

Genetic Mapping of *Cgdef* Gene Controlling Accumulation of 7S Globulin (β -Conglycinin) Subunits in Soybean Seeds

MASAKI HAYASHI, KEISUKE KITAMURA, AND KYUYA HARADA

From the National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki, Japan 305-8602 (Hayashi and Harada); and the Graduate School of Agriculture, Hokkaido University, Kita-ku Kita 9 Nishi 9, Sapporo, Hokkaido, Japan 060-8589 (Kitamura).

Address correspondence to M. Hayashi at the address above, or e-mail: twrees@affrc.go.jp.

Soy protein consists of mainly 7S globulin (β -conglycinin) and 11S globulin (glycinin). The 7S globulin exerts favorable and unfavorable effects on human nutrition, food processing, and human health. Therefore, it is important for the improvement of the soy protein to control the content of 7S globulin. A mutant line lacking the 7S globulin was induced by γ -ray irradiation, and the deficiency is controlled by a single recessive gene, *cgdef*. The *Cgdef* gene, despite its potential for improvement of the soy protein, has not been assigned to a linkage group of a soybean genetic map. We crossed "Mo-shi-dou Gong 503" with plants homozygous or heterozygous for the *Cgdef* allele and screened an F2 mapping population that segregated with the *cgdef* allele to locate the *Cgdef* gene on a soybean genetic map. By linkage analysis, we assigned the *Cgdef* gene to chromosome 19 at the position between the Satt523 and Sat_388 simple sequence repeat (SSR) markers. Six SSR markers (Sat_134, Sat_405, Satt143, Satt398, Sat_195, and Satt694) and 2 amplified fragment length polymorphism markers identified previously were mapped at the same position of the *Cgdef* gene. These markers should enable to conduct map-based cloning of the *Cgdef* gene.

Key words: 7S globulin, β -conglycinin, genetic mapping, seed storage protein, soybean (*Glycine max*), SSR marker

Soybean (*Glycine max*) contains about 40% protein in the seed and is the most important source of vegetable protein in the world. The soy protein is composed of 2 major seed storage protein components, 7S globulin (β -conglycinin) and 11S globulin (glycinin), that account for approximately 70% of the total seed protein (Derbyshire et al. 1976). There is complementarity between these 2 globulins, and the amounts of these 2 are almost equal in seeds (Ogawa et al. 1989). Recently, it has been demonstrated that the 7S

globulin is essentially responsible for the decrease in the cholesterol and triglyceride contents in validated animal models (Duranti et al. 2004; Moriyama et al. 2004). Baba et al. (2004) demonstrated that in human, daily ingesting the 7S globulin will be effective in keeping body fat ratio and serum lipid levels normal and eliminating excessive lipids from the body. The interest in the use of the soy protein for human nutrition and health has considerably increased, and soybean is considered to be a very important crop for human life. On the other hand, the 7S globulin displays unfavorable features, including low contents of sulfur-containing amino acids (Koshiyama 1968), poor gel and film properties (Saio et al. 1974a, 1974b; Fukushima 1991), and allergenicity associated with its subunits (Ogawa et al. 1995; Adachi et al. 2009; Krishnan et al. 2009). Therefore, controlling the content of 7S globulin according to the needs of the user is one of the most important objectives of soy protein improvement.

The 7S globulin is a glycoprotein with a molecular weight of about 150 kDa and consists of 3 major subunits, α' (76 kDa), α (72 kDa), and β (52 kDa), which interact to form a trimeric protein with different subunit compositions (Thanh and Shibasaki 1977, 1978). The β -conglycinin-deficient (*cgdef*) mutant that was developed by γ -ray irradiation of "Yumeyutaka" lacks these 3 subunits simultaneously. The deficiency is associated with a single gene trait controlled by a recessive allele (Kitagawa et al. 1991; Hayashi et al. 2000). In our previous studies, we suggested that a common factor regulating indirectly the expression of the 7S globulin subunit genes in a seed-specific manner might not function well in the *cgdef* mutant (Hayashi et al. 1998). This assumption was also supported by the fact that the α and α' subunit genes are inherited independently (Tsukada et al. 1986) and that these 2 genes and the β subunit genes are encoded by multigene families (Schuler et al. 1982; Harada

et al. 1989). The *cgdef* mutant displays some severe effects on plant growth, referred to as *cgdef* syndrome, such as spotted and twisted leaves, sterility, and lethality. These phenotypes are closely linked to but are inherited independently with the *Cgdef* gene (Hayashi et al. 2000).

