Evaluation of Genetic Diversity in Thai Indigenous and Recommended Soybean Varieties by SSR Markers

O. Chotiyarnwong[1], P. Chatwachirawong[2], S. Chanprame[2] and P. Srinives[2,*]

[1]Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand
[2]Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand

**Corresponding author. Email: agrpss@yahoo.com*

Abstract

Information on genetic diversity and relationship among breeding materials is essential to a plant breeder for an efficient crop improvement. The objective of this study is to evaluate the genetic diversity and to group 160 Thai indigenous and recommended soybean varieties using 18 SSR markers. The SSR loci produced the total of 213 alleles with the average of 11.8 alleles and a mean gene diversity (*H*) of 0.831. Cluster analysis with the UPGMA method and principal component analysis (PCA) separated this population into 2 major groups. This indicates that the indigenous and recommended soybean varieties in Thailand have a small genetic differentiation based on SSR markers. However, they can still serve as a good source of genetic materials for cultivar improvement in Thailand and other tropical countries where soybean genetic resources are limited.

Keywords: *Glycine max*, soybean, SSR marker, genetic diversity, UPGMA, PCA

Introduction

Soybean (*Glycine max* (L.) Merr.) is the world's most important grain legume as a source of protein and oil. It belongs to family Leguminosae, genus *Glycine*, subgenus *soja*. *G. max* is the cultivated soybean while *G. soja* is its wild progenitor. The crop originated in China (Hymowitz, 1970) and has been cultivated for a century in the upper northern part of the Kingdom following rice harvesting in the dry season. The planted area also expanded to the lower northern part, the north-eastern region, and the central plain. Approximately 70% of the soybean planted area is currently in the Northern provinces.

At present, Thailand cannot produce enough soybean grain for domestic consumption. Limiting factors responsible for low yield are poor environment and lack of resistance to insect pests and diseases. Soybean breeders are trying to improve new varieties that give high yield. This can normally be accomplished through the survey, observation, and utilization of genetic diversity in soybean.

Biotechnology is a high potential tool for green revolution in agriculture**.** A benefit that can be derived from biotechnology is the development of DNA fingerprinting. This technique can be used to describe a variety and its genetic purity. The other potential benefits of DNA fingerprinting include gene mapping, marker-assisted selection and assessment of genetic diversity. Much emphasis is currently given to these areas of research.

In Thailand, indigenous soybeans have been collected and evaluated in the fields. A number of them showed the same morphology with different names, or the same name with different morphology. To protect the rights of the plant breeders and the farmers, as well as to expand the use of plant germplasm resources, the Thai Plant Variety Protection Act of 1999 was put into action. New

varieties that can be protected under the Act should show distinct characteristics from those of the current variety and possess the quality of uniformity and stability. Such a variety characterization can be determined from plant morphology. In addition, DNA fingerprinting technology can also help identifying the cultivars more precisely. It is technically possible to develop a protocol for evaluating genetic diversity and relationship among the Thai soybeans to provide an efficient management for a large number of germplasm. Newbury and Ford-Lloyd (1997) stated that the method used in measuring genetic diversity is an important factor determining strategy for conserving genetic resource. In a conservation program, methods based on genetic markers help identifying relationship and structure of the germplasm collected, determining how variation is distributed between individuals and accessions, and locating putative duplicated accessions. This information can provide efficient protocols in designing an optimal procedure to manage extensive germplasm collection and to conserve endangered genetic resources (Lee et al., 2006).

Simple sequence repeat (SSR), also known as microsatellite or short tandem repeat (STR) or simple sequence length polymorphisms (SSLPs) is a repeated nucleotide sequence of 2-6 base pair units. There are plenty of repeated nucleotide sequences which may be di-, tri- or tetra-nucleotide in plants. The SSR technique employs PCR to amplify DNA fragments by repeated cycles of DNA denaturing, annealing and extension using DNA polymerase enzyme. The resulting DNA was separated by banding on gel (Akkaya et al., 1992). SSR analysis has many advantages such as rapid, reliable (Diwan and Cregan, 1997; Ribaut et al., 1997), abundance (Lagercrantz et al., 1993), co-dominant (Akkaya et al., 1992; Brunel, 1994), high heterozygosity (Powell et al., 1996), and high polymorphism (Akkaya et al., 1995). Presently the SSR marker technology is an efficient tool for investigating genetic diversity for facilitating conservation and management of plant genetic resources.

