

## QTL Analysis of Protein Content in Double-haploid Lines of Rice

Yang Qin, Suk-Man Kim, and Jae-Keun Sohn<sup>†</sup>

Department of Agronomy, Kyungpook National University, Daegu, 702-701, Korea

**ABSTRACT** The objective of this study was to characterize the main-effect QTLs, epistatic QTLs and QTL-by-environment interactions (QE), which are involved in the control of protein content. A population of 120 doubled haploid (DH) lines derived from a cross between ‘Samgang’ and ‘Nagdong’, was planted and determined for protein content over three years. Based on the population and a genetic linkage map of 172 markers, QTL analysis was conducted by WinQTLcart 2.5 and QTLMAPPER. Three main-effect QTLs affecting protein content of brown rice were detected from 2004 to 2006 on chromosomes 1 and 11. The qPC11.2 was repeatedly detected across two years. Seven pairs of epistatic loci were identified on eight chromosomes for protein content and collectively explained 39.15% of phenotype variation. These results suggest that epistatic effects might be an even more important component of the genetic basis for protein content and that the segregation of the DH lines for protein content could be largely explained by a few main-effect QTLs and many epistatic loci.

**Keywords :** protein content, main-effect QTLs, epistasis, QE interactions

**Rice** is the principal source of protein in the most rice eating countries. Juliano (2003) conclude that rice is unique among cereals because it has a storage protein principally made of glutelin, which has a more balanced amino-acid profile compared to the prolamin-rich storage proteins. Brown rice prefers to have the higher protein than milled rice, because the aleurone layer, subaleurone layer and embryo are generally relatively rich in protein bodies compared with the starch of rice grains. Even so, there are differences between rice cultivars for loss of protein during polishing. Selection on high protein content retaining genotypes should be possible.

Knowledge accumulated in the past decades indicates that the protein content is controlled by quantitative trait loci (QTLs) showing continuous phenotypic variation in rice progeny (Yano & Sasaki, 1997). Due to the complicated and variable features of QTL, such as magnitude of effect, genomic position, environment effects and interactions, their locations and effects are difficult to characterize. However, the development of DNA markers created opportunities to characterize the quantitative traits. The chromosomal regions contain genes controlling quantitative traits have been identified using QTL mapping based on the linkage maps (Mohan *et al.*, 1997).

As genes, QTL effects may be environmentally sensitive and this sensitivity results in phenotypic plasticity or the ability of the organisms to take on alternative developmental fates, depending on environmental cues. Fan *et al.* (2005) considered that the major and minor genes are subject to environmental modifications, which can cause dramatic difference in the phenotypic effects of the genes. In addition, the epistasis refers to the phenotypic effects of interactions among alleles at multiple loci. Our current understanding of biochemical and physiological genetics, as well as the regulation of gene expression, strongly suggests the ubiquity of interactions among gene products (Xing *et al.*, 2002). To date, the epistasis for amylose content, grain weight and other traits related to yield were many reported (Li *et al.*, 1997; Xing *et al.*, 2002; Fan *et al.*, 2005), but not for the protein content. The characterization regarding the epistasis and QE interaction using primary mapping populations conduce to the identification of putative QTLs for further studies.

The objectives of this study are: 1) QTL identification for protein content of brown rice using a genetic linkage map based on 120 DH population from a cross between ‘Samgang’ and ‘Nagdong’. 2) The epistatic effects and

<sup>†</sup>Corresponding author: (Phone) +82-53-950-5711  
(E-mail) jhsohn@knu.ac.kr <Received November 17, 2008>

QTL-by-environment interactions (QE) of protein content in different environments were carried out to interpret the relationships of genes and environments.

