Identification of Quantitative Trait Loci (QTLs) Associated with Oil and Protein Contents in Soybean (*Glycine max* L.)

Hyeun Kyeung Kim* and Sung Taeg Kang[1]

Yeongnam Agricultural Research Institute, NICS, RDA, Milyang, 627-130, Korea
[1]Research Coordination Division, RDA, Suwon, 441-707, Korea

Received March 11, 2004 / Accepted May 11, 2004

Soybean oil and protein contents are very important as a nutritional component of food. The seed composition as oil and protein are polygenic traits. In this study, the Keunolkong×Iksan10 populations were evaluated with SSR markers to identify QTLs related to oil and protein contents. Three related independent QTLs near the marker satt100 on LG C2, satt546 on LG D1b+W and satt418 on LG L were identified oil contents. The three independent QTLs near the marker satt556 on LG B2, satt414 on LG J and satt238 on LG L were identified of protein contents. In the results of this study, common QTLs on LG L was associated with seed oil and protein contents. In the result of this study, it is believed that the seed composition material as oil and protein contents were mainly controlled by environmental stresses and they are seed size on genotypes.

Key words – oil contents, protein contents, Soybean, QTLs,

Soybean is legume crop and an excellent food and feed source all over the world. Grain legumes are rich and low-cost sources of dietary proteins and nutrients for a large part of the world's populations. The total world production of soybean is about 130 million tons. Soybean seed composition are affected by genotype, location and year effects however, the relative contribution of each of these factors varies with seed component evaluated, soybean type and geographical area [1, 16].

Recent developments in the use of the molecular markers can make plant breeding more efficient [15]. Marker assisted selection can serve as a tool to substantially increase the efficiency of selecting appropriate genotype. The advance in molecular genetics have made possible the genetic dissection and characterization of many quantitatively inherited seed quality traits in soybean.

Since soybean seed is a major source of vegetable protein and oil, numerous researchers have studied the inheritance of protein and oil soybean. Meanwhile, a large number of reports shown on the molecular mapping of characters such as seed oil and protein contents [2,5,8,9,11, 12,17]. The more recently, several researches were conducted using molecular markers to identify genomic regions significantly associated with QTLs controlling sucrose contents [13], seed isoflavone [14].

***Corresponding author**
Tel : +82-55-350-1239, Fax : +82-55-353-3050
E-mail : hkkkim@freechal.com

The objective of this study was to use molecular markers identify and characterize QTLs for controlling seed oil and protein contents in F_2 derived F_{10} RIL populations.

Material and methods

Plant materials and Field Evaluation

Two well-characterized soybean cultivars, Keunolkong and Iksan10 were used as mapping parents. Keunolkong is the pure-line derived from a local variety selected in Korea. Meanwhile, Iksan10 is the typical cultivar released from systemic breeding programs through the deliberate crossings of KW552×Pangsakong.

The cross Keunolkong × Iksan10 generated 115 F_{10} RILs that were derived from individual F_2 plants by single seed descent. This was referred to as the K/I population. The F_{10} seeds of each line were planted in a randomized complete block design with two replications at Yeongnam Agricultural Research Institute, NICS, Milyang, Korea in 2001. Oil and protein contents were measured for the QTLs study.

Measuring oil and protein contents

The protein contents of soybean were determined by auto-kjeldahl system. 0.2 g of ground sample was digested by Buchi B-435 digestion system and Buchi B-412 scrubber with 20 ml of sulfuric acid and 3 g of catalyst ($CuSO_4$: K_2SO_4 = 1 : 9). The percent of nitrogen was calculated by

Buchi B-339 auto-kjeldahl system and then converted to percent protein by multiplication by 6.25. The oil contents were determined by auto-soxhlet method with Buchi B-811 extracted system. The ground samples of 2 g was extracted by hexane for 2 hours, pre-heated for ten minutes, and then dried 1 hour at 105℃. This condition was confirmed in preconditioning experiment (data not shown). The moisture contents were analyzed by oven-dry method with 105℃ for 2 hours, and then all protein contents were converted to dry matter base.

