Theor Appl Genet (2004) 108:1131–1139
DOI 10.1007/s00122-003-1527-2

ORIGINAL PAPER

W.-K. Zhang · Y.-J. Wang · G.-Z. Luo · J.-S. Zhang · C.-Y. He · X.-L. Wu · J.-Y. Gai · S.-Y. Chen

QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers

Received: 30 August 2003 / Accepted: 27 October 2003 / Published online: 22 January 2004

Abstract A set of 184 recombinant inbred lines (RILs) derived from soybean vars. *Kefeng No.1* × *Nannong 1138-2* was used to construct a genetic linkage map. The two parents exhibit contrasting characteristics for most of the traits that were mapped. Using restricted fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs) and expressed sequence tags (ESTs), we mapped 452 markers onto 21 linkage groups and covered 3,595.9 cM of the soybean genome. All of the linkage groups except linkage group F were consistent with those of the consensus map of Cregan et al. (1999). Linkage group F was divided into two linkage groups, F1 and F2. The map consisted of 189 RFLPs, 219 SSRs, 40 ESTs, three *R* gene loci and one phenotype marker. Ten agronomic traits—days to flowering, days to maturity, plant height, number of nodes on main stem, lodging, number of pods per node, protein content, oil content, 100-seed weight, and plot yield—were studied. Using WINQTLCART, we detected 63 quantitative trait loci (QTLs) that had LOD>3 for nine of the agronomic traits (only exception being seed oil content) and mapped these on 12 linkage groups. Most of the QTLs were clustered, especially on groups B1 and C2. Some QTLs were mapped to the same loci. This pleiotropism was common for most of the QTLs, and one QTL could influence at most five traits. Seven EST markers were found to be linked closely with or located at the same loci as the QTLs. EST marker GmKF059a, encoding a repressor protein and mapped on group C2, accounted for about 20% of the total variation of days to flowering, plant height, lodging and nodes on the main stem, respectively.

Communicated by H.F. Linskens

W.-K. Zhang, Y.-J. Wang and G.-Z. Luo contributed equally to this investigation.

W.-K. Zhang · Y.-J. Wang · G.-Z. Luo · J.-S. Zhang · C.-Y. He · X.-L. Wu · S.-Y. Chen (✉)
Plant Biotechnology Laboratory,
Institute of Genetics and Developmental Biology,
The Chinese Academy of Sciences,
100101 Beijing, China
e-mail: sychen@genetics.ac.cn
Tel.: +86-10-64886859
Fax: +86-10-64873428

J.-Y. Gai (✉)
Soybean Research Institute of Nanjing Agricultural University,
National Center for Soybean Improvement,
National Key Laboratory for Crop Genetics
and Germplasm Enhancement,
210095 Nanjing, China
e-mail: sri@njau.edu.cn
Tel.: +86-25-4395405
Fax: +86-25-4395405

Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the most important crop plants in the world, accounting for 48% of the world market in oil crops. It is also one of the more important sources of protein for both human consumption and as a fodder. In China, soybean is a major oilseed crop and is grown on about eight million hectares. Plant breeders have long sought to enhance the yield and quality of soybean. The construction of a genetic linkage map will provide genetic information that will be valuable in improving the agronomic traits. Several genetic linkage maps including molecular markers have been constructed since the early 1990s (Keim et al. 1990; 1997; Muehlbauer et al. 1991; Lark et al. 1993; Akkaya et al.1995; Shoemaker and Specht. 1995; Zhang et al. 1997; Liu et al. 2000; Wu et al. 2001). Using a population derived from a cross of *G. max* × *G. soja*, Shoemaker and Olson (1993) constructed a molecular genetic linkage map of soybean that consisted of 25 linkage groups with 365 restriction fragment length polymorphisms (RFLPs), 11 random amplified polymorphic DNAs (RAPDs), three classical markers, and four isozyme loci. In 1995, Shoemaker and Specht (1995) succeeded in integrating various markers into a common linkage map. Using three populations—the USDA/Iowa State *G. max* × *G. soja* F_2 population, the University of Utah *Minsoy* × *Noir 1* recombinant inbred population, and the University of

Nebraska *Clark* × *Harosoy* F_2 population—Cregan et al. (1999) aligned the 20+ linkage groups derived from each of the three populations into a consensus set of 20 homologous linkage groups presumed to correspond to the 20 pairs of soybean chromosomes.

Based on the construction of a soybean genetic linkage map, quantitative trait locus (QTL) mapping has been reported for a number of agronomic traits in soybean (Keim et al. 1990; Lark et al. 1995; Specht et al. 2001; Tasma et al. 2001; Hoecka et al. 2003). Mansur et al. (1996) used a recombinant inbred line (RIL) population to investigate the genetic basis of 15 agronomic traits. They found that most of the QTLs were clustered, and that the effects of a single locus could be large. For most traits, major loci appeared to act independently and additively. Lee et al. (1996) detected a major locus of *Dt1* associated with plant height on linkage group L. This locus was also associated with lodging and explained 67.7% of the total variation for plant height and 56.4% for lodging. Using two populations, Mian et al. (1996) found that two QTLs for seed weight on linkage groups F and K, respectively, were located at similar genome regions in both populations. Mansur et al. (1993a) mapped a QTL on group A2 for oil content; similar results were also reported by Brummer et al. (1997). Diers et al. (1992) and Brummer et al. (1997) detected a strong QTL on linkage group I for protein content. Occasionally the resistance to bacteria, fungi, and insects in soybean showed a multilocus model, and the QTLs for these traits were also mapped on linkage groups (Yuan et al. 2002).