## MATERIAL AND METHOD

### Plant materials and field trial

One hundred-twenty rice DH lines developed by anther culture of the F<sub>1</sub> from a cross between ‘Samgang’, a Tongil variety, and ‘Nagdong’, a japonica variety, were used as the mapping population in this study. The field trials were carried out in the rice-growing season on the experimental fields at Kyungpook National University in 2004, and Gunwie in 2005 and 2006, Korea. About 30-day-old seedlings of each DH line consisting of three rows with 20 plants per row were grown using a spacing pattern of 15×30 cm between plants and rows. Two parents were grown as the control between every ten DH lines. Field management was conducted according to the normal cultivation practices with an application for fertilizer at the rate of 110 kg N ha<sup>-1</sup>, 45 kg P<sub>2</sub>O<sub>5</sub> ha<sup>-1</sup>, and 57 kg K<sub>2</sub>O ha<sup>-1</sup>.

### Trait evaluation

The harvested paddy rice were collected from the second row of each DH line or the parent, dried naturally, and then stored at room temperature for at least 3 months with 12-13% of moisture controlled before analysis. The damage grains, red grains, green grains and broken grains were removed from the brown rice before test. The protein content (PC) of brown rice of three repeats for each line was determined by a near-infrared spectroscopy (NIRS) spectrophotometer (Foss 6500).

### QTL and data analysis

The composite interval mapping (CIM) was operated for

the whole genome to detect QTLs at a threshold of *LOD* 2.5 by WinQTLcart 2.5, based on a genetic linkage map which consisted of 172 DNA markers with an average interval of 10.51 cM between markers (Qin *et al.*, 2008). Degenic epistasis QTL and QE interaction were identified using maximum-likelihood estimation and best linear unbiased prediction with a threshold of the  $P \leq 0.005$  by QTLMAPPER version 1.6 (Wang *et al.*, 1999). The correlation coefficient, t-test and one-way ANOVA of the protein content in the DH population during three years were calculated by the program of SAS 9.1.

## RESULTS

### Phenotypic analysis

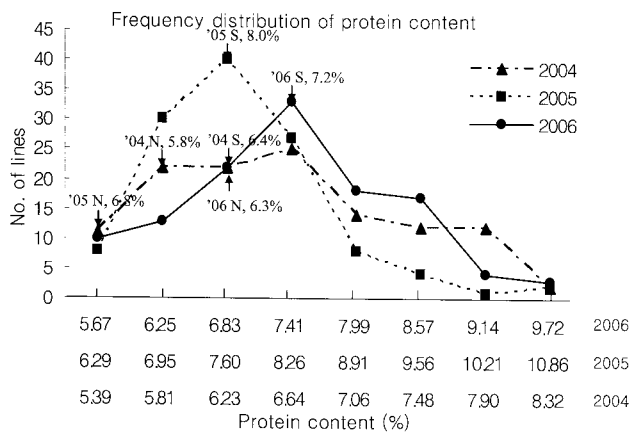
T-test showed significant differences on protein content between two parents during three years. The means of protein content for DH population were higher than those for parents. The significant differences over years were observed on protein content by one-way ANOVA ( $P < 0.001$ ), suggesting that it was easily influenced by environments. The mean of protein content for three years in DH population was 18% higher than that of ‘Nagdong’, but 4% higher than that of ‘Samgang’ (Table 1, Fig. 1).

### Trait correlation

Negative correlations between protein and amylose content ( $r = -0.64^{**}$ ,  $r = -0.57^{**}$ ,  $r = -0.70^{**}$ ) were shown during three years, or lipid content ( $r = -0.36^{**}$ ) only in 2006. Negative correlations were observed between protein content and grain weight ( $r = -0.48^{**}$ ,  $r = -0.40^{**}$ ,  $r = -0.52^{**}$ ), or grain width ( $r = -0.32^{**}$ ,  $r = -0.37^{**}$ ,  $r = -0.42^{**}$ ), or grain thickness ( $r = -0.40^{**}$ ,  $r = -0.43^{**}$ ,  $r = -0.39^{**}$ ) over three years. Thus it can be seen that the protein content has a great influence on the grain appearance of brown