DNA isolation and analysis

Genomic DNA was isolated from healthy leaves following the procedure described by Keim et al. [7]. The DNA solution was diluted to a working concentration with TE buffer (pH 8.0), and quality and intactness of DNA were checked using an agarose gel electrophoresis, and stored at -20℃ until use. A total of 199 soybean SSR markers were used to screen polymorphisms between mapping parents [4]. The primer pairs showing parental polymorphisms against K/I population was used for SSR genotyping in RIL progeny. The PCR reaction was performed in a total volume of 10 μl containing 25 ng of template DNA, 0.15 M of each forward and reverse primer sets, 200 μM of each dNTP, 2 mM $MgCl_2$, 0.1% Triton X-100, 1×reaction buffer (10 mM Tris-HCl pH 8.5, 100 mM KCl) and 0.5 U of *Taq* DNA polymerase. Template DNA was initially denaturated at 94℃ for 2 min, followed by 40 cycles for PCR amplification using the following conditions; denaturation at 94 ℃ for 25 sec, annealing at 47℃ for 25 sec and extension at 68℃ for 60 sec on a 96-Well GeneAmp PCR system 9700 (Applied Biosystem, Forster city, CA, USA). The segregation patterns for each SSR marker in both populations were determined by electrophoresis on polyacrylamide or agarose gels. Following electrophoresis, the gel was stained with the silver sequencing kit (Promega, Madison, WI, USA) or ethidiun bromide and scored for map construction. Pigmentation colors of flower and hilium were also scored as morphological markers.

Map construction and statistical analysis

Based on the segregation data subsets of SSR and morphological markers, linkage maps were individually constructed using the MapManager QT program [10]. Recombination fractions were converted to map distances by applying the Haldane map function [6]. Where possible, linkage groups were named according to the designations of the consensus USDA map [4].

Map positions for QTLs were determined by analysis of variance with Statistical Analysis System (SAS) version 8.03 [18]. Analysis of marker QTL associations was investigated by single factor analysis of variance (SF-ANOVA) in which marker-genotype groups were used as class variance. The means for each marker allelic group were compared using an F-test from the type-III mean squares obtained from the GLM procedure of SAS. When two more markers showing significant association with morphological characters linked to the same chromosome, multiple regression analysis (SLG-Regr) was conducted. Forward and stepwise selection procedures were applied in the regression analysis. All significant markers from SLG-Regr and unlinked single marker identified from SF-ANOVA were combined in a multiple linkage group regression model (MLG-Regr) at $P<0.05$ to determine the proportion of phenotypic variance explained by each QTLs. Phenotypic correlations were calculated with SAS.

Results

Phenotypic evaluation

Variation of seed oil and protein contents of K/I RIL populations are shown in Table 1 and Fig. 1. In the contents of oil and protein, Keunolkong showed the higher protein contents (431.2 mg) and lower oil contents (173.2 mg) than Iksan10 having protein contents (389.0 mg) and oil contents (209.1 mg). The frequency distribution of oil and protein contents among the F_2 derived F_{10} RIL of K/I showed a continuous distribution (Fig. 1). The between seed oil and protein contents showed the highest negative correlation.

Table 1. Variation of oil and protein contents in K/I populations

Traits	Parents		K/I population	
	Keunolkong	Iksan10	Mean±SD	Range
Oil (mg/g)	173.2	209.1	190.7±12.4	159.9~231.9
Protein (mg/g)	431.2	380.9	404.6±13.0	367.3~433.9

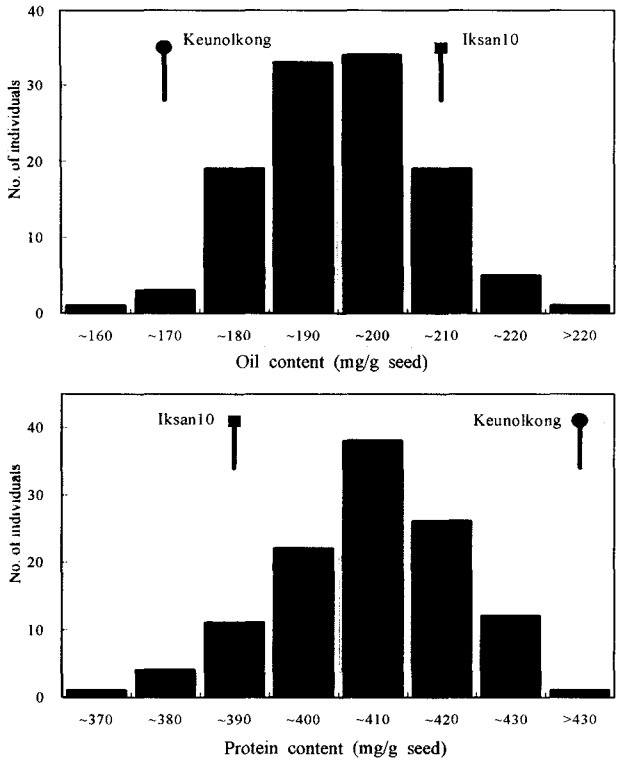

Fig. 1. The frequency distribution of oil and protein contents in K/I population.

