A Simple Procedure for the Analysis of Single Nucleotide Polymorphisms Facilitates Map-Based Cloning in Arabidopsis[1]

Eliana Drenkard, Brent G. Richter[2], Steve Rozen, Lisa M. Stutius, Nathaniel A. Angell, Michael Mindrinos, Raymond J. Cho, Peter J. Oefner, Ronald W. Davis, and Frederick M. Ausubel*

Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114 (E.D., B.G.R., N.A.A., F.M.A.); Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142 (S.R.); Harvard College, Cambridge, Massachusetts 02138 (L.M.S.); and Departments of Biochemistry (M.M., R.J.C., P.J.O., R.W.D) and Genetics (R.W.D), Stanford University School of Medicine, Stanford, California 94305

We developed a modified allele-specific PCR procedure for assaying single nucleotide polymorphisms (SNPs) and used the procedure (called SNAP for single-nucleotide amplified polymorphisms) to generate 62 Arabidopsis mapping markers. SNAP primers contain a single base pair mismatch within three nucleotides from the 3′ end of one allele (the specific allele) and in addition have a 3′ mismatch with the nonspecific allele. A computer program called SNAPER was used to facilitate the design of primers that generate at least a 1,000-fold difference in the quantity of the amplification products from the specific and nonspecific SNP alleles. Because SNAP markers can be readily assayed by electrophoresis on standard agarose gels and because a public database of over 25,000 SNPs is available between the Arabidopsis Columbia and Landsberg *erecta* ecotypes, the SNAP method greatly facilitates the map-based cloning of Arabidopsis genes defined by a mutant phenotype.

Map-based positional cloning in Arabidopsis is a standard but, until recently, time-consuming and expensive procedure for the isolation of genes defined by mutation. The main obstacle encountered in map-based cloning approaches is the insufficient number of PCR-based molecular markers available to perform fine-structure mapping. However, public accessibility to the complete and annotated sequence of the Arabidopsis genome should greatly facilitate map-based cloning, because sequence information will provide the tools necessary for the creation of new molecular markers (Lukowitz et al., 2000). The recent public release by Cereon Genomics (Cambridge, MA) of a database that will soon comprise most of the single nucleotide polymorphisms (SNPs) and small insertions/deletions (InDels) DNA polymorphisms between the Columbia and Landsberg *erecta* ecotypes promises to greatly expedite the process of creating PCR-based markers. Combined with simple methods for detecting SNPs and InDels, this database has the potential to transform map-based cloning from a tedious process to a routine procedure that can be accomplished in a few months.

Based on current estimates, the number of InDels between the Columbia and Landsberg *erecta* ecotypes is approximately 21,000, an average of one InDel every 6.1 kb (S. Rounsley, personal communication). InDel polymorphisms found in the Columbia and Landsberg *erecta* ecotypes are frequently polymorphic in other Arabidopsis accessions. Therefore, it should be possible to use the Columbia/Landsberg InDel markers to find a polymorphism in a particular chromosomal region in any pair of accessions (G. Jander, personal communication). InDels are easily detected by amplifying a small region of the genome containing the insertion/deletion element and determining the size of the amplification product. On the other hand, the main drawback of InDel markers is that the differences in fragment lengths of the insertion/deletion elements are sometimes very small. Therefore, the discrimination of the PCR products corresponding to particular InDel alleles requires more sophisticated methods than standard agarose gel electrophoresis (e.g. denaturing PAGE).

SNPs comprise the largest set of sequence variants in most organisms, including Arabidopsis (Cho et al., 1999). For example, in comparing Arabidopsis Columbia and Landsberg *erecta* genomic sequences, Cereon Genomics identified on average one SNP every 3.3 kb (S. Rounsley, personal communication). This translates into approximately 40,000 SNPs for a 130-Mb genome. Two widely used types of PCR molecular markers based on SNPs are cleaved amplified polymorphic sequences (CAPS; Konieczny and Ausubel, 1993) and derived CAPS (dCAPS; Michaels and Amasino, 1998; Neff et al., 1998). CAPS markers detect polymorphisms that occur in restriction sites and

[1] This work was supported by the National Science Foundation (grant no. MCB–9729599).

[2] Present address: Genomics Collaborative Inc., 99 Erie Street, Cambridge, MA 02139.

* Corresponding author; e-mail ausubel@frodo.mgh.harvard.edu; fax 617–726–5949.

dCAPS makers are created during PCR amplification by introducing a restriction site at the site of an SNP using specially designed primers. SNPs can also be detected using allele-specific PCR primers designed such that the 3′ nucleotide of a primer corresponds to the site of the SNP (Ugozzoli and Wallace, 1991). Thus, the allele-specific primer matches perfectly with one allele (the specific allele) and has a 3′ mismatch with the nonspecific allele. Because mismatched 3′ termini are extended by DNA polymerases with much lower efficiency than correctly matched termini (Petruska et al., 1988), the allele-specific primer preferentially amplifies the specific allele (Cha et al., 1992). Detection of SNPs by allele-specific PCR shares many of the advantages of CAPS and dCAPS, including codominance (as long as an allele-specific primer is used for each allele) and the convenience of detecting the polymorphism on standard agarose gels. Moreover, allele-specific PCR procedures have the additional benefit that restriction digestion is not necessary after PCR amplification. The reason that allele-specific PCR is not widely used to assay SNPs is that in most cases a single base pair change at the 3′ end of the nonspecific allele is not sufficient to create reliable discrimination between the two alleles (Kwok et al., 1990, 1994; Cha et al., 1992).

In this paper, we describe the use of a modified allele-specific PCR procedure for assaying SNPs that is sufficiently robust to easily discriminate between the specific and nonspecific alleles. This new procedure facilitates the rapid and reliable creation and analysis of large numbers of molecular markers using simple methods common to most molecular biology laboratories.

RESULTS

Allele-Specific PCR Strategy

Allele-specific PCR, the basis for the PCR-based mapping strategy described in this paper, is illustrated in Figure 1. The technique utilizes primers with specific mismatches at the 3′ end that allow preferential amplification of one allele relative to another on account of the primers being complementary to the site of a DNA sequence variation (SNP; Ugozzoli and Wallace, 1991; Cha et al., 1992). As shown in Figure 1, primer P1 forms a perfect match with allele 1, but the 3′-terminal nucleotide forms a mismatch with the DNA sequence of allele 2. P1 consequently amplifies allele 1 more efficiently than allele 2 (Fig. 1, A and B). Primer P2 similarly forms a mismatch with the DNA sequence of allele 1 and therefore preferentially amplifies allele 2 (Fig. 1, A and B). Organisms that are heterozygous for the two alleles show comparable levels of amplification with primers P1 and P2 (Fig. 1B). A significant problem with employing this traditional allele-specific amplification method for SNP detection is that under normal PCR conditions different mismatches located at the 3′ end are extended with different efficiencies by *Taq* polymerase (Newton et al., 1989; Kwok et al., 1990; Li et al., 1990). Therefore, some of the primers do not display the required allele specificity (Kwok et al., 1990, 1994; Cha et al., 1992). This represents a major obstacle for the use of this approach in the design of molecular markers.

