Naturally Occurring Genetic Variation in *Arabidopsis Thaliana*

Maarten Koornneef,[1] Carlos Alonso-Blanco,[2] and Dick Vreugdenhil[3]

[1]*Laboratory of Genetics, Wageningen University, Arboretumlaan 4, NL-6703 BD Wageningen, The Netherlands; email: maarten.koornneef@wur.nl*
[2]*Departamento de Biotecnología, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Carretera de A Coruña, Km 7, Madrid 28040, Spain*
[3]*Laboratory of Plant Physiology, Wageningen University, Arboretumlaan 4, NL-6703 BD Wageningen, The Netherlands*

Key Words   quantitative trait loci, adaptation, flowering, plant growth

■ **Abstract**   Currently, genetic variation is probably the most important basic resource for plant biology. In addition to the variation artificially generated by mutants in model plants, naturally occurring genetic variation is extensively found for most species, including *Arabidopsis*. In many cases, natural variation present among accessions is multigenic, which has historically hampered its analysis. However, the exploitation of this resource down to the molecular level has now become feasible, especially in model species like *Arabidopsis*, where several genes accounting for natural variation have already been identified. Dissecting this variation requires first a quantitative trait locus (QTL) analysis, which in *Arabidopsis* has proven very effective by using recombinant inbred lines (RILs). Second, identifying the particular gene and the nucleotide polymorphism underlying QTL is the major challenge, and is now feasible by combining high-throughput genetics and functional genomic strategies. The analysis of *Arabidopsis* natural genetic variation is providing unique knowledge from functional, ecological, and evolutionary perspectives. This is illustrated by reviewing current research in two different biological fields: flowering time and plant growth. The analysis of *Arabidopsis* natural variation for flowering time revealed the identity of several genes, some of which correspond to genes with previously unknown function. In addition, for many other traits such as those related to primary metabolism and plant growth, *Arabidopsis* QTL analyses are detecting loci with small effects that are not easily amenable by mutant approaches, and which might provide new insights into the networks of gene regulation.

CONTENTS

INTRODUCTION

Arabidopsis has been adopted as a major model or reference plant especially suit-
able for genetic and molecular research. This has led to the establishment of a large
research community with important biological and molecular resources available
(104). The presence of the complete genomic sequence (11) and a huge collection
of gene disruptions provide a research resource that is unique for higher plants
(http://www.arabidopsis.org). It enables functional analysis of individual genes by
studying the phenotype of genotypes in which specific genes have been disrupted
by T-DNA or transposon insertions or in which gene-expression is reduced by
RNAi technology or overexpressed. Although studying the effect of individual
genes has impressively increased our knowledge, the genetic analysis of complex
traits requires the study of the phenotypic effect of gene combinations and their sub-
sequent effect on other genes, gene products, and processes indirectly controlled.
This is being achieved by studying these genotypes with genomic, transcriptomic,
metabolomic, and proteomic approaches, leading us into the era of system biology.

In addition to the artificially generated genotypes and high-throughput tech-
nologies, analysis of variants that are found in nature provides an important source
of genetic variation that can be used to gain insight into the control of important
processes in plants. This resource is relevant to discover which specific allelic vari-
ants are present in nature, where they might either be neutral or have a selective
advantage under specific conditions.

Here we describe the analysis and use of this natural variation present among
Arabidopsis wild accessions. The approaches described for *Arabidopsis* can also
be used in other plant species including crop plants, provided that sufficient genetic
and genomic resources are available (54). Presently, this is the case for rice, and it
is expected that these resources will soon also be available for several other crop
plants (54).

Taxonomic Status

Arabidopsis thaliana (L.) Heyhn ($2n = 10$) is a small annual weed in the mustard
family (Brassicaceae or Cruciferae). Until recently little was known about the

systematics of the *Arabidopsis* genus, although the position of *A. thaliana* as a species has never been disputed (10) and natural or fully fertile hybrids with other species have not been reported. Supported by molecular analyses (10), O'Kane & Al-Shehbaz (124) revised the taxonomy of the genus *Arabidopsis* and assigned nine species to this genus while removing others previously assigned to it (10). All other *Arabidopsis* species for which the chromosome number was determined have a higher number than 2n = 10. Some species, such as *A. lyrata* (2n = 16) and *A. halleri* (2n = 16), previously described as *Cardaminopsis* species, are the subject of increasing research. Within this genus, the allotetraploid species *A. suecica* (2n = 26) originates from an ancient hybrid of *A. thaliana* with *A. arenosa* (2n = 16). This hybrid could be reconstructed in the laboratory (28). Vigorous but male-sterile hybrids arose from crossing *A. thaliana* with *A. lyrata*, which could be used as female parents in backcrosses with both parental species. However, recombination between the homologous chromosomes of these species probably does not occur (122).

Geographic Distribution, Habitat, and Ecology

Arabidopsis thaliana, called *Arabidopsis* hereafter, is native to Europe and central Asia and now is naturalized at many places elsewhere in the world (10). Hoffmann (50) (Figure 1) described the biogeography of the species in detail and showed that too-low spring and autumn temperatures, and high temperatures (average month temperature >22°C) with low precipitation in summer, limit its distribution range. Climatic conditions are, on a global scale, sufficient for explaining the distribution boundaries. Hoffmann (50) concluded that laboratory conditions with sufficient water and high mean temperatures are artificial for most accessions except in the southeastern North American range. However, within its limitations, *Arabidopsis* has a wider climatic amplitude than other well-investigated species of the *Brassicaceae*, and it has an impressive latitudinal range from 68°N (North Scandinavia) to 0° (mountains of Tanzania and Kenya), which makes it suitable for analyzing variation in adaptive traits.

Arabidopsis occupies open or disturbed habitats often on sandy or loamy soils such as riverbanks, roadsides, rocky slopes, waste places, and cultivated grounds. It has been found from sea level up to 4250 m (10). Its life history characteristics and occurrence in disturbed habitats indicate that the species is vulnerable to rapid colonization and extinction cycles and is a poor competitor in dense vegetations.

The ecology of *Arabidopsis* was reviewed by Pigliucci (129) and is summarized here. Flowering time and seed dormancy are key traits that determine the timing and length of the *Arabidopsis* natural life cycle. For both traits extensive variation is present among *Arabidopsis* accessions (89), which determine the variation observed in life cycles. In Europe, *Arabidopsis* flowers most abundantly in spring and early summer, and mature seeds are available from May to July (89). However, flowering plants can also be found in late summer up to early fall, although they are much less abundant. These later flowering plants almost certainly germinate in

spring in the same year and presumably are summer annuals. However, the plants flowering early in the year probably germinate in the fall and survive the winter as rosettes and therefore are winter annuals. These differences become obvious when accessions collected from nature are grown in greenhouse conditions without vernalization. In these conditions, winter annual accessions are late flowering but very responsive to vernalization, which under natural conditions synchronizes flowering in springtime. In contrast, accessions that do not require vernalization, and which flower early, only after a limited number (6–20) of rosette leaves form (120 and references therein), are most likely summer annuals. Because many accessions, including recently collected ones (90; M. Koornneef & C. Alonso-Blanco, unpublished data) are early flowering, it is clear that this trait is not specifically found in laboratory strains as Pigliucci suggested (129). In general, most accessions from Northern Europe are late flowering and typical winter annuals; southern populations are of both types, summer and winter annuals (58). In Western Europe, both early and late flowering types are found among the progeny of seeds harvested in early summer (89, 90).

Population Structure and Genetic Variation

Arabidopsis is predominantly a selfing species, and therefore most plants collected in nature represent inbred lines (89), which are practically homozygous. These wild homozygous lines are commonly referred to as ecotypes, a term originally defined as distinct races of a species genetically adapted to particular habitats. However, as previously noted, the use of this word in *Arabidopsis* does not strictly conform to its ecological definition. Because the ecological meaning has been lost, the term accession, as it is often used in germplasm collections, is becoming more common to refer to a sample of a species collected at a specific location (8).

Genetic variation within populations, and especially among populations, has been studied extensively using a variety of tools. Plants raised from seeds collected in nature have been analyzed phenotypically and molecularly for numerous characteristics including morphological traits (84, 161, for early studies in the United Kingdom and Czech Republic see 89); isozyme markers (3); microsatellites (53, 95 161); restriction fragment length polymorphisms, mainly analyzed as cleaved amplified polymorphic *s*equences (CAPS) markers (13, 17, 47, 158), amplified fragment length polymorphism (AFLP) markers (23, 38, 118, 147); and extensive DNA sequencing (144). Most studies have focused on detecting polymorphisms in single-copy nuclear sequences. However, repeated sequences and gene families have been also analyzed, including mtDNA (158), centromere repeats (46), and transposable elements (41). Cytogenetic polymorphisms have also been found, but these have been studied only in a limited number of accessions (reviewed in 76). From all these analyses, we can conclude that within populations of *Arabidopsis* polymorphisms are limited but not absent (84, 89), in agreement with the high degree of selfing reported (3). However, exceptions with higher variation have been reported (3, 84, 89). In addition, the relatively high genetic variation found

among *Arabidopsis* accessions using genome-wide markers such as AFLP, CAPS, and microsatellites (17, 38, 47, 53, 95, 118, 147, 158) shows that in general there is no association (or weak association) between geographical origin and genetic distance. The topology of neighbor-joining trees is often star–like with relatively long, and often similar, branch lengths for individual accessions. This lack of phylogeografic structure is explained by a rapid recent expansion of the species with strong involvement of human-induced migrations. Nevertheless, Sharbel et al. (147) found genetic isolation by distance by analyzing 79 AFLP markers in a worldwide collection of 142 accessions, with a major representation from central Europe. These authors suggested that after the last glaciation *Arabidopsis* colonized Central and Northern Europe from Asia and from Mediterranean Pleistocene refugia.

More than 25 genes have been systematically sequenced and compared in various accessions, and the number of genes for which DNA sequence comparisons among accessions are available is increasing rapidly (4, 33, 52, 62–64, 83, 90, 102, 117, 125, 133, 134, 149, 155). These include mainly floral and meristem developmental genes, pathogen resistance and defense genes, and genes encoding metabolic enzymes. Sequence analyses of individual genes revealed nucleotide diversity values ranging from 0.0006 for *ATTI1* (33) to 0.0558 for *CLV2* (149), with an average value of 0.006. Similar levels of variation have been estimated by analyzing 606 sequence tagged sites (STS) in 12 accessions, where the mean sequence divergence to the Col accession was 0.68%. A well-defined dimorphic haplo-structure with a clear separation in two highly differentiated haplotypes has been found for some genes such as *ADH* (52), *ChiA* (62), *ChiB* (63), *FAH*, *F3H* (4), *TFL* (125), and *RPS2* (102), but not for others such as *CHI* (83), *CAL* (133), *AP3*, *PI* (134), *AP1*, *LFY* (125), or *FRI* (90). The frequent presence of sequence dimorphism could be due to a recent admixture of two differentiated populations (134). However, because allelic dimorphism is seemingly not present at a genome-wide scale, this remains unclear. In addition, the maintenance of a dimorphic pattern of variation might involve selection processes such as balancing selection (52, 102, 149, 155). Nevertheless, as Aguadé (4) argued, sequence dimorphisms could also fit the expectations of a neutral process in an essentially selfing species. Therefore, the role played by demographic events, drift, and selection in the evolutionary history of *Arabidopsis* is currently unclear.

