Theor Appl Genet (1999) 99:785–790

W.P. Young · J.M. Schupp · P. Keim

DNA methylation and AFLP marker distribution in the soybean genome

Received: 5 November 1998 / Accepted: 20 February 1999

Abstract Amplified fragment length polymorphisms (AFLPs) have become important markers for genetic mapping because of their ability to reliably detect variation at a large number of loci. We report here the dissimilar distribution of two types of AFLP markers generated using restriction enzymes with varying sensitivities to cytosine methylation in the soybean genome. Initially, AFLP markers were placed on a scaffold map of 165 RFLP markers mapped in 42 recombinant inbred ($F_{6:7}$) lines. These markers were selected from a map of over 500 RFLPs analyzed in 300 recombinant inbred ($F_{6:7}$) lines generated by crossing BSR101×PI437.654. The randomness of AFLP marker map position was tested using a Poisson-model distribution. We found that AFLP markers generated using *Eco*RI/*Mse*I deviated significantly from a random distribution, with 34% of the markers displaying dense clustering. In contrast to the *Eco*RI/*Mse*I AFLP markers, *Pst*I/*Mse*I-generated AFLP markers did not cluster and were under represented in the *Eco*RI/*Mse*I marker clusters. The restriction enzyme *Pst*I is notably sensitive to cytosine methylation, and these results suggest that this sensitivity affected the distribution of the AFLP markers generated using this enzyme in the soybean genome. The common presence of one *Eco*RI/*Mse*I AFLP cluster per linkage group and the infrequent presence of markers sensitive to methylation in these clusters are consistent with the low recombination frequency and the high level of cytosine methylation observed in the heterochromatic regions surrounding centromeres. Thus, the dense *Eco*RI/*Mse*I AFLP marker clusters may be revealing structural features of the soybean genome, including the genetic locations of centromeres.

Communicated by S. Maroof

W.P. Young · J.M. Schupp · P. Keim (✉)
Department of Biological Sciences,
Northern Arizona University, Flagstaff, AZ 86001-5640, USA
Fax: 520 523 7500
e-mail: paul.keim@van.edu

Key words Soybean · *Glycine max* · AFLP · Methylation · Genetic mapping · Centromeres

Introduction

Genetic maps are extremely important resources for plant and animal genetic analyses. The development of new techniques has made high-density molecular-marker-based maps available for many plant and animal species. This was first made possible with the discovery of restriction fragment length polymorphism (RFLPs; Botstein et al. 1980). Since that time, PCR-based markers such as microsatellites (Dib et al. 1996; Dietrich et al. 1997) and AFLPs (Vos et al. 1995) have become important for linkage analysis in many species. Despite the relatively recent development of the AFLP technique, the widespread acceptance of AFLP markers for mapping is evident by the large number of recent published mapping analyses. The effectiveness of AFLP markers is due largely to their ease, abundance and reproducibility. AFLP markers have been used for mapping in populations of intensively mapped species such as *Arabidopsis thaliana* (Alonso-Blanco et al. 1998), soybean (Keim et al. 1997), rice (Mackill et al. 1996) and barley (Becker et al. 1995), as well as relatively uncharacterized species such as pinyon pine (Travis et al. 1999). The power of the technique comes from the ability to combine different AFLP primer combinations, resulting in a nearly unlimited supply of markers. Because of the ability to screen thousands of polymorphisms, AFLP markers have been particularly useful for physically locating and/or cloning important genes when combined with bulking strategies (Meksem et al. 1995; Cnops et al. 1996; Buschges et al. 1997).

In plant genomes, cytosine (C) methylation of CpG and CpNpG nucleotides varies in frequency along a chromosome and acts to regulate gene expression either at the gene level or else regionally to influence entire regions of chromosomes (for a review see Kass et al. 1997). The regulation of gene expression occurs by

changing local chromatin structure (Davey et al. 1997), preventing the binding of DNA-binding proteins to promoter regions (Inamadar et al. 1991) or as a binding cue for transcriptional repressors (Kass et al. 1997). Regional-methylation functions to inactivate heterochromatin, and elements in or near heterochromatin, by increasing the methylation frequency compared to that in euchromatic regions (Bird AB, 1986).

Like other important crop plants, soybean has a large and complex genome as a result of both abundant repetitive sequences and having a polyploid ancestry. For many years the methylation-sensitive restriction enzyme *Pst*I has been used to clone single-copy sequences for use as RFLP probes for linkage analyses in order to avoid repetitive DNA sequences (Burr et al. 1988). Targeting these hypomethylated regions by using *Pst*I-derived RFLP clones as markers has most probably increased the frequency of markers in genetically active euchromatic regions. A similar targeted approach may be possible using the AFLP technique by utilizing the same methylation-sensitive restriction enzyme, *Pst*I, during the restriction step of the AFLP technique.

