RESEARCH ARTICLE
J. Crop Sci. Biotech. 2008 (June) 11 (2) : 83 ~ 90

Genetic Diversity and Population Structure of Korean Soybean Landrace [*Glycine max* (L.) Merr.]

Gyu-Taek Cho[1,2], Jeongran Lee[2,*], Jung-Kyung Moon[3], Mun-Sup Yoon[2], Hyung-Jin Baek[2], Jung-Hoon Kang[2], Tae-San Kim[2], Nam-Chon Paek[1,*]

[1] *Department of Plant Science, Seoul National University, Seoul 151-921, Republic of Korea*
[2] *National Institute of Agricultural Biotechnology, RDA, Suwon 441-707, Republic of Korea*
[3] *National Institute of Crop Science, RDA, Suwon 441-857, Republic of Korea*

<Received April 15, 2008 / Accepted May 20, 2008>

Abstract

Two hundred and sixty Korean soybean landrace accessions were analyzed for polymorphism at 92 simple sequence repeat (SSR) loci. The 995 identified alleles served as raw data for estimating genetic diversity and population structure. The number of alleles at a locus ranged from three to 27 with a mean of 10.4 alleles per locus. F_{ST} values estimated by analysis of molecular variance (AMOVA) using SSR data set were 0.018, 0.027, and 0.016 for usage, collection site and maturity groups, respectively, indicating little genetic differentiation. The model-based clustering analysis placed the accessions into three clusters ($K = 3$) with 0.0503 of F_{ST}, indicating moderate genetic differentiation. Duncan's Multiple Range Test at $K = 3$ on the basis of 18 quantitative traits revealed that one cluster was mainly differentiated from the other two clusters by seed related traits and the other two clusters were differentiated from each other by biochemical traits. Genetic structure of Korean soybean landraces was differentiated by model-based clustering and supported by their phenotypic traits in part. This preliminary study could be the first step towards more efficient germplasm management and utilization of soybean landraces and helpful in association studies between genotypic and phenotypic traits in Korean soybean landraces.

Key words: *Glycine max*, Genetic diversity, Population structure, SSR

Introduction

Various kinds of soybeans [*Glycine max* (L.) Merr.] have been cultivated in East Asia as a result of being used as diverse food sources, and being adapted in different environment for a long time. Characteristic variations in embryonic axis colors, flower colors, pubescence colors, plant heights, maturity, leaf shapes, seed coat colors, hilum colors, seed sizes, protein contents and lipid contents of soybeans in Korea have been accumulated mainly due to selections of farmers in favor of diverse usages and high yields (Hong et al. 1988; Kwon et al. 1972b; Song et al. 1991). To maintain such genetic diversity, the National Genebank of Rural Development Administration, Korea (RDA-genebank, http://genebank.rda.go.kr), has conserved a large collection of soybean landraces, more than 7,000 accessions, that have been collected from a series of nationwide expeditions by the Korean Atomic Energy Research Institute, National Institute of Crop Science and RDA-genebank since 1969 (Yoon et al. 2003).

However, the vast amount of soybean collections has prevented the effective evaluation and utilization of genetic resources. The presence of highly related individuals may lead to deviation from model hypotheses resulting in unreliable group assignment when population structure is inferred from neutral marker loci. In this respect, Camus-Kulandaivelu et al. (2007) advised to use a core collection or a subsample representing the genetic diversity of the whole panel, removing families of related accessions.

Their high levels of allelic diversity facilitate the detection of the fine structure of diversity more efficiently than an equal number of restriction fragment length polymorphism (RFLP). Thus, amplified fragment length polymorphism (AFLP), single

*** To whom correspondence should be addressed**
Nam-Chon Paek
E-mail: ncpaek@snu.ac.kr
Tel: +82-2-880-4543
Jeongran Lee
E-mail: jrlee19@rda.go.kr
Tel: +82-31-299-1832

nucleotide polymorphism (SNP) loci, and simple sequence repeat (SSR) loci are particularly useful for the study of population structure and demographic history of domesticated species (Garris et al. 2005). Especially, SSR markers have proven to be important tools in soybean genetics (Akkaya et al. 1992; Cregan et al. 1999) and have been widely applied to studies in the genetic diversity of soybean germplasm (Abe et al. 2003; Brown-Guedira et al. 2000; Wang et al. 2006). These diversity studies not only provide useful information for understanding the genetic bases of various soybean gene pools established in different geographic regions, but also facilitate the selection of sources of new genes for yield and quality improvements.

