

Simple Sequence Repeat Markers Linked to Quantitative Trait Loci Controlling Seed Weight, Protein and Oil Contents in Soybean

Hyeun-Kyeong Kim^{†1*}, Sung-Taeg Kang², Myoung-Gun Choung³, Chan-Sik Jung¹, Ki-Won Oh¹, In-Youl Baek¹ and Beung-Gu Son⁴

¹Yeongnam Agricultural Research Institute, NICS, RDA, Miryang 627-130, Korea

²Breeding Resource Development Division, NICS, RDA, Suwon, 441-857, Korea

³Dept. of Pharmacognosy Material Development, Kangwon National University, Sancheok 245-711, Korea

⁴School of Resources & Life Science, Pusan National University, Miryang 627-702, Korea

Received June 29, 2006 / Accepted September 15, 2006

Soybean [*Glycine max* (L.) Merr.] is an important crop, accounting for 48% of the world market in oil crops. Improvement of the quality and quantity of soybean seed constituents is one of the most important objectives in soybean breeding. Protein content and seed size are important properties to determine the quality of tofu and soy sprouts respectively. The objective of this study was to identify quantitative trait loci (QTLs) that control seed weight, protein and oil content in soybean. The 117 F_{2:30} recombinant inbred lines (RIL) developed from a cross of 'Keunolkong' and 'Shinpaldalkong' were used. Narrow-sense heritability estimates based on a plot mean on seed weight, protein and oil content were 0.8, 0.78 and 0.71, respectively. Four independent QTLs for seed weight were identified from linkage group (LG) F, I and K. Five QTL for protein content were located on LG D1b, E, H, I and L. Oil content was related with six QTLs located on LG D1b, E, G, I, J and N. Protein and oil content have three common QTLs on LG D1b, E and I. Thus, we identified major loci improving soybean seed quality.

Key words – Soybean, seed weight, protein, oil, QTL, SSR marker

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. Soy foods prepared from soybean include textured soy protein, soymilk, tofu, natto, denjang, miso, soy sprouts, and soy sauce. Seed traits such as protein content and seed weight have important roles in determining the quality of these soy food items[4,8].

Hardiness, brittleness, and gumminess are important physical properties, which determine the overall quality of tofu. Seed protein content determines these properties in tofu, whereas seed size determines the quality of soy sprouts[14]. Small seeded soybeans are generally desired for high quality soy sprouts, denjang and natto production, and combining higher protein with large seed size is de-

sired for tofu production[20]. Genomic regions for seed weight along with protein and oil quantitative trait loci (QTLs) could be used in marker-assisted selection (MAS) for desired soybean types for soy food application.

Despite moderately high heritabilities[2], it is difficult to improve seed traits, particularly protein and oil content, simultaneously. Also, numerous researchers[1,3,5,6,12,13, 16-18,21,22] have studied the inheritance of seed size, and protein and oil content in soybean. QTLs for oil content are located in many linkage group except linkage group (LG) D1b, M and N. QTLs for protein content are also spread among several linkage groups except LG D1b and D2. Also, several seed size QTLs have been identified but there are very few consistent QTLs across the populations. Breeder's use of such QTLs is more likely only after they are confirmed in diverse populations grown in diverse environments.

The primary objective of the present research was to improve the breeding efficiency of seed quality in soybean. This study was conducted to identify simple sequence repeats (SSR) markers associated with QTLs and marker-assisted selection (MAS) for seed weight, protein and oil contents in soybean.

*Corresponding author

Tel : +82-51-240-7527, Fax : +82-51-247-3549

E-mail : hkkkim@freechal.com

[†]Present address: Bioresources Development Institute, Pusan National University, Miryang 627-702, Korea

Materials and Methods

The well-characterized soybean cultivars, 'Keunolkong' and 'Shinpaldalkong' were used as mapping parents. 'Keunolkong' is the pure-line derived from a local variety selected in Korea. It possesses early maturity, short stem length, and large seed size. 'Shinpaldalkong' is the typical cultivars released from systemic breeding programs through the deliberate crossings, 'Will'×('Elf'× SS74185). 'Shinpaldalkong' showed the late maturity, long stem length, and small seed size.