The integrated linkage map of soybean was constructed (Song et al. 2004), and DNA marker information is readily available through the World Wide Web at <http://soybase.org/resources/ssr.php>. Mapping of a large number of loci responsible for agronomically important traits was achieved (Harada and Xia 2004). Although, the *Cgdef* gene was tagged using amplified fragment length polymorphism (AFLP) markers (Hayashi et al. 2000), its position in the soybean linkage map had not been fully elucidated. In the present study, we assigned the locus to chromosome 19 of the soybean. Positioning of the locus is the first step toward a detailed genetic map of the region and physical isolation of this potential gene for improvement of soy protein quality.

Materials and Methods

Plant Materials

Because the *cgdef* mutant failed to set seeds and died at the middle developmental stage, the F₂ populations were

obtained by selfing of F₁ plants derived from a cross between “Mo-shi-dou Gong 503” and wild-type (Wt) plants, including homozygous or heterozygous plants for the *Cgdef* allele. A weedy soybean, Mo-shi-dou Gong 503, is generally used as a crossing partner for gene mapping in our laboratory because of its high polymorphism to cultivated soybean. An F₂ population composed of 103 individuals that segregated with the *cgdef* allele was selected by a progeny test using sodium dodecyl sulfate–polyacrylamide gel electrophoresis technique and was used as the mapping population for the *Cgdef* gene. Screening of the 7S globulin phenotype was performed as described previously (Hayashi et al. 2000). After the progeny test of F₂ individuals using the distal portion of the seed, the remainder was sowed in a greenhouse. The leaves at an early stage of growth were collected, and DNA extraction was conducted by the hexadecyltrimethylammonium bromide method.

Bulked Segregant Analysis and Linkage Mapping for the *Cgdef* Gene

DNAs from 14 dominant Wt plants, including both homozygous and heterozygous plants for the *Cgdef* allele and from 14 recessive plants homozygous for the *cgdef* allele in the mapping population, were pooled to construct Wt