GENETIC ANALYSIS OF NATURAL VARIATION

Analyzing natural genetic variation involves three overall steps:

1. phenotypic identification of genetic variation for traits of interest,
2. determination of the genetic basis underlying this variation,
3. finding the molecular nature of the allelic differences that account for the genetic differences.

Phenotypic Variation Among Accessions

When different *Arabidopsis* accessions are grown together and compared under similar environmental conditions, genetic variation can be observed for many traits. Phenotypic variation for morphological and physiological traits is abundant and enables almost every *Arabidopsis* accession to be distinguished from other accessions collected at different locations. This variation is of interest from two general points of view. First, analyzing this natural variation makes it possible to identify the function of individual genes. Despite the fact that mutant approaches have been very powerful for functional analysis, the definition of gene functions using these procedures is limited by the small number of genetic backgrounds analyzed. Ultimately, the sort of mutant phenotypes that can be identified depends on the wild-type genotype. For instance, mutant phenotypes of genes for which the wild-type accession carries a natural null allele (either mutated or silenced) or a weak allele might not be detected. Epistatic interactions ensure that some phenotypes appear only in certain genetic backgrounds. One of the best documented examples of the identification of genes using natural variation that could not be identified in the standard laboratory strains are the flowering time loci *FRI* and *FLC* (described in detail below). Late flowering is produced only when both loci carry active alleles, and in early accessions such as L*er*, Col, and Ws, one or both genes is defective. Examples of loss of function and probably null alleles present in accessions are increasing, which is illustrated with the large number of deletions found between the two laboratory accessions L*er* and Col (18). By hybridization of genomic DNA on Affimetrix microarrays designed for gene expression containing part of the Columbia sequence, 111 Col genes were found that were very likely partially or completely deleted in the L*er* accession (18). These genes mainly included transposons and R genes, but also several secondary metabolism genes and genes with unknown function. There are specific alleles in nature that would not be easily recognized in mutant screens because they require very specific amino acid changes and therefore appear at an extremely low frequency (37).

Second, analysis of natural variation has an increasing interest from an ecological and evolutionary perspective (114, 139). Thus, the patterns of phenotypic and molecular variation observed are analyzed aiming to infer the mechanisms generating and maintaining this variation, and to identify which allelic variants are adaptive under specific environmental conditions. It is likely that variation for certain traits, e.g., timing of flowering and germination, or tolerance to biotic and abiotic factors, reflect adaptations to specific environments and bear ecological significance. However, for many other traits it is less clear how variation is maintained. In this respect, one important aspect to understanding natural variation requires the determination of the genetical and physiological trade-offs between different traits. Ultimately, due to the limited resources of the organism, any apparently positive trait involves a cost for the individual to maintain that trait. Thus, for instance, susceptible alleles of resistance loci might survive in nature when no strong selection for resistance occurs (78, 80, 102).

Natural genetic variation was identified for many traits mainly by direct analysis and comparison of accessions. This evaluation is facilitated by the large collection of more than 300 different accessions collected worldwide, which are publicly available in the stock centers (Figure 1). However, certain genetic variation present among *Arabidopsis* accessions is undetectable by comparison of accessions, and can only be revealed when accessions are crossed and segregating individuals of the offspring display phenotypes outside the range of variation of the parents (transgression). Analysis of transgressive segregation has proven particularly useful for detecting variation for homeostatic traits such as the circadian period (152) or for viability loci detected by the segregation distortion observed in most mapping populations (9).

Genetic variation has been found for resistances to biotic factors such as bacteria, fungi, viruses, insects, and mammals (24, 30, 35, 42, 45, 51, 55, 68, 87, 106, 156, 162, 164). Variation for disease resistance genes is large and involves multiple pathogens (51) and many variants of one of the approximately 200 types of plant disease resistance genes of the so-called NBS-LRR classes (106). Large variation has been reported for tolerance to abiotic factors such as freezing temperatures (100), drought (103, 105), UV light (34, 157), high and low carbon dioxide (167), salt (135), metals (49, 73), or oxidative factors (2, 136). Developmental traits showing large variation include flowering time (reviewed in 74, and further discussed below), floral morphology (59), leaf morphology (128), trichome density (88), venation pattern (26), branching (6, 159, 163), plant size (6, 98, 113, 128, 159, 163), and growth rate and cell division (14). Considerable genetic variation has also been analyzed for physiological traits such as seed dormancy (5, 89, 160); circadian period (152); components of plant growth (see below) such as nitrogen use efficiency (98, 138) or plant responses to light and hormone treatments (19, 21, 36, 67, 68, 99, 165). Biochemical traits also differ among accessions, such as the content of seed lipids (110); glucosinolates (25, 70, 116); epicuticular waxes (137); seed oligosaccharides (15); phytate and phosphate (16) or anions (97); and enzyme activities for primary and secondary metabolism (115, 146). Finally, genetic variation has been detected for complex genetic mechanisms such as chiasma frequency (141), DNA methylation (27, 140), and gene expression levels (65).

Finding genetic variation for this increasing list has largely been based on more or less complex standard laboratory phenotypic assays. However, new high-throughput technologies allow automatization of highly specific assays, including morphological measurements (22) or metabolite and transcript profiling (39, 56). It is expected that these and other high-throughput technologies will facilitate and increase the power to find relevant natural genetic variation.

The Genetic Basis of Natural Variation: From Phenotype to Quantitative Trait Locus (QTL)—QTL Mapping

Once genetic variation is found among *Arabidopsis* accessions, we aim to identify how many loci account for it and where they are located in the genome (mapping).

Part of this variation is of a qualitative nature, so the phenotypes in the progeny of experimental crosses can be classified in a limited number of unambiguous discrete classes that correspond to the genotypes of a single segregating locus (monogenic). The genetic analysis of these monogenic traits is similar to that used for the study of mutants, including mapping by Mendelian linkage analysis (detection of cosegregation with markers) and allelism tests by complementation. The first example of such recessive qualitative alleles described among *Arabidopsis* accessions was probably the absence or reduction of trichomes found in a small number of accessions whose segregation was already analyzed in 1943 by Laibach (86). Now it is known that most natural variants with trichome reduction can be assigned to the *GL1* gene (M. Koornneef, unpublished data). The *erecta* mutation is present in several American and one Japanese accession, and at least three semidwarf accessions allelic to the *ga5* mutant have been found (M. Koornneef, unpublished data). Multiple resistance genes (51, 106) and glucosinolate composition (116) loci have been identified in this way. In addition, large differences for the primarily quantitative trait flowering time segregate often as qualitative (described below).

Nevertheless, most variation among accessions is of a quantitative nature due to the effects of allelic variation at several loci (multigenic), which, combined with the environmental effect, determines a continuous (quantitative) phenotypic distribution of the trait in segregating populations. The genotype at these loci cannot be directly known from the single phenotypic value of a plant, but it can only be indirectly inferred from linked marker loci. The estimation of the number and genome position of the segregating QTL in an experimental population is referred to as QTL mapping and requires the following steps: (*a*) the generation of an experimental mapping population; (*b*) its genotyping with markers throughout the genome and the phenotyping for the trait of interest; (*c*) the association analysis between phenotypic values of the trait and genotypic classes of the polymorphic markers. For the latter step specific software is available (http://mapmgr.roswellpark.org/qtsoftware.html). Thus, the number and genetic position of loci that control the trait variation in that population, their relative additive effect, the contribution of genetic interactions between loci (epistasis) and, depending on the population type, the mode of action of each QTL (dominance effects) are calculated. The number of loci identified per analysis varies from 1 to >10 (Table 1), depending on the complexity of the genetic variation under study, including parameters such as the true number of loci segregating, the relative additive effect of each QTL, and the effect of genetic interactions. In addition, this number depends on the heritability of the trait in the assay performed, i.e., the control of the environmental uniformity, the quality and density of genotypic data, the statistical method used to map QTL, and the size of the mapping population.

In principle, any experimental segregating population can be used for QTL mapping. However, there are advantages to using immortal mapping populations such as RILs or introgression lines (ILs), also referred to as near isogenic lines (NILs), backcross inbred lines (BILs), or congenic lines (CLs). Such populations are practically homozygous and therefore phenotypic values can be based on multiple

TABLE 1

Trait	Mapping crosses analyzed	Mapping population	Number of loci identified	References
Resistance and tolerance to biotic factors[a]				
Bacterial and fungi-Mendelian analyses	various	F_2/F_3	>18	(51)
Powdery mildew resistance	Col-*gl* × Kas-1	RIL	3	(165)
Xanthomonas resistance	L*er* × Col;	RIL	3	(24)
Ralstonia resistance	Col-5 × Nd-1	RIL	1	(35)
Rhizobacteria induced resistance	Col × RLD	F_2	1	(156)
Bacterial flagellin induced growth inhibition	Ws × Col; Ws × L*er*	F_2	1	(45)
Long-distance movement of viruses	C24 × Col-3; L*er* × Col	F_2, RIL	1	(30)
Insect resistance (Plutella -P- and Trichoplusia -T-)	L*er* × Col	RIL	3 (T)	(55, 72, 87)
	L*er* × Cvi	RIL	2 (P) + 3 (T)	(72)
Resistance to rabbit herbivory	L*er* × Col	RIL	6	(162)
Resistance to abiotic factors[a]				
Freezing tolerance (plant survival)	L*er* × Cvi	RIL	4	(100)
NaCl tolerance (plant fresh weight, germination)	L*er* × Col	RIL	11	(135)
Aluminium tolerance (root growth)	L*er* × Col	RIL	2	49.73
Developmental traits				
Flowering time: Mendelian genetic analyses	Li-5 × St, L*er* × Sf-2, L*er* × Sy-0	F_2/F_3, BC	3	(74, 120)
Flowering time: Quantitative genetic analyses	HM × WS	F_2	3	(79)
	H51 × L*er*	F_2/F_3	5	(32)
	Li-5 × Nantali	BC	7	(85)
	L*er* × Col	RIL	2–12	(57, 60, 113, 151, 159, 162, 163[b])
	L*er* × Cvi	RIL	8	(7, 60, 159)
	Bay-0 × Sha	RIL	5	(96)
	L*er* × Sha	RIL	6	M.E. El-Lithy, M. Koornneef & C. Alonso-Blanco, unpublished data
Inflorescence morphology	L*er* × Col	RIL	5	(159, 163)
	L*er* × Cvi	RIL	4, 7	(6, 159[b])
Floral morphology	L*er* × Col	RIL	11	(59)
Plant size	L*er* × Col	RIL	2–8	(113, 159, 163, 128[b])
	L*er* × Cvi	RIL	4	(6, 159)
	Bay-0 × Sha	RIL	4–8	(98)
Seed size	L*er* × Cvi	RIL	11	(6)
Leaf shape	L*er* × Col	RIL	16	(128)
Leaf trichome density	L*er* × Col	RIL	1	(88)
Physiological traits				
Seed dormancy	L*er* × Col	RIL	14	(161)
	L*er* × Cvi	RIL	7	(5)
Circadian period of leaf movement	L*er* × Col	RIL	2	(152)
	L*er* × Cvi	RIL	4	(152)
Tissue culture response	L*er* × Col	RIL	4	(143)

(*Continued*)

TABLE 1   (*Continued*)

Trait	Mapping crosses analysed	Mapping population	Number of loci identified	References
Water and anion content	Bay-0 × Sha	RIL	19	(97)
Nitrogen use efficiency (growth, content)	Bay-0 × Sha	RIL	18	(98)
Nitrogen response (growth)	Ler × Col	RIL	5	(138)
Cotyledon unfolding fluence responses	Ler × Col	RIL	2	(165)
	Ler × Cvi	RIL	3	(21)
Hypocotyl growth light inhibition	Ler × Cvi	RIL	12	(19, 21, 99)
	Col × Be-0	F_2	1	(127)
Hypocotyl growth hormone response	Ler × Cvi	RIL	5	(19)
Glucosinolate accumulation jasmonate response	Ler × Col	RIL	>10	(68)
Chemical contents and enzyme activities				
Glucosinolate composition and quantity	Limburg-5 × H51	F_2/F_3	1	(116)
	Ler × Col	RIL	2, >10	(25, 71, 116)
	Ler × Cvi	RIL	15	(69)
Seed oligosaccharide content	Ler × Cvi	RIL	4	(15)
Phytate and phosphate content	Ler × Cvi	RIL	5	(16)
Nitrogen/nitrate/amino acid content	Bay-0 × Sha	RIL	7–8 per trait	(98)
Anion (water, chloride, phosphate)	Bay-0 × Sha	RIL	6–8 per trait	(97)
Primary and secondary metabolism enzyme activities	Ler × Col	RIL	1–3 per enzyme	(115)
PGM activity	Ler × Cvi	RIL	6	(146)
Sugar-induced b-amylase	Ler × Col	F_2	1	(112)
Genetic mechanisms				
Cytosine methylation	Ler × Cvi	RIL	5	(140)
Viability (segregation distortion)	Nd × Ler	F_2	1	(35)
	Ler × Cvi	RIL	2	(9)

[a]The specific trait measured in tolerances to environmental factors is indicated between brackets.

