A Physical Amplified Fragment-Length Polymorphism Map of Arabidopsis

Janny L. Peters[1]*, Hans Constandt, Pia Neyt, Gerda Cnops, Jan Zethof, Marc Zabeau, and Tom Gerats[1]

Flanders Interuniversity Institute for Biotechnology, University of Ghent, Department of Plant Genetics, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium

We have positioned amplified fragment-length polymorphism (AFLP) markers directly on the genome sequence of a complex organism, Arabidopsis, by combining gel-based AFLP analysis with in silico restriction fragment analysis using the published genome sequence. For placement of the markers, we used information on restriction fragment size, four selective nucleotides, and the rough genetic position of the markers as deduced from the analysis of a limited number of Columbia (Col)/Landsberg (L*er*) recombinant inbred lines. This approach allows for exact physical positioning of markers as opposed to the statistical localization resulting from traditional genetic mapping procedures. In addition, it is fast because no extensive segregation analysis is needed. In principle, the method can be applied to all organisms for which a complete or nearly complete genome sequence is available. We have located 1,267 AFLP Col/L*er* markers resulting from 256 *Sac*I+2, *Mse*I+2 primer combinations to a physical position on the Arabidopsis genome. The positioning was verified by sequence analysis of 70 markers and by segregation analysis of two leaf-form mutants. Approximately 50% of the mapped Col/L*er* AFLP markers can be used for segregation analysis in Col/C24, Col/Wassilewskija, or Col/Cape Verde Islands crosses. We present data on one such cross: the localization of a viviparous-like mutant segregating in a Col/C24 cross.

Forward genetics involves the position-dependent cloning of genes underlying phenotypes of interest. A prerequisite for successful positional, or map-based, cloning is the availability of an accurate high-density genetic map. Genetic maps are traditionally constructed by analyzing linkage relationships between markers in (large) segregating populations. The estimated genetic distances between the markers depend on the type and size of the population used to construct the map. In Arabidopsis, recombinant inbred lines (RILs), essentially representing homozygous genotypes, constitute permanent mapping populations. More than 1,200 markers (http://nasc.nott.ac.uk/new_ri_map.html), including 237 single nucleotide polymorphism (SNP) markers recently localized by Cho et al. (1999), have been analyzed in the Columbia (Col)/Landsberg (L*er*) RILs (Lister and Dean, 1993). Alonso-Blanco and colleagues (1998) developed an amplified fragment-length polymorphism (AFLP)-based linkage map containing 385 and 321 AFLP markers by using the Col/L*er* and L*er*/Cape Verde Islands (Cvi) RIL populations, respectively. Even though good mapping populations are available for Arabidopsis, marker locations in the resulting genetic maps can be given only relative values because map distance is expressed in terms of recombination frequency. Unfortunately, units of genetic recombination (centiMorgans; cM) cannot be translated into physical distances, i.e. distances in base pairs. Moreover, for each marker to be placed, at least 100 individuals must be scored, a time-consuming and error prone process. With the availability of the Arabidopsis genome sequence, ecotype Col-0 (The Arabidopsis Genome Initiative, 2000), an increasing number of genetic markers can be allocated to a physical map position by sequence analysis; about 800 Col markers have already been located physically (The Arabidopsis Information Resource Web site: http://www.Arabidopsis.org/).

Here we propose a new procedure for the construction of high-resolution maps that overcomes many of the shortcomings of traditional linkage analysis. This procedure combines in silico AFLP analysis of the available Arabidopsis genomic sequence with gel-based AFLP analysis on Arabidopsis ecotypes, and results in the physical placement of Col/L*er* AFLP markers on the genome sequence. By avoiding the use of an extensive mapping population, this method considerably accelerates the normally time-consuming procedure for building genetic maps. In addition, it benefits from the fact that it allows for precise physical placement, which implies that the order of markers that would be clustered in traditional genetic maps can now be deciphered. The availability of the Arabidopsis genome sequence allows us to perform in silico restriction digests, which provides us with a set of Col AFLP fragments; whereas experimental AFLP analysis identifies Col AFLP fragments that are polymorphic with L*er* (or any other Arabidopsis ecotype). We can correlate experimental and in silico AFLP analysis by using the restriction fragment sizes and selective nucleotides used in AFLP analysis as discriminative parameters.

[1] Present address: Dept. of Experimental Botany, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands.

* Corresponding author; e-mail japet@sci.kun.nl; fax 32–24–3652787.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.010504.

In addition to these two parameters, knowledge about the approximate position of the AFLP markers on the genome significantly increases the number of markers that can be positioned unequivocally. We therefore analyzed a number of Col/L*er* RILs (Lister and Dean, 1993) together with the ecotypes. The restriction fragment size, selective nucleotides, and approximate position on the genome as afforded by the limited RIL analysis provide the information used to place the experimentally obtained Col AFLP polymorphisms directly onto the published Arabidopsis sequence. This procedure can in principle be applied to all organisms for which a complete or nearly complete genome sequence is available. In this article we show that the analysis of all 256 possible *Sac*I+2, *Mse*I+2 AFLP primer combinations on Col, L*er*, and seven Col/L*er* RILs leads to the physical placement of 1,267 Col/L*er* AFLP markers on the Arabidopsis genome sequence, i.e. an average density of one marker per 100 kb. This means that the application of a limited number of AFLP reactions leads to the physical positioning of more markers than currently available via the TAIR Web site. Additional sets of AFLP markers can be obtained by using other restriction enzymes, implying that very high-density physical AFLP maps can be produced in this way.

Well-characterized mutants are often created in backgrounds other than Col or L*er*. To facilitate the process of positional cloning, we have included three more ecotypes in our analysis. In our map we display information on the Col/L*er* AFLP markers that are polymorphic with the ecotypes C24, Wassilewskija (Ws), and/or Cvi as well. This results in maps of 644, 607, and 663 Col/L*er* AFLP markers that are also polymorphic between Col/C24, Col/Ws, and Col/Cvi, respectively. We present data on the localization of a viviparous-like mutation in a Col/C24 segregating population.

RESULTS

Strategy for Positioning Experimentally Obtained AFLP Markers on the Arabidopsis Genome

The analysis of a non-complex genomic sequence, Arabidopsis bacterial artificial chromosome (BAC) clone F26P21 (data not shown), and work on *Escherichia coli* (Arnold et al., 1999) demonstrated that experimental AFLP fragments can be predicted in silico. A more complex organism like Arabidopsis, requires the use of selective nucleotides for a genome-wide experimental AFLP analysis. We performed *Sac*I+2, *Mse*I+2 primer combinations for our experimental AFLP analysis. Unlike the approach chosen for *E. coli* (Arnold et al., 1999), we did not aim to predict all AFLP fragments visualized on gels. Instead we identified the unambiguous Col/L*er* AFLP polymorphisms from the gels and aimed to define their corresponding in silico AFLP fragments, with the intention to produce a physical AFLP map. For in silico AFLP analysis, the available Arabidopsis genome sequence (Col-0) was downloaded and digested in silico with the restriction enzymes *Sac*I and *Mse*I. This resulted in 41,217 predicted *Sac*I/*Mse*I fragments. The criteria used to define the in silico fragment(s) corresponding to each Col/L*er* AFLP marker obtained by experimental AFLP analysis are described in detail in "Materials and Methods." In silico fragments that comply with the criteria are considered candidates for the experimental Col/L*er* AFLP markers. In a preliminary analysis, positioning a set of 955 experimentally obtained Col/L*er* AFLP markers on the genomic sequence based solely on the size of the markers and the selective nucleotides enabled us to assign only 246 (26%) of the Col/L*er* AFLP markers to a unique candidate position on the Arabidopsis genome. The great majority of the remaining AFLP markers had more than one possible corresponding in silico fragment. Thus, additional information was needed to increase the number of uniquely assigned markers. We therefore investigated the impact of (1) additional sequence information and (2) knowledge of the rough genetic position of the AFLP markers on the genome.