**Table 1.** Mean values of protein content of parents and DH population

Year	Parents			DH lines		
	Samgang	Nagdong	T value	Mean±SD	Mean±SD	Range
2004	6.4±0.08	5.8±0.09	8.82**	6.1±0.07	6.8±0.76	5.4~8.5
2005	8.0±0.33	6.8±0.23	10.13**	7.4±0.65	8.0±0.88	6.3~11.4
2006	7.2±0.29	6.3±0.21	6.31**	6.8±0.49	7.7±0.97	5.7~10.3

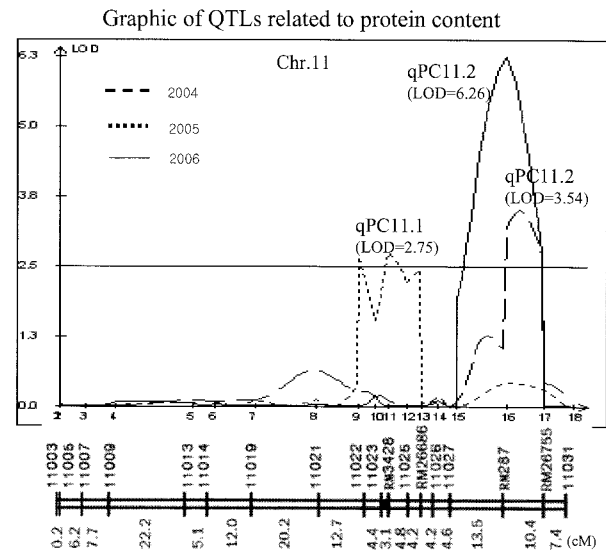


**Fig. 1.** The phenotypic distributions of protein content among 120 DH rice lines derived from a cross ‘Samgang/Nagdong’ over three years. ‘04S and ‘04N, or ‘05S and ‘05N, or ‘06S and ‘06N indicate the protein contents of ‘Samgang’ and ‘Nagdong’ in 2004, 2005 and 2006, respectively.

rice. In addition, the significantly negative correlations between protein content and culm length (CL,  $r = -0.32^{**}$ ), or moment (Mo,  $r = -0.35^{**}$ ), or lodging index (LI,  $r = -0.27^{*}$ ) respectively were shown. Besides, there are no significant correlations between protein content and days to heading, or plant height, or other agronomic traits.

#### QTLs for protein content

Two QTLs associated with protein content were identified on chromosome 11 during three years, consisting of qPC11.2, across two years at a peak marker of RM287 (Table 2, Fig. 2.) and qPC11.1, only detected in 2005 with flanking markers of RM3428 and 11025. They averagely accounted



**Fig. 2.** Graphic of QTLs related to protein content on chromosome 11 over three years.

for 22.10% and 6.92% of phenotype variation at LOD scores of 4.90 and 2.75, respectively (Table 2). The ‘Samgang’ allele increased the protein content by 0.40% and 0.23%, respectively. The qPC1.1 was detected on chromosome 1 with a LOD value of 3.11 flanked by 1008 and RM575 only in 2006. It accounted for 8.22% of phenotype variation with an increase effect of 0.28% from ‘Samgang’ allele.

#### Epistatic QTLs and QE interactions

The significant interaction was observed between qPC1.1 and the environment, which explained 4.32% of phenotype variation (Table 2). In addition, five pairs of two-locus combination for protein content showed negative effects

**Table 2.** Main-effect QTLs and the QE interactions for protein content with the DH population from the cross ‘Samgang / Nagdong’ over three years

QTLs	Years	Chr.	Interval markers	LOD	R <sup>2</sup> (%)	Add.	Increase allele	AEi1	AEi2	AEi3	AEiR <sup>2</sup> (%)
qPC 1.1	2006	1	1008-RM575	3.11	8.22	0.28	Samgang	-0.151	0.229**	-0.078	4.32
qPC 11.1	2005	11	RM3428-11025	2.75	6.92	0.23	Samgang	-0.134	0.024	-0.101	1.70
qPC 11.2	2004	11	11027-RM287	6.26	22.98	0.36	Samgang	-0.124	0.072	0.052	0.40
	2006		RM287-RM26755	3.54	21.21	0.44					

<sup>†</sup>Chr. means chromosomes; R<sup>2</sup> means percentage of phenotypic variation explained; Add. means additive effect. The underline markers indicate the peak markers regarding the QTLs.