To gain a better understanding of genetic relationship among germplasm of indigenous and recommended soybeans in Thailand, we evaluated their genetic diversity by examining the length polymorphism of alleles found in 18 SSR loci from different linkage groups. The information will be useful for soybean genetics and breeding not only in Thailand but also in the other tropical countries where availability of adapted germplasm is very limited.

Materials and Methods

Plant Materials

The plant materials were 160 Thai soybean accessions consisting of 149 indigenous and 11 recommended varieties. Among the indigenous ones, 62 were collected from upper northern part, 32 from lower northern part, 16 from central plain, 8 from northeastern part and 31 with no record of their origin. The 11 recommended varieties are Nakhon Sawan 1 (NS1), Chiang Mai 2 (CM2), CM3, CM60, SJ1, SJ2, SJ3, SJ4, SJ5, Sukhothai 1 (ST1) and Chakkrabhandhu no.1 (CKB1).

Analysis of SSRs

DNA was extracted from unfolded leaflets of 5 plants per accession, using the method described by Christiansen et al. (2002). Eighteen tri-nucleotide repeat SSRs that were located in each of 18 different molecular linkage groups selected from Cregan et al. (1999) (Table 1) were used in the present study. The PCR reaction mixture contained 30 ng of total genomic DNA, 0.25 μM of 5' and 3' end primers, 200 μM of each dNTP, 0.4 units of *Taq* polymerase (Fermentas), 1.5 mM of $MgCl_2$, and 1x PCR buffer in a total volume of 10 μl. PCR reactions were performed with a PTC-100™ Programmable Thermal Controller (MJ Research, Inc.) using the following program: 32 cycles at 94°C for 30s, 55°C for 30s and 72°C for 30s (Abe et al., 2003). After the amplification, a half volume of sequencing dye was added in the PCR product, the 10 μl of the product was loaded and separated by 4% agarose gel electrophoresis, using 50 bp DNA ladder (GeneRuler™) as the size standard.

Data Analysis

Allelic frequencies at each of the 18 SSR loci were roughly estimated based on the ratio of peak area of SSR fragments. The allelic frequency data was subjected to the Principal Component Analysis (PCA) in order to visualize the pattern of genetic variation observed at the marker loci.

Table 1 Number of alleles, allele size range and genetic diversity index at 18 SSR loci when tested on 160 Thai soybean varieties.

Locus	Linkage group	No. of alleles	Allele size range (bp)	Genetic diversity index (*H*)
Satt236	A1	16	177-270	0.9136
Satt228	A2	10	193-262	0.7980
Satt197	B1	10	127-195	0.8320
Satt565	C1	10	142-195	0.8389
Satt277	C2	20	123-270	0.9036
Satt184	D1a+Q	11	128-196	0.8695
Satt458	D2	22	134-260	0.9160
Satt045	E	6	122-157	0.6952
Satt114	F	9	72-117	0.7380
Satt038	G	8	144-195	0.7986
Satt253	H	9	119-166	0.8097
Satt367	I	12	167-229	0.8323
Satt431	J	10	180-232	0.8134
Satt055	K	15	64-148	0.9014
Satt156	L	9	184-233	0.8494
Satt463	M	13	106-153, 203-227	0.8250
Satt009	N	16	139-256	0.8509
Satt262	O	7	227-262	0.7762
Average		11.83		0.8312

The genetic diversity index (*H*) based on allele frequencies was calculated for each SSR using the formula of Nei (1973): $H = 1 - \Sigma P_i^2$ where P_i is the frequency of the i^{th} SSR allele present in the examined accessions. A dendrogram was constructed based on Euclidean distance of product size. The distance matrix was subjected to cluster analysis using the UPGMA (unweighted pair-group method with arithmetic average) with the help of SHAN-Clustering from NTSYS-pc version 2.0 (Rohlf, 1998).