To overcome this problem, we used a modification of the original allele-specific PCR methodology, in which an additional base pair change is introduced within the last four bases of the primer (Newton et

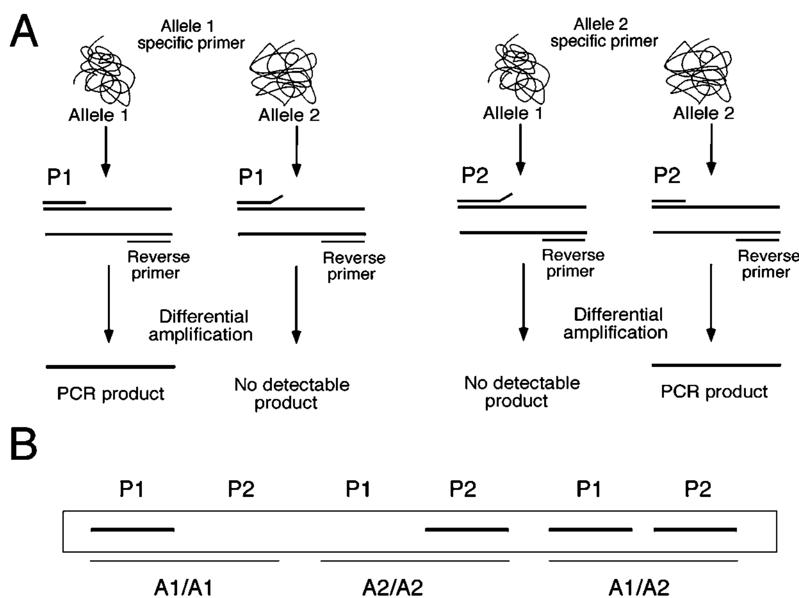

Figure 1. Schematic representation of the allele-specific codominant PCR strategy. Oligonucleotide primers with 3′ nucleotides that correspond to an SNP site are used to preferentially amplify specific alleles. A, Primer P1 forms a perfect match with allele 1 but forms a mismatch at the 3′ terminus with the DNA sequence of allele 2. Primer P2 similarly forms a perfect match with allele 2 and a 3′ terminus mismatch with allele 1. B, Schematic of agarose gel analysis showing the expected outcome for the amplification of organisms homozygous and heterozygous for both alleles using primers P1 and P2. P1, Primer 1; P2, primer 2; A1, allele 1; A2, allele 2.

al., 1989). The addition of the extra mismatch, coupled with the presence of the natural mismatch at the 3′ end, produces a dramatic reduction in the PCR product yield of the nonspecific allele but has a relatively minor effect on the amplification of the specific allele (Kwok et al., 1990). This modification produces a significant increase in the specificity of the primers (Cha et al., 1992; Kwok et al., 1994). We call the PCR-based markers generated using this modified allele-specific amplification methodology SNAP for single nucleotide amplified polymorphisms (E. Drenkard, S. Rozen, M. Mindrinos, B.G. Richter, and F.M. Ausubel, unpublished data).

Design of Allele-Specific Primers

The main obstacle to the application of the SNAP procedure in the generation of molecular markers is the determination of which additional mismatches to introduce to obtain the required primer specificity. To facilitate the process of primer design, a computer program was written based on a set of empirical data that evaluates the effect of the addition of different mismatch alternatives on PCR amplification (Drenkard et al., article in preparation; http://patho.mgh.harvard.edu/ausubelweb). The program, called SNAPER, generates a list of up to 32 possible primers per SNP site (16 alternatives for each allele) that contain an additional mismatch within the three bases closest to the 3′ end. Along with the allele-specific primers for the SNP site, the program also generates a second (reverse) primer that contains no mismatches. The program provides information concerning the likelihood that the primer will be allele-specific, predicted by empirical data, and the position and type of base pair change introduced to generate the additional mismatch in the primer.

Because the work described in this paper was initiated well before the release of the Cereon SNP database, we based the design of our initial SNAP primers on a set of 487 loci containing SNPs between the Columbia and Landsberg *erecta* ecotypes identified by Cho et al. (1999). We used the SNAPER program to design SNAP primers for both the Columbia and Landsberg alleles for 43 SNPs.

Testing of Allele-Specific Primer Sets

During previous work, we determined experimental conditions that would assess primer specificity given the range of template DNA concentrations typically encountered in map-based cloning. We found that primers that showed specificity in two sets of PCR reactions that differed by 10 PCR cycles were specific over a 1,000-fold range of template DNA concentration. Therefore, we could test primer specificity by running two identical sets of PCR reactions per primer pair, in which the PCR reactions differ only in the number of cycles used during the amplification process (Drenkard et al., article in preparation). We verified experimentally that primer pairs displaying the presence of a PCR product on agarose gels for the specific allele and the absence of a PCR product for the nonspecific allele in both sets of PCR reactions were suitable for use as molecular markers (E. Drenkard, S. Rozen, M. Mindrinos, B.G. Richter, and F.M. Ausubel, unpublished data).

A total of 331 SNAP primer pairs corresponding to 43 SNPs were tested using both 28 and 38 cycles of PCR amplification. From the list of primers generated by the SNAPER program for each SNP, we selected and tested approximately 8 primer pairs per allele, which represented approximately 16 primer pairs per marker (for some SNPs fewer alternatives were generated by the program). The products amplified by PCR were analyzed by agarose gel electrophoresis, and the presence or absence of bands in both sets of reactions was scored for each primer pair. Figure 2 shows the results obtained with primer pairs 1-4, 1-6, 4-3, and 2-6 using 28 and 38 cycles of PCR amplification. Primer pairs 1-4 (Landsberg-specific) and 1-6 (Columbia-specific) generated an amplification product for the specific allele and no amplification product for the nonspecific allele at both 28 and 38 cycles, indicating that the primers have the required specificity. Primer pair 4-3 (Landsberg-specific) generated amplification products for both the specific and nonspecific allele at 38 cycles, suggesting that at conditions of high template DNA concentration the primer could produce a false positive. Primer pair 2-6 (Columbia-specific) failed to generate an amplification product for the specific allele at 28 cycles, indicating that when using low template DNA concentrations the primer could produce a false negative.