Extra sequence information can be obtained by performing four additional AFLP reactions on the Col parent with a fifth selective nucleotide (*Sac*I+2, *Mse*I+3). Upon this analysis we could assign 554 of 955 Col/L*er* AFLP markers to a unique position. Although a significant improvement, we considered 58% placement still insufficient, particularly because the AFLP analysis becomes more complicated and labor intensive due to usage of the four *Mse*I+3 primers.

As an alternative strategy, we determined the rough genetic position of the markers by including a limited number of Col/L*er* RILs (Lister and Dean, 1993) in the experimental AFLP analysis. With the inclusion of these lines, AFLP markers could theoretically be assigned to specified genetic segments. For segment definition, 370 genetically placed markers with full scores for the seven RILs used were selected from the recombinant inbred (RI) map, as published on February 4, 2000 (http://nasc.nott.ac.uk/new_ri_map.html). This resulted in the definition of 42 genetic segments covering the Arabidopsis genome, 36 of which have a unique genetic segment code for the seven RILs used (Fig. 1; Table I). Table I presents the number of Nottingham RI map markers that define each genetic segment. Based on the scores for the seven RILs included in the experimental AFLP analysis, Col/L*er* AFLP polymorphisms can now be classified to one of the 36 genetic segment codes. The resulting limitation in possible genomic location(s) of each AFLP marker led to the unambiguous assignment of a high number of the markers: addition of seven RILs to the AFLP analysis led to the placement of 712 (75%) of the 955 AFLP markers.

For placement of the experimentally obtained Col/L*er* AFLP markers on the Arabidopsis genome sequence we thus use the information on fragment size,

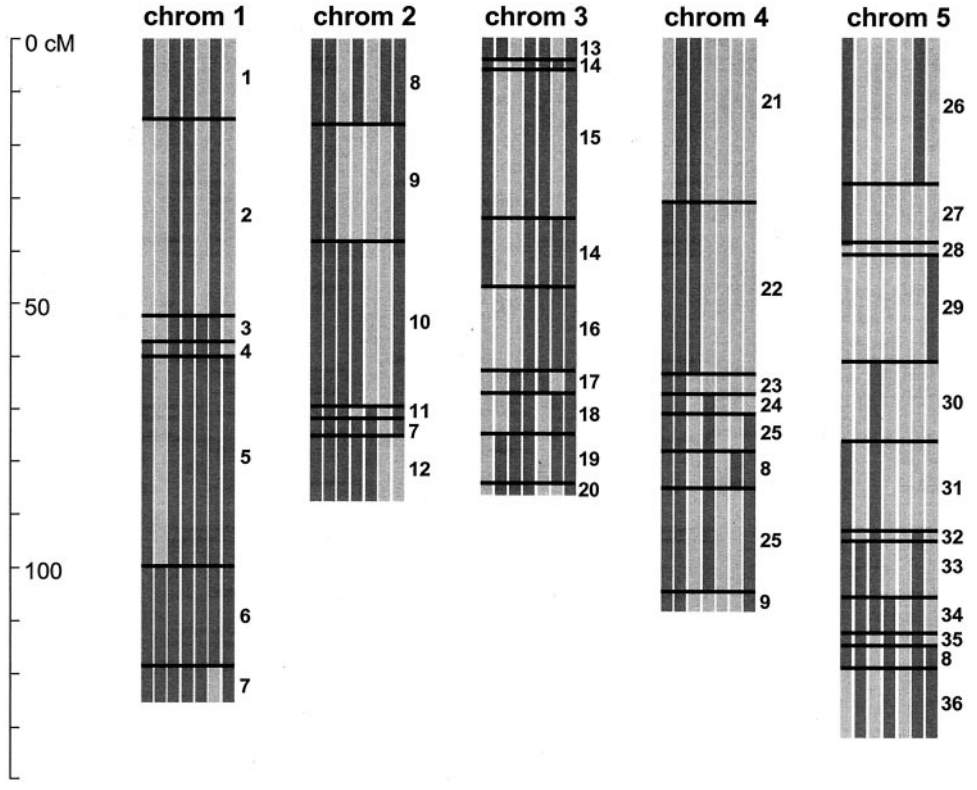

Figure 1. Genetic segments of the Arabidopsis genome as defined by the analysis of seven Col/L*er* RILs (nos. 5, 17, 62, 79, 240, 302, and 390, respectively). The recombination breakpoints were estimated based on the constitution of 370 selected RIL markers from the Nottingham RI map (http://nasc.nott.ac.uk/new_ri_map.html). Forty-two segments could be defined, 36 of which have a unique code that represents the parental origin of seven RILs. Dark gray represents Col, whereas light gray represents L*er*.

four selective nucleotides, and the rough genetic position of the markers deduced from the analysis of seven RILs.

Positioning the *Sac*I+2, *Mse*I+2 AFLP Polymorphisms on the Arabidopsis Genome

Performing AFLP analysis with the 256 possible *Sac*I+2, *Mse*I+2 primer combinations resulted in the definition of 1,623 Col/L*er* AFLP markers (Table I). As described above, the scores for the seven RILs included in the experimental AFLP analysis enable us to classify each Col/L*er* AFLP marker to one of the 36 defined genetic segment codes (Fig. 1). This consequently limits the possible position(s) of each AFLP marker on the Arabidopsis genome. Out of a total number of 1,623 Col/L*er* polymorphisms, a vast majority of 1,601 (98.6%) could be assigned to 34 of the 36 defined genetic segment codes.

The remaining 22 AFLP markers (1.4%) identified four genetic segment codes not exhibited by RI markers (Table I). Therefore, these four groups of AFLP markers could not be assigned a genetic map position. Their location might be identified, however, by positioning them on the total genomic sequence. One group containing seven AFLP markers can be allocated to the top of chromosome 2, whereas a second group of seven markers belongs to chromosome 5 (Table I). For the remaining two groups of four markers each, no unambiguous position could be identified.

