An Informative Linkage Map of Soybean Reveals QTLs for Flowering Time, Leaflet Morphology and Regions of Segregation Distortion

Naoki YAMANAKA,1,* Seishi NINOMIYA,2 Masako HOSHI,1 Yasutaka TSUBOKURA,1 Masahiro YANO,3 Yoshiaki NAGAMURA,3 Takuji SASAKI,3 and Kyuya HARADA1

Graduate School of Science and Technology, Chiba University, 648 Matsudo, Matsudo City, Chiba 271-8510, Japan,1 National Institute of Agro-Environmental Sciences, 3-1-1 Kannondai, Tsukuba City, Ibaraki 305-8666, Japan,2 and National Institute of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba City, Ibaraki 305-8602, Japan3

(Received 26 February 2001; revised 30 March 2001)

Abstract

A genetic linkage map covering a large region of the genome with informative markers is essential for plant genome analysis, including identification of quantitative trait loci (QTLs), map-based cloning, and construction of a physical map. We constructed a soybean genetic linkage map using 190 F2 plants derived from a single cross between the soybean varieties Misuzudaizu and Moshidou Gong 503, based on restriction-fragment-length polymorphisms (RFLPs) and simple-sequence-repeat polymorphisms (SSRPs). This linkage map has 503 markers, including 189 RFLP markers derived from expressed sequence tag (EST) clones, and consists of 20 major linkage groups that may correspond to the 20 pairs of soybean chromosomes, covering 2908.7 cM of the soybean genome in the Kosambi function. Using this linkage map, we identified 4 QTLs—FT1, FT2, FT3, and FT4—for flowering time, the QTLs for the 5 largest principal components determining leaflet shape, 6 QTLs for single leaflet area, and 18 regions of segregation distortion. All 503 analyzed markers identified were located on the map, and almost all phenotypic variations in flowering time were explained by the detected QTLs. These results indicate that this map covers a large region of the soybean genome.

Key words: Glycine max (L.) merrill; linkage map; RFLP; QTL; segregation distortion

1. Introduction

Various DNA markers and molecular linkage maps have been developed in a large number of plants and make effective tools for phylogenetic analysis, marker-assisted selection (MAS) in breeding programs, genetic analysis of quantitative traits, cloning of the genes responsible for various phenotypes (map-based cloning), and construction of a physical map. Soybean (Glycine max (L.) Merrill) is one of the major crops cultivated around the world. Its capacity for protein and oil production makes it a significant contributor to human nutrition, and its characteristic symbiosis with root bacteroids makes it a very important crop in research. Many reports of the construction of soybean genetic linkage maps have been published.1–7 However, a functional linkage map with cDNA markers covering a large region of the soybean genome has thus far not been constructed because of soybean’s large genome size (1.12×10⁹ bp; 1.81×10⁹ bp; 1.29×10⁹ bp) and its ancient polyploid genome origin.11 Cregan et al. integrated a soybean linkage map into 20 homologous linkage groups by using simple-sequence-repeat (SSR) markers based on their correspondence with some known linkage maps.12 SSR markers are single-locus markers with multiple alleles; it is easy to assay the genotype of each locus by using the polymerase chain reaction (PCR). Thus, through the design of highly specific primers, these markers came to be the tools of choice for genome analysis of plants that have many duplicated arrangements. A number of SSR markers of soybean have been developed and have proven to be of practical utility.12,13

In our previous study,14 we described QTL analysis for flowering time in soybean by using published soybean RFLP markers4 and the soybean RFLP markers originally developed.14 Almost all phenotypic variations in flowering time of the tested population could be ex-
explained by the QTLs based on these markers. Thus, we concluded that the linkage map constructed for this analysis covered a comparatively large region of the soybean genome. However, the number of linkage groups (33) was greater than the number of chromosomes (20), and the map included 4 unlinked markers.

Therefore, the first objective of the present study was to extend the construction of the linkage map to cover a larger region of the soybean genome by using SSR markers and RFLP markers that include cDNA markers. The second objective was to incorporate new information such as loci for leaflet shape, single leaflet area loci, regions of segregation distortion, mapped cDNA clone sequencing, and reanalysis of flowering time loci, into the soybean map. The third objective was to document the validity of cDNA markers in soybean genome analysis.

Leaflet shape is a quantitative, highly complex morphological trait in soybean; it is thought that many genes are involved in its control. Furuta et al.\(^\text{15}\) reported that the contour shape of soybean leaflets could be represented by elliptic Fourier series expansion, and that principal component scores about standardized elliptic Fourier coefficients give quantitative measures to evaluate soybean leaflet shape. In the present study, we analyzed this trait by the same method and identified the loci for each principal component that makes a large contribution to leaflet shape. To our knowledge, no similar QTL analysis of soybean leaflet shape has been previously reported. We analyzed another quantitative morphological trait of leaves, single leaflet area, and identified some loci associated with this quantitative trait.