Most of the genetic linkage maps were constructed with RFLP, amplified (A)FLP, RAPD and simple sequence repeat (SSR) markers. All of these markers are the fragments of genome DNA. Recently, expressed sequence tags (ESTs) have been used as markers and incorporated into linkage maps. The growing EST database has provided a valuable resource for EST markers. The association of EST markers with phenotypes can increase our understanding of the biochemical pathways and mechanisms affecting agronomically important traits (Matthews et al. 2001). Using the EST database and QTL mapping, the genes related to agronomic traits may be identified. Several ESTs have been found to be linked with QTLs for flowering time, leaflet shape, and leaf area (Yamanaka et al. 2001; Matthews et al. 2001). However, soybean ESTs related to QTLs for other agronomic traits remain to be identified.

Three genetic linkage maps (Zhang et al. 1997; Liu et al. 2000; Wu et al. 2001) have been constructed in our laboratory. The 2001 linkage map consists of 24 linkage groups, including 792 markers of which more than half are AFLP markers. Although AFLP markers are efficient for constructing maps, they can easily form clusters that result in large gaps on the map. In the investigation reported here, a RIL population derived from two cultivars that differ greatly for most of the agronomic traits was used to construct a genetic linkage map and for QTL analysis of ten agronomic traits. While most of the markers were RFLPs and SSRs, some EST markers were also used and their associations with QTLs were evaluated.

Materials and methods

Plant materials and agronomic traits

The RIL population used was derived from the cross between *Kefeng No.1* and *Nannong 1138-2*. The original population consisted of 206 $F_{2:7:10}$ lines; this was reduced to 184 lines using the method of simulated population sampling criteria (SPSC; Wang 2001). The RILs and their parents were planted in a randomized block design with 0.7×0.7-m^2 hill-plots, eight replications, at Jiangpu Station, Nanjing Agricultural University, Nanjing, China. Ten agronomic traits—days to flowering (FD), days to maturity (MD), plant height (HT), number of nodes on main stem (SN), lodging (LD), number of pods per node (PN), protein content (PT), oil content (OL), 100-seed weight (SW), and plot yield (YD)—were measured (Table 1). Protein and oil content were determined by near-infrared reflectance on two replicates randomly selected from the eight replicates.

Phenotype markers and resistance loci

Flower color (*W*) and three loci conferring resistances to different strains of soybean mosaic virus (*Rn1*, *Rn3*, and *Rsa*) were used to construct the genetic map. The resistance was determined at Nanjing Agricultural University.

Table 1 Statistical analysis of traits of the RIL population and the parents (*SD* standard deviation)

Trait	Unit	RIL population			Parents	
		Average	Range	SD	Kefeng No.1	Nannong 1138-2
Days to flowering (FD)	Days	50.3	37.6–60.7	6.1	47.3	60.1
Days to maturity (MD)	Days	114.8	94.0–139.1	10.4	106.3	125.1
Plant height (HT)	Centimeters	52.5	23.8–97.1	14.8	36.1	66.1
Nods on mail stem (SN)	Nodes	14.1	7.7–21.7	2.9	10.7	16.3
Lodging (LD)	Score	1.8	1.0–3.6	0.6	1.0	1.2
100-seed weight (SW)	Grams	12.1	7.9–17.0	1.7	8.1	19.3
Plot yield (YD)	Grams	61.8	7.4–123.4	18.1	23.7	83.7
Pods per node (PN)	Seeds	3.0	1.7–4.9	0.6	3.3	2.9
Protein content (PT)	Percentage	38.9	34.5–44.5	1.7	40.6	44.5
Oil content (OL)	Percentage	19.6	16.7–24.8	1.4	16.7	18.7

RFLP analysis

The genomic DNAs of the RIL individuals and the two parents were extracted from young leaves. The parents and RIL individuals were surveyed for polymorphisms using restriction enzymes *Eco*RI, *Eco*RV, *Dra*I, *Bam*HI, *Hin*dIII, and *Taq*I, respectively. Three kinds of RFLP probes derived from soybean genomic DNA, soybean cDNA clones, and cloned genes were used for RFLP analysis. Most of the RFLP markers derived from genomic DNA were provided by Prof. Shoemaker (Iowa State University, USA); the others were obtained from genomic clones digested by *Pst*I, telemore sequences and RAPD products (Zhang et al. 1997; Liu et al. 2000; Wu et al. 2001).

SSR markers

SSR primers were synthesized according to the sequences published on the SOYBASE website. The amplifications were performed in 15-μl aliquots of reaction mix containing of 2.5 m*M* Mg^{2+}, 0.5 m*M* each of dNTPs, 0.1% Tween-20, 10 pmol of each primer, 2 U *Taq* polymerase and 50 ng genomic DNA. The initial step was performed at 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 46°C for 1 min and 68°C for 1 min, with a final extension for 10 min at 72°C. The products of amplifications were separated on 1% agarose or 4% polyacrylamide gels, and the fragments were visualized by ethidium bromide or silver staining, respectively.

EST analysis

EST markers were obtained from a cDNA library constructed using leaves of *Kefeng No.1*. (He et al. 2003). The cDNAs inserted in the plasmid pBluescript were amplified using T3 and T7 promoter primers. The PCR products were gel-purified and diluted in super pure water for RFLP analysis.

Three additional cloned genes were also used as EST markers in this study: *GmKR4* [a candidate of R genes with a nucleotide binding site (NBS) domain], *Gmpti1* (a *pti*-liked gene of soybean), and *GmHZ* (a transcription factor with a conserved homeodomain and leucine zipper).

