Positional Cloning in Arabidopsis. Why It Feels Good to Have a Genome Initiative Working for You[1]

Wolfgang Lukowitz*, C. Stewart Gillmor, and Wolf-Rüdiger Scheible

Department of Plant Biology, Carnegie Institution of Washington, 260 Panama Street, Stanford, California 94305 (W.L., C.S.G., W.-R.S.); and Department of Biological Sciences, Stanford University, Gilbert Hall, Stanford, California 94305 (C.S.G.)

Positional (or map-based) cloning techniques are widely used to identify the protein products of genes defined by mutation. In Arabidopsis the information generated by the Genome Initiative is giving this approach a decisive boost. A wealth of sequence polymorphisms and molecular markers is now available and can be exploited for fine mapping with technically simple and robust polymerase chain reaction-based methods. As a result it has become possible to complete positional cloning projects in a short time and with relatively little effort.

Over the past decade the use of Arabidopsis genetics has become increasingly popular for the study of plant biochemistry, physiology, pathology, and developmental biology (for review, see Meinke et al., 1998). Much information can be gathered from analysis of genetic variation without knowing anything about the underlying molecular causes. However, most researchers would ultimately like to know which genes have been altered to produce a given phenotype. If a mutation is caused by the insertion of a T-DNA or transposon, the inserted sequence provides a tag pointing directly to the gene. Most genetic variability, however, is not of this kind. Chemical agents and radiation are widely used to induce mutations for a number of good reasons. Ethyl methanesulfonate, for example, is easy to use, more effective than many other mutagens, and since it typically causes single basepair exchanges, more likely to create mutations with special properties such as weak, dominant, or conditional alleles. Apart from induced mutations there is a wealth of naturally occurring genetic variability in traits like resistance to pathogens, flowering-time, seed size, and many others. This variability is likely to reflect rather subtle changes in the DNA sequence. In all of these cases there is little alternative but to isolate the genes involved based on their map position—a process that until recently has been considered time-consuming and expensive. With public accessibility of the complete and annotated sequence of the Arabidopsis genome anticipated in only a few months, this situation has radically changed. Chromosome walks, formerly the most tedious and technically problematic steps in the process, are a thing of the past. Molecular mapping can be vastly enhanced by systematically exploiting the available sequence information. As a consequence it now takes just a few rather basic molecular biology routines and as little as a few months to isolate (almost) any mutation that can be mapped.

In this paper we discuss how to approach a positional cloning project in Arabidopsis with an emphasis on techniques that have been found to be reliable, simple, and inexpensive. Protocols for many of the procedures mentioned have been deposited on an accompanying Web page (Table I). Our hope is to convince researchers who have little or no experience in positional cloning that now is a good time to capitalize on the achievements of the Arabidopsis Genome Initiative and find out what your favorite mutant is all about.

MAPPING RESOLUTION

In contrast to gene tagging, positional (or map-based) cloning is an essentially indirect approach: mapping will narrow down the genetic interval containing a mutation by successively excluding all other parts of the genome. Once an interval is defined, other criteria have to be employed to find out which of the genes within the interval is mutated. Naturally, this is easier if the interval is small or the mapping resolution is high. For the purpose of this paper we will refer to the mapping resolution as the average distance between two recombination breakpoints in a given mapping population. In Arabidopsis a genetic distance of 1% recombination corresponds, on average, to a physical distance of about 250 kb. However, the ratio between genetic and physical distance is by no means constant and it varies with respect to position on the chromosome as well

[1] W.L. was partially supported by a postdoctoral fellowship (LT 594/96) from the Human Frontier Science Program Organization and by a fellowship from the Carnegie Institution of Washington. C.S.G. was supported by a U.S. Department of Energy/National Science Foundation/U.S. Department of Agriculture TRI-Agency Training Grant, and W.-R.S. was supported by a postdoctoral fellowship (Sche 548 1/1) from the Deutsche Forschungsgemeinschaft.

* Corresponding author; e-mail lubo@andrew2.stanford.edu; fax 650–325–6857.

Table I. *Internet resources*

Web Site	Internet Address
Supplemental material for this paper	http://carnegiedpb.stanford.edu/methods/ppsuppl.html
Nottingham Stock Centre (U.K.)	http://nasc.nott.ac.uk/
Recombinant Inbred map	http://nasc.nott.ac.uk/new_ri_map.html
Ohio Stock Center (U.S.A.)	http://aims.cps.msu.edu/aims/
TAIR database[a], homepage	http://www.arabidopsis.org
Recombinant Inbred map (mirror site)	http://www.arabidopsis.org/cgi-bin/maps/Riintromap
CAPS markers	http://www.arabidopsis.org/aboutcaps.html
Sequence table	http://www.arabidopsis.org/cgi-bin/maps/Seqtable.pl
SNP collection	http://www.arabidopsis.org/SNPs.html
CEREON collection of polymorphisms	http://www.arabidopsis.org/cereon
SSLP markers	http://genome.bio.upenn.edu/SSLP_info/SSLP.html
TIGR, genome annotations	http://www.tigr.org/tdb/athl/htmls/index.html
Database of Ler sequences	http://www.tigr.org/tdb/atgenome/Ler.html
Kasuza DNA Research Institute, genome annotations	http://www.kazusa.or.jp/kaos/
MIPS genome annotations	http://websvr.mips.biochem.mpg.de/proj/thal/
SINS database of transposon insertions	http://www.jic.bbsrc.ac.uk/sainsbury-lab/jonathan-jones/jjhome.htm

[a] The Arabidopsis Information Resource (TAIR) monitors a large variety of information relevant for researchers working with Arabidopsis and provides links to most of the web-sites listed here. TAIR plans to incorporate all available information about polymorphic sequences and molecular markers at one central location in the future.

as with respect to different mapping populations. For example, a substantial difference was observed in two mapping populations involving different alleles of the *KNOLLE* gene: although the relative distances of genetic markers were similar, the overall recombination frequency varied 2-fold (Lukowitz et al., 1996). Thus, the resolution of a given mapping population has to be established for each individual experiment.

Mapping resolution is mainly determined by the size of a mapping population. As shown in Table II, resolutions in the range of 10 to 40 kb can routinely be obtained in mapping populations of about 1,000 plants (or 2,000 chromosomes). Typically, a DNA fragment of this length contains between two and 10 genes. This level of resolution might seem like more than is required, especially in cases where a prediction about the gene product can be made. However, even well grounded predictions will occasionally be misleading, and often it is difficult to make any prediction at all. Once an effective way of fine mapping a mutation has been set up, on the other hand, it requires only a little more effort to analyze 1,000 or more plants instead of a few hundred.

