African Journal of Biotechnology Vol. 9(31), pp. 4857-4866, 2 August, 2010
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2010 Academic Journals

Full Length Research Paper

Assessing the genetic diversity of cultivars and wild soybeans using SSR markers

Wang Min, Li Run-zhi, Yang Wan-ming and Du Wei-jun*

College of Agronomy, Shanxi Agricultural University, Taigu Shanxi 030801, P.R. China.

Accepted 2 July, 2010

Increasing the diversity of the soybean germplasm base could introduce new genes affecting agronomic traits. In this study, we demonstrated the differences of genetic diversity level among 40 soybean accessions of cultivars, landraces and wild soybeans collected in the Shanxi Agricultural University using 40 simple sequence repeat (SSR) primer pairs. The structure based on model result showed that the cultivars, landraces and wild soybeans could be divided into three groups. Comparison of three types of soybeans showed that wild soybeans and landraces showed higher genetic diversity level than cultivars. The average genetic diversity index of wild soybeans and landraces was 1.5421 and 1.2864, while that of cultivars was 1.0981. A total number of alleles in wild soybeans were 224, while those in cultivars and landraces were 182 and 148, respectively, which were 81.25 and 66.07% of wild soybeans. The higher genetic distance (0.6414) and genetic differentiation (0.1200) and the lower genetic identity (0.5265) and gene flow (1.8338) between wild soybeans and cultivars were found. The proportion of low frequency alleles (allele frequency < 0.15) was the highest in wild soybeans (57.5%), followed by landraces (42%) and cultivars (29.8%). The UPGMA results also showed that wide soybean were of more abundant genetic diversity than cultivars. These results indicated that wild soybeans and landraces possessed greater allelic diversity than cultivars and might contain alleles not present in the cultivars which can strengthen further conservation and utilization.

Key words: Soybean, simple sequence repeat, genetic diversity.

INTRODUCTION

Soybean is one of the most important oil and protein crops in the world. Soybean originated in and is extensively cultured in China, which has abundant sources of soybean germplasm. However, the loss of genetic diversity, in part due to conventional breeding programs associated with modern agronomic and agricultural practices, has been dramatic for many cultivated species (Wilkes, 1983; Plucknett et al., 1983). For example, Gai and Zhao, (2001) reported that out of the 308 ancestral varieties, 38 provided approximately 54.18 and 56.84% of nuclear and cytoplasmic genetic material, respectively, of the 651 soybean cultivars released from 1923 to 1995 in China. The use of only a few introduced plants and intensive plant breeding has narrowed the genetic diversity among North America elite soybean cultivars (Gizlice et al., 1994). 258 cultivars bred in North America from 1947-1988 were descended from 35 ancestral parents and were identified by Gizlice et al.(1994) who proposed that soybean in North America had a very limited genetic basis. Today, the narrowing of the crop germplasm base is a threat to sustained breeding improvement and also increases the vulnerability of crops to pathogens and pests.

Several researchers have investigated the potential of exploiting the wild relatives of cultivated crops as a source of genetic material. The annual wild soybean (*Glycine soja*) is the direct ancestor of cultivated soybean and therefore serves as a valuable gene pool for soybean improvement (Lu, 2004). It has many advantages, including high reproduction rate and strong resistance to different adverse surroundings. Landraces are important sources of genetic variation when specialized attributes are needed to tolerate different environmental stresses (Srivastava and Damania, 1989). Therefore, introducing genetic diversity from the exotic gene pool of wild soybeans and

*Corresponding author. E-mail: wangmin3502@126.com. Tel: +86-0354-6288374.

Table 1. Forty soybean materials and their origins in this study.

S/N	Entry	Origin (county)	S/N	Entry	Origin (county)
1	Jin dou1(cultivar)	Shanxi	21	huguan daheidou (landrace)	Shanxi (changzhi)
2	Jin dou14 (cultivar)	Shanxi	22	Linxian xiaoheidou (landrace)	Shanxi (Linxian)
3	Jin dou19(cultivar)	Shanxi	23	Yushe banye (landrace)	Shanxi (Yushe)
4	Jin dou514(cultivar)	Shanxi	24	SNWS0208 (wild)	Shanxi (Xiyang)
5	Jin da47(cultivar)	Shanxi	25	SNWS0179 (wild)	Shanxi (Pingyao)
6	Jin da53(cultivar)	Shanxi	26	SWNS0224 (wild)	Shanxi (Wenshui)
7	Jin da70(cultivar)	Shanxi	27	SNWS0263 (wild)	Shanxi (Taigu)
8	Jin da73(cultivar)	Shanxi	28	SNWS0027 (wild)	Shanxi (Hequ)
9	Jin han125(cultivar)	Shanxi	29	SNWS0038 (wild)	Shanxi (Linxian)
10	Xi ye501(cultivar)	Shanxi	30	SNWS0159 (wild)	Shanxi (Hejin)
11	Qi huang22(cultivar)	Shandong	31	SNWS0349(wild)	Shanxi (Hongtong)
12	Qian jindou(cultivar)	unknown	32	SNWS0045 (wild)	unknown
13	Hucun huang dou (cultivar)	unknown	33	SNWS0048 (wild)	Shanxi (Zuoyun)
14	SNSZ0092 (cultivar)	unknown	34	SNWS0054 (wild)	Shanxi (Gujiao)
15	Kefeng1(cultivar)	Beijing	35	SNWS0056 (wild)	Shanxi (Yangqu)
16	Xingxian xiaohuangdou (landrace)	Shanxi (Xingxian)	36	ZYD05779 (wild)	Shanxi (Taiyuan)
17	Jinbei xiaoheidou (landrace)	Shanxi (Shuozhou)	37	SNWS0293 (wild)	Shanxi (Jinci)
18	Mo shidou (landrace)	unknown	38	SNWS0273 (wild)	Shanxi (Qingxu)
19	Zuoquan daheidou (landrace)	Shanxi (Zhuoquan)	39	SNWS0278(wild)	Shanxi(Taiyuan)
20	Liulin baidou (landrace)	Shanxi (Linxian)	40	SNWS0295 (wild)	Shanxi (Loufan)

landraces to broaden the genetic background of soybean cultivars might facilitate an increase in soybean yield, in addition to disease and pest resistance.