Seed oil contents

The seed oil contents of Keunolkong was average 35.9 mg/g smaller than that of Iksan10. There were F_2-derived F_{10} lines with up to 13.3 mg oil contents than Keunolkong, and 22.8 mg larger oil contents than Iksan10 (Table 1). Based on the SF-ANOVA in K/I population, seven markers were detected as potentially linked to oil contents (Table 2). Individually, these markers accounted from 4.02% to 6.67% of the phenotypic variation. The three markers alleles of Iksan10 on linkage groups (LG) J and L increased the oil contents while the others were controlled by the Keunolkong allele. The SLG-Regr analysis identified markers satt100 on LG C2 and satt418 on LG L related to oil contents. When the markers on each linkage group were subjected to MLG-Regr analysis, three independent QTLs near the marker satt100 on LG C2, satt546 on LG D1b+W and satt418 on LG L were identified with accounting of 4.97%, 6.10% and 7.47% of phenotypic variation, respectively. Moreover, their individual QTLs showed a relatively low phenotypic explanation though they accounted total phenotypic variation of 18.51% for oil contents.

Seed protein contents

Keunolkong and Iksan10 differed by 50.3 mg/g in seed protein contents. Wide variation (367.3~433.9 mg) occurred in protein contents among progenies (Table 1). The SF-ANOVA analysis identified twelve markers as potentially associated with protein contents (Table 3). Individually, these markers accounted for 3.93% to 11.12% of the protein contents variation. The positive allele of the LG A2, J, L and M were derived from Keunolkong, whereas the other LGs (B2 and O) were derived from Iksan10. In the SLG-Regr model, the marker interval of satt414 on LG J was identified to be associated with protein contents. However, three putative QTLs associated with protein contents were detected by MLG-Regr analysis, three independent QTLs near the marker satt556 on LG B2, satt414 on LG J and satt238 on LG L were identified with accounting of 4.27%, 4.09% and 7.80% of phenotypic variation, respectively.

Table 2. Marker linked to QTLs associated with oil contents in F_2 derived F_{10} RILs of Keunolkong×Iksna10

Markers	LG	SF-ANOVA[a]		Allelic means		SLG-Regra		MLG-Regra	
		P	R^2 (%)	K/K[c]	I/I[c]	*P*	R^2 (%)	*P*	R^2 (%)
satt460	C2	0.0318	4.20	193.3	188.2	-	-	-	-
satt100	C2	0.0156	5.59	193.7	187.7	0.0069	7.13	0.0210	4.94
satt546	D1b+W	0.0070	6.67	193.9	187.5	NA[b]	-	0.0125	6.10
satt183	J	0.0165	5.16	188.5	194.2	NA	-	-	-
satt523	L	0.0413	4.02	189.9	190.1	-	-	-	-
satt418	L	0.0093	6.27	189.2	197.1	0.0202	5.49	0.0074	7.47
satt152	N	0.0331	4.10	194.1	189.0	NA	-	-	-
Total									18.51

[a]SF-ANOVA: single factor analysis of variance.
SLG-Regr : multiple regression with markers on each linkage group.
MLG-Regr: multiple regression with all significant markers from the SLG-Regr model.
[b]NA: Not applicable. Not linked to other markers.
[c]K/K : Keunolkong, I/I : Iksan10.

Table 3. Marker linked to QTLs associated with protein contents in F_2 derived F_{10} RILs of Keunolkong×Iksna10

Markers	LG	SF-ANOVA[a]		Allelic means		SLG-Regra		MLG-Regra	
		P	R^2 (%)	K/K[c]	I/I[c]	P	R^2 (%)	P	R^2 (%)
satt187	A2	0.0154	5.26	407.5	401.6	NA[b]	-	-	-
satt556	B2	0.0050	7.08	400.0	407.0	NA	-	0.0324	4.27
satt596	J	0.0143	5.91	408.4	402.1	-	-	-	-
satt529	J	0.0341	4.02	407.2	402.0	-	-	-	-
satt414	J	0.0164	5.08	407.2	401.4	0.0014	10.57	0.0328	4.09
satt380	J	0.0203	4.76	407.2	401.6	-	-	-	-
satt183	J	0.015	5.3	407.0	401.0	-	-	-	-
sct 001	J	0.0053	6.86	407.0	399.8	-	-	-	-
satt547	J	0.0398	3.93	401.1	406.2	-	-	-	-
satt238	L	0.0003	11.12	407.5	398.1	NA	-	0.0049	7.80
satt590	M	0.0034	7.73	407.6	400.3	NA	-	-	-
satt243	O	0.0215	4.84	401.8	407.5	NA	-	-	-
Total									16.17

[a]SF-ANOVA, SLG-Regr, MLG-Regr, [b]NA, [c]K/K and I/I are represented as shown in Table 1.