The objectives of this study are to evaluate the genetic diversity of Korean soybean landraces using SSR markers, to analyze and characterize population structure within Korean soybean landraces, and then to examine degree of genetic differentiation of morphologically and genetically defined groups.

Materials and Methods

Plant materials and grouping criteria

Two hundred and sixty germplasm accessions of a core collection developed from 2,765 Korean soybean landraces using six SSR markers were used in this study. To evaluate the degree of differentiation among the groups, the accessions were grouped by usages, collecting sites, maturity group and the model-based clustering. Usages were grouped as sauce soybean, sprouting soybean, soybean for cooking with rice and others based on the criteria of Hong et al. (1988) and Yoon (1998). Collecting sites were grouped by five provinces of South Korea, namely, Gyeonggi, Gangwon, Chungcheong, Jeolla, and Gyeongsang provinces. In this case, only 129 accessions possessing the collection site data out of the core collection were used. Maturity group was divided into early (< 113 days), intermediate (113-140 days), and late (> 140 days) by trisecting the days to maturity observed in the field. The model-based clustering was identified by using STRUCTURE software (Pritchard et al. 2000).

Phenotypic data

Each accession was sown as single row spaced 60 cm apart. Plant-to-plant spacing within the row was 15 cm. The cultivation was applied at Suwon, Korea, from 2005 to 2006. The characters were observed for 11 morphological traits according to the soybean descriptors of genetic resources (RDA, 1986). Days to flowering were recorded as the mean emergence days to the date when 50% plants started flowering. Days to maturity were recorded as the mean dates to the date when 95% pods matured and most of leaves faded and fallen. Plant height (cm) was measured. The number of branches per plant, nodes per plant, pods per plant, and the number of seeds per pod were counted as a mean of ten plants. After harvest, seed weight was recorded as 100-seed weight (g), seed length (mm), seed width (mm) and seed thickness (mm) were measured as means of ten seeds using calipers.

Crude compositions of protein, oil, oligosaccharide, sucrose, oleic acid, linoleic acid and linolenic acid were analyzed by NIRSystem model 6500 (FOSS NIRSystems Inc., Raurel, MD, USA) followed by Choung et al. (2001) using the seed samples of 10 g ground by PERTEN LABORATORY MILL 3303 (Seedburo Equipment Co., IL, USA).

DNA extraction

Genomic DNA was extracted from fresh leaf tissues of seedlings (10-12 days old) for each accession. Sampled leaf tissues were ground in liquid nitrogen. DNA was extracted by powder of frozen tissue using DNeasy Plant Mini Kits (Qiagen #69106). The precipitated DNA was dissolved in 10mM TE buffer and stored at -20 ℃ until use. The final concentration was adjusted to 20 ng/㎕ for PCR reaction.

SSR genotyping

A total of 92 SSR markers were used in this study (Table 1). These loci were uniformly distributed across the 20 soybean genetic linkage groups (Cregan et al. 1999; Song et al. 2004). PCR amplification was performed on each of the 260 soybean genomic DNA using primers for each SSR locus. Reaction mixture contained 20 ng of soybean genomic DNA, 4 ㎕ of dNTP mixture containing 10 mM of each nucleotide (final concentration of 0.2 mM each) (Promega), 1X PCR buffer containing 50 mM of KCl, 0.15 ㎛M 3' and 5' end primers, and 1 unit of Taq DNA polymerase (Promega). All PCR reactions consisted of 1 cycle of 12 s with incubation at 94 ℃, 32 cycles of 25 s for denaturation at 94 ℃, 25 s for annealing at 46 ℃, and 25 s for extension at 68 ℃, and the final step held constant at 4 ℃, on PTC-100 thermocycler (MJ Research, MA, USA). After PCR amplification, SSR alleles were resolved on an ABI-PRISM 3100 DNA sequencer (Applied Biosystems, CA, USA) using GENESCAN v.3.7 software, and sized precisely against 6-carboxy-X-rhodamine (ROX) molecular size standards using GENOTYPER v.3.7 software (Applied Biosystems, CA, USA).

Table 1. List of SSR primers used in this study. Marker designations are from Song et al. (2004).