No AFLP markers were assigned to segments 11 and 13 (Table I). Both segments are very small and comprise two and one marker(s), respectively. It is not unlikely that the segments appear due to mis-scoring. Alternatively, they do not harbor any AFLP markers. For example, segment 11 contains two RI markers at 71.24 and 71.39 cM of chromosome 2 (Table I), which score b (= L*er*) for RIL79. The RI map, as published February 4, 2000 (http://nasc.nott.ac.uk/new_ri_map.html) shows that the markers just above (at 71.12 cM) and just below (at 71.84 cM) score a (= Col) for RIL79 and therefore could be assigned a segment 7 (= aaaaaaba) score, which is the segment lying just below segment 11 (Fig. 1; Table I). Because no AFLP markers could be assigned to this segment, mis-scoring in the RI map is a likely explanation for the appearance of segment 11. Alternatively, segment 11 might indeed not harbor AFLP markers.

Positioning the *Sac*I+2, *Mse*I+2 AFLP Polymorphisms on the Arabidopsis Sequence

All experimentally identified Col/L*er* AFLP markers (Table I) were subsequently positioned on the available

Table I. *The number of Nottingham RI map markers, total number of Col/Ler AFLP markers, and number and percentage of Col/Ler AFLP markers that can be positioned on the Arabidopsis genome unambiguously and ambiguously (in brackets) per genetic segment*

The definition of the genetic segments is explained in Figure 1. For the RI map markers only those that show complete scores for the seven Col/L*er* RILs used to define the genetic segments were included in this table. The Col/L*er* AFLP markers resulted from performing 256 *Sac*I+2, *Mse*I+2 primer combinations. In the genetic segment code, the parental origin of each RIL is represented by a, Col, and b, L*er*; numbers of the RILs used are 5, 17, 62, 79, 240, 302, and 390, respectively.

Genetic Position of RI Map Markers	Genetic Segment: Code	RI Map Markers Used for Segment Definition	Col/L*er* AFLP Markers Total	Col/L*er* AFLP Markers	
				Positioned	Positioned
cM					%
chrom1: 2.71–12.55	1: abaabab	23	33	29 (1)	87.9 (90.9)
chrom1: 18.02–51.89	2: bbaabab	33	59	51 (5)	86.4 (94.9)
chrom1: 52.95–56.69	3: bbaaaab	4	12	11	91.7
chrom1: 57.73–58.77	4: abaaaab	2	23	19 (1)	82.6 (87.0)
chrom1: 61.17–96.65	5: abaaaaa	33	154	121 (13)	78.6 (87.0)
chrom1: 102.17–117.86	6: aaaaaaa	15	30	24	80.0
chrom1: 119.25–125.44	7: aaaaaba[a]	5	19	9	89.5
chrom2: ?	aaaabaa	0	7	6	85.7
chrom2: 7.40–13.38	8: aababaa[b]	5	61	18 (3)	82.0 (88.5)
chrom2: 18.60–36.77	9: aabbbba[c]	17	191	139 (13)	74.3 (81.2)
chrom2: 39.32–67.97	10: aaaabba	17	48	38 (4)	79.2 (87.5)
chrom2: 71.24–71.39	11: aaababa	2	0		
chrom2: 71.84–74.36	7: aaaaaba[a]	4		8	
chrom2: 76.11–87.55	12: aaaaabb	8	8	8	100.0
chrom3: 4.20	13: aabaaba	1	0		
chrom3: 4.24–5.83	14: abbaaaa[d]	3	42	4	88.1 (92.9)
chrom3: 6.87–30.86	15: abbaaba	26	54	38 (8)	70.4 (85.2)
chrom3: 36.35–43.77	14: abbaaaa[d]	8		33 (2)	
chrom3: 50.55–61.38	16: bbbaaaa	7	206	147 (21)	71.4 (81.6)
chrom3: 64.57	17: bbaaaba	1	11	11	100.0
chrom3: 70.5473.94	18: bbaabaa	4	11	11	100.0
chrom3: 76.14–82.86	19: baaabaa	8	18	15 (1)	83.3 (88.9)
chrom3: 86.41	20: baaabba	1	3	3	100.0
chrom4: 1.48–29.64	21: baabbbb	21	141	106 (11)	75.2 (83.0)
chrom4: 31.73–60.92	22: aaabbbb	23	87	69 (3)	79.3 (82.8)
chrom4: 65.70–65.72	23: aabbbbb	2	16	11 (2)	68.8 (81.3)
chrom4: 69.37	24: aababbb	1	7	4	57.1
chrom4: 72.21–76.76	25: aababba[e]	6	29	13 (2)	86.2 (93.1)
chrom4: 80.08–83.41	8: aababaa[b]	8		20	
chrom4: 86.27–104.04	25: aababba[e]	4		12	
chrom4: 104.73–108.58	9: aabbbba[c]	4		3	
chrom5: 2.52–25.42	26: abbbbab	16	35	26	74.3
chrom5: 29.58–38.65	27: abbbbbb	12	39	32 (1)	82.1 (84.6)
chrom5: 39.60	28: bbbbbbb	1	7	1	14.3
chrom5: 42.22–57.06	29: bbbbbba	5	18	14 (1)	77.8 (83.3)
chrom5: ?	bbabbba	0	7	4 (1)	57.1 (71.4)
chrom5: 65.16–72.26	30: bbabbbb	6	89	70 (6)	78.7 (85.4)
chrom5: 79.89–93.07	31: ababbbb	13	54	48 (3)	88.9 (94.4)
chrom5: 93.68	32: ababbab	1	1	1	100.0
chrom5: 96.87–104.63	33: aaabbab	5	50	40 (4)	80.0 (88.0)
chrom5: 106.55–111.73	34: aaaabab	4	20	18	90.0
chrom5: 113.72–114.25	35: aababab	5	4	3	75.0
chrom5: 115.91	8: aababaa[b]	1		12 (1)	
chrom5: 123.23–132.33	36: bababaa	5	21	17	81.0
?	aaabbba	0	4	0	
?	bbaaaaa	0	4	0	
Total		370	1623	1267 (107)	78.1% (84.7%)

[a–e] Represent the genetic segment codes present in the genome more than once.

