Characterization and Identification of (CT)\textsubscript{n} Microsatellites in Soybean Using Sheared Genomic Libraries

Khwaja G. HOSSAIN,* Hishashi KAWAI, Masaki HAYASHI, Masako HOSHI, Naoki YAMANAKA, and Kyuya HARADA

Faculty of Horticulture, Chiba University, 648 Matsudo, Chiba 271-8510, Japan

(Received 6 October 1999; revised 25 January 2000)

Abstract

Three small insert (300 ~ 600 bp) sheared genomic libraries were constructed by pipetting and DNase I treatment of soybean DNA. About 15,000 clones from each library were screened for CT- simple sequence repeats (CT-SSRs). The CT-SSRs were abundant in the soybean genome at an estimated frequency of approximately one SSR per 110 kb of genomic DNA. Following the sequencing of 129 positive clones, the repeat types and frequency of CT repeats among the positive clones were characterized. Forty-nine primer pairs were designed and preliminarily evaluated for their ability to amplify genomic DNA from a set of six varieties, including parents of a mapping family. Amplified products were analyzed by 10% PAGE. Eighty-eight percent of the designed primers were able to amplify all these genomic DNAs using a single PCR profile of 53°C annealing temperature, of which 22 (45%) were polymorphic in the six varieties, and 14 of them were polymorphic in the parents of the mapping family. The polymorphic primer sets were further assessed for allelic information using DNA from 16 soybean cultivars. The average number of alleles was 4, ranging from 2 to 7 with the highest polymorphism information content value 0.84. Fourteen of these SSRs were mapped, using an existing soybean RFLP map. The findings presented here will advance our understanding of the soybean genome, and assist in the mapping genome and discrimination of closely related varieties of this species.

Key words: SSRs; CT repeats; random shearing; primer design; DNA amplification; PIC

1. Introduction

Simple sequence repeats (SSRs) or microsatellites, consist of stretches of mono-, di-, tri-, tetra-, penta-, or hexanucleotide tandemly repeated motifs. SSRs are extremely abundant, are randomly distributed throughout the eukaryotic genome and exhibit considerable length variation. The DNA sequences flanking microsatellites are, however, conserved within individuals of the same species, and thus allow for the selection of primers for the polymerase chain reaction (PCR) and the amplification of the intervening microsatellites in all genotypes. Considerable length variation is observed among the amplified products, resulting from the variation between alleles in the number of tandemly repeated units at the locus.\textsuperscript{1} The presence and hypervariability of SSRs have been studied in many species, including soybean,\textsuperscript{2,3} Brassica,\textsuperscript{4} Grapevine,\textsuperscript{5} Arabidopsis,\textsuperscript{6} maize\textsuperscript{7} and rice.\textsuperscript{8,9}

Plant microsatellites have provided a new source of markers for the construction of genetic maps\textsuperscript{8} and for unambiguously distinguishing cultivars in different species.\textsuperscript{10} In addition to the high level of polymorphism, the application of microsatellites as sequence-tagged sites is more convenient than the use of classical restriction fragment length polymorphisms (RFLPs), once the procedures for their isolation and detection have been identified. Both the RFLPs and random amplified polymorphic DNA (RAPDs) have been used extensively in plant genetic analysis\textsuperscript{11,12} but, despite their usefulness for selected applications, both methods suffer from limitations, like misrepresentation or misinterpretation. Because of the limited number of alleles per locus in soybean, it has also been reported that the use of RFLP probes for soybean genetic analysis is limited, and that SSRs are the preferred alternative for genetic analysis of this species.\textsuperscript{10}

Analysis of the published DNA sequences of 54 plant species for possible microsatellites, suggested that (AT)\textsubscript{n} was the most abundant SSR type, followed by (A)\textsubscript{n}, (CT)\textsubscript{n}, (AAT)\textsubscript{n}, (ATT)\textsubscript{n}, (AAC)\textsubscript{n}, (AGC)\textsubscript{n}, (AAG)\textsubscript{n}, (AATT)\textsubscript{n}, (AAAT)\textsubscript{n}, and (AC)\textsubscript{n}.\textsuperscript{13} The frequency of AT microsatellites was estimated at one per 40 kb in the soybean genome.\textsuperscript{2} This type of repeat in plant
genomes is technically complex to confirm by hybridization analysis, because of the difficulties in using such self-complementary AT repeat sequences as probes. CA repeats are less abundant in soybean compared to human, and as the repeat lengths are significantly shorter, they may not offer sufficient polymorphism for use as genetic markers in soybean. In a recent study, it was observed that the CA repeat is also less polymorphic in maize and difficult to type in this species. The characterization and identification of CT microsatellites have been reported for many species, with frequencies of one per 225–330 kb in rice, per 440 kb in wheat, per 168–710 kb in maize, per 244 kb in Arabidopsis, and one in every 100 kb in Brassica.