Construction of a genetic linkage map and QTL mapping

The program MAPMAKER/EXP VER. 3.0 (Lander et al. 1987; Lincoln et al. 1993) was used to analyze the linkage between the markers. The parameters included the Kosambi function, and the linkage criteria were a LOD score greater than 3.0 and a maximum genetic distance of 37.2 cM. The error detection probability level was set at 5%. The markers were first grouped at an LOD of 6.0 and then ordered to construct the core linkage groups. Other markers were added using the INSERT command as the LOD score was decreased to 3.0. The RIPPLE command was used to test the order with a window size of 5 after each ORDER command. Linkage groups were named using the designations of the consensus map of Cregan et al. (1999).

QTL analysis was performed using WINQTLCART (Zeng 1993). Model 6 was adopted, and the control marker number and window size was 5 and 10 cM, respectively. The walk speed was 1 cM, and the forward regression method was selected. LOD score peaks greater than 3.0 indicated the existence of QTLs. The averages of the eight agronomic traits were used in the QTL analysis. For oil and protein content, the averages were of the two replicates.

Results

Construction of the genetic linkage map

A total of 522 markers were identified as having polymorphisms between the two parents, *Kefeng No.1* and *Nannong 1138-2*, including 214 RFLPs, 53 EST markers, 251 SSRs, one phenotype trait and three resistance loci. Based on these markers and their segregation in the RIL population, we used MAPMAKER/EXP. VER. 3.0 (Lander et al. 1987; Lincoln et al. 1993) to construct a genetic linkage map that covered 3,595.9 cM and consisted of 21 linkage groups (Fig. 1). Of the 522 markers, only 452 (189 RFLPs, 219 SSRs, 40 ESTs, three *R* loci and *W*) were mapped; most of the linkage groups were consistent with those of Cregan et al.'s map (Cregan et al. 1999) except for the F linkage group. In this study, the F linkage group was divided into two linkage groups, F1 and F2 (Fig. 1). The three *R* gene loci were all mapped onto group D1b+W and the *W* locus was mapped on group F1.

The distribution of the molecular markers on the linkage groups was also examined. The average number of markers on each group is 21.5. The linkage group with the most markers (45) and longest genetic distance (368.8 cM) is group G followed by groups N and K. The group with the least markers (seven) and the shortest genetic distance (36.1 cM) is Group H. The average length of the linkage groups is 1,71.23 cM, and the average distance between markers is 7.96 cM.

Most of the markers mapped in the 184 RILs showed Mendelian segregation (1:1). However, there were 77 markers (17.0%) that showed a significant deviation from the expected ratio.

EST mapping and homology search

We sequenced about 30,000 soybean ESTs, of which150 were selected and examined, and only 28 showed polymorphisms between the two parents. Using the BLASTX algorithms, we compared the corresponding amino acid sequences with those in GeneBank (Altschul et al. 1997) (Table 2).

The 28 ESTs together with three previously cloned genes, *GmKR4*, *GmHZ and GmPti1*, produced 53 pairs of polymorphic bands between the two parents. These polymorphic bands were then used as EST markers for segregation analysis in the population. Following linkage analysis, 41 markers were mapped onto 14 linkage groups (Fig. 1). Group B1 had more EST markers (nine) than any of the other groups, and these markers were mapped onto the bottom of B1 successively. Linkage groups A1, B2, C2 D1a+Q, F2, L, and N have only one to five EST markers on each group.

Seventeen ESTs were found to generate only one polymorphic locus for each, whereas the other 11 ESTs were found to have two or three pairs of polymorphic bands between the two parents. Some of the EST markers

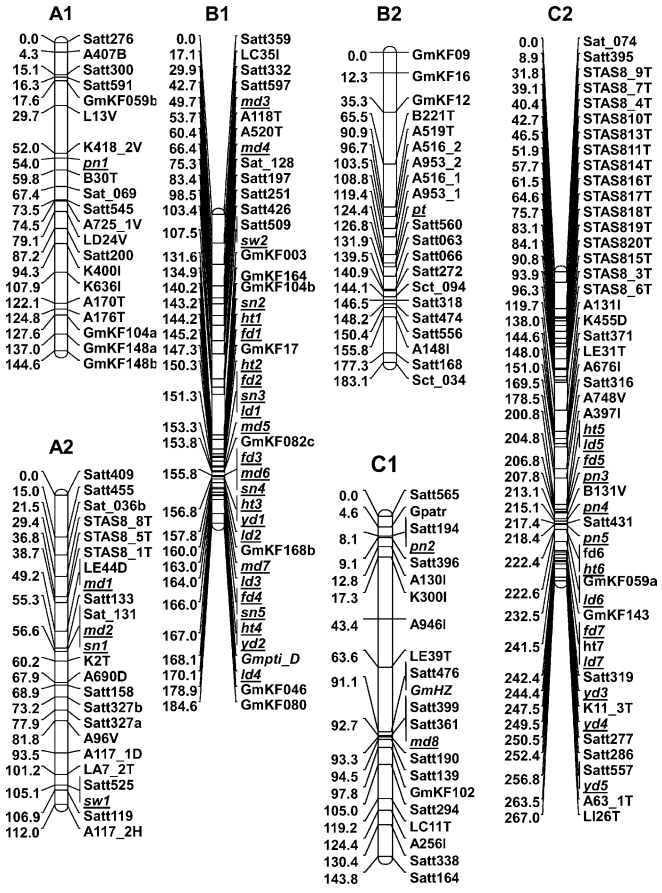

Fig. 1 Soybean linkage map based on a RIL populations derived from Kefeng No.1×Nannong 1138-2. Markers designated with *GmKF* are EST markers. The linkage groups were named using the designations of Cregan's consensus map (Cregan et al. 1999). QTL loci are *italicized* and *underlined*

Fig. 1 (continued)

derived from the same cDNA were mapped onto two linked loci; for example, GmKF148a and GmKF148b on linkage group A1; GmKF008a and GmKF008b on linkage group D1a+Q; GmKF167a and GmKF167b on linkage group F2; GmKF125a and GmKF1258b on linkage group L. Others were mapped onto different linkage groups: for example, GmKF059a was mapped onto group C2, whereas GmKF059b was mapped on group A1 (Fig. 1).