For 43 different SNPs examined, the SNAPER program had an overall success rate of approximately 53% in generating primers with the desired range of specificity. Table I shows a list of 33 SNAP markers that we have generated to date using the program. Information about the markers generated is also available at the Ausubel laboratory web site (http://patho.mgh.harvard.edu/ausubelweb) and on The Arabid-

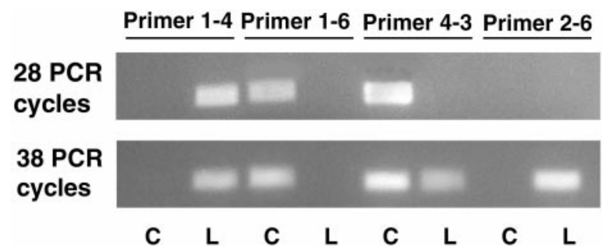

Figure 2. Analysis of specificity for representative SNAP primer pairs. Lanes 1 and 2, Primer pair 1-4 (Landsberg-specific); lanes 2 and 3, primer pair 1-6 (Columbia-specific); lanes 4 and 5, primer pair 4-3 (Landsberg-specific); lanes 5 and 6, primer pair 2-6 (Columbia-specific). PCR reactions for each allele-specific primer were performed using Columbia (C) and Landsberg *erecta* (L) DNA for 28 or 38 PCR cycles.

opsis Information Resource (TAIR) web site (http://www.Arabidopsis.org/). Seventeen of the markers listed in Table I were developed prior to the creation of the SNAPER program and are part of the empirical database used to generate the rules underlying the SNAPER algorithm.

Mapping of SNAP Markers onto the Columbia-Landsberg Recombinant Inbred (RI) Map

The chromosomal locations of 33 of the SNPs corresponding to SNAP markers in Table I were previously mapped, using an Affymetrix-based mapping technique, to unique chromosomal positions and integrated into the existing Arabidopsis RI linkage map (Table I; Cho et al., 1999). To demonstrate the feasibility of using SNAP primers in a large-scale mapping experiment, 17 SNAP markers with unknown map locations were mapped using 94 of the Lister/Dean RI lines (Lister and Dean, 1993). Figure 3 shows a representative example of the type of amplification data that were obtained in this large-scale experiment that involved 3,264 PCR reactions. The PCR reactions were carried out for 35 cycles because empirical testing of a subset of the markers at 34, 35, and 36 cycles showed that 35 cycles yielded the highest level of discrimination. Amplified products were evaluated by agarose gel electrophoresis, scored, and analyzed using the MAPMAKER software (Lander et al., 1987). Among the 3,264 PCR reactions, only 90 (2.7%) produced ambiguous data.

Map positions for all 50 markers are shown in Table I and are also available at the Nottingham Arabidopsis Stock Centre (http://nasc.nott.ac.uk/RI_data/RI_menu.html). The 50 markers are in general well scattered throughout the five Arabidopsis chromosomes with an average of 10 markers per chromosome (Table I). BAC, YAC, and P1 clones containing the markers described in this paper were identified in the TAIR database (Table I; http://www.Arabidopsis.org/blast/). Approximate physical map positions for these clones, obtained from The Institute of Genomic Research web site (http://www.tigr.org/), were compared with the genetic map positions obtained in this paper (Table I) for all 5 chromosomes. The correlation coefficients (r^2) between the genetic and physical map positions were all high for the markers located on chromosomes 1, 2, 3, and 4 (0.999, 0.977, 0.986, and 0.990, respectively; Fig. 4A). We were unable to obtain good estimates for physical map positions on chromosome 5 because information on the extent of the overlap that exists between Arabidopsis clones was not available. Nevertheless, based on a less rigorous analysis, we found some discrepancies between the genetic and physical maps of chromosome 5. Genetic map positions were reversed with respect to physical map positions for at least 3 of the markers located below 93.07 cM (Fig. 4B). Moreover, although markers SGCSNP101 and m558a reside in the same clone (MUA2) they differ considerably in their genetic map positions (142.02 and 113.80 cM, respectively).

Mapping of *edr5-1*

To assess the feasibility of the use of SNAP markers in map-based cloning approaches, we localized the mutation *edr5-1*, which causes enhanced disease resistance to the virulent bacterial pathogen *Pseudomonas syringae* pv *maculicola* ES4326 and to the obligate fungal pathogen *Erysiphe orontii*. The *edr5-1* mutant was isolated in the Ausubel laboratory in a genetic screen to identify Arabidopsis mutants in the Columbia accession with altered susceptibility to *P. syringae* pv *maculicola* ES4326 (Volko, 1998). The mutated *edr5-1* gene was shown to segregate as a single recessive Mendelian trait by backcrossing to the wild-type Columbia parent. To generate the mapping population, homozygous *edr5-1* plants were crossed to Landsberg *erecta*, and 507 F_2 plants displaying the resistant phenotype were collected. Some of these F_2 plants were reconfirmed for resistance in the F_3 generation and were used for mapping. A total of 31 resistant F_2 plants and pooled F_3 families were used to initially map *edr5-1* to the long arm of chromosome 4. This initial map position was obtained using a total of 10 CAPS and simple sequence length polymorphism (SSLP) markers found in the TAIR (http://www.Arabidopsis.org/aboutcaps.html) and Arabidopsis Genome Center (http://genome.bio.upenn.edu/SSLP_info/SSLP.html) web sites, respectively.

Once the mutation was positioned on chromosome 4, a group of SNAP (SGCSNP24, SGCSNP64, and SCGSNP102; Table I), CAPS (g4539, AG, and RPS2), and SSLP (nga 1111) markers on the long arm of chromosome 4 were used to determine a more defined map position for *edr5-1*. The analysis of the data obtained from those markers allowed us to narrow down the position of *edr5-1* to a 1.7-Mb region (which corresponds to 19 BAC clones) between markers SGCSNP24 and SGCSNP64 (Fig. 5A). For 18 of these BAC clones, two SNPs per BAC clone were selected from the Cereon SNP database (http://www.Arabidopsis.org/cereon) for conversion into SNAP markers. Primer pairs designed by the SNAPER program were tested for specificity under the conditions described previously (28 and 38 PCR cycles). Four primer pairs were tested per allele, representing an average of eight primer pairs tested per SNP site. Because primers/markers were tested in groups, the results obtained from the analysis of the groups tested initially allowed us to reduce the number of SNPs to be used in subsequent rounds, since more precise map positions were obtained after each round of markers was analyzed. A total of 144 primers (which corresponded to 18 SNPs) were tested, and 12 new SNAP markers lo-