The complementary use of database searches and screening of different types of libraries have been suggested for the characterization and identification of microsatellites in a particular genome. A more usual approach, however, is to construct and screen genomic libraries with extremely small inserts (300–600 bp). Such an approach has made it possible to obtain most of the necessary sequence information in a single sequence run and to avoid the necessity for sub-cloning. Usually, combinations of different restriction enzymes have been used for genomic DNA fractionation for constructing a library. However, this method allows only a subset of the inserts to be represeneted in the library. An alternative and preferable method of the use of sheared DNA in library construction was suggested and the use of which has been demonstrated by the isolation of microsatellites in rice.

We have initiated a study to determine the frequency and distribution, and the identification of different microsatellites in the soybean genome using several sheared genomic libraries. The objectives of the present study were: i) to characterize the structure and abundance of (CT)_n microsatellites in the soybean genome, ii) to design primers flanking the microsatellite repeats, and iii) to evaluate the primers for amplifying genomic DNA and for detecting polymorphism in different varieties as well as in a mapping family.

2. Materials and Methods

2.1. Construction of genomic library

A 44 μl (~20 μg) aliquot of genomic DNA from Norin No. 2 was pipetted 250 times using a Pipetman - 200 micropipette with intervals of cold treatment. DNase I (0.01 U) was added to the sheared DNA, and incubated for 1 min at 15°C. After incubation of the enzyme, the DNA was precipitated with ethanol and resuspended in TE. Inserts of 300 ~ 600 bp were carefully isolated by electrophoretic separation through 2% agarose gel using DEAE paper. Blunting and repairing of insert ends were performed by incubation for 30 min at 37°C in a 30 μl reaction containing 5 × forward buffer, 100 mM ATP, 6 U T4 polynucleotide kinase, 4 U T4 DNA polymerase and 2.5 mM dNTP. Inserts were ligated to EcoR1 adapters, following the manufacturer’s protocol (Amersham Pharmacia Biotech), and inserted into a Lambda ZAP II vector using the supplier’s protocol (Stratagene). The ligated vector was packaged into Gigapack III Gold Packaging Extracts (Stratagene), and the packaged products were plated onto 90 mm NZY agar plates followed by blue/white color selection. Three sheared genomic libraries named 138, 510 and 910 were constructed.

2.2. Screening of libraries for clones containing (CT)_n

Phage clones were transferred to the Hybond-N membranes (Amersham Pharmacia). The (CT)_10 oligonucleotides were labeled with DIG - 11 - dUTP (Boehringer Mannheim) and used as probes for screening clones containing microsatellite repeats. Pre-hybridization and hybridization were carried out for 3 hr each in a solution consisting of 5 × SSC, 1% blocking agent, 0.1% lauryl sarcosine, and 0.01% SDS. The membranes were washed three times, each for 10 min with 2 × SSC, 0.1% SDS. First two washes were carried out at room temperature while the third one was at the hybridization temperature. Detection was performed according to the DIG Luminescent detection system protocol (Boehringer Mannheim). About 2000 clones per membrane and a total of 15,000 clones per library were screened. Putative positive clones were screened a second time but at a 2°C higher than the hybridization temperature.

2.3. DNA sequencing

Purified phage clones in Lambda ZAP II were excised in vivo according to the manufacturer’s protocol. Inserts of the resulting plasmid clones were amplified by PCR using M13 forward and reverse primers and purified on Suprec - 02 column (TakaRa). The amplified inserts were used as templates for automated sequencing (ABI Prism 310, Applied Biosystems). In most cases, complete sequences could be obtained in a single run. In other cases, sequencing from both ends was required because of the difficulties encountered in sequencing through SSRs.