[b]Different assays applied by different authors.

replicates, reducing the environmental effects and increasing the power to detect QTL. They can be analyzed in multiple environments without the need of further genotyping, and thus, the effects of each QTL in different environments can be precisely estimated and tested for QTL by environment interactions. These mapping populations can be genotyped very efficiently in *Arabidopsis* using microsatellites, AFLP markers, or any technique to detect *s*ingle-*n*ucleotide *p*olymorphisms (SNPs) such as CAPS markers. However, the availability of polymorphism and sequence data from several accessions (http://arabidopsis.org) (18, 144) is accelerating the development and implementation of high-throughput protocols to genotype microsatellites (132), indels (18), or SNPs (20).

Currently, five different RIL populations (9, 35, 94, 96, 164) are publicly available. Their usefulness has stimulated the development of many other well-genotyped RIL populations, with an increasing population size (see http://natural-eu.org). The power of this approach can be seen from the many studies that have

been performed using mainly two RIL populations that are publicly available for a few years (Table 1). Specifically, the L*er* × Cvi RIL population has been analyzed for more than 40 traits (Table 1 and references therein). Figure 2 shows a summary of the QTL detected in this cross, giving an impression of the amount of genetic variation present between Cvi and L*er* that resulted in the identification of QTL. In addition, the analysis of many traits in the same mapping population enables comparisons of QTL map positions. Although the inaccuracy of QTL mapping should be taken into account, colocation of different QTL is a first indication of pleiotropic effects, even for a priori unrelated traits. Using single-gene mutants (as shown for *FLC* and *ER*) or transgenic plants expressing only one gene from one accession in the other accession (as shown for *CRY2*) has allowed an unambiguous assignment for QTL of different traits to the same gene. Such comparative QTL mapping will be particularly useful in the near future, when high-throughput techniques are used for QTL mapping, such as metabolomic assays that simultaneously measure large numbers of metabolites (39) or transcriptomic protocols that measure gene expression at a genome-wide scale, as reported for other organisms (142).

By means of QTL mapping the position of each QTL is assigned to a genetic interval of 5 to 50 cM, depending on its relative effect and the quality of the QTL mapping assay. These genetic distances correspond, on average, to 1.2 to 12 Mb of *Arabidopsis* DNA, containing 240 to 2400 predicted open reading frames (ORFs). Therefore, after QTL mapping, ILs with single introgressions around a particular QTL can be used to confirm the effects of the QTL and to generate a new mapping population for further fine QTL mapping. Crossing such IL with the recurrent parent will derive a secondary population that avoids the complexity of multigenic inheritance because a single QTL will segregate and, in addition, will provide the polymorphisms needed around the locus under study. Alternatively, more heterogeneous second-generation populations can be derived from particular individuals of the first mapping population, such as two RILs, selected as parentals on the basis of predicted single QTL differences. To date, fine (or high-resolution) mapping has been performed to identify a limited number of *Arabidopsis* QTL involved in flowering time (37), phytate content (16), seed dormancy (L. Bentsink, unpublished data), insect resistance and glucosinolate content (80, 81).

Recently, the potential of another alternative strategy for identifying and mapping loci accounting for natural variation, so-called linkage disequilibrium (LD) mapping, is being theoretically and empirically evaluated in *Arabidopsis* and many other plants (reviewed in 20, 40). In the QTL linkage mapping described above, segregating experimental populations are developed from F1 hybrids to maximize LD between close loci, which will be reduced in further generations. In contrast, LD mapping aims to exploit the LD between very close loci, naturally persisting in *Arabidopsis* accessions as a consequence of their shared ancestry and evolutionary history. For that, a collection of practically unrelated accessions is thoroughly genotyped with markers at a very high density and phenotyped for the trait of interest. Then, marker-trait associations are directly searched, with expectations

that markers closely linked to a QTL will show significant association. Thus, the variation present in a relatively large collection of *Arabidopsis* accessions is directly and simultaneously analyzed. The potential of LD mapping will depend on the physical distance for LD to decay, which in *Arabidopsis* has been estimated to be within 250 kb (equivalent to 1 cM) (123). In addition, the persistence of LD between two markers will depend on the age of the alleles, the history of the mutations in the region, the recombination rate, and breeding system, among other things (19, 40). Despite these factors, associating traits with the actual genes causing the phenotypic differences is expected to be significant, and thus, association studies using candidate genes within a particular region where a QTL is located have a higher potential (82, 154). Therefore, LD mapping seems a promising tool to complement standard QTL mapping in *Arabidopsis*, due to its relatively small genome and high frequency of polymorphisms (20). However, there are important issues regarding statistical significance and power, population structure and selection, and experimental design that must first be explored (40).

The Molecular Basis of Natural Variation: From QTL to Quantitative Trait Nucleotide (QTN)—QTL Cloning

Once a QTL is mapped, identifying the molecular variation underlying this QTL is a major challenge. This includes the identification of the particular gene and the discovery of the nucleotide polymorphisms within the genes that determine the phenotypic differences. The difficulties behind this are shown by the very limited number of QTL that have been cloned in plant species (summarized in 126). However, *Arabidopsis* offers several specific advantages due to the efficiency of its genetics and the resources available. As with mutants, the positional cloning of QTL is an efficient strategy to pursue this goal with principles similar to map-based cloning of genes identified by induced mutations. It basically consists in the identification of closely linked recombination events, requiring analysis of a large number of segregating progeny with molecular markers covering the critical region at high density. As with mutants, unambiguously inferring the genotype of each plant from the phenotype is the main requirement to avoid misinterpretations. This is accomplished by analyzing monogenic fine mapping populations, which can be derived from ILs or NILs, as described above. In addition, the phenotyping should be done as accurately as possible, often requiring the analysis of a large progeny from particular individuals. Preselecting recombinants around the locus using molecular markers reduces the number of individuals for which this accurate phenotyping should be performed. Finding markers in the region is very much facilitated by the availability of sequence data from multiple accessions (http://www.Arabidopsis.org) (20, 144). Ultimately, the resolution of the fine mapping will depend on the number of plants analyzed. For instance, on average, 1000 *Arabidopsis* gametes (0.1 cM) will allow a resolution of 50 kb, which will contain about 10 ORFs. In addition to fine mapping, several functional strategies are available in *Arabidopsis* to select relevant candidate genes for the QTL. First, knowledge of the complete genome sequence allows the search of such candidates

on the bases of the predicted gene functions. Nevertheless, the function of many *Arabidopsis* ORFs is still unknown at the cellular and/or phenotypic level and, therefore, one cannot always find obvious candidates from the genome sequence. An additional complexity arises when gene families of tandemly repeated genes are found, as observed for three methylthioalkylmalate synthase (*MAM*) enzymes, suggested as candidates for the glucosinolate biosynthesis locus *GS-Elong* (80, 81), or three *CBF* genes found for a freezing tolerance QTL (100). Sequence comparison of several accessions, and association analysis using the sequence of numerous accessions that are phenotypically well characterized, provides a complementary tool to test candidate genes. For instance, indel polymorphisms deleting one or the other of two *MAM* genes in different accessions enabled the functional distinction of both genes (80). Further sequence analysis using population genetics parameters strongly suggested that the *MAM2* gene, only present in L*er* and not in Col, is responsible for the increased resistance of L*er* to the generalist herbivore *Spodoptera exigua* (80).

Another resource for finding candidate genes is analzying gene expression in the vicinity of the QTL. This can be done using standard assays for a limited number of candidate genes, or using high-throughput genome-wide techniques, such as microarrays (20). When the functional allelic variation results in gene expression differences, as found for several natural alleles of the *FRI* and *FLC* loci (58, 107, 148), this may give a clear indication of which is the candidate gene.

Identifying an artificially induced mutant showing phenotypic effect in the trait of interest provides a unique functional argument to select a candidate gene. The availability of T-DNA insertion mutants for almost any *Arabidopsis* gene and the efficiency of tilling procedures to identify mutations in numerous candidate genes (48) provide efficient strategies to analyze knockout phenotypes of (nearly) all genes in a QTL region. Nevertheless, most collections of mutants are in the laboratory backgrounds L*er* and Col, which do not necessarily carry functional alleles at the gene of interest and, consequently, will not always show a distinct phenotype when mutated. Therefore, loss of function mutants can also be induced by mutagenesis with standard chemical or physical agents of particular lines such as NILs carrying alleles different from the common laboratory accessions. This approach is especially useful when identifying novel alleles that are dominant over laboratory backgrounds, which suggests they are active, as shown for the *FLC* gene (107).

Ultimately, the proof for the identification of a QTL gene should come from complementation experiments by plant transformation. The transfer of an allele from one parent to the other and vice versa, or the transfer of either allele in a null background, should show the predicted effects of the QTL. Further sequencing of both alleles of the corresponding QTL gene will identify DNA polymorphisms. However, to find the precise nucleotide polymorphism underlying the QTL (or QTN) requires further work. Currently, about 15 genes accounting for natural variation for various traits have been identified in *Arabidopsis* (Table 2). Most of them participate in flowering time or pathogen-related traits, for either bacteria or insect pathogens, and encode for photoreceptors, transcription factors, R-genes,

TABLE 2 Summary of the genes accounting for natural variation identified in *Arabidopsis*, including the molecular characteristics of their natural alleles

Locus	Trait	Molecular function	Functional polymorphism	Functional alteration	References
CAL	Inflorescence morphology	MADS TF	SNP	Altered protein	(66)
FRI	Flowering	Unknown	INDELs (1-345 bp)	Truncated protein	(43, 58, 79)
FLC	Flowering	MADS TF	INDELs (1.2 to 4.2 kb)	Expression level	(43, 109)
EDI	Flowering	CRY2 photoreceptor	SNP	AA substitution	(37)
PHYA	Hypocotyl length	PHYA photoreceptor	SNP	AA substitution	(99)
PHYD	Flowering/hypocotyl length	PHYD photoreceptor	INDEL (14 bp)	Truncated protein	(12)
AOP2	Glucosinolate biosynthesis	2-oxoglutarate-dioxygenase	INDEL (5 bp)	Expression level	(71)
AOP3	Glucosinolate biosynthesis	2-oxoglutarate-dioxygenase	Unknown	Expression level	(71)
ESP	Glucosinolate hydrolysis	Epithiospecifier protein	INDELs, SNP	Expression level	(87)
MAM1	Glucosinolate biosynthesis	Methylthioalkylmalate synthase	INDELs, (several kb)	Deleted gene	(80)
MAM2	Glucosinolates/insect resistance	Methylthioalkylmalate synthase	INDELs, (several kb)	Deleted gene	(80)
RTM1	Virus resistance	Jacalin-like protein	SNP	Truncated protein	(30)
RPS2	Resistance to pseudomonas	LRR protein	INDEL (10 bp), SNP	Truncated protein	(111)
RPM1	Resistance to pseudomonas	LRR protein	INDEL (2.8 kb)	Deleted gene	(150)

or glucosinolate biosynthesis genes. Several types of molecular polymorphisms underlying the phenotypic variation have been reported, including SNPs that generate single amino acid substitutions as found for *CRY2* and *PHYA* (37, 99); small deletions that give rise to truncated proteins, as reported for the *FRI* and *PHYD* loci (12, 43, 58, 90); large deletions spanning the complete gene, as those found for *RPM1*, *MAM1*, and *MAM2* (80, 150); and large transposon-related insertions in noncoding regulatory regions, found in *FLC* (43, 109).