Results and Discussion

The scatter diagram of 160 soybean accessions on two-dimensional plotting obtained from Principal Component Analysis of allele frequencies is presented in Figure 1. The proportion of the variation explained by each component to the whole variation was 39.44% for the first-dimension (horizontal axis) and 16.74% for the second-dimension (vertical axis). The first-dimension clearly separated the 160 soybean accessions into two groups, i.e. completely splitted the 31 accessions from 129 accessions. This result also reflected in the dendrogram of UPGMA (Figure 2), where each group was formed into each cluster. The efficiency of grouping of 160 soybean accessions tested with cophenetic correlation showed in Figure 3. The highly significant cophenetic correlation of r = 0.906 indicated that the germplasm was very well grouped.

The present analysis detected a high level of length polymorphism at 18 SSR loci in the soybean accessions. In each locus, the length of the alleles varies in steps of three bases (Table 1), and were nearly continuously distributed, except for Satt463

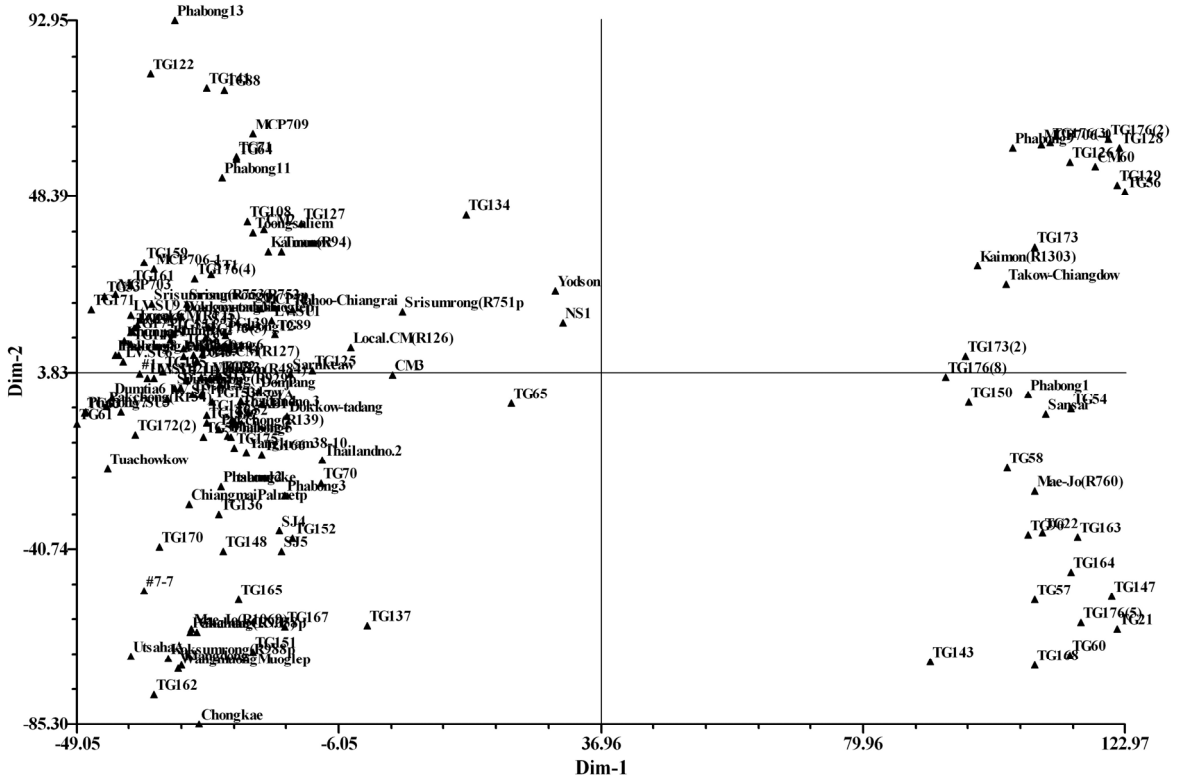

Figure 1 The scatter plot of 160 soybean accessions on the first-dimension (horizontal (X) axis) and the second-dimension (vertical (Y) axis) obtained from a Principal Component Analysis of allelic frequencies at 18 SSR loci.