Linkage Groups	Number of markers per linkage group	Markers used
A1	4	Satt165, Satt042, Satt155, Satt225
A2	6	Satt390, Satt187, Satt329, Satt409, Satt228, Satt429
B1	1	Satt197
B2	5	Satt577, Satt168, Satt070, Satt063, Satt560
C1	3	Satt194, Satt190, Satt161
C2	6	Satt227, Satt322, Satt286, Satt134, Satt307, Satt357
D1a+Q	5	Satt532, Satt179, Satt077, Satt071, Satt129
D1b+W	18	Satt558, Satt634, Satt296, Satt542, Satt266, Satt141, Satt290, Satt506, Satt005, Satt600, Satt537, Satt579, Satt189, Satt350, Satt428, Satt041, Satt172, Satt271
D2	8	Sctt008, Satt135, Satt014, Satt002, Satt154, Satt082, Satt186, Satt031
E	3	Satt384, Satt185, Satt045
F	4	Satt146, Satt160, Satt072, Sat_074
G	3	Satt309, Satt012, Satt288
H	2	Satt314, Satt181
I	6	Satt571, Satt367, Satt239, Satt354, Satt270, Satt148
J	2	Satt285, Satt183
K	4	Satt242, Satt178, Satt046, Satt001
L	3	Satt182, Satt143, Satt166
M	2	Satt175, Satt336
N	3	Satt152, Satt009, Satt022
O	4	Satt358, Satt173, Satt478, Satt153
Total	92	

Data analysis

Basic statistics including the number of alleles, gene diversity, and polymorphism information content (PIC) were calculated using PowerMarker v.3.0 (Liu and Muse 2005). F_{ST} (pairwise estimates of the correlation of alleles within subpopulations) for groupings by usage, collection site (province), maturity group, and the model-based (Bayesian) cluster was calculated by using an analysis of molecular variance (AMOVA) approach in ARLEQUIN v.3.11 (Excoffier et al. 2005) with the exclusion of one marker of missing value over 0.5% (Satt134). The genetic structure of the germplasm analyzed was also investigated with an alternative approach by using the model-based clustering algorithm (STRUCTURE v.2.2), which identifies subgroups with distinct allele frequencies (Pritchard et al. 2000). Differently from the cluster analysis based on the calculation of a pairwise distance matrix and on a non-overlapping graphical representation in the model-based method each accession is allowed to have membership in several different subgroups with membership coefficients totaling one. The model-based clustering analysis was performed using all the 260 accessions and five independent runs of STUCTURE for each K value (the number of subpopulations) from two to 13 without prior population information. Runs were carried out by setting for 200,000 iterations, of which only the last 50,000 were recorded, and assuming an admixture model with correlated allele frequencies (Falush et al. 2003). For each K value, the runs showing the highest posterior probability of data were considered. The true value of K was detected by an ad hoc quantity based on the second order rate of change of the likelihood function with respect to K (ΔK) (Evanno et al. 2005);

$$\Delta K = m(|L(K+1) - 2L(K) + L(K-1)|)/s[L(K)]$$

where, L(K) is Ln P(D), the posterior probability of the data for a given K, Pr($X|K$) in STRUCTURE output, s[L(K)] is the standard deviation of L(K), and m is mean in the parenthesis. ΔK shows a clear peak at the true value of K. Analysis of variance (ANOVA) and Duncan's Multiple Range Test were analyzed by using SAS software (SAS Institute 2004).

Results and Discussion

Phenotypic diversity

Phenotypic characteristics are provided in Table 2. Large variation was observed in morphological and biochemical traits. The computed coefficient of variation was the highest for the 100-seed weight (36.7%) followed by the number of pods per plant (32.73%) and plant height (31.57%), whereas the lowest

Table 2. Basic statistics for morphological and biochemical traits.

