in BlhCol and ShaCol, one on chr5 at ~5 Mb that was mapped in BurCol and ShaCol (the Col-0 allele has opposite effects in these two RIL sets), and one on chr5 at ~26 Mb in CtCol, CviCol, and ShaCol. The latter was mapped at the same position as a flowering-time QTL in these three populations and the Col-0 allele consistently delays flowering and increases the rosette size simultaneously. A number of other DIA QTL colocalize with FLO ones (Figure 2) and their effects are always in the same direction: the earliest plants are the smallest.

For FIT, one to five QTL were found depending on the RIL set, which accounted for 3–28% of the total phenotypic variation. Seven QTL are population specific, but one (chr5, ~3.5 Mb) was detected in all crosses but CviCol, with the Col-0 allele at this locus reducing the amount of seeds in all cases. In BurCol and CviCol, some of the FIT QTL colocalize with QTL for FLO or FLO/DIA but their effects can be in either the same direction (CviCol) or the opposite (BurCol).

Analysis of QTL × QTL interactions revealed the occurrence of a few weak epistatic relationships (R^2 ranging from 1 to 4%) between pairs of loci only in the BurCol and ShaCol populations. Interestingly, only minor-effect QTL were epistatic. In BurCol, four loci were involved in interactions: chr1 at ~13.5 Mb and chr4 at ~0.6 Mb interact in the control of FLO ($R^2 = 4\%$) and FIT ($R^2 = 2\%$), chr1 at ~13.5 Mb and chr4 at ~8.6 Mb interact for FIT ($R^2 = 1\%$), and chr4 at ~8.6 Mb and chr5 at ~3.7 Mb also interact for FIT ($R^2 = 3\%$). In ShaCol, only two significant interactions were found between DIA QTL, involving chr4 at ~0.2 Mb and chr5 at ~5.2 Mb ($R^2 = 2\%$) and chr5 at ~5.2 Mb and chr5 at ~26.2 Mb ($R^2 = 3\%$).

QTL analysis in core populations: The same QTL analyses were performed using only the phenotypic data obtained with the subsets of 164 lines selected for each core population. The QTL detected are indicated by stars in Figure 2. In core populations, we were able to detect QTL with R^2 as low as 5% of within-population variation (respectively 2% in whole sets), corresponding to very weak additive effects ($2a$) (supplemental Table C). More than half of the QTL found with the complete sets were also found with the core populations and all

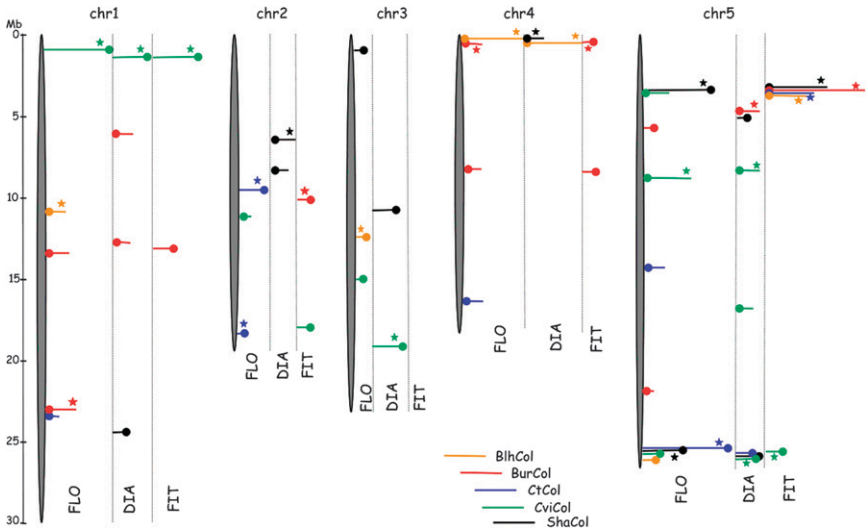

FIGURE 2.—Physical map of the QTL detected in the five RIL populations for the three traits analyzed: flowering time (FLO), rosette diameter (DIA), and total seed weight (FIT). The framework map is built from the consensus SNPlex markers, with physical positions being linearly extrapolated between adjacent markers from the respective genetic positions. The colored bars representing the QTL are located at the most probable QTL positions and the length of the bar is proportional to the percentage of variance explained by the QTL (R^2). The sign of the allelic effect is indicated by the position of the dots on the bars: a dot to the right (left) end of the bar indicates a positive (negative) allelic effect ("$2a$": mean effect of the replacement of both non-Columbia alleles by Columbia alleles at the QTL). A star beside a bar indicates that the QTL was also detected in the core population of 164 lines.

