DNA Research **16**, 213–225, (2009) doi:10.1093/dnares/dsp010

High-density Integrated Linkage Map Based on SSR Markers in Soybean

Tae-Young Hwang[1,2], Takashi Sayama[1], Masakazu Takahashi[3], Yoshitake Takada[4,†], Yumi Nakamoto[1], Hideyuki Funatsuki[1], Hiroshi Hisano[5], Shigemi Sasamoto[5], Shusei Sato[5], Satoshi Tabata[5], Izumi Kono[6], Masako Hoshi[7], Masayoshi Hanawa[7], Chizuru Yano[7], Zhengjun Xia[8], Kyuya Harada[8], Keisuke Kitamura[2], and Masao Ishimoto[1,*]

National Agricultural Research Center for Hokkaido Region, 1 Hitsujigaoka, Toyohira, Sapporo, Hokkaido 062-8555, Japan[1]; Graduate School of Agriculture, Hokkaido University, Kita 9 Nishi 9, Kita, Sapporo, Hokkaido 060-8589, Japan[2]; National Agricultural Research Center for Kyushu Okinawa Region, 2421 Suya, Koshi, Kumamoto 861-1192, Japan[3]; National Agricultural Research Center for Tohoku Region, 297 Uenodai, Kariwano, Daisen, Akita 019-2112, Japan[4]; Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan[5]; Institute of Society for Techno-Innovation of Agriculture, Forestry, and Fisheries, 446-1 Kamiyokoba, Tsukuba, Ibaraki 305-0854, Japan[6]; Faculty of Horticulture, Chiba University, 648 Matsudo, Chiba 271-8510, Japan[7] and National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan[8]

(Received 2 April 2009; accepted 25 May 2009; published online 16 June 2009)

Abstract

A well-saturated molecular linkage map is a prerequisite for modern plant breeding. Several genetic maps have been developed for soybean with various types of molecular markers. Simple sequence repeats (SSRs) are single-locus markers with high allelic variation and are widely applicable to different genotypes. We have now mapped 1810 SSR or sequence-tagged site markers in one or more of three recombinant inbred populations of soybean (the US cultivar 'Jack' × the Japanese cultivar 'Fukuyutaka', the Chinese cultivar 'Peking' × the Japanese cultivar 'Akita', and the Japanese cultivar 'Misuzudaizu' × the Chinese breeding line 'Moshidou Gong 503') and have aligned these markers with the 20 consensus linkage groups (LGs). The total length of the integrated linkage map was 2442.9 cM, and the average number of molecular markers was 90.5 (range of 70–114) for the 20 LGs. We examined allelic diversity for 1238 of the SSR markers among 23 soybean cultivars or lines and a wild accession. The number of alleles per locus ranged from 2 to 7, with an average of 2.8. Our high-density linkage map should facilitate ongoing and future genomic research such as analysis of quantitative trait loci and positional cloning in addition to marker-assisted selection in soybean breeding.

Key words: EST-derived SSR marker; integrated linkage map; microsatellite marker; polymorphism information content

1. Introduction

Soybean [*Glycine max* (L.) Merrill] is one of the most important grain legumes because it is a staple source of high-quality vegetable protein and oil for food products and industrial material. Over the past 20 years, the global growing area of soybean has increased by a factor of 1.7, whereas the average yield has increased

Edited by Masahiro yano

† Present address: National Agricultural Research Center for Western Region, 1-3-1 Senyuu, Zentsuuji, Kagawa 765-8508, Japan

* To whom correspondence should be addressed. Tel. +81-11-857-9312. Fax. +81-11-859-2178. E-mail: ishimoto@affrc.go.jp

by a factor of 2.3, reaching 2.5 tons/hectare in 2006.[1] However, the rate of growth in soybean yield has declined during the last 10 years, increasing by only 8% in this period. It is therefore imperative to develop new technologies and resources that will allow the supply of soybean to meet the large growth in demand anticipated in the near future.

An accurate and well-saturated genetic linkage map is fundamental to modern plant breeding because it allows both the identification of agronomic trait loci, including quantitative trait loci, and an understanding of genetic diversity and genome structure of genetic resources. Furthermore, such a linkage map is required for construction of a physical map. Since the first genetic map of soybean was constructed with phenotypic traits,[2] several linkage maps have been developed either alone or in combination with various types of molecular markers such as restriction fragment length polymorphism (RFLP), random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers.[3–7] Among these molecular markers, SSRs (also known as microsatellite markers) possess several favorable features: (i) they are inherited in a codominant manner at individual loci; (ii) they exhibit high levels of polymorphism and exist in multiple alleles; (iii) they are distributed evenly and randomly in the genome and (iv) they can be readily analyzed by the polymerase chain reaction (PCR) and subsequent gel electrophoresis.[5,6,8,9] Multiallelic molecular markers with a high level of polymorphism are useful for the detection of allelic differences among many genetic resources.[10–12] SSR markers are applicable to analysis of many of the segregating populations derived from the hybridization of any given genotypes. In addition, most PCR primer pairs for SSR markers yield a single amplification product for each soybean genotype,[13] allowing the use of a variety of means for the detection of SSR length polymorphisms. The amplification products are usually analyzed by electrophoresis on a polyacrylamide or agarose gel, but they can also be analyzed with a genetic sequencer in combination with a labeling protocol.[6,10,14] Moreover, SSR markers can be genotyped more rapidly and cost efficiently by simultaneous detection of multiple loci with the use of multiplex PCR analysis.[15] Such analysis is ideal for high-throughput and repetitive genotyping applications for which common sets of SSR marker loci are advantageous.

Given that SSR markers were found to map to individual loci with the same order in three different mapping populations, 606 such loci were integrated and aligned with RFLP, RAPD, AFLP and classical markers into a consensus set of 20 linkage groups (LGs) corresponding to the 20 pairs of soybean chromosomes.[16] The integrated linkage map was subsequently updated with 420 newly developed SSR markers and two additional mapping populations, giving a total of 1849 molecular markers including 1015 SSR, 709 RFLP, 73 RAPD, 6 AFLP and 46 classical markers.[17,18] The framework of 20 consensus LGs with an average of 50 SSR loci per group has provided a resource for construction of a linkage map for genetic analysis of qualitative and quantitative traits. However, there remain several genomic intervals of at least 20 cM that contain no SSR markers, and the number of SSR markers is not sufficient for application to all hybridizing combinations of germplasm, given that the genome size of soybean is 1.1 Gb.[19]

Xia et al.[20] recently developed 702 new SSRs from analysis of genomic DNA or by surveying expressed sequence tag (EST)-derived SSRs in the public database, and 121 of these SSR markers were mapped in an F_2 population derived from a cross between the Japanese cultivar Misuzudaizu and the Chinese experimental line Moshidou Gong 503. Additionally, 6920 primer pairs were designed to amplify SSRs from 63 676 non-redundant soybean ESTs available in the public database, and 668 EST-derived SSR markers were assigned to a molecular linkage map constructed with F_8 recombinant inbred lines (RILs) of Misuzudaizu $\times$ Moshidou Gong 503.[21] These new SSR markers enrich and saturate the soybean linkage map and provide a framework for analysis of the entire genome. We have now analyzed polymorphism of the new SSRs between parents of three mapping populations including the Misuzudaizu $\times$ Moshidou Gong 503 RIL population, generated three molecular linkage maps in combination with the public SSR markers[18] and integrated the new SSR markers into the framework of 20 consensus LGs. In addition, we investigated allelic polymorphism for the mapped EST-derived SSR markers in comparison with the public SSR markers, and we consider the transferability of the new SSR markers in soybean germplasm.