2. Materials and Methods

2.1. Plant materials

We used 2 soybean (\textit{Glycine max.}) varieties, Misuzudaizu and Moshidou Gong 503 as ovule and pollen parents, respectively. We used 190 \(F_2\) plants derived from this cross as a segregating population for linkage map construction. We used variety Norin No. 2 as the source of cDNA clones. The conditions for cultivating each variety and for cDNA library construction have been described previously.\(^\text{14}\) The rice (\textit{Oryza sativa L.}) varieties Nipponbare (japonica) and Kasalath (indica) were used for RFLP analysis. These 2 varieties are parents of the mapping population in the Rice Genome Research Program (RGP) in Japan.\(^\text{16,17}\)

2.2. RFLP analysis

RFLP analysis of soybean parental varieties, the \(F_2\) plants, and the two rice varieties were performed by Southern hybridization, as in our previous report.\(^\text{14}\) We used four kinds of RFLP probes—soybean RFLP markers\(^\text{4}\) derived from soybean genomic DNA,\(^\text{18}\) soybean cDNA clones derived from green leaves,\(^\text{14}\) cloned soybean genes,\(^\text{14}\) and a wheat clone (45S rDNA)—for RFLP analysis.

2.3. SSR markers

We used three kinds of soybean SSR markers for mapping, developed at the USDA,\(^\text{12}\) DuPont Corporation, and Chiba University.\(^\text{13}\) For the markers developed at the USDA, PCR was performed under the following conditions: 15 min at 95°C, then 33 cycles of 1 min at 92°C, 1 min at 46°C, and 1 min at 68°C. PCR products were electrophoresed in 10% polyacrylamide gel, then polymorphisms were detected with ethidium bromide staining. The SSR marker mapping protocol used for the other SSRs is described in Hossain et al.\(^\text{13}\)

2.4. Linkage map construction

For the RFLP analysis of the segregating population, 190 \(F_2\) plants were used for mapping. For the SSR marker analysis, an arbitrary 120 of the 190 \(F_2\) plants were used. Mapping for 4 qualitative traits was described in our previous study.\(^\text{14}\) One qualitative trait that corresponds to stem growth habit was added in this study (\(D_t1\): indeterminate or determinate stem). The linkage analysis of the above markers was performed with the program MAPMAKER/EXP. ver. 3.0\(^\text{19}\) using the Kosambi function. The linkage criteria of the markers were an LOD score of > 3.0 and a maximum distance of 37.2 cM.

In the case where two markers had a distance > 37.2 cM, the linkage between these two markers is indicated by dotted lines in Fig. 1.

2.5. QTL analyses for flowering time, leaflet shape, and single leaflet area

We used MAPMAKER/QTL ver. 1.\(^\text{20,21}\) to analyze the QTLs for flowering time, the five largest principal components of leaflet shape, and single leaflet area, their locations and effects. One QTL for flowering time, \(FT1\), has a very large effect and was described previously.\(^\text{14}\) Therefore, the other QTLs for flowering time were analyzed by first excluding the effect of \(FT1\) based on the multiple-QTL model.

Fifteen central leaflets of fully expanded compound leaves from the parental varieties and 1 to 6 (\(Av. = 4.36\)) central leaflets of each of 153 \(F_2\) plant were used for principal component analysis. This analysis is based on the elliptic Fourier descriptor according to the method of Furuta et al.\(^\text{15}\) The values of the five largest components were multiplied by 1000 and then used for QTL analysis. The same samples were also used for QTL analysis of single leaflet area. All QTLs for leaflet shape and single leaflet area were identified based on the single QTL model. All the effects of each QTL were estimated by comparing those of the Moshidou Gong 503 genotype with those of the Misuzudaizu genotype.
Figure 1. Soybean genetic linkage map constructed in this study. The name of each linkage group is indicated on the top of the bar. Distances between markers are shown on the left of each linkage group by the Kosambi function. The total length of each linkage group is shown on the bottom of the group. In the name of the linkage group, '+' means the integration of two or three previous linkage groups. The label in parentheses shown beside the name of the linkage groups (e.g., a1, b2) indicates correspondence to the linkage map of Cregan et al. Marker types and QTLs are indicated in the figure.
2.6. Sequencing analysis and homology search

The cDNA fragments inserted in the plasmid vector, pBluescript SK II’ (Stratagene) were amplified by PCR by using ‘M13 primer M4’ and ‘M13 primer RV’ (TaKaRa). After the PCR products were purified, the sequencing reaction was performed using d-Rhodamine cycle sequencing kits (Perkin-Elmer) and ‘M13 primer M3’ (TaKaRa). The sequencing reaction products were purified in Centri-Sep spin columns (Princeton Separations), and 5’ partial sequences were detected using an ABI 310 genetic analyzer (Perkin-Elmer). Similarities with known
nucleotide sequences were analyzed with the BLASTN program (http://www.blast.genome.ad.jp).