The purpose of the present study was to compare the genomic distribution of AFLP markers generated using a methylation- and a non-methylation-sensitive restriction enzyme. AFLP markers were generated using the restriction enzymes *Eco*RI and *Mse*I (*Eco*-AFLP markers) and *Pst*I and *Mse*I (*Pst*-AFLP markers). *Eco*RI and *Pst*I differ greatly in their ability to cut restriction sites containing methylated C; *Pst*I (5′CTGCAG3′) is greatly inhibited by C methylation whereas *Eco*RI (5′GAATTC3′) is relatively insensitive to C methylation. By comparing the genomic distribution of these two types of AFLP markers, we have concluded that the restriction enzyme used in AFLP analysis can bias marker distribution to different regions, probably due to genomic methylation patterns. We have observed that usually only one AFLP marker cluster exists per linkage group, which is consistent with their association with centromeres. These results potentially provide the first systematic mapping of soybean centromeric regions and also suggest an effective strategy for creating highly saturated genetic maps (1 marker/cM).

Materials and methods

AFLP markers were produced as described by Vos et al. (1995) and modified by Travis et al. (1996). Following the restriction/ligation step, adenine was used as the pre-selective (+1) nucleotide in all cases. Subsequent amplifications utilized three selective nucleotides (+3/+3) including the +1 adenine. Markers were named as described in Keim et al. (1997). The first three letters represent the *Eco*RI+3 selective nucleotides, the last three letters represent the *Mse*I +3 nucleotides and the number is the molecular weight of the PCR product.

The *Eco*-AFLP markers have been previously published as a soybean genetic map that included 165 RFLPs, 650 *Eco*-AFLP and 30 RAPD markers (Keim et al. 1997). Although this map was based on the segregation of markers in a population of 42 recombinant inbred lines (RILs), it was constructed using a scaffold of 165 RFLP markers selected from over 500 public and proprietary RFLP markers. These markers were analyzed in a recombinant inbred (RI) population of 300 $F_{6:7}$ lines (PI437.654×BSR-101).The *Eco*-AFLP and RAPD markers were subsequently mapped using the genotype data from the 42 RILs while maintaining the order of the RFLP markers. The *Pst*-AFLP markers were added to the map using the 42 RILs and the published order of the *Eco*-AFLP markers as a framework, in this case maintaining the order of the *Eco*-AFLP markers. Therefore the order presented here is based on a framework of over 500 RFLPs in a population of 300 $F_{6:7}$ lines.

Linkage analysis was done using Mapmaker/Exp version 3.0 (Lander et al. 1987). The *Pst*-AFLP markers were first sorted to linkage groups using the group command (LOD 4.0, theta 25), then placed in their most-likely position using the try command. If a marker could not be uniquely positioned to an interval with a LOD>3.0 it was tried in the most-likely positions and scrutinized for disruption of the framework markers. The most common form of disruption was extensive map extension of an interval or decreased LOD scores for surrounding markers. This was attributed to genotyping errors or, more commonly, to the placement of a marker in a marker-dense interval that resulted in the placement being determined by a small number of recombinants. *Pst*-AFLP markers placed in densely populated regions mapped with a similar LOD probability to all closely linked markers, so the exact placement could not be identified. However, these markers generally did not result in a great deal of map expansion and were left in the analysis. Markers that contained numerous genotyping errors could be identified because they mapped with low LOD likelihood to two locations on the linkage group, an internal interval and at the end of the linkage group. Since these markers could not be correctly placed there were dropped from the analysis.

Eco-AFLP marker distribution was analyzed using a Poisson distribution function ($P(x)=e^{-\mu}\mu^x/x!$ (μ=1.86 markers/10 cM interval). The minimum number of markers in a cluster was determined from the Poisson distribution of the marker data. Analysis indicated that a density of seven or more markers in a 10-cM interval represented the lower threshold of the significant deviation from the expected number of markers/interval (χ^2=464.6, 6 *df*, P<0.001). Clusters on the map were identified by sliding a 10-cM interval over the linkage groups. At any point where the number of markers in the interval was greater than or equal to seven, that spot was designated as a cluster. Only the *Eco*-AFLP markers were analyzed by this method.

Because of the smaller number of *Pst*-AFLP markers in this study, their distribution was compared to that of the *Eco*-AFLP markers using a chi-square test. The expected frequency of *Pst*-AFLP markers in the clusters being equal to the frequency of *Eco*-AFLP markers found in clusters. Significant deviation would indicate an alternate distribution of the *Pst*-AFLP markers compared to that of the *Eco*-AFLP markers.

