

Genetic Diversity and Population Structure of Korean Soybean Collection Using 75 Microsatellite Markers

Gi-An Lee*, Yu-Mi Choi*, Jung-Yoon Yi*, Jong-Wook Chung*, Myung-Chul Lee*, Kyung-Ho Ma*,
Sok-young Lee*, Jin-Woong Cho**†, and Jung-Ro Lee*†

*National Agrobiodiversity Center, National Academy of Agricultural Science, RDA, Jeonju-si, Jeollabuk-do, Korea

**College of Agricultural and Life Science, Chungnam National University, Daejeon 305-764, Korea

ABSTRACT Soybean (*Glycine max* L.) is crucial legume crop as source of high quality vegetable protein and oil, and Korea is regarded as a part of center of soybean origin. To expand the information of conserved genetic diversity, we analyzed the genetic variability of soybean collection mainly introduced Korean accessions using 75 microsatellite markers. A total of 1,503 alleles with an average value of 20.0 alleles were detected among 644 accessions. Korean collection revealed average allele number of 13.4 while Chinese, Japanese and Southeast Asian accessions showed 9.0, 5.4 and 6.5 mean alleles, respectively. Especially, Korean accessions showed more number of private allele per locus as 3.4 contrary to other geographical groups. The mean expected heterozygosity and polymorphic information content was 0.654 and 0.616, respectively, and expected heterozygosity values were not significantly distinguished according to the geographical groups. The phylogenetic dendrogram and deduced population structure based on DNA profiles of 75 SSR loci showed Korean accessions formed distinct gene pool against Chinese accessions, and could be divided into five subpopulations. Korean soybean accessions have specific genetic diversity and might be serve the valuable alleles for bio-industry as a part of the center of soybean origin.

Keywords : microsatellites, genetic diversity, soybean, *Glycine max* L.

Soybean (*Glycine max* L.) is cultivated over the world and especially has been a crucial legume crop in eastern Asia region including Korea, China, and Japan because this crop has been important source of high quality vegetable protein and oil for food products. Recently, the usage of soybean for biodiesel has made this crop more economically important,

and the whole-genome sequencing of Williams 82 (*G. max*) and IT182932 (*G. soja*) have been completed accordingly (Hisano *et al.*, 2007; Kim *et al.*, 2010; Schmutz *et al.*, 2010). Yu & Kiang (1993) reported that Korea might be one of the major center of soybean origins based on electrophoretic assays of wild soybean (*G. soja*) and cultivated soybeans (*G. max*). Lee *et al.* (2008) also suggested that Korea is a major center of diversity for wild soybean and potentially could serve useful alleles not found in other parts of the world while Dong *et al.* (2004) regarded northeastern region of China as major center of soybean origin. Furthermore, the collection of Korea and China revealed divergent genotypes while accessions of Japan was genetically similar to Korean accessions (Lee *et al.*, 2011; Li & Nelson, 2001).

The soybean cultivars, derived from a few accessions, show narrow genetic background, and this means cultivated varieties might be vulnerable to new biotic and abiotic stress along with global environmental changes (Hyten *et al.*, 2006; Mulato *et al.*, 2010). The genetic diversity of soybean including landraces and wild relatives is needed for breeding program and bio-industry as a valuable allele source. Especially, landraces collected in center of origins are useful for genetic improvement of cultivars in that wild relatives generally reveal many unwanted traits and show low introgression rate (Song *et al.*, 2013). Genetic analysis using molecular markers have been consistently studied to understand the genetic variability of soybeans, and have revealed the genetic structure and diversity of various soybean collections (Choi *et al.*, 1999; Dong *et al.*, 2004; Guan *et al.*, 2010; Lee *et al.*, 2008; Li *et al.*, 2008; Wang *et al.*, 2006; Yoon *et al.*, 2009).