CONTRIBUTIONS OF THE ANALYSIS OF *ARABIDOPSIS* NATURAL VARIATION: TWO CASE STUDIES

The analysis of natural variation, from determining its genetic magnitude within *Arabidopsis* to identifying the molecular variants accounting for it, is providing unique knowledge. These contributions include the discovery of novel functions

of genes affecting a particular trait and the further characterization of previously identified genes. In addition, it is starting to make available information on the ecological function of genes at the population and species level. To illustrate these contributions, we describe the current status of the analysis of natural variation in *Arabidopsis* for two processes: the initiation of flowering and plant growth. Both aspects differ in their general complexity because the initiation of flowering is a specific developmental transition that is already well-studied in *Arabidopsis*, whereas growth might be considered at any developmental stage and has been studied less extensively in *Arabidopsis*. Due to these differences, knowledge and progress have been qualitatively different for both case studies, showing the future potential and limitations of these analyses for two different sorts of traits.

Natural Variation Affecting Flowering Time

Natural environmental conditions differ throughout the year, and therefore, the timing of the transition from the vegetative to reproductive phase of a plant is a major determinant of its reproductive success. The genetic and molecular dissection of this developmental transition is being undertaken through the analysis of artificially induced mutants in *Arabidopsis* (74, 119). These analyses have identified around 100 genes that affect this transition, and these genes are classified in different pathways that integrate the environmental signals (such as photoperiod, light intensity, light quality, or temperature) and the endogenous signals (including hormones and metabolites). Basically, four interacting pathways have been described as the photoperiod response pathway, the vernalization response pathway, the autonomous pathway, and the gibberellin pathway, whose signals ultimately regulate the expression of genes involved in flower development (74, 119).

Additionally, the timing of the flowering transition is genetically differentiated among natural populations of *Arabidopsis*, as shown by the large genetic variation observed for this trait among *Arabidopsis* accessions. For instance, when grown under long day photoperiods, some accessions will flower in three weeks, others will require four months, and some will be unable to flower unless they are treated with low, nonfreezing temperatures during several weeks (vernalization) (61; C. Alonso-Blanco, unpublished data). As indicated above, two extreme natural life cycles have been described in annual plants such as *Arabidopsis*, e.g., summer and winter–annuals which differ in the timing of germination and flowering; winter annuals having a considerably larger vegetative phase (late flowering time) than summer annuals. This variation is presumed to reflect adaptations to different environments because environmental conditions determined by biotic and abiotic factors are very different along the distribution area of *Arabidopsis* (Figure 1). The study of natural variation for flowering initiation has attracted a great deal of interest from the earliest work of *Arabidopsis* research (reviewed in 74). Much of the pioneering work aiming to estimate the number of loci determining the variation between winter and summer annuals was accomplished by Napp-Zinn, who

systematically analyzed the cross between the early flowering accession Limburg-5 and the late flowering Stockholm (summarized in 120). He identified four loci and vernalization was able to overcome the effect of the late flowering alleles. Late alleles at the *FRIGIDA* (*FRI*) locus showed the largest effect, were dominant, and interacted with the late alleles of smaller effect at the *KRYOPHILA* (*KRY*) and *JUVENALIS* (*JUV*) loci. The advent of molecular markers in combination with Mendelian genetic analyses allowed the mapping of Napp-Zinn's *FRI* locus on top of chromosome 4 (31) and the identification and location on top of chromosome 5 of another locus named *Flowering Locus C* (*FLC*) (75, 91). Late dominant alleles at these two loci interact synergistically to account for a large proportion of the late flowering of several *Arabidopsis* accessions (74). Using similar approaches, a third locus named *AERIAL ROSETTE1* (*ART1*) was identified in the middle of chromosome 5, whose dominant late allele also interacts with *FRI* and *FLC* (131).

Although classical Mendelian analysis allowed the finding and mapping of three large effect loci, most of the flowering variation among *Arabidopsis* accessions is multigenic and has required the quantitative analysis of crosses to identify and map the corresponding QTL. Currently, seven different crosses have been analyzed in this way (see Table 1), two of them involving an early flowering parental accession and a late one (32, 85), whereas the rest were derived from two early flowering accessions (7, 57, 96). The latter crosses show considerable transgression despite the fact that the parents differed little in their flowering behavior, in some cases providing as much variation as the crosses obtained from phenotypically distinct parents (7, 96). Five to twelve QTL per cross have been detected in the crosses derived from early accessions, a number that is similar to, if not larger than, the five to seven QTL detected in crosses involving late flowering accessions (Table 1). Comparing the QTL map positions identified in the various crosses aiming to determine whether they are the same or different loci is not straightforward. Due to heterogeneity, it is not possible to decide if QTL found at similar map positions in different crosses recognize similar or different effect alleles that might provide a natural allelic series. Conservatively, a minimum number of 14 different QTL have been identified, accounting for the flowering time differences between *Arabidopsis* accessions. Some of them show different effects in different environments, measured by significant QTL by environment interactions that have been estimated in those analyses performed in multiple controlled environments differing in photoperiod or vernalization treatment (7, 32, 57, 96, 151, 163). Significant interactions among QTL have been detected in several cases as well (7, 32, 60, 85, 159). Among these QTL, the strong effect loci *FRI* and *FLC* were clearly detected in several crosses. In addition, three other QTL identified in the Landsberg *erecta* (L*er*) × Cape Verde Islands (Cvi) mapping population and one QTL in the L*er* × Columbia (Col) cross were confirmed by characterizing introgression lines carrying particular alleles at a single genomic region into the reference genetic backgrounds, L*er* or Col (7, 77). However, most of the smaller effect QTL described so far await confirmation. Developing other introgression lines in

reference backgrounds will allow not only confirmation and further characterization but also complementation tests aiming to determine whether different introgressed alleles belong to the same locus. Comparing map positions of QTL and known flowering genes has enabled discovery of several candidate genes for most of the QTL described (7, 60, 96, 159). However, further molecular, physiological, and genetical analysis of the QTL is required to evaluate those candidate genes on the basis of functional information.

The molecular characterization of the first loci causing natural variation for flowering has begun with the isolation of the large effect loci *FRI* and *FLC* (58, 107) (see Table 2). Isolating these loci has identified two novel genes involved in the regulation of flowering, which could not be found by mutant analyses because the common laboratory early flowering strains carry loss of function alleles at *FLC* and/or *FRI*, and given their synergistic interaction, further loss of function alleles could not be easily detected in these accessions. *FRI* encodes a protein with no significant homology to previously identified proteins (58), whereas *FLC* encodes a transcription factor of the MADS box family (107). FRI positively regulates the expression of *FLC*, whereas the FLC protein negatively regulates the expression of other transcription factors involved in the regulation of flowering such as *SOC1* (108). Further molecular characterization of *FLC* is showing that it is a central integrator of flowering signals from the autonomous and the vernalization pathways through a complex regulation of its expression (148).

In addition, the *early*, *day-length insensitive* (*EDI*) locus was recently isolated, being the first *Arabidopsis* locus identified as a QTL that was cloned (37). *EDI* affects the flowering response to photoperiod and is the blue light photoreceptor gene *cryptochrome 2* (*CRY2*), which is mutated in the photoperiod pathway mutant *fha* (119). A single amino acid substitution in the CRY2 protein of Cvi compared to Ler affects the CRY2 light-induced down regulation in short days and alters the CRY2 flowering promotion.

Natural variation has also been found in the phytochrome genes *PHYD* and *PHYA*, which are other photoreceptors known to affect flowering (12, 99), by using different candidate gene approaches. Based on the identification of the Wassilewskija (WS) accession as a bearer of a deletion in *PHYD*, it was demonstrated that the lack of function of this phytochrome determines several pleiotropic phenotypes including early flowering (12). In contrast, based on detailed phenotypic analyses, the Lm-2 accession was identified as resembling plants carrying the loss of function alleles of *PhyA* (99). It was demonstrated that a single amino acid substitution in the PhyA protein of Lm-2 reduces the far-red light response measured by hypocotyl elongation, compared to that of Col. Starting from a similar reverse approach, the flowering gene *FWA* has been suggested to account for the different flowering response to demethylation treatments shown by Ler, Col, and C24 (44). Two amino acid substitutions in a very conserved region of this transcription factor were found in C24 compared to Ler, suggesting a C24 loss of function. However, this needs to be proven because any downstream *FWA* target might be nonfunctional in C24. Analysis of these last three genes illustrates

the power of selecting candidate genes on the basis of sequence variation and/or well-defined phenotypic differences between accessions.

Identifying genes responsible for the natural variation in flowering time enabled the characterization of the natural molecular variation present in these genes in relation to their function and in relation to the ecological and evolutionary forces that maintain it. Nine different *FRI* loss of function, probably null, alleles have been identified produced by insertion/deletion events of 1 to 376 bps (see Table 2) (43, 58, 90). Two different alleles with reduced function of *FLC* have been identified, generated by the insertion of transposon-related sequences in the first intron (Table 2) (43, 109). Thus, the early flowering behavior of *Arabidopsis* summer annual accessions derived from winter annual ancestor accessions by loss of function of *FRI* and/or *FLC*. The large number of rare alleles in the *FRI* gene suggests that early flowering has derived in multiple independent occasions. This flowering variation might be maintained as local adaptations to different environments.

Despite this enlightening contribution of the *FRI* and *FLC* analysis, there is still much flowering genetic variation among accessions carrying the same functional classes of *FRI* and *FLC* alleles. This additional flowering genetic variation is determined by the large number of QTL identified so far. To fully understand the flowering variation requires the identification of many other naturally varying loci and the simultaneous analysis of all these genes. In addition, the interactions among the various loci such as that described for *FRI* and *FLC*, as well as for several other loci, need to be implemented (7, 60, 75, 91, 131, 159). Taking into account the genetic interactions will help in understanding the patterns of molecular variation. For instance, unlike the *FRI* alleles, the *FLC* loss of function alleles are not null alleles. This might be related to their genetic interaction determined by the functional relationship between these genes, *FLC* being downstream in the regulatory pathway (58, 107, 148).