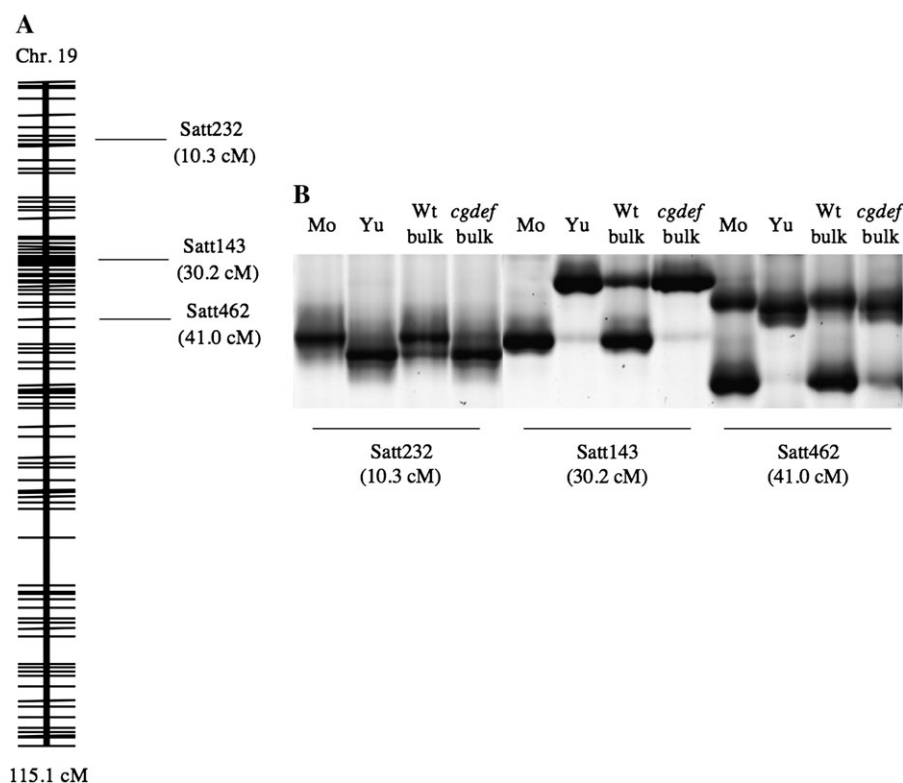


Figure 1. BSA on pools of wild- and *cgdef* mutant type F₂ genotypes. (A) Linkage group L (renamed chromosome 19) of the integrated soybean linkage map quoted from Song et al. (2004). (B) Three portions of the BSA profiles generated with 3 SSR markers (Satt232, Satt143, and Satt462) located on chromosome 19. Mo, Mo-shi-dou Gong 503; Yu, Yumeyutaka; Wt bulk, bulked samples from 14 F₂ individuals of wild type; and *cgdef* bulk, bulked samples from 14 F₂ individuals of *cgdef* mutant type.

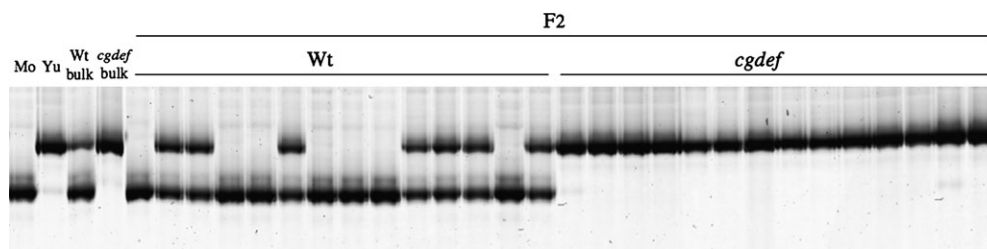


Figure 2. Segregation pattern of Satt143 marker in a set of 28 F2 individuals used for BSA. Mo, Mo-shi-dou Gong 503; Yu, Yumeyutaka; Wt, 14 F2 individuals of wild type; and *cgdef*, 14 F2 individuals of *cgdef* mutant type.

and *cgdef* bulks, respectively. The pooled DNAs were used for bulked segregant analysis (BSA; Michelmore et al. 1991) using simple sequence repeat (SSR) markers mapped on the integrated soybean linkage map (Song et al. 2004). The SSR markers were surveyed at intervals of 10–20 cM, and the SSR primers were designed, based on sequence information through the World Wide Web at <http://soybase.org/resources/ssr.php>. The polymerase chain reaction (PCR) conditions were as follows: 33 cycles of 30 s at 94 °C, 30 s at 47 °C, and 1 min at 72 °C. The PCR products were separated on a 13% nondenaturing polyacrylamide gel with discontinuous buffer system of the Ornstein–Davis type (Hayashi et al. 2001). After running, the electrophoretic pattern was detected by SYBR Gold (Invitrogen) staining with Typhoon 9410 (Amersham Biosciences). Marker linkage analysis was performed with MapMaker Macintosh V2.0 (Lander et al. 1987). The Kosambi function was used for calculating the map distance, and a logarithm of the odds score of 3.0 was used as the threshold for the confirmation of linkage.

Results and Discussion

Of 103 F2 plants derived from the crossing, 89 plants belonged to the Wt and 14 plants to the *cgdef* mutant type. BSA using the Wt and the *cgdef* bulks was conducted to identify SSR markers that cosegregated with the *Cgdef* gene. We found that the Satt143 marker located at the position of 30.2 cM on linkage group L (renamed chromosome 19) of the map of Song et al. (2004) was cosegregated with the *Cgdef* gene (Figure 1). To confirm the result, we performed segregant analysis for individuals consisting of the DNA pools with the SSR marker, Satt143. The Satt143 marker evidently cosegregated with the *Cgdef* gene (Figure 2), and we successfully located the *Cgdef* gene on chromosome 19 of the soybean. Six plants exhibiting recombination between Satt523 and the *Cgdef* gene were identified and the recombination between Sat_388 and the *Cgdef* gene was detected in only 1 plant in the individuals of the mapping population. These 2 markers, which were codominant, enable to accurately construct a local map around the *Cgdef* gene and to estimate genetic distances. We did not detect recombination between the *Cgdef* gene and 6 SSR markers located between Satt523 and Sat_388 in the integrated the