The best way to identify a mutated gene within a genetically defined interval is much dependent on circumstances, and it is a worthwhile exercise to spend some time (and imagination) to re-assess this problem in each individual case (see Table II and refs. therein for some of the more common approaches). If the interval is small enough, the annotations to the DNA sequence (Table I) will in most cases provide enough information to choose a candidate gene and determine the DNA sequence of mutant allele(s). Alternatively, the mutated gene can be identified by molecular complementation, that is, by transforming overlapping pieces of wild-type DNA into the mutant to determine which sequences are capable of restoring the mutant trait to wild type. While molecular complementation of recessive mutants is typically straightforward, gain-of-function mutants with a dominant or semidominant inheritance can be problematic. In these cases, however, it might be possible to transfer a dominant mutant allele into wild-type plants to copy the mutant phenotype (for example, see Leung et al., 1994; Meyer et al., 1994). Often a gene can be identified by virtue of RFLPs associated with mutant alleles. This approach is most successful when radiation-induced (commonly x-ray or fast neutron) alleles are available, but it has also been taken with ethyl methanesulfonate-induced alleles (for example, see Li and Chory, 1997). It is also possible to scan the complete DNA sequence of the genetic interval for changes that have caused the mutation by any of several methods, such as enzymatic or chemical cleavage of mismatched bases (for review, see Taylor, 1999), analysis of single-strand conformational polymorphisms (for review, see Nataraj et al., 1999), heteroduplex analysis (Hauser et al., 1998), or denaturing HPLC (O'Donovan et al., 1998). Although at present at least some of these methods seem technically challenging, it is likely that they will provide powerful and robust tools in the future.

RESOURCES AND TOOLS FOR MOLECULAR MAPPING

Mapping with a high resolution requires a high density of genetic markers. Several Arabidopsis accessions, or ecotypes, are sufficiently divergent to support the design of molecular markers at this high density. Wassilewskija (Ws), the genetic background of a large collection of T-DNA lines (Feldmann, 1991), and Niederzenz, according to early reports, a highly diverged strain (Chang et al., 1988), have both successfully been used in positional cloning projects.

Table II. *Recent examples of positional cloning projects in chronological order*

Gene	Size of Mapping Population	Estimated Mapping Resolution	Smallest Genetic Interval Containing the Gene[b]	Markers Used	Gene Identified By	Reference
	plants		*kb*			
SUP	2,513	<10[a]	<10 [8]	Visible, CAPS, RFLP	RFLPs associated with EMS alleles, sequencing of EMS alleles, molecular complementation	Sakai et al. (1995)
KN	1,348	30	100 [4]	CAPS, RFLP	RFLPs associated with x-ray alleles, sequencing of EMS alleles	Lukowitz et al. (1996)
GN	529	50[a]	100 [3]	CAPS, RFLP	RFLP associated with x-ray allele, sequencing of EMS alleles	Busch et al. (1996)
*BRI*1	~1,000	50	150 [3]	CAPS, SSLP	RFLPs associated with EMS alleles, sequencing of EMS alleles	Li and Chory (1997)
*NIM*1	1,138	30[a]	70 [2]	SSLP, AFLP[c]	Molecular complementation, sequencing of EMS alleles	Ryals et al. (1997)
WUS	1,575	<10[a]	40 [7]	CAPS, RFLP	Molecular complementation, sequencing of EMS alleles	Mayer et al. (1998)
MP	898	20[a]	45 [3]	CAPS, RFLP	RFLPs associated with x-ray alleles, sequencing of EMS alleles	Hardtke and Berleth (1998)
ZLL	1,250	20[a]	100 [6]	Visible, RFLP, CAPS	Sequencing of EMS alleles	Moussian et al. (1998)
*IFL*1	652	30[a]	65 [2]	CAPS, RFLP	Molecular complementation, sequencing of EMS alleles	Zhong and Ye (1999)
*CYT*1	2,842	<10	10 [2]	Visible, SSLP	Sequencing of EMS alleles	W. Lukowitz and C.R. Somerville (unpublished data)
*IXR*1	1,056	10	50 [6]	SSLP, CAPS	Sequencing of EMS alleles, molecular complementation	W.-R. Scheible and C.R. Somerville (unpublished data)
KNF	468	20	160 [8]	SSLP	RFLPs associated with fast neutron and x-ray alleles, sequencing of EMS alleles	C.S. Gillmor and C.R. Somerville (unpublished data)

[a] Estimated mapping resolutions were calculated from the published data. [b] Nos. in brackets indicate the no. of remaining recombinants within the smallest genetically defined interval containing the gene. [c] Amplified fragment length polymorphism (Vos et al., 1995).

The most commonly used combination for mapping purposes, however, is Landsberg *erecta* X Columbia (Ler X Col). These two accessions have been estimated to differ in four to 11 positions every 1,000 bp (Chang et al., 1988; Konieczny and Ausubel, 1993; Hardtke et al., 1996). Many valuable resources have been accumulated that facilitate mapping Ler X Col populations. The sequence of the Columbia accession (with a few exceptions) is being determined by the Arabidopsis Genome Initiative. Various libraries made from wild-type plants of this accession have been deposited at the Arabidopsis stock centers (Table I, nos. 2 and 3), including genomic and cDNA phage-libraries, and libraries of yeast- and bacterial artificial chromosomes and of transformation-competent artificial chromosomes, which may be used for molecular complementation (Liu et al., 1999). Extensive sequence information is available for the Landsberg *erecta* accession facilitating the design of new molecular markers. Most existing mutations, including mutations causing visible phenotypes that can be employed as genetic markers, have been induced either in a Columbia or a Landsberg *erecta* background. Recombinant-inbred lines from a Ler X Col cross have been generated (Lister and Dean, 1993) and used to create a genetic map that currently features about 1,200 molecular markers, more than 80 of which are PCR-based (Table I, nos. 2 and 4). All these resources make Ler X Col the combination of choice.

The molecular markers most widely used in mapping experiments at present are simple sequence length polymorphisms (SSLPs), cleaved amplified polymorphic sequences (CAPS) and derived CAPS (dCAPS). As illustrated in Figure 1, they share two important properties. (a) They are codominant, which means that both chromosomes of a plant may be genotyped and this allows the maximum amount of information to be gathered from a mapping population. (b) They are PCR-based and can be analyzed on agarose gels, which makes them easy to use and inexpensive. Several methods have been described for the extraction of plant DNA for PCR purposes from small tissue samples (see refs. in Table II). Two of them are routinely used in our lab with good success, a simplified CTAB-extraction procedure (Lukowitz et al., 1996) and a variation of an alkaline-lysis procedure (Klimyuk et al., 1993; see Table I, Supplemental material). The first method yields comparatively high quality DNA, but includes a chloro-