2.7. Regions of segregation distortion

We surveyed whether the genotype frequencies of all markers on the linkage map correspond to each theoretical frequency of dominant and co-dominant markers based on the chi-square test. When the adaptable probabilities of the markers were < 0.05, the markers were considered to have segregation distortion. Next we classified the segregation distortion markers into distorted pattern classes, in which 1 parental genotype frequency is greater than the theoretical frequency, or the heterozygote genotype frequency is greater, or both parental genotype frequencies are greater. The linkage regions of the same distortion pattern classes were identified as regions of segregation distortion.

3. Results

3.1. RFLP analysis

The results of RFLP analysis of Misuzudaizu and Moshidou Gong 503 are shown in Table 1. When known soybean RFLP markers derived from soybean genomic DNA were used as the probes, 173 of 248 (69.8%) showed RFLPs for at least 1 of 8 restriction enzymes that can recognize 6-base sequences. When cDNA clones were employed, 348 of 542 (64.2%) showed RFLPs. Therefore, 521 of 790 (65.9%) probes used in this study showed RFLPs. No large difference in the level of polymorphism was revealed between the paired parental varieties by using probes from different sources. Table 2 shows the polymorphism frequencies of parental varieties according to the restriction enzymes used for RFLP analysis. Few probes showed RFLPs when we used Apa I and Kpn I, which yielded relatively long DNA fragments. This result indicates that there are few target sequences for these enzymes in the soybean genome. On the other hand, many probes showed RFLPs when we used Eco RV, HindIII, Dra I, and EcoRI, which yield relatively short DNA fragments, indicating that there is a larger number of target sequences of these enzymes in the soybean genome. This tendency was observed commonly in three kinds of probes.

We also attempted to hybridize two kinds of probes with DNAs derived from the monocot plant rice, which is a relatively distant species. When we used 165 random genomic DNA fragments as probes, only a few probes (10.9%) hybridized with the rice DNAs under the same conditions. However, when we used 197 cDNA fragments as probes, hybridization signals could be detected with a relatively large numbers of probes (42.9%).

### Table 1. RFLPs detected between the parents using 2 kinds of probes.

| Probe Type          | No. of analyzed probes | No. of probes showing RFLPs |
|---------------------|------------------------|-----------------------------|
| Known RFLP markers  | 248                    | 173 (69.8%)                 |
| cDNA clones         | 542                    | 348 (64.2%)                 |

- a) RFLP was detected by at least one of eight restriction enzymes.
- b) Including a lectin gene, 5 cycline cDNAs from root suspension culture cells, and 12 nodule-specific cDNA clones.

| Probe Type          | Percentage of probes showing RFLPs |
|---------------------|-----------------------------------|
| Probe              | BamHI | BglII | EcoRV | HindIII | ApaI | DraI | EcoRI | KpnI |
| Known RFLP markers  | 22.2  | 27.0  | 39.1  | 31.5    | 5.2  | 21.8 | 35.1  | 12.9 |
| cDNA clones        | 17.6  | 28.9  | 29.7  | 27.3    | 5.1  | 22.7 | 31.6  | 13.7 |
| Cloned soybean genes | 28.1  | 46.9  | 56.2  | 53.1    | 0.0  | 50.0 | 18.8  | 21.9 |
| Total              | 20.3  | 29.1  | 35.6  | 30.8    | 4.9  | 23.9 | 32.5  | 13.8 |

3.2. Linkage map

We have now added 212 new markers to our previous map and reduced the number of linkage groups from 33 to 20 major and 1 minor groups (Fig. 1). The linkage map in Fig. 1 contains the loci for five phenotypic traits and one random amplified polymorphic DNA (RAPD) marker, as well as for 401 RFLP markers and 96 SSR markers (Table 3). The map contains no unlinked markers and covers 2908.7 cM of the soybean genome in the Kosambi function. The average distance between all types of markers is 6.0 cM (Table 3). The average length of the major linkage groups was calculated at 144.9 cM. The average number of markers in all the major linkage groups was 24.9. The longest linkage group was LG 12 (202.2 cM), and the groups containing the greatest number of markers were LG 9 + 16 and LG 13 + 29, with both having 41 markers. The smallest major linkage group was LG 7 (102.4 cM) and the group with the smallest number of markers (14) was LG 10 + 26. The highest and lowest densities of markers were found in LG 9 + 16 (Av. Distance = 3.8 cM) and LG 12 (Av. Distance = 11.1 cM), respectively. The correspondence between all the linkage groups in this map and the linkage groups of the soybean genetic map of Cregan et al. is clearly evident (shown in Fig. 1).
3.3. Mapping of QTLs for flowering time, leaflet shape, and single leaflet area