The genetic diversity patterns of wild soybean have been evaluated on the basis of enzymes in natural populations in Japan (Kiang et al., 1992; Fujita et al., 1997), China (Pei et al., 1996) and South Korea (Yu and Kiang, 1993). The patterns have also been revealed by RFLP, RAPD, SSR and AFLP markers to probe the genetic differences between wild and cultivated soybeans or for the origin and dissemination of soybeans (Maughan et al., 1995; Rongwen et al., 1995; Kisha et al., 1998; Shimamoto et al., 1998, 2000; Thompson et al., 1998; Brown-Guedira et al., 2000; Tian et al., 2000; Li and Nelson, 2001; Xu and Zhao, 2002; Abe et al., 2003). These studies have revealed higher levels of genetic diversity in wild soybean. Though many data from different researchers have discussed the genetic differences among landraces, wild and cultivated soybeans from quantitative-traits, chromosomes, isozymes and nucleotide loci, there is still a long way to explain thoroughly and systematically the diversity level among different types of soybean germplasms.

To this end, we have used SSR (simple sequence repeat) analysis to study genetic variability in cultivars, landraces and wild soybeans. Microsatellites have been found to vary in the polymorphism they detect depending on the length and sequence of the repeat motif they contain and their location in coding or non-coding segments of the genome (Thoquet et al., 2002; Temnykh et al., 2000, 2001; Eujay et al., 2002). Our primary objective was to compare the genetic diversity among landraces, cultivated and wild soybeans. In addition, this work will provide evidence for soybean genetic differentiation and valuable information for further conservation and breeding programs in soybeans.

MATERIALS AND METHODS

Plant materials

The 40 Chinese soybean accessions, including cultivars, landraces and wild soybeans were mostly from Shanxi province except for several other provinces abroad and of unknown origin; as control and these were chosen because they are representative of the main ecological types in China. The cultivars from Shanxi studied here represented modern Shanxi soybean breeding and were released during 1970 to 2003. The landraces and wild soybeans were collected from a total of 20 counties in Shanxi, encompassing all 3 of the major growing regions of Shanxi: northern Shanxi, central Shanxi and southern Shanxi (Table 1).

DNA extraction

Total genomic DNA was isolated from fresh young leaves using modified CTAB extraction method (Eujayl et al., 2002).

PCR analysis

PCR was carried out in a 20 µl volume that contained template DNA (50 - 100 ng), 1×PCR buffer (including $MgCl_2$), 0.2 mmol dNTPs, 0.2 µmol SSR primers and 1 U Taq DNA polymerase. The 40 SSR primers represented a variety of repeat types from dinucleotide to

Table 2. Genetic diversity revealed by 40 SSR primer pairs.

Primer	Diversity index	PIC	The number of alleles	Primer	Diversity index	PIC	The number of alleles
Satt197	1.9887	0.8500	8	Satt530	1.6566	0.7694	6
Satt177	1.4986	0.7525	5	Satt307	1.7339	0.7656	7
Satt590	1.8550	0.8313	7	Satt267	1.5854	0.7638	6
Satt431	1.8511	0.8325	7	Satt268	1.7616	0.7963	7
Satt300	1.8019	0.8138	7	Satt352	1.6237	0.7750	6
Satt571	1.5529	0.7400	6	Satt184	1.8307	0.7950	8
Satt565	1.5538	0.7650	6	Satt168	1.5438	0.7313	6
Satt596	1.7390	0.7913	7	Satt453	1.6586	0.7875	6
Satt281	1.9499	0.8425	8	Satt022	1.4301	0.7338	5
Satt226	1.4798	0.7538	5	Satt346	1.7744	0.7894	7
Satt185	1.9635	0.8500	8	Satt236	1.3196	0.7150	4
Satt586	1.8974	0.8438	7	Satt429	1.6781	0.7963	6
Satt487	1.1582	0.5850	5	Satt279	1.6069	0.7763	6
Satt386	1.4601	0.7156	6	Satt005	1.9010	0.8450	7
Satt434	1.5456	0.7588	6	Satt309	1.7945	0.7544	7
Satt239	1.6409	0.7625	7	Satt607	1.4222	0.6838	6
Satt216	1.8108	0.8213	7	Satt534	1.8945	0.8225	8
Satt588	1.7864	0.8100	7	Satt514	1.4226	0.7275	5
Satt345	1.8016	0.7963	8	Satt574	2.0454	0.8269	8
Satt373	1.9861	0.8225	8	Satt528	1.6635	0.7938	6
				Mean	1.6917	0.7800	6.55

hexanucleotide motifs and were distributed throughout the genome (Song et al., 2004). Most of them were core loci (Xie et al., 2003) and a few were selected that have generated stable and clear DNA fragments in our laboratory for many years. These SSR loci were evenly distributed on the 20 genetic linkage groups (LGs) of soybean, with an average of two loci per LG. The primer sequences of all SSR loci were obtained from SoyBase, the USDA-ARS sponsored genome database (http://129.186.26.94/SSR.html) and were synthesized by AuGCT Biotechnology.