Col-0 sequence (The Arabidopsis Genome Initiative, 2000). Software was developed in-house to perform in silico AFLP analysis. With the help of Microsoft Access (Microsoft, Redmond, WA), in silico AFLP fragments fitting the characteristics of each AFLP marker were putatively identified on the BACs cited in the sequence tables from TAIR (http://www.Arabidopsis.org/). The criteria used to identify the in silico fragments corresponding to each Col/L*er* AFLP marker are described in detail in "Materials and Methods." The successfully positioned AFLP markers have been deposited at the TAIR Web site and can be viewed at http://www.Arabidopsis.org/search/marker_search.html. In addition to the *Sac*I/*Mse*I AFLP markers, we placed 140 of the *Eco*RI/*Mse*I Col/L*er* AFLP markers (Kuiper, 1998; Alonso-Blanco et al., 1998) on the physical map; they are also available from the TAIR Web site. A compact disk containing the gel images of all 256 *Sac*I+2, *Mse*I+2 will be made available upon request. Table I demonstrates that 1,267 (78%) of the 1,623 AFLP markers can be assigned a unique position on the Arabidopsis genome. Only 1,644 kb of the telomeric regions are not covered by these markers. Thus 98.6% of the sequenced Arabidopsis genome (115,400 kb) is covered with AFLP markers. A considerable number of additional markers (107) is assigned to more than one position either within one genetic segment or even within a single BAC clone. Although these markers are not included in the tally of the uniquely positioned markers, the correct position of most of them can be solved over time. For 80 markers we find two, and for three markers three possible locations within the same genetic segment, whereas 24 markers have two or more possible positions within a single BAC clone. The in silico sequences that correspond to the markers with more than one possible localization within a genetic segment are mostly (i.e. in about 90% of the cases) totally different from one another (data not shown). By sequencing these AFLP markers or by analyzing additional RILs, most of the markers that now have two or three possible positions within one genetic segment can be assigned to a unique position. For the markers with two or more possible locations within a single BAC clone, the possible positions are by definition genetically very close, whereas the size and signature of the corresponding in silico fragments are very similar or identical (data not shown). In 15 of 24 cases these repeats appear twice, whereas six markers are repeated up to nine times within a single BAC. For the remaining three markers the in silico sequences are completely different; thus their correct positions could also be solved by sequencing the AFLP markers.

In addition to Col and L*er*, three other ecotypes were included in our analysis. Of the 1,267 Col/L*er* AFLP markers that were positioned on the Arabidopsis genome, 1,121 had complete scores, i.e. no missing data, for all additional three ecotypes. Thus, when a mutant of interest is created in a background other than L*er* (or Col), still a large percentage of these AFLP markers can be used in mapping projects (Table II). For example, 644 (57.4%), 607 (54.1%), and 663 (59.1%) of the 1,121 Col/L*er* AFLP markers were polymorphic with C24, Ws, and Cvi, respectively. This implies that a physical map of over 600 Col/L*er* AFLP markers is available for mapping projects involving mutants in either of these three backgrounds crossed into a Col background. As an example of such a cross we present data on the analysis of 44 viviparous-like mutant individuals (Table III). The F2-mutant individuals were selected from a cross between the mutant in C24 and the wild type in Col background. A first analysis indicated that the mutation was located on chromosome 1. Four selected primer combinations were subsequently performed and resulted in the analysis of 15 Col/L*er* AFLP markers, nine of which could be used to localize the viviparous-like mutant. Despite the fact that not all Col/L*er* AFLP markers can be used to localize the mutant locus, the viviparous-like mutation was successfully positioned on chromosome 1 between the AFLP markers SM233_177.7 and SM18_84.5 (Table III), comprising an interval of 30 BACs or about 2,600 kb.

Verification of the Physical AFLP Map

To ascertain that our physical AFLP map is correct, two independent approaches were followed. First, we isolated and sequenced 70 randomly chosen AFLP markers from our physical map. The experimentally obtained sequences were compared with the sequences predicted by in silico analysis. In all cases, the experimentally obtained sequence matched the predicted sequence (data not shown).

Second, mutant alleles located on the two chromosomes for which the complete sequence was available at the time of the experiment, chromosomes 2 and 4 (Lin et al., 1999; Mayer et al., 1999), were used to analyze the degree of linkage for putatively placed AFLP markers. The two leaf-form mutants used, angusta4 (*ang4*) and rotunda2-1 (*ron2-1*; Berná et al., 1999), map to chromosome 2 and 4, respectively, which is in good agreement with the position as determined by Robles and Micol (2001). Thirty and 29 AFLP markers spread along chromosomes 2 and 4, were scored for two sets of 21 mutant individuals each

Table II. *Percentage of polymorphism between pairs of five Arabidopsis ecotypes*

One hundred percent polymorphism between Col and L*er* signifies the 1,121 (out of a total of 1,267) Col/L*er* AFLP markers positioned on the Arabidopsis genome (see Table VI). Other ecotypes included are C24, Ws, and Cvi.

	Col	L*er*	C24	Ws	Cvi
Col	●	100	57.4	54.1	59.1
L*er*		●	42.6	45.9	40.9
C24			●	30.1	33.3
Ws				●	36.0
Cvi					●

Table III. *Linkage of Col/Ler AFLP markers on chromosome 1 with the viviparous-like locus*

The 44 F2-mutant individuals used for the linkage analysis resulted from a cross of a viviparous-like mutant (C24) and wild type (Col). Not all Col/L*er* AFLP markers are polymorphic for Col/C24.

Marker Name	Genetic Segment[a]	Position on Chromosome 1	Col	C24
		Mb		
SM233_306.8	1	3.5	8	36
SM233_81.5	1	4.3	8	36
SM28_319.1[b]	2	5.9		
SM233_177.7[c]	2	7.1	4	40
SM218_200.9[b]	2	9.4		
SM18_237.3[b]	2	9.7		
SM18_84.5[c]	2	9.7	2	42
SM218_338.5[b]	2	9.9		
SM18_97.8	2	10.0	2	42
SM233_327.6	2	11.1	3	41
SM28_302.3	2	12.7	7	37
SM218_160.1[b]	4	15.2		
SM28_334.2	5	17.1	17	27
SM218_156.8	5	18.2	19	25
SM18_192.2[b]	5	26.3		

[a] The definition of the genetic segments is explained in Figure 1. [b] Col/L*er* AFLP markers that are not polymorphic for Col/C24. [c] The viviparous-like mutation is located between these markers.

(Tables IV and V). These mutants were selected from an F_2 population resulting from a cross between the L*er* mutant and Col wild type. AFLP markers adjacent to the mutant loci will be in linkage disequilibrium and completely linked markers will score L*er* in all 21 mutant individuals. The ANG4 locus shows total linkage at a chromosomal position between 18.3 and 19.3 Mb, whereas the RON2-1 locus shows total linkage between 12.6 and 15.2 Mb (Tables IV and V). The recombination frequency increases regularly on both sides of the defined interval for both loci, culminating in 50% recombinants, i.e. no linkage disequilibrium. Because AFLP analysis is not a codominant technique, the expected ratio for non-linkage is 3:1 Col:L*er*, which means roughly 16:5 in a population of 21 individuals. Independent segregation of markers and mutant loci is reached at a distance of roughly 12 Mb from the ANG4 and RON2-1 loci. Taken together, the sequence analysis and the mapping data show the robustness of the proposed physical map.

The Distribution of AFLP Markers over the Arabidopsis Genome

Performing all 256 *Sac*I+2, *Mse*I+2 primer combinations allowed for the physical localization of 1,267 Col/L*er* AFLP markers on the Arabidopsis genome. As Arabidopsis is estimated to have a 125 Mb genome (The Arabidopsis Genome Initiative, 2000), this represents approximately one marker per 100 kb on average. Figure 2 shows the distribution of the AFLP markers over the five Arabidopsis chromosomes. The table included in Figure 2 indicates that the average distance between markers differs per chromosome; it is highest in chromosome 1 and lowest in chromosome 4, the largest and smallest chromosome, respectively.