Table I. *Compilation of 50 SNAP markers*

Marker Name	Map Position		Columbia-Specific Primer Pair (5′-forward-3′ and 5′-reverse-3′)	Landsberg-Specific Primer Pair (5′-forward-3′ and 5′-reverse-3′)
	Genetic[a]	Physical[b]		
SGCSNP131[c]	1/0.00	F22L4	5′-gactctttggtcaagtcaatcacatgctctatg-3′ 5′-attaatttttccagcaatgcgtgc-3′	5′-tccggcgtgaactgcaagtaaaatctt-3′ 5′-tagtctgcacatgtataaacgtacgtcctccac-3′
SGCSNP5[c]	1/2.61	F22L4	5′-actgtgaagtcgtgcactgtaggcatctc-3′ 5′-attttgcaaccaaaaagtacaatcggcc-3′	5′-ctgtgaagtcgtgcactgtaggcatctg-3′ 5′-gggattaatggaccgaccatgtcc-3′
SGCSNP151[c]	1/3.35	F19P19	5′-gaaccaatggtggcgaaattaccaaac-3′ 5′-ctacagctttaccaacacgcaatgcctt-3′	5′-gaaccaatggtggcgaaattaccactt-3′ 5′-tccggaatgtgaagctccggc-3′
SGCSNP170[c]	1/7.66	T25N20	5′-acataacgaatacaatagagagaatgcaagagatgt-3′ 5′-tcgtaggttttaccagtggtccacgat-3′	5′-acataacgaatacaatagagagaatgcaagagaatc-3′ 5′-cagtggtccacgatagactgaatattgcg-3′
SGCSNP107[c]	1/11.86	F14J9	5′-catcacgcatcagcatatcgccc-3′ 5′-taccttgtgtggctccacccttagagc-3′	5′-ggagcatcacgcatcagcatatcacat-3′ 5′-ggctccacccttagagcaccgc-3′
SGCSNP138[c]	1/14.69	F21M12	5′-cgcaccttcttcagccttcatttatatagcaa-3′ 5′-tctttgaacggatcctccctgatcg-3′	5′-attcttttcactatataaatagtctcatccgtgcg-3′ 5′-tgcctctgctactggggcttcctt-3′
SGCSNP109[c]	1/61.13	F13A11	5′-ttatgggagtcaatgagcaagactatggtgac-3′ 5′-aaaaagactggaatccttacttggtctccg-3′	5′-aactttaaacggacctaataatgttagcgacatta-3′ 5′-cagagaccctcttaagcttccatgggc-3′
SGCSNP17	1/61.21	F1I21	5′-ttgtcacaagtaagagttggccta-3′ 5′-ttgtatcattggggtctaattggg-3′	5′-ttgtcacaagtaagagttggtctc-3′ 5′-ttgtatcattggggtctaattggg-3′
SGCSNP143[c]	1/70.73	ND[d]	5′-ccccaagcaaagcaatacttgccattc-3′ 5′-gtggcggttggatacggaactgag-3′	5′-ccccaagcaaagcaatacttgccatct-3′ 5′-actagaccacggagttgtggcggtt-3′
SGCSNP69[c]	1/81.84	F15I1	5′-tacctattcttcgacggacgtaac-3′ 5′-gttgaacggtggctgacgttgggg-3′	5′-tacctattcttcgacggacgtgat-3′ 5′-gttgaacggtggctgacgttgggg-3′
SGCSNP186	1/115.32	T9N14	5′-agtagtcctactgaggctattacg-3′ 5′-ttcaccaagcaacaacgtctgttg-3′	5′-agtagtcctactgaggctattgcc-3′ 5′-ttcaccaagcaacaacgtctgttg-3′
SGCSNP95[c]	1/129.49	YUP8H12R	5′-gcaaggtttgtagtaactgaaaacggcgtt-3′ 5′-attccggcgagagttaccgcc-3′	5′-gggcaaggtttgtagtaactgaaaacggtctc-3′ 5′-gtacacaagaagagcacctccggca-3′
SGCSNP184[c]	2/18.52	T6P5	5′-gtcactatctttccaagagtcatcaatgaaacaaag-3′ 5′-atcacaaataccaacggtcccagg-3′	5′-gtcactatctttccaagagtcatcaatgaaacatta-3′ 5′-gaatgcccaccagcactctaaacttgc-3′
SNAP80	2/37.14	F5H14	5′-caagctcctcctgcggtttcactc-3′ 5′-ttggctctggctttactccaccgt-3′	5′-caagctcctcctgcggtttcctta-3′ 5′-ttggctctggctttactccaccgt-3′
SGCSNP60	2/44.72	T28I24	5′-atgatcgctttatcctcctctaca-3′ 5′-ttcactacaaatccgaatggcacc-3′	5′-atgatcgctttatcctcctctctg-3′ 5′-ttcactacaaatccgaatggcacc-3′
SNAP108	2/47.97	F17H15	5′-cattctcttccgtcaaccttaaga-3′ 5′-tccaatcatgagatcgacctcgag-3′	5′-cattctcttccgtcaaccttaatg-3′ 5′-tccaatcatgagatcgacctcgag-3′
SGCSNP56	2/64.12	T28P16	5′-tattgagtttttggaaacatggct-3′ 5′-ggggtaggagcagcattggaagtg-3′	5′-tattgagtttttggaaacatgagc-3′ 5′-ggggtaggagcagcattggaagtg-3′
SGCSNP48	2/64.12	F25I18	5′-acatgcagagcacggataattcatccc-3′ 5′-gcgttcttcttgagcatctccgacc-3′	5′-tttttttgttgttgtttgatttttgttcgcttt-3′ 5′-acagaatagaaaagatagcgacagaccctca-3′
SGCSNP106[c]	2/72.11	T16B24	5′-tcaatcattaaatgaccgcttcaattttacgtc-3′ 5′-tgaggagactggtaatgccgtgtgaac-3′	5′-atatcaatcattaaatgaccgcttcaattttacgtg-3′ 5′-ggcatgaggagactggtaatgccg-3′
SGCSNP37[c]	2/72.45	F12L6	5′-tattgaatcggcctaagcaagcttgttctc-3′ 5′-cagtaaagaaccagaagcagtttgaccaatgtaac-3′	5′-aaggtgttaatagttgcacagagatatcaatagcgc-3′ 5′-acattggcgattagaatatcaaacattgcaaa-3′
SGCSNP9[c]	2/78.06	T14P1	5′-cattcaccttttagtcattatattgtcaagctgagg-3′ 5′-tggatcctaaaatagaatcggatttggtgatt-3′	5′-cattcaccttttagtcattatattgtcaagctaaac-3′ 5′-atgcatttgcaggtgaagctttctgg-3′
SGCSNP115[c]	3/3.32	T1B9	5′-tgcatgatcaccatggtccattgc-3′ 5′-gattactctatgatcttttcgagg-3′	5′-tgcatgatcaccatggtccaagga-3′ 5′-gattactctatgatcttttcgagg-3′
SGCSNP54	3/12.01	T16O11	5′-ttggggtccaccgcctcggtacta-3′ 5′-tagtccaattcaagccggttcacag-3′	5′-ttggggtccaccgcctcggtactg-3′ 5′-tagtccaattcaagccggttcacag-3′
SGCSNP11[c]	3/14.78	F9F8	5′-cgtctctcaatttctatctctgtctctggatcg-3′ 5′-aacaataggtatagggtgtgaaacattgaaacg-3′	5′-atcgtctctcaatttctatctctgtctctggttga-3′ 5′-gaaacaataggtatagggtgtgaaacattgaaacgt-3′
SGCSNP224[c]	3/43.65	MLJ15	5′-gtatattttctctttgttgctcctatcttctgcaag-3′ 5′-agaagacgactgcccaaaactgg-3′	5′-gtatattttctctttgttgctcctatcttctgcaaa-3′ 5′-gacgactgcccaaaactggccat-3′
SGCSNP225[c]	3/43.65	MLJ15	5′-caaagacccgatcttttagctacttctagtgattac-3′ 5′-gtcgaatgctccttatcacggagatcaa-3′	5′-catcatcagagattcgccataaacggaga-3′ 5′-ggtttatgctacattgaagcaaagatggttacg-3′
SGCSNP237[c]	3/47.16	K16N12	5′-gtaagtgtatacagtgcataaacaaactggttaagg-3′ 5′-ttatgatgggttttcgtgacacgtgct-3′	5′-gtaagtgtatacagtgcataaacaaactggttaaga-3′ 5′-ttatgatgggttttcgtgacacgtgct-3′
SNAP77	3/51.95	MOD1	5′-cagtttccgaagtaagtccccgtg-3′ 5′-cgatgtcgtgttagccacgctagt-3′	5′-cagtttccgaagtaagtccccgta-3′ 5′-cgatgtcgtgttagccacgctagt-3′
SNAP100	3/65.78	T24C20	5′-cgaaatccatcggttgattctccg-3′ 5′-catgtttaaacccttagagaacac-3′	5′-cgaaatccatcggttgattcactt-3′ 5′-catgtttaaacccttagagaacac-3′