The RIL populations were obtained from the crosses of 'Keunolkong' and 'Shinpaldalkong'. The cross 'Keunolkong'×'Shinpaldalkong' (K/S) generated 117 F₁₀ RILs derived from individual F₂ plants by single seed descent (SSD). The F₁₀ 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. Each entry was planted in a 1.5 m long paired-row plot with two seeds per hill. Spacing was 60 cm between rows and 10 cm between plants. Seed weight was determined by weighting 100 seeds per plot.

Protein and oil contents

The protein contents of soybean were determined by automated kjeldahl method. point two gram of ground seed sample was digested by a B-435 digestion system and a B-412 scrubber (Buchi) with 20 ml of sulfuric acid and 3 grams of catalyst (CuSO₄ : K₂SO₄ = 1 : 9). Percent nitrogen was calculated by a B-339 auto-kjeldahl system (Buchi) and then converted to percent protein by multiplication with the constant of 6.25.

The oil content was determined by an automated soxhlet method with a B-811 extraction system (Buchi). Two grams of ground sample was extracted by hexane for 2 h, pre-heated for 10 min, and then dried 1 h at 105°C. The moisture contents were analyzed by an oven-dry method with 105°C for 2 h, and then all data were converted to dry matter basis.

DNA isolation and analysis

Genomic DNA was isolated from fresh leaves following the procedure described by Keim et al. [11]. A total of 199 soybean SSR markers (<http://soybase.agron.iastate.edu/ssr.htm>) were used to screen for polymorphisms between mapping parents. The primer pairs showing parental poly-

morphisms were used for SSR genotyping in RIL progenies. The PCR reaction was performed in a total volume of 10 μ l containing 25 ng of template DNA, 0.15 M of each forward and reverse primers, 200 M of each dNTP, 2 mM MgCl₂, 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 (BioBasic *Taq* Polymerase, Applied Bio Basic, Canada). Template DNA was initially denaturated at 94°C for 2 min, followed by 40 cycles of PCR amplification using the following conditions; denaturation at 94°C for 25 sec, annealing at 47°C for 25 sec and extension at 68°C for 60 sec on a 96-Well GeneAmp PCR system 9700 (Applied Biosystems, USA).

Map construction and statistical analysis

Means of traits, correlation, and analysis of variance were determined by SAS program (SAS, 2002). Narrow-sense heritability was calculated on a per-plot basis, using estimates of the variance components [7]. Based on the segregation data subsets for SSRs and morphological markers, we constructed a linkage map with MapManager QT version 2.8 software [15]. Recombination fractions were converted to map distances by applying the Haldane map function [9].

The association between marker and QTL was tested according to the interval mapping methods of Whittaker et al. [27], using MapManager QT and single-factor ANOVA (SF-ANOVA). For each SSR and morphological marker, the class means for seed weight, protein and oil concentration were compared for significance ($P < 0.05$) using an *F*-test from the Type III mean squares, as obtained from the General Linear Model (GLM) procedure of SAS.

If SF-ANOVA identified two or more linked markers associated with the seed weight, protein and oil content, multiple regression analysis was conducted by including all the significant markers on that linkage group in the model (SLG-Regr). All significant markers from the SLG-Regr analysis were then combined into a multiple-linkage group regression (MLG-Regr) at $P < 0.05$ to determine the combination of independent markers those were explaining the greatest amount of phenotypic variation in a given trait. The probability level of 0.05 was selected to enhance our ability to detect QTLs associated with seed weight, protein and oil concentration. Finally, the coefficient of determination (R^2) obtained from MLG-Regr was used to provide an estimate of the percentage phenotypic variation explained by the markers.

Results

Seed weight, protein and oil contents

Variation of seed weight, protein and oil content in the F_{2:10} RIL population is presented in Table 1 and Fig. 1. 'Keunolkong' shows large seed weight and high protein content, while 'Shinpaldalkong' is characterized with small seed weight and high oil content. There was significant difference ($p < 0.05$) among the RIL for each traits in the population. Transgressive segregation for seed weight, protein and oil content, was found a few lines being significantly greater or lower than high and low parents (Fig. 1). Seed weight ranged from 11.6 to 30.35 g (per 100 seeds weight) and protein content ranged from 35.57 to 45.04%, and oil content ranged 16.82 to 24.56%. Narrow-sense heritability of seed weight, protein and oil content were 0.80, 0.78 and 0.71, respectively (Table 1). There was a strongly negative correlation between protein and oil content ($r = -0.57$, $p < 0.001$). On the other hand, protein content was strongly positively correlated with seed weight ($r = 0.454$, $p < 0.001$) (Table 2).