A better understanding of the functional, ecological, and evolutionary meaning of flowering variation will require the simultaneous analysis of multiple traits because genes affecting flowering time also affect other vegetative and reproductive traits. The naturally occurring flowering variation has been studied in relation to other life history traits such as morphological and physiological plasticity (166), plant growth (113), seed size and seed number (6), inflorescence morphology (159), circadian rhythm (152), drought resistance (103), and resistance and tolerance to herbivory (162). These reports show the large amount of pleiotropic effects of the primarily so-called flowering loci, and the various trade-offs between flowering time variation and other adaptation mechanisms. Knowing the distinct pleiotropic effects of each gene will assist the understanding of its pattern of variation, like for *FRI* and *FLC*, which have been also associated with circadian rhythm (152) or dehydration avoidance (103).

Ultimately, understanding the ecological significance of the flowering variation for plant adaptation will require its analysis under well-defined natural field environments because so far most of the work has been performed under laboratory

conditions. However, natural environments are far more complex due to the number of environmental factors involved and to spatial and temporal heterogeneity. This is illustrated in the first analysis of *Arabidopsis* flowering variation under natural conditions, which was recently reported (162, 163). The authors found considerable QTL by environment interactions, several flowering QTL appearing only under controlled conditions with their effects overridden in natural environments. In addition, it is suggested that genes determining variation in ecological relevant conditions might not be identified under laboratory conditions because various QTL were identified only under natural conditions. Further field experiments will identify the precise environments in which variation at particular loci plays a role and thus, the relevant factors shaping this variation.

Flowering variation is an adaptive mechanism common to many annual plants including wild and domesticated species. To know if the genes accounting for flowering variation under particular environments are conserved among plant species will be of great relevance, from basic and applied perspectives. Currently, flowering variation is mainly studied among varieties of cultivated species. This variation present in cultivated species probably had a natural origin, but it involves different evolutionary forces to maintain it due to the participation of artificial selection during domestication and breeding. Nevertheless, its comparison with the variation observed in *Arabidopsis* is allowing a first estimation of this conservation. Three major QTL were recently isolated in rice, *Hd1*, *Hd3a*, and *Hd6*, involved in photoperiod sensitivity (reviewed in 54). *Hd1* and *Hd3a* are orthologues of the *Arabidopsis CO* and *FT* flowering genes, respectively, both implicated in the photoperiod pathway, whereas *Hd6* encodes the CK2 casein kinase (references in 54). Characterizing *FLC*-like genes in *Brassica napus* (153) and *B. rapa* (145) suggests that *FLC* orthologues account for the different vernalization response among cultivars of *Brassica* species. Association analysis of the candidate *Dwarf8* gene, orthologue of the *Arabidopsis* flowering related gene *GAI*, suggests that this gibberellin signaling gene accounts for part of the flowering time variation found in maize (154). Similar analyses in *Brassica nigra* suggest that an orthologue of the *Arabidopsis CO* flowering gene underlies a flowering QTL (82). These results show that there is considerable conservation of the molecular mechanisms that regulate flowering among plant species. In addition, it suggests that among species of the crucifer family, there is partial conservation of the genes accounting for within species variation. In contrast genome comparisons between rice and *Arabidopsis* have failed to identify rice orthologues of several *Arabidopsis* vernalization pathway genes (54). Because vernalization has not been reported in rice, it is suggested that these genes have been lost in rice during evolution (54). Therefore, substantial differences among species will also occur depending on the flowering mechanisms and genes available in each species. The isolation of more QTL in *Arabidopsis*, crop plants, and wild species, combined with their detailed analysis described above, will allow an integrative understanding of the mechanisms regulating flowering at the species level, including comprehension of the evolutionary forces that maintain flowering variation and its relevance for plant adaptation.

Natural Variation Affecting Plant Growth

Although *Arabidopsis* has been used mainly for studies related to developmental and general biology, it will also be useful for dissecting complex traits that are essential for plant breeding and ecology, of which growth and reproductive success are probably the most important traits.

Plant growth can be defined in an absolute manner in many different ways, depending on different parameters used to measure it, such as size, fresh weight, dry weight, reproductive yield, or plant height. In addition, growth can be estimated using a time-component, which provides a more integrative parameter referred to as relative growth rate (RGR), defined as the relative increase in a certain parameter per time-unit. Absolute growth and RGR are the integrative results of a complex set of parameters, including directly related components such as carbon fixation, water and mineral uptake, partitioning of reserves and respiration, and more indirect components like stress tolerances, senescence, flower formation, and so on. Thus, growth and RGR must be highly multigenic traits determined by the numerous genes affecting one or several of the various components, and large environmental effects will strongly affect its output. Due to this complexity, measuring overall growth will often lead to low heritabilities and the detection of relatively small effect loci, from which many might easily escape detection. However, measuring particular components or subprocesses contributing to plant growth will increase the power to detect loci affecting this trait.

Arabidopsis natural genetic variation has been reported for overall growth parameters and for several of its components. Variation has been analyzed for various traits related to overall growth of the plants such as rosette diameter (128), plant height (130), fresh or dry weight (1, 93, 98), number of lateral branches (130), and leaf shape (128). The range of variation for some of these parameters was substantial, such as for dry weight of the whole plant at final harvest (93) or plant height and number of basal branches (130), where a sixfold variation or a more than tenfold variation were reported, respectively. However, lower variation has been reported for other traits, such as root elongation rate (fourfold variation) (14) or above-ground vegetative biomass (twofold variation) (1).

In addition, genetic variation has also been described using a noninvasive method to estimate plant growth rate of *Arabidopsis* plants by determining the area of the rosette, using digital imaging (92; M.E. El-Lithy, M.E. Clerkx, G. Ruys, M. Koornneef, D. Vreugdenhil, unpublished data).

Several integrative parameters related to plant growth and water use efficiency have been analyzed such as carbon stable isotope discrimination $\delta^{13}C$, which measures transpiration efficiency, or chlorophyll fluorescence. Considerable variation among accessions has been found for carbon stable isotope discrimination $\delta^{13}C$ (101, 103), which was correlated with flowering time (103). In contrast, chlorophyll fluorescence, which is a nondestructive measurement for plant performance, shows more limited variation under normal conditions (M.E. El-Lithy, J.F.H. Snel, M. Koornneef, D. Vreugdenhil, unpublished data).

Parameters estimating overall growth have been correlated with particular components. For instance, a negative correlation has been found between maximum plant size (above-ground vegetative biomass) and fecundity (number of seeds per plant weight) (1), or between carbon stable isotope discrimination $\delta^{13}C$ and flowering time (103). In addition, a positive correlation was found between root elongation rate and the activity of the cyclin-dependent kinase (CDKA), suggesting that cell-cycle activity might be an important determinant for growth differences of roots (14).

In addition, genetic variation among *Arabidopsis* accessions has also been analyzed for various biochemical and metabolic components of plant growth such as plant mineral contents (29, 97, 98, 121, 138) and seed sugar contents (15). Large variation has been described for phosphate acquisition efficiency (PAE) (121), revealing different underlying mechanism affecting PAE: root and root hair morphologies, phosphate uptake kinetics, organic acid release, and rhizosphere acidification (121).

Finally, the genetic variation affecting plant growth has been analyzed in various environments differing in mineral contents, mainly phosphorus and nitrogen (29, 98, 138). Thus, the availability of P has profound effects on root morphological characteristics, e.g., root length, the number of lateral roots, root hair length, and density. Significant variation was found for the growth responses of these parameters to P-availability, comparing P-starved and control plants (29). Similarly, it has been reported that different sources of nitrogen strongly affect dry matter accumulation (98, 138).

Despite this considerable variation, little is known about the genetic basis of plant growth in *Arabidopsis*. Recently, several QTL mapping efforts focusing on growth-related traits were initiated using the resources available in *Arabidopsis*. The traits that were determined and for which loci were mapped are descriptive traits, not including a time-component.

QTLs for plant height have been mapped in two RIL populations (see Table 1). Not surprisingly, the major QTL for this trait colocated with the position of the *ERECTA* gene in both RIL populations (6, 159) because both populations involve Landsberg carrying the *erecta* mutation, as parental line. In addition, 2 to 4 other loci affecting plant height were found in both populations (Table 1). Above-ground biomass was estimated by determining dry weight (98) or rosette area (128, 159). For these traits major QTL were also found around the *ERECTA* locus in the populations having L*er* as one of the parents, in agreement with the pleiotropic effect of this gene. Five additional major QTL affecting the rosette area were found on chromosomes 3, 4, and 5 (6, 159, 163), and eight loci affecting above-ground dry matter (98).

QTL mapping experiments have been also performed for several biochemical and metabolic components of plant growth. For instance, sugar content has been analyzed in mature seeds in the L*er* × Cvi RIL population revealing a major QTL affecting sucrose, stachyose, and raffinose content, which suggest pleiotropic effects by altering a step in the conversion from sucrose into raffinose and stachyose (15).

In addition, Mitchell-Olds & Pedersen (115) studied the activities of several enzymes involved in primary and secondary metabolism in leaves of the L*er* × Col RIL population. For 7 out of 10 enzymes studied one or more QTL were mapped. Some of these mapped very close to enzyme-coding loci. They also mapped a locus that might be a joint regulator of activities of three different enzymes, all involved in primary carbon metabolism. PGM activities have been also analyzed in seedlings of the L*er* × Cvi RIL population (146). Besides QTL colocating with all five putative PGM genes, other loci affecting PGM activities were also mapped, pointing toward possible regulating genes. QTL have also been mapped for leaf phosphate content in several populations (see Table 1) (16, 97). Six loci were identified on four chromosomes in the Bay × Sha population, two of them colocating with loci affecting chloride levels and/or water content, suggesting a role in osmotic regulations. In the L*er* × Cvi population QTL were mapped for phosphate and phytate (inositol-1,2,3,4,5,6-hexakisphosphate, the major storage form of P in plants) in leaves and seeds (16). A major QTL at the top of chromosome 3 explained most of the variation observed in seeds and leaves, both for phosphate and phytate, which again suggested pleiotropic effects.

QTL for plant growth parameters have been mapped in environments with different nitrogen sources (98, 138). Rauh et al. (138) studied biomass and root-shoot ratios in relation to nitrogen type (ammonium or nitrate) and availability using the L*er* × Col RIL population. A total minimum number of 14 QTL was identified affecting the various traits in four N-treatments. Interestingly, most QTL were not shared across treatments indicating strong QTL by environment interactions. Similarly, Loudet et al. (98) analyzed a large RIL population (415 lines) derived from the cross Bay × Sha to map QTL for N-related traits, by growing the plants at high or low nitrate supply. They found at least 18 loci affecting one or more traits, and again, the regulation of the traits measured was mostly specific for the N-treatments.

These works show the feasibility of identifying QTL affecting plant growth and/or plant growth components, and inferring putative roles of these QTL on the basis of pleiotropy. However, in contrast to the analysis of natural variation for flowering time, the fine mapping of these QTL has just started, and only candidate genes have been reported without definite proof for any of these. Thus, many candidate genes are available based on colocation of QTL with genes involved in the processes studied. For instance, as described above, QTL affecting enzyme activities (115, 146) often colocate with the position of the structural genes, which makes it likely that some QTL are due to different activities of the structural genes, either by differences in the promoter region, resulting in variation in transcription, or in activities of the resulting enzyme. In addition, several QTL affecting plant growth parameters such as total biomass and root-shoot ratios, have been colocated with the positions of genes associated with nitrogen metabolism (98, 138) e.g., a cytosolic glutamine synthase gene, or with developmental genes, like the *ERECTA* gene, suggesting these as candidates. It is expected that the combination of approaches described before will soon identify the molecular bases for this important and complex variation.