†Corresponding author: (Phone) +82-63-238-4880 (E-mail) jrmail@korea.kr
(Phone) +82-42-821-5725 (E-mail) jwcho@cnu.ac.kr

<Received 22 September 2014; Accepted 22 October 2014>

Microsatellite or simple sequence repeats (SSRs) markers has been used for genetic diversity analysis because microsatellites are plentiful variations over the genome and highly reproducible and easy for genotyping as a codominant markers. The high level of polymorphism was detected in several soybean collections; 18.3 alleles (6 microsatellite loci) in Korean soybean accessions (Yoon *et al.*, 2009), 25.5 alleles (46 microsatellite loci) in wild soybean collection (Lee *et al.*, 2008), 12.2 alleles (60 microsatellite loci) in Chinese soybean accessions (Wang *et al.*, 2006), 19.7 alleles (59 microsatellite loci) in soybean landraces of China (Li *et al.*, 2008). Using microsatellite markers, Choi *et al.* (1999) analyzed the collection acquired from five major rivers in Korea and they supported that SSR markers are useful tool to classify cultivated soybeans and wild soybeans, and Dong *et al.* (2004) analyzed genetic diversity of cultivated soybean accessions collected from several provinces of China and proposed one possible diversity center.

In this paper, we analyzed the genetic diversity of soybean collection mainly introduced Korean accession from USDA National Plant Germplasm System of USA using genome-wide 75 microsatellite markers.

Materials and Methods

Plant materials and DNA extraction

A total of 644 soybean and wild soybean accessions were used in this study; 619 soybean (*G. max*) accessions (507 - Korea, 57 - China, 19 - Japan, 36 - Southeastern Asia) and 25 wild soybean (*G. soja*) accessions (6 - Korea, 9 - China, 10 -

Table 1. Summary information on soybean (*Glycine max* & *G. soja*) accessions used in this study.

Species	Origin	Accession number
Soybean (<i>Glycine max</i>)	Korea	507
	China	57
	Japan	19
	Southeastern Asia	36
	Sub total	619
Wild soybean (<i>Glycine soja</i>)	Korea	6
	China	9
	Japan	10
	Sub total	25
	Total	644

Japan) were included for diversity analysis (Table 1, supplementary Table S1). The majority of Korean soybean accessions were re-introduced from USDA National Plant Germplasm System of USA because these accessions were not conserved in Korean National Agrobiodiversity Center before.

All plant materials were distributed from the National Agrobiodiversity Center, National Academy of Agricultural Science (NAAS), Rural Development Administration (RDA), Republic of Korea. Total genomic DNA was extracted by modified cetyltrimethylammonium bromide (CTAB) method (Dellaporta *et al.*, 1983) using young leaf tissue after grinding in liquid nitrogen. DNA concentration and quality were determined with a NanoDrop ND-1000 spectrometer (NanoDrop Technologies, Wilmington, DE, USA), and DNA stocks were finally diluted to working solutions of 20 ng/ μ l.

Genotyping with microsatellite markers

Genome-wide 75 microsatellite markers were chosen along the soybean genetic linkage map. The reproducibility of PCR product size was evaluated by performing three replications using two different samples for each SSR primer. The M13F-tail PCR method of Schuelke (2000) was used to measure the size of PCR products, as described previously (Lee *et al.*, 2014). The PCR master mix was prepared as follows: Each 20 μ l reaction mixture contained 20 ng template gDNA, 1X PCR buffer, 2 uM specific primer, 200 uM dNTPs, 4 uM M13 universal primer, 6 uM normal reverse primer, and 0.5 U Taq-polymerase (SolGent, Korea). PCR was performed as follows: denaturation at 94°C (3 min) then 30 cycles each of 94°C (30 s), 55°C (45 s), and 72°C (1 min), followed by 10 cycles of 94°C (30 s), 53°C (45 s), and 72°C (1 min), and a final extension at 72°C for 10 min with a PTC-200 thermocycler (MJ Research, Waltham, MA, USA). Amplified PCR products were resolved using an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) with GeneScan 500 ROX (6-carboxy-rhodamine) as molecular size standard (35–500 bp).