GmKR4, a candidate of R genes with a NBS domain (B.J. Wang and S.Y. Chen, unpublished results), was mapped on linkage group E. The transcription factor gene, *GmHZ*, originally identified by cDNA-AFLP analysis and possibly related to defense response, was mapped on linkage group C1. The *GmPti1* gene was identified from soybean (*G. max* L.) by its similarity to the tomato *Pti1* (A.G. Tian AG and S.Y. Chen, unpublished results). *GmPti1* was found to have five pairs of polymorphic bands between the parents. However, only two of them were mapped, with Gmpti_C at the top of linkage group I and Gmpti_D at the bottom of group B1.

Statistical analysis of ten agronomic traits

Ten agronomic traits (see Materials and methods) were evaluated. The results of our statistical analysis of the ten traits of the RIL population and the parents are shown in Table 1. For all traits except oil content, the differences between the two parents were significant. The ranges in the values for traits like days to flowering, days to maturity, plant height, nodes on main stem, pods per node, 100-seed weight, and plot yield, exceeded those of the parents. The transgressive phenomena were prominent for plant height, nodes on main stem, plot yield, and pods per node.

QTL mapping of agronomic traits

The QTL mapping for the ten traits was performed using WINQTLCART (Basten et al. 1994, Wang et al. 2001). Sixty-four QTLs were detected and mapped on 12 linkage groups (Table 3, Fig. 1). QTLs with LOD scores greater than 3 were detected for all of the ten agronomic traits

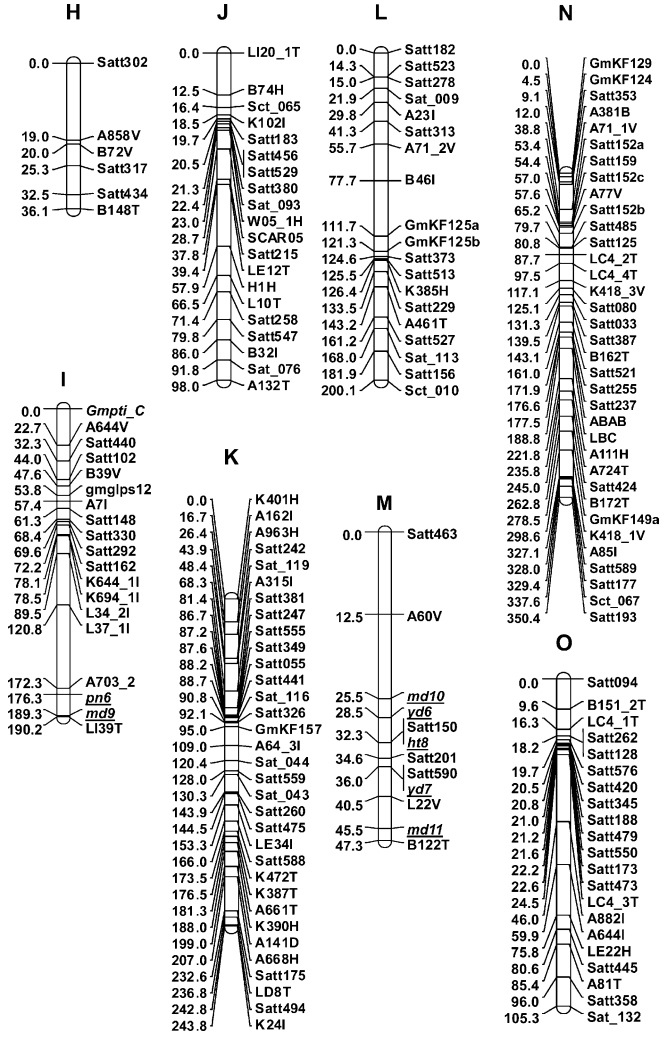

Fig. 1 (continued)

except seed oil content. At least four QTLs were found for each of eight traits. When the LOD score criterion was reduced to 2.5, one QTL was found on linkage group D1b+W for oil content. Other traits had various numbers of QTLs—from four to ten.

Days to flowering had eight QTLs distributed over linkage groups B1, C2, and E. Three QTLs that mapped on group C2 (*fd5*, *fd6* and *fd7*) had larger effects than others. Each of them accounted for more than 20% of the total variation, respectively. Days to maturity had 11 QTLs distributed over linkage groups A2, B1, C1, I, and M. About one-half of these were located on group B1 of which *md3* and *md4* were the most important, explaining 27.5% and 24.1% of the total variation, respectively.

Eight to ten QTLs for each of plant height, lodging and nodes on the main stem were found, Most of the QTLs for these three traits were located at the same region on groups B1 and C2. The QTLs with the larger effect (accounting for more than 18% variation) for the three traits were all mapped on group C2.