The reaction was denatured at 95℃ for 5 min, followed by 35 cycles, each consisting of denaturing at 94℃ for 1 min, annealing at 52℃ for 1 min, extension at 72℃ for 1 min and a final extension was performed at 95℃ for 5 min. The amplification products were separated on 8% denaturing polyacrylamide gels and detected by silver staining.

Data analysis

Microsatellite allele sizes for the 40 loci were scored for all genotypes on the basis of comparison to an allele matrix which were prepared from this dataset.

Cluster analysis was performed with NTSYS-pc Ver 2.0 (Rohlf, 1997) software package based on UPGMA (Unweighted Pair Group Method with Arithmetic mean). Polymorphism information content (PIC) analysis was used to evaluate markers so that the most appropriate can be selected for genetic mapping, phylogenetic analysis, or association genetics (Anderson et al., 1992). The extent of diversity was evaluated by polymorphic information content, which was calculated, respectively as: PIC = $1-\sum P_i^2$, in which P_i is the frequency of an allele (*i*). Genetic variation within and among the groups and subgroups detected was analyzed with POPGENE software (Yeh et al., 1997) using parameters such as the number of alleles, the index of Shannon and Weaver (1949; H), Nei's (1978) coefficient of gene differentiation (GST), gene flow (Nm), genetic distance (GD) and genetic identity (I). The relationship between populations (K) was evaluated with the software structure (Pritchard et al., 2000) based on populations of K = 2 to K = 3. Genetic structure was evaluated using the analysis of molecular variance model (AMOVA) (Excoffier et al., 1992) in the ARLEQUIN Version 3.0 software package (Excoffier et al., 2005). Other data was calculated using EXCEL 2007.

RESULTS

Genetic diversity of 40 soybean germplasms by SSR

Analysis of 40 materials by 40 SSR primer pairs identified a total of 262 alleles, with 6.55 alleles per locus (Table 2). Higher numbers of alleles were scored from loci Satt197, Satt281, Satt185, Satt373, Satt184 and Satt534; all had 8 alleles. Lower numbers of alleles were found in loci Satt177, Satt226, Satt487, Satt022 and Satt514; all had 5 alleles. Satt236 had the lowest numbers of alleles (4).

Polymorphic information contents (PIC) varied from 0.8500 to 0.5850, which corresponded to the primer pairs Satt197 and Satt487, with an average of 0.7800. Shannon-Weaver's information indices ranged from 1.1528 to 2.0454 and averaged at 1.6917, which corresponded to the primer pairs Satt574 and Satt487. The results also

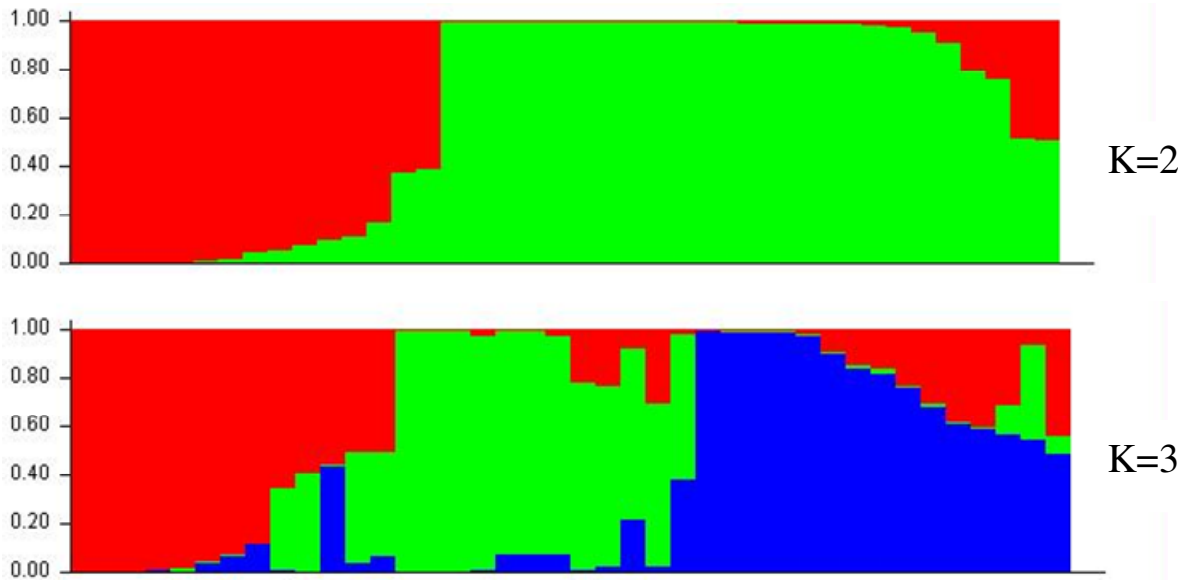

Figure 1. Genetic structure of 40 soybean germplasm.

Table 3. Genetic variation statistics for 40 SSR loci in soybean.

Statistics	Cultivars	Landraces	Wild soybeans
Sum of alleles	182	148	224
Mean of alleles	4.55	3.70	5.60
PIC	0.6684	0.6042	0.7497
Genetic diversity index	1.2864	1.0981	1.5421

showed that the 40 materials had abundant genetic diversity.

Genetic structure analysis of 40 soybean germplasms

Genetic structure analysis of the 40 soybean germplasms were performed with STRUCTURE based on model. When K = 2, *G. soja* (wild soybeans) and *G*lycine *max* (cultivars and landrace) were divided into 2 genetic populations (Figure 1).