Using the progressive linkage map constructed in this study, we reanalyzed the three main QTLs $FT1$, $FT2$ and $FT3$ for flowering time (distribution is shown in Fig. 2) from our previous study.\(^{14}\) $FT1$ had $-9.4$ days as an additive effect and $+3.8$ days as a dominance effect, and was located on LG 3 (the same location as in our previous study; Table 4). However, a new SSR marker, AG36, had been mapped to the same location as the marker nearest to $FT1$ (Table 4, Fig. 1). The second most effective QTL, $FT2$, was detected on a new, tightly linked SSR marker, 138GA26 (Table 4, Fig. 1). The reanalyzed QTL $FT2$ had $+3.6$ days as an additive effect and $-0.2$ days as a dominance effect. The variance that accounted for the total phenotypic variation with $FT1$ and $FT2$ was 80.6%. For the QTL $FT3$, a new, tightly linked SSR marker, Satt373, was detected (Table 4, Fig. 1). After reanalysis, $FT3$ had $-2.1$ days as an additive effect and $+1.7$ days as a dominance effect. We focused on the QTLs with minute effects for reanalysis. As a result, we detected 1 additional QTL, $FT4$, located on LG 2-1 (Fig. 1). Although this QTL had $-1.7$ days as an additive effect, little dominance effect was observed (Table 4). The variance that accounted for total phenotypic variance by $FT1$ and $FT4$ was 72.9% (Table 4).

We could also detect large numbers of QTLs associated with leaflet morphology. Regarding the five largest principal components of leaflet shape (distributions are shown in Fig. 2), we detected four QTLs for the 1st principal component ($LS1$ to $LS4$), one QTL for the 2nd principal component ($LS5$), two QTLs for the 4th principal component ($LS6$ and $LS7$), and one QTL for the 5th principal component ($LS8$) (Table 5, Fig. 1). No QTL was detected for the 3rd principal component. Six loci ($LA1$ to $LA6$) for single leaflet area (distribution is shown in Fig. 2) were localized on the linkage map (Table 6, Fig. 1). Five of the six loci decreased the single leaflet area in Moshidou Gong 503 genotypes; the other loci increased the area. The most effective QTL ($LA1$) among the six loci was located at the same position as $LS1$ and $LS4$.

3.4. Sequencing analysis

We determined 5’ partial sequences of 113 among 167 cDNA clones (corresponding to 189 markers) used for construction of the linkage map. The BLASTN search yielded 94 clones (83.2%) showing significant similarities with previously known nucleotide sequences. Nineteen clones (16.8%) had little or no similarity with known sequences. Among the 94 clones, 78 (83.0%) matched those derived from soybean, and the remaining 16 (17.0%) hit nucleotide sequences derived from other species. Among the 78 soybean clones, those whose putative functions were determined are listed in Table 7.

3.5. Regions of segregation distortion

For 32 of 503 markers, the actual segregation ratios were significantly different from the theoretical ones (Fig. 3). These markers were distributed in 18 regions of ten linkage groups. Among the 18 segregation distortion regions, 3 regions showed high frequencies of heterozygosity and 2 showed high genotypic frequencies for both parents. There were 8 regions where only Misuzudaizu homozygote frequencies were high and three regions where only Moshidou Gong 503 genotypic homozygote frequencies were high. There was only 1 region each where Moshidou Gong 503 or Misuzudaizu homozygous and heterozygous dominance frequencies were found.

4. Discussion

DNA markers with high-frequency polymorphisms are very useful. Generally, the DNA regions that code for genes have lower-frequency polymorphisms than those not coding for genes. In this study, however, cDNA clones used as probes gave almost the same frequencies of polymorphisms in RFLP analysis as random genomic DNA clones (Table 1). Therefore, we consider that: 1) RFLPs of non-coding and flanking regions on the genes were detectable by cDNA probes, and 2) because the soybean genome is highly duplicated, the polymorphisms of repeated genes or pseudogene-like DNA regions derived from genes, as well as genes themselves, were also detected at the same time by using cDNA probes. Thus,
Figure 2. The frequency distributions of seven quantitative traits analyzed in this study. The proportions of the five largest principal components were determined from the parental varieties and 153 F2 plants. The phenotypes of parental varieties are shown as by Mi, Misuzudaizu; Mo, Moshidou Gong 503. The 1st, the 2nd, the 3rd and 4th, and the 5th principal components represent the aspect ratio of the leaflet shape, the location of the centroid of leaflets along the midrib, the straightness or distortion of leaflet shape, and the roundness of leaflet shape, respectively.
**Table 4.** QTLs of soybean flowering time mapped in this study.