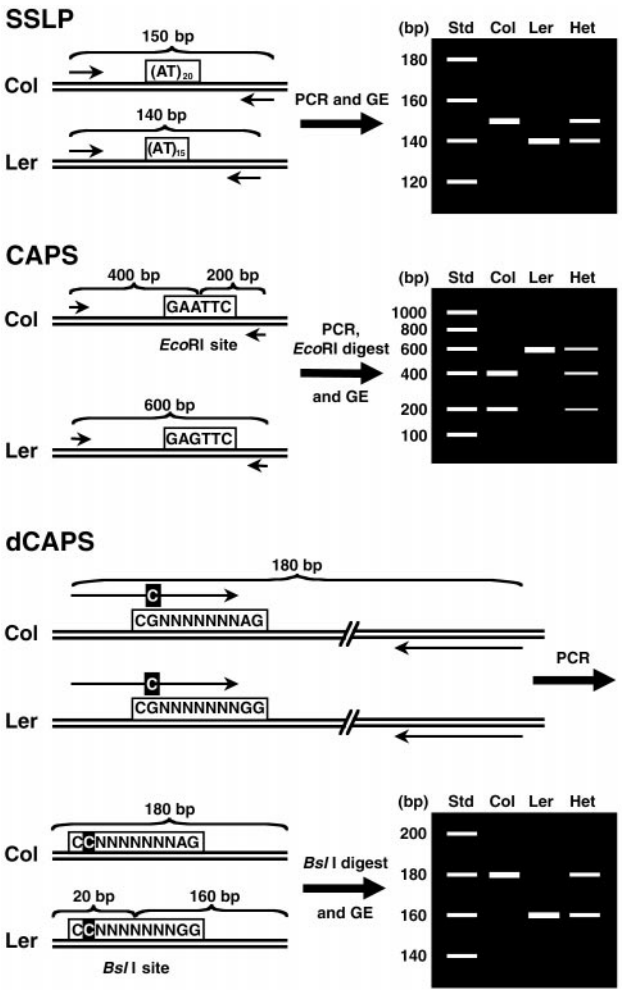

Figure 1. Molecular markers commonly used for mapping. SSLP (Bell and Ecker, 1994), SSLP markers exploit the variability of short repetitive sequences for mapping purposes. A primer pair (arrows) is used to amplify a fragment containing a short repetitive element, such as the (AT)-dinucleotide repeat shown in the example. The length of this repeat differs between the two accessions. In the example Columbia (Col) has an $(AT)_{20}$ repeat whereas Landsberg *erecta* (Ler) has an $(AT)_{15}$ repeat. Consequently, the PCR products from Ler and Col DNA also differ in length. In the example, the Col product is 150 bp and the Ler product is 140 bp long. This size difference can be visualized by agarose gel-electrophoresis (GE) with a standard (Std) as comparison. If a plant is heterozygous for the marker (Het), both products are amplified resulting in two bands. Typically the size difference created by short repetitive sequences is small. To facilitate their detection, SSLP primers usually amplify short fragments (80 to 250 bp), and the PCR products are resolved on high-percentage gels often made with special agarose blends. CAPS (Konieczny and Ausubel, 1993), CAPS marker exploits polymorphic restriction sites for mapping purposes. PCR primers (arrows) are used to amplify a genomic sequence that contains a restriction site in one of the accessions, but not in the other. In the example the fragment is 600 bp long; the Col fragment contains an *Eco*RI site, but the Ler fragment does not. The PCR products are then digested with the respective enzyme (*Eco*RI) and the products analyzed on an agarose gel. In the example two bands (400 and 200 bp) are seen in Col, whereas a single band (600 bp) is detected in Ler. A heterozygous plant (Het) gives all three bands. dCAPS (Michaels and Amasino, 1998; Neff et al., 1998), dCAPS markers can exploit almost all single nucleotide changes for mapping purposes. To achieve this a mismatched PCR primer is designed next to the polymorphic position such that an artificial restriction site is created with the sequence variant of one accession, but not with the other. In the example the mismatch is in the forward primer, which has a C instead of G eight nucleotides away from its 3′ end. Consequently, at this position the sequence of all PCR products is changed with respect to the genomic sequence. As a result the Ler PCR product, but not the Col product, contains a *Bsl*I site (CCNNNNNNNGG). Restriction sites with interrupted palindromes are frequently used in the design of dCAPS markers because they allow to position the mismatched nucleotide at a distance from the 3′ end of the oligonucleotide where it is less likely to interfere with the priming of *Taq* polymerase. The size difference after digestion is rather small (20 bp in the example) and essentially determined by the length of the mismatched primer. To facilitate detection, the PCR products are usually designed to be short (180 bp in the example).

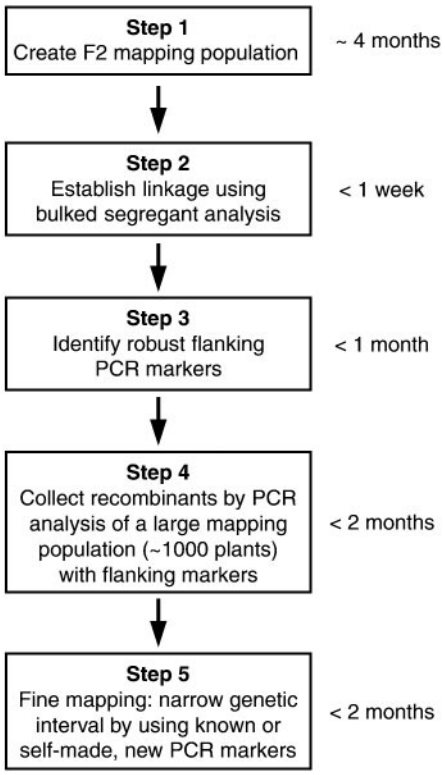

Figure 2. Procedure of a typical fine mapping experiment. The time that should be calculated to complete each of the steps, assuming no complications are encountered, is shown to the right.

form extraction step and is thus somewhat difficult to adapt to a microtiter plate format. The second method yields a rather crude extract, which occasionally causes problems when large DNA fragments need to be amplified. However, everything is done in a single tube making this method ideally suited for high throughput applications in a microtiter plate format.

OUTLINE OF A FINE-MAPPING EXPERIMENT

A typical fine-mapping experiment proceeds in a number of successive steps (Fig. 2). To begin, a mu-