linkage map of Song et al. (2004). Five codominant markers, Sat_134, Sat_405, Satt143, Satt398, and Sat_195, and a dominant marker, Satt694, were found to tightly flank the *Cgdef* gene within a genetic distance of 0 cM (Figure 3). In addition, 2 AFLP markers, E-ACT/M-CTA and E-ACA/M-CTT, identified previously (Hayashi et al. 2000) were also completely linked to the *Cgdef* gene (Figure 3). The fact that these 8 markers and the *Cgdef* gene were mapped within a genetic distance of 0 cM indicates that genetic recombination was severely repressed in the region of the *Cgdef* gene in the same way as in the centromeric regions. The physical distance between Satt523 and Sat_388 is approximately 24.1 Mbp, and the *Cgdef* gene and the 8 markers lie within approximately 7.6 Mbp genomic region

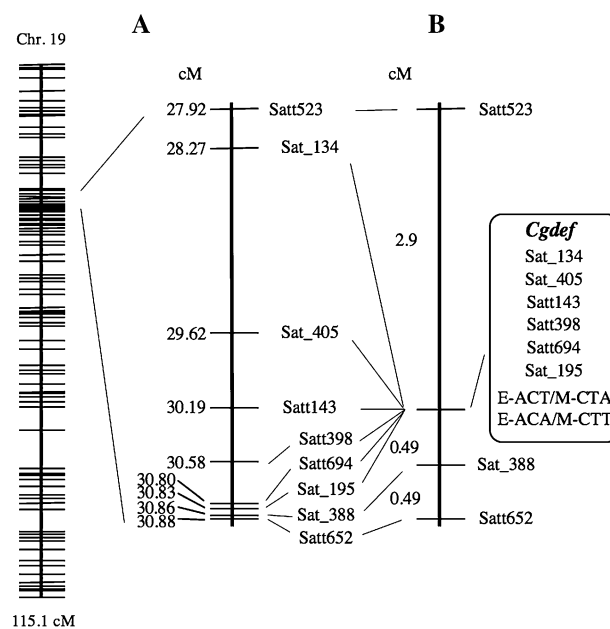


Figure 3. Location of the *Cgdef* gene on chromosome 19 of the soybean. (A) Linkage group L (renamed chromosome 19) of the integrated soybean linkage map quoted from Song et al. (2004) and enlargement of part of the target region. (B) Genetic linkage map showing the positions of the *Cgdef* locus, the 8 tightly linked markers, and 3 linked SSR markers (Satt523, Sat_388, and Satt652).

(<http://www.phytozome.net/soybean> and <http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/>).

The *cgdef* mutant displays some unfavorable effects on plant growth, referred to as *cgdef* syndrome, including spotted and twisted leaves, sterility, and lethality. However, we have found a sterile *Cgdef* plant in the mapping population, the fertile plant lacking the 7S globulin subunits have not been found (Hayashi et al. 2000). If we generate a larger scale population, the 8 tightly linked markers will be useful to screen the fertile *cgdef* plants effectively, and the recombinants between the *Cgdef* gene and these markers will be found. In addition, these markers will be available to isolate the *Cgdef* gene by a physical method.

Another 7S globulin-deficient mutant, QT2, controlled by a single dominant gene, *Scg-1* (suppressor of β -conglycinin), does not display the abnormal traits (Hajika et al. 1996, 1998). The *Scg-1* was located in the same region on chromosome 20 as the α and β subunit genes, which are tightly linked to each other (Tsubokura et al. 2006b). The deficiency is associated with the cluster region of the 7S globulin subunit genes (Teraishi et al. 2001; Tsubokura et al. 2006a). The subunit genes, including the α' subunit gene located on chromosome 10, share highly conserved region. Therefore, Teraishi et al. (2001) and Tsubokura et al. (2006a, 2006b) correctly predicted that the structure of the *Scg-1* locus is involved in RNA silencing of the 7S globulin subunit genes; that the location of the *Cgdef* gene is different from the chromosomes 20 and 10; and the mechanism of the deficiency in the *cgdef* mutant must be different from that in the *Scg-1* mutant.