2. Materials and methods

2.1. *Mapping populations*

To develop an integrated linkage map, we used the following three mapping populations. (i) The MM population. This population originates from a cross between the Japanese cultivar Misuzudaizu and the Chinese weedy experimental line Moshidou Gong 503, consists of 165 RILs and has been described previously.[22,23] The RILs were available from the National BioResource Project of *Lotus* and *Glycine* (http://www.legumebase.agr.miyazaki-u.ac.jp/index.jsp). We studied 94 individual plants representing the F_8 generation of the RILs to construct a linkage map.[21]

(ii) The JF population. This population is derived from a cross between the US cultivar Jack and Fukuyutaka, a leading Japanese cultivar, and comprises 179 RILs. We randomly selected 91 individual plants representing the F_7 generation of the RILs and used for mapping. (iii) The PA population. This population is created from a cross between the Chinese cultivar Peking and the Japanese cultivar Akita. We randomly selected 68 individual plants representing the F_9 generation of 120 RILs.

2.2. DNA extraction and marker analysis

Total genomic DNA was extracted from young leaves or seed flour of individual samples with the use of an Automatic DNA Isolation System PI-50a (Kurabo, Osaka, Japan) according to Plant DNA Extraction Protocol (version 2). The concentration of DNA in solution was quantified by comparison with a serial dilution of lambda phage DNA (Promega, Madison, WI, USA) and was then adjusted to 10 ng/μL with sterile water. PCR amplification was performed in a reaction mixture (10 μL) containing 10 ng of total genomic DNA, 0.4 μM of each primer and 5 μL of GoTaq Green Master Mix (Promega) and with the use of a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). The amplification protocol comprised an initial denaturation for 2 min at 95°C; 30 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 47°C[13] or 55°C,[20] and extension for 30 s at 72°C; and a final extension for 5 min at 72°C. The EST-derived SSR markers were amplified with the modified 'Touchdown PCR' protocol.[24] PCR products were resolved by 10% polyacrylamide gel electrophoresis in Tris-borate-EDTA (TBE) buffer. The gel was stained with ethidium bromide, and the bands were visualized and photographed under ultraviolet light.

2.3. Screening and genotyping of DNA markers

SSR markers developed by USDA-ARS (United States Department of Agriculture, Agricultural Research Service) and DuPont Corporation[16,18] as well as SSR and sequence-tagged site (STS) markers developed by the Chiba University[20] were screened to detect polymorphism between parents of the JF and PA populations in addition to the MM population. The polymorphic markers were mapped with each of the three mapping populations. A total of 6920 EST-derived SSR primer pairs developed by the Kazusa DNA Research Institute[21] were used to detect polymorphisms between parents of the two mapping populations (JF and PA), and the polymorphic markers were analyzed in the two populations. The genotype data of 693 EST-derived SSR markers in the MM population were used from the previous study by Hisano et al.[21]

2.4. Construction of an integrated linkage map

We first used JoinMap (version 4.0)[25] to group and order molecular marker loci within each of the JF, PA and MM populations. The segregated markers were grouped in LGs on the basis of an LOD (logarithm of the odds ratio for linkage) score of ≥ 7.0 and referral to previously reported LGs of the public SSR marker loci.[16,21] Marker order was subsequently determined on the basis of a minimum LOD score of 1.0 and recombination threshold of 0.4 in each LG. Markers were tested for deviation from expected Mendelian segregation by the chi-squared test performed with the JoinMap software under the 'Locus Genotypic Frequency' command. Markers were sorted on the basis of the chi-squares test with a *P*-value of <0.05.

The three maps of each LG were integrated with the used of a minimum LOD score of 3.0 and recombination threshold of 0.4 under the 'Combine Groups for Map Integration' command of the JoinMap software, and they were reorganized with reference to the 20 consensus linkage map[16] and a previously reported map of EST-derived SSR markers.[21] The large interval between the two markers Sat_235 and GMES1325 in LG C1 was calculated with the use of MAPMAKER/EXP, version 3.0.[26] Recombination values were converted to genetic distance (cM) with the use of the Kosambi mapping function.[27] Each map of the 20 LGs was graphically visualized with MapChart.[28]

2.5. Evaluation of mapped markers and allele frequencies

Polymorphism analysis was performed with the same 24 soybean cultivars and lines used in a previous study,[21] namely 15 Japanese cultivars and lines (Enrei, Fukuyutaka, Hayahikari, Himeshirazu, Ibarakimame 7 gou, Koitozairai, Misuzudaizu, Nourin 2 gou, Saikai No. 20, Suzuyutaka, Tamahomare, Toyokomachi, Toyomusume, Yukihomare and Toiku No. 237), two Japanese breeding material lines (Tokei 758 gou and To-8E), three US cultivars (Adams, Harosoy and Jack), two Chinese lines (Peking and Moshidou Gong 503), an Indonesian cultivar (Wilis) and the Japanese wild accession (Funaba 3). The number of alleles and polymorphism information content (PIC) were analyzed for 304 USDA-ARS SSR markers on a 3% agarose gel[12] using Tris-acetate-EDTA buffer and for 934 EST-derived SSR markers mapped in this study on a 10% polyacrylamide gel using TBE buffer. The PIC for each SSR marker locus (i) was calculated according to the formula described by Keim et al.:[29] $\mathrm{PIC}(i) = 1 - \sum P_{ij}2$, where P_{ij} is the frequency of the jth allele of

Table 1. Numbers of SSR or STS markers analyzed, found to be polymorphic and mapped in each population of RILs

Status	RIL population											
	Jack × Fukuyutaka				Peking × Akita				Misuzudaizu × Moshidou Gong 503			
	ES[a]	CS[b]	US[c]	Total	ES	CS	US	Total	ES[d]	CS	US	Total
Primer pairs	6920	370	1015	8305	6920	370	1015	8305	6920	370	1015	8305
Polymorphic	360	100	402	862	406	103	320	829	680	NA[e]	NA[e]	NA[e]
Mapped	344	97	398	839	396	99	315	810	693	9	299	1006

[a]EST-derived SSR markers developed by Kazusa DNA Research Institute.[21]
[b]SSR or STS markers developed by Chiba University.[20,22]
[c]Public SSR markers developed by USDA-ARS and DuPont Corporation.[16,18]
[d]Genotype data were used from the previous study by Hisano et al.[21]
[e]Not applicable (NA); not all primer pairs analyzed.

the *i*th SSR locus and summation extends over *n* alleles.

3. Results and discussion

3.1. Screening of polymorphic molecular markers

We performed PCR amplification of template DNA from the parents of two mapping populations, the JF and PA populations, in addition to the MM population with EST-derived SSR primer pairs.[21] Amplification of template DNA from the parents Jack and Fukuyutaka yielded products from both parents with 6079 primer pairs and from only one parent with 31 pairs. Obvious polymorphism between the parents was observed with 360 (5.2%) of the primer pairs (Table 1). Similarly, PCR products were obtained from both parents Peking and Akita with 5517 primer pairs and from only one parent with 82 pairs. Unambiguous polymorphism between the parents was observed with 406 (5.9%) of the EST-derived SSR primer pairs. The polymorphism frequencies obtained for the parents of the JF and PA populations (5.2 and 5.9%, respectively) are approximately half of that previously reported (9.8%) for the parents (Misuzudaizu and Moshidou Gong 503) of the MM population (680 polymorphisms among the 6920 used for analysis).[21] A total of 1127 (16.3%) EST-derived SSR primer pairs gave polymorphisms in at least one of each parental pair of the three mapping populations.