2.4. Primer design and amplification of the soybean genome

Primer pairs were designed manually from the sequences flanking a minimum of eight CT repeats, with the following criteria: i) nucleotide length of 18–21 bases ii) a Tm value of 50–60°C, iii) the 3' end base with a G or C, and iv) an amplified fragment size of 85–260 bp. The name of primers was prefixed with the source library. A single annealing temperature of 53°C was used to amplify about 30 ng of genomic DNA from six different varieties of soybean, and parents and progeny of a mapping family. The genomic DNA was amplified in a reaction volume of
20 µl consisting of 2 µl of 10 x reaction buffer, 2 µl of 2.5 mM dNTP, 3 µl of 10 µM primers and 2.5 U of Hot-StarTaq DNA polymerase (Qiagen). Amplifications were carried out with a PCR profile of 95°C for 10 sec, 95°C for 15 min and 33 cycles of 92°C for 1 min, 53°C for 1 min and 68°C for 1 min. Separation of amplified products was performed with 10% non-denaturing PAGE for at least 3 hr at 100 volts.

2.5. Data analysis
The genomic distribution frequency of the SSR was estimated by dividing the estimated total length of the genomic clones of the library by the observed number of the SSRs in the clones. The informativeness of the primer pairs was evaluated by calculating the number of alleles and the polymorphism information content (PIC) for each of the primer sets. The PIC value was computed according to the formula:

\[
PIC = 1 - \sum p_{ij}^2
\]

Where, \( p_{ij} \) is the frequency of the \( j \)th pattern for marker \( i \) and the summation extends over \( n \) patterns.

The polymorphic SSR loci in the mapping parents were integrated into the RFLP map, using the same mapping software for constructing the RFLP map, Mapmaker/Exp., Version 3.0 on a UNIX computer and the same mapping criteria of LOD 3.0 and minimum of 37.2 recombination percentage.

2.6. Plant materials
Norin No. 2, a standard soybean variety in Japan, was used for constructing genomic libraries. Genetic linkage analysis was performed on 100 individuals of an F2 mapping population which was derived from a cross between a Japanese cultivar, Misuzudaizu and a Chinese line, Moshidou Gong 503. Informativeness of the SSR loci was evaluated on 16 soybean cultivars including Norin No. 2 and the parents of the mapping population. The rest of the cultivars were selected randomly from 70 soybean cultivars and accessions.

3. Results and Discussion
3.1. Frequency of CT-SSRs in soybean genome
Of the 45,000 clones screened using the (CT)10 oligonucleotide probes, 174 clones were isolated as positive for CT microsatellites. Based on the average insert size of 450 bp, we estimate on average one CT microsatellite per 116 kb DNA in the soybean genome we used for constructing the libraries. Assuming a genome size of 1.29 x 10^9 bp per haploid genome of soybean, there is approximately 1.1 x 10^4 CT microsatellite regions in the soybean genome. In our study, the frequency of CT microsatellites (one in every 116 kb of the genome) was higher than those reported earlier for other species, but similar to the frequency of CT microsatellites in Brassica. There is no reliable data on the frequency of this microsatellite in soybean since, in most of the cases, estimates were based either on the database searches or by screening libraries constructed using different combinations of restriction enzymes. The results presented here, based on screening randomly sheared libraries, which usually offer a more comprehensive evaluation of the complete genome, therefore provides the possibility of a more accurate estimate of microsatellite frequency.

The CT microsatellite is reasonably abundant in many species including soybean. Although AT microsatellites are the most abundant repeat type in plant genomes, there are inherent difficulties in their isolation and characterization. The CA microsatellite is not only less frequent but also, in many species, has been reported to be uninformative. On these bases, CT microsatellites may be used as the di-nucleotide repeat candidate for plant genome analysis.

3.2. Sequence analysis
Of the 174 clones positive for CT microsatellites 129 were sequenced. The repeat structures of the sequenced CT clones were categorized based on the method of Weber and are presented in Fig. 1. The repeat number in the CT clones varied from 5 to 44, with an average of 11 repeats per clone. The distribution of repeat numbers among clones is presented in Fig. 2. In the 129 CT-positive clones, we found 10 clones were chimeric with adapter sequences, and that the microsatellite repeat sequences of 30 (23%) clones were either at the end of, or too close to, either end of the inserts. It has been speculated that the blunt-ending process may cross ligate adapters with the microsatellites and generated chimeric clones. The presence and location of microsatellite repeats too close to, or at the either ends of inserts, are well-documented and pre-screening procedures are available to exclude these inserts from sequence analysis. In our study, we did not perform any pre-screening step as we intended to characterize CT repeat sequences in the soybean genome. In Norway spruce, even using a pre-screening step to exclude inserts with repeat sequences at either end or close to either end, it was reported that 13% of clones were useless as they were too close to or at either ends.