QTL with $R^2 > 7\%$ in the complete sets were identified with the core populations. On average, 72% of the phenotypic variation that is allocated to significant QTL detected with the complete set was also mapped accordingly with the core population. Although the loci found to interact within the whole populations did not have significant additive effects within the core populations, some of their epistatic interactions were still significant: in BurCol, chr1 at ~13.5 Mb and chr4 at ~0.6 Mb still interact for FLO ($R^2 = 5\%$) and chr4 at ~8.6 Mb and chr5 at ~3.7 Mb for FIT ($R^2 = 4\%$).

DISCUSSION

A powerful tool for the study of complex traits: In this work, we created an integrated set of five new large populations of RILs optimized for QTL mapping, from crossing a pivot accession (Col-0) as a male parent to different accessions. To maximize the number of different alleles that would segregate in the populations, the female parents (Blh-1, Bur-0, Ct-1, Cvi-0, and Sha) all belong to the Arabidopsis core-collection minimal set that presents tremendous genetic and phenotypic diversity (McKhann *et al.* 2004; Reboud *et al.* 2004). The accuracy of QTL mapping benefits from a high-resolution genetic map, which is mainly a function of the number of evenly distributed markers and the quality of the genotyping as well as the size of the RIL sets (Darvasi and Soller 1994; Charmet 2000). Although some existing Arabidopsis RIL populations either have more lines or are mapped with more markers, these ones are to date, to our knowledge, those cumulating the largest number of lines (319–377), the highest density of evenly distributed markers (1.3–1.5 Mb or 4.4–6 cM on average between two consecutive markers with no gaps >11.7–15.5 cM), and the greatest number of anchored consensus markers (52 markers common to the five maps plus 28 common to four of them). Once QTL have been mapped, the next step is to identify the genes responsible for these QTL. As illustrated below, these new RIL sets offer the advantage of allowing more direct candidate gene studies (due to a more accurate QTL localization) and QTL colocalization analyses (due to the numerous consensus markers).

Genetic maps and recombination rates: All markers were mapped in accordance to their physical order on the genome. Most of the adjacent markers that could not be separated from each other are located in the centromeric regions that are known to undergo nearly no recombination. Indeed, a decrease in recombination frequency is observed around the centromeres of all the chromosomes. In the ShaCol population, no recombination event occurred between the markers c3_02968 and c3_05141 in a rather large noncentromeric region of nearly 2.2 Mb, which obviously makes this population unsuitable for map-based cloning in that specific region. This is also observed in the Bay-0 × Sha RIL set created by Loudet *et al.* (2002) and is then likely due to a structural chromosomal change in the accession Shahdara compared to Col-0 and Bay-0, such as an inversion of this region of chromosome 3. A number of chromosome rearrangements have been described and should be increasingly detected as genetic and cytogenetic analyses are performed on more and more accessions (Koornneef *et al.* 2003). For example, Fransz *et al.* (2000) described an inversion on the short arm of chromosome 4 in the accessions L*er* and WS with respect to Col-0 that suppresses recombination in the concerned interval. We also observed a decrease of recombination in our five RIL sets in this region of chromosome 4, but, as it is located very close to the centromere, our marker density is not high enough to attribute this suppression of recombination either to an inversion between Col and the other parental accessions or to the proximity of the centromere.