| QTL | Nearest marker | Linkage group | LOD score | Variance (%) explained | Additive effect (days) | Dominance effect (days) |
|-----|---------------|---------------|-----------|------------------------|-----------------------|------------------------|
| FT1 | AG360         | 3             | 48.79     | 69.7                   | -9.4                  | +3.8                   |
| FT2 | 13G92A26      | 15+25         | 63.33     | 80.6                   | +3.6                  | -0.2                   |
| FT3 | Sat373        | 9+16          | 54.97     | 74.7                   | -2.1                  | +1.7                   |
| FT4 | GM0201        | 2+1           | 51.86     | 72.9                   | -1.7                  | -0.2                   |

a) Marker with highest LOD score.
b) For FT1, the threshold LOD score is 2.0. For FT2, FT3 and FT4, the threshold is 50.79.
c) For the total F2 phenotype. For FT2, FT3 and FT4, the values are obtained in combination with the effect of FT1.
d) Co-segregated with GM133, GM169 and GmN93.

**Table 5.** QTLs of soybean leaflet shape mapped in this study.

| Component | QTL | Nearest marker | Linkage group | LOD score | Variance (%) explained | Additive effect (X10^3) | Dominance effect (X10^3) |
|-----------|-----|---------------|---------------|-----------|------------------------|-------------------------|-------------------------|
| 1         | LSI-1| A489         | 9+16          | 6.84      | 18.6                   | -49.46                  | -33.83                  |
| LSI-2     | GM339 | 3             | 5.25          | 14.6      | -53.16                 | -2.18                   |                         |
| LSI-3     | GM354 | 1-1           | 2.65          | 7.7       | -15.17                 | +37.30                  |                         |
| LSI-4     | GM018 | 10+26         | 2.45          | 7.1       | +28.47                 | +4.32                   |                         |
| 2         | LSI-2| K385         | 9+16          | 3.32      | 9.6                    | -6.18                   | -3.186                  |
| 4         | LSI-4| A489         | 9+16          | 5.66      | 15.7                   | -5.34                   | -0.992                  |
| LSI-4     | A590 | 7             | 2.60          | 7.5       | -3.030                 | +0.155                  |                         |

a) Marker with highest LOD score.
b) The threshold LOD score of each component is 2.0.
c) For the total F2 phenotype variation of each component.
d) Co-segregated with AG36, GM169 and GmN93.
e) Co-segregated with B170.
f) Co-segregated with K384.

The value of a linkage map depends on four main factors: 1) precision of the linkage map, 2) coverage of the map over the genome, 3) usefulness of the markers mapped on that linkage map, and 4) whether the map contains important genetic information. The F2 population used for construction of the linkage map in the present study consisted of 190 plants derived from a single cross between two soybean varieties. Because heterozygote genotypes of the co-dominant marker can be distinguished in the F2 population, a large amount of information can be obtained by analyzing one population. In the F2 population the influence of genotype selection in raising the population is smaller than in other kinds of populations. That is, in a double haploid (DH) population, greater genotype selection may apply to genes related to regeneration and to regions linked to those genes. Similarly, in recombinant inbred lines, the effects of selection may accumulate for genes associated with the gametophyte or sterility and for regions linked to these genes. Because many markers located on this linkage map were co-dominant and large numbers of plants were used for mapping, we consider this linkage map to be highly precise. The number of main linkage groups developed in this study is equal to the 20 soybean chromosome pairs, on which all the markers analyzed were successfully mapped. Almost all phenotypic varieties in flowering time were explained by four QTLs. Therefore, it can be considered that the present linkage map covers a large region of the soybean genome.
Table 7. Putative functions of cDNA clones mapped on the linkage map established in this study.