The result indicated that *G. max* had obvious genetic differences from *G. soja* in the evolution progress, although *G. soja* is the direct ancestor of *G. max*. When K = 3, cultivars, landraces and wild soybeans were divided into 3 genetic populations. The result showed that SSR is one of the best methods for detecting genetic differences among cultivars, landraces and wild soybeans.

Comparison of genetic diversity among cultivars, landraces and wild soybeans

Analysis of cultivars, landraces and wild soybeans by 40 SSR primer pairs identified a total of 224 alleles in wild soybeans, with 5.60 alleles per locus. The average genetic diversity index was 1.5421, 1.2864 and 1.0981 of wild soybeans, cultivars and landraces, respectively (Table 3). T test result showed that there was a highly significant difference between cultivars and wild soybeans ($t = 4.44$, $P < 0.01$), landraces and wild soybeans ($t = 6.04$, $P < 0.01$) and cultivars and landraces ($t = 2.58$, $P < 0.05$). The results indicated that the genetic diversity of wild soybeans was significantly higher than cultivars and landraces. The total number of alleles was 182 and 148 in cultivars and landraces, respectively, which was 81.25 and 66.07% of wild soybeans. The results indicated that many alleles were losing in the evolution of wild soybean (*G. soja*) to *G. max*.

Genetic differentiation, gene flow, genetic distance and genetic identity among three types of soybean are shown in Tables 4 and 5. The genetic differentiation ranged from 0.0818 between cultivars and landraces to 0.1200 between wild soybeans and cultivars. The gene flow was calculated according to genetic differentiation. It was 1.8338 between wild soybeans and cultivars, 2.2191 between wild soybeans and landraces and 2.8055 between cultivars and landraces. Similar gene flow (higher than 2.6) was found between the cultivars and landraces. In contrast, different gene flow (lower than 2.6) was found between wild soybeans and other types. It indicated that

Table 4. Genetic differentiation (below diagonal) and gene flow (above diagonal) in soybean.

POP ID	Cultivars	Landraces	Wild soybeans
Cultivars	-	2.8055	1.8338
Landrace	0.0818	-	2.2191
Wild soybeans	0.1200	0.1012	-

Table 5. Genetic identity (above diagonal) and genetic distance (below diagonal) in soybean.

POP ID	Cultivars	Landrace	Wild soybeans
cultivars	-	0.5715	0.5265
landrace	0.5594	-	0.5472
wild soybeans	0.6414	0.6030	-

Table 6. Analysis of molecular variance (AMOVA) in soybean.

Sample	Source of variance	*Df*	Variance component	Percentage of variation (%)	P
Cultivars vs landrace	Between populations	1	1.88006	11.25	< 0.001
	Within populations	44	14.83251	88.75	
Cultivars vs wild soybean	Between populations	1	2.44448	15.17	< 0.001
	Within populations	62	13.66439	84.83	
Landrace vs wild soybeans	Between populations	1	2.43040	14.26	< 0.001
	Within populations	48	14.61213	85.74	

wild soybeans has fewer genetic information flowed to cultivars. Correspondingly, the genetic distance was 0.5594 between cultivars and landraces, 0.6030 between landraces and wild soybeans and 0.6414 between cultivars and wild soybeans. Genetic identity was 0.5715 between cultivars and landraces, 0.5472 between landraces and wild soybeans and 0.5267 between cultivars and wild soybeans. It indicated that the genetic backgrounds varied and there was a more distant genetic relationship between the wild soybeans and two other types.

AMOVA was conducted to describe variance components of wild soybeans, cultivars and landraces. The results showed that molecular variance was 15.17% between wild soybeans and cultivars, 14.26% between wild soybeans and landraces and 11.25% between cultivars and landraces, which were highly significant ($P < 0.001$) (Table 6). It indicated that there were obviously genetic differentiation in cultivars, landraces and wild soybeans.

The trend lines of allele frequency among cultivars, landraces and wild soybeans are in close agreement (Figure 2) and the allele number decreased as the allele frequency increased. The proportion of low frequency alleles (allele frequency < 0.15) was the highest in wild soybeans (57.5%), followed by landraces (42%) and cultivars (29.8%). In contrast, the proportion of high frequency alleles (> 0.65) was highest in cultivars, followed by landraces and was lowest in wild soybean. These results indicated that wild soybeans contained alleles that were missing from cultivars.

The specific alleles are those that only exist in one type of germplasm. Of the SSR loci tested, the cultivars had 14 specific alleles, the landraces had six and the wild soybeans had 50 specific alleles, representing 19.08% of the 262 alleles. These results indicated that wild soybeans contain many rare genes (Table 7).

In the progress of evolution from the annual wild soybean (*G. soja*) to *G. max*, the lost number of alleles in *G. max* was 46, which was 18% of 40 SSR loci tested in germplasms and new in *G. max* were 38 alleles, which were 14% of all tested (Figure 3). The number of alleles in *G. soja* was 86% of the total alleles in germplasms. It indicated that most of the genetic information in *G. max* was from *G. soja*.