Of the more than 1000 SSR or STS primer pairs developed by the Chiba University,[20,22,30] we selected 370 pairs on the basis of their mapping information and banding patterns in order to detect polymorphism between each pair of the parents of the JF and PA populations. Obvious polymorphism between the parents of the JF or PA population was observed with 100 (27.0%) and 103 (27.8%) of the primer pairs, respectively (Table 1). A total of 145 (39.2%) primer pairs yielded polymorphisms for at least one of the parental pairs of the JF and PA populations.

The public soybean SSR markers developed by USDA-ARS and DuPont Corporation[16,18] were examined for polymorphism between each pair of the parents of the three populations in order to construct a framework for the 20 consensus LGs. Clear polymorphism between the parents of the JF or PA population was observed with 402 (39.6%) and 320 (31.5%) of the 1015 SSR primer pairs, respectively (Table 1). In addition, 164 primer pairs for the 377 public SSR markers analyzed previously[12] as well as the 300 SSR markers previously reported[21] yielded unambiguous polymorphism between the parents of the MM population. Consequently, obvious polymorphism between at least one of each parental pair of the three populations was observed with a total of 614 (60.5%) SSR markers.

3.2. Construction of each linkage map

A total of 862 polymorphisms were scored in the JF population, with each primer pair yielding polymorphic bands at a single locus. After exclusion of 16 unlinked markers and 7 unsuccessfully positioned markers, 839 marker loci, which included 344 EST-derived SSR markers, 97 SSR or STS markers developed by Chiba University and 398 public SSR markers, were assembled into 26 LGs with a total genetic length of 2363.6 cM. Sixteen of these LGs were matched with A1, A2, B1, B2, D1a, D1b, D2, E, F, G, I, J, K, L, M and N of the 20 consensus LGs.[18,31] The other four consensus LGs (C1, C2, H and O) were divided into two or three fragments.

After exclusion of 12 unlinked markers and 7 unsuccessfully positioned markers, a total of 810 marker loci, including 396 EST-derived SSR markers, 99 SSR or STS markers, and 315 public SSR markers, were grouped into 32 LGs in the PA population. Twelve of these LGs corresponded to A1, A2, B2, D1a, D1b, D2, F, I, J, M, N and O of the 20 consensus LGs. The other eight consensus LGs (B1, C1, C2, E, G,

H, K and L) were divided into 2–4 fragments. All but one of the primer pairs yielded polymorphic bands at a single locus; the primer pair for CSSR534 detected two loci. The linkage map covered a genetic distance of 1575.9 cM.

Genotype data for 693 EST-derived SSR loci in the MM population were used from the previous study by Hisano et al.[21] In addition to these markers, 308 marker loci, including 9 SSR or STS markers and 299 public SSR markers, were assembled into 22 LGs with a total genetic length of 2187.2 cM, after the elimination of one unlinked marker. Eighteen of these LGs matched 18 of the 20 consensus LGs. The other two consensus LGs (C1 and E) were each fragmented into two short pieces. After the addition of two public SSR markers, Sat_137 and Satt684, of LG A1, five EST-derived SSR markers (GMES4205a, GMES6481, GMES2036, GMES4016 and GMES2497) mapped at one end of LG A1, though they were previously assigned to LG C2.[21] Twenty-four and one primer pairs detected double and quadruple loci, respectively. Segregation distortion was observed in the same region of consensus LG I in two mapping populations, JF (CSSR366, CSSR529, GMES1137, GMES2783, GMES6339 and Sat_189) and MM (CSSR529, GMES0289, GMES2783, GMES4080, GMES6339 and GMES6428).

3.3. Construction of an integrated linkage map

Since 648 of the mapped marker loci were common among at least two of the mapping populations, JoinMap analysis of the 1074 EST-derived SSR including 387 newly mapped marker loci, 141 SSR or STS including 81 newly mapped marker loci and 595 public SSR markers resulted in their successful merger into a genetic linkage map comprised the 20 consensus LGs and spanning 2442.9 cM of Kosambi map distance (Fig. 1, Table 2). The average genetic length of the 20 LGs was 122.1 cM, although the length ranged from 175.8 (LG A2) to 77.4 (LG J) cM (Table 2). The number of marker loci for the LGs ranged from 114 (LG F) to 70 (LG N), with an average of 90.5. The average genetic distance between markers varied between 0.94 (LG J) and 1.78 (LG A2) cM. A cluster of five EST-derived SSR markers (GMES4205a, GMES6481, GMES2036, GMES4016 and GMES2497) was previously assigned to LG C2,[21] but this cluster together with GMES0706, CSSR1,[20] GMES5782, GMES3776 and GMES1678 was assembled at one end of LG A1 in the present integrated linkage map as well as individual linkage maps (Fig. 1). Most marker loci were relatively evenly distributed along the LGs, although there were some regions with a high or low marker density (Fig. 1, Supplementary Fig. S1). The LG C1 still contained a large genetic interval (35.3 cM) between Sat_235 and GMES1325 for which there was no marker locus (Table 3). There were additional six large genetic intervals with a gap of >10 cM (Table 3).

The integrated linkage map contains 132 dominant marker loci, including 83 of the 1074 EST-derived SSR markers, 27 of the 141 SSR or STS markers, and 22 of the 595 public SSR markers, in at least one of the mapping populations. However, amplified products of each of 77 of these 132 marker loci exhibited polymorphism with bands of different sizes among 23 soybean cultivars and 1 wild soybean line, indicating that most of these loci could be used as codominant markers dependent on the combination of genotypes. Among the mapped molecular markers, 32 EST-derived SSR markers and 2 SSR markers developed by the Chiba University detected double loci in one or more than two of the mapping populations. One EST-derived SSR marker, GMES4205, gave four polymorphic bands positioned in four different LGs, whereas two SSR and one STS markers developed by the Chiba University were mapped to positions different from those reported previously.[20] These multiple loci were discriminated by appending a lowercase letter, as in GMES4205a and GMES4205b (Fig. 1, Supplementary Fig. S1).

3.4. Polymorphism of USDA-ARS SSR and EST-derived SSR markers

We examined allelic variation among the 24 soybean genotypes with the mapped SSR markers on the integrated linkage map. We selected 934 EST-derived SSR markers and 304 public SSR markers on the basis of both their unambiguous banding patterns and their distribution throughout the entire genome.[12,21] Though the two groups of the SSR markers were analyzed by the different electrophoresis systems on either polyacrylamide or agarose gel, the resolutions of amplified products were almost equivalent in both systems. For confirmation, the electrophoresis system for each marker was indicated in Supplementary Fig. S1. The number of alleles per locus ranged from 2 to 7 with an average of 2.8 for the EST-derived SSR markers and 3.1 for the public SSR markers (Fig. 2, Table 4). The most frequent allele numbers were 2 and 3 for the EST-derived SSR and public SSR markers, respectively. The PIC value, a measure of allelic diversity at a locus, ranged from 0.08 to 0.83, with an average value of 0.40 for all markers; the average was 0.38 for the EST-derived SSR markers and 0.46 for the public SSR markers (Fig. 2, Table 4). Marker loci with PIC values between 0.40 and 0.50 were the most common among both groups. The highest and