that the allele lengths were distributed in two different ranges (106-153 bp and 203-227 bp). Each of the loci had between 6 and 22 alleles (11.83 on average). Satt458, located on linkage group D2 had the highest number of alleles (22), while Satt045 on linkage group E had the lowest number (6). The genetic diversity index (*H*) ranged from 0.695 for Satt045 to 0.916 for Satt458, with an average of 0.831. Similar to that reported by Huang et al. (2002) our study revealed that the genetic diversity index seemed to be positively correlated with the number of alleles at a given locus. Thus the number of alleles may be used for evaluation of genetic diversity. Each SSR locus had different allele size because of the number of repeats of each fragment resulted from slippage (slip-strand mispairing) in DNA replication (Schlötterer and Tautz, 1992) or unequal crossing-over (Smith, 1976). Replication slippage is considered a major factor affecting the repeated number for short tandem repeat sequences, whereas unequal crossing-over is thought to cause a very large number of alleles for long tandem repeat arrays (Lee et al., 2006).

The high level of allelic variation in this study facilitates detection of genetic diversity and genetic relationship among these soybean accessions. Although soybean is a self-pollinated crop, the variation among individuals within each accession can occur from natural out crossing, mutation and seed admixture. However, such a variation was not evaluated in this study.

A cluster analysis with UPGMA indicated ranges of Euclidean distance between 12.37 to 167.03. At the Euclidean distance of 128.37, the germplasm could be divided into 2 groups whereas at the distance of 89.70, they could be divided into 14 groups (Figure 2). The variety names in each group were listed in Table 2. Although most of these varieties were introduced into Thailand from various unknown sources in the past, a sizable genetic diversity existed considering the SSR variation. The results obtained from this study can be used to explain the relationship between Thai soybean germplasm as well as to select for parents in a breeding program, to avoid genetic relatedness. This finding can be added into the scarcely available information on tropical soybean genetic resource.