| Clone     | Putative function                                                                 |
|-----------|----------------------------------------------------------------------------------|
| GM004     | Malate dehydrogenase                                                             |
| GM005     | Photosystem I reaction centre subunit V                                          |
| GM010     | Aquaporin 1                                                                       |
| GM012, GM026 | Protein involved in starch metabolism                                           |
| GM016     | Cysteine protease                                                                 |
| GM028     | Thaumatin-like protein precursor MDTL1                                             |
| GM035     | Aminomethyltransferase                                                            |
| GM036, GM161 | Fructose-bisphosphate aldolase                                                     |
| GM041     | Beta-tubulin (tubB3)                                                              |
| GM043     | PTO kinase interactor 1                                                            |
| GM047, GM102 | Catalase (cat4)                                                                   |
| GM048     | Hypothetical 55.0 kD protein                                                       |
| GM049     | Oxygen-evolving enhancer protein 1                                                |
| GM053     | Glutamine synthetase                                                              |
| GM055     | *Glycine max* isoprenylated protein                                               |
| GM060     | Pyrophosphate-dependent phosphofructokinase beta subunit                         |
| GM064     | Thiol protease isoform A                                                           |
| GM072     | *Glycine max* seed maturation protein                                             |
| GM073     | Proteasome 27 kD subunit                                                          |
| GM082     | Soybean phytoene desaturase                                                        |
| GM085     | Ferredoxin--NADP reductase                                                        |
| GM095     | Hydroperoxide lyase                                                                |
| GM096, GM099 | Thioredoxin                                                                      |
| GM106     | Calreticulin                                                                      |
| GM112     | Carbonic anhydrase                                                                |
| GM117     | *Glycine max* brassinosteroid-regulated protein                                   |
| GM118     | Beta-amyrase                                                                      |
| GM131     | Actin depolymerizing factor 1                                                      |
| GM134, GM135 | Aspartic protease                                                                |
| GM154     | Hypothetical protein F19D11.10                                                    |
| GM159     | 40S ribosomal protein S12                                                          |
| GM162     | PHI-1-like protein                                                                |
| GM165     | GA protein                                                                        |
| GM178     | Cytosolic aldehyde dehydrogenase                                                   |
| GM195     | Profucosidase                                                                     |
| GM210     | S-adenosyl-l-methionine decarboxylase proenzyme                                   |
| GM214     | DNA binding protein S1FA                                                           |
| GM221     | Sugar transport protein                                                            |
| GM227     | Chloroplast 30S ribosomal protein S10                                              |
| GM249     | Inorganic pyrophosphatase                                                          |
| GM251     | Ribulose bisphosphate carboxylase small chain 1                                   |
| GM256, GM284 | 3,4-dihydroxy-2-butanone kinase                                                   |
| GM267     | *Glycine max* FAD2-2 microsomal omega-6 desaturase                                |
| GM268     | Glyceraldehyde 3-phosphate dehydrogenase                                          |
| GM270     | *Arabidopsis thaliana* RNA binding protein                                        |
| GM282     | Hypothetical protein T26J13.1                                                     |
| GM288     | Ferredoxin-dependent glutamate synthase (glu)                                      |
| GMS016    | Hypothetical 27.0 kD protein                                                       |
| GMS039    | 60S ribosomal protein L34                                                          |
| GMS082    | Serine hydroxy methyltransferase                                                   |
| GMS087    | Ribosomal protein S26                                                              |
| GMS091    | Hypothetical 43.5 kD protein                                                       |
| GMS114    | Hypothetical 47.3 kD protein                                                       |
| GMS128    | 40S ribosomal Protein S5                                                           |
| GMS211    | Ribosomal protein L2                                                               |
In our previous study, we suggested that \( FT1 \) corresponds to the flowering and maturity locus \( E1 \) because of the linkage between \( E1 \) and the pubescence color locus \( T \). In the present study, we strengthened our evidence that \( FT1 \) corresponds to \( E1 \), and obtained further information on the relationship between QTLs and the \( E \) loci. \( E1 \) is mapped to the position of 4 cM from \( T \).
and 33 cM from the \( \beta \)-amylase variant locus, \( \text{Sp} \). The cDNA clone, GM118, has an identical partial sequence to the \( \text{Glycine max} \) \( \beta \)-amylase mRNA sequence (Table 7). The nearest marker for \( FT1 \), AG36, is mapped at a distance of 26.3 cM from GM118 and 2.9 cM from \( T \) (Fig. 1). Cregan et al.\(^{12} \) integrated some molecular linkage maps and a classical genetic map, and indicated that \( E2 \) belongs to linkage group O, which corresponds to LG 15+25 in this study. Cober and Voldeng\(^{24} \) reported that \( E3 \) is mapped at a distance of 27.32 ± 3.23 cM from the growth habit locus \( D t1 \). The nearest marker for \( FT3 \), Satt373, is mapped at a distance of 37.4 cM from \( D t1 \). Therefore \( FT1 \) and \( FT3 \) are predicted to correspond to \( E1 \) and \( E3 \), respectively. Also, there is a possibility that \( FT2 \) may correspond to \( E2 \). The correspondence between \( FT4 \) and the other \( E \) loci has not yet been clarified because of a lack of information on the other \( E \) loci.