Construction of linkage map based on RIL

Based on the RIL F₁₀ generation, a genetic linkage map with 108 SSR markers and two morphological markers (total 110) was constructed. The map covered a distanced 1,890 cM of the soybean genome using the Haldane function, comprising 19 linkage groups. On average, this map revealed a marker density of 1.0 per 17.2 cM. The order of most of the markers was in agreement with the public soybean molecular linkage map [23].

QTL analysis for seed traits

The SF-ANOVA analysis identified fifteen markers (data not show) as potentially associated with seed weight. MLG-Regr analysis showed that four QTL markers, satt146 on LG F, satt571 on LG I and satt137 and satt417 on LG K were significantly associated with seed weight (Table 3).

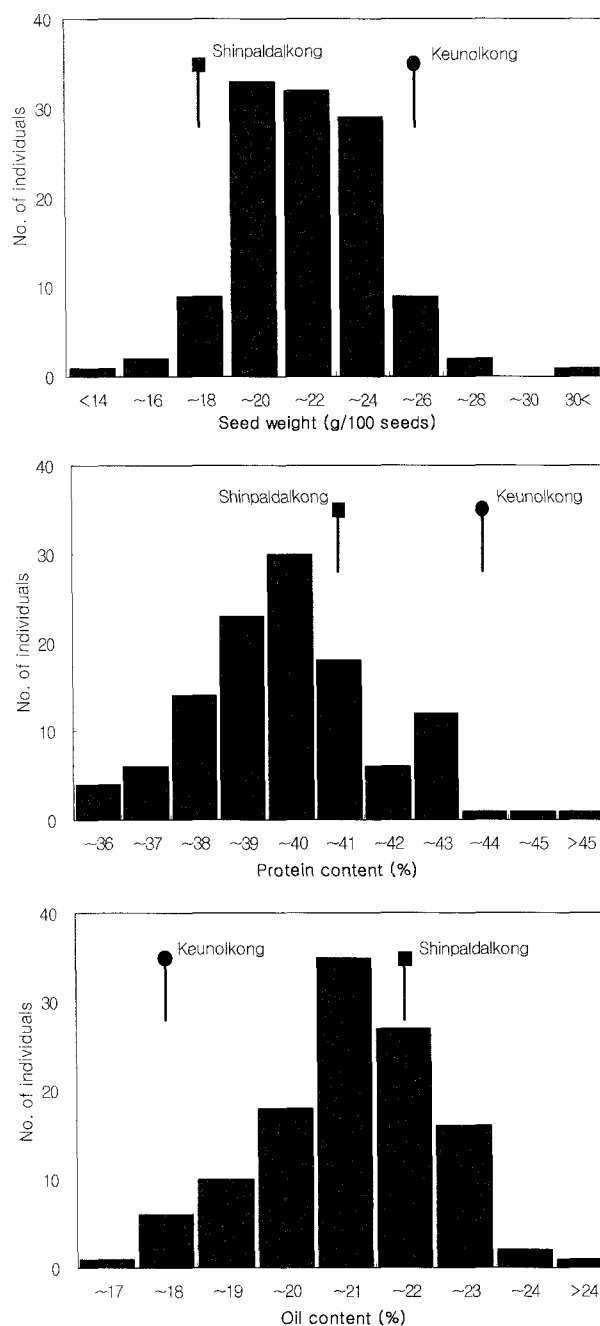


Fig. 1. Frequency distribution of seed weight (g/100 seeds), protein and oil contents (%) in 117 RILs of 'Keunolkong' × 'Shinpaldalkong'.

Table 1. Descriptive statistics for seed weight, protein and oil contents in 117 RIL population of 'Keunolkong' × 'Shinpaldalkong'

Traits	Parents		RIL population		h^{2*}
	Keunolkong (mean ± SD)	Shinpaldalkong (mean ± SD)	Range	Mean ± SD	
Seed weight (g/100 seeds)	24.70 ± 0.29	16.83 ± 0.87	11.60 ~ 30.35	21.02 ± 2.54	0.80
Protein (%)	43.12 ± 0.56	40.10 ± 0.70	35.57 ~ 45.04	39.45 ± 1.86	0.78
Oil (%)	17.32 ± 0.37	20.10 ± 0.80	16.82 ~ 24.56	20.60 ± 1.48	0.71

* Heritability on a per-plot basis

Table 2. Correlation coefficients among seed weight, protein and oil contents in 117 RILs of 'Keunolkong' × 'Shinpaldalkong'

	Seed weight	Protein
Protein	0.454 ^{***}	
Oil	-0.059 ^{ns}	-0.570 ^{***}

*** Significant at 0.001 probability level.
ns, Not significant ($P > 0.05$)

Individual QTLs explained relatively low phenotypic variation (3.63~8.11%) though they accounted for total phenotypic variations of 23.49% for seed weight.