Results

Clustering of *Eco*RI AFLP markers

Previous linkage analysis etablished the placement of the 650 *Eco*-AFLP markers used in this investigation. That composite map including the *Eco*-AFLP markers, RFLPs and RAPDs, contained 28 linkage groups covering 3441 cM (see Keim et al. 1997). This represented approximately 1 *Eco*-AFLP marker/5 cM. However, given the high degree of *Eco*-AFLP marker clustering, many large gaps actually remained. Apparent clustering of the *Eco*-AFLP markers was noted by Keim et al. (1997), but no quantitative analysis was performed.

We have analyzed the distribution of the *Eco*-AFLP markers and determined that 34% of the markers segregated into clusters of greater than seven markers within a

10-cM interval. In contrast, no *Pst*-AFLP clusters were observed and only 10.9% of these markers were found within the *Eco*-AFLP cluster regions. Linkage group L20 was omitted from this analysis due to extremely skewed segregtion and no *Pst*-AFLP markers mapped to this linkage group. The number of AFLP markers in each linkage group, the number of *Eco*-AFLP markers in the clusters, and the number of *Pst*-AFLP markers in the *Eco*-AFLP-defined clusters are presented in Table 1. Assuming a random marker distribution, the number of 10-cM intervals containing a given number of markers is expected to follow a Poisson distribution. Overall, our *Eco*-AFLP marker results deviated significantly from Poisson expectations (χ^2=494.5, 7 *df*, *P*<0.0001). Of particular significance were the extreme deviations in the number of intervals containing ≥7 markers (obs=20; exp=1) or zero markers (obs=116; exp=50). Figure 1 shows the observed and expected number of *Eco*-AFLP markers/10-cM interval for the entire map, excluding linkage group L20.

Eco-AFLP clusters were identified on 15 linkage groups including 13 out of the 18 linkage groups that could be correlated with previously published maps based on the identification of common RFLP markers (Fig. 2). Identifiable linkage groups that could be correlated with other maps were designated using capital letters consistent with those used in the website SOYBASE (http:\\129.186.26.94/default.html). Linkage groups that could not be matched were arbitrarily designated using 'L' and a number. Linkage group G illustrates the dense

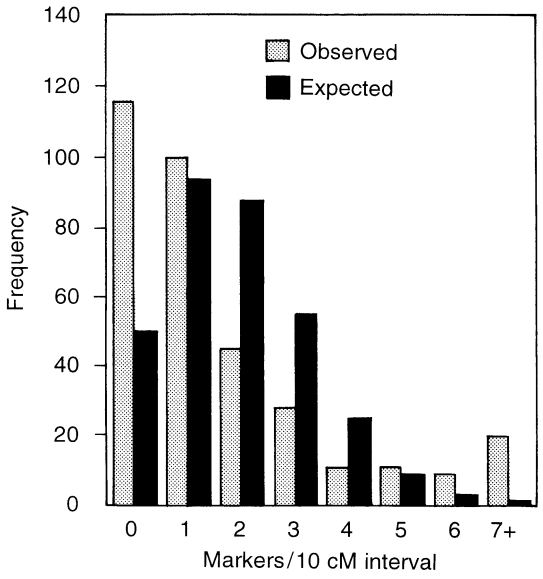

Fig. 1 The marker distribution of the *Eco*-AFLP markers compared to that expected assuming a Poisson distribution. Intervals with the same marker counts were summed and the resulting frequencies compared against the expected frequencies generated from the Poisson distribution function $P(x)=e^{-\mu}\mu^x/x!$: μ is equal to the average marker density across the entire map (1.86 markers/10 cM; σ=6.66). The two distributions were compared using a χ^2 goodnes of fit test. Linkage group L20 was omitted from the analysis due to extremely skewed segregation patterns

Table 1 AFLP marker composition of soybean linkage groups

Linkage group[a] (cM)	AFLP markers	*Eco*-AFLPs	*Pst*-AFLPs	cM/AFLP	Clustered *Eco*-AFLPs[b]	*Pst*-AFLPs in clusters[c]
A2 (250)	53	42	11	4.7	9	2
B1 (193)	44	39	5	4.4	23	1
C2 (100)	31	28	3	3.2	16	1
C1 (96)	52	48	4	1.8	25	1
D1a (21)	56	44	12	3.8	15	2
E (206)	31	27	4	6.6	8	0
F (180)	29	20	8	6.2	0	0
G (320)	62	53	9	5.2	18	0
H (138)	26	20	6	5.3	7	0
I (139)	27	20	7	5.1	0	0
J (166)	35	29	6	4.7	8	0
K (136)	42	33	9	3.2	22	3
L (100)	7	6	1	14.3	0	0
M (185)	40	35	6	4.6	8	0
N (135)	37	30	7	3.6	17	1
P (65)	15	12	3	4.3	0	0
S (114)	25	23	2	4.6	12	0
W (75)	27	17	9	2.8	0	0
L14 (205)	42	32	10	4.9	18	1
L15 (27)	4	2	2	6.8	0	0
L16 (16)	2	1	1	8	0	0
L18 (47)	5	2	3	9.4	0	0
L19 (16)	2	2	0	8	0	0
L21 (15)	3	3	0	5	0	0
L22 (25)	6	5	1	4.2	0	0
L23 (33)	11	8	3	3	0	0
L26 (64)	32	26	6	2	15	3
Total (3256)	746	607	138	4.4	221	15