Traits	Mean	SD	Min	Max	Range	CV[a]
100-seed weight (g)	21.30	7.82	6.9	42.5	35.6	36.70
Seed length (mm)	8.06	1.13	5.22	11.15	5.93	14.06
Seed width (mm)	7.17	1.00	5.01	9.52	4.51	13.96
Seed thickness (mm)	5.84	0.82	3.34	8.06	4.72	13.98
Days to flowering (days)	56.46	4.77	39.5	69.0	29.5	8.46
Days to maturity (days)	130.40	9.87	85.0	156.0	71.0	7.57
Plant height (cm)	69.86	22.06	32.8	206.4	173.6	31.57
No. of branches per plant	4.03	1.08	1.4	10.2	8.8	26.73
No. of nodes per plant	16.93	2.76	9.3	27.2	17.9	16.28
No. of pods per plant	60.48	19.80	22.6	135.8	113.2	32.73
No. of seeds per pod	2.01	0.15	1.6	2.74	1.14	7.70
Protein (%)	43.04	2.25	36.61	49.38	12.76	5.24
Oil (%)	14.74	1.53	9.52	18.58	9.06	10.39
Oligosaccharide (%)	7.71	0.94	4.71	10.25	5.54	12.17
Sucrose (%)	5.50	1.01	2.76	8.20	5.44	18.35
Oleic acid (%)	25.29	4.17	14.76	42.97	28.21	16.50
Linoleic acid (%)	50.62	3.45	34.42	57.77	23.35	6.82
Linolenic acid (%)	12.41	1.48	8.41	15.90	7.497	11.91

[a] SD: standard deviation, Min: minimum value, Max: maximum value, Range: difference from maximum value to minimum value, CV: coefficient of variance.

was for protein content (5.24%). This wide range of phenotypic variation for the Korean soybean landraces can be explained by the human selection focusing on the usage of interest, accumulation of diverse variation through long cultivation history and a series of nationwide expeditions, which have led to a conservation of variation in RDA-Genebank (Kwon et al. 1972a; Kwon et al. 1972b; Song et al. 1991; Yoon et al. 2003).

Allelic diversity

In all the tested alleles, 955 alleles were detected among the 260 accessions at the 92 SSR loci on 20 homologous linkage groups (Table 3). The number of alleles per locus varied from three (Satt071, Satt072, Satt082, Satt227, Satt271, Satt322, and Satt384) to 27 (Sat_074) with a mean of 10.4 alleles. The estimate of 10.4 alleles per locus implied that a high level of diversity was present in Korean soybean landraces, compared with 7.5 alleles per locus among 91 Korean soybean cultivars using 20 SSR loci (Kim et al. 2006) and 6.3 alleles per locus using 37 SSR loci among 45 Canadian soybean cultivars released from 1934 to 2001, and 37 exotic germplasm accessions (Fu et al. 2007). These facts corresponded to the reports that modern crop breeding creates genetic bottlenecks (Hyten et al. 2006). The genetic bottlenecks can decrease genetic diversity, change allele frequencies, and lose many genes or alleles that exist in ancestral lines. This has taken place in the U.S. during the last 40 years through the extensive breeding (Kisha et al. 1998). There are also reports which detected 10.2 alleles per locus among 39 elite genotypes and 40 plant introductions using 74 SSR loci (Narvel et al. 2000), and 12.2 alleles per locus among 129 Chinese accessions including 122 landraces and seven cultivars at the 60 SSR loci (Wang et al. 2006). The significant level of diversity in 129 Chinese accessions might have resulted from those collected from seven ecotypes from 2183′ 59″ N to 49° 81′ 09″ N and analyzed with sixty SSRs including four AT or CT motifs which generally show high polymorphism.

Based on the results of phenotypic and allelic variation studies, we expect that many useful genes or alleles may exist in Korean soybean landraces. Burham et al. (2002) also reported a genetic distinction of South Korean soybean accessions from the US germplasm, recommending breeders to evaluate the South Korean accessions for resistance to *Phytophthora sojae*.

Table 3. Summary statistics by linkage groups.

LG [a]	No [b]	No. of Alleles [c]	Gene Diversity [d]	PIC [e]
A1	4	8.500	0.469	0.436
A2	6	10.833	0.625	0.594
B1	1	13.000	0.844	0.829
B2	5	11.200	0.655	0.624
C1	3	10.333	0.693	0.663
C2	6	7.000	0.631	0.575
D1a+Q	5	7.200	0.631	0.583
D1b+W	18	9.222	0.626	0.584
D2	8	8.375	0.618	0.572
E	3	10.667	0.718	0.680
F	4	13.750	0.752	0.718
G	3	13.000	0.772	0.743
H	2	6.000	0.615	0.562
I	6	9.000	0.625	0.581
J	2	6.500	0.529	0.458
K	4	11.500	0.755	0.732
L	3	7.333	0.628	0.576
M	2	10.500	0.792	0.764
N	3	16.000	0.746	0.720
O	4	11.250	0.603	0.585
Mean [f]		10.38	0.655	0.615

[a] LG: linkage group.
[b] No: number of markers used on each linkage group.
[c] No. of Alleles: number of alleles per linkage group.
[d] Gene Diversity: Nei's gene diversity (1973).
[e] PIC: Polymorphic Information Content.
[f] Mean per locus.