CONCLUSIONS

Natural variation is becoming an important tool for functional biology, and it is at the core of evolutionary biology and plant breeding (20). Currently, the large amount of genetic variation found among *Arabidopsis* accessions collected at different regions of the world can be efficiently analyzed up to the molecular identification of the genes and polymorphism underlying QTL of relative large effect. This has become feasible thanks to the genetic features of *Arabidopsis*, such as short generation time, autogamy, and the small and low complexity genome, combined with the unique tools developed by the *Arabidopsis* research community. The traits analyzed thus far deal with important adaptive plant characteristics such as seed dormancy, flowering, and tolerance to abiotic and biotic factors, and are relevant in ecology and plant breeding. It is expected that the methods and knowledge generated in *Arabidopsis* will assist to speed up the analysis of similar traits and genes in other plants, including wild and domesticated crop species. Current *Arabidopsis* QTL analyses enable the study of loci with small effects, due to the experimental control of the environmental conditions and the use of genetically well-defined *Arabidopsis* lines. It is expected that the application of these functional resources will soon lead to the identification and characterization of small effect alleles, which will be relevant to understand and artificially fine-tune the networks of plant gene regulation.

ACKNOWLEDGMENTS

The *Annual Review of Plant Biology* is online at http://plant.annualreviews.org

LITERATURE CITED

1. Aarssen LW, Clauss MJ. 1992. Genotypic variation in fecundity allocation in *Arabidopsis thaliana*. *J. Ecol.* 80:109–14

2. Abarca D, Roldan M, Martin M, Sabater B. 2001. *Arabidopsis thaliana* ecotype Cvi shows an increased tolerance to photo-oxidative stress and contains a new chloroplastic copper/zinc superoxide isoenzyme. *J. Exp. Bot.* 52:1417–25

3. Abbot RJ, Gomes MF. 1989. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* 62:411–18

4. Aguadé M. 2001. Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 18:1–9

5. Alonso-Blanco C, Bentsink L, Hanhart CJ, Blankestijn-deVries H, Koornneef M. 2003. Analysis of natural allelic variation at seed dormancy loci *of Arabidopsis thaliana*. *Genetics* 164:711–29

6. Alonso-Blanco C, Blankenstijn-de Vries H, Hanhart CJ, Koornneef M. 1999. Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 96:4710–17

7. Alonso-Blanco C, El-Assal SE-D, Coupland G, Koornneef M. 1998. Analysis of natural allelic variation at flowering time loci in the Landsberg *erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. *Genetics* 149:749–64

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

9. Alonso-Blanco C, Peeters AJM, Koornneef M, Lister C, Dean C, et al. 1998. Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. *Plant J.* 14:259–71

10. Al-Shehbaz IA, O'Kane SL. 2002. Taxonomy and phylogeny of Arabidopsis (Brassicaceae). See Ref. 149a, pp. 22

11. *Arabidopsis* Genome Initiative (AGI). 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–15

12. Aukerman MJ, Hirschfeld M, Wester L, Weaver M, Clack T, et al. 1997. A deletion in the *PHYD* gene of the *Arabidopsis* Wassilewskija ecotype defines a role for phytochrome D in red/far-red light signalling. *Plant Cell* 9:1317–26

13. Barth S, Melchinger AE, Lübberstedt T. 2002. Genetic diversity in *Arabidopsis thaliana* L. Heynh. investigated by cleaved amplified polymorphic sequence (CAPS) and inter-simple sequence repeat (ISSR) markers. *Mol. Ecol.* 11:495–505

14. Beemster GTS, De Vusser K, De Tavernier E, De Bock K, Inzé D. 2002. Variation in growth rate between *Arabidopsis* ecotypes is correlated with cell division and A-type cyclin-dependent kinase activity. *Plant Physiol.* 129:854–64

15. Bentsink L, Alonso-Blanco C, Vreugdenhil D, Tesnier KJY, Groot SPC, Koornneef M. 2000. Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. *Plant Physiol.* 124:1595–604

16. Bentsink L, Yuan K, Koornneef M, Vreugdenhil D. 2003. The genetics of phytate and phosphate accumulation in seeds and leaves of *Arabidopsis thaliana*, using natural variation. *Theor. Appl. Genet.* 106:1234–43

17. Bergelson J, Stahl E, Dudek S, Kreitman M. 1998. Genetic variation within and among populations of *Arabidopsis thaliana*. *Genetics* 148:1311–23

18. Borevitz JO, Liang D, Plouffe D, Chang H-S, Zhu T, et al. 2003. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* 13:513–23

19. Borevitz JO, Maloof JN, Lutes J, Dabi T, Redfern JL, et al. 2002. Quantitative trait loci controlling light and hormone response in two accessions of *Arabidopsis thaliana*. *Genetics* 160:683–96

20. Borevitz JO, Nordborg M. 2003. The impact of genomics on the study of natural variation in *Arabidopsis*. *Plant Physiol.* 132:718–25

21. Botto JF, Alonso-Blanco C, Garzarón I, Sánchez RA, Casal JJ. 2003. The Cvi allele of cryptochrome 2 enhances cotyledon unfolding in the absence of blue light in *Arabidopsis*. *Plant Physiol.* 133:1547–56

22. Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, et al. 2001. Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* 13:1499–10

23. Breyne P, Rombaut D, van Gysel A, van Montagu M, Gerats T. 1999. AFLP analysis of genetic diversity within and

between *Arabidopsis thaliana* ecotypes. *Mol. Gen. Genet.* 261:627–34

24. Buell CR, Somerville SC. 1997. Use of *Arabidopsis* recombinant inbred lines reveals a monogenic and a novel digenic resistance mechanism to *Xanthomonas campestris* pv *campestris*. *Plant J.* 12:21–29

25. Campos de Quiros H, Magrath R, McCallum D, Kroymann J, Schnabelrauch D, et al. 2000. α-Keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana*. *Theor. Appl. Genet.* 101:429–37

26. Candela M, Martinez-Laborda A, Micol JL. 1999. Venation pattern formation in *Arabidopsis thaliana* vegetative leaves. *Dev. Biol.* 205:205–16

27. Cervera MT, Ruiz-Garcia L, Martinez-Zapater JM. 2002. Analysis of DNA methylation in *Arabidopsis thaliana* based on methylation–sensitive AFLP markers. *Mol. Genet. Genomics* 268: 543–52

28. Chen ZJ, Comai L, Pikaard CS. 1998. Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. *Proc. Natl. Acad. Sci. USA* 95:14891–96

29. Chevalier F, Pata M, Nacry P, Doumas P, Rossignol M. 2003. Effects of phosphate availability in the root system architecture: large-scale analysis of the natural variation between Arabidopsis accessions. *Plant Cell Environ.* 26:1839–50

30. Chisholm ST, Mahajan SK, Whitham S, Yamamoto ML, Carrington JC. 2000. Cloning of the Arabidopsis *RTM1* gene, which controls restriction of long-distance movement of tobacco etch virus. *Proc. Natl. Acad. Sci. USA* 97:489–94

31. Clarke JH, Dean C. 1994. Mapping *FRI*, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 242:81–89

32. Clarke JH, Mithen R, Brown JKM, Dean C. 1995. QTL analysis of flowering time in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 248:555–64

33. Clauss MJ, Mitchell-Olds T. 2003. Population genetics of tandem trypsin inhibitor genes in *Arabidopsis* species with contrasting ecology and life history. *Mol. Ecol.* 12:1287–99

34. Cooley NM, Higgins JT, Holmes MG, Attridge TH. 2001. Ecotypic differences in responses of *Arabidopsis thaliana L.* to elevated polychromatic UV-A and UV-B+A inhibition and growth rate. *J. Photochem. Photobiol.* 60:143–50

35. Deslandes L, Pileur F, Liaubet L, Camut S, Can C, et al. 1998. Genetic characterization of *RRS1*, a recessive locus in *Arabidopsis thaliana* that confers resistance to the bacterial soilborne pathogen *Ralstonia solanacearum*. *Mol. Plant Microbe Interact.* 11:659–67

36. Eichenberg K, Hennig L, Martin A, Schäfer E. 2000. Variation in dynamics of phytochrome A in *Arabidopsis* ecotypes and mutants. *Plant Cell Environ.* 23:311–19

37. El-Assal SE-D, Alonso-Blanco C, Peeters AJM, Raz V, Koornneef M. 2001. A QTL for flowering time in *Arabidopsis* reveals a novel allele of CRY2. *Nat. Genet.* 29:435–40

38. Erschadi S, Haberer G, Schoninger M, Torres-Ruiz RA. 2000. Estimated genetic diversity of *Arabidopsis thaliana* ecotypes with amplified fragment length polymorphisms (AFLP). *Theor. Appl. Genet.* 100:633–40

39. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. 2000. Metabolite profiling for functional genomics. *Nat. Biotechnol.* 18:1157–61

40. Flint-Garcia S, Thornsberry JM, Buckler ES. 2003. Structure and linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54:357–74

41. Frank MJ, Preuss D, Mack A, Kuhlmann TC, Crawford NM. 1998. The Arabidopsis transposable element *Tag1* is widely

distributed among *Arabidopsis* ecotypes. *Mol. Gen. Genet.* 257:478–84

42. Fuchs H, Sacristan MD. 1996. Identification of a gene in *Arabidopsis thaliana* controlling resistance to clubroot (*Plasmodiophora brassicae*) and characterization of the resistance response. *Mol. Plant Microbe Interact.* 9:91–97

43. Gazzani S, Gendall AR, Lister C, Dean C. 2003. Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol.* 132:1107–14

44. Genger RK, Peacock WJ, Dennis ES, Finnegan EJ. 2003. Opposing effects of reduced DNA methylation on flowering time in Arabidopsis thaliana. *Planta* 216:461–66

45. Gómez-Gómez L, Felix G, Boller T. 1999. A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*. *Plant J.* 18:277–84

46. Hall SE, Kettler G, Preuss D. 2003. Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res.* 13:195–205

47. Hardtke CS, Muller J, Berleth T. 1996. Genetic similarity among *Arabidopsis thaliana* ecotypes estimated by DNA sequence comparison. *Plant Mol. Biol.* 32:915–22

48. Henikoff S, Comai L. 2003. Single-nucleotide mutations for plant functional genomics. *Annu. Rev. Plant Biol.* 54:375–401

49. Hoekenga OA, Vision TJ, Shaff JE, Monforte AJ, Lee GP, et al. 2003. Identification and characterization of aluminium tolerance loci in *Arabidopsis* (Landsberg *erecta* × Columbia) by quantitative trait locus mapping. A physiologically simple but genetically complex trait. *Plant Physiol.* 132:936–48

50. Hoffmann MH. 2002. Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *J. Biogeogr.* 29:125–34

51. Holub EB. 2001. The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2:516–27

52. Innan H, Tajima F, Terauchi R, Miyashita NT. 1996. Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana*. *Genetics* 143:1761–70

53. Innan H, Terauchi R, Miyashita NT. 1997. Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. *Genetics* 146:1441–52

54. Izawa T, Takahashi Y, Yano M. 2003. Comparative biology comes into bloom: genomics and genetic comparison of flowering pathways in rice and Arabidopsis . *Curr. Opin. Plant Biol.* 6:113–20

55. Jander G, Cui JP, Nhan B, Pierce NE, Ausubel FM. 2001. The *TASTY* locus on chromosome 1 of Arabidopsis affects feeding of the insect herbivore *Trichoplusia ni*. *Plant Physiol.* 126:890–98