[a] Linkage groups from SOYBASE (see text) except for L-groups which could not be correlated to known previously identified groups
[b] The total number of *Eco*-AFLP markers found in clusters of >7 markers/10 cM
[c] The total number of *Pst*-AFLP markers found in the *Eco*-AFLP clusters

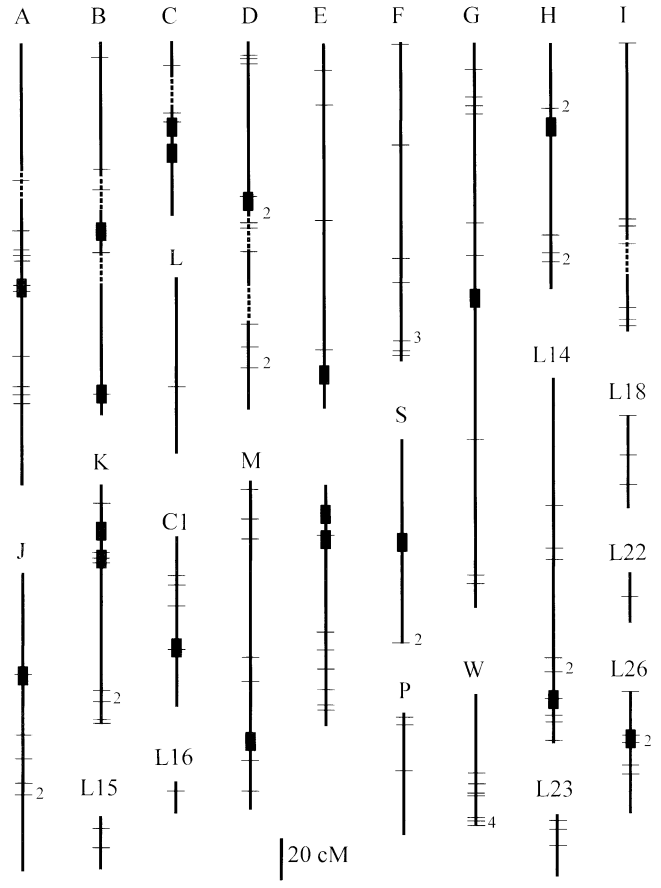

Fig. 2 *Pst*-AFLP marker distribution on the soybean linkage map. The *Pst*-AFLP markers are represented by *crosshatches*. Exact locations can be found on the website SOYBASE (see text). *Shaded boxes* represent the locations of the *Eco*-AFLP marker clusters. Only the positions of the *Pst*-AFLP markers are included on the map. *Numbers* indicate two or more *Pst*-AFLP markers mapping to the same position. Linkage groups L19, L20 and L21 were omitted from the figure because no *Pst*-AFLP markers were identified on them