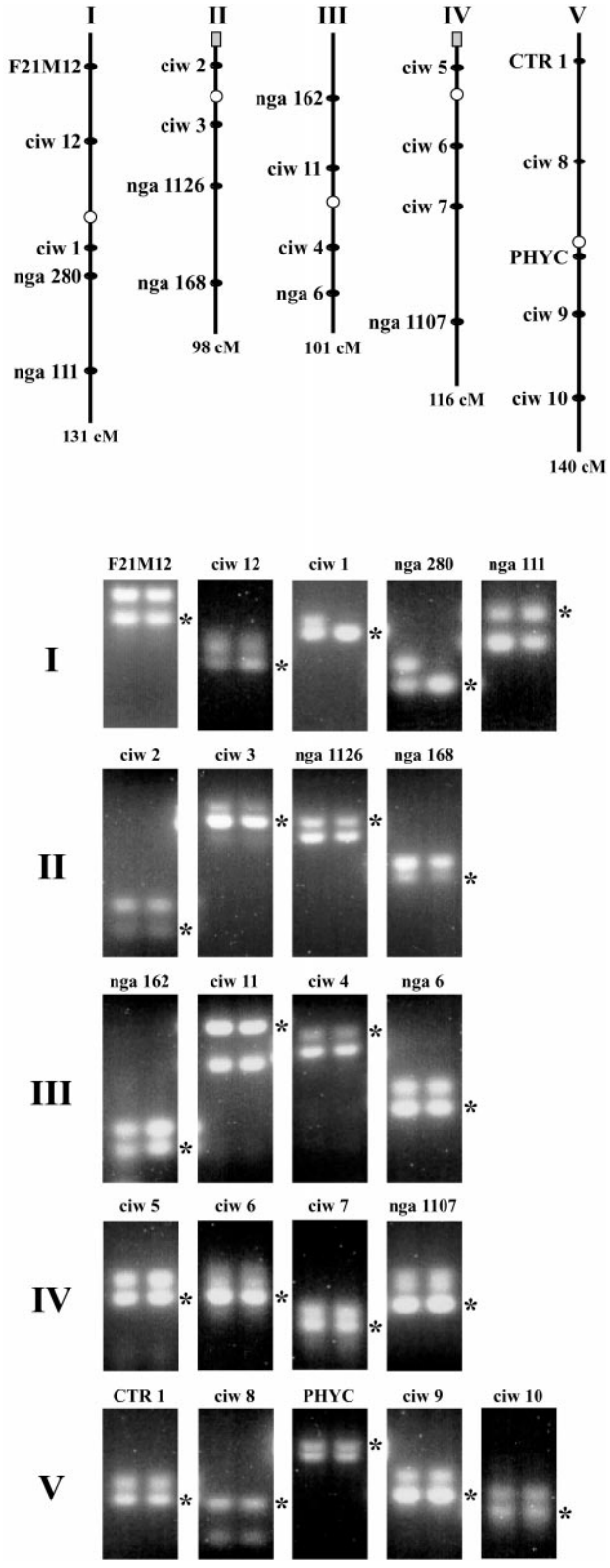

Figure 3. Bulked segregant analysis of a seedling mutation. Top, Schematic representation of SSLP marker positions used in the genetic mapping experiment. Markers with the prefix nga, as well as CTR1 and PHYC, were obtained from the Arabidopsis Genome Center (AtGC) Web page at the University of Pennsylvania (Table I; Bell and Ecker, 1994), and ciw markers were created by C.S.G. Markers are placed on the chromosomes to scale as determined by their position on the RI map (AtGC markers) or as estimated from the physical map (ciw markers). Centromeres are represented by open circles, and nucleolar organizing regions by shaded boxes. The length of each chromosome according to the recombinant inbred map (in centimorgans, cM) is listed below the chromosome. Bottom, Gel electrophoresis of PCR products for each SSLP marker. In each panel the heterozygous control sample (named A in the text) is shown on the left, and the pooled F_2 mutant sample (named B in the text) on the right. The Ler specific bands are marked with an asterisk. The mutation (created in the Ler accession) is linked to the markers, ciw1 and nga 280, as judged by the clear bias toward amplification of the Ler band from the genomic DNA of F_2 mutants. At all other loci the ratio of Col to Ler amplification is approximately the same between the heterozygous control and the F_2 mutants, indicating the mutation is unlinked to these loci. See Table I, supplemental material, for experimental details.

tation is crossed to one or several diverged accessions to create a mapping population. It is advisable to prepare more than one mapping cross for the rare cases that complications, perhaps due to second-site modifiers, may be encountered with the first population (see below for discussion). Once the F_2 generation of a mapping cross is available, the mutation is assigned a rough position on the genetic map by identifying linked genetic markers. Next, two markers are found that are about 5% recombination apart and define a genetic interval containing the mutation. Utilizing these flanking markers, a large mapping population is screened for recombinants in the vicinity of the mutation. Finally, the genetic interval containing the mutation is narrowed down as much as possible by creating and analyzing new markers in the region—ideally, markers that are only one recombinant apart from the mutation are identified.

Each of these steps has its own, well-defined goal and, as a consequence, poses different experimental problems. Below, we will discuss these problems and how to approach them separately. Figure 2 also gives an estimate on how much time should be calculated for the completion of each step, provided that no complications are encountered. The most common type of such complications is inconsistencies in the mapping data. In our experience these inconsistencies are usually due to mix-ups or errors in scoring the mutant phenotype—for example, many phenotypes are not fully penetrant or strongly affected by growth conditions. In these cases, the genotypes of the relevant F_2 plants have to be re-examined in the F_3 or F_4 generation—an additional and relatively time-consuming step.

DETECTING LINKAGE BY BULKED SEGREGANT ANALYSIS

An effective way to identify markers that are genetically linked to a mutation is by bulked segregant analysis (Michelmore et al., 1991). In this method pools of DNA are examined rather than many indi-

vidual samples. This substantially reduces the number of PCR reactions required to establish linkage. Figure 3 shows the result of a bulked segregant analysis applied to a recessive seedling mutation. The mutation was induced in a Landsberg *erecta* background and crossed to Columbia to generate a mapping population. Two DNA samples were prepared from the following tissues. Sample A was from leaves of the F_1 plant used to generate the F_2 mapping population, and sample B was from a pool of about 100 mutant F_2 seedlings. Sample A is heterozygous for all genetic loci and serves as a control—any other DNA sample that is heterozygous for all loci can be used in its place. Sample B, which was made from mutant seedlings, is homozygous Landsberg *erecta* at the mutation and therefore mostly Landsberg *erecta* in the vicinity of the mutation, but essentially heterozygous for unlinked markers. Both samples were examined with a collection of 22 SSLP markers spaced evenly over the entire genome in intervals of 10% to 40% recombination. The molecular markers ciw1 and nga280 show a clear bias toward the Landsberg *erecta*-specific band in the mutant pool as compared to the heterozygous control. This indicates that the mutation maps to the lower arm of chromosome I.

Instead of using DNA prepared from a pool of mutant tissue it is often advantageous to first prepare DNA from individual mutants and then pool aliquots of it for the analysis. This way the individual DNA samples can later be used for three-point mapping (see below). A bulked segregant analysis can also be performed in cases where it is impossible or inconvenient to prepare DNA from mutants, for example, because the mutants arrest early during embryogenesis. In this case a pool of DNA from homozygous wild-type individuals can be used to map the wild-type allele instead of the mutant allele.

The SSLP markers used in the above example were specifically compiled with the aim to optimize bulk segregant analysis of Ler X Col populations. Some of them were taken from a published collection (Table I; Bell and Ecker, 1994). The majority, however, were newly designed for the purpose. Detailed information about the primer sequences, the PCR conditions, as well as the sizes of the PCR fragments generated from the Columbia, Landsberg *erecta*, and Ws accessions can be accessed online (Table I, no. 1; the primer sequences have also been deposited at The Arabidopsis Information Resource (TAIR). Since the same PCR profile can be used to amplify all markers and the same gel to separate the products, it is possible to complete a bulked segregant analysis within a few hours.