Until now, the lengths of the genetic maps from existing RIL populations have been reported to be roughly similar (Lister and Dean 1993; Alonso-Blanco *et al.* 1998b; Loudet *et al.* 2002; Clerkx *et al.* 2004; El-Lithy *et al.* 2006; Torjek *et al.* 2006). Here, the five new maps are again of comparable sizes; however, the level of resolution of the genetic maps and the high number of common markers used in this study allow a more precise comparison. It can be seen that the rate of recombination differs not only between different regions of the chromosomes but also between the chromosomes and from one cross to another. Recombination occurred more frequently in the CviCol population than in BurCol (18.6 crossings over per F_6 line on average *vs.* 13.4), leading to a 40% increase in the genetic map length with very similar numbers of markers and lines (Tables 2 and 3). This is in agreement with cytogenetic data from Sanchez-Moran *et al.* (2002) that show significant variation in meiotic recombination frequency between diverse accessions. As also observed by these authors in different accessions, the variation in recombination rate among the different chromosomes depends on the cross (Table 3). The recombination rate is higher on chromosome 4 than on the other chromosomes in BlhCol, CtCol, and ShaCol, as already observed by Loudet *et al.* (2002) in the Bay-0 × Sha cross. It was previously suggested that this phenomenon could be due to chromosome 4 being the smallest in physical length, combined with the requirement for one crossing over per chromosome arm at meiosis (Copenhaver *et al.* 1998). However, the recombination rate is rather low on chromosome 4 in BurCol and not higher than that on chromosomes 1 and 2 in CviCol. Moreover, recombination frequency is not particularly high in any cross for chromosome 2, which is more or less the same size as chromosome 4. Surprisingly, chromosome 3 seems to undergo less recombination events than the others in every population, an observation that cannot

be explained by the chromosome size. A more precise study of these RILs will provide new insights into recombination features in *A. thaliana*.

Segregation distortion of parental alleles and linkage disequilibrium: Despite the care taken to avoid any artificial selection during the SSD steps of RIL generation, regions of the genome with a significant distortion in the segregation of parental alleles were found in each set. Nonetheless, due to the very large sizes of the populations, that distortion does not impair QTL detection, which actually relies on the number of RILs representing each allele. For example, the strongest segregation distortion was detected in CviCol at the marker c5_02900 (ratio of 2.4:1), but there are still >100 lines representing the less frequent genotype in this region, which is largely sufficient to allow QTL analysis even in an epistatic context. The segregation distortion could be due to the effect of a number of genetic and/or environmental factors cumulating across generations. For instance, selection resulting from environmental conditions being unfavorable to some genotypes (at the germination stage, for example) cannot easily be avoided.

Some of the local segregation distortions can also be explained by negative epistatic interactions between different loci. Specific combinations of parental alleles at different regions of the genome can be unfavorable and then counterselected, or even lethal, resulting in those loci not segregating independently from each other and behaving as in LD. Such a situation was recently described by Torjek *et al.* (2006) in a population derived from the cross Col-0 × C24, where a specific combination of alleles at two distant loci led to a reduced fertility. In our RIL sets, three cases of transchromosomal LD were found. Two occur in the CviCol population and seem to be responsible for all of the segregation distortions observed in this set. Interestingly, Alonso-Blanco *et al.* (1998b) reported a similar pattern of distortion within the Ler × Cvi-0 RIL set involving regions very close to the loci involved in one of the transchromosomal LD reported here in CviCol (bottom of chromosome 1 and top of chromosome 5). The third case was observed in the ShaCol population and also corresponds to regions with distortions of segregation. This pair of interacting loci colocalizes with that reported by Torjek *et al.* (2006). The three pairs of loci are currently under investigation for fine mapping and cloning.

QTL mapping: The female parents used to generate the RIL populations were chosen for being genetically distant (Nordborg *et al.* 2005; Ostrowski *et al.* 2006) and show important variation for most morphological traits (Reboud *et al.* 2004). While the QTL identified in a single RIL population concern only a fraction of the genes that potentially affect a trait in the species, the analysis of these multiple-RIL sets is expected to reveal different QTL, depending on the combination of alleles present in the parental accessions. A total of 19, 13, and 8 different QTL were identified for FLO, DIA, and FIT, respectively, whereas individual populations segregated at the most for 6, 7, and 5 QTL. The interest of using several populations was also underscored in particular by Symonds *et al.* (2005) for trichome density and El-Lithy *et al.* (2006) for flowering-time variation. We found transgression even in RIL sets derived from crosses between accessions showing similar values for the trait studied (FLO in CtCol, CviCol, and ShaCol) since different combinations of positive and negative alleles can result in the same phenotype, and we were able to identify QTL with very limited allelic effects.