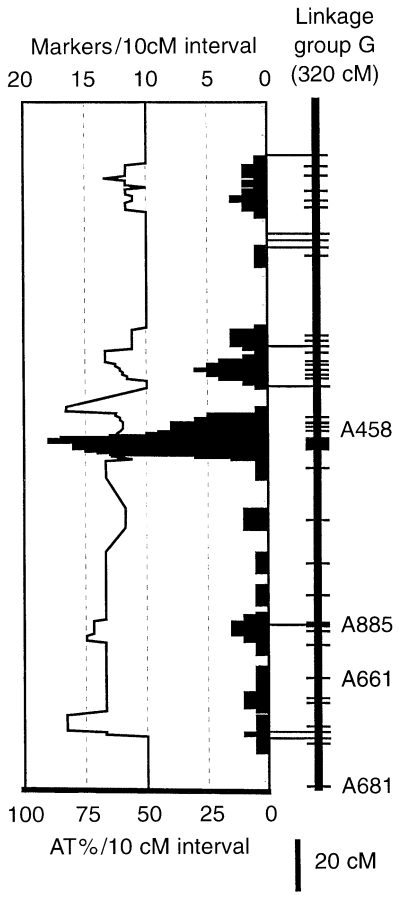

Fig. 3 AFLP marker density and selective AT% on linkage group G. AFLP marker density along linkage group G was determined by shifting a 10-cM interval along the map in 1-cM increments. *Horizontal bars* indicate the number of AFLP markers in the given interval as indicated by the top axis. Each bar is centered across from is corresponding interval. The *solid line* indicates the average selective AT% of the markers in the corresponding 10-cM interval, as indicated by the bottom axis. For illustrative purposes, intervals lacking markers appear as a *straight portion* of the solid line between two flanking intervals that contain markers. Public RFLP marker postions and names are included only for reference purposes and were not included in the analysis. *Short horizontal lines* across the linkage group indicate *Eco*-AFLP marker positions. *Long horizontal lines* across the linkage group indicate the position of *Pst*-AFLP markers. Note that multiple AFLP markers may occupy the same map position

clustering of *Eco*-AFLP markers (Fig. 3). The marker-frequency histogram was aligned with the genetic map to show the *Eco*-AFLP marker clustering that was similarly observed on all other linkage groups that contained clusters. The histogram was produced by moving a 10-cM "window" along the length of the linkage group in 1-cM steps. The large peak near the center of the linkage group represents a region of high-marker density. Because marker density might be a function of A+T/G+C content, we have plotted the A+T% of the +3 selective nucleotides. This provides one estimate of genomic nucleotide composition. No correlation between the percent A+T and clustering was observed and this value varied in the marker clusters over a range of 33–83% A+T (Fig. 3). In addition, the overall A+T percentage of the clustered *Eco*-AFLP markers was approximately equivalent to that of all *Eco*-AFLP markers, 63.6 and 64.4%, respectively.

Two separate *Eco*-AFLP marker clusters were identified on five of the linkage groups. Examination of these linkage groups revealed that three contained very tightly linked clusters separated by less than 5 cM (C2, K and N). The two remaining linkage groups had clusters that were separated by relatively large map distances (B1 and L14). All other linkage groups with clusters contained only a single high-density marker cluster. The AFLP markers that flank the clusters and, when available, the public RFLP marker closest to the cluster, are presented in Table 2.

*Pst*I AFLP marker distribution

A total of 189 *Pst*-AFLP markers were identified using 41 primer combinations. Of these, 137 *Pst*-AFLP mark-

Table 2 Markers near the AFLP clusters on soybean linkage groups

Linkage group[a]	Nearest RFLP[b]	Flanking AFLP markers[c]	
		Cluster 1 (# of AFLPs)[d]	Cluster 2 (# of AFLPs)[d]
A2 (250)	Bng121H	aacagc225 (9) aggact44	
B1 (193)	A808	aggatc180 (8) aaaaac210	acgaca79 (15) aaaaca230
C2 (100)	A748	aagagt182 (8) aaaact170	acgaat370 (8) aagagt250
C1 (96)	K477v	acgata232 (25) aaaagc340	
D1a (210)	K226	aaaacg182 (15) aaaagc195	
E (206)	–	aaaagg124 (8) aaaact256	
G (320)	A458	acgaat290 (18) aggact40	
H (138)		aacacc400 (7)	
J (166)	–	actagt276 aaaaac295 (8)	
K (136)	–	aacacal172 aacact232 (9) acgaaa445	acaagal136 (13) actata99
M (185)	A715V	accagc315 (8) aggatg104	
N (135)	A181	aggaga205 (8) aggata400	actaag475 (9) aacata90
S (114)	–	aaaagt475 (12) aacagt210	
L14 (205)	K265	aaaagt113 (11) aagaag98	aggatt61 (7) accact400
L26 (64)	K79	aacatg168 (15) acaagt245	

[a] Linkage groups from SOYBASE (see text) except for L-groups which could not be correlated to previously identified groups
[b] The public RFLP markers closest to the *Eco*-AFLP cluster
[c] *Eco*-AFLP markers that flank the AFLP cluster. See text for description of marker names
[d] The total number of *Eco*-AFLP markers found in the clusters

ers were placed on the map, 12 were linked to other AFLP markers that were not on the map (both *Eco*- and *Pst*-AFLP markers), 12 were unlinked, and 28 were dropped from the analysis because of skewed segregation ratios or because they could not be accurately positioned. Linkage analysis determined that only 10.9% of the *Pst*-AFLP markers mapped in the *Eco*-AFLP-defined clusters. This was a significant deviation from the expected frequency (34%) assuming that the frequency of the *Pst*-AFLP markers in the clusters was equal to that of the *Eco*-AFLP markers (χ^2=21.4, 1 *df*, P<0.0001). Figure 3 shows the location of the *Pst*-AFLP markers in relationship to the *Eco*-AFLP clusters on linkage group G. This was similar to what was observed on the other linkage groups. *Pst*-AFLP markers generally mapped to the lower marker-density regions, similar to the RFLP markers. Indeed, many *Pst*-AFLP markers mapped to the tips of the existing linkage groups. The approximate positions of the *Pst*-AFLP markers are presented in Fig. 2. Exact positions can be obtained from the website SOYBASE (http:\\129.186.26.94/default.html).