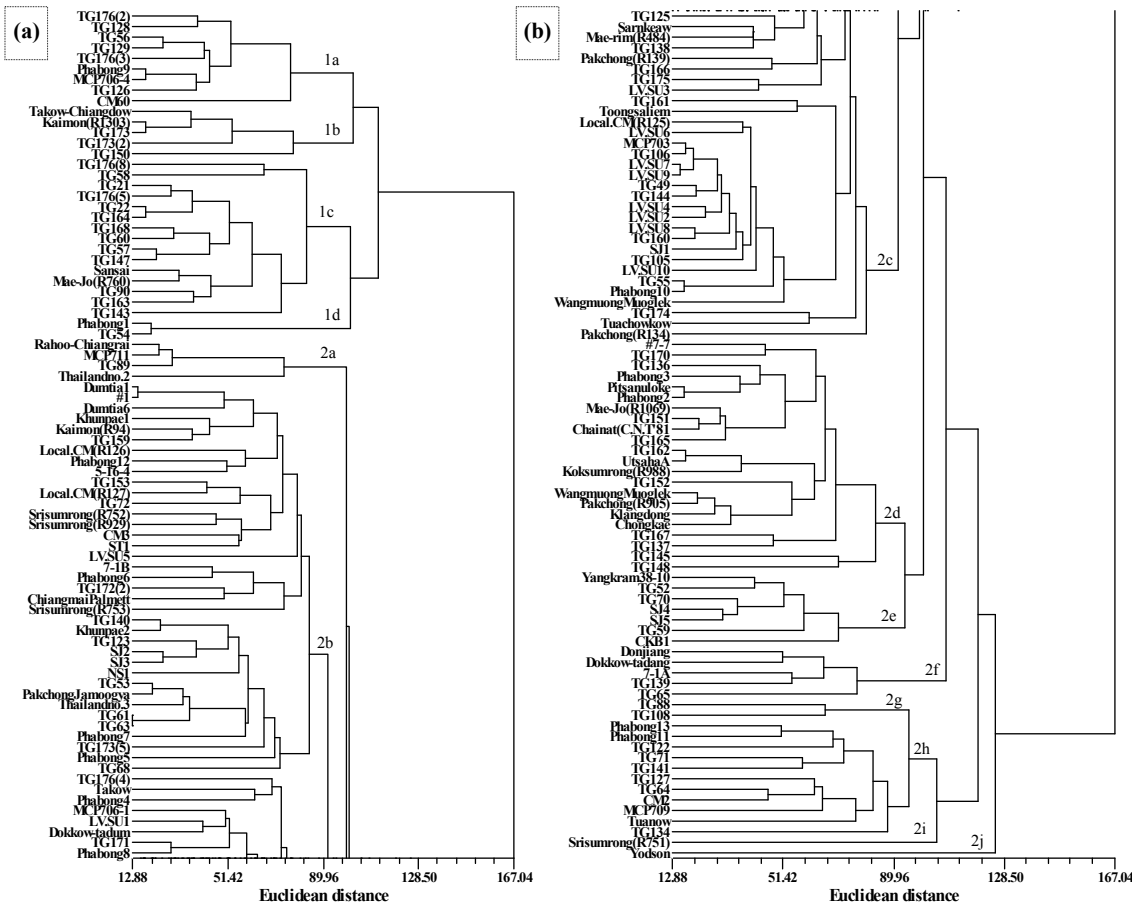

Figure 2 A UPGMA dendrogram representing genetic relationship among 160 soybean accessions. The dendrogram was constructed based on Euclidean distances between accessions.

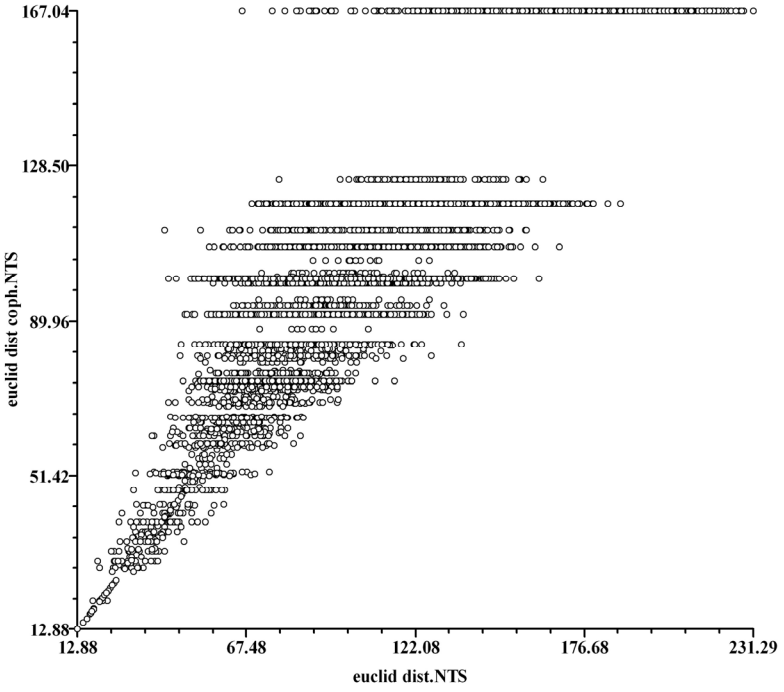

Figure 3 A scatter diagram showing efficiency of grouping the 160 soybean accessions as tested by cophenetic correlation.

The information on genetic diversity and relationship among the breeding materials is essential to plant breeders for the efficient improvement of crop species. Genetic similarity estimates among cultivars are helpful in selecting parental combinations as well as in maintaining genetic diversity in breeding programs. In general, high yielding ability and genetic dissimilarity from another parent to be crossed with are considered as two preferring characteristics of a good cross combination. Crossing between genetically diverse parents are expected to show a larger genetic variation among progenies than that between closely related parents (Messmer et al., 1993). However, the repeated crossing of

Table 2 List of 160 soybean varieties assigned into 2 major groups (1 and 2) based on UPGMA dendrogram of Figure 1. Both groups are further subdivided into 4 (a to d) and 10 (a to j) minor groups, respectively.