In rice, it was proposed that the sheared library may have been subjected towards physical breakage within extended repeat motifs, since the longest CT repeat motif was found to be (CT)_{39} compared to (CT)_{64} of an enzyme-digested library. In this study, the longest motif (CT)_{44} was shorter than that of an enzyme-digested library but larger than that of the sheared library in rice. In other enzyme-digested libraries the longest CT repeats were reported to be (CT)_{39} in Norway spruce.
(CT)$_n$ Microsatellites in Soybean Genome

3.3. Primer design

Forty-nine primers could be designed from the sequence information of 129 clones containing CT repeats. The primer sequences, PCR conditions, repeat types, and PCR product sizes, allele numbers and PIC values for the polymorphic primer sets are presented in Table 1. The approximate product size of primer sets were estimated based on the amplified product of the cultivar

Norin No. 2, from which the genomic libraries were constructed. Primers could not be designed from the rest of the clones because either the SSR motifs were too close to either of the cloning sites, or compatible primer pairs could not be designed, or some clone sequences were chimeric with adapter sequences. In this study, primers were designed from sequences flanking CT microsatellites that ranged from 8 to 33 repeats, and were subsequently analyzed for use in PCR amplification. Although microsatellites of more than 10 repeats are usually polymorphic in humans, we found that the repeat number was not an indicator of polymorphism. Furthermore, as the lower limit of repeats required for generating polymorphism has not been studied, the use of primers with less than 8 repeats for generating polymorphism can not be ruled out. Moreover, repeat structures did not seem to be a factor in variability, as variable microsatellites were composed of different types of repeat structures. Some of the SSR motifs identified were compound repeats associated with (AT)$_n$, (CA)$_n$ or (A)$_n$. A similar situation was observed in the case of AC repeats in Arabidopsis. However, unlike the Arabidopsis example, this class of SSR in soybean proved to be neither re-
calcitrant to PCR nor less polymorphic than other types.

3.4. PCR amplification

PCR amplification was performed using a single PCR profile for the amplification of all primers. Six (12%) of the primers could not amplify genomic DNA. Twenty-two (45%) were found to be polymorphic, 9 (18%) were monomorphic and the rest were found to result in non-specific amplification (bad amplification). Fourteen primers were found to generate polymorphism between the parents of a mapping population. Among the polymorphic primers, 14 were designed from sequences flanking a perfect repeat structure, 4 were from imperfect compound structures, and 4 were found to flank compound structures either with AT, TA or A. In the case of polymorphic primers, 14 were designed from sequences flank- ing a perfect repeat structure, 4 were from imperfect

In a study of CT microsatellites in rice with a pre- screening step, half of the positive clones were eliminated, 31% of the designed primers were reportedly amplified, and 40% of the total primers were polymorphic. In an analysis of a sheared library, they reported a 70% failure in obtaining clear PCR products. In our study, however, 88% of our designed primers could be used for amplification, and about 70% had clear amplifica-

tions (polymorphic+monomorphic). The rate of successful amplification from primers in this study was higher than that of an enzyme digested library of rice (73%) and that of a sheared library (30%) from the same species.

A further set of primers, designed from the clones from which initial primers could not amplify or from which non-specific products were amplified, could plausibly increase the number of polymorphic primers.

We have not observed the banding complexities identified in Norway spruce or ‘Ghost’ bands observed in the analysis of AT or CA repeats in soybean that could lead to confusion in detecting the expected PCR products. All the polymorphic primers gave a simple variable pattern and showed a co-dominant segregation of two bands when tested on the progeny of a mapping family.

The abundance, high level of polymorphism, and ease of genotyping make SSRs an excellent molecular marker system for many types of genetic analyses, including linkage mapping, germplasm surveys and phylogenetic studies. The wider use of this class of molecular markers, however, will depend on the speed and cost-effectiveness of developing SSR markers and the ability to genotype large numbers of individuals. The primers reported here successfully amplified genomic DNA using a single PCR profile which will help to expedite data collection by conducting multiplex PCR, in which two or more primers

Table 1. Primer sequence, repeat type, amplified product size, number of allele and PIC values of CT-SSR Loci in soybean.