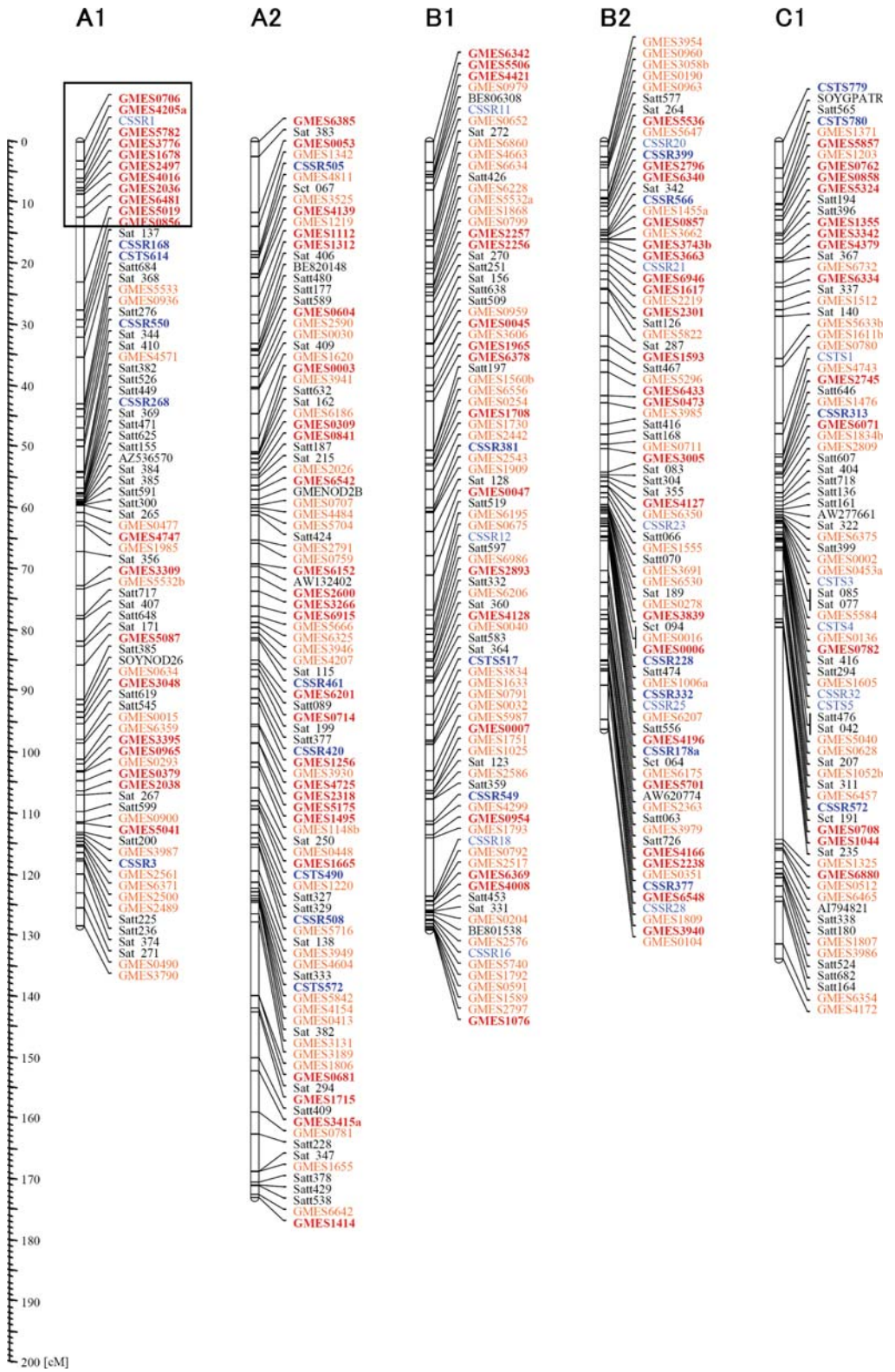

Figure 1. Genetic linkage map of soybean constructed with SSR and STS markers. The linkage map was visualized graphically with MapChart. The LGs were constructed from 595 public SSR markers developed by USDA-ARS and DuPont Corporation (black characters), 1074 EST-derived SSR markers developed by Kazusa DNA Research Institute (red), and 141 SSR or STS markers developed by the Chiba University (blue). A total of 468 newly mapped PCR-based markers were polymorphic in one or more of the three mapping populations (bold). The EST-derived SSR markers GMES4205a, GMES6481, GMES2036, GMES4016 and GMES2497 were previously assigned to LG C2,[21] but the JoinMap analysis here positioned these markers as well as GMES0706, CSSR1, GMES5782, GMES3776 and GMES1678 at one end of LG A1 (boxed region).