Population structure

As fixation indices (F_{ST}) measure the amount of differentiation among subpopulations derived from the subdivision of an original population (Wright 1978), values for F_{ST} range from 0 for non-differentiation to 1 for complete differentiation between an original population and its subpopulations, respectively. Wright (1978) suggested qualitative guidelines for the interpretation of F_{ST}. The range from 0 to 0.05 may be considered as indicating little genetic differentiation, the ranges from 0.05 to 0.15, 0.15 to 0.25, and values of F_{ST} above 0.25 indicate moderate, great, and very great genetic differentiation, respectively. In this study, F_{ST} values were 0.018, 0.027, and 0.016 for usage

Table 4. AMOVA by usage, collection site, and maturity groups.

Categories	Number of Accessions	Variation among populations (F_{ST})	Variation within populations	*P*-value*
Usage groups [a]	260	0.018	0.982	< 0.001
Collection sites [b]	129	0.027	0.972	< 0.001
Maturity groups [c]	260	0.016	0.984	

The type of hierarchical AMOVA implemented here was with genotypic data, one group of populations and no. within individual level.
* Significance Level = 0.05.
[a] Usage groups: sauce soybean (88 accessions), sprouting soybean (80 accessions), soybean for cooking with rice (87 accessions) and others (80 accessions).
[b] Collection sites: five provinces of South Korea, namely, Gyeonggi (5 accessions), Gangwon (16 accessions), Chungcheong (15 accessions), Jeolla (26 accessions), and Gyeongsang (71 accessions) province.
[c] Maturity groups: maturity was grouped as early (17 accessions), intermediate (147 accessions), and late (96 accessions) by trisecting the maturity observed in the field.

groups, collection site groups, and maturity groups, respectively, representing little genetic differentiation (Table 4). Burnham et al. (2002) assumed that there is a distinct genetic diversity within South Korean soybean landraces but this might not be related to geographical location along a latitudinal or longitudinal gradient based on the SSR markers and accessions used in their study. All Korean soybean landraces adapted to the Korean environments probably formed one group. It is supposed that the genetic base of Korean soybean landraces is so complex and diverse that the accessions could not be clustered with some phenotypic traits.

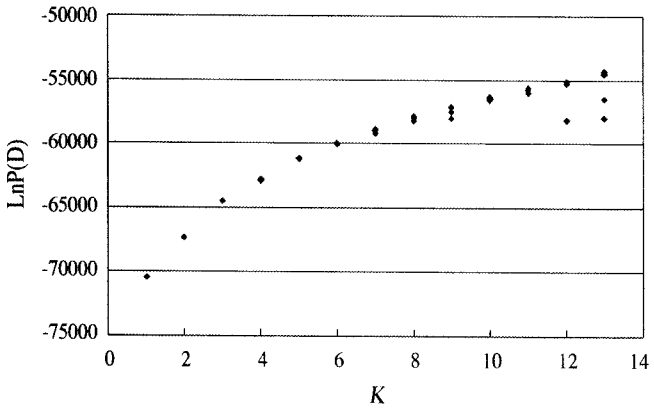

Fig. 1. The Bayesian posterior probability of data [Ln P(D)] with increasing *K*. The Bayesian posterior probability of data steadily improved until $K = 3$, to a little lower extent, until $K = 13$, but not clear.

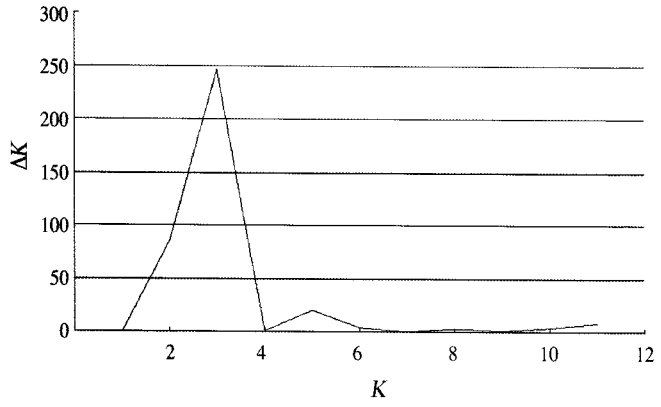

Fig. 2. Magnitude of ΔK as a function of *K*. The modal value of this distribution is the true *K* or the uppermost level of structure, here three clusters.