Cluster results of 40 soybean germplasms

A UPGMA dendrogram was constructed for the 40 soybean materials (Figure 4). The cluster analysis showed significant genetic variation among the included cultivars, landraces and wild soybeans, as assessed by similarity coefficients. The dendrogram revealed 7 distinct groups

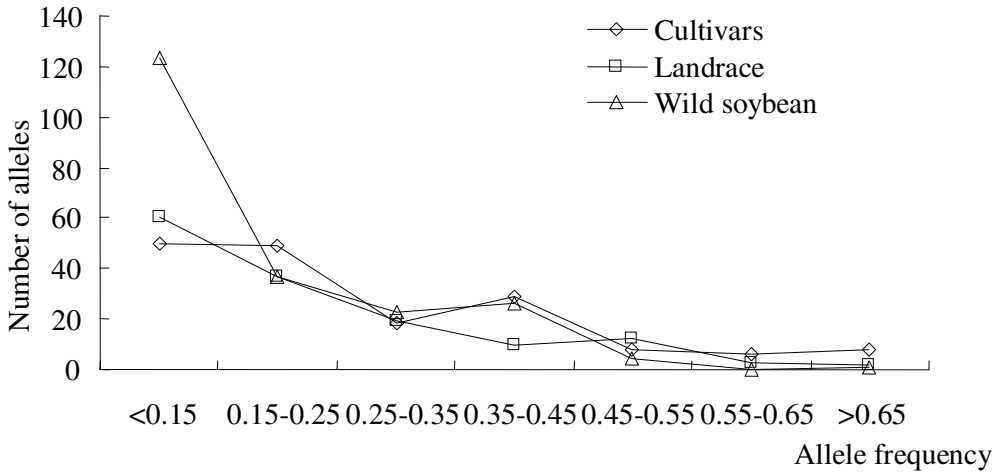

Figure 2. Allele frequency plot among improved varieties landrace and wild soybeans by SSR.

Table 7. Specific allele in soybean.

	Cultivars	Landraces	Wild soybeans
Specific allele	14	6	50

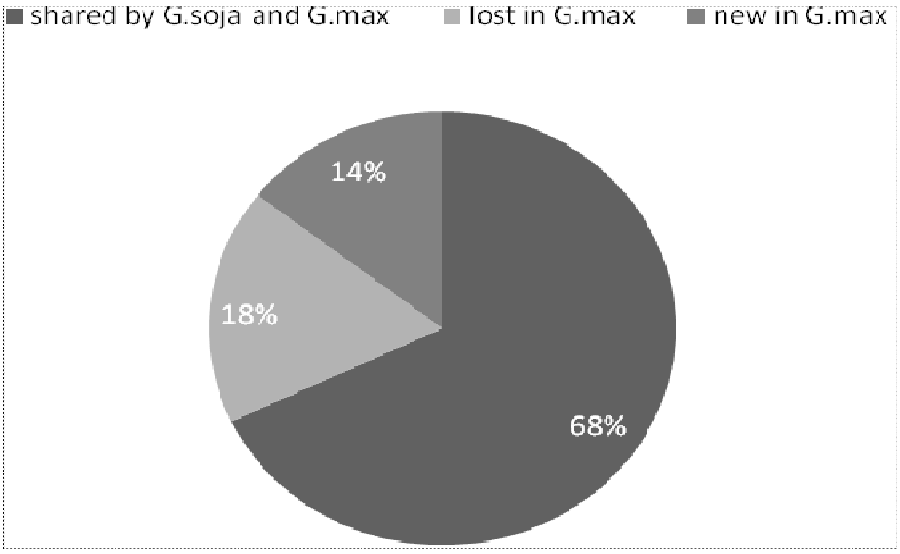

Figure 3. Composition of allele among soybean germplasms.

at the similarity coefficient level of 0.755 (Figure 1). 17 materials, including 14 cultivars and 3 wild soybeans were included in Group 1, which was further divided into 2 subgroups at the similarity coefficient level of 0.763. The first subgroup comprised 12 materials, which were improved varieties. The second subgroup comprised 5 materials, including 2 cultivars and 3 wild soybeans. The dendrogram indicated that modern varieties have a narrow genetic basis and low genetic variation. 9 accessions, including all the landraces and a cultivar named Kefeng 1 were included in Group 2. Wild soybeans were scattered among the other 5 groups in the dendrogram. SNWS0263, SNWS0027, SNWS0038, SNWS0349 and SNWS0045 were included in Group 3, which are from different regions in Shanxi province. SNWS0159 was included in Group 4, which is from Hejin. SNWS0056, SNWS0273, SNWS0278 and SNWS0295 were included in Group 5, which are from the Taiyuan region of Shanxi