Moreover, their individual QTLs showed a relatively low phenotypic explanation though they accounted total phenotypic variation of 16.17% for protein contents.

To identify the interaction between each marker, two-way ANOVA was conducted with all possible two-way combinations (Table 4). Four combinations (satt556/satt596, satt529/sct001, satt380/sct001 and satt547/satt238) showed significant interactions. The interaction effects between ssatt556/satt596 resulted in higher protein contents for lines with Iksan10/Keunolkong alleleic configurations. For combinations between satt529 and sct001, and satt380 and sct001; the allele combination of Keunolkong/Iksan10 resulted in higher protein contents. Otherwide, the interaction with satt547/satt238 a line with Keunolkong allele at satt547, and Iksan10 allele at satt547 resulted lower protein contents.

Discussion

Soybean seed protein and oil concentrations were more polygenic traits [2], and are highly influenced by environmental factors [17]. Molecular markers can be used to characterize QTLs conditioning these traits. However, only limited information is available on association of DNA markers and soybean seed component traits [5,8,9,11,12,17] data showed that different marker combinations could have different efficiencies in marker assisted selection (MAS). In theory, the more markers used the more reliable MAS

Table 4. Epistatic interactions between two markers associated with the protein contents in Keunolkong×Iksan10

SSR locus	Allele	Allele/locus		P	R^2(%)
		Keunolkong	Iksan10		
satt556 / satt596		satt556		0.0277	18.95
satt596	Keunolkong	400.3	413.2		
	Iksan10	400.4	402.4		
satt529 / sct001		satt529		0.0156	12.08
sct001	Keunolkong	406.6	408.4		
	Iksan10	420.5	398.6		
satt380 / sct001		satt380		0.0056	13.45
sct001	Keunolkong	406.5	408.5		
	Iksan10	423.6	398.4		
satt547 / satt238		satt547		0.0327	20.66
satt238	Keunolkong	405.0	408.7		
	Iksan10	388.3	402.8		

would be useful because various markers could simultaneously select desirable soybean genotypes based on all these different markers. This may eliminate the selection of false positive based on only one marker. In breeding practices, consideration has to be given to the number of markers employed and the efficiency of different marker combinations may provide valuable information on choosing desirable marker combination of optimizing MAS in soybean breeding program.

In the results of this study, using the MLG-Regr analysis, three independent QTLs near the marker satt100 on LG C2, satt546 on LG D1b+W and sat418 on LG L were identified with oil contents. In the protein contents, three independent QTLs near the marker satt556 on LG B2, satt414 on LG J and satt238 on LG L were identified. LG L was QTLs that appeared commonly in oil and protein contents. Mansur et al. [12] reported that oil and protein contents identified with common marker on LG U7 and L formerly described as LG U14).

On the other hand, Diers et al. [5] reported that oil contents identified nine marker loci (LG E and F), and protein contents identified eight marker loci (LG E, F and) in an interspecific G. *max*×G. *soja* cross as having significant effects. Lee et al. [9] reported that the two populations were various common markers on LG C1, G, J. K and L identified for seed oil contents, and the common markers on LG C1, E, J, K, and N identified for seed protein contents. In this way, some investigators reported same result and a little different from result of this study.

These a litter of common markers for seed oil and protein contents in the Keunolkong×Iksan10 would explain the lack of association between seed oil and protein contents in this population, but does not preclude the possibility that existing markers associated with both traits were not uncovered in this study. If colligated result of this study, it is believed that the seed composition material as oil and protein contents were mainly controlled by environmental stresses and seed size on genotypes.