Table 2 List of ESTs showing polymorphism between the parents

cDNA clone	Homologous gene	Source
GmKF003	Poly (A)-binding protein	*Nicotiana tabacum*
GmKF008	Casein kinase II beta chain	*Arabidopsis thaliana*
GmKF012	Wound-induced protein	*Mesembryanthemum crystallinum*
GmKF022	Eukaryotic initiation factor	*Brassica oleracea*
GmKF046	Vacuolar H^+-ATPase catalytic subunit	*Vigna radiata*
GmKF050	Translation initiation factor	*Arabidopsis thaliana*
GmKF058	Translation elongation factor	*Glycine max*
GmKF059	Repressor protein	*Glycine max*
GmKF080	Glutathione peroxidase	*Arabidopsis thaliana*
GmKF082	Water channel protein	*Glycine max*
GmKF086	Ca^{2+}-binding binding protein	*Phaseolus vulgaris*
GmKF090	Abscisic stress-ripening protein	*Prunus armeniaca*
GmKF102	Initiation factor NeIF-5A2	*Arabidopsis thaliana*
GmKF104	Vacuolar H^+-ATPase E subunit	*Phaseolus acutifolius*
GmKF120	Lipid-transfer protein	*Fragaria ananassa*
GmKF124	BTF3-like transcription factor	*Arabidopsis thaliana*
GmKF125	RNA binding protein	*Arabidopsis thaliana*
GmKF129	Elongation factor 1-alpha	*Zea mays*
GmKF138	Integral membrane protein	*Arabidopsis thaliana*
GmKF143	bZIP transcription factor 2	*Phaseolus vulgaris*
GmKF148	Transcriptional co-activator	*Arabidopsis thaliana*
GmKF149	Protein serine/threonine kinase	*Sorghum bicolor*
GmKF157	Protein phosphatase	*Rubus idaeus*
GmKF164	Transcription factor ARF6	*Arabidopsis thaliana*
GmKF167	Isoflavone synthase 2	*Trifolium repens*
GmKF168	Heat-shock protein 80	*Euphorbia esula*
GmKF169	Calmodulin-like protein kinase	*Glycine max*
GmKF177	Cytosolic chaperonin, delta-subunit	*Glycine max*
GmKR4	Resistance gene candidate	*Nicotiana tabacum*
GmHZ	Transcription factor	*Arabidopsis thaliana*
GmPti1	Pti-like kinase	*Lycopersicon esculentum*

Pods per node and seed weight are yield component traits. Six QTLs were detected for pods per nodes. Of these, three were positioned on group C2 and the others on groups A1, C1, and I. The variation accounted for by each of these six QTLs ranged from 7.1% to 10.2%. Four QTLs for seed weight were mapped on groups A2, B1, and D2. The most significant one on group D2 accounted for 11.4% variation. Three linkage groups, B1, C2 and M, were associated with yield and seven QTLs were mapped on them. Most of these QTLs accounted for about 10% of the total variation.

Only one QTL was identified on linkage group B2 for seed protein content. The LOD score was 3.5, and the QTL accounted for 12.4% variation. One QTL with a LOD of 2.5 was detected for oil content on group D1b+W; 7.4% variation was accounted for by this QTL.

Twelve of the linkage groups were found to have QTLs for the ten traits investigated (A1, A2, B1, B2, C1,