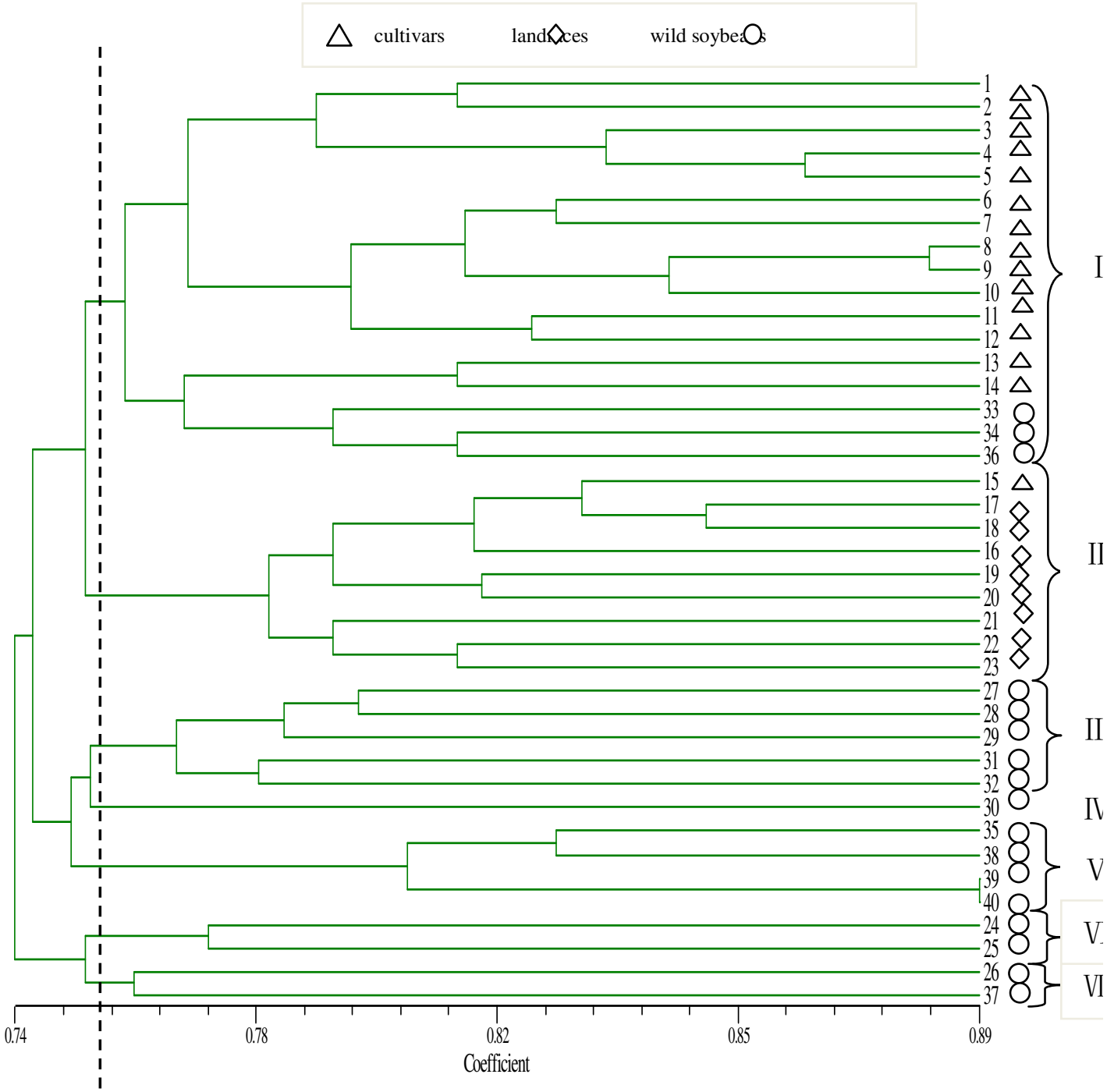

Figure 4. Dendrogram of 40 soybean genetic materials analyzed by SSR.

province. SNWS0208 and SNWS0179 were included in Group 6, which are from the Jinzhong region of Shanxi province. SNWS0224 and SNWS0293 were included in Group 7, which are from Wenshui and Jinci, respectively. It is evident that there are far genetic distances between wild soybeans and the others; thus wild soybeans should be further studied and effectively utilized by breeders.

DISCUSSION

Determination of the genetic diversity of soybean germplasms by SSR

Earlier studies evaluated the relationship among cultivars, landraces and wild soybeans. Lin (2003) identified that

varieties can be associated with landraces and Maughan et al. (1995) identified that cultivars can be clearly associated with wild soybeans, using SSR markers. Lin (2005) had showed that 18 soybeans in Fujian could be divided into 3 groups; vegetable soybean, cross-breeding and local varieties and varieties with semi-wild kindred relationships. Zhao et al., (2001) showed that 44 soybeans could be divided into 2 groups at a similarity coefficient level of 0.130 by SSR. In this study, we showed using 40 SSR markers, that there is large genetic diversity among the 40 soybean materials. Cultivars, landraces and wild soybeans were effectively divided into different groups. Xu et al., (1999) also showed that SSR markers were effective tools for evaluating genetic diversity.

Of the 40 SSR loci tested, 7 (Satt168, Satt307, Satt184, Satt005, Satt309, Satt588 and Satt022) were previously examined by Narvel et al. (2000). Narvel et al. (2000) found that there was a total of 110 alleles and an average marker diversity of 0.659 in 79 Northern American elite cultivars and introductions, whereas we detected a total of 262 alleles and an average marker diversity of 0.780 in 40 materials, which are higher than those that were identified by Narvel et al. (2000). The results show that there was abundant genetic variation among the 40 materials and a large degree of genetic differentiation among cultivars, landraces and wild soybeans at the molecular level.

Comparison of genetic diversity among landraces, cultivated and wild soybeans

Maughan et al. (1995) studied genetic diversity of 94 soybeans with 5 SSRs. The results indicated that the polymorphism information content of wild soybeans and cultivars were 0.55 and 0.87, respectively. Rongwen et al. (1995) analyzed 96 soybeans with SSR and showed that the mean polymorphism information content in soybeans was 0.87 and in the cultivars it was 0.74. Wu et al., (2001) studied 37 soybeans with SSR and showed 71.5% of them were present in *G. soja*. Zhao et al., (2001) reported that the mean genetic distance among wild soybeans and among cultivars were 0.176 and 0.150, respectively. The results suggested a large genetic diversity for wild soybeans and a lower diversity for cultivars. Xu and Zhao, (2002) analyzed 326 soybeans using 6 cpSSR markers and demonstrated genetic variation indices for wild soybeans and cultivars of 0.964 and 0.421, respectively, again indicating higher genetic diversity in wild soybeans. In this study, the genetic diversity of wild soybeans was significantly higher than cultivars and landraces. Meanwhile, wild soybeans had higher genetic distance and lower genetic identity with cultivars. In addition, the cluster result showed that the wild soybeans were scattered among 5 different groups. This indicated that the genetic backgrounds varied and there was a more distant genetic relationship between the wild soybeans and the other 2 types. The results of this study agree with those of the previous studies.