Among the numerous genes identified to regulate flowering, *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) are key factors in the variation of flowering time (Roux *et al.* 2006). Four of our parental accessions (Col-0, Ct-1, Cvi-0, and Sha) are early flowering due to nonfunctional alleles of *FRI* and/or *FLC*, whereas Blh-1 and Bur-0 are late flowering (Shindo *et al.* 2005; Werner *et al.* 2005). While Blh-1 has functional alleles for both *FRI* and *FLC*, Werner *et al.* (2005) showed that the Bur-0 *FLC* allele is inactive but that this is masked by the presence of other late-flowering loci. In the BlhCol and BurCol populations, FLO QTL were found to colocalize with *FRI* (chr4, 0.3 Mb), with the Col-0 allele accelerating flowering in both cases, as expected. In CviCol and ShaCol, a flowering-time QTL, also reported in L*er* × Cvi-0 (Alonso-Blanco *et al.* 1998a) and in Bay-0 × Sha (Loudet *et al.* 2002), was identified at ~3.5 Mb on chromosome 5 near the *FLC* locus (3.2 Mb). However, as both Col-0 and Cvi-0 have functional *FLC* alleles, at least in the CviCol population this QTL likely does not correspond to FLC, and indeed a number of genes involved in the control of flowering are located in this region of chromosome 5 (Koornneef *et al.* 1998).

Using populations whose parents possess nonfunctional *FRI* and *FLC* alleles and thus where these large-effect alleles do not segregate eased the identification of other QTL that contribute to flowering-time variation. A number of the other FLO QTL detected here colocalized with QTL previously found in long-day conditions in other mapping populations. Moreover, the accuracy of the map allowed us to suggest probable candidate genes, on the basis of their position. At 1 Mb on chromosome 1, a major QTL with a positive effect found only in CviCol most likely corresponds to *CRY2* (located at 1.2 Mb). A QTL was also identified at this location in L*er* × Cvi-0 (Alonso-Blanco *et al.* 1998a), and El-Assal *et al.* (2001) showed that a punctual mutation in *CRY2* specific to Cvi-0 was responsible for earlier flowering in this accession. A possible candidate for the QTL at 11.1 Mb on chromosome 1 in BlhCol is the FRIGIDA-like *FLC* activator *FRL1-2* (11.4 Mb). In BurCol and CtCol, a FLO QTL around 24 Mb on chromosome 1, also found by El-Lithy *et al.* (2006) in L*er* × Kondara, colocalizes with the floral pathway integrator *FT* (24.3 Mb) that

promotes flowering. On chromosome 2, a QTL at 9.7 Mb in CtCol, also found in Bay-0 × Sha (Loudet *et al.* 2002), colocalizes with the floral repressor *SVP* (9.6 Mb), a QTL at 11.4 Mb in CviCol colocalizes with the light-dependent pathway gene *ELF3* (11.1 Mb), and a QTL at 18.7 Mb in CtCol colocalizes with the suppressor of overexpression of *CONSTANS*, *SOC1* (18.8 Mb). In ShaCol, the QTL on the top of chromosome 3 was also described in the three populations studied by El-Lithy *et al.* (2006). The QTL at 15.2 Mb on chromosome 3 in CviCol could correspond to *VIP3* (14.6 Mb). The QTL mapped in BurCol at 8.4 Mb on chromosome 4 could be the same as that in L*er* × Kas-2 (El-Lithy *et al.* 2006). Good candidates for the QTL on chromosome 5 at 5.9 Mb in BurCol and at 14.5 Mb in CtCol are, respectively, the floral repressor *TFL2* (5.8 Mb) and the *PHYC* photoreceptor gene (14 Mb) (Balasubramanian *et al.* 2006). Although not so close to the FLO QTL found in CviCol around 9 Mb on chromosome 5, the *HUA2* gene (located at 7.8 Mb) could be a candidate, as Wang *et al.* (2007) recently showed that natural changes in this gene have implications for the control of flowering induction. Finally, the QTL found in all our populations but BurCol at the very end of chromosome 5 (~26 Mb), a region also indicated in L*er* × Cvi-0 (Alonso-Blanco *et al.* 1998a), in L*er* × Sha (El-Lithy *et al.* 2004), and in L*er* × Kas-2 and L*er* × Kond (El-Lithy *et al.* 2006), could be the floral repressor *MAF2*. Among our 19 FLO QTL, 15 were population specific and 10 were not identified in previous long-day studies. Indeed, in contrast to *FRI* where loss-of-function appears in multiple ways, many alleles of other genes that modulate flowering responses could be rare (El-Assal *et al.* 2001; Maloof *et al.* 2001; Werner *et al.* 2005) and further characterizations of the variation underlying QTL with small effects are needed to fully understand the global architecture of the trait.