<table>
<thead>
<tr>
<th>SSR Loci</th>
<th>Primer sequences: Forward/Reverse (5' - 3')</th>
<th>Repeat Type</th>
<th>Product size (bp)</th>
<th>No. of allele</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>138ct04</td>
<td>ACAATTTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(AG)_1</td>
<td>224</td>
<td>2</td>
<td>0.49</td>
</tr>
<tr>
<td>138ct21</td>
<td>TGGCGACATATGTCGCGGACATCGTCGGCGGATGTAATCC</td>
<td>(CT)_1</td>
<td>85</td>
<td>5</td>
<td>0.81</td>
</tr>
<tr>
<td>138ct24</td>
<td>TCATTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(CT)_1</td>
<td>188</td>
<td>3</td>
<td>0.53</td>
</tr>
<tr>
<td>138ct25</td>
<td>TAAGAGGTCTGATATGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(CT)_1(CT)_1</td>
<td>105</td>
<td>5</td>
<td>0.76</td>
</tr>
<tr>
<td>138ct27</td>
<td>TGCATTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(CT)_1</td>
<td>188</td>
<td>2</td>
<td>0.42</td>
</tr>
<tr>
<td>138ct31</td>
<td>CCCCTCTTTGAGAAGATATGACATCGTCGGCAAGAGGAG</td>
<td>(TC)_2</td>
<td>113</td>
<td>2</td>
<td>0.49</td>
</tr>
<tr>
<td>138ct32</td>
<td>TACATTTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(AG)_1</td>
<td>146</td>
<td>3</td>
<td>0.65</td>
</tr>
<tr>
<td>138ct34</td>
<td>CTGCAACATTTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(CT)_2</td>
<td>175</td>
<td>3</td>
<td>0.61</td>
</tr>
<tr>
<td>138ct39</td>
<td>CCATTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(TC)_2</td>
<td>100</td>
<td>4</td>
<td>0.42</td>
</tr>
<tr>
<td>138ct40</td>
<td>ACCATTTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(AG)_1</td>
<td>85</td>
<td>4</td>
<td>0.60</td>
</tr>
<tr>
<td>138ct41</td>
<td>TGATTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(GA)_1</td>
<td>190</td>
<td>5</td>
<td>0.75</td>
</tr>
<tr>
<td>138ct44</td>
<td>ATATATATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(CT)_1</td>
<td>90</td>
<td>3</td>
<td>0.59</td>
</tr>
<tr>
<td>138ct46</td>
<td>ATATATATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(GA)_1</td>
<td>98</td>
<td>4</td>
<td>0.64</td>
</tr>
<tr>
<td>138ct47</td>
<td>ATATATATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(GA)_1</td>
<td>95</td>
<td>4</td>
<td>0.65</td>
</tr>
<tr>
<td>138ct57</td>
<td>ACCATTTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(CT)_1(CT)_1</td>
<td>108</td>
<td>3</td>
<td>0.57</td>
</tr>
<tr>
<td>138ct58</td>
<td>CATATTTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(A)_1(GA)_2</td>
<td>135</td>
<td>7</td>
<td>0.84</td>
</tr>
<tr>
<td>138ct63</td>
<td>TATATATATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(AG)_1</td>
<td>110</td>
<td>4</td>
<td>0.56</td>
</tr>
<tr>
<td>138ct71</td>
<td>ATATATATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(CT)(TC)_3(CT)(TC)(CT)(TC)(CT)</td>
<td>214</td>
<td>4</td>
<td>0.73</td>
</tr>
<tr>
<td>510ct28</td>
<td>CACCATTTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(GA)_(2)(Ga)_1</td>
<td>200</td>
<td>3</td>
<td>0.62</td>
</tr>
<tr>
<td>510ct34</td>
<td>ATATATATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(TC)_2(CT)_1</td>
<td>145</td>
<td>3</td>
<td>0.62</td>
</tr>
<tr>
<td>910ct23</td>
<td>GACATTTATAGTGACAGCAATCGTCGGCGGATGTAATCC</td>
<td>(CT)_1(CT)_1</td>
<td>118</td>
<td>2</td>
<td>0.49</td>
</tr>
<tr>
<td>910ct38</td>
<td>CGAAGACGACACACCACTGCATGGTACAGTGCAAGGACAGTC</td>
<td>(CT)_1</td>
<td>95</td>
<td>3</td>
<td>0.62</td>
</tr>
</tbody>
</table>
1.2.3.4.5.6.7.8.9.10.11.12.13.14.15.16.17.18