Implications for collection and conservation of wild soybeans and landraces

As a direct and indirect consequence of artificial selection for traits that improved agronomic qualities, favorable alleles at loci controlling agronomic traits were brought to fixation in the population during the domestication period while wild species and domestic breeds and strains that owned some distinct favorable alleles responding to environmental changes are disappearing at an alarming rate. The continued practice of selective and genetic bottleneck effect speeded the process. Some anticipated consequences are a genome wide loss of diversity at unselected genes. Therefore, the loss of genetic diversity is becoming an increasingly central topic in conservation genetics. Kisha et al. (1998) speculated that many of the genes or alleles that exist in ancestral lines have been lost in the extensive breeding that has taken place in the U.S. It has been estimated that the diversity found in cultivated populations has declined by as much as 50% during the domestication bottleneck (Hyten et al., 2006). It is reported that the most significant loss of diversity occurred during domestication, the introduction of bottleneck where there was a large loss of rare alleles present in *G. soja* and the Asian Landraces. Our study also revealed the domestication bottleneck that was responsible for 18% reduction in diversity and the elimination of 27.7% of rare alleles present in *G. soja*. Some study reported that the low nucleotide diversity in modern elite soybean cultivars is mainly due to an unusually low level of genetic variability in the wild progenitor. However, more and more data showed agricultural expansion as the main threat to wild plant species, followed by over grazing, environmental pollution and modern breeding practice. These processes have resulted in a drastic reduction of genome diversity. These alleles are likely to benefit future soybean improvement. Expansion of the currently low number of *G. soja* accessions available to soybean geneticists and breeders should be considered a high priority. In our study, the selected landraces and wild soybean came from the new collections in recent years and represented 20 counties in Shanxi, encompassing all three of the major growing regions of Shanxi. Our study found some materials owned by some rare and unique alleles not found in the available *G. max* germplasm collections and the elite cultivars. These materials will be of great value to future genetic gains in cultivar productivity.

Investigation and utilization of Shanxi soybean germplasms

It is interesting to note that cultivar Kefeng 1 and land-

races from Shanxi, which are similar in phenotype, are gathered in one group. The findings here support the concept that cultivars trace their pedigree to landraces and Kefeng 1 might have come from landrace in Shanxi Province. SNWS0048, SNWS0054 and ZYD05779 were similar in appearance to wild soybeans, but were genetically closer to cultivars, indicating the accessions might have come from the natural hybridization of cultivars and wild soybeans. Hucun huangdou and SNSZ0092 are of unknown origin and were gathered in one subset group, indicating that these materials have a distant genetic relationship with cultivars from Shanxi. These materials should be further studied for their potential in broadening the genetic background of the cultivars from Shanxi. A few materials had kindred and greater genetic similarity coefficients but were scattered in different groups, such as Xiye and Xiye 501. Xiye 501 was bred by a cross of Xiye and Jindou 501 and was further developed by pedigree selection according to human demands, thus it was gradually lost by parental coenogenesis. Further research should be performed to discover the lost superordinary genes.

Some wild soybeans from different regions in Shanxi province were also included in the same groups [SNW S0263, SNWS0027, SNWS0038, SNWS0349 and SNWS0045 (Group 3)] which are distributed from north to south in Shanxi Province. SNWS0224 and SNWS0293 (Group 4) are from Wenshui and Jinci, respectively. These soybeans may have been as a result of genetic shift. Some wild soybeans from same regions in Shanxi province were included in the same group (SNWS0056, SNWS0273, SNWS0278 and SNWS0295 (Group 5) which are from Taiyuan. SNWS0208 and SNWS0179 (Group 6) are from Jinzhong. However, some wild soybeans from same regions in Shanxi province are included in different groups. For example, SNWS0263 from Jinzhong was not included in Group 6 and SNWS0349 from Taiyuan was not included in Group 5. The results indicated that most of the wild soybeans from same region had close relationships but a few had far relationships. These resources require further research for their utility in breeding programs.

More than 60 years of breeding have improved Chinese soybean production substantially, raising yield levels and providing resistance to potentially devastating diseases. One unintended consequence, however, is the dramatic loss of genetic diversity for many cultivated species and in part to the conventional breeding programs associated with modern agronomic and agricultural practices. Pedigree analysis has shown that many Chinese cultivars are as closely related as half sibs. One straight forward method to avert further erosion of diversity in China is to add new breeding stocks to applied programs. The successful introgression of landraces and wild soybeans into Chinese cultivars will show whether landraces and wild soybean could be used to increase soybean diversity in China.

ACKNOWLEDGMENTS

The project was financed by the National Natural Science Foundation of China (30471094), the Key Program of Science and Technology of Shanxi Province (No.051017) and the Scientific Research Foundation for Youth Academic Leaders from University in Shanxi Province (No. 200425).

REFERENCES

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

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1992). Optimizing parental selection for genetic linkage maps. Genome, 36: 181-186.

Brown Guedira GLJ, Thompson A, 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.

Hyten DL, Song QJ, Zhu YL, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006). Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. USA, 103: 16666-16671.

Eujay I, Sorrells ME, Baum M, Wolters P, Powell W (2002). Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor. Appl. Genet. 104: 399-407.