The SF-ANOVA revealed that thirty-three markers (data not show) were detected as potentially associated with protein content. MLG-Regr analysis identified five QTLs on LG D1b, E, H, I and L (Table 3). Two major QTLs, satt151 on LG E and satt440 on LG I, explained 29.85 and 13.83% of the phenotypic variation, respectively.

Based on the SF-ANOVA analysis in th RIL population, twenty-six markers (data not show) were found significantly ($p < 0.05$) associated with oil content. Six QTLs were identified on LG D1b, E, G, I, J and N. In the MLG-Regr analysis, those QTLs explained 52.04% of total phenotypic variation of the oil content. The QTL located near the marker satt216 on LG D1b and satt151 on LG E were detected as major QTL contributing to $R^2 = 19.03$ and 11.53%, respectively (Table 3).

Discussion

Seed weight, protein and oil contents are main factors determining seed quality in soybean. Smaller seed weight is preferred in Asian markets for quality sprouts and natto, whereas large seed weight with high protein content in preferred for tofu and denjang (fermented soybean) production[14,24].

Our population is a possibility of increasing the frequency of desirable alleles for protein or oil and for smaller and larger seed weight, since the narrow-sense heritabilities for these traits are moderately high (Table 1). The narrow-sense heritabilities for seed weight, protein and oil contents found in K/S population indicated that much (0.80, 0.78 and 0.71, respectively) of the variation was genetic. The narrow-sense heritabilities for seed weight, protein and oil contents are higher than that of Panthee et al. [20] reported while Chang et al. [3] reported higher heritabilities for protein and oil (0.89 and 0.84, respectively). The heritability observed in our population for indicated that selection response would be reasonable for achieving genetic gain.

The negative phenotypic correlation ($r = -0.57$, $P < 0.001$) between protein and oil content and the positive phenotypic correlation ($r = 0.454$, $P < 0.001$) between protein and seed weight are agreed with previous report [3,17,20]. According to our result and previous reports, selection for the three characteristics is somehow difficult to improve a

Table 3. Marker distributions and QTLs associated with seed weight, protein and oil contents in 117 RIL of 'Keunolkong' × 'Shinpaldalkong'

Traits	LG	Markers	SF-ANOVA [†]		Allelic means		MLG-Regra	
			P	R ² (%)	K/K [‡]	S/S [‡]	P	R ² (%)
Seed weight (g/100 seeds)	F	satt146	0.024	4.84	21.60	20.46	0.006	8.11
	I	satt571	0.009	5.79	20.43	21.64	0.018	5.71
	K	satt137	0.029	4.10	21.50	20.48	0.014	5.84
	K	satt417	0.003	7.60	21.63	20.31	0.041	3.83
Protein (%)	D1b	satt157	0.001	11.66	40.13	38.81	0.015	6.06
	E	satt151	<0.0001	13.14	40.30	38.92	<0.0001	29.85
	H	satt442	0.026	4.35	39.82	39.05	0.017	5.25
	I	satt440	0.002	8.22	38.95	40.03	0.001	13.83
	L	satt238	<0.0001	16.49	40.22	38.75	0.025	4.29
Oil (%)	D1b	satt216	<0.0001	18.91	20.05	21.05	<0.0001	19.03
	E	satt151	<0.0001	15.09	19.89	21.07	0.001	11.53
	G	satt324	<0.0001	14.69	20.11	21.27	0.007	6.32
	I	satt571	0.016	5.04	20.27	20.93	0.010	4.66
	J	satt285	0.003	7.70	20.20	21.02	0.009	5.42
	N	satt521	0.004	7.27	20.23	21.02	0.014	4.49

[†] SF-ANOVA: single-factor analysis of variance, MLG-Regr: multiple regression with all significant markers from the SLG-Regr model

[‡] K/K: Keunolkong, S/S: Shinpaldalkong.

trait without affect other traits.