Data analysis

As a indicator of genetic diversity, the major allele frequency (M_{AF}), number of alleles (N_A), observed heterozygosity (H_O), expected heterozygosity (H_e) and polymorphism information content (PIC), were calculated using the genetic analysis

package PowerMarker V3.25 (Liu & Muse, 2005). Phylogenetic dendrogram was constructed with an unweighted pair-group method of the arithmetic average (UPGMA) based on genetic distance matrix using the same software. The Structure software (V2.2) (Pritchard *et al.*, 2007) was utilized to detect possible subpopulations (K=1 to K=10) with a model allowing for admixture and correlated allele frequencies using a burn-in period of 50,000 and MCMC repeats of 50,000 followed by 3 iterations. The optimal number of populations corresponds to the highest peak in the ΔK graph (Evanno *et al.*, 2005), and the accessions with membership probabilities $\geq 70\%$ were assigned to each subpopulation. The molecular variance for geographical regions was calculated by molecular variance (AMOVA) approach using the GenALEX software (ver. 6.1) (Peakall & Smouse, 2006).

Results

Genetic variability of soybeans

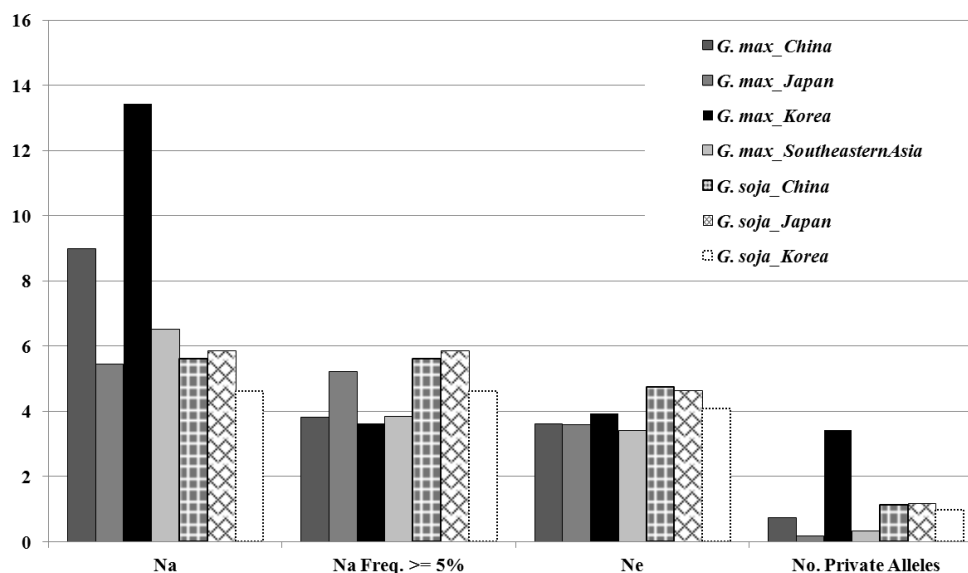
To reveal the genetic diversity of retained Korean soybean accessions in 75 SSR loci, major allele frequency (M_{AF}), number of alleles (N_A), expected heterozygosity (H_E), observed

heterozygosity (H_o) and polymorphic information content (PIC) are analyzed and presented in Table 2 and supplementary Table S2. The mean value of M_{AF} was 0.476 (satt276; 0.111~satt328; 0.913) and 0.278 (satt316 and satt157; 0.091~satt229; 0.833) in soybean and wild soybean accessions, respectively. A total of 1,503 alleles (soybean; 1,200 alleles, wild soybean; 835 alleles) with an average value of 20.0 alleles (soybean; 16.0 mean alleles, wild soybean; 11.4 mean alleles) were detected in 75 SSR loci. Among the tested soybean samples, The Korean accessions revealed average allele number of 13.4 (2~39) while Chinese, Japanese and Southeast Asian accessions showed 9.0, 5.4 and 6.5 mean alleles, respectively. Interestingly, the mean private allele number of Korean group was 3.4 contrary to other soybean geographical groups ranged from 0.2 to 0.7 mean alleles (Fig. 1). The mean expected heterozygosity (H_E) and polymorphic information content (PIC) among soybean accessions was ranged from a 0.162 on satt328 to 0.954 on satt276 (mean = 0.654) and from 0.156 on satt328 to 0.952 on satt323 (mean = 0.616), respectively. The average H_E values were not significantly distinguished according to the geographical group as Korea of 0.628, China of 0.637, Japan of 0.638 and Southeast Asia of 0.600 while wild soybeans