The log likelihood from model-based clustering increases until the real *K* is reached and then levels off or continues to increase slightly (Evanno et al. 2005). According to the STRUCTURE results, the log likelihood steadily improved until $K = 3$, to a little lower extent, until $K = 13$, but not clear (Fig. 1). Based on the four steps for the graphical method allowing detection of the true number of groups *K* suggested by Evanno et al. (2005), the real structure showing a clear peak of the accessions was set at $K = 3$ (Fig. 2). It means that the 260 accessions should be divided into three subpopulations, namely, POP1, POP2, and POP3. In total, 148 accessions (56.9% out of 260 accessions) were clearly assigned to each single subpopulation, where 80% of their inferred ancestry derived from one of the model-populations, whereas 112 accessions (43.1% out of 260 accessions) in the sample were categorized as having admixed ancestry (Fig. 3).

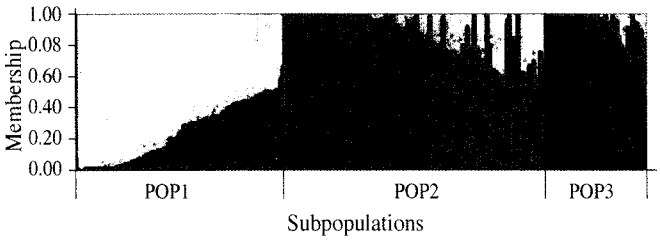

Fig. 3. Barplot showing genetic diversity structure for 260 Korean soybean accessions using the program STRUCTURE (v.2.2). Each accession is divided into a number of hypothetical sub-populations based on the proportional membership coefficients totaling 1 at $K = 3$. Each subgroup is represented by a different color as listed: red (POP1), green (POP2), blue (POP3).

Table 5. AMOVA for the three clusters identified in Fig. 2.

Source of variation	degree of freedom	Sum of squares	Variance components	Percentage of variation	*P*-value*
Among populations	P - 1 = 2	543.680	1.49705	5.03	< 0.001
Within populations	2N - P = 517	14627.122	28.29231	94.97	< 0.001
Total	2N - 1 = 519	15170.802	29.78936		

The type of hierarchical AMOVA implemented here was with genotypic data, one group of populations, and no. within individual level. P: Total number of populations. N: Total number of individuals for genotypic data.
* Significance Level = 0.05.

The overall pairwise F_{ST} value of the model-based clusters was 0.0503 of the variation ($P < 0.0001$), indicating a moderate degree of differentiation (Wright 1978; Table 5). POP3 showed higher gene diversity than replace with "POP1 and POP2". In contrast, POP1 had lower gene diversity than replace with "POP2 and POP3", suggesting that POP1 may be more homogeneous than the other POPs. Pairwise F_{ST} values were estimated to explain genetic differentiation among the subpopulations, i.e. POP1, POP2, and POP3. The sub-populational pairwise F_{ST} values ranged from 0.067 between POP1 and POP2, to 0.038 between POP2 and POP3 (Table 6).

Table 6. Genetic diversity and pairwise differentiation (F_{ST}) among the three sub-populations using the model-based clusters.

POP	Diversity				Pairwise F_{ST} values		
	No [a]	No. of Alleles [b]	Gene Diversity [c]	PIC [d]	1	2	Overall
1	119 (45)	7.913	0.602	0.558			
2	96 (69)	6.533	0.634	0.591	0.067*		
3	45 (34)	8.054	0.642	0.601	0.038*	0.060*	
							0.050*

[a] No: Number of accessions. Number in parenthesis refers to accession numbers with a probability higher than 80% of their inferred ancestry derived from each of the model-populations.
[b] Number of Alleles: Number of alleles per linkage group.
[c] Gene Diversity: Nei's gene diversity (1973).
[d] PIC: Polymorphic Information Content.
*F_{ST} values are all significant at $P < 0.0001$.