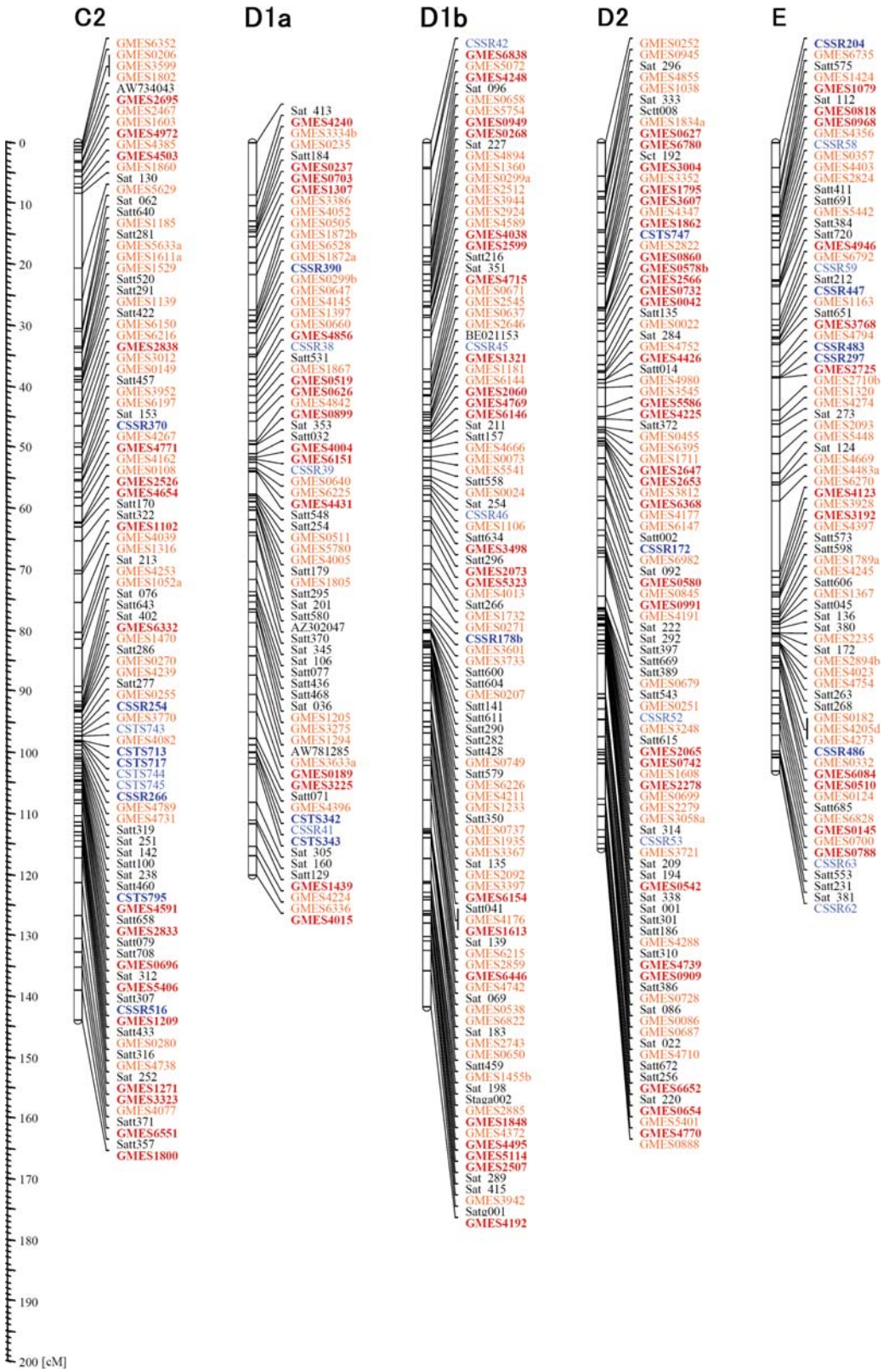

Figure 1. (Continued)

lowest average PIC values for the EST-derived SSR markers were detected in LG G (0.49) and LGs D2 and M (0.34), respectively. In contrast, the highest and lowest such values for the public SSR markers were detected in LG D2 (0.52) and LG H (0.34), respectively (Table 4). Clusters of marker loci with a

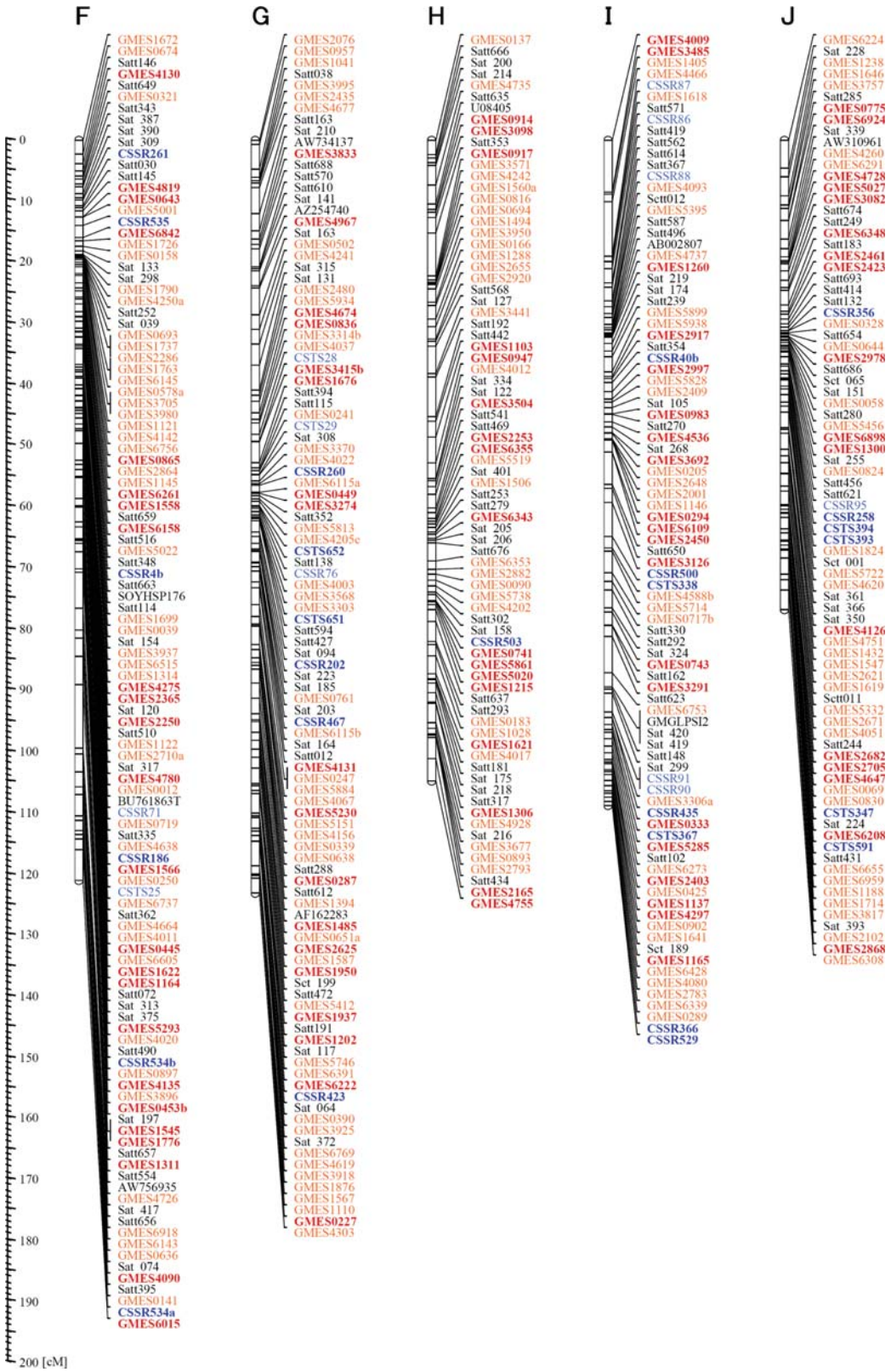

Figure 1. (Continued)

higher than average PIC value, such as those located between GMES1587 and GMES4303 (average PIC value of 0.58), were detected in LG G (Supplementary Fig. S1). In contrast, marker loci located between GMES0688 and GMES4161 in LG O showed an average PIC value of only 0.19.