Figure 3. Segregation of amplified fragments of the loci, 138ct41, among progenies of an F2 mapping family. Lanes 1 and 18, size marker φ×174/Hae III; lanes 2 and 3, parents; lanes 4 to 17 progenies, in which 8, 9 and 11 show the heterozygous state of hybrids. The separation of amplified fragments was performed by electrophoresis through a 17% polyacrylamide gel for 12 hr at 100 V.

could be used simultaneously in one reaction. If a single problem such as the position of microsatellite sequences at either end of inserts could be overcome, the efficiency of using this method could be increased twofold. We are, thus, focusing on further analysis using a relatively larger insert size (~500-900 bp) for microsatellite analysis, which may, to some extent exclude the problem associated with the position of repeat sequences in the inserts.

3.5. Informativeness of SSR loci

In order to determine the level of variability of the developed SSR loci, genomic DNAs from 16 different soybean cultivar were used as templates for PCR amplification of each locus. The PCR products were electrophoresed on 17% denaturing polyacrylamide gel with 6 M urea and 1 x TBE.

Table 1, summarizes the number of alleles and the PIC values for each of the primer sets. The number of alleles ranged from 2 to 7 with an average of 4 alleles per primer set and the PIC values varied from 0.42 to 0.84 with a mean value of 0.61. The highest number of alleles was observed among the 16 cultivars on loci 138ct58. Figure 4 is presented with the polymorphism of this SSR loci among 12 of the 16 cultivars.

In a study of seven SSRs of (AT)n and (ATT)n, as many as ten alleles were reported in a group of 26 North American soybean cultivars whose PIC values ranged from 0.52 to 0.88 with a mean of value 0.74. However, the primer pairs, we are reporting here are much more informative than the mean PIC value 0.32 of 132 RFLP probe analyzed in soybean.24

A part of the allelic information of the CT-SSRs in this study may be masked, as the evaluating varieties were randomly selected from a group of 70 varieties and the PCR products were separated on polyacrylamide/urea gels of 20 cm×16 cm×1 mm size, whereas 40 cm×20 cm×0.5 mm sequencing gels visualized by image analyzer are usually used. However, using the CT-SSRs with four or more alleles coupled with higher PIC values, a detailed genetic diversity analysis of 70 varieties using sequencing gels and image analyzer is underway in our laboratory. The findings of this analysis may represent correct allelic information of the SSRs.

3.6. Linkage analysis

Linkage analysis was performed to evaluate the distribution of developed SSR loci in the soybean genome. After preliminary assessment, 14 primer sets were found to display polymorphism in the parents of an F2 mapping population. A genetic map of 29 linkage groups was constructed in this population using 247 RFLP markers including 97 cDNA markers developed in our laboratory.30 Using the same mapping criteria of the RFLP map, the segregation of the SSR loci were analyzed among the 100 individuals of the mapping population. All of the SSR loci were positioned among 8 of the 29 linkage groups; these linkage groups are presented in Fig. 5. The linkage groups, LG-18 and LG-20 were found merged together by the integration of the loci 138ct25. Six of the 14 loci were found positioned into a single linkage group LG-25, where 2 and 3 of the loci were clustered with the interruption of a single RFLP locus. The SSR loci identified in this study were developed from random genomic fragments of soybean, and if the SSR loci were truly distributed randomly in the soybean genome, the SSRs would have more evenly distributed in the linkage groups. In a study of integrating 40 SSR loci, a cluster of five loci was reported.31 In another study of
mapping 500 SSR loci, portions of many linkage groups were found clustered with SSR loci. Clustering of SSR loci with the interruption of groups of RFLP loci was also observed when these SSR loci were mapped in other mapping populations. Considering the developmental basis of RFLP and SSR markers, they speculated that the clustering could be associated with RFLP loci. In our study of only 14 SSR loci, although it is difficult to conclude the reason for clustering, it is more likely that the clustering events are associated with SSR loci considering the developmental basis of our markers. However, we are expecting to develop a consensus map of soybean genome in the near future, when the developed markers would have a definite role in mapping the genome and the developed procedures will help in identifying more markers.

Acknowledgments: We are thankful to Dr. John Forster of La Trobe University, Australia for his technical advice regarding microsatellite analysis. This research was supported by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid of the Ministry of Education, Science, Sports and Culture of Japan, and funds from the Ministry of Agriculture, Forestry and Fisheries of Japan.

References
6. Bell, C. J. and Ecker, J. R. 1994, Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*, *Ge-
nomics, 19, 137–144.