<sup>‡</sup>AEi1, AEi2 and AEi3 indicate the interactions between main-effect QTLs and years; AEiR<sup>2</sup> indicates the percentage of phenotypic variation explained by the interaction between main-effect QTLs and years.

**Table 3.** The epistatic QTLs and QE interactions for protein content

C-N (ei)	Interval (ei)	C-N (ej)	Interval (ej)	LOD	Ai	AiR <sup>2</sup> (%)	Aj	AjR <sup>2</sup> (%)	AAij	AAijR <sup>2</sup> (%)	AEiR <sup>2</sup> (%)	AEjR <sup>2</sup> (%)	AAEijR <sup>2</sup> (%)
1-9	1019-RM488	9-6	RM105-RM434	5.67	-0.028	-	0.035	-	-0.227**	4.79	-	-	1.20
1-13	1026-1028	3-1	RM282-3023	9.40	-0.063	-	0.112	-	-0.298**	8.08	-	-	-
2-10	2027-2030	2-14	RM318-2038	8.43	-0.124*	1.40	0.105	-	-0.323**	9.33	-	-	0.55
8-13	8028-8030	12-3	RM247-12011	7.58	0.079	-	0.089	-	-0.309**	8.87	-	-	-
10-8	10025-10027	10-12	RM5620-10030	5.81	0.282**	7.18	-0.095	-	-0.255**	5.76	-	-	-
11-11	RM3428-11025	11-15	11027-RM287	9.12	0.122	1.38	0.177*	2.91	0.159*	2.32	0.03	2.23	-
11-15	11027-RM287	12-1	12003-12008	8.60	0.344**	10.99	0.202**	-	0.097	-	1.94	-	-
Total						A R <sup>2</sup> (%) = 23.86			AA R <sup>2</sup> (%) = 39.15		AE R <sup>2</sup> (%) = 4.20		1.75

<sup>†</sup>C-N means Chromosome-marker Number. ei and ej mean the epistatic loci tested in the analysis.

<sup>‡</sup>Ai and Aj mean the additive effects of the point ei and ej, respectively; The positive value of Ai or Aj implies that the 'Samgang' genotype has a positive effect; AAij means the epistatic effect of additive×additive interaction between points of ei and ej; a positive value indicates that the parental two-locus genotypes have a positive effect. AiR<sup>2</sup>, AjR<sup>2</sup> and AAijR<sup>2</sup> indicate the percentage of phenotypic variations explained by Ai, Aj and AAij, respectively.

<sup>§</sup>AEiR<sup>2</sup> and AEjR<sup>2</sup> indicate the percentage of phenotypic variations explained by the interactions between additive effects of point ei or ej and environments, respectively. AAEijR<sup>2</sup> indicates the percentage of phenotypic variations explained by the interactions between epistatic effects of two points (ei and ej) and environments.

implied that the protein content was decreased by two-locus interactions. Two main-effect QTLs controlling protein content were involved in epistatic effects, whereas these two pairs combination showed positive effects implied that protein content was increased by two-locus interactions (Table 3).

On the whole, the additive effects of two loci with epistatic effects affecting protein content collectively explained 23.86% of phenotype variation, but interactions between additive effects of two loci and environments totally explained 4.20% of phenotype variation of protein content. Furthermore, the additive-by-additive effects collectively accounted for 39.15% of phenotype variation for protein content, while the interactions between environment and epistatic effects totally accounted for 1.75% of phenotype variation of protein content.