Table 3 QTLs of agronomic traits in soybean

Trait[a]	QTL	Linkage group	Marker or interval	Position (cM)	LOD	R^2(%)	Additive effect
FD	*fd1*	B1	GmKF104b-GmKF177	145.2	5.7	11.2	2.1
	fd2	B1	GmKF177-GmKF082c	151.3	8.4	16.2	2.5
	fd3	B1	GmKF082c-GmKF168b	155.8	7.3	12.9	2.3
	fd4	B1	GmKF168b-Gmpti_D	166.0	6.9	12.3	2.2
	fd5	C2	A397I-B131 V	206.8	10.1	22.6	3.0
	fd6	C2	Satt431-GmKF059a	222.4	13.4	21.6	3.1
	fd7	C2	GmKF143-Satt319	241.5	9.6	20.0	2.9
	fd8	E	Satt496-A374H	113.6	4.3	14.6	–2.4
MD	*md1*	A2	LE44D	49.2	4.1	6.1	–2.6
	md2	A2	Sat_131	56.6	3.4	5.1	–2.4
	md3	B1	Satt597-A118T	49.7	13.1	27.5	5.8
	md4	B1	A520T-Sat_128	66.4	8.6	24.1	5.3
	md5	B1	GmKF177-GmKF082c	153.3	5.3	8.4	3.2
	md6	B1	GmKF082c-GmKF168b	155.8	6.0	10.8	3.7
	md7	B1	GmKF168b-Gmpti_D	163.0	4.4	8.4	3.2
	md8	C1	Satt399	92.7	3.1	4.2	2.2
	md9	I	A703_2B-LI39T	189.3	3.7	5.3	–2.5
	md10	M	A60V-Satt150	25.5	3.1	7.3	2.8
	md11	M	L22V-B122T	45.5	3.3	4.8	2.3
HT	*ht1*	B1	GmKF104b-GmKF177	144.2	3.3	5.8	3.8
	ht2	B1	GmKF177-GmKF082c	150.3	3.3	5.8	4.1
	ht3	B1	GmKF082c-GmKF168b	156.8	7.3	13.3	5.7
	ht4	B1	GmKF168b-Gmpti_D	167.0	5.2	7.8	4.6
	ht5	C2	A397I-B131 V	204.8	11.0	21.4	7.0
	ht6	C2	Satt431-GmKF059a	222.4	16.4	24.3	7.7
	ht7	C2	GmKF143-Satt319	241.5	12.6	23.6	7.6
	ht8	M	Satt150	32.3	3.4	5.0	3.4
LD	*ld1*	B1	GmKF177-GmKF082c	151.3	4.1	8.5	0.2
	ld2	B1	GmKF082c-GmKF168b	157.8	4.8	9.4	0.2
	ld3	B1	GmKF168b-Gmpti_D	164.0	4.9	9.3	0.2
	ld4	B1	Gmpti_D-GmKF046	170.1	4.5	8.5	0.2
	ld5	C2	A397I-B131 V	204.8	8.3	18.2	0.3
	ld6	C2	GmKF059a	222.6	11.4	18.9	0.3
	ld7	C2	GmKF143-Satt319	241.5	7.2	14.8	0.3
	ld8	F2	Satt114	102.8	3.2	4.8	0.1
SN	*sn1*	A2	Sat_131	56.6	3.2	4.6	–0.7
	sn2	B1	GmKF104b-GmKF177	143.2	5.3	10.8	1.0
	sn3	B1	GmKF177-GmKF082c	151.3	5.7	9.9	1.0
	sn4	B1	GmKF082c-GmKF168b	155.8	5.3	8.8	0.9
	sn5	B1	GmKF168b-Gmpti_D	166.0	5.1	8.2	0.9
	sn6	C2	A397I-B131 V	205.8	9.2	20.1	1.4
	sn7	C2	GmKF059a	222.6	12.9	19.1	1.4
	sn8	C2	GmKF143-Satt319	241.5	10.4	18.9	1.4
	sn9	F1	Satt252-Satt269	27.6	5.3	10.2	1.0
	sn10	I	A703_2B-LI39T	186.3	3.7	6.1	–0.7
PN	*pn1*	A1	K418_2V-B30T	54.0	3.6	7.9	–0.2
	pn2	C1	Satt194	8.1	3.9	9.3	0.2
	pn3	C2	A397I-B131V	207.8	3.3	8.1	–0.2
	pn4	C2	B131V-Satt431	215.1	3.2	7.1	–0.2
	pn5	C2	Satt431-GmKF059a	218.4	3.1	7.3	–0.2
	pn6	I	A703_2B-LI39T	176.3	3.4	10.2	–0.2
SW	*sw1*	A2	Satt525	105.1	3.4	6.7	0.4
	sw2	B1	Satt509	107.5	3.8	10.2	0.5
	sw3	D2	A611D-B146H	31.0	4.5	9.2	0.5
	sw4	D2	B146H-Satt458	35.4	4.8	11.4	0.6
YD	*yd1*	B1	GmKF082c-GmKF168b	156.8	4.4	9.2	5.6
	yd2	B1	GmKF168b-Gmpti_D	167.0	5.4	10.2	5.9
	yd3	C2	Satt319-K11_3T	244.4	6.0	12.3	6.4
	yd4	C2	K11_3T-Satt277	249.5	5.9	11.5	6.1
	yd5	C2	Satt557	256.8	6.9	12.6	6.5
	yd6	M	A60V-Satt150	28.5	3.8	9.8	5.5
	yd7	M	Satt590	36.0	3.5	6.6	4.5
PT		B2	A953_1H-Satt560	124.4	3.5	12.4	–0.6
OL		D1b+W	A481V-A725_3V	10.0	2.5	7.4	–0.3

[a] See Table 1 for abbreviations of traits

C2, D1B+W, D2, E, F2, I, and M). Most of the QTLs were clustered, especially on groups B1 and C2 (Fig. 1). There were 24 QTLs on linkage group B1, and these could be related to days to flowering, days to maturity, plant height, nodes on main stem, lodging, seed weight, and yield. On group C2, 15 QTLs were detected and subsequently associated with days to flowering, plant height, pods per node, lodging, and yield. In the intervals between GmKF082c and GmKF168b of group B1 and between GmKF168b and Gmpti_D of group C2, there were six QTLs associated with different traits in each of the respective intervals. Other linkage groups had one to four QTLs.

It was possible to relate one QTL to more than one trait. Several QTLs of different traits were mapped on the same loci (Table 3, Fig. 1). On linkage group A2, *md2* and *sn1* were located on the same locus, 56.6 cM from the top. On linkage group B1, 12 QTLs were mapped on five loci—*fd2*, *sn3*, *ld1* on a locus at 151.3 cM; *fd3*, *md6*, *sn4* on a locus at 155.8 cM; *ht3*, *yd1* on a locus at 156.8 cM; *fd4*, *sn5* on a locus at 166.0 cM; *ht4*, *yd2* on a locus at 167.0 cM. For linkage group C2, eight QTLs for three traits were located at three loci: the locus at 204.8 cM from the top has two QTLs,*ht5* and *ld5*; the locus at 222.4 cM from the top has *fd6* and *ht6*; the locus at 241.5 cM from the top has three QTLs, *fd7, ht7, ld7*.

Association of ESTs with QTLs of agronomic traits

Several ESTs were found to be closely linked with the QTLs of agronomic traits. These were all mapped on linkage groups B1 and C2. Five EST markers (GmKF104b, GmKF177, GmKF082c, GmKF168b, Gmpti_D) that mapped on group B1 were associated with several traits—days to flowering, nodes on main stem, days to maturity, plant height, lodging, and yield. Two EST markers (GmKF059a, GmKF143) mapped on C2. GmKF059a, a repressor protein gene, was mapped on the same locus as QTLs *ld6* and *sn7* and had large effects, accounting for 18.9% and 19.1% of the total variation of lodging and nodes on main stem, respectively. It was also closely linked with QTLs *fd6* and *ht6* (0.2 cM) and accounted for more than 20% of the total variation, respectively. GmKF143, a putative bZIP transcription factor gene, was also related to the four traits above. However, the variations explained were less than 10% for all of them together

Discussion

We report here the construction of a genetic linkage map using RFLP, SSR and EST markers. The map covers about 3,600 cM and consists of 21 linkage groups. Although the markers are substantially evenly distributed among the linkage groups, gaps still remain on the linkage groups, especially between L37–1I and A703–2B on linkage group I.