As an alternative way to establish linkage with a minimal number of PCR reactions, a protocol has been described that allows the analysis of individual F_2 plants with 21 SSLP markers simultaneously (Ponce et al., 1999). This protocol involves PCR reactions with fluorescently labeled primers and separation/detection of the products with an automated sequencer.

COLLECTING RECOMBINANTS IN THE VICINITY OF THE MUTATION

Searching a large mapping population for chromosomes with a recombination in the vicinity of the mutation can be done effectively if two markers are known that are closely linked and that flank the mutation on both sides. In a large F_2 population the majority of plants will have the same genotype at both flanking markers. If the two markers, for example, are 5% recombination apart, about 900 of 1,000 plants (or 1,800 of 2,000 chromosomes) can be expected to show the same genotype at both markers. These plants are either not recombinant in the interval between the two markers or, in very rare cases, carry two recombination events. For the most part they are not informative for mapping the mutation and can be discarded without further analysis. Thus, the bulk of the mapping population (90% in the example) is only analyzed with two flanking markers and not considered further. A minority of the F_2 plants will show different genotypes for the two flanking markers. Obviously, these plants carry a chromosome with a recombination close to the mutation and are informative for further mapping.

If the mutation is recessive, viable, and fertile, it is possible to utilize only plants that show the mutant phenotype for fine mapping. Since all these plants should be homozygous mutant for the gene of interest, no further analysis of the F_3 generation is required. Alternatively, all F_2 plants can be included in the mapping population. In this case the genotype with respect to the mutation must be determined for all plants that have been found to carry a recombination event in the vicinity of the gene. Obviously, this often requires analysis of the F_3 generation.

Although a bulked segregant analysis is a very effective way to detect linkage, it usually does not allow determination of the order of closely linked loci on the chromosome. This can only be done by analyzing three-point mapping data. To find flanking markers that can be used to select recombinant chromosomes as described above, it is therefore necessary to examine individual F_2 plants with markers from the region. A small mapping population of about 50 to 100 F_2 plants, giving a mapping resolution of roughly 1% recombination, is usually sufficient for the purpose. It is advantageous if this small mapping population is not subjected to any selection because in this case, it can also serve as a control, addressing some basic genetic issues: Is the mutation segregating in a Mendelian fashion or is the segregation distorted? How is the mutant phenotype affected in the hybrid genetic background? For the efficient selection of recombinant chromosomes, two markers should be identified that are technically robust, less

than 10% recombination apart, and define a genetic interval containing the mutation. There is a fair chance of finding appropriate markers in the existing SSLP- and CAPS-collections (Table I). However, if the existing markers prove to be less than ideal, for example because PCR amplification is unreliable, restriction digest with an expensive enzyme is required, or if the polymorphisms are difficult to score, adequate new markers should be created (see below for how to approach this problem).

Analyzing 1,000 or more plants with two flanking PCR markers involves some amount of repetitive work. The effort can be substantially reduced if a microtiter plate format is adopted. Using commercially available inserts, 96 plants can be grown individually in a single tray, largely eliminating the need for labeling. Tissue samples of young rosette leaves can be collected in 96-well microtiter plates establishing a one-to-one correspondence between plants and DNA samples. If this is done while the plants are young, preliminary mapping can be finished before the plants bolt. The alkaline-lysis protocol for DNA extraction allows a single individual to process several plates, or several hundred samples, per day (Klimyuk et al., 1993; Table I, Supplemental material). It thus becomes fairly easy to complete the analysis of a large mapping population in about a month. To avoid problems due to cross-contamination or sampling mistakes, the genotype of all recombinant plants retained for further analysis should be confirmed using a second DNA preparation. Similarly, seed samples of all relevant plants should be saved.

MAKING USE OF VISIBLE AND SELECTABLE MARKERS

In the above scheme a large mapping population is analyzed only with molecular markers. Often, however, there are convenient visible markers that allow selection of recombinant chromosomes with more ease. A number of mapping strains that contain multiple visible markers on the same chromosome are available from the stock centers (Table I). Using such marker strains, mutations in the *CYT1* gene have been mapped between the visible markers *as* and *cer8*, an interval of about 10% recombination on chromosome II. Since the *cyt1* mutation is in a Ws background, whereas the *as cer8* marker chromosome is a mixture of Landsberg *erecta* and Columbia, it was possible to assay molecular markers and visible markers in the same mapping population. By selecting plants that were mutant for *as* but wild-type looking for *cer*, more than 500 chromosomes with a recombination event close to the *cyt1* mutation were collected. Subsequent analysis of molecular markers mapped the *cyt1* mutation to a DNA fragment of less than 10 kb (W. Lukowitz and C. Somerville, unpublished results). In the case of the *CONSTANS* gene, chromosomes were created carrying visible markers and the *constans* mutation in cis. These chromosomes were then crossed to the Niederzenz accession to generate a mapping population where visible and where molecular markers could be examined (Putterill et al., 1993). Whether or not the use of visible markers is advantageous depends very much on circumstances, however, there are several possible pitfalls. The genetic background of marker strains is often difficult to reconstruct; typically it is Landsberg *erecta* or a mixture of Landsberg *erecta* and Columbia. In most cases visible and molecular markers have been mapped using different mapping populations. Consequently, the map position of visible markers with respect to molecular markers is often not exactly known and needs to be confirmed or determined from first principles before they can be used in a fine mapping experiment. Finally, generating a second mapping population is time consuming, especially if appropriately marked chromosomes have to be constructed first (as in the case of the *CONSTANS* gene). In most cases these drawbacks will not outweigh the potential benefits.

Dominant selectable markers, such as resistance genes present in T-DNAs and transposons, have also been used in fine mapping experiments. To find recombinations in the vicinity of the *AGR1* gene, for example, the mutant (Ws accession) was crossed to a transgenic line (Nossen accession) in which a stable Dissociation element conferring hygromycin resistance was inserted close to the *AGR1* locus (Chen et al., 1998). In the F_2, *AGR1* mutants were selected that were also resistant to hygromycin and thus recombinant between the *AGR1* gene and the flanking Dissociation element. At present, large databases containing sequenced insertion sites of T-DNAs and transposons are being generated as a tool for reverse genetics (Parinov et al., 1999; Speulman et al., 1999; Tissier et al., 1999; Table I, no. 9). This information will also make a large number of dominant selectable genes spread throughout the genome available, most of which are stable and could be used for mapping.