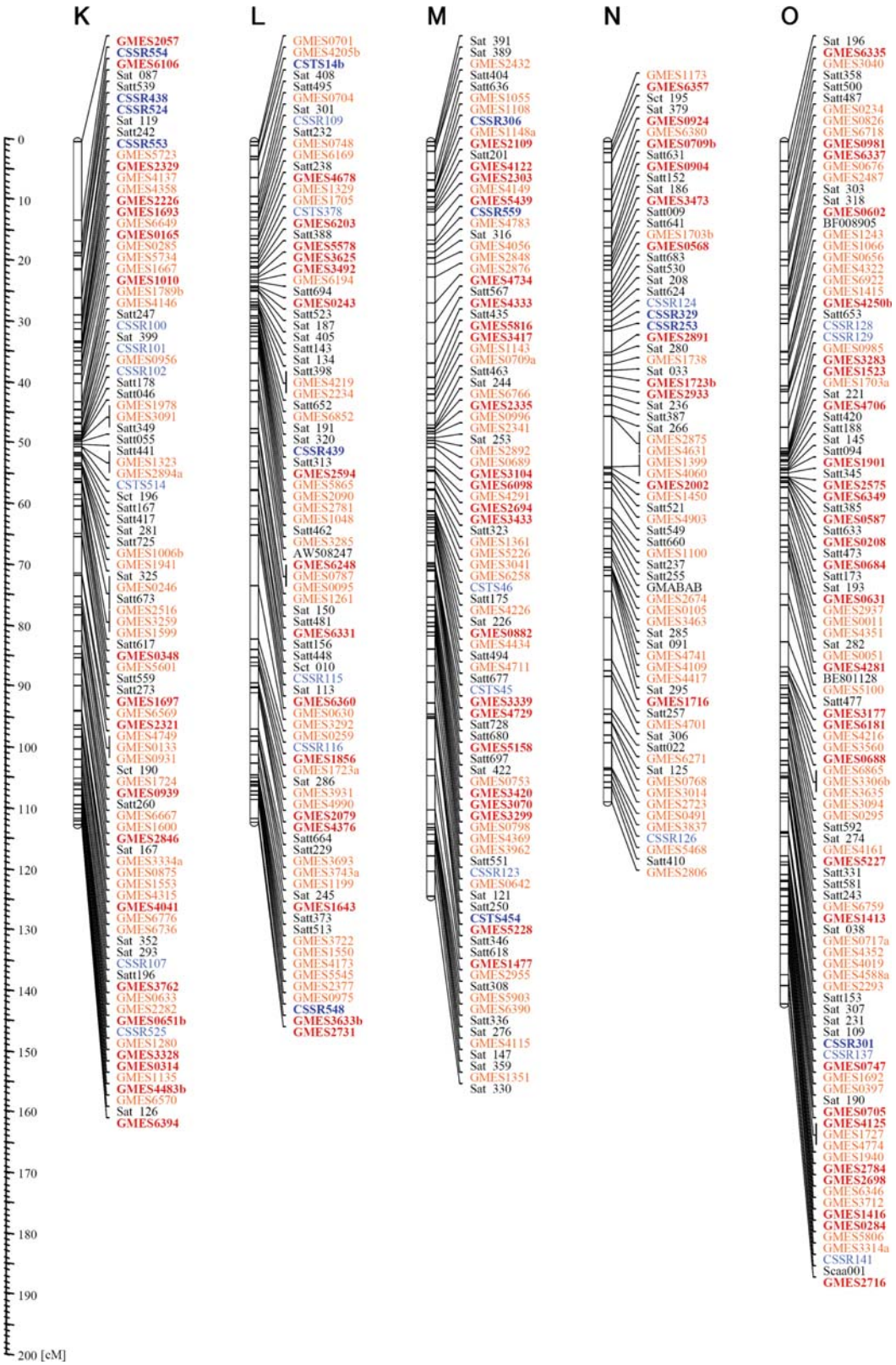

Figure 1. (Continued)

3.5. General discussion

Several SSR-based linkage maps have recently been developed independently for different mapping populations.[16,17,20,21] Although these previous studies accumulated almost 2000 SSRs mapped to the 20 consensus LGs, the positional relations

Table 2. Summary of the integrated linkage map and mapped markers

LG	Number of markers				Length (cM)	Average interval (cM)
	ES[a]	CS[a]	US[a]	Total		
A1	38	6	35	79	130.4	1.65
A2	61	6	32	99	175.8	1.78
B1	59	7	21	87	131.3	1.51
B2	49	11	21	81	97.8	1.21
C1	42	9	32	83	136.1	1.64
C2	56	10	34	100	146.3	1.46
D1a	43	6	24	73	122.0	1.67
D1b	70	4	32	106	144.1	1.36
D2	64	4	31	99	117.7	1.19
E	46	9	23	78	104.7	1.34
F	69	8	37	114	122.0	1.07
G	63	9	34	106	124.1	1.17
H	46	1	30	77	105.6	1.37
I	48	12	29	89	109.6	1.23
J	48	7	27	82	77.4	0.94
K	58	10	28	96	112.2	1.17
L	50	7	31	88	111.9	1.27
M	56	6	31	93	124.1	1.33
N	37	4	29	70	108.0	1.55
O	71	5	34	110	141.8	1.29
Total	1074	141	595	1810	2442.9	
Average	53.7	7.1	29.8	90.5	122.1	1.36

[a]Abbreviations are as in Table 1.

Table 3. Marker loci at the boundaries of genomic intervals of >10 cM containing no SSR or STS markers

LG	Flanking SSR loci	Interval distance (cM)
A1	GMES6481–GMES5019	10.8
A2	GMES3189–GMES1806	12.3
C1	Sat_235–GMES1325	35.3
C2	Sat_130–GMES5629	12.5
E	GMES6270–GMES4123	11.7
F	AW756935–GMES4726	10.4
K	GMES2057–CSSR554	12.9

among SSRs have remained unclear. We have mapped a total of 1810 SSR or other kinds of PCR-based markers in one or more than three mapping populations and have generated an integrated linkage map with the use of JoinMap 4.0 software. In addition to 1343 known SSR markers, 387 EST-derived SSR markers and 81 SSR or STS markers developed by the Chiba University were newly positioned on the map, allowing determination of the relative position for each microsatellite marker. The total length of the present integrated linkage map is 2442.9 cM, which is almost identical to that of an updated integrated map (2524.0 cM)[16–18] including SSR, RFLP, RAPD, AFLP and classical markers as well as to that of a transcript map (2550.3 cM) with SNP markers.[7] The genetic length of individual LGs in the present integrated map correlates well with that of those in the previous maps ($r = 0.73$ and 0.72, respectively). Whereas LGs A1 and E are >20 cM longer in the present map than in the previous maps, LGs B2, D2 and F are >20 cM shorter. The estimated length of these LGs, with the exception of LG A1, agrees relatively well with that of those in the latest version of the integrated linkage map.[32] In contrast, several new SSR marker loci were mapped outside the existing SSR loci in LG A1, extending its genetic length by >20 cM. The present integrated linkage map thus appears to achieve high coverage of the entire soybean genome.

The present integrated linkage map is composed of 1810 molecular markers including 1074 EST-derived SSR markers, 141 SSR or STS markers developed by the Chiba University and 595 public SSR markers with an average marker density of 0.74/cM. Although the markers are distributed relatively evenly, randomly and densely throughout the genome, there remain seven gaps of >10 cM between markers. In particular, LG C1 contains the largest gap of 35.3 cM between Sat_235 and GMES1325, an interval for which no SSR markers were reported in previous studies.[17,21,31] This gap may reflect either a lack of polymorphic markers in a highly homozygous region or the presence of hot spots of recombination that enlarge the genetic distance corresponding to a short physical distance.

Public SSR markers, which have been developed by USDA-ARS and DuPont Corporation,[17,31] were strictly standardized to yield a single amplification product for a given soybean genotype. All the primer pairs for these markers thus corresponded to a single locus in the present mapping populations. In contrast, EST-derived SSR markers sometimes gave multiple bands, which were mapped separately to different LGs in some cases. Thirty-two and one primer pairs for these SSRs detected double and quadruple loci, respectively. A total of 84 public SSR markers were common to the three individual linkage maps and mapped to the 20 consensus LGs described previously.[17] Most markers were arranged in the same order and with similar genetic intervals in the JF, PA, MM and present integrated linkage maps as well as in the consensus integrated linkage map. However, reversions or altered orders of markers relative to the consensus linkage map were observed for LGs D1a, D1b, E, J and N. Each RIL population used in this study was derived from a cross between a