Table 2. Summary of genetic diversity indices in 644 soybean accessions using 75 microsatellite markers.

Group	Subgroup	N	M_{AF}	N_A	H_E	H_o	PIC
	Sub-total	619	0.476 (0.111~0.913)	16 (3~41)	0.654 (0.162~0.954)	0.019 (0~0.13)	0.616 (0.156~0.952)
Soybean (<i>G. max</i>)	Region-1 (Korea)	507	0.502 (0.115~0.944)	13.427 (2~39)	0.628 (0.106~0.951)	0.008 (0~0.141)	0.589 (0.101~0.949)
	Region-2 (China)	57	0.507 (0.125~0.889)	8.987 (3~22)	0.637 (0.201~0.928)	0.038 (0~0.109)	0.602 (0.187~0.924)
	Region-3 (Japan)	19	0.483 (0.167~0.842)	5.44 (2~13)	0.638 (0.266~0.898)	0.02 (0~0.368)	0.589 (0.231~0.89)
	Region-4 (Southeastern Asia)	36	0.522 (0.167~0.972)	6.52 (2~21)	0.6 (0.054~0.914)	0.137 (0~0.371)	0.559 (0.053~0.908)
	Sub-total	25	0.278 (0.091~0.833)	11.373 (3~20)	0.826 (0.299~0.941)	0.045 (0~0.261)	0.81 (0.289~0.938)
Wild soybean (<i>G. soja</i>)	Region-1 (Korea)	6	0.368 (0.167~1)	4.627 (1~7)	0.71 (0~0.847)	0.092 (0~0.5)	0.668 (0~0.828)
	Region-2 (China)	9	0.356 (0.111~1)	5.627 (1~9)	0.742 (0~0.889)	0.017 (0~0.2)	0.71 (0~0.878)
	Region-3 (Japan)	10	0.378 (0.111~0.9)	5.88 (2~10)	0.733 (0.18~0.889)	0.044 (0~0.222)	0.702 (0.164~0.878)

*Abbreviation : M_{AF} , major allele frequency; N_A , number of alleles, H_o , observed heterozygosity; H_E , expected heterozygosity; PIC, polymorphic information content.



*Abbreviation : Na, No. of alleles; Na Freq. $\geq 5\%$, mean number of alleles with frequency greater than 0.05 per locus; Ne, average effective number of alleles per locus; No. private alleles; Equivalent to the number of alleles unique to a single population in the data set.

Fig. 1. Allele variability according to the origin of tested 644 soybean accessions.