For application of the map in map-based cloning projects it is important that there are relatively few marker gaps in the map. The median and largest distance between two markers give an indication of the existing marker gaps (Fig. 2). More specifically, a majority of 69.4% of the markers exhibit a pair-wise distance of less than 100 kb, whereas 30.6% of the AFLP markers represent markers separated by more than 100 kb. Only 2.6% of the markers exhibit pair-wise distances of more than 500 kb.

There are two limitations to our approach. First, roughly 10 Mb of the 125-Mb Arabidopsis genome consists of unsequenced centromeric and rDNA repeat regions (The Arabidopsis Genome Initiative, 2000). Without sequence information we are not able to physically locate any markers in these regions. In Figure 2 these regions are indicated by the double

Table IV. *Linkage of 30 Col/Ler AFLP markers of chromosome 2 with the angusta4 locus*

Twenty-one F2-mutant individuals resulting from a cross of the mutant (L*er*) and wild type (Col) were tested for the AFLP markers described below.

Marker Name	Genetic Segment[a]	Position on Chromosome 2	Col	L*er*
		Mb		
SM143_142.1	10	3.2	16	5
SM143_204.1	10	3.7	16	5
SM20_322.4	10	3.8	16	5
SM13_281.2	10	4.3	15	6
SM13_388.5	10	4.7	15	6
SM61_475.7	10	5.8	14	7
SM54_166.4	10	6.5	13	8
SM244_146.8	10	6.8	13	8
SM89_323.1	10	7.5	13	8
SM89_484.2	10	9.1	13	8
SM239_73.7	11	12.7	12	9
SM34_231.6	11	13.4	11	10
SM13_353.5	11	14.3	9	12
SM116_200.5	11	14.4	9	12
SM240_448.2	11	14.9	5	16
SM239_188.2	11	15.2	4	17
SM20_93.3	7	15.8	3	18
SM168_330.8	7	16.0	3	18
SM232_211.1	7	16.6	1	20
SM143_183.2	7	16.7	1	20
SM239_289.7	7	17.6	1	20
SM89_120.4	7	17.8	1	20
SM26_495.4	12	18.3	1	20
SM54_397.3	12	18.3	0	21
SM33_202.4	12	18.5	0	21
SM120_153.8	12	18.6	0	21
SM61_107.7	12	18.9	0	21
SM117_193.1	12	18.9	0	21
SM190_142.6	12	19.3	1	20
SM244_283.8	12	19.5	2	19

[a] The definition of the genetic segments is explained in Figure 1.

horizontal lines that interrupt the chromosomes. Second, the file that was used to determine the position of the markers on the continuous sequence AGI. nonredundant (see "Materials and Methods") is, in fact, redundant because it still contains BAC overlaps. Therefore the distances between some of the markers placed on the continuous sequence undoubtedly appear greater than they are in reality.

DISCUSSION

The completion of the genome sequence of Arabidopsis and its public accessibility have a profound impact on basic research in plant biology. Many new tools are being developed to exploit the availability of the genome sequence and eventually every gene, and its function is expected to be characterized in this model plant. As part of this effort, we utilized the Arabidopsis genome sequence to construct a high-resolution physical AFLP map that can facilitate the cloning of genes underlying interesting mutations. Our approach combines in silico AFLP analysis of the available Arabidopsis genomic sequence with experimental AFLP analysis on Arabidopsis ecotypes. This method appears to be not only very reliable but rapid as well, because the use of an extensive mapping population is avoided. Moreover, markers are given an exact physical position (in base pairs) as opposed to a genetic position (in centiMorgans). The fact that experimentally obtained AFLP fragments can be predicted in silico with very high confidence provides strong evidence for the robustness of the AFLP method itself. Earlier work by Arnold and colleagues (1999) demonstrated that experimental AFLP fragments can be predicted in silico in *E. coli*. Our results verify that the same can be done for a more complex organism. In principle, our method can be applied to all completely or nearly completely se-

Table V. *Linkage of 29 Col/Ler AFLP markers of chromosome 4 with the rotunda2-1 locus*

Twenty-one F2-mutant individuals resulting from a cross of the mutant (L*er*) and wild type (Col) were tested for the AFLP markers described below.

Marker Name	Genetic Segment[a]	Position on Chromosome 4	Col	L*er*
		Mb		
SM214_89.2	21	0.8	14	7
SM101_172.7	21	0.8	14	7
SM19_219.8	21	1.9	14	7
SM204_278.2	21	2.2	14	7
SM68_192.3	21	3.2	14	7
SM194_255.5	21	3.2	14	7
SM57_179.5	21	3.6	14	7
SM180_192.2	21	3.9	14	7
SM20_173.3	21	4.0	14	7
SM214_261.9	22	5.1	12	9
SM57_279.4	22	6.9	9	12
SM180_272.9	22	7.0	9	12
SM204_88.7	22	7.9	8	13
SM57_175.9	22	8.7	8	13
SM238_171.6	23	9.8	6	15
SM204_155.1	24	11.1	4	17
SM180_342.4	25	12.2	3	18
SM232_308.8	25	12.6	3	18
SM194_204.8	25	12.6	0	21
SM101_215.7	25	13.0	0	21
SM57_196.6	25	13.2	0	21
SM19_144.9	8	14.1	0	21
SM214_373.8	8	14.8	0	21
SM238_283.9	8	15.2	1	20
SM96_154.8	8	15.7	1	20
SM68_376.4	25	16.5	2	19
SM194_334.6	25	16.6	2	19
SM20_165.3	25	16.9	3	18
SM235_444.4	9	17.2	4	16

[a] The definition of the genetic segments is explained in Figure 1.

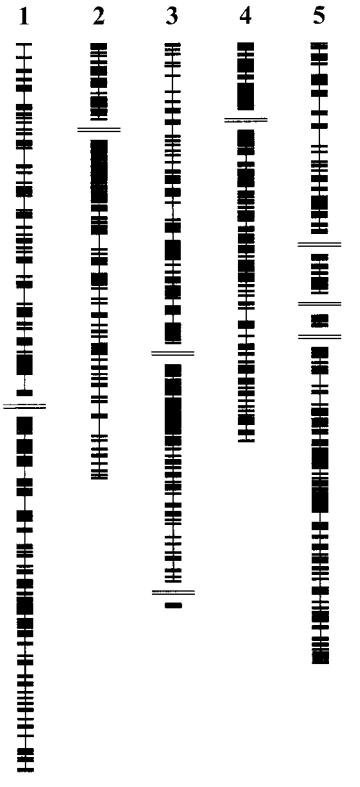

Figure 2. Distribution of Col/L*er* AFLP markers over the five chromosomes of Arabidopsis. For each chromosome, the total number of Col/L*er* AFLP markers (A), the average (B), the median (C), and the largest distance (in kb) (D) between two adjacent markers is given as well as the percentage of markers that have distances <100 kb (E), >100 kb (F), >200 kb (G), >300 kb (H), >400 kb (I), and >500 kb (J) between them. Each bold horizontal line represents a marker, whereas the eight double horizontal lines disrupting the chromosomes represent gaps in the sequence.

quenced organisms, including other complex organisms, as for example *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998), fruitfly (*Drosophila melanogaster*; Adams et al., 2000), and, in the near future, rice (*Oryza sativa*).