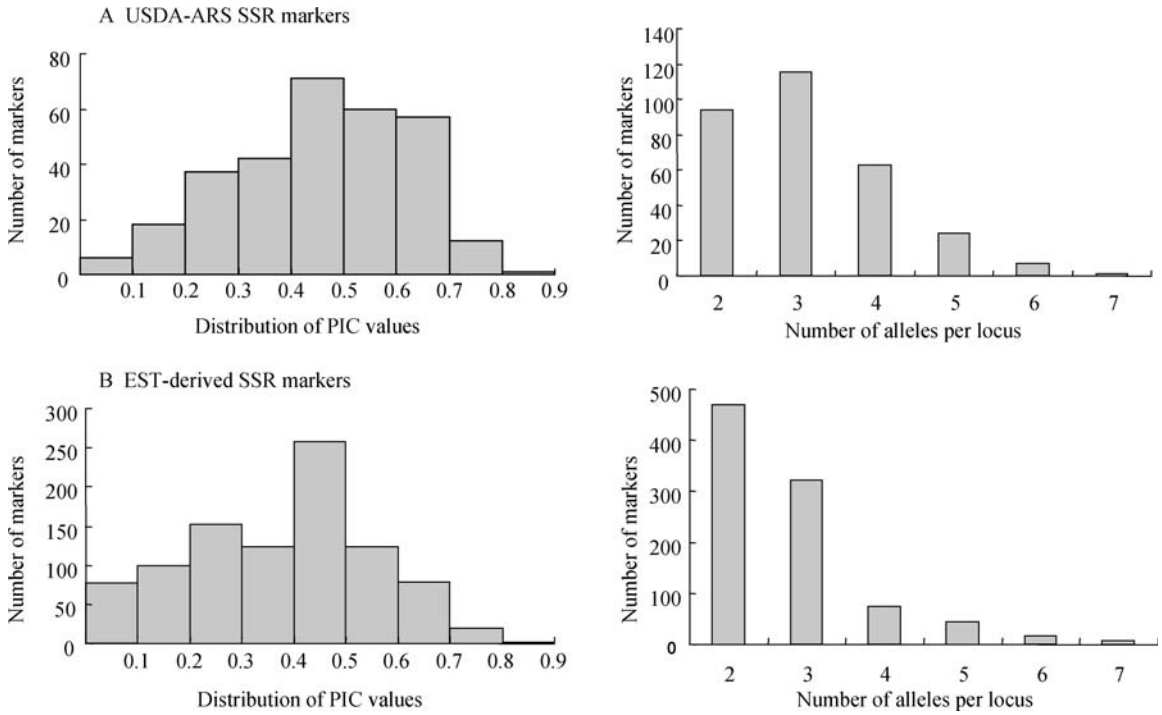

Figure 2. Comparison of allele frequencies and PIC values for markers in 23 soybean cultivars and one wild accession. (**A**) Public SSR markers developed by USDA-ARS and DuPont Corporation. (**B**) EST-derived SSR markers.

Table 4. Average numbers of alleles and PIC values for mapped markers in each LG

LG	ES[a]			US[a]			Total		
	Number	Average PIC value	Average number of alleles	Number	Average PIC value	Average number of alleles	Number	Average PIC value	Average number of alleles
A1	35	0.39	2.9	16	0.48	3.2	51	0.42	3.0
A2	56	0.39	2.7	15	0.47	3.1	71	0.41	2.8
B1	51	0.36	2.6	12	0.44	2.8	63	0.38	2.7
B2	40	0.43	2.8	13	0.50	3.2	53	0.45	2.9
C1	35	0.37	2.8	18	0.48	3.1	53	0.41	2.9
C2	50	0.36	2.8	20	0.48	3.4	70	0.40	2.9
D1a	38	0.42	2.9	13	0.45	2.8	51	0.42	2.9
D1b	58	0.37	2.7	13	0.47	3.3	71	0.39	2.8
D2	53	0.34	2.5	14	0.52	3.9	67	0.38	2.8
E	38	0.41	2.7	13	0.42	2.8	51	0.41	2.7
F	58	0.39	2.9	16	0.47	3.0	74	0.41	2.9
G	58	0.49	3.0	21	0.47	3.4	79	0.48	3.1
H	43	0.41	2.9	11	0.34	2.5	54	0.39	2.9
I	42	0.38	2.6	17	0.48	3.4	59	0.41	2.8
J	41	0.38	2.9	18	0.45	3.1	59	0.40	2.9
K	51	0.39	3.0	16	0.42	3.0	67	0.40	3.0
L	45	0.37	2.8	16	0.46	3.3	61	0.39	2.9
M	49	0.34	2.4	13	0.49	3.1	62	0.37	2.5
N	34	0.35	2.5	14	0.46	3.0	48	0.38	2.6
O	59	0.35	2.6	15	0.46	3.1	74	0.37	2.7
Total	934	0.38	2.8	304	0.46	3.1	1238	0.40	2.8

[a]Abbreviations are as in Table 1.

Japanese cultivar and foreign cultivar or experimental line. All the parents were genetically classified into distant groups by cluster analysis of length polymorphism of SSR markers.[12] In contrast, with one exception, previous consensus integrated linkage maps were constructed on the basis of mapping populations derived from US domestic cultivars.[16,18] The discrepancy in marker order may be due to inversion, translocation, insertion or deletion of genomic regions specific to mapping populations with different genetic backgrounds.[7] Similarly, the frequency of recombination has been found to vary substantially among populations,[33] resulting in divergence of genetic distance among mapping populations.

In the present study, we analyzed allelic variation for a large number of SSR loci including 934 EST-derived and 304 public SSRs. The overall average PIC value for the EST-derived markers (0.38) was lower than that for the public SSR markers (0.46), consistent with previous observations.[21,30] The PIC values and their distribution are indicative of the transferability of mapped SSR markers to different genotypes. The most frequent number of alleles per EST-derived marker was two, suggesting that several such markers might reflect insertion–deletion polymorphisms (indels), as is the case for STS markers.[21] The PIC values of EST-derived markers in individual LGs did not correlate with those of the public SSR markers ($r = -0.23$). The public SSR marker loci with the highest and lowest average PIC values were found in LG D2 (0.52) and LG H (0.34), respectively, whereas the highest and lowest average PIC values for the EST-derived SSR markers were detected in LG G (0.41) and both LG D2 and LG M (0.34), respectively. The reason for this difference between the two SSR marker groups of different origin remains unclear.

Soybean is a member of the phaseoloid legumes, which include numerous crops of economic importance such as common bean (*Phaseolus vulgaris* L.), cowpea [*Vigna unguiculata* (L.) Walp.], mungbean [*V. radiata* (L.) Wilczek] and azuki bean [*V. angularis* (Willd.) Ohwi et Ohashi]. The availability of an integrated linkage map with markers, such as EST-derived SSR markers, that are transferable to closely related species would allow comparative genetic analysis between soybean and other legume crops as a strategy for applied agriculture.[21,34] Comparative genetic mapping and synteny analysis would facilitate gene discovery through the development of markers that are closely linked genetically to the target gene.[35] In addition, information on EST-derived SSR markers provides a basis for evaluation of the structural diversity and evolution of these various species.[36]

A chromosome-based assembly of sequence data in soybean, Glyma1.0, is available at Phytozome.[37] The accuracy and resolution of genetic linkage maps and physical maps (whole-genome sequencing) can thus be confirmed and refined reciprocally. The availability of physical maps does not diminish the importance of genetic linkage maps, with the two map types having a synergistic influence on modern soybean genomics as well as marker-assisted selection for target traits. The large number of SSR markers positioned in the present and previous studies promote understanding of the soybean genome at the structural and functional levels.

Acknowledgements: We thank Asuka Ueki, Motomi Suzuki, Takako Yamamoto and Ayako Suzuki for technical assistance.