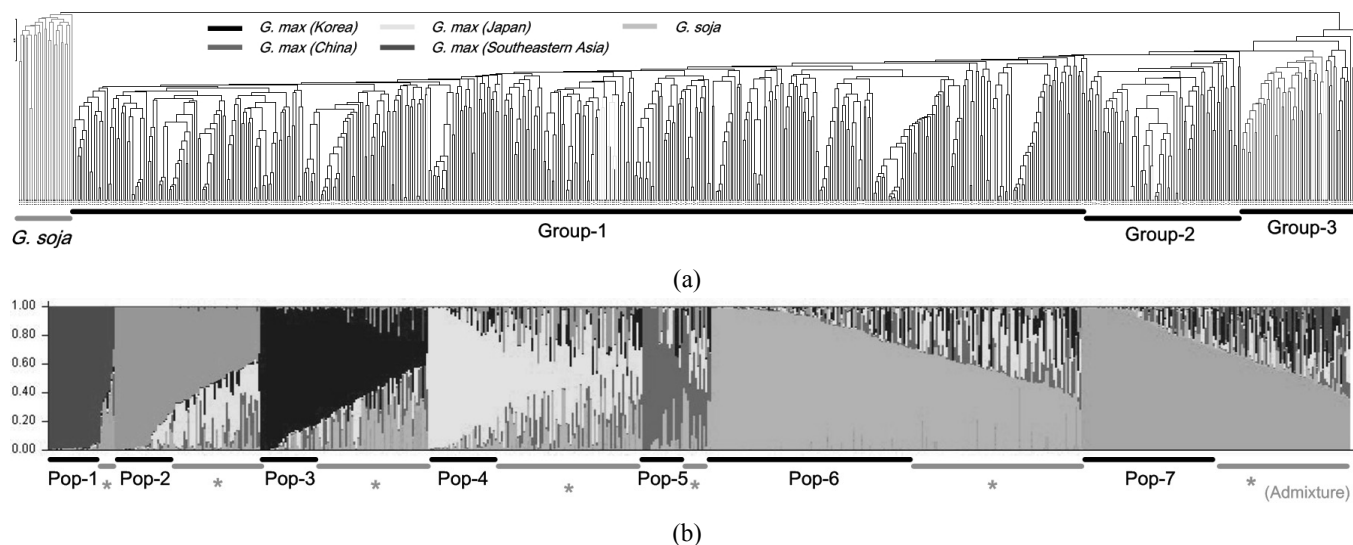


Fig. 2. Phylogenetic dendrogram (a) and model-based population structure based on SSR profiles of 644 soybean accessions.

revealed higher mean H_E values of 0.826 ranged from 0.299 to 0.941.

Genetic relationship and population structure

To reveal genetic relationship of tested 644 accessions, we constructed the phylogenetic dendrogram based on DNA profiles of 75 SSR loci (Fig. 2). The soybean and wild soybean were critically divergent, and soybean accessions could be

divided into three groups in a genetic distance value of 0.72; Group-1 was mainly composed of Korean and Japanese soybean accessions, main Chinese accessions were appertained into Group-2 and Group-3 included mainly Southeast Asian accessions. According to the phylogenetic analysis Korean and Chinese soybean gene pools were divergent, and Japanese accessions tied into Korean gene pool while Southeast Asian accessions showed near genetic relationship to Chinese accessions.

The model-based Bayesian clustering method was used to detect the possible subpopulations along the tested soybean gene pools, and we could find significant population structure. For each K value (K = 1 to 10), we determined the optimal value of K by calculating ΔK . As we used soybean and wild soybean accessions in this study, we surveyed the highest ΔK value over “K=3”. The seven subpopulations, referred to as Pop-1~Pop-7, were observed with the highest ΔK value (data not shown), and 331 accessions were admixture type with a membership probability < 70% (Fig. 2). Most accessions were distinguished according to the geographical origin; soybean accessions could be divided into six subpopulations (Pop-2~Pop-7) while wild soybean accessions were tied into Pop-1 group. Korean accessions were sub-divided into five subpopulations (Pop-2~Pop-6), and the Pop-3 included three Japanese accessions. Chinese and Southeast Asian accessions were located in Pop-7, and some Korean accessions were included in this group. The AMOVA analysis showed that the molecular variance between the geographical groups accounted for 13% and the remaining 87% were due to the differences within the geographical groups.

Discussion

Soybean is important legume crop for protein and oil source, but the recent cultivated varieties reveals narrow genetic background and weak adaptability to global environment changes (Hyten *et al.*, 2006; Mulato *et al.*, 2010). As Korea is one of the center of soybean origins, we analyzed the genetic diversity of Korean soybean collection using 75 microsatellite markers for a sustainable conservation and utilization.