In this study, several DIA and/or FIT QTL were found to colocalize with FLO ones, explaining trait correlations. For instance, in CviCol, we found a positive correlation between FLO and DIA and a high positive correlation between DIA and FIT (Table 5); *i.e.*, the earliest plants were the smallest and produced the smallest amount of seeds. This is due to major QTL with allelic effects in the same direction responsible for the variation of these traits (on chr1 at ~1 Mb and on chr5 at ~9 and ~26 Mb). On the contrary, in BurCol, we found a negative correlation between FLO and FIT, here as well explained by colocalized QTL responsible for the variation of both traits, but with opposite allelic effects. This is likely due to our experimental growth conditions: watering was arbitrarily stopped 20 days after flowering individually for each plant, which is probably far from optimal for the latest-flowering plants of the BurCol population (duration of the reproductive period for very late-flowering plants often exceeds that of earlier-flowering plants).

Core populations: Of the 53 QTL identified, 28 (among which are all the major ones) were found using solely the core populations, despite our intended minimal phenotyping display (no repetition). These core populations thus make it feasible to perform QTL analyses with reduced cost and saved time without impairing the power of detection of major- and medium-effect QTL. Indeed, QTL that until now were considered to be amenable to cloning are mostly large-effect ones; however, if the aim is not only to map main QTL to clone them, but also to fully describe the complete genetic architecture of trait variation, including weak epistatic interactions, it is advisable to use complete populations. In both cases, it is important to keep in mind that using more RILs is almost always more powerful than performing more repetitions (see also Keurentjes *et al.* 2006). This is due to the fact that RILs are already partial repetitions of each other, so that phenotyping more RILs not only increases the number of informative recombination breakpoints analyzed, but also improves the estimation of phenotypic values for a each given genotypic class the RIL participates in. Our strategy, consisting of genotyping a very large set of RILs to then select the most informative (recombined) lines for phenotyping, seems efficient as it reconciles the need for QTL detection power and the inherent difficulty in phenotyping a large number of individuals.

Until now, most quantitative approaches used to study complex traits have been conducted in a limited number of mapping populations, which harbor a very small part of the existing allelic variation. This identifies only a fraction of the loci involved in the control of the traits. The different populations surveyed in this work segregated for different loci, depending on the genetic composition of their parental accessions. This confirms that the use of multiple RIL sets originating from different crosses is still needed to get insight into the diversity of the species and to dissect the global genetic architecture of traits. In this aim, in addition to the five RIL sets surveyed here, we generated 10 supplementary RIL populations following the same strategy and using the same markers: 6 are described and currently available at http://dbsgap.versailles.inra.fr/vnat (female parents: Bla-1, Can-0, Ge-0, Nok-1, Ri-0, and Tsu-0) and 4 more are currently being genotyped and will be made available soon (female parents: Ita-0, Jea, Oy-0, and Yo-0).

We thank all the Resource Centre team: J. Babillot, L.Laroche, J. Legay, P. Marie, B. Trouvé, and C. Sallé for producing the RILs and Roger Voisin for his help in taking care of the plants. SNP genotyping was partly financed by the Institut National de la Recherche Agronomique Department of Genetics and Plant Breeding.

LITERATURE CITED

Alonso-Blanco, C., and M. Koornneef, 2000 Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. Trends Plant Sci. **5:** 22–29.