We identified a total of 1,623 Col/L*er* AFLP markers in the analysis of 256 *Sac*I+2, *Mse*I+2 primer combinations. To identify the matching counterpart in the sequence by in silico AFLP analysis, we analyzed the experimentally obtained information for (1) restriction fragment size, (2) selective nucleotides, and (3) rough genetic position of the markers as deduced from the concomitant analysis of seven Col/L*er* RILs. We thus could assign 1,267 Col/L*er* AFLP markers a unique position on the available Arabidopsis sequence (Table I). These 1,267 markers fail to cover only 1,644 kb at telomeric regions of the five chromosomes. In other words, 98.6% of the sequenced Arabidopsis genome is covered by the identified markers. Although the localized AFLP markers are not evenly distributed over the Arabidopsis genome, a majority of about 70% of the markers exhibit a pair-wise distance of less than 100 kb (Fig. 2). Only a small percentage (2.6%) of the markers exhibit pair-wise distances of more than 500 kb. Because there are relatively few large marker gaps, our AFLP map is useful for application in map-based cloning projects.

In Arabidopsis, clustering of *Eco*RI/*Mse*I AFLP markers around the centromeres was reported by Alonso-Blanco et al. (1998). Such clusterings also appear in other plant AFLP linkage maps as in potato (van Eck et al., 1995), barley (Powell et al., 1997), soybean (Keim et al., 1997), and maize (Vuylsteke et al., 1999). Because the above clustering has been observed in genetic maps, it is not possible to distinguish whether it results from a reduced recombination rate around the centromeres, a higher frequency of AFLP markers in these regions, or a combination of the two. Our work provides information on the precise physical location of the AFLP markers. Calculations reveal that genome-wide there are about 10 AFLP markers per megabase pair (Mb). However, the centromeric regions as defined by Copenhaver et al. (1999) contain twice as many AFLP markers per Mb. Dividing the genome in 1-Mb intervals shows that in the pericentromic region, the number of AFLP markers per Mb are well above average as well. These data are nicely visualized in Figure 2. The double horizontal lines in this figure represent gaps in the sequence and signify the centromeric region in chromosomes 1, 2, and 4. The upper and lower gaps signify the centromeric region in chromosomes 3 and 5, respectively. Chromosomes 2, 3, and 4 clearly show an increased density of markers around the centromeres. To a lesser extent, the same can be said for chromosomes 1 and 5. Although our data concern *Sac*I/*Mse*I AFLP markers as opposed to the *Eco*RI/*Mse*I AFLP markers of Alonso-Blanco et al. (1998), we may conclude that clustering of AFLP markers in the centromeric region is at least partly due to an increased frequency of AFLP markers in this region. Because the occurrence of in silico AFLP fragments is not increased in the (peri) centromeric region (data not shown), this suggests that the frequency of mutations is increased in this region, which may be explained by the fact that the rate of nucleotide substitutions is higher in non-coding than in coding sequences.

In addition to the 1,267 AFLP markers in unique positions, another 107 markers could be assigned more than one position on the genome (Table I). We do not yet have enough information to place these markers unequivocally on the Arabidopsis genome. Including markers with more than one possible position results in placement of 85% of the experimentally identified markers, approaching the estimated percentage of the genome sequenced (90%). An obvious prediction is that the majority of the non-placed markers will be located in the non-sequenced part of the genome. In addition, it is noteworthy that 48% of the non-placed markers are either large (>500 bp) or small (<100 bp), as compared with 11% for placed markers. Due to the usage of a 10-bp ladder between 100 and 500 bp, markers with sizes beyond this range are less accurately sized, and therefore the chance that they deviate too far from their in silico counterpart to be assigned the correct position on the physical map is higher.

Of the 107 AFLP markers that could be assigned more than one position on the genome, 83 could be assigned more than one possible position within a genetic segment, whereas 24 markers have two or more possible positions within a single BAC clone. In about 90% of the cases, the in silico sequences that correspond to the group of 83 markers are totally different from each other. Therefore, their correct position could, in principle, be elucidated by sequencing the markers or by analyzing additional RILs such that the genetic segments change in size and constitution. The in silico-obtained sequences of 21 of the 24 AFLP markers with more than one position within a single BAC clone are very similar, although usually not identical. Despite being present more than once, these Col fragments are still markers, i.e. they are polymorphic with L*er*. Presence of the repeats in both Col and L*er* would mean that the L*er* repeats must be mutated both (or all) to become polymorphic with Col. Therefore, the easiest explanation for the polymorphic repeats is that the formation of the polymorphisms preceded the duplication and that the duplication is unique to Col.

Two independent approaches verify the accuracy of our physical AFLP map: (1) sequence analysis of 70 AFLP markers and (2) mapping data for two L*er* mutants (Tables IV and V) and one C24 mutant (Table III). Besides offering proof for the validity of our physical AFLP map, these data also indicate that our map is likely to be a useful tool in map-based cloning projects. Until recently, map-based or positional cloning in Arabidopsis was a time-consuming procedure. The number of easy-to-use, inexpensive molecular markers was insufficient to perform the required fine-mapping.

However, with the public accessibility of the complete Arabidopsis sequence, the tools necessary for the production of a sufficient number of molecular markers are available. Cereon Genomics (Cambridge, MA) recently released a database with SNPs and insertions/deletion DNA polymorphisms between Col and L*er*. Among others, two recent publications discuss the usefulness of the information generated by the Arabidopsis Genome Initiative and Cereon for positional cloning (Drenkard et al., 2000; Lukowitz et al., 2000). Our approach has been to take advantage of the available sequence to produce a physical AFLP map with a coverage of around 98% of the sequenced genome. Almost 70% of the markers locate at pair-wise distances of 100 kb or less and only 2.6% of the markers exhibit pairwise distances of 500 kb or more. Because the Arabidopsis genome is estimated to be 125 Mb with a genetic length of approximately 600 cM, 1 cM on average equals 210 kb. Although actual recombination rates vary considerably throughout the genome, we conclude that the produced physical map can be used to efficiently perform the initial steps in map-based cloning. An additional advantage of our physical AFLP map is that, besides the preferred parents Col and L*er*, it can handle crosses involving C24, Ws, and Cvi, albeit with a lower resolution (see Tables II and III).

In practice, the physical map has provided AFLP markers that delineate an interval, harboring the locus of interest, of 2,600 kb at maximum and 1,000 kb at minimum (Tables III, IV, and V). In these experiments we performed AFLP analysis on very few mutant individuals, 21 for the *ang4* and *ron2-1* mutants and 44 for the viviparous-like mutant (Tables III, IV, and V), isolated from segregating F_2 populations. With this low number of individuals, we have a number of AFLP markers left to further delineate the region of interest for each of the three cases mentioned. Preliminary data indicate that F2-mutant individuals selected from a relatively small F_2 population of about 500 individuals will make optimal use of the produced AFLP map and delineate the locus of interest to a 400- to 700-kb region. After exhausting the available AFLP markers in the region of interest, it is recommended to switch to a source of easy-to-use, inexpensive polymorphisms for the selection of more recombinants in the region of interest. The most obvious choice for this purpose is the Cereon SNP collection.