In this study, Korean soybean accessions revealed average allele number (N_A) of 13.4, H_E of 0.628 and PIC of 0.589, and the N_A of Korean accession was higher than other regional groups (China - 9.0, Japan - 5.4, Southeast Asia - 6.5). As we analyzed the more Korean samples than other geographical regions, detection of more alleles in Korean collection is expectable results, but the average private allele number of Korean accessions was 3.4 contrary to other soybean geographical groups ranged from 0.2 to 0.7 mean private alleles. This means that the Korean accessions conserve distinct alleles and might be valuable sources for breeding program and bio-industry with the regional specific alleles. In previous studies, (Li *et al.*,

2011) reported average 14.2 alleles and 0.806 genetic diversity in Chinese core soybean collection at 55 microsatellite loci, and (Guan *et al.*, 2010) reported 16.2 alleles with a mean gene diversity of 0.84 in Chinese and Japanese soybean collection at 46 SSR loci. Our results showed a little low allele numbers and genetic diversity values, but determining the diversity level is inappropriate because of different microsatellite loci and collection. The wild soybean (*G. soja*) revealed higher mean H_E values of 0.826 than soybean (*G. max*) of 0.654, and this is reasonable in that wild relative generally conserve more genetic diversity than domesticated landraces.

The phylogenetic dendrogram and deduced population structure based on DNA profiles of 75 SSR loci showed that Korean and Chinese soybean accessions formed divergent gene pools. The AMOVA for geographical groups revealed that 13% and 87% of total variation were due to the differences between and within geographical groups, respectively, and this indicated that the significant geographical differentiation was present. Japanese accessions showed high genetic relationship to Korean soybean gene pools, whereas Southeast Asian accessions were similar to Chinese accessions. As previous studies (Lee *et al.*, 2011; Lee *et al.*, 2008; Li & Nelson, 2001; Yu & Kiang, 1993), soybean germplasm of Korea region conserve distinct genetic diversity compared to Chinese accessions, and Japanese and Korean accessions were more similar to each other. By model-based bayesian clustering method showed that Korean soybean accessions could be subdivided into five subpopulation (Pop-2~Pop6). The results in this study indicate that Korean soybean accessions have specific genetic diversity and might be serve the valuable alleles for bio-industry as a part of the center of soybean origin.

Continuous utilization of limited elite cultivars reduces the genetic background of cultivated soybean, and the adaptability against rapidly changing circumstance might decline accordingly. This study serves that Korean soybean accessions show distinct genetic diversity and is composed of several subpopulations. The detected specific alleles and population structures could elevate the utilization Korean soybean genetic resources.

Acknowledgements

This study was supported by the Rural Development Administration (RDA), a grant (Code # PJ008625) from the

National Academy of Agricultural Science, RDA, Republic of Korea.