(*Table I continues on next page.*)

Table I. (*Continued from previous page.*)

Marker Name	Map Position		Columbia-Specific Primer Pair (5′-forward-3′ and 5′-reverse-3′)	Landsberg-Specific Primer Pair (5′-forward-3′ and 5′-reverse-3′)
	Genetic[a]	Physical[b]		
SGCSNP41[c]	4/5.15	T18A10	5′-ttcattttcgaatatctgctcttttgtgtttgact-3′	5′-cattttcgaatatctgctcttttgtgtttgaca-3′
			5′-gtgtatgaaccgggtttagaatcacgtttacc-3′	5′-gtgtatgaaccgggtttagaatcacgtttacc-3′
SGCSNP24[c]	4/30.88	F24G24	5′-agacttacgattttgcacatgagagcacaag-3′	5′-caactccttttgcttgcttcgatttgatg-3′
			5′-gggaccagtagtatcgaacgacccaga-3′	5′-aattgaagccacagaagtaaacacagcatga-3′
SGCSNP64	4/50.90	FCA0	5′-gaagtcagatggtgttcctgcaacattttt-3′	5′-gcggcgtggtcttgagtggtg-3′
			5′-ttaacaccagtggtagtttcctcagagacacc-3′	5′-cataccgctcgtgcagcgtgtt-3′
SGCSNP102[c]	4/55.57	FCA5	5′-tcgctagcctacaccttggtatttacggtttaa-3′	5′-gctagcctacaccttggtatttacggtccac-3′
			5′-aaagaggaaagagcggccaaggag-3′	5′-cgggattacgcaatatttaggaagagagga-3′
SNAP58	4/66.16	F9F13	5′-ggtacggtggctacagggcgtctc-3′	5′-ggtacggtggctacagggcgtgat-3′
			5′-tgaataccccctaagcaacaaccct-3′	5′-tgaataccccctaagcaacaaccct-3′
SGCSNP149[c]	4/75.69	M7J2	5′-agcggcgaacatttaagaccaacctg-3′	5′-gcaagcggcgaacatttaagaccaaatta-3′
			5′-ttagaccccaatctcttaccatatcactgactttg-3′	5′-caatctcttaccatatcactgactttgcttccc-3′
SGCSNP152[c]	4/76.81	T13J8	5′-gagataggttcacaattgcggcag-3′	5′-gagataggttcacaattgcgactc-3′
			5′-tcctccgatcatccgtccccgtct-3′	5′-tcctccgatcatccgtccccgtct-3′
SGCSNP79[c]	4/110.17	T19P19	5′-gcctacttcaattgtgagctacaatcacttgtga-3′	5′-cctacttcaattgtgagctacaatcacttgagc-3′
			5′-ctgtgaactgacgaacctctccctcg-3′	5′-gaacctctccctcggatttgaactgg-3′
SGCSNP53[c]	4/110.40	T5J17	5′-aagcaacttgctactaaggttttgttccttcatg-3′	5′-aagcaacttgctactaaggttttgttccttcctt-3′
			5′-ctaagacagcatggctctggaaacg-3′	5′-tcttgagcaatctcacggactagcct-3′
SGCSNP4[c]	5/18.86	MJJ3	5′-atgccgtcactgggaggcaagaa-3′	5′-atgccgtcactgggaggcatgag-3′
			5′-gcttaccaggagctcttgtcaccgc-3′	5′-aatttaattgcagatcggcctttgac-3′
SGCSNP123[c]	5/18.87	MJJ3	5′-gagaacttggtggccaagttaaat-3′	5′-gagaacttggtggccaagttactc-3′
			5′-aaggggtctgcaaatgttcgccgc-3′	5′-aaggggtctgcaaatgttcgccgc-3′
SGCSNP126[c]	5/90.13	MJB21	5′-tcctcgagactatacttgaagttaaaccccacg-3′	5′-cattgaattcagttccaagaagctgacgaaa-3′
			5′-cctatcactcgtacgtgtgtgtgcttcttc-3′	5′-gcaaagttcggtcgattcaaagacagc-3′
SGCSNP97[c]	5/93.07	K9D7	5′-cttcctcaagagctcagataaagcccgt-3′	5′-cagcttcctcaagagctcagataaagcacag-3′
			5′-ctggtgctgtacttatgccatataggtgagcta-3′	5′-ttcttagaaatgatgccgggcttgc-3′
SNAP54	5/93.75	MBD2	5′-gcaagtttttctctgataacgagt-3′	5′-gcaagtttttctctgataacaaag-3′
			5′-agattcagcctagagttgtttcc-3′	5′-agattcagcctagagttgtttcc-3′
SNAP18	5/93.75	MMG4	5′-tagtcctttccgttctcggttcga-3′	5′-tagtcctttccgttctcggttctg-3′
			5′-aaatggtggtattgataccgaagc-3′	5′-aaatggtggtattgataccgaagc-3′
SGCSNP44	5/113.80	MTI20	5′-ggaactggcatcaatcaggcctccta-3′	5′-cacacgtaataagcccttgcattaaggca-3′
			5′-gcatctggaacgggttaggccct-3′	5′-caagctggccctgggcttacg-3′
SGCSNP124	5/119.05	MAF19	5′-aggaagaaggcgtgacagagaacg-3′	5′-aggaagaaggcgtgacagagaact-3′
			5′-ggttttgatagcagcaagctgggt-3′	5′-ggttttgatagcagcaagctgggt-3′
SGCSNP29[c]	5/121.87	MGI19	5′-ccagcaccaaccatcaaagtcttttcaa-3′	5′-gctactcagatgggatactatgttaaggattcagca-3′
			5′-gccatgcctcagttgtttgccact-3′	5′-aatggattgctatattcaactgcaatcccaa-3′
SGCSNP19	5/123.85	MAC9	5′-aaaacttggttttgtaatatgttgcctgagaatcta-3′	5′-atccagaaaggacaatctaaaccaagagaagtctg-3′
			5′-cagactgaaccaatgtagatccttacttggtgac-3′	5′-ggaagctcatgcttccctacactgctc-3′
SGCSNP84[c]	5/140.75	MHJ24	5′-gatggttggatctgcggtttgtacg-3′	5′-tgtggtgtattcaggaacaggctattcaatactaca-3′
			5′-tttagcattagccgctcgtgacacagt-3′	5′-tacggatttccccattagctccggt-3′
SGCSNP101[c]	5/142.02	MUA2	5′-gacactctctgatttctcggtctccgattatc-3′	5′-aacgacagagagacataagtaagaggaaagtacaaa-3′
			5′-tagcaaaagcagggcgttttcgc-3′	5′-tgtttctcttagtgggccctagcagac-3′