In summary, the physical AFLP map can help to locate a mutant of interest in a very brief time span, by performing a limited number of AFLP reactions on mutants resulting from a relatively small F_2 population.

MATERIALS AND METHODS

Plant Material

Five Arabidopsis ecotypes, {Col-4 (Nottingham Arabidopsis Stock Center [NASC] stock no. N933), L*er*-0 (NASC stock no. NW20), C24 (NASC stock no. N906), Ws-2, (NASC stock no. N1601), and Cvi (Alonso-Blanco et al., 1998)} and seven Col/L*er* RILs (nos. 5, 17, 62, 79, 240, 302, and 390; Lister and Dean, 1993; http://nasc.nott.ac.uk/new_ri_map.html) were used for this study.

To verify our map we measured the degree of linkage of two leaf-form mutants with AFLP markers spread along the completely sequenced chromosomes 2 and 4. The angusta4 (*ang4*) and rotunda2-1 (*ron2-1*) mutants (Berná et al., 1999), represent mutations on chromosomes 2 and 4, respectively (Robles and Micol, 2001). For both mutants, the F_2 seeds resulting from a cross of the mutant (L*er*) and wild type (Col) were obtained from Dr. José L. Micol (Universidad Miguel Hernández, Alicante, Spain).

The viviparous-like data result from a cross of the mutant in C24 and the wild type in Col background. DNA samples from the F2-mutant individuals, used in the mapping experiment were kindly provided to us by Dr. Sergei Kushnir (University of Ghent, Ghent, Belgium).

Experimental AFLP Analysis

DNA was prepared from leaves using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). AFLP analysis was performed according to Vos et al. (1995) using the restriction enzymes *Sac*I and *Mse*I, the *Sac*I adapter:

CTCGTAGACTGCGTACAAGCT

CATCTGACGCATGT

and the *Mse*I adapter:

GACGATGAGTCCTGAG

TACTCAGGACTCAT

AFLP reactions were carried out using primers specific for the *Sac*I (GACTGCGTACAAGCTC) and *Mse*I (GATGAGTCCTGAGTAA) adapters, each containing two selective nucleotides. All possible 256 *Sac*I+2 and *Mse*I+2 primer combinations were applied (Table VI). Amplification reactions were performed in a PerkinElmer 9600 thermocycler (PerkinElmer, Norwalk, CT). *Sac*I+2 primers were radioactively phosphorylated using [γ-^{33}P]ATP (specific activity >92 TBq mmol^{-1}; Amersham, Buckinghamshire, UK) and the DNA fragments were separated in a Sequi-Gen GT sequencing cell (Bio-Rad, Hercules, CA). As a size marker, the SequaMark 10 base-ladder (Research Genetics, Huntsville, AL) was used. For each primer combination, the following samples were used and loaded in this order: Col, L*er*, C24, Ws, Cvi, RIL5, RIL17, RIL62, RIL79, RIL240, RIL302, and RIL390. After electrophoresis, gels were dried on a Heto dry GD-1 slab gel dryer (Heto Lab Equipment, Allerod, Denmark) and visualized using a PhosphorImager 445 SI (Molecular Dynamics, Sunnyvale, CA). Markers were scored by AFLP Quantar software (Keygene Products B.V., Wageningen, The Netherlands). AFLP markers were named with the code of the corresponding primer combination (Table VI) followed by the molecular size of the fragment as estimated by Quantar software.

Table VI. *Nomenclature of the primer combinations used in the AFLP analysis*

Selective Nucleotides *Sac*I	Selective Nucleotides *Mse*I															
	AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
AA	SM1	SM2	SM3	SM4	SM5	SM6	SM7	SM8	SM9	SM10	SM11	SM12	SM13	SM14	SM15	SM16
AC	SM17	SM18	SM19	SM20	SM21	SM22	SM23	SM24	SM25	SM26	SM27	SM28	SM29	SM30	SM31	SM32
AG	SM33	SM34	SM35	SM36	SM37	SM38	SM39	SM40	SM41	SM42	SM43	SM44	SM45	SM46	SM47	SM48
AT	SM49	SM50	SM51	SM52	SM53	SM54	SM55	SM56	SM57	SM58	SM59	SM60	SM61	SM62	SM63	SM64
CA	SM65	SM66	SM67	SM68	SM69	SM70	SM71	SM72	SM73	SM74	SM75	SM76	SM77	SM78	SM79	SM80
CC	SM81	SM82	SM83	SM84	SM85	SM86	SM87	SM88	SM89	SM90	SM91	SM92	SM93	SM94	SM95	SM96
CG	SM97	SM98	SM99	SM100	SM101	SM102	SM103	SM104	SM105	SM106	SM107	SM108	SM109	SM110	SM111	SM112
CT	SM113	SM114	SM115	SM116	SM117	SM118	SM119	SM120	SM121	SM122	SM123	SM124	SM125	SM126	SM127	SM128
GA	SM129	SM130	SM131	SM132	SM133	SM134	SM135	SM136	SM137	SM138	SM139	SM140	SM141	SM142	SM143	SM144
GC	SM145	SM146	SM147	SM148	SM149	SM150	SM151	SM152	SM153	SM154	SM155	SM156	SM157	SM158	SM159	SM160
GG	SM161	SM162	SM163	SM164	SM165	SM166	SM167	SM168	SM169	SM170	SM171	SM172	SM173	SM174	SM175	SM176
GT	SM177	SM178	SM179	SM180	SM181	SM182	SM183	SM184	SM185	SM186	SM187	SM188	SM189	SM190	SM191	SM192
TA	SM193	SM194	SM195	SM196	SM197	SM198	SM199	SM200	SM201	SM202	SM203	SM204	SM205	SM206	SM207	SM208
TC	SM209	SM210	SM211	SM212	SM213	SM214	SM215	SM216	SM217	SM218	SM219	SM220	SM221	SM222	SM223	SM224
TG	SM225	SM226	SM227	SM228	SM229	SM230	SM231	SM232	SM233	SM234	SM235	SM236	SM237	SM238	SM239	SM240
TT	SM241	SM242	SM243	SM244	SM245	SM246	SM247	SM248	SM249	SM250	SM251	SM252	SM253	SM254	SM255	SM256

In Silico AFLP Analysis

For in silico AFLP analysis, all BACs mentioned in the TAIR sequence table (http://www.Arabidopsis.org/cgi-bin/maps/Seqtable.pl) were retrieved from the EMBL database. With the help of in-house developed software, the sequences were digested in silico with the restriction enzymes *Sac*I and *Mse*I. All 41,217 resulting *Sac*I/*Mse*I fragments were stored in a Microsoft Access database. For each individual AFLP fragment, the database contains information about the size (in base pairs) and sequence. Furthermore, the chromosome number and the accession number of the BAC clone on which the fragment is located and its location within the BAC are included in the database.