CREATING NEW MOLECULAR MARKERS

To take full advantage of recombinant chromosomes for fine mapping, it is necessary to create new molecular markers closer and closer to the mutation. There are two ways of approaching the problem: either guessing or comparing the DNA sequence of the two accessions involved. Since simple repetitive elements are hypervariable (Tautz, 1989), guessing works surprisingly well in the case of SSLP markers. Single-, di-, or trinucleotide repeats of 20 or more bp in length are on average found every 10 or 20 kb and provide good targets for new SSLP markers. As shown in Table III, there is a likelihood of about 40% that such fragments will be polymorphic between different accessions when analyzed on high-resolution agarose gels. Most groups of the Arabi-

Table III. *Variability of short repetitive sequences between the Columbia and the Landsberg erecta accessions*[a]

Type of Repeat	Length of Repeat	Loci Tested	Polymorphic Loci
	nucleotides	*no.*	
$(A)_n$ or $(T)_n$	<20	9	2
	≥20	41	11
$(AT)_n$	<20	7	6
	≥20	91	50
$(GA)_n$ or $(CT)_n$	<20	6	1
	≥20	10	4
$(GT)_n$ or $(CA)_n$	<20	2	0
	≥20	2	2
Others[b]	<20	3	0
	≥20	7	4

[a] Data generated by the authors. [b] Trinucleotide repeats, AT- or purine-rich stretches >30 nucleotides.

dopsis Genome Initiative are identifying these sequences in their annotations, and TAIR displays all simple-sequence repeats that are longer than 30 nucleotides as part of the sequence tables (Table I; the repeats are listed with 200 bp of flanking DNA to facilitate the design of PCR primers). In most cases the flanking DNA is complex enough to allow PCR-amplification of a short fragment that includes the repeat.

An even larger source of polymorphisms is provided by nucleotide exchanges. For the Ler X Col combination, sequence information deposited in various databases can be utilized to search for such polymorphisms. The sequence of the Columbia accession is, of course, determined by the Arabidopsis Genome Initiative. The Stanford DNA Sequence and Technology Center (Stanford, CA) has deposited a collection of 412 simple nucleotide polymorphisms (SNPs) between Landsberg *erecta* and Columbia at TAIR (Table I). All of these polymorphisms have been confirmed and many of them have been used for mapping (Cho et al., 1999). The Institute for Genomic Research (Rockville, MD) has assembled a collection of about 15,000 random sequences of approximately 500 bp from the Landsberg *erecta* accession that can be searched with a BLAST tool (Table I, no. 6). However, most of these sequences are single-pass and the polymorphisms have not been further confirmed (indeed, we have had mixed experiences when using these data). Cereon Genomics (Cambridge, MA) has made an effort to sequence the Landsberg *erecta* accession in a whole-genome shotgun approach (Rounsley et al., 1999). The average coverage was about 3-fold, suggesting that the sequence information is near complete. This data has been used to identify approximately 40,000 SNPs between Landsberg *erecta* and Columbia, including single basepair changes and small insertion/deletion events. This truly extensive collection of molecular polymorphisms has been made accessible for the academic and non-profit sector through TAIR and will be updated as the sequence of the Columbia accession is being completed (Table I).

For all other accessions, there is very little existing sequence information, so that it becomes necessary to amplify introns or stretches of intragenic regions by PCR and determine their sequence to find polymorphisms. In many cases the polymorphisms will be in a restriction site creating a target for a CAPS marker; about 30% of the SNPs present in the TAIR database, for example, affect a restriction site (Table I). Most other point mutations can be exploited for mapping by designing dCAPS markers. This technique makes use of partly mismatched PCR primers to artificially create a restriction site at a polymorphic position (Fig. 1), which then can be analyzed in the same way as in the case of CAPS markers. Detailed descriptions of how to create dCAPS markers have been published (Michaels and Amasino, 1998; Neff et al., 1998), and a computer program that finds appropriate enzymes is also available (requests should be addressed to M. Neff, Washington University, St. Louis, MO 63130; e-mail mneff@biology.wustl.edu).

COMPLICATIONS AND TROUBLESHOOTING

The complication most likely to be encountered in the analysis of naturally occurring variations is that a given trait is influenced by more than one locus. Resistance to powdery mildew in a cross between the Kashmir-1 (resistant) and Columbia (susceptible) accessions, for instance, was found to involve at least three genetic loci acting in an additive fashion (I. Wilson, C. Schiff, and S. Somerville, personal communication). Fine mapping any one of these resistance genes requires reduction of the genetic complexity of the mapping population, for example, by creating recombinant-inbred lines in which only one of the loci remains polymorphic. A large variety of traits are influenced by one or more genetic loci in crosses between Arabidopsis accessions, including flowering time, seed size, dormancy, circadian rhythms, secondary metabolism, and trichome density (for review, see Alonso-Blanco and Koornneef, 2000). Whenever natural or induced mutations affecting these traits are mapped, second site modifiers might interfere with the analysis.

Epigenetic mutations, a term describing heritable changes in the expression and function of a gene that are not due to changes in the DNA sequence (for review, see Wolffe and Matzke, 1999), are another potential complication for positional cloning projects. Well-documented examples are the epigenetic *clark kent* alleles of the flower development gene *SUPERMAN* (Jacobson and Meyerowitz, 1997). These alleles are heritable but unstable and revert with a low frequency. All of them show similar patterns of cytosine methylation within the DNA sequence of the *SUPERMAN* gene and possibly as a consequence, exhibit reduced expression of *SUPERMAN* tran-

scripts. None of them is associated with a change in the *SUPERMAN* DNA sequence; however, they can be complemented by a transgene carrying a copy of the *SUPERMAN* gene. At present, not much is known about how such epigenetic mutations arise and with what frequency they can be expected.

The ratio between physical and genetic distance varies with respect to the location on the chromosome. In general these variations are small and have relatively little impact on mapping resolution (Copenhaver et al., 1998). However, there is evidence that some chromosomal regions are exceptional. For example, positional cloning of the *GURKE* gene, which maps close to the centromere of chromosome I, has proven very difficult; recombination was severely suppressed in the vicinity of the centromere, making fine mapping efforts extremely ineffective. Furthermore, the prevalence of repetitive DNA elements in this region made identification of interspersed single-copy sequences that can be used to generate genetic markers problematic (R. Torres-Ruiz, personal communication). This observation is confirmed by a systematic analysis of the ratio between genetic and physical distance on chromosome II (Lin et al., 1999). For almost the entire sequence of the chromosome, a genetic distance of 1% recombination corresponds to a physical distance of 100 to 400 kb, with an average of 250 kb. A striking exception, however, is the centromere region, where 1% recombination appears to correspond to 1,000 to 2,500 kb. It seems noteworthy to point out that none of the five Arabidopsis centromeres is fully covered in the existing physical maps. A recent analysis of the centromere regions indicates that they contain mostly repetitive DNA and very few expressed genes (Copenhaver et al., 1999). Thus, there should be few Arabidopsis genes that are not amenable to a positional cloning strategy because of their proximity to the centromere.