Comparison of population structure derived from the model-based clusters with phenotypic traits

Duncan's multiple range test (DMRT) was employed to determine whether there was any correlation between phenotypic traits and the three subpopulations of Korean soybean landraces grouped by STRUCTURE. On the basis of ANOVA, 12 phenotypic traits out of 18 revealed significant differences (α = 0.05). Multiple means comparisons using DMRT revealed that the 100-seed weight and seed thickness were the traits showing significant differences from all of the three subpopulations. Phenotypic characters such as seed length, seed width, oligosaccharide content, and sucrose contents were smaller in POP3 than in the other two POPs whereas the number of branches per plant and the number of pods per plant were higher. Linoleic acid contents and oleic acid contents showed significant differences only in POP2 and the phenotypic trait showing a significant difference only in POP1 was oil contents (Table 7). Genetic differentiation of POP3 from the other two POPs was mainly supported by seed related traits, complying with the report of Kwon et al. (1972b) that soybean landraces grown by farmers in Korea have been divergently selected on the basis of seed size. Korean farmers have preferred larger seeds for soybean sauce and tofu, and smaller seeds for soybean sprouts. POP1 and POP2 differed from each other in biochemical traits such as contents of oil, linoleic acid, and oleic acid.

The genetic structure present in a collection can be crop- or species-dependent, and for some crops, geographic origin and other criteria associated with genetic differentiation have been used as a basis for grouping (Erskine and Muehlbauer 1991; Fukunaga et al. 2005; Garris et al. 2005). In the recent studies applied with the model-based clustering, Jun et al. (2007) identified the existence of some population structures of soybean germplasm based on three origins (China, Korea, and Japan) and maturity groups and Agrama et al. (2007) identified eight main clusters for the rice accessions corresponding to the major geographic regions. The results of this study suggest that the population structure of Korean soybean landraces on a molecular level can be better differentiated by the model-based clustering rather than by the phenotypic or geographic clustering. And the model-based clustering is supported by their phenotypic traits in part.

Table 7. Comparison of means for phenotypic traits among sub-populations by Duncan's Multiple Range Test at $K = 3$.

Traits	POP1	POP2	POP3	F value [a]
100-seed weight (g)	24.11 b	26.60 a	15.30 c	70.08***
Seed length (mm)	8.46 a	8.55 a	7.33 b	40.61***
Seed width (mm)	7.55 a	7.74 a	6.41 b	64.34***
Seed thickness (mm)	6.13 b	6.48 a	5.18 c	86.34***
Days to flowering (days)	56.54 a	55.46 a	56.83 a	1.32
Days to maturity (days)	131.55 a	130.41 a	128.97 a	1.81
Plant height (cm)	66.12 b	69.91 ab	74.52 a	3.91*
No. of branches per plant	3.90 b	3.66 b	4.37 a	8.59***
No. of nodes per plant	16.49 b	16.77 ab	17.56 a	4.15*
No. of pods per plant	53.71 b	53.69 b	72.19 a	32.64***
No. of seeds per pod	1.99 a	2.02 a	2.02 a	1.56
Protein (%)	42.86 a	42.71 a	43.43 a	2.31
Oil (%)	15.14 a	14.42 b	14.40 b	7.83***
Oligosaccharide (%)	7.96 a	8.06 a	7.23 b	23.46***
Sucrose (%)	5.82 a	5.92 a	4.90 b	33.64***
Oleic acid (%)	24.95 b	27.61 a	24.64 b	9.26***
Linoleic acid (%)	51.14 a	48.51 b	51.00 a	11.41***
Linolenic acid (%)	12.30 a	12.64 a	12.43 a	0.93

Means with same letters in rows are not statistically different, with $\alpha = 0.05$.
[a] F value of ANOVA.
*** Significant at $P < 0.001$. ** Significant at $P < 0.01$. * Significant at $P < 0.05$.

This preliminary study could be the first step towards more efficient germplasm management and utilization of soybean landraces and helpful in association studies between genotypic and phenotypic traits in Korean soybean landraces. Further characterization of the accessions in this core collection will help to provide more insights into genetic relationships among accessions and to utilize them for various purposes by breeders and researchers in public and private sectors.

Acknowledgements

This work was supported by a grant (NIAB-08-05) from National Institute of Agricultural Biotechnology, Rural Development Administration, Republic of Korea.