56. Jansen RC, Nap JP. 2001. Genetical genomics: the added value from segregation. *Trends Genet.* 17:388–91

57. Jansen RC, Van Ooijen JW, Stam P, Lister C, Dean C. 1995. Genotype by environment interaction in genetic mapping of multiple quantitative trait loci. *Theor. Appl. Genet.* 91:33–37

58. Johanson U, West J, Lister C, Michaels SD, Amasino RM, et al. 2000. Molecular analysis of *FRIGIDA*, a major determinant of natural varition in Aradopsis flowering time. *Science* 290:344–47

59. Juenger T, Purugganan MD, Mackay TFC. 2000. Quantitative trait loci for floral morphology in *Arabidopsis thaliana*. *Genetics* 156:1379–92

60. Juenger TE, Sen S, Stowe KA, Simms EL. 2003. Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica.* In press

61. Karlsson BH, Sills GR, Nienhuis J. 1993. Effects of photoperiod and vernalization on the number of leaves at flowering in

32 *Arabidopsis thaliana* (*Brassicaceae*) ecotypes. *Am. J. Bot.* 80:646–48

62. Kawabe A, Innan H, Terauchi R, Miyashita NT. 1997. Nucleotide polymorphism in the acidic chitinase locus (ChiA) region of the wild plant *Arabidopsis thaliana*. *Mol. Biol. Evol.* 14:1303–15

63. Kawabe A, Miyashita NT. 1999. DNA variation in the basic chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. *Genetics* 153:1445–53

64. Kawabe A, Yamane K, Miyashita NT. 2000. DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. *Genetics* 156:1339–47

65. Kehoe DM, Villand P, Somerville S. 1999. DNA microarrays for studies of higher plants and other photosynthetic organisms. *Trends Plant Sci.* 4:38–41

66. Kempin SA, Savidge B, Yanofsky MF. 1994. Molecular basis of the *cauliflower* phenotype in *Arabidopsis*. *Science* 267:522–25

67. King JJ, Stimart DP. 1998. Genetic analysis of variation for auxin-induced adventitious root formation among eighteen ecotypes of *Arabidopsis thaliana* (L.) Heynh. *J. Hered.* 89:481–87

68. Kliebenstein DJ, Figuth A, Mitchell-Olds T. 2002. Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*. *Genetics* 161:1685–96

69. Kliebenstein DJ, Gershenzon J, Mitchell-Olds T. 2001. Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics* 159:359–70

70. Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, et al. 2001. Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* 126:811–25

71. Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T. 2001. Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell* 13:681–93

72. Kliebenstein DJ, Pedersen D, Barker B, Mitchell-Olds T. 2002. Comparative analysis of quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in *Arabidopsis thaliana*. *Genetics* 161:325–32

73. Kobayashi Y, Koyama H. 2002. QTL analysis of Al tolerance in recombinant inbred lines of *Arabidopsis thaliana*. *Plant Cell Physiol.* 43:1526–33

74. Koornneef M, Alonso-Blanco C, Peeters AJM, Soppe W. 1998. Genetic control of flowering time in *Arabidopsis*. *Annu. Rev. Plant Physiol.* 49:345–70

75. Koornneef M, Blankestijn-de Vries H, Hanhart CJ, Soppe W, Peeters AJM. 1994. The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg *erecta* wild-type. *Plant J.* 6:911–19

76. Koornneef M, Fransz P, de Jong H. 2003. Cytogenetic tools for *Arabidopsis thaliana*. *Chromosome Res.* 11:183–94

77. Koumproglou R, Wilkes TM, Wang XY, Beynon J, et al. 2002. STAIRS: a new genetic resource for functional genomic studies of Arabidopsis. *Plant J.* 31:355–64

78. Kover PX, Schaal BA. 2002. Genetic variation for disease resistance and tolerance among *Arabidopsis thaliana* accessions. *Proc. Natl. Acad. Sci. USA* 99:11270–74

79. Kowalski SP, Lan TH, Feldmann KA, Paterson AH. 1994. QTL mapping of naturally-occurring flowering time of *Arabidopsis thaliana*. *Mol. Gen. Genet.* 245:548–55

80. Kroymann J, Donnerhacke S, Schnabelauch D, Mitchell-Olds T. 2003. Evolutionary dynamics of an *Arabidopsis*

insect resistance QTL. *Proc. Natl. Acad. Sci. USA.* 10.1073/pnas.1734046100 (Colloquium) 1–6

81. Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, et al. 2001. A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiol.* 127:1077–88

82. Kruskopf-Osterberg M, Shavorskaya O, Lascoux M, Lagercrantz U. 2002. Naturally occurring indel variation in the *Brassica nigra COL1* gene is associated with variation in flowering time. *Genetics* 161:299–306

83. Kuittinen H, Agaudé M. 2000. Nucleotide variation at the chacone isomerase locus in *Arabidopsis thaliana*. *Genetics* 155:863–72

84. Kuittinen H, Mattila A, Savolainen O. 1997. Genetic variation at marker loci and in quantitative traits in natural populations in *Arabidopsis thaliana*. *Heredity* 79:144–52

85. Kuittinen H, Sillanpaa MJ, Savolainen O. 1997. Genetic basis of adaptation: flowering time in *Arabidopsis thaliana*. *Theor. Appl. Genet.* 95:573–83

86. Laibach F. 1943. *Arabidopsis thaliana* (L.) Heynh. als Object für genetische und entwicklungs-physiologische Untersuchungen. *Bot. Arch.* 44:439–55

87. Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein DJ, Gershenzon J. 2001. The Arabidopsis epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. *Plant Cell* 13:2793–807

88. Larkin JC, Young N, Prigge M, Marks MD. 1996. The control of trichome spacing and number in Arabidopsis. *Development* 122:997–1005

89. Lawrence MJ. 1976. Variations in natural populations of *Arabidopsis thaliana* (L.) Heynh. In *The Biology and Chemistry of the Cruciferae*, ed. JG Vaughan, AJ Macleod, BMG Jones, pp. 167–90. London/New York/San Francisco: Academic

90. Le Corre V, Roux F, Reboud X. 2002. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Mol. Ecol. Evol.* 19:1261–71

91. Lee I, Michaels SD, Masshardt AS, Amasino RM. 1994. The late-flowering phenotype of *FRIGIDA* and mutations in *LUMINIDEPENDENS* is suppressed in the Landsberg *erecta* strain of Arabidopsis. *Plant J.* 6:903–9

92. Leister D, Varotto C, Pesaresi P, Niwergall A, Salamini F. 1999. Large-scale evaluation of plant growth in Arabidopsis thaliana by non-invasive image analysis. *Plant Physiol. Biochem.* 37:671–78

93. Li B, Suzuki J-I, Hara T. 1998. Latitudinal variation in plant size and relative growth rate in *Arabidopsis thaliana*. *Oecologia* 115:293–301

94. Lister C, Dean C. 1993. Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* 4:745–50

95. Loridon K, Cournoyer B, Goubely C, Depeiges A, Picard G. 1998. Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of *Arabidopsis thaliana*. *Theor. Appl. Genet.* 97:591–604

96. Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F. 2002. Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theor. Appl. Genet.* 104:1173–84

97. Loudet O, Chaillou S, Krapp A, Daniel-Vedele F. 2003. Quantitative trait loci analysis of water and anion contents in interaction with nitrogen availability in *Arabidopsis thaliana*. *Genetics* 163:711–22

98. Loudet O, Chaillou S, Merigout P,

Talbotec J, Daniel-Vedele F. 2003. Quantitative trait loci analysis of nitrogen use efficiency in *Arabidopsis*. *Plant Physiol.* 131:345–58

99. Maloof JN, Borevitz JO, Dabi T, Lutes J, Nehring RB, et al. 2001. Natural variation in light sensitivity of Arabidopsis. *Nat. Genet.* 29:441–46

100. Martinez-Zapater JM, Gomez-Mena C, Medina J, Llorente F, Salinas J, et al. 2002. Naturally-occurring variation for freezing tolerance in Arabidopsis. *Plant*, *Animal*, *Microbe Genomes X Conf.*, *San Diego*, *CA*

101. Masle J, Shin JS, Farquhar GD. 1993. Analysis of restriction fragment length polymorphisms associated with variation of carbon isotope discrimination among ecotypes of *Arabidopsis thaliana*. In *Stable Isotopes and Plant Carbon-Water Relations*, ed. JR Ehleringer, AE Hall, GD Farquhar, pp. 371–86. San Diego, CA: Academic

102. Mauricio R, Stahl EA, Korves T, Tian D, Kreitman M, Bergelson J. 2003. Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. *Genetics* 163:735–46

103. McKay JK, Richards JH, Mitchell-Olds T. 2003. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol. Ecol.* 12:1137–51

104. Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. 1998. *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282:662–82

105. Meyer D, Leonardi A, Brisson G, Vartanian N. 2001. Drought-adaptive mechanisms involved in the escape/tolerance strategies of Arabidopsis Landsberg *erecta* and Columbia ecotypes and their F1 reciprocal progeny. *J. Plant Physiol.* 158:1145–52

106. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15:809–34

107. Michaels SD, Amasino RM. 1999. *FLOWERING LOCUS C* encodes a novel MAD domain protein that acts as a repressor of flowering. *Plant Cell* 11:949–56

108. Michaels SD, Amasino RM. 2001. Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell* 13:935–41

109. Michaels SD, He Y, Scortecci KC, Amasino RM. 2003. Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behaviour in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 100:10102–7

110. Millar AA, Kunst L. 1999. The natural genetic variation of the fatty-acyl composition of seed oils in different ecotypes of *Arabidopsis thaliana*. *Phytochemistry* 52:1029–33

111. Mindrinos M, Katagiri F, Yu GL, Ausubel FM. 1994. The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* 78:1089–99

112. Mita S, Murano N, Akaike M, Nakamura K. 1997. Mutants of *Arabidopsis thaliana* with pleiotropic effects on the expression of the gene ß-amilase and on the accumulation of anthocyanin that are inducible by sugars. *Plant J.* 11:841–51

113. Mitchell-Olds T. 1995. Interval mapping of viability loci causing heterosis in *Arabidopsis*. *Genetics* 140:1105–9

114. Mitchell-Olds T. 2001. *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* 16:693–700

115. Mitchell-Olds T, Pedersen D. 1998. The molecular basis of quantitative genetic variation in central and secondary

metabolism in Arabidopsis. *Genetics* 140:1105–9

116. Mithen R, Clarke J, Lister C, Dean C. 1995. Genetics of aliphatic glucosinolates. III. Side chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. *Heredity* 74:210–15

117. Miyashita NT, Innan H, Terauchi R. 1996. Intra- and interspecific variation of the alcohol dehydrogenase locus region in wild plants *Arabis gemmifera* and *Arabidopsis thaliana*. *Mol. Biol. Evol.* 13:433–36

118. Miyashita NT, Kawabe A, Innan H. 1999. DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. *Genetics* 152:1723–31

119. Mouradov A, Cremer F, Coupland G. 2002. Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14:S111–30

120. Napp-Zinn K. 1987. Vernalization. Environmental and genetic regulation. In *Manipulation of Flowering*, ed. JG Atherton, pp. 123–32. London: Butterworths

121. Narang RA, Bruene A, Altmann T. 2000. Analysis of phosphate acquisition efficiency in different Arabidopsis accessions. *Plant Physiol.* 124:1786–99

122. Nasrallah ME, Yogeeswaran, Snyder S, Nasrallah JB. 2000. *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiol.* 124:1605–14