ESTs are derived from cDNA clones sequenced partially or completely and provide more information on gene function than other molecular markers. They can also be converted to PCR-based markers such as sequence-tagged sites (STS) and single nucleotide polymorphisms (SNPs). When ESTS are used to construct a genetic map, more information will be provided than when other markers are applied (Matthews et al. 2001). In total, 150 cDNA clones were surveyed in this study, but only 28 (18.6%) showed polymorphisms between the parents. Coding sequences, especially those of housekeeping genes, are conserved among many plant species. Therefore, ESTs can be used in different plant species for map construction and QTL mapping (Yamanaka et al. 2001).

The two parents differed greatly for most of the traits studied in the present work. Ten traits were studied for their QTLs, and almost all of these had QTLs with large effects; most were located on linkage groups B1 and C2. In previous studies *E1* was found to inhibit flowering under natural day length and the locus of *E1* was located proximal to a QTL for flowering time on linkage group C2 (Bernard 1971; Tasma et al. 2001). Four QTLs for flowering time—*FT1*, *FT2*, *FT3*, *FT4*— have also been found and mapped on linkage groups C2, O, L, and B1, respectively (Yamanaka et al. 2001). We detected seven loci for flowering time in the present study, with QTLs *fd1*, *fd2*, *fd3*, and *fd4* mapping on group B1, while QTLs *fd5, fd6*, and *fd7* were located at the same region as *FT1*, on group C2. There were no QTLs detected on linkage groups O and L. From this comparison, we propose that the major gene(s) controlling flowering time may be located on linkage groups C2 and/or B1.

Five linkage groups were found to relate to days to maturity—A2, B1, C1, I, M. The QTLs on linkage group A2 and M were consistent with those reported by Mansur et al. (1996) and Orf et al. (1999). Early reports indicated that maturity locus *E3* mapped on linkage group L (Cregan et al. 1999), however we did not find any QTLs for maturity on group L in the present investigation.

Plant height, lodging and nodes on the main stem are important agronomic traits. Although low plant height can prevent lodging, which may subsequently decrease the yield, it is often concomitant with reduced nodes on the main stem. Thus, these three traits were closely related. The QTLs of these three traits coexist on linkage groups B1 (for *ht1* to *ht4*, *ld1* to *ld4*, *sn2* to *sn5*) and C2 (*ht5* to *ht7*, *ld5* to *ld7*, *sn6* to *sn8*). Mian et al. (1998) found a QTL of plant height near the locus detected by probe A397 on group C2. In the present investigation we also detected one of the QTLs (*ht5*) for plant height in the same region.

Pods per node is the only trait for which *Nannong 1138-2* has a lower value than *Kefeng No.1*. Six QTLs for this trait were detected on four linkage groups of Nannong 1138-2: the alleles of five QTLs reduced the number of pods per node, whereas, the allele of *pn2* (linkage group C1) could increase it. Tanksley and Nelson (1996) obtained similar results. This phenomenon implies

that the target alleles could be found in the germplasm with a reverse phenotype.

Seed weight is a yield component trait. Four QTLs on three linkage groups (A2, B1, D2) were related to seed weight. Similar QTLs have also been detected in other investigations (Mansur et al. 1996; Mian 1996). Yield is the most important trait for most crops. Early reports showed that the QTLs for yield mapped to linkage group C2, M, and L (Mansur et al. 1996; Spechet et al 2001). We also detected several QTLs for yield on linkage groups C2 and M, but none on L. The QTLs on linkage group M have been proven to be related to yield by different groups. Spechet et al. (2001) found that the QTL on linkage group M was the strongest, accounting for 33–38% of the total variation. We also found two QTLs, yd6 and yd7, on linkage group M. However, these QTLs accounted for only 9.8% and 6.6% of the total variation, respectively. Further saturation of linkage group M may result in the detection of more major QTLs on it.

The differences between the parents were not significant with respect to protein and oil content. Therefore, only one QTL of strong effects was detected for protein content. A QTL for oil content was also revealed when the LOD was set to be larger than 2.5.

The association between ESTs and quantitative traits may increase our knowledge of the genes influencing the agronomic traits. Seven ESTs were found to be associated with quantitative traits like days to flowering, plant height, lodging, and nodes on the main stem. The LOD scores on the loci of the EST markers were at least 3.2. GmKF059 encodes a repressor protein and had two copies in the genome. GmKF059a was mapped on linkage group C2, at the same locus as QTLs *ld6* and *sn7*, which account for 18.9% and 19.1% of the total variation of lodging and nodes on the main stem, respectively. It was also closely linked to the QTLs of *fd6* and *ht6* (0.2 cM). The GmKF059a allele of *Nannong 1138-2* delayed days to flowering by 3 days, increased plant height by 7.6 cm, and resulted in 1.4 nodes on the main stem, but it does not have any influence on yield. In rice, a homologous protein OsDr1 has been isolated and found to encode a nuclear repressor protein. This protein can interact with another protein, OsDrAp1, and with a TATA binding protein/DNA complex (Song et al. 2002). Further study of GmKF059 should reveal its possible functions in soybean.

Acknowledgements This work was supported by National 973 project (G1998010209 and 2002CB111301) and National High Technology Program (2002AA211051).