A user-friendly version of the in-house developed software to restrict the Arabidopsis sequence is under development. The version we have used to obtain the data presented here can be made available upon request. For cDNA-AFLP a comparable program, GenEST, has been described by Qin et al. (2001).

Positioning of Col/L*er* AFLP Markers on the Arabidopsis Genome Sequence

A file with all experimentally obtained Col/L*er* AFLP polymorphisms was created. The Col/L*er* AFLP markers were named with the code of the corresponding primer combination (Table VI) followed by the molecular size of the fragment as estimated by AFLP Quantar software (Keygene Products B.V., Wageningen, The Netherlands). The code for the primer combination includes information about both the restriction enzymes and the selective nucleotides. To account for the length of the primers used in the AFLP reaction, 22 bp was subtracted from the experimentally obtained fragment sizes. Furthermore, the file contains the scores for the seven Col/L*er* RILs used in this study. With the help of Microsoft Access, these experimental AFLP data were used to identify the corresponding AFLP fragments from the in silico AFLP database. In silico predicted AFLP fragment(s) that correspond to each Col/L*er* AFLP marker were selected by taking into account a standard gel mobility deviation of 1%. To identify the smaller AFLP fragments, it is essential to take in account an additional ±2-bp deviation around the estimated size of each AFLP marker. In other words, in silico predicted AFLP fragments that fall within the range [(experimental AFLP fragment length in base pairs − 22) + (experimental AFLP fragment length in base pairs − 22) × 0.01 + 2] and [(experimental AFLP fragment length in base pairs − 22) − (experimental AFLP fragment length in base pairs − 22) × 0.01 − 2] are considered candidate(s) for the identified Col/L*er* AFLP markers. Col/L*er* AFLP markers that are tentatively positioned at a wrong genetic position can be detected by their score for the seven RILs.

Determining the Sequence of AFLP Fragments

To compare the experimentally obtained sequences with those predicted from the in silico database, 70 randomly chosen markers were cut from AFLP gels. The gel pieces, including Whatman paper (Whatman, Clifton, NJ), were put into 100 μL of water and were eluted for 1 h with vortexing every 10 min. Gel and paper pieces were spun down (5 min, maximal speed), and the watery phase was removed and used as template DNA to re-amplify the AFLP marker. For the re-amplification reaction, 2 μL of eluted DNA, 1 μL of *Sac*I primer (50 ng/mL), 1 μL of *Mse*I primer (50 ng/mL), 2 μL of dNTPs (5 mM), 5 μL of 10× PCR buffer (100 mM Tris-HCl, 15 mM $MgCl_2$, and 500 mM KCl, pH 8.3), and 0.2 μL of *Taq* DNA polymerase (5 units/mL; Roche, Basel) were brought to a final volume of 50 μL with water. Amplification reactions were performed in a PerkinElmer 9600 thermocycler: 20 s at 94°C, 30 s at 65°C to 56°C (Δt = 0.7°C), and 60 s at 72°C for 13 cycles and 15 s at 94°C, 30 s at 56°C, and 35 s at 72°C for 30 cycles. Five microliters of the PCR reaction was analyzed on an agarose gel to check for the correct fragment size; samples were

directly sequenced using the same primers as in the re-amplification reaction described above.

Distribution of AFLP Markers over the Arabidopsis Genome

To determine the distribution of AFLP markers over the Arabidopsis genome, the file named AGI.nonredundant (November, 26, 2000) from the TAIR Web page (ftp://tairpub:tairpub@ftp.Arabidopsis.org/home/tair/AGI/) was used. In this way, all markers are placed on the continuous DNA sequence. It should be noted that the file AGI.nonredundant is, in fact, redundant because it contains BAC overlaps. Therefore, the assigned distances between some of the markers that were placed on the continuous sequence are undoubtedly greater than they are in reality.

ACKNOWLEDGMENTS

We thank Dr. José L. Micol for providing F_2 seeds for the leaf-form mutants, angusta4 and rotunda2-1; Dr. Sergei Kushnir for the viviparous-like mutant material; Gert van Hoecke and the sequence group for determining the sequences of most of the 70 AFLP markers tested; Dr. Mieke Van Lijsebettens for coordinating the leaf developmental program; and Prof. Judy Strommer and Dr. Marnik Vuylsteke for critically reading the manuscript.

Received June 7, 2001; returned for revision August 2, 2001; accepted September 10, 2001.

LITERATURE CITED

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF et al. (2000) The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195

Alonso-Blanco C, Peeters AJ, Koornneef M, Lister C, Dean C, van den Bosch N, Pot J, Kuiper MT (1998) Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. Plant J **14:** 259–271

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815

Arnold C, Metherell L, Clewley JP, Stanley J (1999) Predictive modelling of fluorescent AFLP: a new approach to the molecular epidemiology of *E. coli*. Res Microbiol **150:** 33–44

Berná G, Robles P, Micol JL (1999) A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana*. Genetics **152:** 729–742

The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012–2018

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N et al. (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat Genet **23:** 203–207

Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD et al. (1999) Genetic definition and sequence analysis of Arabidopsis centromeres. Science **286:** 2468–2474

Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000). A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis.* Plant Physiol **124:** 1483–1492

Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L, Ferreira A, Webb DM (1997) A high-density soybean genetic map based on AFLP markers. Crop Sci **37:** 537–543

Kuiper M (1998) Building a high-density genetic map using the AFLP technology. *In* J Martinez-Zapater, J Salinas, eds, Methods in Molecular Biology: *Arabidopsis* Protocols, Vol 82. Humana Press, Totowa, NJ, pp 157–171

Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M et al. (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

Lukowitz W, Gillmor CS, Scheible WR (2000) Positional cloning in *Arabidopsis*: why It feels good to have a genome initiative working for you. Plant Physiol **123:** 795–806

Mayer K, Schuller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Dusterhoft A, Stiekema W, Entian KD, Terryn N et al. (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature **402:** 769–777

Powell W, Thomas WTB, Baird E, Lawrence P, Booth A, Harrower B, McNicol JW, Waugh R (1997) Analysis of quantitative traits in barley by use of amplified fragment length polymorphisms. Heredity **79:** 48–59

Qin L, Prins P, Jones JT, Popeijus H, Smant G, Bakker J, Helder J (2001) GenEST, a powerful bidirectional link between cDNA sequence data and gene expression profiles generated by cDNA-AFLP. Nucleic Acids Res **29:** 1616–1622

Robles P, Micol JL (2001) Genome-wide linkage analysis of *Arabidopsis* genes required for leaf development. Mol Genet Gen **266:** 12–19

van Eck HJ, van der Voort JR, Draaistra JR, van Zandvoort P, van Enckevort E, Seger B, Peleman J, Jacobsen E, Helder J, Bakker J (1995) The inheritance of chromosomal localization of AFLP markers in non-inbred potato offspring. Mol Breed **1:** 397–410

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

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P et al. (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet **99:** 921–935