Seed weight, measured as mass per seed, is an important yield component of soybean and is generally positively correlated with seed yield[2]. Seed size of soybean is one of the most important traits for the criteria of soybean usage, soy sprouts, tofu, denjang, natto and so on. Although the heritability is high, seed weight is very easily affected by environmental conditions. Previous reports showed that polygenic influence results in a large numbers of QTLs[1,13,26]. There was four molecular markers associated with seed weight QTLs in K/S population located on three LG F, I and K. Compared with the previous study, LG F, I and K might be almost identical with QTLs previously reported[10,17,19].

Results of this study using the MLG-Reg analysis, six independent QTLs on LG D1b, E, G, I, J and N were significantly related with oil content. In the protein contents, five independent QTLs on LG D1b, E, H, I and L were identified. LG E appeared to be the major QTL control of these two compounds. However, oil and protein contents were controlled by satt216-satt157 on LG D1b and satt212-satt263 on LG E. QTLs for oil content were reportedly located in many linkage groups except LG D1b, M, and N . QTLs of protein contents were reported in all LG except in LG D1b and D2[1,5,6,13,16-18,21,22,25].

For both protein and oil content, we observed that different allele effects depended on the chromosomal background. For example, the 'Keunolkong'- derived marker allele on LG D1b and I increased the contents of protein. On the other side, reverse effect was seen with the oil contents. This result is exactly agreed with the result that the protein and oil contents are negatively correlated with each other. The both those contents are controlled in the same direction under the same genes. Nevertheless, we detected no tentative QTLs in LG D1b that were related to protein and oil contents. Therefore, further study is needed to confirm whether those QTLs are real or that, instead, environmental factors affected those results. Finally, we hope that the QTLs identified from this study and recent germplasm line that we have relased will benefit breeders in accumulation favorable alleles for improvement in soybean seed quality.

References

1. Brummer, E. C., G. L. Graef, J. Orf, J. R. Wilcox and R. C. Shoemaker. 1997. Mapping QTL for seed protein and oil content eight soybean populations. *Crop Sci.* **37**, 370-378.
2. Burton, J. W. 1987. Quantitative genetics: Results relevant to soybean breeding. In J. R. Wilcox (ed.) Soybeans: Improvement, production and uses, 2nd ed. Agron. Monogr. 16 ASA, CSSA, ANd SSSA, Madison, WI.
3. Chung, J., H. L. Babka, G. L. Graef, P. E. Staswick, D. J. Lee, P. B. Cregan, R. C. Shoemaker and J. E. Specht. 2003. The seed protein, oil and yield QTL on soybean linkage group I. *Crop Sci.* **43**, 1053-1067.
4. Clarke, E. J. and J. Wiseman. 2000. Development in plant breeding for improved nutritional quality of soybeans I. Protein and amino acids content. *J. Agric. Sci.* **134**, 111-124.
5. Csanadi, G, J. Vollmann, G. Stift and T. Lelley. 2001. Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor. Appl. Genet.* **103**, 912-919.
6. Diers, B. W., R. Keim, W. R. Fehr and R. C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. *Theor. Appl. Genet.* **83**, 608-612.
7. Frey, K.J. and T. Horner. 1957 Heritability in standard units. *Agron. J.* **49**, 59-62.
8. Friedman, M. and D. L. Brandon. 2001. Nutritional and health benefits of soy proteins. *J. Agric. Food Chem.* **49**, 1069-1086.
9. 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.
10. Hoeck, J. A., W. R. Fehr, R. C. Shoemaker, G. A. Welke, S. L. Johnson and S. R. Cianzio. 2003. Molecular marker analysis of seed size in soybean. *Crop Sci.* **43**, 68-74.
11. Keim, P., T. C. Olson and R. C. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. *Soybean Genet. Newsl.* **15**, 150-154.
12. Kim, H. K. and S. T. Kang. 2004. Identification of quantitative trait loci associated with oil and protein content in soybean (*Glycine max* L.). *J. Life Sci.* **14**, 453-458.
13. Lec, 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.
14. Liu, K. 1997. Soybeans: Chemistry, technology and utilization. Chapman and Hall, New York.
15. Mainly, K. F. and J. M. Olson. 1999. Overview of QTL mapping software and introduction to Map Manager QT. *Mamm. Genome* **10**, 327-334.
16. 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.
17. 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.