Although the *Pst*-AFLP markers demonstrated an alternate distribution pattern compared to the non-random pattern observed for the *Eco*-AFLP markers, the genome coverage of the *Pst*-AFLP markers appeared to be similar based upon the number of markers mapped on each linkage group. To prove this, a regression analysis comparing the number of *Eco*-AFLP markers per linkage group with the number of *Pst*-AFLP markers per group was performed. The significant result (r^2=0.49, P<0.01) indicated that the number of AFLP markers of either type found on a particular linkage group was correlated, most likely to the size of the linkage group. Keim et al. (1997) demonstrated that the number of bands and the number polymorphisms per primer combination were positively correlated with the A/T percentage of the +3/+3 selective nucleotides for the *Eco*-AFLP markers. Likewise, the *Pst*-AFLP markers showed a similar correlation. However, the *Pst*-AFLP markers on average produced approximately half as many bands and polymorphisms as did the *Eco*-AFLP markers for a given +3/+3 selective nucleotide percentage (data not shown). The reduction in fragments, and hence polymorphic markers, is consistent with the lesser number of restriction fragments generated by *Pst*I genomic digests (Zhu et al. 1994).

Discussion

We have investigated the genetic distribution of *Pst I*/MseI- and *Eco*RI/*Mse*I-generated AFLP markers in the soybean cross BSR101×PI437.654. Distribution analysis revealed that the *Eco*-AFLP markers deviated signifi-

cantly from Poisson expectations, confirming that the *Eco*-AFLP markers were not uniformly distributed on the genetic map. As a result, the linkage map consisting of the *Eco*-AFLP markers contained many clusters of up to 25 markers and many large gaps (see Keim et al. 1997). In contrast, the *Pst*-AFLP markers generally mapped to regions containing a lower density of markers, similar to that of the RFLP markers, with significantly fewer *Pst*-AFLP markers mapping to the *Eco*-AFLP-defined clusters (Fig. 2). These results suggested that the enzymes used to generate the AFLP markers influenced their genomic distribution.

The shared feature of the *Pst*-AFLP and -RFLP markers was the use of the restriction enzyme *Pst*I for marker generation which, in contrast to *Eco*RI, is highly sensitive to C methylation. Localized variation in methylation frequency has been determined to be present in many complex plant genomes, including soybean (Gruenbaum et al. 1981; Hepburn et al. 1987; Zhu et al. 1994). The highly significant relationship between marker clustering and restriction-enzyme methylation sensitivity suggested that the two AFLP types preferentially identified markers in genomic regions exhibiting dissimilar C methylation frequencies. Similar to *Pst*I-generated RFLP markers, the *Pst*-AFLP markers were most likely located in regions of the genome where the frequency of C methylation was low. In contrast, the placement of *Eco*-AFLP markers was not significantly influenced by the level of methylation, and the clusters resulted from markers being located in regions were the frequency of recombination was low (see below). It has been suggested that the apparent clustering of *Eco*RI/*Mse*I-generated AFLP markers was caused at least in part by the relativelyA+T-rich recognition sequence of *Mse*I that biased the placement of these markers to chromosomal regions with higher A+T contents (Alonso-Blanco et al. 1998). In spite of the fact tht noncoding soybean sequences had a higher A+T content than coding sequences (Zhu et al. 1994), examination of the selective +3/+3 nucleotide composition of marker clusters was inconsistent with such a model (Fig. 2).