[a] Chromosome/genetic map position in centiMorgans. [b] Clone designation. [c] Markers mapped using the Affymetrix-based mapping technique (Cho et al., 1999). [d] ND, Not determined.

cated in the region of interest were generated. Markers CER426330, CER426890, CER442145, CER444061, CER444203, CER446565, CER447954, CER447956, CER447203, CER465981, CER466066, and CER466198 are described on the Ausubel laboratory web site (http://patho.mgh.harvard.edu/ausubelweb) and on TAIR (http://www.Arabidopsis.org/). A total of 266 F_2 plants and pooled F_3 families were used to analyze 10 of the 12 markers that were generated (using 35 PCR cycles). As of the preparation of this manuscript, *edr5-1* has been mapped to a 315-kb region on the long arm of chromosome 4 between SNAP markers CER447954 and SGCSNP64 (Fig. 5B).

DISCUSSION

The recent release of a list of approximately 25,000 SNPs (out of 40,000 predicted) between the Columbia and Landsberg *erecta* accessions in conjunction with the SNAP procedure described in this paper allow the design of a considerable number of molecular markers targeted specifically to regions of interest. Sufficient resolution for fine mapping of a mutation consequently can be achieved in a short period of time. In the SNAP procedure, SNP alleles are assayed using specially designed allele-specific primers, which generate allele-specific patterns (i.e. detectable

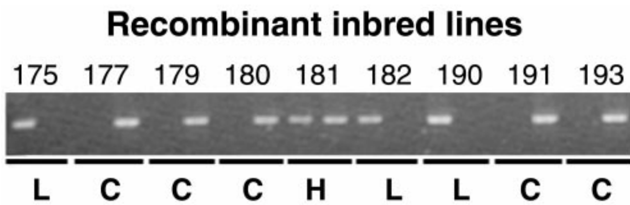

Figure 3. Mapping of SNAP markers with the Columbia × Landsberg RI lines. Allele-specific primers corresponding to SNAP markers were used to amplify 94 RI lines, and patterns of amplification were scored as Landsberg (L; e.g. line 175), Columbia (C; e.g. line 177), or heterozygous (H; e.g. line 181) for each of the RI lines. For each line, the left lane shows the amplification product obtained with the Landsberg-specific SNAP primer and the right lane shows the amplification product obtained with the Columbia-specific SNAP primer. A representative portion of an agarose gel is shown.

amplification product only from one allele) that are rapidly and reliably scored by simple analytical methods. We used a modification of the original allele-specific PCR methodology, the introduction of an additional mismatch within the last four bases of the primer (Newton et al., 1989), to increase primer specificity and allow reliable discrimination between alleles (Cha et al., 1992; Kwok et al., 1994). The use of that modification combined with an appropriate method to test for primer specificity (i.e. using two different sets of PCR amplification) allowed us to generate primers with sufficient selectivity for use as molecular markers in a typical map-based cloning project.

The design of SNAP primers is greatly facilitated by the use of the SNAPER program. The success rate of the SNAPER program, which uses empirical rules to design the primers, was approximately 53% out of a total of 331 primers that corresponded to 43 SNPs examined. Testing an average of eight primer pairs per allele we obtained specific primer pairs for both alleles of a given SNP in 27 out of 43 cases. Specific primer pairs for only one of the alleles were obtained in 14 cases, and two SNPs failed to generate any allele-specific primers. The failure to generate allele-specific primers in some cases is consistent with previous reports that indicated that mismatch extension can vary significantly (approximately 5–100-fold) depending on the sequences surrounding the mismatch (Mendelman et al., 1989; Huang et al., 1992). We speculate that the inability to obtain primers that showed the required degree of specificity for some SNP sites could be due to particular characteristics of the sequence surrounding the SNP.