Alonso-Blanco, C., S. E. El-Assal, G. Coupland and M. Koornneef, 1998a Analysis of natural allelic variation at flowering time loci in the Landsberg *erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. Genetics **149:** 749–764.

Alonso-Blanco, C., A. J. Peeters, M. Koornneef, C. Lister, C. Dean *et al.*, 1998b Development of an AFLP based linkage map of Ler, Col and Cvi Arabidopsis thaliana ecotypes and construction of a Ler/Cvi recombinant inbred line population. Plant J. **14:** 259–271.

Balasubramanian, S., S. Sureshkumar, M. Agrawal, T. P. Michael, C. Wessinger *et al.*, 2006 The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of Arabidopsis thaliana. Nat. Genet. **38:** 711–715.

Basten, C. J., B. S. Weir and Z. B. Zeng, 2000 *QTL CARTOGRAPHER*, Version 1.14. North Carolina State University, Raleigh, NC.

Borevitz, J. O., and J. Chory, 2004 Genomics tools for QTL analysis and gene discovery. Curr. Opin. Plant Biol. **7:** 132–136.

Charcosset, A., and A. Gallais, 1996 Estimation of the contribution of quantitative trait loci (QTL) to the variance of a quantitative trait by means of genetics markers. Theor. Appl. Genet. **93:** 1193–1201.

Charmet, G., 2000 Power and accuracy of QTL detection: simulation studies of one-QTL models. Agronomie **20:** 309–323.

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

Clark, R. M., G. Schweikert, C. Toomajian, S. Ossowski, G. Zeller *et al.*, 2007 Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science **317:** 338–342.

Clerkx, E. J. M., M. E. El-Lithy, E. Vierling, G. J. Ruys, H. Blankestijn-De Vries *et al.*, 2004 Analysis of natural allelic variation of Arabidopsis seed germination and seed longevity traits between the accessions Landsberg erecta and Shakdara, using a new recombinant inbred line population. Plant Physiol. **135:** 432–443.

Copenhaver, G. P., W. E. Browne and D. Preuss, 1998 Assaying genome-wide recombination and centromere functions with Arabidopsis tetrads. Proc. Natl. Acad. Sci. USA **95:** 247–252.

Darvasi, A., and M. Soller, 1994 Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. **89:** 351–357.

El-Assal, S., C. Alonso-Blanco, A. J. Peeters, V. Raz and M. Koornneef, 2001 A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. Nat. Genet. **29:** 435–440.

El-Lithy, M. E., E. J. M. Clerkx, G. J. Ruys, M. Koornneef and D. Vreugdenhil, 2004 Quantitative trait locus analysis of growth-related traits in a new Arabidopsis recombinant. Plant Physiol. **135:** 444–458.

El-Lithy, M. E., L. Bentsink, C. J. Hanhart, G. J. Ruys, D. I. Rovito *et al.*, 2006 New Arabidopsis recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. Genetics **172:** 1867–1876.

Fransz, P. F., S. Armstrong, J. H. de Jong, L. D. Parnell, C. van Drunen *et al.*, 2000 Integrated cytogenetic map of chromosome arm 4S of A. thaliana: structural organization of heterochromatic knob and centromere region. Cell **100:** 367–376.

Keurentjes, J. J., L. Bentsink, C. Alonso-Blanco, C. J. Hanhart, H. Blankestijn-De Vries *et al.*, 2006 Development of a near isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics **175:** 891–905.

Koornneef, M., C. Alonso-Blanco, A. J. Peeters and W. Soppe, 1998 Genetic control of flowering time in Arabidopsis. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49:** 345–370.

Koornneef, M., P. Fransz and H. de Jong, 2003 Cytogenetic tools for Arabidopsis thaliana. Chromosome Res. **11:** 183–194.

Koornneef, M., C. Alonso-Blanco and D. Vreugdenhil, 2004 Naturally occurring genetic variation in Arabidopsis thaliana. Annu. Rev. Plant Biol. **55:** 141–172.

Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

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

Loudet, O., S. Chaillou, C. Camilleri, D. Bouchez and F. Daniel-Vedele, 2002 Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis. Theor. Appl. Genet. **104:** 1173–1184.

Loudet, O., S. Chaillou, P. Merigout, J. Talbotec and F. Daniel-Vedele, 2003 Quantitative trait loci analysis of nitrogen use efficiency in Arabidopsis. Plant Physiol. **131:** 345–358.

Maloof, J. N., J. O. Borevitz, T. Dabi, J. Lutes, R. B. Nehring *et al.*, 2001 Natural variation in light sensitivity of Arabidopsis. Nat. Genet. **29:** 441–446.

McKhann, H. I., C. Camilleri, A. Berard, T. Bataillon, J. L. David *et al.*, 2004 Nested core collections maximizing genetic diversity in Arabidopsis thaliana. Plant J. **38:** 193–202.

Nordborg, M., and J. Bergelson, 1999 The effect of seed and rosette cold treatment on germination and flowering time in some Araidopsis thaliana (Brassicaceae) ecotypes. Am. J. Bot. **86:** 470–475.

Nordborg, M., T. T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian *et al.*, 2005 The pattern of polymorphism in Arabidopsis thaliana. PLoS Biol. **3:** e196.

Ostrowski, M. F., J. David, S. Santoni, H. McKhann, X. Reboud *et al.*, 2006 Evidence for a large-scale population structure among accessions of Arabidopsis thaliana: possible causes and consequences for the distribution of linkage disequilibrium. Mol. Ecol. **15:** 1507–1517.

Reboud, X., V. Le Corre, N. Scarcelli, F. Roux, J. L. David *et al.*, 2004 Natural variation among accessions of *Arabidopsis thaliana*: beyond the flowering date, what morphological traits are relevant to study adaptation, pp. 135–142 in *Plant Adaptation: Molecular Genetics and Ecology*, edited by Q. C. B. Cronk, J. Whitton, R. H. Ree and I. E. P. Taylor. NRC Research Press, Ottawa, Ontario, Canada.

Roux, F., P. Touzet, J. Cuguen and V. Le Corre, 2006 How to be early flowering: an evolutionary perspective. Trends Plant Sci. **11:** 375–381.

Sanchez-Moran, E., S. J. Armstrong, J. L. Santos, F. C. Franklin and G. H. Jones, 2002 Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana*. Genetics **162:** 1415–1422.

Shindo, C., M. J. Aranzana, C. Lister, C. Baxter, C. Nicholls *et al.*, 2005 Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. Plant Physiol. **138:** 1163–1173.

Symonds, V. V., A. V. Godoy, T. Alconada, J. F. Botto, T. E. Juenger *et al.*, 2005 Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. Genetics **169:** 1649–1658.

Tonsor, S. J., C. Alonso-Blanco and M. Koornneef, 2005 Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in Arabidopsis thaliana. Plant Cell Environ. **28:** 2–20.

Torjek, O., H. Witucka-Wall, R. C. Meyer, M. von Korff, B. Kusterer *et al.*, 2006 Segregation distortion in Arabidopsis C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. Theor. Appl. Genet. **113:** 1551–1561.

Vision, T. J., D. G. Brown, D. B. Shmoys, R. T. Durrett and S. D. Tanksley, 2000 Selective mapping: a strategy for optimizing the construction of high-density linkage maps. Genetics **155:** 407–420.

Wang, Q., U. Sajja, S. Rosloski, T. Humphrey, M. C. Kim *et al.*, 2007 HUA2 caused natural variation in shoot morphology of A. thaliana. Curr. Biol. **17:** 1513–1519.

Weigel, D., and M. Nordborg, 2005 Natural variation in Arabidopsis. How do we find the causal genes? Plant Physiol. **138:** 567–568.

Werner, J. D., J. O. Borevitz, N. H. Uhlenhaut, J. R. Ecker, J. Chory *et al.*, 2005 FRIGIDA-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. Genetics **170:** 1197–1207.

Xu, Z., F. Zou and T. Vision, 2005 Improving quantitative trait loci mapping resolution in experimental crosses by the use of genotypically selected samples. Genetics **170:** 401–408.

Communicating editor: R. W. Doerge