Excoffier L, Laval G, Schneider S (2005). Arlequin Ver. 3.0: an integrated software package for population genetics data analysis. Evol. Bioinform. Online, 1: 47-50.

Excoffier L, Smouse P, Quattro J (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, 131: 479-491.

Fujita R, Ohara M, Okazaki K, Shimamoto Y (1997). The extent of natural cross-pollination in wild soybean (*Glycine soja*). J. Hered., 88: 124-128.

Gai JY, Zhao TJ (2001). The core ancestors of soybean cultivars in China. J. Nanjing Agric. Univ. 24(2): 20-23.

Gizlice Z, Carter Jr. TE, Burton JW (1994). Genetic base for north American public soybean cultivars released between 1947-1988. Crop Sci. 34: 1143-1551.

Kiang YT, Chiang YC, Kaizuma N (1992). Genetic diversity in natural populations of wild soybean in Iwate Prefecture, Japan. J. Hered. 83: 325-329.

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

Li ZL, Nelson RL (2001). Genetic diversity among soybean accessions from three countries measured by RAPD. Crop Sci. 41: 1337-1347.

Lin FY, Qiu LJ, Chang RZ, He BR (2003). Genetic diversity of landrace and bred varieties of soybean in Shanxi**.** Chin. J. Oil Crop Sci. 25(3): 24-28.

Lin GQ, Zhu HS, Huang K (2005). Genetic diversity analysis of soybean cultivars in Fujian Province by RAPD. J. Plant Genet. Resour. 6(2): 130-134.

Lu BR (2004). Conserving biodiversity of soybean gene pool in the biotechnology era. Plant Species Biol. 19: 115-125.

Maughan PJ, Saghai Maroof MA, Buss GR (1995). Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome, 38: 715-723.

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

Pei YL, Wang L, Ge S, Wang LZ (1996). Study on genetic diversity of *Glycine soja*-isozyme variation in four populations (Chin.). Soybean Sci. 15: 302-309.

Plucknett DL, Smith NJH, Williams JT, Anishetti N (1983). Crop germplasm conservation and developing countries. Science, 220: 163-169.

Pritchard JK, Stehens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics, 155: 945-959.
Rohlf FJ (1997). NTSYS-pc version 2.0. Exeter Press, East Setauket, New York, USA.
Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB (1995). The use of micro satellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90(1): 43-48.
Shimamoto Y, Abe J, Gao Z, Gai JY, Thseng FS (2000). Characterizing the cytoplasmic diversity and phyletic relationship of Chinese landraces of soybean, *Glycine max*, based on RFLPs of chloroplast and mitochondrial DNA. Genet. Resour. Crop Evol. 47: 611-617.
Shimamoto Y, Fukushi H., Abe J, Kanazawa A, Gai J Y, Gao Z and Xu D H. (1998). RFLPs of chloroplasm and mitochondrial DNA in wild soybean, *Glycine soja*, growing in China. Genet. Resour. Crop Evol. 45: 433-439.
Song QJ, Marek LF, Shoemaker RC (2004). A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109: 122-128.
Srivastava JP, Damania AB (1989). Use of collections in cereal improvement in semi-arid areas. In: Brown AHD, Marshall DR, Frankel OH, Williams JT (eds) The use of plant genetic resources. Cambridge University Press, Cambridge, pp. 88-104.
Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch SR (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome, 11: 1441-1452.
Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000). Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100: 697-712.
Thompson JA, Nelson RL, Vodkin LO (1998). Identification of diverse soybean germplasm using RAPD markers. Crop Sci. 38: 1348-1355.
Thoquet P, Ghérardi M, Journet EP, Kereszt A, Ané JM, Prosperi JM, Huguet T (2002). The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes BMC. Plant Biol. 2: 1.
Tian QZ, Gai JY, Yu DY, Jia JZ (2000). A study of amplified fragment length polymorphism (AFLP) in soybean (Chin.) Soybean Sci. 19: 210-217.
Wilkes G (1983). Current status of crop plant germplasm. CRC Crit. Rev. Plant Sci. 1: 133-181.
Wu XL, He CY, Chen SY, Zhuang BC, Wang KJ, Wang XC (2001). SSR markers with a soybean genetic evolutionary relationships between species. J. Genet. Genomics, 28(4): 359-366.
Xie H, Chang RZ, Chao YS, Zhang MH, Feng ZF, Qiu LJ (2003). Selection of Core SSR Loci by Using Chinese Autumn Soybean (*Glycine max* L. Merr). Scientia Agricultura Sinica, 36(4): 360-366.
Xu L, Zhao GF (2002). Microsatellite DNA marker and its application in genetic diversity research. Acta Botanica Boreali Occidentalia Sinica, 22(3): 714-722.
Xu ZY, Qiu LJ, Chang RZ, Li XH, Zheng CM, Liu LH, Guo B (1999). Using SSR Markers to Evaluate Genetic Diversity of Soybean in China. Scientia Agricultura Sinica, 32: 40-48.
Yeh FY, Boyle R, Ye T, Mao Z (1997). POPGENE, the user-friendly shareware for population genetic analysis, version 1.31. Molecular Biology and Biotechnology Centre, University of Alberta, Alberta.
Yu H, Kiang YT (1993). Genetic variation in South Korean natural populations of wild soybean (*Glycine soja*). Euphytica, 68: 213-221.
Zhao H K, Wang Y M, Li Q Y, Zhang M, and Zhuang BC (2001). SSR analysis of wild soybean (*G. soja*) and cultivated soybean from different latitude in China. Soybean Sci. 20(3): 172-176.