Reference

1. Bhardwaj, H. L., A. S. Bhagsari, J. M. Joshi, M. Rangappa, V. T. Sapra and M. S. Rao. 1999. Yield and quality of soymilk and tofu made from soybean genotypes grown at four locations. *Crop Sci.* **39**, 401-405
2. Brummer, E. C., G. L. Graef, J. Orf, J. R. Wilcox and R. C. Shoemaker. 1997. Mapping QTL for seed protein and oil contents of eight soybean populations. *Crop Sci.* **37**, 370-378.
3. Burton, J. W. 1987. Quantitative genetics: results relevant to soybean breeding. pp211-247. *In*: Wilcox J.R. (ed) Soybeans: improvement, production, and use. (Agronomy, series, vol. 16) American Society of Agronomy Madison Wis.
4. Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, N. Kaya, T. T. VanToai, D. G. Lohnes, J. Chung and J. E. Specht. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci.* **39**, 1464-1490.
5. Diers, B. W., R. Keim, W. R. Fehr and R. C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil contents. *Theor. Appl. Genet.* **83**, 608-612.
6. Haldane, J. B. S. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* **8**, 299-309.
7. Keim, P., T. C. Olson and R. C. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. *Soybean Genet. Newsl.* **15**, 150-154.
8. Lark, K. G., J. Orf and L. M. Mansur. 1994. Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. *Theor. Appl. Genet.* **88**, 486-489.
9. Lee, S. H., M. A. Bailey, M. A. R. Mian, T. E. Carter, E. R. Shipe, D. A. Ashley, W. A. Parrott, R. S. Hussey and H. R. Boerma. 1996. RFLP loci associated with soybean seed protein and oil content across populations and locations. *Theor. Appl. Genet.* **93**, 649-657.
10. Mainly, K. F. and J. M. Olson. 1999. Overview of QTL mapping software and introduction to Map Manager QT. *Mamm. Genome* **10**, 327-334.
11. Mansur, L. M., K. G. Lark, H. Kross and A. Oliveira. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). *Theor. Appl. Genet.* **86**, 907-913.
12. Mansur, L. M., J. H. Orf, K. Chase, T. Jarvik, P. B. Cregan and K. G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean [*Glycine max* (L.) Merr.]. *Crop Sci.* **36**, 1327-1336
13. Maughan, P. J., M. A. Saghai Maroof and G. R. Buss. 2000. Identification of quantitative trait loci controlling sucrose content in soybean (*Glycine max*). *Molecular Breeding* **6**, 105-111.
14. Meksem, K., V. N. Njit, W. J. Banz, M. J. Iqbal, My. M. Kassem, D. L. Hyten, J. Yuang, T. A. Winters and D. A. Lightfoot. 2001. Genomic regions that underlie soybean seed isoflavone content. *J. Biomed Biotech.* **1**, 38-44.
15. Paterson. A. H. and S. D. Tanksley. 1991. DNA markers in plant improvement. *Adv. Agron.* **46**, 39-89.
16. Poysa, V. and L. Woodrow. 2002. Stability of soybean seed composition and its effect on soymilk and tofu yield and quality. *Food Research International* **35**, 337-345.
17. Qiu B. X., P. R. Arell and D. A. Sleper. 1999. RFLP markers associated with soybean cyst nematode resistance

and seed composition in a 'Peaking'×'Essex' population. *Theor. Appl. Genet.* **98**, 356-364.

18. SAS Institute, Inc. 1990. *SAS/STAT user's guide, version 6.4th ed., vol.2.* SAS institute, Cary, North Carolina, USA.

초록 : 콩에서 Microsatellite marker를 이용한 oil 및 단백질 함량의 양적형질 유전자좌의 분석

김현경* · 강성택[1]

(농촌진흥청 작물과학원 영남농업연구소, [1]농촌진흥청 연구개발국 연구관리과)

콩의 oil 및 단백질은 식품에서 매우 중용한 영양학적인 구성요소이다. Oil 및 단백질과 같은 종자 구성물질들은 polygenetic 형질들로 되어있다. 본 시험은 큰올콩과×익산10호의 RIL 계통과 SSR marker를 이용하여 유전자 지도를 작성하고, 이를 바탕으로 oil 및 단백질 함량과 관련된 양적형질 유전자좌(QTLs)를 탐색하였다. Oil 함량과 관련된 QTLs는 연관군 C2와 satt100과 연관군 D1b+W의 satt546 및 연관군 L의 satt418의 세 개의 독립적인 QTLs를 확인하였다. 단백질 함량에 있어서는 연관군 B2와 J 및 L에 각각 satt556과 satt414 및 satt238의 marker에서 독립적인 QTLs를 확인하였다. 본시험의 결과, oil 및 단백질 함량과 관련된 공통의 QTL은 연관군 L이었다. 한편, oil 및 단백질과 같은 종자구성물질은 주로 환경적인 stress 및 종자의 크기 등에 의해서 구성되어지는 것으로 생각된다.