The analysis of the expression of seed storage protein genes that is tightly regulated during seed development by both genetic and epigenetic mechanisms could offer a powerful model system to study the regulation of gene expression during plant embryogenesis. Regulators of seed storage protein genes mediated by or interacting with other factors have been well-characterized for *Arabidopsis* 2S albumin, napin (Kroj et al. 2003; To et al. 2006), and for bean 7S globulin, phaseolin (Li et al. 1999). Transcription from the phaseolin promoter requires 2 major events: chromatin remodeling, mediated by PvALF, a B3 domain transcription factor, and activation by an ABA-induced signal transduction cascade (Li et al. 1999). Ng et al. (2006) revealed that chronological changes in histone modifications were required for chromatin remodeling and activation of the phaseolin promoter. We suggest that soybean has a similar mechanism for the regulation of 7S globulin subunit genes expression, including chromatin remodeling and histone modifications. The study of the *Cgdef* gene might lead to the elucidation of the regulation mechanism of 7S globulin subunit genes expression in soybean developing seeds and provide us with a clue to improve the soy protein.

Funding

Program for Promotion of Basic Research Activities for Innovative Biosciences.

References

- Adachi A, Horikawa T, Shimizu H, Sarayama Y, Ogawa T, Sjolander S, Tanaka A, Moriyama T. 2009. Soybean beta-conglycinin as the main allergen in a patient with food-dependent exercise-induced anaphylaxis by tofu: food processing alters pepsin resistance. *Clin Exp Allergy*. 39:167–173.
- Baba T, Ueda A, Kohno M, Fukui K, Miyazaki C, Hirotsuka M, Ishinaga M. 2004. Effects of soybean beta-conglycinin on body fat ratio and serum lipid levels in healthy volunteers of female university students. *J Nutr Sci Vitaminol*. 50:26–31.
- Derbyshire E, Wright DB, Boulter D. 1976. Legumin an vicilin, storage protein of legume seeds. *Phytochemistry*. 15:3–24.
- Duranti M, Lovati MR, Dani V, Barbiroli A, Scarafoni A, Castiglioni S, Ponzoni C, Morazzoni P. 2004. The α' subunit from soybean 7S globulin lowers plasma lipids and upregulates liver β -VLDL receptors in rats fed a hypercholesterolemic diet. *J Nutr*. 134:1334–1339.
- Fukushima D. 1991. Recent progress of soybean protein foods: chemistry, technology, and nutrition. *Food Rev Int*. 7:323–351.
- Hajika M, Takahashi M, Sakai S, Igita M. 1996. A new genotype of 7S globulin (β -conglycinin) detected in wild soybean (*Glycine soja* Sieb. et Zucc.). *Breed Sci*. 46:385–386.
- Hajika M, Takahashi M, Sakai S, Matsunaga R. 1998. Dominant inheritance of a trait lacking β -conglycinin detected in a wild soybean line. *Breed Sci*. 48:383–386.
- Harada JJ, Barker SJ, Goldberg RB. 1989. Soybean β -conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell*. 1:415–425.
- Harada K, Xia Z. 2004. Soybean genomics: efforts to reveal the complex genome. *Breed Sci*. 54:215–224.
- Hayashi M, Harada K, Fujiwara T, Kitamura K. 1998. Characterization of a 7S globulin-deficient mutant of soybean (*Glycine max* (L.) Merrill). *Mol Gen Genet*. 258:208–214.
- Hayashi M, Nishioka M, Kitamura K, Harada K. 2000. Identification of AFLP markers tightly linked to the gene for deficiency of the 7S globulin in soybean seed and characterization of abnormal phenotypes involved in the mutation. *Breed Sci*. 50:123–129.
- Hayashi M, Miyahara A, Sato S, Kato T, Yoshikawa M, Taketa M, Hayashi M, Pedrosa A, Onda R, Imaizumi-Anraku H, et al. 2001. Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F2 population. *DNA Res*. 8:301–310.
- Kitagawa S, Ishimoto M, Kikuchi F, Kitamura K. 1991. A characteristic lacking or decreasing remarkably 7S globulin subunits induced with γ -ray irradiation in soybean seeds. *Jpn J Breed*. 41(Suppl 2):460–461.
- Koshiyama I. 1968. Chemical and physical properties of a 7S protein in soybean globulins. *Cereal Chem*. 45:394–404.
- Krishnan HB, Kim WS, Jang S, Kerley MS. 2009. All three subunits of soybean β -conglycinin are potential food allergens. *J Agric Food Chem*. 57:938–943.
- Kroj T, Savino G, Valon C, Giraudat J, Parcy F. 2003. Regulation of storage protein gene expression in *Arabidopsis*. *Development*. 130:6065–6073.
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L. 1987. MapMaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*. 1:174–181.
- Li G, Bishop KJ, Chandrasekharan MB, Hall TC. 1999. β -Phaseolin gene activation is a two-step process: pvALF facilitated chromatin modification followed by abscisic acid-mediated gene activation. *Proc Natl Acad Sci USA*. 96:7104–7109.
- Michelmore RW, Paran I, Kesseli RV. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid

- method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA*. 88:9828–9832.
- Moriyama T, Kishimoto K, Nagai K, Urade R, Ogawa T, Utsumi S, Maruyama N, Maebuchi M. 2004. Soybean β -conglycinin diet suppresses serum triglyceride levels in normal and genetically obese mice by induction of β -oxidation, downregulation of fatty acid synthase, and inhibition of triglyceride absorption. *Biosci Biotechnol Biochem*. 68:352–359.
- Ng DWK, Chandrasekharan MB, Hall TC. 2006. Ordered histone modifications are associated with transcriptional poising and activation of the *phaseolin* promoter. *Plant Cell*. 18:119–132.
- Ogawa T, Bando N, Tsuji H, Nishikawa K, Kitamura K. 1995. α -subunit of β -conglycinin, an allergenic protein recognized by IgE antibodies of soybean-sensitive patients with atopic dermatitis. *Biosci Biotechnol Biochem*. 59:831–833.
- Ogawa T, Tayama E, Kitamura K, Kaizuma N. 1989. Genetic improvement of seed storage proteins using three variant alleles of 7S globulin subunits in soybean (*Glycine max* L.). *Jpn J Breed*. 39:137–147.
- Saio K, Sato I, Watanabe T. 1974a. Food use of soybean 7S and 11S. Part II. High temperature expansion characteristics of gels. *J Food Sci*. 39:777–782.
- Saio K, Sato I, Watanabe T. 1974b. Functional properties of heat-induced gel prepared from crude fractions of soybean 7S and 11S. *Nippon Shokuhin Kogyo Gakkaishi*. 21:234–238.
- Schuler MA, Schmitt ES, Beachy RN. 1982. Closely related families of genes code for the α and α' subunits of the soybean 7S storage protein complex. *Nucleic Acids Res*. 10:8225–8244.
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB. 2004. A new integrated genetic linkage map of the soybean. *Theor Appl Genet*. 109:122–128.
- Teraishi M, Takahashi M, Hajika M, Matsunaga R, Uematsu Y, Ishimoto M. 2001. Suppression of soybean β -conglycinin genes by a dominant gene, *Sgg-1*. *Theor Appl Genet*. 103:1266–1272.
- Thanh VH, Shibasaki K. 1977. β -Conglycinin from soybean proteins. Isolation and immunological and physicochemical properties of the monomeric forms. *Biochim Biophys Acta*. 490:370–384.
- Thanh VH, Shibasaki K. 1978. Major proteins of soybean seeds. Subunit structure of β -conglycinin. *J Agric Food Chem*. 26:692–695.
- To A, Valon C, Savino G, Guillemot J, Devic M, Giraudat J, Parcy F. 2006. A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell*. 18:1642–1651.
- Tsubokura Y, Hajika M, Harada K. 2006a. Molecular characterization of a β -conglycinin deficiency soybean. *Euphytica*. 150:249–255.
- Tsubokura Y, Hajika M, Harada K. 2006b. Molecular markers associated with β -conglycinin deficiency in soybean. *Breed Sci*. 56:113–117.
- Tsukada Y, Kitamura K, Harada K, Kaizuma N. 1986. Genetic analysis of subunits of two major storage proteins (β -conglycinin and glycinin) in soybean seeds. *Jpn J Breed*. 36:390–400.

Received February 24, 2009; Revised May 10, 2009;
Accepted June 16, 2009

Corresponding Editor: Reid G. Palmer