References

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB (1995) Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci 35:1439–1445

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Basten CJ, Weir BS, Zeng Z-B (1994) ZMAP—a QTL cartographer. In: Smith C, Gavora JS, Benkel B, Chesnais J, Fairfull W, Gibson JP, Kennedy BW, Burnside EB (eds) Proc 5th World Congr Genet Appl Livestock Prod: Computing Strategies Software. Organizing Committee, 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ont., vol 22, pp 65–66

Bernard RL (1971) Two genes for time of flowering and maturity in soybean Crop Sci 11:242–244

Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC (1997) Mapping QTL fro seed protein and oil content in eight soybean populations. Crop Sci 37:370–378

Cregan PB, Jarvik T, Bush AL, et al. (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Diers BW, Keim P, Fehr WR, Shoemaker RC (1992) RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83:608–612

He CY, Tian AG, Zhang JS, Zhang ZY, Chen SY (2003) Isolation and characterization of a full-length resistance gene homolog from soybean. Theor Appl Genet 106:786–793

Hoecka JA, Fehr WR, Shoemaker RC, Welkea GA, Johnson SL, Cianzioa SR (2003) Molecular marker analysis of seed size in soybean. Crop Sci 43:68–74

Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735–742

Keim P, Schupp JM, Travis SE, et al. (1997) A high-density soybean genetic map based on AFLP markers. Crop Sci 37:537–543

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lark KG, Weisemann JM, Matthews BF, Palmer R, Chase K, Macalma T (1993) A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: Minsoy and Noir 1. Theor Appl Genet 86:901–906

Lark KG, Chase K, Adler F, Mansur LM, Orf JH (1995). Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc Natl Acad Sci USA 92:4656–4660

Lee SH, Bailey MA, Mian MAR, Shipe ER, Ashley DA, Parrot WA, Hussey RS, Boerma HR (1996) Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor Appl Genet 92:516–523

Lincoln SE, Lander SL (1993) MAPMAKER/EXP 3.0 and MAPMAKER/QTL 1.1. Technical report. Whitehead Institute of Medical Research, Cambridge, Mass.

Liu F, Zhuang B-C, Zhang J-S, Chen S-Y (2000) Construction and analysis of soybean genetic map. Acta Bot Sin 27:1018–1026

Mansur LM, Lark KG, Kross HK, Oliveira L (1993) Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.) Theor Appl Genet 86:907–913

Mansur LM, Orf JH, Chase K, Jarvik T, Cregan PB, Lark KG (1996) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci 36:1327–1336

Matthews BF, Devine TE, Weisemann JM, Beard HS, Lewers KS, McDonald MH, Park Y-B, Maiti R, Lin J-J, Kuo J, Pedroni MJ, Cregan PB, Saunders JA (2001) Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. Crop Sci 41:516–521

Mian MAR, Bailey MA, Tamulonis JP, Shipe ER, Carter TE, Parrott JWA, Ashley DA, Hussey RS, Boerma HR (1996) Molecular markers associated with seed weight in two soybean populations. Theor Appl Genet 93:1011–1016

Mian MAR, Ashly DA, Vencil WK, Boerma HR (1998) QTLs conditioning early growth in a soybean population segregating for growth habit. Theor Appl Genet 97:1210–1216

Muehlbauer GJ, Staswick PE, Specht JE, Graef GL, Shoemaker RC, Keim P (1991) RFLP mapping using near-isogenic lines in the soybean [*Glycine max* (L.) Merr.]. Theor Appl Genet 81:189–198

Orf JH, Chase K, Adler FR, Mansur LM, Lark KG (1999) Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Sci 39:1652–1657

Shoemaker RC, Olson TC (1993) Molecular linkage map of soybean (*Glycine max* L. Merr.). In: O'Brien SJ (ed) Genetic maps: locus maps of complex genomes. Cold Spring Harbor Press, Cold Spring Harbor, pp 6.131–6.138

Shoemaker RC, Specht JE (1995) Integration of the soybean molecular and classical genetic linkage groups. Crop Sci 35:436–446

Song W, Solimeo H, Rupert RA, Yadav NS, Zhu Q (2002) Functional dissection of a rice Dr1/DrAp1 transcriptional repression complex Plant Cell 14:181–195

Specht JE, Chase K, Macrander M, Graef GL, Chung JU, Markwell JP, Germann M, Orf JH, Lark KG (2001) Soybean response to water: a QTL analysis of drought tolerance. Crop Sci 41:493–509

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis in a cross between an elite proceeding line of tomato and its wild relative *L. pimpinellifolium.* Theor. Appl. Genet. 92:213–224

Tasma IM, Lorenzen LL, Green DE, Shoemaker RC (2001) Mapping genetic loci for flowering times, maturity and photoperiod insensitivity in soybean. Mol Breed 8:25–35

Wang SC, Basten, CJ, Zeng ZB (2001) WINDOWS QTL CARTOGRAPHER, VER. 1.21. Department of Statistics, N.C. State University, Raleigh, N.C.

Wang YJ (2001) Establishment and adjustment of RIL population and its application to map construction, mapping genes resistant to SMV and QTL analysis of agronomic and quality traits in soybeans. PhD thesis, Nanjing Agricultural University, China

Wu XL, He CY, Wang YJ, Zhang ZY, Dongfang Y, Zhang JS, Chen SY, Gai JY (2001) Construction and analysis of a genetic linkage map of soybean. Acta Genet Sin 28:1051–1061

Yamanaka N, Ninomiya S, Hoshi M, Tsubokura Y, Yano M, Nagamura Y, Sasaki T, Harada K (2001) An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. DNA Res 8:61–72

Yuan J, Njiti VN, Meksem K, Iqbal MJ, Triwitayakorn K, Kassem MA, Davis GT Schmidt ME, Lightfoot DA (2002) Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop Sci 42:271–277

Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976

Zhang DS, Dong W, Hui DW, Chen SY, Zhuang BC (1997) Construction of a soybean linkage map using an F2 hybrid population from a cultivated variety and a semi-wild soybean. Chin Sci Bull 42:1326–1330