The nearest markers to \( FT1 \), \( FT2 \), \( FT3 \), and \( FT4 \) were AG36, 138GA26, Satt373, and GM021, respectively. However, the interval mapping result of the QTLs indicates that each putative location corresponds to the intervals from \( T \) to AG36, Satt331 to 138GA26, K385 to Satt373, and A520 to GM021. Therefore, further precision in QTL localization for flowering time will be possible by mapping DNA markers in these regions.

The 2-dimensional expansion of the plant leaf is regulated genetically. For example, in \( \text{Arabidopsis} \), the \( \text{ROTUNDIFOLIA} \) gene is connected with the polar elongation of leaf cells; as a result, leaf blade length and leaf petiole length are altered.\(^{25} \) Furuta et al.\(^{15} \) described that the leaflet shape of soybean is almost completely determined by five main components (whose cumulative contribution is > 96\%), and the 1\(^{\text{st}} \) principal component contributed the most to this trait, (in this study, the proportion of the first 1\(^{\text{st}} \) principal component was 0.87; Fig. 2). Four indications support the fact that the analyzed QTLs are the primary determinants of leaflet shape. First, the plant leaf shape is regulated genetically. Second, the contribution of the 1\(^{\text{st}} \) principal component is in all cases exceedingly high. Third, four QTLs of the 1\(^{\text{st}} \) principal component were detected on linkage map covering a large region of the soybean genome. Last, \( LS1-1 \), \( LS1-2 \), \( LS1-3 \), and \( LS1-4 \) had variance-explained values of 18.6\%, 14.6\%, 7.7\%, and 7.1\%, respectively, for total phenotypic variation of this principal component. Thus, it is clear that these QTLs are the main genetic factors that determine the leaflet shape of soybean. The leaflet morphological trait \( Ln \) is mapped on linkage group 4 in the classical genetic map.\(^{23} \) The classical linkage group 4 corresponds to LG 23+24 in our map, but no QTL was located on this linkage group.

The most effective QTL for single leaflet area \( LA \) \( I \) was mapped to the same position as \( LS \) \( I-1 \) (Fig. 1). \( LS \) \( I-1 \) is the most dominant QTL among the four QTLs of the 1\(^{\text{st}} \) principal component (Table 5), which shows the highest proportion among all components (Fig. 2). Thus, the gene that determine both leaflet shape and single leaflet area might exist around this region. The other five QTLs did not match to loci for leaflet shape.

We detected some segregation distortion regions on the linkage map constructed in this study. The reasons for segregation distortion include the possibility of mistakes in genotyping because of band duplications and selection in the populations. Xu et al.\(^{26} \) described segregation distortion in variable populations in rice. In the \( F_2 \) population, many segregation distortion regions exist at gametophyte and sterility loci. The same results were described in another study.\(^{27} \) The same phenomenon might have occurred at the segregation distortion regions in the present study. The segregation distortion in LG 7 is the most remarkable among all the segregation distortion regions we detected (Fig. 3). Fourteen of 20 markers in this linkage group showed segregation distortion. Among four segregation distortion regions, three regions, excluding the marker 510CT28 (the only marker included in this region), showed segregation distortion in favor of the Misuzudaizu homozygote genotype. The markers with the highest chi-square values for segregation ratios in each of the three segregation distortion regions were AG45, B166 and GM085a. These three markers might be linked tightly to the factors that cause segregation distortion in these regions.

We succeeded in constructing an informative soybean genetic linkage map covering a large region of the soybean genome and incorporating important information such as QTLs and segregation distortion regions. We also demonstrated the validity of RFLP markers derived from cDNA clones. However, this map has some marker-rare regions or gaps and the marker density is not necessarily sufficient. Thus, we need to map more markers, and to convert the EST markers into cleaved amplified polymorphic sequence markers or other PCR-based markers to analyze their genotype more easily.

**Acknowledgements:** We would like to thank the members of the Rice Genome Research Program, NIAR/STAFF, for their technical assistance, advice, and encouragement. This study was partly supported by a Grant-in-Aid (No. 9356001) from the Ministry of Education, Culture, Sports, Science and Technology of Japan; by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan (Comprehensive Studies on Soybean Improvement, Production and Utilization in South America, Rice Genome Project DM-2109); by grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences; and by a Grant-in-Aid (No. 06413) for Research Fellows of the Japan Society for the Promotion of Science for Young Scientists.
References

1. Keim, P., Diers, B. W., Olson, T. C. et al. 1990, RFLP mapping in soybean: association between marker loci and variation in quantitative traits, *Genetics*, 126, 735–742.

2. Rafalski, A. and Tingey, S. 1993, RFLP map of soybean (*Glycine max*) 2N = 40. In: O’Brien, S. J. (ed.) Genetic Maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, New York, pp. 6.149–6.156.