## DISCUSSION

It is generally considered that the protein content of brown rice was a composition with low heritability and easy to be affected by environmental conditions and the level of nitrogen fertilization. In a few previous researches,

the QTLs controlling protein content of rice grain were collectively mapped covering ten rice chromosomes except chromosome 9 and 12 (Yoshida, *et al.*, 2002; Hu *et al.*, 2004; Aluko *et al.*, 2004; Cho *et al.*, 2006). In this study, three QTLs were detected on two chromosomes for protein content of brown rice grain. A QTL *pro11* with flanking markers of RM209-RM338 was mapped by Aluko *et al.* (2004) using a DH population from the BC<sub>3</sub>F<sub>1</sub> of an interspecific cross of *O. sativa* / *O. glaberrima* on the vicinity of qPC11.2 in the long arm of chromosome 11 (Fig. 3). In addition, Yoshida *et al.* (2002) reported that a QTL with a peak marker of RM206 for brown and polished rice was located on the end of chromosome 11, which was different locus to qPC11.1 and qPC11.2 in comparison with two genetic maps.

Aluko *et al.* (2004) located the QTL of *pro1* in the intervals RM226-RM297 for grain protein content on the end of chromosome 1, as well as Yoshida *et al.* (2002) with a peak marker of F16722 and Cho *et al.* (2006) with flanking markers RG317-ESTI-2. No main-effect QTL was detected in that locus, but a pair of epistatic loci in the intervals of 1026-1028 on chromosome 1 and RM282-3023 on chromosome 3 was detected on this chromosomal region