123. Nordborg M, Borewitz JO, Bergelson J, Berry CC, Chory J, et al. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 30:190–93

124. O'Kane SL, Al-Shehbaz IA. 1997. A synopsis of *Arabidopsis* (Brassicaceae). *Novon* 7:323–27

125. Olsen KM, Womack A, Garrett A, Suddith JI, Purugganan M. 2002. Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. *Genetics* 160:1641–50

126. Paran I, Zamir D. 2003. Quantitative traits in plants: beyond the QTL. *Trends Genet.* 19:303–6

127. Pepper AE, Corbett RW, Kang N. 2002. Natural variation in *Arabidopsis* seedling photomorphogenesis reveals a likely role for *TED1* in phytochrome signalling. *Plant Cell Environ.* 25:591–600

128. Pérez-Pérez JM, Serrano-Cartagena J, Micol JL. 2002. Genetic analysis of natural variations in the architecture of *Arabidopsis thaliana* vegetative leaves. *Genetics* 162:893–15

129. Pigliucci M. 2002. Ecology and evolutionary biology of Arabidopsis. See Ref. 149a, pp. 20

130. Pigliucci M, Schlichting CD. 1995. Reaction norms of *Arabidopsis* (*Brassicaceae*). III. Response to nutrients in 26 populations from a worldwide collection. *Am. J. Bot.* 82:1117–25

131. Poduska B, Humphrey T, Redweik A, Grbic V. 2003. The synergistic activation of *flowering locus C* by *Frigida* and a new flowering time gene *Aerial rosette 1* underlies a novel morphology in Arabidopsis. *Genetics* 163:1457–65

132. Ponce MR, Robles P, Micol JL. 1999. High-throughput genetic mapping in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 261:408–15

133. Purugganan MD, Suddith JI. 1998. Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. *Proc. Natl. Acad. Sci. USA* 95:8130–34

134. Purugganan MD, Suddith JI. 1999. Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. *Genetics* 151:839–48

135. Queseda V, Garcia-Martinez S, Piqueras P, Ponce MR, Micol JL. 2002. Genetic architecture of NaCl tolerance in

Arabidopsis. Plant Physiol. 130:951–63

136. Rao MV, Davis KR. 1999. Ozone-induced cell death occurs via two distinct mechanisms in *Arabidopsis*: the role of salycilic acid. *Plant J.* 17:603–14

137. Rashotte AM, Jenks MA, Nguyen TD, Feldmann KA. 1997. Epicuticular wax variation in ecotypes of *Arabidopsis thaliana. Phytochemistry* 45:251–55

138. Rauh BL, Basten C, Buckler ES IV. 2002. Quantitative trait loci analysis of growth response to varying nitrogen sources in *Arabidopsis thaliana. Theor. Appl. Genet.* 104:743–50

139. Remington DL, Purugganan MD. 2003. Candidate genes, quantitative trait loci, and functional trait evolution in plants. *Int. J. Plant Sci.* 164:S7–20

140. Riddle NC, Richards EJ. 2002. The control of natural variation in cytosine methylation in *Arabidopsis. Genetics* 162:355–63

141. Sanchez-Moran E, Armstrong SJ, Santos JL, Franklin FCH, Jones GH. 2002. Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana. Genetics* 162:1415–22

142. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, et al. 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422:297–302

143. Schiantarelli E, De la Pena A, Candela M. 2001. Use of recombinant inbred lines (RILs) to identify, locate and map major genes and quantitative trait loci involved with in-vitro regeneration ability in *Arabidopsis thaliana. Theor. Appl. Genet.* 102:335–41

144. Schmid KJ, Rosleff Sörensen T, Stracke R, Törjek O, Altmann T, et al. 2003. Large-scale identification and analysis of genome-wide single nucleotide polymorphs for mapping in *Arabidopsis thaliana. Genome Res.* 13:1250–57

145. Schranz ME, Quijada P, Sung S-B, Lukens L, Amasino R, Osborn TC. 2002. Characterization and effects of the repli-cated flowering time gene *FLC* in *Brassica rapa. Genetics* 162:1457–68

146. Sergeeva LI, Vonk J, Keurentjes JJB, van der Plas LHW, Koornneef M, Vreugdenhil D. 2004. Histochemical analysis reveals organ-specific QTLs for enzyme activities in *Arabidopsis thaliana. Plant Physiol.* 134:237–45

147. Sharbel TF, Haubold B, Mitchell-Olds T. 2000. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol. Ecol.* 9:2109–18

148. Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES. 2000. The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC*). *Proc. Natl. Acad. Sci. USA* 97:3753–58

149. Shepard KA, Purugganan MD. 2003. Molecular population genetics of the *Arabidopsis CLAVATA2* region. The genomic scale of variation and selection in a selfing species. *Genetics* 163:1083–95

149a. Somerville CR, Meyerowitz EM, eds. 2002. The *Arabidopsis Book.* Rockville, MD: Am. Soc. Plant Biol. http://www.aspb.org/publications/arabidopsis

150. Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. 1999. Dynamics of disease resistance polymorphism at the *RPM1* locus of *Arabidopsis. Nature* 400:667–71

151. Stratton DA. 1998. Reaction norm functions and QTL-environment interactions for flowering time in *Arabidopsis thaliana. Heredity* 81:144–55

152. Swarup K, Alonso-Blanco C, Lynn JR, Michaels S, Amasino RM, et al. 1999. Natural allelic variation identifies new genes in the *Arabidopsis* circadian system. *Plant J.* 20:1–11

153. Tadege M, Sheldon CC, Helliwell CA, Stoutjesdijk P, Dennis ES, et al. 2001. Control of flowering time by *FLC* orthologues in *Brassica napus. Plant J.* 28:545–53

154. Thornsbury JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, et al. 2001. Dwarf 8 polymorphisms associate with variation in flowering time. *Nat. Genet.* 28:7922–27

155. Tian D, Araki H, Stahl E, Bergelson J, Kreitman M. 2002. Signature of balancing selection in Arabidopsis. *Proc Natl. Acad. Sci. USA* 99:11525–30

156. Ton T, Pieterse CMJ, van Loon LC. 1999. Identification of a locus in *Arabidopsis* controlling both the expression of rhizobacteria-mediated induced systemic resistance (ISR) and basal resistance against *Pseudomonas syringae* pv. tomato. *Mol. Plant Microbe Interact.* 12:911–18

157. Torabinejad J, Caldwell MM. 2000. Inheritance of UV-B tolerance in seven ecotypes of *Arabidopsis thaliana* (L.) Heynh. and their F1 hybrids. *J. Hered.* 91:228–33

158. Ullrich H, Lattig K, Brennicke A, Knoop V. 1997. Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 *Arabidopsis thaliana* ecotypes. *Plant Mol. Biol.* 33:37–45

159. Ungerer MC, Halldorsdottir SS, Modliszewski JL, Mackay TFC, Purugganan MD. 2002. Quantitaive trait loci for inflorescence development in *Arabidopsis thaliana*. *Genetics* 160:1133–51

160. Van der Schaar W, Alonso-Blanco C, Léon-Kloosterziel KM, Jansen RC, van Ooijen JW, Koornneef M. 1997. QTL analysis of seed dormancy in *Arabidopsis* using recombinant inbred lines and MQM mapping. *Heredity* 79:190–200

161. Vander Zwan C, Brodie SA, Campanella JJ. 2000. The intraspecific phylogenetics of *Arabidopsis thaliana* in worldwide populations. *Syst. Bot.* 25:47–59

162. Weinig C, Stinchcombe JR, Schmitt J. 2003. QTL architecture of resistance and tolerance traits in *Arabidopsis thaliana* in natural environments. *Mol. Ecol.* 12:1153–63

163. Weinig C, Ungerer MC, Dorn LA, Kane NC, Toyonaga Y, et al. 2002. Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics* 162:1875–84

164. Wilson IW, Schiff CL, Hughes DE, Somerville SC. 2001. Quantitative trait loci analysis of powdery mildew disease resistance in the *Arabidopsis thaliana* accession Kashmir-1. *Genetics* 158:13019

165. Yanovsky MJ, Casal JJ, Luppi JP. 1997. The VLF loci, polymorphic between ecotypes Landsberg *erecta* and Columbia, dissect two branches of phytochrome A signal transduction that correspond to very-low-fluence and high-irradiance responses. *Plant J.* 12:659–67

166. Zhang J, Lechowicz MJ. 1994. Correlation between time of flowering and phenotypic plasticity in *Arabidopsis thaliana* (Brassicaceae). *Am. J. Bot.* 81:1336–42

167. Zhang J, Lechowicz MJ. 1995. Responses to CO_2 enrichment by two genotypes of *Arabidopsis thaliana* differing in their sensitivity to nutrient availability. *Ann. Bot.* 75:491–99

Figure 1   Geographical distribution of *Arabidopsis thaliana*. Green shaded areas correspond to the species distribution (8) modified according to (50). The ~300 accessions collected from different locations and publicly available through stock centers are plotted as red dots.
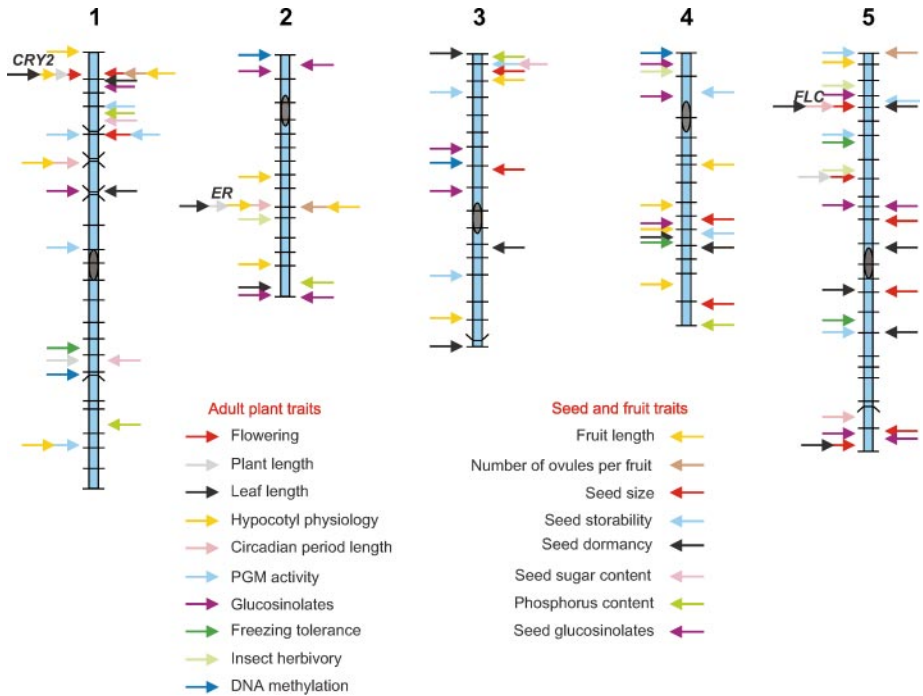
Figure 2 Map position of QTL found for various traits in the L*er* x Cvi population based on published analyses (5–7, 15, 16, 19, 21, 70–72, 100, 140, 147, 152, 159). The approximate position of each QTL is depicted by an arrow pointing to the maximum LOD score reported. *CRY2*, *ER*, and *FLC* indicate genes proven to correspond to pleiotropic QTL (see text).

$\overset{A}{R}$*Annual Review of Plant Biology*
Volume 55, 2004

CONTENTS