Groups (number of variety)	Variety name[1/]
1a (9)	TG 176 (2), TG 128, TG 56, TG 129, TG 176 (3), Phabong 9, MCP 706-4, TG 126, CM60
1b (5)	Takow-Chiangdow, TG 173 (2), Kaimon (R1303), TG 173, TG 150
1c (15)	TG 176 (8), TG 21, TG 176 (5), Sansai, Mae-Jo (R760), TG 22, TG 164, TG 90, TG 58, TG 143, TG 168, TG 60, TG 57, TG 147, TG 163
1d (2)	Phabong 1, TG 54
2a (4)	Rahoo-Chiangrai, MCP 711, TG 89, Thailand no.2
2b (37)	Dumtia 1, #1, Khunpae 1, Kaimon (R94), TG 159, LV.SU 5, Dumtia 6, Local.CM (R126), Phabong 12, 7-1-B, TG 153, TG 172 (2), TG 140, Khunpae 2, Phabong 6, TG 53, PakchongJamoogyai (R1273), TG 123, Srisumrong (R753), Local.CM (R127), TG 68, TG 72, Srisumrong (R752), Srisumrong (R929), 5-16-4, TG 173 (5), Phabong 7, TG 61, TG 63, Phabong 5, Thailand no.3, Chiangmai Palmetto, NS1, SJ2, SJ3, CM3, ST1
2c (39)	TG 176 (4), TG 161, MCP 706-1, Toongsaliem, TG 175, LV.SU 3, Local.CM (R125), MCP 703, TG 106, LV.SU 10, LV.SU 4, LV.SU 2, LV.SU 7, LV.SU 9, Pakchong (R139), LV.SU 1, TG 166, LV.SU 6, Takow, TG 125, LV.SU 8, TG 105, TG 160, Sarnkeaw, TG 49, TG 138, TG 174, Dokkow-tadum, Phabong 4, TG 144, TG 171, Phabong 8, Mae-rim (R484), TG 55, Phabong 10, Wangmuong Muoglek (R930), Pakchong (R134), Tuachowkow, SJ1
2d (22)	# 7-7, TG 136, Phabong 3, Pitsanuloke, Phabong 2, TG 170, TG 167, Mae-Jo (R1069), TG 148, TG 145, TG 162, Utsaha A, Koksumrong (R988), TG 137, TG 152, Wangmuong Muoglek (R1053), Chongkae, Pakchong (R905), Klangdong, TG 151, Chainat (C.N.T '81), TG 165
2e (7)	Yangkram 38-10, TG 70, TG 59, TG 52, SJ4, SJ5, CKB1
2f (5)	Donjiang, 7-1A, Dokkow-tadang, TG 65, TG 139
2g (2)	TG 88, TG 108
2h (11)	Phabong 13, Phabong 11, TG 127, MCP 709, TG 71, TG 134, Tuanow, TG 141, TG 122, TG 64, CM2
2i (1)	Srisumrong (R751)

[1/] In each group, the variety names are listed from top to bottom, for example, the first and the last accessions of group 1a are TG 176 (2) and CM60, respectively.

closely related elite parents to develop new cultivars is common in many crop species, especially tropical soybeans which suitable germplasm is extremely limited. This practice is likely narrowing their genetic base and increasing their vulnerability to potentially losses from widespread diseases and pests (Walsh, 1981). Therefore, the preservation of genetic diversity in breeding programs has been a serious issue (Jong et al., 2006). For developing countries with a limited facility to store germplasm, the idea of core collection should be initiated. The idea is to keep a core genetic stock which includes most recommended varieties, varieties with certain useful traits, and those with genetically diverse (based on SSR markers in our case). This helps reducing the storage space and field facilities for seed rejuvenation, yet makes sure that the plant breeders still maintain all the useful genes required in the future breeding activities.