References

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

Agrama HA, Eizenga GC, Yan W. 2007. Association mapping of yield and its components in rice cultivars. Mol. Breed. 19: 341-356

Akkaya MS, Bhagwat AA, Cregan PB. 1992. Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132: 1131-1139

Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML. 2000. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Sci. 40: 815-823

Burnham KD, Francis DM, Dorrance AE, Fioritto RJ, Martin SK. 2002. Genetic diversity patterns among Phytophthora resistant soybean plant introductions based on SSR markers. Crop Sci. 42: 338-343

Camus-Kulandaivelu L, Veyrieras JB, Gouesnard B, Charcosset A, Manicacci D. 2007. Evaluating the reliability of STRUCTURE outputs in case of relatedness between individuals. Crop Sci. 47: 887-890

Choung MG, Baek IY, Kang ST, Han WY, Shin DC, Moon HP, Kang KH. 2001. Determination of protein and oil contents in soybean seed by near infrared reflectance spectroscopy. Korean J. Crop Sci. 46: 106-111

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE. 1999. An integrated genetic linkage map of the soybean. Crop Sci. 39: 1464-1490

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611-2620

Erskine W, Muehlbauer FJ. 1991. Allozyme and morphological variability, out crossing rate and core collection formation in lentil germplasm. Theor. Appl. Genet. 83: 119-125

Excoffier, Laval LG, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol. Bioinform. Online 1: 47-50

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data. Linked loci and correlated allele frequencies. Genetics 164: 1567-1587

Fu YB, Peterson GW, Morrison MJ. 2007. Genetic diversity of Canadian soybean cultivars and exotic germplasm revealed by simple sequence repeat markers. Crop Sci. 47: 1947-1954

Fukunaga F, Hill J, Vigouroux Y, Matsuoka Y, Jesus Sanchez G, Liu K, Buckler ES, Doebley J. 2005. Genetic diversity and population structure of teosinte. Genetics 169: 2241-2254

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169: 1631-1638

Hong EH, Kim SD, Hwang YH, Lee HS. 1988. Breeding strategy for 21st Century: Leguminous grain crops. Korean J. Breed. 20(S): 13-18

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. PNAS 103: 16666-16671

Jun TH, Van K, Kim NY, Lee SH, Walker DR. 2007. Association analysis using SSR markers to find QTL for seed protein content in soybean. Euphytica DOI 10.1007/s10681-007-9491-6

Kim SH, Jung JW, Moon JK, Woo SH, Cho YG, Jong SK, Kim HS. 2006. Genetic diversity and relationship by SSR markers of Korean soybean cultivars. Korean J. Crop Sci. 51: 248-258

Kisha TJ, Diers BW, Hoyt JM, Sneller CH. 1998. Genetic diversity among soybean plant introductions and North American germplasm. Crop Sci. 38: 1669-1680

Kwon SH, Im KH, Kim JR. 1972a, Diversity of protein and oil contents of the Korean soybean land races (I). Korean J. Breed. 4: 29-32

Kwon SH, Kim JR, Song HS, Im KH. 1972b. Characteristics of important agronomic traits of Korean local soybean collections. Korean J. Breed. 6: 67-70

Liu K, Muse SV. 2005. PowerMarker: Integrated analysis environment for genetic marker data. Bioinform. 21: 2128-2129

Narvel JM, Fehr WR, Chu WC, Grant D, Shoemaker RC. 2000. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Sci. 40: 1452-1458

Nei M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70: 3321-23

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959

Reif JC, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M, Hoisington D, Bohn M, Melchinger AE. 2005. Wheat genetic diversity trends during domestication and breeding. Theor. Appl. Genet. 110: 859-864

Rural Development Administration. 1986. Descriptors of genetic resources.

SAS Institute Inc. 2004. SAS Enterprize Guide 3.0.0.369. Cary, NC, USA

Song HS, Lee YI, Kwon SH. 1991. Study on the agronomic traits of Korean native soybean (*Glycine max*). Korea Soybean Digest 8: 1-16

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

Wang L, Guan R, Liu Z, Chang R, Qiu L. 2006. Genetic diversity of Chinese cultivated soybean revealed by SSR markers. Crop Sci. 46: 1032-1038

Wright S. 1978. Evolution and the genetics of populations Vol 4. Variability within and among natural populations, University of Chicago Press, Chicago, pp 580.

Yoon MS. 1998. Variation of morphological and biochemical characteristics of soybean [*Glycine max* (L.) Merrill] germplasm oriented from Korea, China and Japan. Ph.D. thesis. Kyunghee University, Seoul

Yoon MS, Baek HJ, Lee J, Kim HH, Cho YH, Ahn JW, Kim CY. 2003. The major morphological characteristics and variations of soybean landraces. Kor. J. Intl. Agric. 15: 294-303