In the 14 cases where specific primers were not obtained for one of the alleles after testing eight primer pairs, we tested more primers. We ultimately obtained the second allele-specific primer for six out of seven SNPs examined. However, because of the large number of primers tested, we concluded that the most efficient approach is to simply abandon the SNPs that fail to yield specific primers for both alleles after testing a limited number of primer pairs. Designing primers for new SNPs reduces the time and cost required to generate a marker, and the availability of SNPs is not a limitation. In the case of the mapping of *edr5-1*, we tested only four primer pairs per allele for each SNP and were able to obtain 12 markers out of 18 SNPs tested. These data suggest that it is only necessary to test a relatively small number of primers to obtain a molecular marker for a specific SNP (approximately four primer pairs per allele). Moreover, we predict that the success rate of the program will increase as the data generated testing the primers is added to the empirical database used to generate the rules underlying the SNAPER algorithm.

The 50 SNAP markers generated are evenly distributed throughout the Arabidopsis chromosomes, showing only two gaps at the bottom of chromosome

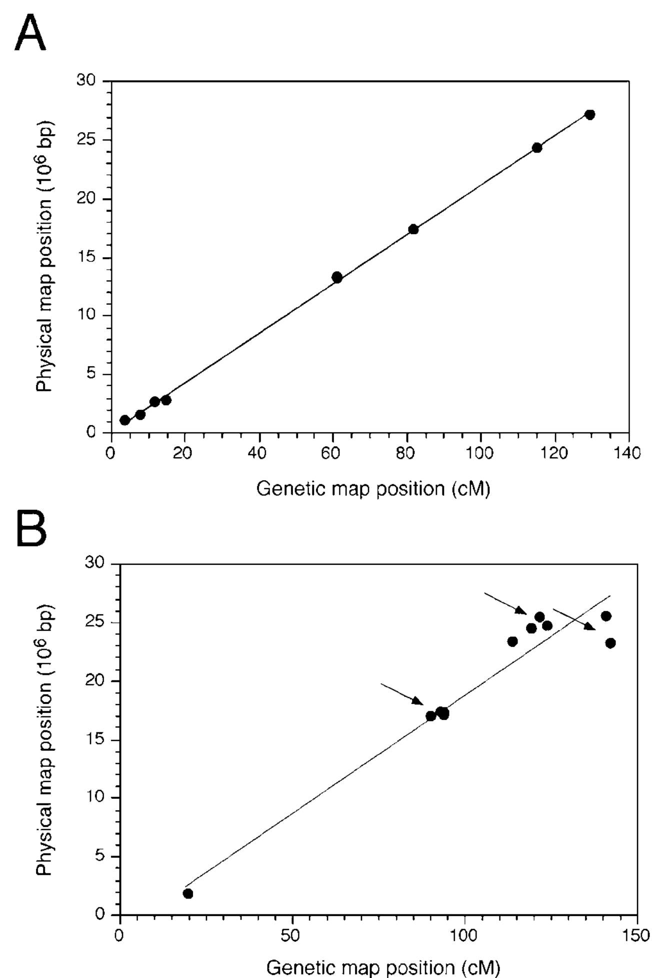

Figure 4. Correlation between SNAP positions on the genetic and physical maps for chromosomes 1 (A) and 5 (B). Physical map positions for the SNP sites were obtained by identifying the corresponding sequences in overlapping BAC, P1, and YAC clones. Genetic map positions were determined using the RI lines. The arrows indicate markers for which genetic map positions are reversed relative to the physical map.

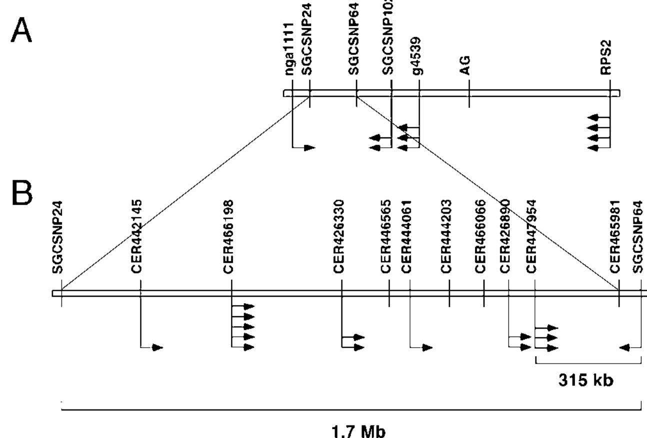

Figure 5. Mapping of *edr5-1*. A, A group of seven SNAP, CAPS, and SSLP markers were used to obtain an initial map position for the *edr5-1* mutation. B, Ten new SNAP markers were generated using the SNPs from the Cereon database and the SNAPER program to further delimit the position of the mutation. The arrows indicate the sites of recombination events that occurred in F_2 plants.

3 and the top of chromosome 5 (Table I). The maximum distance between any two of the 50 SNAP markers described in Table I is approximately 70 cM (SGCSNP123 and SGCSNP126) at the top of chromosome 5 (Table I). Because there is widespread interest in using RI lines to clone quantitative trait loci, we evaluated the correspondence between the physical map of the Columbia genome with the genetic map derived from the Columbia × Landsberg RI lines. It is interesting that the analysis showed a strong correlation between the genetic and physical maps for chromosomes 1, 2, 3, and 4 (r^2 = 0.999, 0.977, 0.986, and 0.990, respectively; Fig. 4A). On the other hand, a less rigorous comparison for chromosome 5 showed that the genetic and physical map positions differed considerably in several places (Fig. 4B). Although it is not possible to definitively attribute a physical map position to a marker until the sequence of the genome is complete (because of potential duplications), the data suggest that caution must be used in determining a physical map position based on genetic data.

In the fine-structure mapping of the *edr5-1* mutation, we created a total of 12 new SNAP markers, and using 10 of the 12 markers generated we were able to delimit the mutation to a 315-kb region on the long arm of chromosome 4 (Fig. 5B) in approximately a 3-week period (not counting the time it took for primers to be synthesized). The rapid mapping of the *edr5-1* mutation (which is still in progress) can be credited to a combination of factors: first, the current abundance of SNPs makes it extremely easy to target a region of interest and generate a large number of markers that can be used in fine-structure mapping; second, the SNAPER program facilitates design of primer alternatives for the creation of PCR-based markers using existing SNP sites; and third, the methods used to analyze the markers generated (PCR and agarose gel electrophoresis) are relatively simple.

Compared with some of the methods currently used for high-throughput SNP detection in genome centers, such as pyrosequencing (Ahmadian et al., 2000; Alderborn et al., 2000; Nordstrom et al., 2000), Taq-Man (Livak, 1999) and fluorescence resonance energy transfer (Chen et al., 1998), the SNAP methodology has the advantage that it does not require special equipment or sophisticated methodologies to detect SNPs. Since SNAP marker detection requires a simple gel-based assay, it is accessible to any molecular biology laboratory. Moreover, the generation and analysis of the mapping markers used by our method is relatively inexpensive. In any case, because the SNAP method is based on the generation or lack of generation of an amplified DNA sequence, it should be relatively straightforward to design high-throughput methods for detecting SNAP markers using simple DNA-DNA hybridization technologies.