Supplementary data: Supplementary data are available online at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported by the Ministry of Agriculture, Forestry, and Fisheries of Japan [Genomics for Agricultural Innovation (DD-3260) and Research Project for Utilizing Advanced Technologies in Agriculture, Forestry, and Fisheries (18063)].

References

1. FAS 2008, *Oilseeds: World Markets and Trade. Foreign Agricultural Service (FAS)*, United States Department of Agriculture: Washington, DC.
2. Palmer, R. G. and Kilen, T. C. 1987, Qualitative genetics and cytogenetics, In: Wilcox, J. R. (ed.), *Soybeans: Improvement, Production, and Uses*, 2nd Ed., Agronomy: Madison, Wisconsin, pp. 135–209.
3. Keim, P., Diers, B. W., Olson, T. C. and Shoemaker, R. C. 1990, RELP mapping in soybean: association between marker loci and variation in quantitative traits, *Genetics*, **126**, 735–742.
4. Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. 1990, DNA polymorphisms amplified by arbitrary primers are useful as genetic markers, *Nucleic Acids Res.*, **18**, 6531–6535.
5. Akkaya, M. S., Bhagwat, A. A. and Cregan, P. B. 1992, Length polymorphisms of simple sequence repeat DNA in soybean, *Genetics*, **132**, 1131–1139.
6. Akkaya, M. S., Shoemaker, R. C., Specht, J. E., Bhagwat, A. A. and Cregan, P. B. 1995, Integration of simple sequence repeat DNA markers into a soybean linkage map, *Crop Sci.*, **35**, 1439–1445.
7. Choi, I. Y., Hyten, D. L., Matukumalli, L. K., et al. 2007, A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis, *Genetics*, **176**, 685–696.
8. Morgante, M. and Olivieri, A. M. 1993, PCR-amplified microsatellites as markers in plant genetics, *Plant J.*, **3**, 715–723.

9. Maughan, P. J., Saghi Maroof, M. A. and Buss, G. R. 1995, Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean, *Genome*, **38**, 715–723.
10. Diwan, N. and Cregan, P. B. 1997, Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean, *Theor. Appl. Genet.*, **95**, 723–733.
11. Chen, Y., Wang, D., Arelli, P., Ebrahimi, M. and Lelson, R. L.. 2006, Molecular marker diversity of SCN-resistant sources in soybean, *Genome*, **49**, 938–949.
12. Hwang, T. Y., Nakamoto, Y., Kono, I., et al. 2008, Genetic diversity of cultivated and wild soybeans including Japanese elite cultivars as revealed by length polymorphism of SSR markers, *Breed. Sci.*, **58**, 315–323.
13. Cregan, P. B., Bhagwat, A. A., Akkaya, M. S. and Rongwen, J. 1994, Microsatellite fingerprinting and mapping of soybean, *Methods Mol. Cell. Biol.*, **5**, 49–61.
14. Yu, Y. G., Saghai Maroof, M. A., Buss, G. R., Maughan, P. J. and Tolin, S. A. 1994, RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance, *Phytopathology*, **84**, 60–64.
15. Tang, S., Kishore, V. K. and Knapp, S. J. 2003, PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower, *Theor. Appl. Genet.*, **107**, 6–19.
16. Cregan, P. B., Jarvik, T., Bush, A. L., et al. 1999, An integrated genetic linkage map of the soybean genome, *Crop Sci.*, **39**, 1464–1490.
17. Cregan, P. B. 2003, An integrated genetic linkage map of the soybean July 2003, http://bldg6.arsusda.gov/~pooley/soy/cregan/soymap.htm (July 2003).
18. Song, Q. J., Marek, L. F., Shoemaker, R. C., et al. 2004, A new integrated genetic linkage map of the soybean, *Theor. Appl. Genet.*, **109**, 122–128.
19. Arumaganthan, K. and Earle, E. 1991, Nuclear DNA content of some important plant species, *Plant Mol. Biol. Rep.*, **9**, 208–218.
20. Xia, Z., Tsubokura, Y., Hoshi, M., et al. 2007, An integrated high-density linkage map of soybean with RFLP, SSR, STS, and AFLP markers using a single F_2 population, *DNA Res.*, **14**, 257–269.
21. Hisano, H., Sato, S., Isobe, S., et al. 2007, Characterization of the soybean genome using EST-derived microsatellite markers, *DNA Res.*, **14**, 271–281.
22. Yamanaka, N., Ninomiya, S., Hoshi, M., et al. 2001, An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion, *DNA Res.*, **8**, 61–72.
23. Watanabe, S., Tajuddin, T., Yamanaka, N., Hayashi, M. and Harada, K. 2004, Analysis of QTLs for reproductive development and seed quality traits in soybean using recombinant inbred lines, *Breed. Sci.*, **54**, 399–407.
24. Sato, S., Isobe, S., Asamizu, E., et al. 2005, Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.), *DNA Res.*, **12**, 301–364.
25. Van Ooigen, J. W. 2006, *JoinMap 4.0 Software for the Calculation of Genetic Linkage Maps in Experimental Populations*, Plant Research International: Wageningen, the Netherlands.
26. Lincoln, S. E., Daly, M. J. and Lander, E. S. 1992, *Constructing Genetic Maps with MAPMAKER/EXP 3.0, Whitehead Institute Technical Report*, 3rd Ed., Whitehead Institute: Massachusetts, Cambridge.
27. Kosambi, D. D. 1944, The estimation of map distance from recombination values, *Ann. Eugen.*, **12**, 172–175.
28. Voorrips, R. E.. 2002, MapChart: software for the graphical presentation of linkage maps and QTLs, *J. Hered.*, **93**, 77–78.
29. Keim, P., Beavis, W., Schupp, J. and Freestone, R. 1992, Evaluation of soybean RFLP marker diversity in adapted germ plasm, *Theor. Appl. Genet.*, **85**, 205–212.
30. Hossain, K. G., Kawai, H., Hayashi, M., Hoshi, M., Yamanaka, N. and Harada, K. 2000, Characterization and identification of (CT)n microsatellites in soybean using sheared genomic libraries, *DNA Res.*, **7**, 103–110.
31. Cregan, P. B., Mudge, J., Fickus, E. W., Danesh, D., Denny, R. and Young, N. D. 1999, Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by rhg1 locus, *Theor. Appl. Genet.*, **99**, 811–818.
32. SoyBase 2008, Soybean chromosome number to linkage group assignments, http://soybase.org/LG2Xsome.php (30 July 2008).
33. Pfeiffer, T. W. and Vogt, S. D. 1990, Variability for recombination frequencies in the AP12 soybean population, *Crop Sci.*, **30**, 545–549.
34. Bushman, B. S., Larson, S. R., Mott, I. W., et al. 2008, Development and annotation of perennial Triticeae ESTs and SSR markers, *Genome*, **51**, 779–788.
35. Hwang, T. Y., Moon, J. K., Yu, S., et al. 2006, Application of comparative genomics in developing molecular markers tightly linked to the virus resistance gene Rsv4 in soybean, *Genome*, **49**, 380–388.
36. Choi, H. K., Mun, J. H., Kim, D. J., et al. 2004, Estimating genome conservation between crop and model legume species, *Proc. Natl Acad. Sci. USA*, **101**, 15289–15294.
37. The Department of Energy's Joint Genome Institute and the Center for Integrative Genomics 2008, Phytozome Glyma 1, http://www.phytozome.net/soybean.php (8 December 2008).