REFERENCES

- Choi, I. Y., J. H. Kang, H. S. Song, and N. S. Kim. 1999. Genetic diversity measured by simple sequence repeat variations among the wild soybean, *Glycine soja*, collected along the riverside of five major rivers in Korea. *Genes Genet. Syst.* 74(4) : 169-177.
- Dellaporta, S. L., J. Wood, and J. B. Hicks. 1983. A plant DNA miniprep: version II. *Plant Mol. Biol. Rep.* 1(4) : 19-21.
- Dong, Y., L. Zhao, B. Liu, Z. Wang, Z. Jin, and H. Sun. 2004. The genetic diversity of cultivated soybean grown in China. *Theor. Appl. Genet.* 108(5) : 931-936.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14(8) : 2611-2620.
- Guan, R., R. Chang, Y. Li, L. Wang, Z. Liu, and L. Qiu. 2010. Genetic diversity comparison between Chinese and Japanese soybeans (*Glycine max* (L.) Merr.) revealed by nuclear SSRs. *Genet. Resour. Crop Evol.* 57(2) : 229-242.
- Hisano, H., S. Sato, S. Isobe, S. Sasamoto, T. Wada, A. Matsuno, T. Fujishiro, M. Yamada, S. Nakayama, and Y. Nakamura. 2007. Characterization of the soybean genome using EST-derived microsatellite markers. *DNA Res.* 14(6) : 271-281.
- Hyten, D. L., Q. Song, Y. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht, R. C. Shoemaker, and P. B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl. Acad. Sci. USA* 103(45) : 16666-16671.
- Song, J. Y., X. M. Piao, Y. M. Choi, G. A. Lee, J. W. Chung, J. R. Lee, Y. I. Jung, H. J. Park, and M. C. Lee. 2013. Evaluation of genetic diversity and comparison of biochemical traits of soybean (*Glycine max* L.) germplasm collections. *Plant Breed Biotechnol* 1(4) : 374-384.
- Kim, M. Y., S. Lee, K. Van, T. H. Kim, S. C. Jeong, I. Y. Choi, D. S. Kim, Y. S. Lee, D. Park, and J. Ma. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc. Natl. Acad. Sci. USA* 107(51) : 22032-22037.
- Lee, G. A., J. S. Sung, S. Y. Lee, J. W. Chung, J. Y. Yi, Y. G. Kim, and M. C. Lee. 2014. Genetic assessment of safflower (*Carthamus tinctorius* L.) collection with microsatellite markers acquired via pyrosequencing method. *Mol. Ecol. Resour.* 14(1) : 69-78.
- Lee, J. D., T. D. Vuong, H. Moon, J. K. Yu, R. Nelson, H. T. Nguyen, and J. G. Shannon. 2011. Genetic diversity and population structure of Korean and Chinese soybean [*Glycine max* (L.) Merr.] accessions. *Crop Sci.* 51(3) : 1080-1088.
- Lee, J. D., J. K. Yu, Y. H. Hwang, S. Blake, Y. S. So, G. J. Lee, H. Nguyen, and J. G. Shannon. 2008. Genetic Diversity of Wild Soybean (*Glycine soja* Sieb. and Zucc.) Accessions from South Korea and Other Countries. *Crop Sci.* 48(2) : 606-616.
- Li, Y. H., M. J. Smulders, R. Z. Chang, and L. J. Qiu. 2011. Genetic diversity and association mapping in a collection of selected Chinese soybean accessions based on SSR marker analysis. *Conserv. Genet.* 12(5) : 1145-1157.
- Li, Y., R. Guan, Z. Liu, Y. Ma, L. Wang, L. Li, F. Lin, W. Luan, P. Chen, and Z. Yan. 2008. Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. *Theor. Appl. Genet.* 117(6) : 857-871.
- Li, Z., and R. L. Nelson. 2001. Genetic diversity among soybean accessions from three countries measured by RAPDs. *Crop Sci.* 41(4) : 1337-1347.
- Liu, K., and S. V. Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21(9) : 2128-2129.
- Mulato, B. M., M. Möller, M. I. Zucchi, V. Quecini, and J. B. Pinheiro. 2010. Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. *Pesqui. Agropecu. Bras.* 45(3) : 276-283.
- Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6(1) : 288-295.
- Pritchard, J., X. Wen, and D. Falush. 2007. Documentation for structure software: Version 2.2. Department of Human Genetics, University of Chicago; Department of Statistics, University of Oxford.
- Schmutz, J., S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, Q. Song, J. J. Thelen, and J. Cheng. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463(7278) : 178-183.
- Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18(2) : 233-234.
- Wang, L., R. Guan, L. Zhangxiong, R. Chang, and L. Qiu. 2006. Genetic diversity of Chinese cultivated soybean revealed by SSR markers. *Crop Sci.* 46(3) : 1032-1038.
- Yoon, M. S., J. Lee, C. Y. Kim, J. H. Kang, E. G. Cho, and H. J. Baek. 2009. DNA profiling and genetic diversity of Korean soybean (*Glycine max* (L.) Merrill) landraces by SSR markers. *Euphytica* 165(1) : 69-77.
- Yu, H., and Y. T. Kiang. 1993. Genetic variation in South Korean natural populations of wild soybean (*Glycine soja*). *Euphytica* 68(3) : 213-221.