18. Orf, J. H., K. Chase, T. Jarvik, L. M. Mansur, P. B. Cregan, F. R. Alder and L. G. Lark. 1999a. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Sci.* **39**, 1642-1651.
19. Orf, J. H., K. Chase, F. R. Alder, L. M. Mansur and L. G. Lark. 1999. Genetics of soybean agronomic traits: II. Interaction between yield quantitative trait loci in soybean. *Crop Sci.* **39**, 1652-1657.
20. Panthee, D. R., V. R. Pantalone, D. R. West, A. M. Saxton and C. E. Sams. 2005. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci.* **45**, 2015-2022.
21. 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.
22. Sebort, A. M., R. C. Shoemaker and D. W. Diers. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci.* **40**, 1438-1444.
23. Song, Q. J., L. F. Marek, R. C. Shoemaker, K. G. Lark, V. C. Concibido, X. Delannay, J. E. Specht and P. B. Cregan. 2004. A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* **109**, 122-128
24. Son, Y. K., J. J. Hwang, S. L. Kim, Y. H. Ryu, D. C. Shin and J. Y. Yoo. 1997. Effect of soybean cultivars Korean traditional deonjang (soybean paste) processing. *Korea Soybean Digest* **14**, 27-36.
25. Specht, J. E., K. Chase, M. Macrander, G. L. Graef, J. Chung, J. P. Markwell, M. Germann, J. H. Orf and K. G. Lark. 2001. Soybean response to water: A QTL analysis of drought tolerance. *Crop Sci.* **41**, 493-509.
26. Watanabe S, T. Tajuddin, N. Yamanaka, M. Hayashi and K. Harada. 2004. Analysis of QTLs for reproduction development and seed quality traits in soybean using recombinant inbred lines. *Breed. Sci.* **54**, 399-407.
27. Whittaker, J. C., R. Thompson, D. M. Vissche. 1996. On the mapping of QTL by regression of phenotype on marker type. *Heredity* **77**, 23-32.

초록 : 콩에서 종실의 무게와 oil 및 단백질 함량을 조절하는 양적 형질 유전자좌와 연관된 simple sequence repeat marker

김현경*¹ · 강성택² · 정명근³ · 정찬식¹ · 오기원¹ · 백인열¹ · 손병구⁴

(¹농촌진흥청 작물과학원 영남농업연구소, ²농촌진흥청 작물과학원 작물기능개발과, ³강원대학교 한방보건복지대학 생약자원개발학과, ⁴부산대학교 생명자원과학부)

콩은 전 세계 시장의 48%를 차지하는 중요한 유류작물이다. 콩 종자를 구성하는 양적인 부분과 질적인 부분의 개선은 콩 육종 목표의 중요한 부분이다. 단백질함량과 종실의 크기는 두부와 콩나물의 질을 평가하는 중요한 특성이다. 따라서 본 연구는 콩에서 종실의 크기와 단백질 및 oil 함량을 조절하는 양적형질유전자좌(QTLs)를 확인하기 위하여 실시하였다. 시험재료로는 큰올콩과 싹발달콩을 교배한 후 F₂유래 F₁₀세대의 RIL을 이용하였으며, 이를 바탕으로 종실의 무게와 oil 및 단백질 함량과 관련된 QTLs를 탐색하였다. 종실의 무게와 단백질 및 oil 함량의 협의의 유전력은 각각 0.8과 0.78 및 0.71을 나타내었다. 종실의 무게와 관련된 QTLs는 연관군 F, I와 K의 세 개의 독립적인 QTLs를 확인하였다. 단백질함량과 관련된 QTLs는 연관군 D1b, E, H, I와 L의 다섯 개의 독립적인 QTLs를 확인하였다. 그리고 oil 함량과 관련된 QTLs는 연관군 D1b, E, G, I, J와 N의 여섯 개의 독립적인 QTLs를 확인하였다. oil 및 단백질 함량과 관련된 QTLs는 연관군 D1b, E와 I에서 공통적으로 나타났다. 따라서 이들 주요 QTL은 콩 품종 육성과정에서 품질의 개선하기 위한 선발 마커로서 활용가치가 높은 것으로 판단된다.