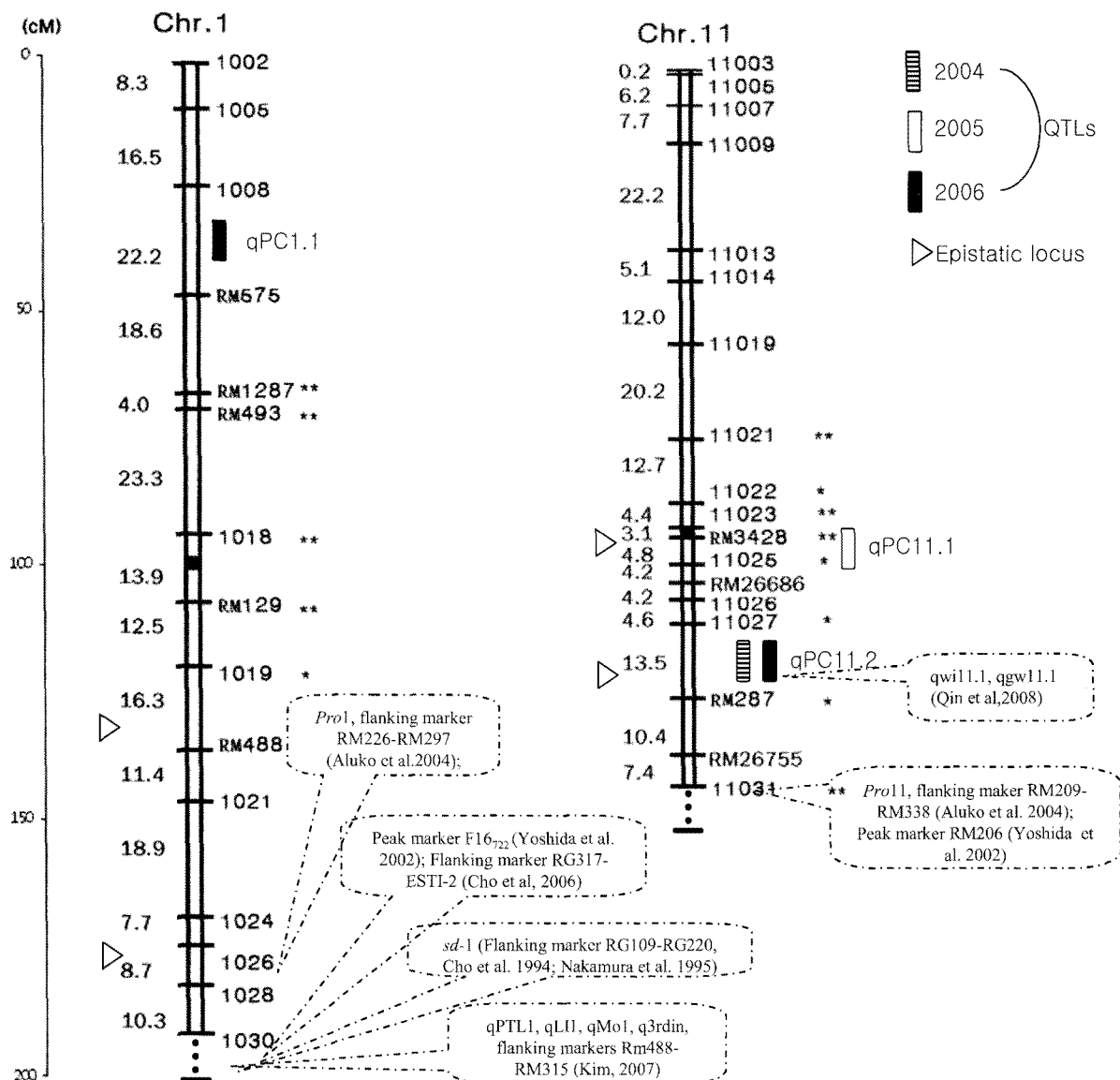


Fig. 3. A reference review of QTL information on chromosome 1 and 11

with a high LOD score of 9.40 in this study, even though the additive-by-additive effect of that locus might decrease the protein content. In addition, the QTLs associated with culm length, the third internode ( $3^{\text{rd}}$ ), moment and lodging index, which closely connected to the plant lodging, were identified on the near regions by Kim (2006). Classical quantitative genetics assumes that trait correlations are the result of either pleiotropic effects or the tight linkage of genes. The negative correlations between protein content

and CL, or  $3^{\text{rd}}$  ( $r = -0.20$ ), or Mo, or LI, respectively, which probably indicated that the genes related to the five traits are tight linkage. Cho *et al.* (1994) identified the *sd-1* gene with flanking markers RG109-RG220 on the long arm of chromosome 1 (Fig. 3). Nakamura *et al.* (1995) reported the tight linkages of seed glutelin (*Glu-1*), semidwarfism (*sd-1*) and shattering habit (*sh-2*) in rice (*Oryza sativa* L.). There was not obvious evidence to prove *Glu-1* and *sh-2* tightly linked to the *sd-1* gene, but it is an undisputable fact

that an epistatic locus related to protein content and the QTLs influencing lodging were immediately located on the region near the *sd-1* gene in the present study.

Remarkably, qPC11.1 might be a summation of two or more main-effect and epistatic QTLs according to the peak form of QTL (Fig. 2). Also, the QTLs regarding grain weight and width were co-localized on the same locus to qPC11.2 with flanking markers of 11027 and RM287 in the previous study (Qin *et al.*, 2008). There is a possibility either there is a cluster of QTLs or genes controlling protein content, grain weight and width from centromere to the end of chromosome 11, or pleiotropy is the genetic basis of high trait correlations. Certainly, it is to be easily understood the negative correlations between protein content and grain weight or width. Since the storage proteins were mainly dispersed in the protein bodies of embryo, aleurone layer and starchy endosperm, the bigger and wider rice grains tend to lose much more aleurone layer during rice processing.

Hu *et al.* (2004) detected five main-effect QTLs related to protein content using 81 DH lines. A major QTL qRPC-5 was mapped in the interval of RG435-RG172a on chromosome 5. It accounted for 35% of the phenotype variation with a LOD of 16.7. The results demonstrated that protein content could be mainly controlled by relatively few QTLs with large main-effects. However, from the standpoint of the present research, protein content is easily influenced by digenic epistatic effects rather than environments. Moreover, the additive-by-additive effects for protein content explained 39.15% of phenotype variation. These results suggest that epistatic effects might be an even more important component of the genetic basis for protein content and that the segregation of the DH lines for protein content could be largely explained by a few main-effect QTLs and many epistatic loci. The epistatic effects and QE interactions are detected to further understand the relationships between loci and how to manipulate these relations in order that they can serve for quality breeding more effectively on the gene level. In addition, how to apply for the epistatic QTLs with increasing or decreasing effects for grain qualities in marker-assisted selection system, however, the vast amount of data are the best approach to getting out the reality.