Keim et al. (1997) previously noted the apparent nonuniform marker distribution of the *Eco*-AFLP markers compared to the relatively evenly spaced RFLP markers and speculated that this was the result of the markers being located in regions of reduced recombination, most likely near centromeres. Greatly suppressed recombination in centric regions would result in apparently tight clustering of markers while elevated recombination would exhibit widely dispersed markers. Therefore, the *Eco*-AFLP markers in the clusters were not necessarily physically close, but more likely were spread across a region exhibiting little or no recombination. Although the average rate of recombination is approximately 500 kb/cM in the soybean genome, a region on linkage group G has been revealed to have a level of 150 kb/cM (Nevin Young, personal communication). Given this highly variable recombination rate across the genome, it is likely that the regions containing the marker clusters exhibit little or no recombination.Other plant species show similar patterns. In wheat, the distribution of recombination is highly variable on the chromosome, with approximately 60% of the recombination events occurring on only 18% of the arm length and with no recombination within the closest 20% of the arm surrounding the centromere (Gill et al. 1996). Similarly, a 10-fold reduction in recombination frequency was observed in the centromeric regions compared to the telomeric regions in the tomato genome. In addition to the observed clusters of markers in these regions, karyological analysis determined that they were closely associated with the centromeres (Tanksley et al. 1992). In contrast, a highly saturated map of the rice genome found marker clustering on only a few chromosomes, with regions of recombination-suppression covering approximately 5% of the chromosome arm (Harushima et al. 1998). This discrepancy in marker clustering and in influence of the centromere on recombination might be due to the very small size of the rice genome and the greatly reduced satellite DNA in heterochromatin. Because the soybean genome is more similar in size to wheat and tomato, it is reasonable to assume a similar pattern of centromeric influence.

In addition to a reduced recombination rate, a second characteristic of centromeric regions is a high level of C methylation in the surrounding heterochromatin. The clustering of *Eco*-generated AFLP markers and the deficiency of either *Pst*I-generated AFLP markers or RFLPs in these clusters is consistent with the high levels of methylation in the regions marked by the clusters. The use of *Pst*I for RFLP marker generation has been used to avoid these hypermethylated heterochromatic regions (Burr et al. 1988, Keim and Shoemaker, 1988). Our results imply that the *Pst*I-generated AFLP markers behave in a similar fashion.

A majority of the major linkage groups contained an *Eco*-AFLP marker cluster. This is also consistent with the clusters marking the centromeric regions of the chromosomes. Because the haploid number in soybean is 20, a complete map should contain 20 linkage groups, each with a single cluster/centromere. Thirteen of the eighteen major groups, and two of the groups that could not be positively correlated to known soybean chromosomes (labeled L groups), contained *Eco*-AFLP clusters. Of the 15 groups that contained clusters, all but two contained either one cluster or two closely spaced clusters. The presence of two closely spaced clusters most likely characterizes a single centromeric region that contains a segment either lacking in markers (e.g., very simple repetitive DNA) or else exhibiting a higher level of recombination than the flanking regions. Groups B1 and L14 each contained two widely spaced clusters. However, in each of these examples, one of the clusters contained a marginally significant number of markers, seven and eight, respectively. The presence of an additional region of suppressed recombination on the chromosome has been observed in other plant species (Tanksley et al. 1992). Therefore, the presence of a single cluster on each of the major groups is consistent with the expectation of

one centromeric region per chromosome. Because this genetic map did not coalesce into 20 linkage groups, linkage groups without major marker clusters (i.e., centromeres) may represent distal chromosomal regions that are actually part of a centromere-containing linkage group. Alternatively, some chromosomes may not have suppressed centric recombination. Gill et al. (1996) demonstrated that the centric recombination frequency in wheat varies from chromosome to chromosome and doesn't depend solely on the proximity to the centromere.

In spite of the fact that *Eco*RI/*Mse*I-AFLP clusters have been observed in potato (van Eck et al. 1995), barley (Becker et al. 1995; Castiglioni et al. 1998) and soybean (Keim et al. 1997), *Eco*-AFLP clusters have only been physically localized on the centric regions of the chromosomes in *Arabidopsis* (Alonso-Blanco et al. 1998). The clustering of *Eco*-AFLP markers on *Arabidopsis* chromosomes was attributed to a combination of the centromeric suppression of recombination and the preferential placement of AFLP markers in these regions. Centromeric suppression of recombination was responsible for the clustering of RFLP (Tanksley et al. 1992; Gill et al. 1996) and AFLP markers (Alonso-Blanco et al. 1998) in other species, so it is apparent that it can have a major influence on marker spacing. The preferential placement of AFLP markers in recombinationally suppressed regions implies that some aspect of the AFLP technique biases these markers to a specific region of the genome, specifically centromeric regions. Reasons given for this included a higher frequency of polymorphisms in these regions, because they are mainly noncoding sequences, and the use of *Mse*I that cuts more frequently in high-A+T regions due to its restriction sequence (Alonso-Blance et al. 1998). Our results demonstrated that the use of the enzymes *Pst*I and *Mse*I for AFLP marker generation does not result in ALFP marker clustering. This implies that *Mse*I, at least in combination with *Pst*I, does not greatly influence marker placement. Additional genetic and physical mapping of the regions marked by the clusters will be necessary in order to determine if a higher rate of marker polymorphism, in addition to suppressed recombination, contributed to our results.