Apart from the centromere, there is a second short segment on chromosome II where a genetic distance of 1% recombination appears to correspond to 1,000 kb or more. It has been speculated that the observed low recombination rates might be due to rearrangements in the DNA sequence of the accessions used for the mapping analysis (Lin et al., 1999; Mayer et al., 1999). A comparison of the DNA sequences of chromosome II and IV has revealed several stretches that are duplicated between the two chromosomes (one of them is 4.6 Mbp in size), as well as a recent transfer of DNA from the mitochondrial genome to chromosome II (Lin et al., 1999). These findings clearly demonstrate that the structure of the Arabidopsis genome is subject to continual changes. Accordingly, the genetic variability between the accessions might very well extend beyond point mutations and include DNA rearrangements, a potential cause of severe problems for positional cloning projects. If, for example, a sequence of about 500 kb was inverted between two accessions used to create a mapping population, all recombination events within the inversion would create non-viable meiotic products. Consequently, it would be impossible to map mutations within the inverted sequence. No such DNA rearrangements between the commonly used accessions have been reported at present, but this may well be because they are difficult to detect. In a mapping experiment their presence would most likely go unnoticed until the last steps.

Occasionally, T-DNA insertions and irradiation have also been observed to cause DNA rearrangements (Shirley et al., 1992; Nacry et al., 1998; Laufs et al., 1999; Ogas et al., 1999). Thus, similar difficulties might occur when mutations are mapped that were generated by these methods. In such cases, however, there is at least a fair probability that the mutations are associated with one or both breakpoints of the rearrangement.

PERSPECTIVE

We hope it has become clear that positional cloning in Arabidopsis is no longer for a few dedicated (and enduring) specialists, but rather for everybody. Over the past years many inexpensive but powerful tools have been developed and a vast amount of information has been assembled in freely accessible databases. Utilizing these resources, most positional cloning projects should now be predictable, straightforward, and simple. The situation will further improve, as our growing knowledge about the structure and flux of the Arabidopsis genome will help to eliminate some of the remaining complications mentioned above, or at least make them more manageable. We are now in a position where genes defined by naturally occurring variations or induced point mutations can be cloned effectively and fast, but this, of course, is only the beginning of the story.

ACKNOWLEDGEMENTS

We are thankful to C. Somerville for his encouragement, generous support, and helpful discussions. Many thanks also to I. Wilson, C. Schiff, S. Somerville, and R. Torres-Ruiz for communicating unpublished results and to A. Krapp, S. Rhee, R. Silady, and an anonymous reviewer for helpful comments on the manuscript. Our apologies to all colleagues whose work was not included due to limited space.

Received March 15, 2000; accepted April 27, 2000.

LITERATURE CITED

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

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics **19:** 137–144

Busch M, Mayer U, Jürgens G (1996) Molecular analysis of the *Arabidopsis* pattern formation gene *GNOM*: gene structure and intragenic complementation. Mol Gen Genet **250:** 681–691

Chang C, Bowman JL, DeJohn AW, Lander ES, Meyerowitz EM (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc Natl Acad Sci USA **85:** 6856–6860

Chen R, Hilson P, Sedbrook J, Rosen E, Caspar T, Masson PH (1998) The *Arabidopsis thaliana AGRAVITROPIC 1* gene encodes a component of the polar-auxin-transport efflux carrier. Proc Natl Acad Sci USA **95:** 15112–15117

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang W-H, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat Genet **23:** 203–207

Copenhaver GP, Browne WE, Preuss D (1998) Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads. Proc Natl Acad Sci USA **95:** 247–252

Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin XY, Bevan M, Murphy G, Harris B, Parnell LD, McCombie WR, Martienssen RA, Marra M, Preuss D (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science **286:** 2468–2474

Feldmann KA (1991) T-DNA insertional mutagenesis in *Arabidopsis*: mutational spectrum. Plant J **1:** 71–82

Hardtke CS, Berleth T (1998) The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. EMBO J **17:** 1405–1411

Hardtke CS, Müller J, Berleth T (1996) Genetic similarity among *Arabidopsis thaliana* ecotypes estimated by DNA sequence comparison. Plant Mol Biol **32:** 915–922

Hauser M-T, Adhami F, Dorner M, Fuchs E, Glössl J (1998) Generation of co-dominant PCR-based markers by duplex analysis on high-resolution gels. Plant J **16:** 117–125

Jacobson SE, Meyerowitz EM (1997) Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. Science **277:** 1100–1103

Klimyuk VI, Carroll BJ, Thomas CM, Jones JDG (1993) Alkali treatment for rapid preparation of plant material for reliable PCR analysis: technical advance. Plant J **3:** 493–494

Konieczny A, Ausubel F (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J **4:** 403–410

Laufs P, Autran D, Traas J (1999) A chromosomal paracentric inversion associated with T-DNA integration in *Arabidopsis*. Plant J **18:** 131–139

Leung J, Bouvier-Durand M, Morris P-C, Guerrier D, Chefdor F, Giraudat J (1994) *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulated protein phosphatase. Science **264:** 1448–1452

Li J, Chory J (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell **90:** 929–938

Lin X, Kaul S, Rounsley S, Shea TP, Benito M-I, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, Feldblyum TV, Buell CR, Ketchum KA, Lee J, Ronning CM, Koo HL, Moffat KS, Cronin LA, Shen M, Pai G, Van Aken S, Umayam L, Tallon LJ, Gill JE, Venter JC (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature **402:** 761–768

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

Liu YG, Shirano Y, Fukaki H, Yanai Y, Tasaka M, Tabata S, Shibata D (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. Proc Natl Acad Sci USA **96:** 6535–6540

Lukowitz W, Meyer U, Jürgens G (1996) Cytokinesis in the *Arabidopsis* embryo involves the syntaxin-related *KNOLLE* gene product. Cell **84:** 61–71