## ACKNOWLEDGEMENTS

This work was supported by a grant (Code 20050301-034-483-090-08-00) from the Biogreen 21 R&D Program, Rural Development Administration, Republic of Korea.

## REFERENCE

- Aluko, G., C. Martinez, J. Tohme, C. Castano, C. Bergman, and J. H. Oard. 2004. QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. *Theor. Appl. Genet.* 109 : 630-639.
- Fan, C. C., X. Q. Yu, Y. Z. Xing, C. G. Xu, L. J. Luo, and Q. F. Zhang. 2005. The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. *Theor. Appl. Genet.* 110 : 1445-1452.
- Hu, Z. L., P. Li, M. Q. Zhou, Z. H. Zhang, L. X. Wang, L. H. Zhu, and Y. G. Zhu. 2004. Mapping of quantitative trait loci (QTLs) for rice protein and fat content using doubled haploid lines. *Euphytica.* 135 : 47-54.
- Cho, Y. G., M. Y. Eun, S. R. McCouch, and Y. A. Chae. 1994. The semidwarf gene, *sd-1*, of rice (*Oryza sativa* L.). II. Molecular mapping and marker-assisted selection. *Theor. Appl. Genet.* 89 : 54-59.
- Cho, Y. G., H. J. Kang, Y. T. Lee, S. R. McCouch, and M. Y. Eun. 2006. QTL analysis of rice quality using a recombinant inbred population. *J. Agr. Sci., Chungbuk Nat'l Univ.* 23 : 147-164.
- Juliano, B. O. 2003. Rice chemistry and quality, 3<sup>rd</sup> ed., pp. 25-39, Philippine rice research institute, Manila.
- Kim, S. M. 2007. Molecular mapping and DNA marker-assisted selection of resistance genes to brown planthopper in rice. Ph.D Thesis. Kyungpook National University, Korea.
- Li, Z. K., R. M. Pinson, D. William, A. H. Paterson, and J. W. Stansel. 1997. Epistasis for three grain yield components in rice. *Genetics.* 145 : 453-465.
- Mohan, M., S. Nair, A. Bhagwat, T. G. Krishna, M. Yano, C. R. Bhatia, and T. Sasaki. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed.* 3 : 87-103.
- Nakamura, A., S. Komatsu, B. S. Xia, and H. Hirano. 1995. Linkage analysis of the gene loci for seed glutelin (*Glu-1*), semidwarfism (*sd-1*) and shattering habit (*sh-2*) in rice (*Oryza sativa* L.). *Breeding Sci.* 45 : 185-188.
- Qin, Y., S. M. Kim, and J. K. Sohn. 2008. Detection of Main-effect QTLs, Epistatic QTLs and QE Interactions for Grain Appearance of Brown Rice (*Oryza sativa* L.), *J. Crop Sci. Biotech.* 11(2) : 151-156.
- Wang, D. L., J. Zhu, Z. K. Li, and A. Paterson. 1999. Mapping

- QTLs with epistatic and QTL x environmental interactions by mixed linear model approaches. *Theor. Appl. Genet.* 99: 1255-1264.
- Xing, Y. Z., Y. F. Tan, J. P. Hua, X. L. Sun, C. G. Xu, and Q. F. Zhang. 2002. Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. *Theor. Appl. Genet.* 105 : 248-257.
- Yano, M. and T. Sasaki. 1997. Genetic and molecular dissection of quantitative traits in rice. *Plant Mol. Biol.* 35 : 145-153.
- Yoshida, S., M. Ikegami, J. Kuze, K. Sawada, Z. Hashimoto, T. Ishii, C. Nakamura, and O. Kamihima. 2002. QTL analysis for plant and grain characters of sake-brewing rice using a doubled haploid population. *Breeding Sci.* 52 : 309-317.