3. Lark, K. G., Weisemann, J. M., Matthews, B. F. et al. 1993, A genetic map of soybean (*Glycine max*) using an intraspecific cross of two cultivars: ‘Minosy’ and ‘Noir 1’, *Theor. Appl. Genet.*, 86, 901–906.

4. Shoemaker, R. C. and Specht, J. E. 1995, Integration of soybean molecular and classical genetic linkage groups, *Crop Sci.*, 35, 436–446.

5. Mansur, L. M., Orf, J. H., and Chase, K. et al. 1996, Genetic mapping of agronomic traits using recombinant inbred lines of soybean, *Crop Sci.*, 36, 1323–1336.

6. Ferreira, A. R., Foutz, K. R., and Keim, P. 2000, Soybean genetic map of RAPD markers assigned to an existing scaffold RFLP map, *J. Heredity*, 91, 392–396.

7. Keim, P., Schupp, J. M., Travis, S. E. et al. 1997, A high-density soybean genetic map based on AFLP markers, *Crop Sci.*, 37, 537–543.

8. Arumuganathan, K. and Earle, E. D. 1991, Nuclear DNA content of some important plant species, *Plant Molecular Biology Reporter*, 9, 208–218.

9. Goldberg, R. B. 1978, DNA sequence organization in the soybean plant, *Biochem. Genet.*, 16, 45–68.

10. Gjaltema, J. C., Hepsburn, A. G., and Key, J. L. 1979, Sequence organization of the soybean genome, *Biochem. Biophys. Acta*, 561, 167–183.

11. Shoemaker, R. C., Polzin, K., Labate, J. et al. 1996, Genome duplication in soybean (*Glycine subgenus soja*), *Genetics*, 144, 329–338.

12. Cregan, P. B., Jarvik, T., Bush, A. L. et al. 1999, An integrated genetic linkage map of the soybean, *Crop Sci.*, 39, 1464–1490.

13. Hossain, K. G., Kawai, H., Hayashi, M. et al. 2000, Characterization and identification of (CT)n microsatellites in soybean using sheared genomic libraries, *DNA Res.*, 7, 103–110.

14. Yamanaka, N., Nagamura, Y., Tsubokura, Y. et al. 2000, Quantitative trait locus analysis of flowering time in soybean using a RFLP linkage map, *Breed. Sci.*, 50, 109–115.

15. Furuta, N., Ninomiya, S., Takahashi, N. et al. 1995, Quantitative evaluation of soybean (*Glycine max* L. Merr.) leaflet shape by principal component scores based on elliptic Fourier descriptor, *Breed. Sci.*, 45, 315–320.

16. Kurata, N., Nagamura, Y., Yamamoto, K. et al. 1994, A 300 kilobase interval genetic map of rice including 883 expressed sequences, *Nature Genet.*, 8, 365–372.

17. Harushima, Y., Yano, M., Shomura, A. et al. 1998, A high-density rice genetic linkage map with 2275 markers using a single F2 population, *Genetics*, 148, 479–494.

18. Keim, P. B. and Shoemaker, R. C. 1988, Construction of a random recombinant DNA library that is primarily single copy sequence, *Soybean Genet. Newsl.*, 15, 147–148.

19. Lander, E. S., Green, P., Abrahamson, J. et al. 1987, MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations, *Genetics*, 115, 174–181.

20. Lander, E. S. and Botstein, D. 1989, Mapping mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics*, 121, 185–199.

21. Lincoln, S., Daly, M., and Lander, E. 1993, Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1: a tutorial and reference manual, 2nd edn. Whitehead Institute Technical Report.

22. Altschul, S. F., Madden, T. L., Schaffer, A. A. et al. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 25, 3389–3402.

23. Palmer, R. G. and Hedges, B. R. 1993, Linkage map of soybean (*Glycine max* L. Merr.) (2N = 40). In: O’Brien, S. J. (ed.) Genetic Maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, New York, pp. 6.139–6.148.

24. Cober, E. R. and Voldeng, H. D. 1996, *E3* and *Dt1* linkage, *Soybean Genet. Newsl.*, 23, 56–57.

25. Kim, G. T., Tsukaya, H., Saito, Y. et al. 1999, Changes in the shapes of leaves and flowers upon overexpression of cytochrome P450 in Arabidopsis, *Proc. Natl. Acad. Sci. USA*, 96, 9433–9437.

26. Xu Y., Zhu, L., Xiao, J. et al. 1997, Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.), *Mol. Gen. Genet.*, 253, 535–545.

27. Harushima, Y., Kurata, N., Yano, M. et al. 1996, Detection of segregation distortion regions in an *indica-japonica* rice cross using a high-resolution molecular map, *Theor. Appl. Genet.*, 92, 145–150.