Implications for genetic mapping

Our results provide a great deal of support for the fact that the AFLP clusters observed on the soybean genetic map mark the heterochromatic regions surrounding the centromeres. The centromeric boundaries represented by the flanking AFLP markers and the closest RFLP markers provide important landmarks for the centromeres on the map (Table 2). The RFLP markers are common to many elite strains and therefore should be useful as centromeric markers in many crosses. The relatively difficult cytological and aneuploid mapping studies will be necessary to confirm the centric location of these clusters.

The main advantage of the AFLP technique is its ability to rapidly generate a large number of markers for segregation analysis. Given the relatively common occurrence of AFLP clustering in mapping analyses in other organisms, it would be beneficial to utilize enzymes with varying sensivities to methylation. Our findings further enhance this technique by demonstrating that specific chromosomal regions can be targeted using AFLP markers. Since the vast majority of expressed genes are associated with hypo-methylated regions, it is extremely important to saturate these regions with markers. The *Pst*-AFLP markers mapped in this analysis were significantly less likely to map to the clusters, and therefore increased the density of markers in previously marker-less regions. In addition, many mapped to small groups and on the ends of existing groups, thus increasing coverage of the genome. However, the use of non-methylation-sensitive enzymes such as *Eco*RI for AFLP marker generation should not be completely avoided. It is likely that the physical distance separating the markers in the clusters was actually quite large and therefore a large number of markers combined with a large mapping population would be necessary to genetically map and physically locate genetic factors in these regions. A mapping strategy combining methylation-sensitive and -insensitive enzymes with the AFLP technique would provide complete map coverage and allow for the targeting of specific regions based on genomic methylation patterns.

Acknowledgments This research was supported by The E. Raymond and Ruth Reed Cowden Endowment and the United Soybean Board to W.P.Y. and P.K.

References

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

Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 249:65–73

Bird AP (1986) CpG-rich islands and the function of DNA methylation. Nature 321:209–213

Botstein D, White R, Skolnick M, Davis R (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88:695–705

Castiglioni P, Pozzi C, Heun M, Terzi V, Muller KJ, Rohde W, Salamini F (1998) An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. Genetics 149:2039–2056

Cnops G, den Boer B, Gerats A, Van Montagu M, Van Lijsebettens M (1996) Chromosome landing at the *Arabidopsis* TORNADO1 locus using an AFLP-based strategy. Mol Gen Genet 253:32–41

Davey C, Pennings S, Allen J (1997) CpG methylation re-models chromatin structure in vitro. J Mol Biol 267:276–288

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-boles D, Husin Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ, Evans CA, Deangelis MM, Levinson DM, Kruglyak L, Goodman N, Copeland NG, Jenkins NA, Hawkins TL, Stein L, Page DC, Lander ES (1997) A comprehensive genetic map of the mouse genome. Nature 380:149–151

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Mrc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5246 microsatellites. Nature 380:152–154

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

Gill DS, Gill BS, Endo TR, Boyko EV (1996) Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics 143:1001–1012

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292: 860–862

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yakamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148: 479–494

Hepburn AG, Belanger JC, Mattheis JR (1987) DNA methylation in plants. Dev Genet 8:475–493

Inamdar NM, Ehrlich KC, Ehrlich M (1991) CpG methylation inhibits binding of several sequence-specific DNA-binding proteins from pea, wheat, soybean and cauliflower. Plant Mol Biol 17:111–123

Kass SU, Pruss D, Wolffe AP (1997) How does DNA methylation repress transcription. Trends Genet 13(11):444–449

Keim P, Shoemaker RC (1988) Construction of a random recombinant DNA library that is primarily single-copy sequence. Soybean Genet Newslett 15:147–148

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

Lander ES, Green P, Abrahmson 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

Mackill DJ, Zhang Z, Redona ED, Colowit PM (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. Genome 39:969–977

Meksum K, Leister D, Peleman J, Zabeau M, Salamini F, Gebhardt C (1995) A high-resolution map of the vicinity of the R1 locus on chromosome V of potato based on RFLP and AFLP markers. Mol Gen Genet 249:74–81

Tanksley SD, Granal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Foulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AG, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High-density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Travis SE, Maschinski J, Keim P (1996) An analysis of genetic variation in *Astragalus cremnophylax* var. cremnophylax, a critically endangered plant, using AFLP markers. Mol Ecol 5:735–745

Travis SE, Ritland K, Whitham TG, Keim P (1998) A genetic linkage map of pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. Theor Appl Genet 97:871–880

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

Zhu T, Schupp JM, Oliphant A, Keim P (1994) Hypomethylated sequences: characterization of the duplicate soybean genome. Mol Gen Genet 244:638–645