MATERIALS AND METHODS

Primer Design

Allele-specific primers that corresponded to SNP sites were designed using an empirically determined algorithm called SNAPER (E. Drenkard, S. Rozen, M. Mindrinos, B.G. Richter, and F.M. Ausubel, unpublished data). The algorithm was implemented in the PERL language for the automatic design of oligonucleotide primers. Optimization of melting temperature, oligonucleotide length, and length of amplified products was achieved using the Primer3 program (S. Rozen and H.J. Skaletsky, code available at http://www-genome.wi.mit.edu/genome software/other/primer3.html). Primer sequences were screened against an Arabidopsis library of repetitive elements (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) to minimize mis-priming. Quantitation of primer concentrations was performed using the Quant program (developed by B.G. Richter).

Primer Testing

Testing of the primers designed by the SNAPER program was performed on a PTC-225 DNA Engine Tetrad (MJ Research, Watertown, MA) using 20-μL reactions in 384- or 96-well formats. Thirty nanograms of Columbia or Landsberg *erecta* genomic DNA, isolated using standard procedures (Ausubel et al., 1992), was used as template for PCR amplification. PCR reactions (20 μL) contained 5 pM of each primer, 1 unit AmpliTaq Gold DNA polymerase (Perkin Elmer, Branchburg, NJ), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) gelatin, and 200 pmol deoxyribonucleotides. Cycling conditions started with initial denaturation at 94°C for 5 min, followed by 94°C for 30 s, 64°C for 1 min, for 28 cycles or 38 cycles. A final extension reaction was performed at 72°C for 10 min.

Ramp speeds were set at 1.4°/s to 64° and 2°/s to 94° for intra-block consistency. PCR products were separated by gel electrophoresis on 1.3% (w/v) agarose.

RI Mapping

Template DNA from 94 recombinant inbred lines generated using the Columbia × Landsberg *erecta* cross (Arabidopsis Biological Resource Center, Ohio State University, Columbus) was isolated using standard methods (Ausubel et al., 1992). PCR reactions were performed essentially as described above using 5 ng of template DNA and 35 cycles. PCR products were separated by gel electrophoresis on 3% (w/v) agarose. Special equipment was used for the large-scale experiments including the rapid 96/384 well plate liquid pipette robot (Zymark Corp., Hopkinton, MA) to aliquot the DNA. Map positions were analyzed using the MAPMAKER software (Lander et al., 1987) at the Nottingham Arabidopsis Stock Centre (Nottingham, UK).

Mapping of *edr5-1*

Samples were prepared using a plant genomic miniprep (Edwards et al., 1991) or the DNeasy plant kit miniprep (Qiagen, Valencia, CA). Analysis of CAPS and SSLP markers was performed as described previously (Konieczny and Ausubel, 1993; Bell and Ecker, 1994). Primers used to generate the SNAP markers were tested as described above. PCR reactions for SNAP markers were performed essentially as described above using 2 μL of DNA sample and 60 ng of each primer. Cycling conditions were the same as for the mapping of the SNAP markers. PCR products were separated by gel electrophoresis on 1.5% (w/v) agarose.

ACKNOWLEDGMENTS

We thank Sean May and Keith Bradnam from the Nottingham Arabidopsis Stock Centre for the analysis of the RI mapping data; Julie Stone and Mary Wildermuth (Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital) for help in preparation of the manuscript; Sigrid Volko, Julie Stone, and Jacinto Villanueva (Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital) for help in mapping *edr5-1*; and Eric Lander and David Altshuler (Whitehead Institute/Massachusetts Institute of Technology Center for Genome Research, Cambridge, MA) for the use of specialized equipment in the large-scale mapping of the SNAP markers.

Received September 8, 2000; accepted September 22, 2000.

LITERATURE CITED

Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M, Lundeberg J (2000) Single-nucleotide polymorphism analysis by pyrosequencing. Anal Biochem **280:** 103–110

Alderborn A, Kristofferson A, Hammerling U (2000) Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Res **10:** 1249–1258

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1992) Short Protocols in Molecular Biology. Greene Publishing Associates and John Wiley & Sons, New York

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

Cha RS, Zarbl H, Keohavong P, Thilly WG (1992) Mismatch amplification mutation assay (MAMA): application to the c-H-*ras* gene. PCR Methods Appl **2:** 14–20

Chen X, Livak KJ, Kwok PY (1998) A homogeneous, ligase-mediated DNA diagnostic test. Genome Res **8:** 549–546

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

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res **19:** 1349

Huang M-M, Arnheim N, Goodman MF (1992) Extension of base mispairs by *Taq* DNA polymerase: implications for single nucleotide discrimination in PCR. Nucleic Acids Res **20:** 4567–4573

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

Kwok S, Chang S-Y, Sninsky JJ, Wang A (1994) A guide to the design and use of mismatched and degenerate primers. PCR Methods Appl **3:** S39–S47

Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic Acids Res **18:** 999–1005

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181

Li H, Cui X, Arnheim N (1990) Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction. Proc Natl Acad Sci USA **87:** 4580–4584

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

Livak KJ (1999) Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. Genet Anal **14:** 143–149

Lukowitz W, Gillmor CS, Scheible W-R (2000) Positional cloning in Arabidopsis: why it feels good to have a

genome initiative working for you. Plant Physiol **123:** 795–805

Mendelman LV, Boosalis MS, Petruska J, Goodman MF (1989) Nearest neighbor influences on DNA polymerase insertion fidelity. J Biol Chem **264:** 14415–14423

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

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

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acid Res **17:** 2503–2516

Nordstrom T, Ronaghi M, Forsberg L, de Faire U, Morgenstern R, Nyren P (2000) Direct analysis of single-nucleotide polymorphism on double-stranded DNA by pyrosequencing. Biotechnol Appl Biochem **31:** 107–112

Petruska J, Goodman MF, Boosalis MS, Sowers LC, Cheong C, Tinoco I (1988) Comparison between DNA melting thermodynamics and DNA polymerase fidelity. Proc Natl Acad Sci USA **85:** 6252–6256

Volko SM (1998) Isolation and characterization of *Arabidopsis* mutants with altered disease susceptibility. PhD thesis. University of Basel, Switzerland

Ugozzoli L, Wallace RB (1991) Allele-specific polymerase chain reaction. Methods Enzymol **2:** 42–48