Mayer K, Schüller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Düsterhöft A, Stiekema W, Entian K-D, Terryn N, Harris B, Ansorge W, Brandt P, Grivell L, Rieger M, Weichselgartner M, de Simone V, Obermaier B, Mache R, Müller M, Kreis M, Delseny M, Puigdomenech P, Watson M, Schmidtheini T, Reichert B, Portatelle D, Perez-Alonso M, Boutry M, Bancroft I, Vos P, Hoheisel J, Zimmermann W, Wedler H, Ridley P, Langham S-A, McCullagh B, Bilham L, Robben J, Van der Schueren J, Grymonprez B, Chuang Y-J, Vandenbussche F, Braeken M, Weltjens I, Voet M, Bastiaens I, Aert R, Defoor E, Weitzenegger T, Bothe G, Ramsperger U, Hilbert H, Braun M, Holzer E, Brandt A, Peters S, van Staveren M, Dirkse W, Mooijman P, Klein Lankhorst R, Rose M, Hauf J, Kötter P, Berneiser S, Hempel S, Feldpausch M, Lamberth S, Van den Daele H, De Keyser A, Buysshaert C, Gielen J, Villarroel R, De Clercq R, Van Montagu M, Rogers J, Cronin A, Quail M, Bray-Allen S, Clark L, Doggett J, Hall S, Kay M, Lennard N, McLay K, Mayes R, Pettett A, Rajandream M-A, Lyne M, Benes V, Rechmann S, Borkova D, Blöcker H, Scharfe M, Grimm M, Löhnert T-H, Dose S, de Haan M, Maarse A, Schäfer M, Müller-Auer S, Gabel C, Fuchs M, Fartmann B, Granderath K, Dauner D, Herzl A, Neumann S, Argiriou A, Vitale D, Liguori R, Piravandi E, Massenet O, Quigley F, Clabauld G, Mündlein A, Felber R, Schnabl S, Hiller R, Schmidt W, Lecharny A, Aubourg S, Chefdor F, Cooke R, Berger C, Montfort A, Casacuberta E, Gibbons T, Weber N, Vandenbol M, Bargues M, Terol J, Torres A, Perez-Perez A, Pernelle B, Bent E, Johnson S, Tacon D, Jesse T, Heijnen L, Schwarz S, Scholler P, Heber S, Francs P, Bielke C, Frishman D, Haase D, Lemcke K, Mewes HW, Stocker S, Zaccaria P, Bevan M, Wilson RK, de la Bastide M, Habermann K, Parnell L, Dedhia N, Gnoj L, Schutz K, Huang E, Spiegel L, Sehkon M, Murray J, Sheet P, Cordes M, Abu-Threideh J, Stoneking T, Kal-

icki J, Graves T, Harmon G, Edwards J, Latreille P, Courtney L, Cloud J, Abbott A, Scott K, Johnson D, Minx P, Bentley D, Fulton B, Miller N, Greco T, Kemp K, Kramer J, Fulton L, Mardis E, Dante M, Pepin K, Hillier L, Nelson J, Spieth J, Ryan E, Andrews S, Geisel C, Layman D, Du H, Ali J, Berghoff A, Jones K, Drone K, Cotton M, Joshu C, Antonoiu B, Zidanic M, Stong C, Sun H, Lamar B, Yordan C, Ma P, Zhong J, Preston R, Vil D, Shekher M, Matero A, Shah R, Swaby I'K, O'Shaughnessy A, Rodriguez M, Hoffman J, Till S, Granat S, Shohdy N, Hasegawa A, Hameed A, Lodhi M, Johnson A, Chen E, Marra M, Martienssen R, McCombie WR (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature **402:** 769–777

Mayer KFX, Schoof H, Häcker A, Lenhard M, Jürgens G, Laux T (1998) Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. Cell **95:** 805–815

Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M (1998) *Arabidopsis thaliana*: a model plant for genome analysis. Science **282:** 662–682

Meyer K, Leube MP, Grill E (1994) A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. Science **264:** 1452–1455

Michaels SD, Amasino RM (1998) A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. Plant J **14:** 381–385

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA **88:** 9828–9832

Moussian B, Schoof H, Häcker A, Jürgens G, Laux T (1998) Role of the *ZWILLE* gene in the regulation of central shoot meristem cell fate during *Arabidopsis* embryogenesis. EMBO J **17:** 1799–1809

Murray MG, Thompson WF (1980) Rapid isolation of high-molecular-weight plant DNA. Nucleic Acids Res **8:** 4321–4325

Nacry P, Camillieri C, Courtial B, Caboche M, Bouchez D (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. Genetics **149:** 641–650

Nataraj AJ, Olivos-Glander I, Kusukawa N, Highsmith WE (1999) Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. Electrophoresis **20:** 1177–1185

Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. Plant J **14:** 387–392

O'Donovan M, Oefner PJ, Roberts SC, Austin J, Hoogendoorn B, Guy C, Speight G, Upadhyaya M, Sommer SS, McGuffin P (1998) Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. Genomics **52:** 44–49

Ogas J, Kaufmann S, Henderson J, Somerville C (1999) *PICKLE* is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. Proc Natl Acad Sci USA **96:** 13839–13844

Parinov S, Sevugan M, Ye D, Yang W-C, Kumaran M, Sundaresan V (1999) Analysis of flanking sequences from *Dissociation* insertion lines: a database for reverse genetics in *Arabidopsis*. Plant Cell **11:** 2263–2270

Ponce MR, Robles P, Micol JL (1999) High-throughput mapping in *Arabidopsis thaliana*. Mol Gen Genet **261:** 408–415

Putterill J, Robson F, Lee K, Coupland G (1993) Chromosome walking with YAC clones in *Arabidopsis*: isolation of 1700 kb of contiguous DNA on chromosome 5, including a 300 kb region containing the flowering-time gene *CO*. Mol Gen Genet **239:** 145–157

Rounsley S, Subramaniam S, Cao Y, Bush D, DeLoughery C, Field C, Phillips C, Iartchouk O, Last R, Wiegand R, Fischhoff D, Timberlake B (1999) Whole genome shotgun sequencing of *Arabidopsis*: analysis and insight (abstract 3-1). Proceedings of the 10th International Conference on *Arabidopsis* Research, July 4-8, 1999, University of Melbourne, Australia

Ryals J, Weymann K, Lawton K, Friedrich L, Ellis D, Steiner HY, Johnson J, Delaney TP, Jesse T, Vos P, Uknes S (1997) The *Arabidopsis* NIM1 protein shows homology to the mammalian transcription factor inhibitor I κ B. Plant Cell **9:** 425–439

Sakai H, Medrano LJ, Meyerowitz EM (1995) Role of *SUPERMAN* in maintaining *Arabidopsis* floral whorl boundaries. Nature **378:** 199–203

Shirley BW, Hanley S, Goodman HM (1992) Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis transparent-testa* mutations. Plant Cell **4:** 333–347

Speulman E, Metz PLJ, van Arkel G, Hekkert BTL, Stiekema WJ, Pereia A (1999) A two-component *Enhancer-Inhibitor* transposon mutagenesis system for functional analysis of the *Arabidopsis* genome. Plant Cell **11:** 1853–1866

Tautz D (1989) Hypervariability of simple sequences as a general source of polymorphic DNA markers. Nucleic Acids Res **17:** 6463–6471

Taylor GR (1999) Enzymatical and chemical cleaving methods. Electrophoresis **20:** 1125–1130

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JDG (1999) Multiple independent defective *Suppressor-mutator* transposon insertions in *Arabidopsis*: a tool for functional genomics. Plant Cell **11:** 1841–1852

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res **23:** 4407–4414

Wolffe AP, Matzke MA (1999) Epigenetics: regulation through repression. Science **286:** 481–486

Zhong R, Ye Z-H (1999) *IFL*1, a gene regulating interfascicular fiber differentiation in *Arabidopsis*, encodes a homeodomain-leucine zipper protein. Plant Cell **11:** 2139–2152