Conclusions

The SSR loci produced an average of 11.83 alleles and a mean genetic diversity index (*H*) of 0.831 among 160 indigenous and recommended Thai soybean varieties. Cluster analysis with the UPGMA method gave a Euclidean distance between 12.37 to 167.03, and the cophenetic correlation gave a highly goodness of fit. The multivariate analyses of the 18 SSR loci clearly allocated the observed diversity into two major groups. The results obtained in this study indicated that the germplasm has a rather small genetic differentiation based on SSR markers. However, considering the overall marker variation, it is sufficient to identify suitable parents for soybean hybridization and selection. Yet the information is useful for breeding of tropical soybean, since only a very little research has been done regarding germplasm collection and diversity evaluation.

Acknowledgments

This research is supported by the Center for Agricultural Biotechnology, the Thailand Research Fund, and the National Research Council of Thailand. We are thankful to Chiang Mai Field Crops Research Center, Department of Agriculture for providing seed materials used in this study.

References

Abe, J., D.H. Xu, Y. Suzuki, A. Kanazawa and Y. Shimamoto. 2003. Soybean germplasm pools in Asia revealed by nuclear SSRs. Theor. Appl. Genet. 106: 445-453.

Akkaya, M.S., A.A. Bhagwat and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Akkaya, M.S., R.C. Shoemaker, J.E. Specht, A.A. Bhagwat and P.B. Cregan. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. 35:1439-1445.

Brunel, D. 1994. A microsatellite marker in *Helianthus annuus* L. Plant Mol. Biol. 24:397-400.

Christiansen, M.J., S.B. Andersen and R. Ortiz. 2002. Diversity changes in an intensively bred wheat germplasm during the 20th century. Mol. Breed. 9: 1-11.

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 soybean genome. Crop sci. 39:1464-1490.

Diwan, N. and P.B. Cregan. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. 95:723-733.

Huang, X.O., A. BÖrner, M.S. RÖder and M.W. Ganal. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 105:699-707

Hymowitz, T. 1970. On the domestication of the soybean. Econ. Bot. 24:408-421

Jong, S.K., S.H. Kim, S.H. Woo and H.S. Kim. 2006. Comparison of pedigree- and DNA-based genetic similarity among Korean soybean cultivars. Korean J. Breed. 38:19-25.

Lagercrantz, U., H. Ellegran and L. Anderson. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucl. Acids Res. 21:1111-1115.

Lee, J.K., J.W. Chung, Y.J. Park, K.H. Ma, H.K. Kang, J.R. Lee, J.G. Kwag, N.S. Kim, T.S. Kim and S.Y. Lee. 2006. Assessment of genetic diversity of Korean landrace rice accessions (*Oryza sativa* L.) by microsatellite analysis. Korean J. Breed. 38(2):75-82

Messmer, M.M., A.E. Melchinger, R.G. Herrmann, and J. Boppenmaier. 1993. Relationships among early European maize inbreds: II. Comparison of pedigree and RFLP data. Crop Sci. 26:672-676.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70:3321-3323

Newbury, H. and B.V. Ford-Lloyd. 1997. Esimation of genetic diversity, pp 192-206. *In* N. Maxted, B.V. Ford-Lloyd and J.G. Hawkes, (eds.), Plant Genetic Conservation the *In situ* Approach. Chapman and Hall, London

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey and A. Rafalski. 1996. The comparison of

RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2:225-238.

Ribaut, J.M., X. Hu, D. Hoisington and D. González-de-León. 1997. Use of STSs and SSRs as rapid and reliable preselection tools in a marker-assisted selection-backcross scheme. Plant Mol. Biol. Rep. 15:154-162.

Rohlf, F.J. 1998. Numerical Taxonomy and Multivariate Analysis System, NTSYS-pc Program Manual Version 2.0. Applied Biostatistics Inc., 3 Heritage Lane, Setauket, NY 11733

Schlötterer, C. and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. Nucl. Acids Res. 20:211-215.

Smith, G.P. 1976. Evolution of repeated DNA sequences by unequal crossover. Science 191:528-535.

Walsh, J. 1981. Genetic vulnerability down on the farm. Science 214:161-164.

Manuscript received 10 September 2007, accepted 26 April 2008