Botanical Studies (2006) 47: 13-21.

Restriction fragment length polymorphisms in the USDA soybean germplasm from central China

Myong Gi CHUNG[1,*], Mi Yoon CHUNG[1], April D. Clikeman JOHNSON[2], and Reid G. PALMER[3]

[1]*Department of Biology, Gyeongsang National University, Jinju 660-701, Republic of Korea*

[2]*Department of Agronomy, Iowa State University, Ames, IA 50011, USA*

[3]*USDA-ARS, Corn Insects and Crop Genetics Research Unit, Project 3769, Department of Agronomy, Iowa State University, Ames, IA 50011, USA*

(Received April 06, 2005; Accepted October 14, 2005)

ABSTRACT. To evaluate levels of genetic diversity in USDA soybean germplasm from central China, 107 accessions were examined at 46 RFLP loci. We compared genetic diversity in randomly selected accessions with pre-selected accessions based upon root tip fluorescence, pubescence morphology, and isozyme patterns at ten enzyme systems. We also evaluated levels of genetic diversity of the central Chinese accessions (n = 107) by comparing previously studied ancestors and milestone cultivars (NAC, n = 64) in the USA. Finally, we estimated the degree of genetic differentiation among six Chinese provinces (Anhui, Gansu, Henan, Jiangsu, Shaanxi, and Shanxi). There was significant difference between pre-selected and random accessions in terms of the mean number of alleles per locus (A, 2.44 vs. 2.13) and allelic richness (2.26 vs. 2.10). However, the former (H_e = 0.393) maintained levels of gene diversity or expected heterozygosity (H_e) similar to the latter (H_e = 0.394). This is attributed to the fact that many alleles found in pre-selected accessions were present at very low frequencies (mean effective number of alleles, A_e = 1.72). A broader range of alleles detected in the pre-selected accessions suggests that pre-selection of accessions screened from isozyme data may be useful for selecting germplasm collections with a greater number of RFLP alleles. The central Chinese accessions maintained a significantly higher level of RFLP genetic diversity than the NAC (H_e = 0.405, A = 2.50 for central China vs. H_e = 0.339, A = 2.08 for the USA). We detected significant genetic differentiation among the six provinces (mean G_{ST} = 0.133). These results suggest that Chinese germplasm accessions from various regions or provinces in the USDA germplasm collection could be used to enhance the genetic diversity of US. cultivars.

Keywords: Genetic variation; *Glycine max*; Pre-selection; RFLPs; USDA soybean germplasm collection.

INTRODUCTION

The low level of genetic diversity within US soybean cultivars has brought more attention to the use of plant introductions or accessions from other countries, in particular, the regions where the crop evolved. Plant introductions, although often agronomically undesirable, can be used as parents to enhance genetic diversity. Methods to quickly evaluate plant introductions maintained in the National Plant Germplasm System for diversity may increase the use of the collection by plant breeders as parents to improve modern cultivars. Efforts have been made towards identifying "core collections" (Brown, 1989). The goal of core collections is to maximize genetic variability in a smaller, but representative, group of accessions (Crossa et al., 1993; Bretting and Weidrlechner, 1995; Wang et al., 1998).

*Corresponding author: E-mail: mgchung@nongae.gsnu.ac.kr

Soybean [*Glycine max* (L.) Merr.] is a crop of major economic importance in China, Japan, Korea, and North and South America. China is considered to be the soybean center of origin and the center of diversity (Smartt and Hymowitz, 1985). Therefore, soybean accessions from China should include novel genetic diversity and may be a rich source of alleles from which to identify a core collection. One concern in evaluating new accessions of the USDA soybean germplasm collection is how to most efficiently detect genetic diversity. One approach is to determine if use of isozyme data can increase the efficiency of selection of plant introductions. If pre-selected accessions of Chinese soybean germplasm based upon root tip fluorescence (Torkelson and Palmer, 1997), pubescence tip morphology of legumes, and isozyme data (Liao and Palmer, 1997a, b, c, d) have a wider range of variation than randomly selected accessions, it would be predicted that measures at the DNA level maintain more alleles in pre-selected accessions than those from randomly selected accessions.

Allozyme markers have been used in soybean to evaluate genetic diversity in accessions from diverse geographic regions (Kiang et al., 1987; Perry et al., 1991; Griffin and Palmer, 1995; Yeeh et al., 1996; Chung et al., unpublished data). Since many selectively neutral DNA markers for soybean are currently available, several studies measuring genetic relationships at the DNA-level of soybean introductions (accessions from China) with the ancestors in the U.S. have been conducted to obtain information useful to breeding programs for selection of diverse parents (e. g., Keim et al., 1989, 1992; Lorenzen and Shoemaker, 1996; Brown-Guedira et al., 2000; Li et al., 2001). Restriction fragment length polymorphisms (RFLPs) provide a very large number of genetic markers for detecting and analyzing genetic diversity in plants (e. g., Helentjaris et al., 1985; Zhang et al., 1993; Dubreuil and Charcosset, 1998). RFLPs have been used for this purpose in soybean (e. g., Grabau et al., 1989; Lorenzen et al., 1995; Kisha et al., 1997). Since RFLPs can detect variation in both coding and non-coding regions, genetic variation was greater for RFLPs than for allozymes in several studies of crop plants (Messmer et al., 1991; McGrath and Quiros, 1992; Zhang et al., 1993; Dubreuil and Charcosset, 1998). For example, Dubreuil and Charcosset (1998) analyzed ten populations of maize (*Zea mays* L.) from European and north U.S. germplasm and detected that the mean number of alleles per locus (A = 6.3) and total genetic diversity (H_e = 0.60) for RFLPs were higher than those for allozymes (A= 2.4 and H_e = 0.23).

Most of the ancestors of U.S. soybean cultivars were introduced from China in the early part of the 20th century. In the USA, over 400 publicly released cultivars were developed from approximately 80 soybean ancestral lines (Gizlice et al., 1994). Twenty-eight introductions (ancestors) and seven first progenies (U.S. - developed cultivars with uncertain pedigrees) have contributed over 95% of the genes in public cultivars released between 1947 and 1988 (Gizlice et al., 1994). This strongly suggests that the number of ancestors that constitute the genetic base of soybean breeding programs in the USA is very limited. Considering the smaller number of the U.S. ancestors (64) relative to the 348 Chinese ancestral lines (Cui et al., 1999), we predict that Chinese germplasm collections should harbor higher levels of genetic variation than those for the U.S. ancestors. To date, little is known about quantification of the DNA-level genetic diversity using standard genetic parameters (Berg and Hamrick, 1997; Hamrick and Godt, 1997) in the Chinese soybean accessions at the provincial level and in the U.S. ancestors. This information also would aid soybean breeders in selecting parents to enhance the performance of future soybean cultivars in the USA.

In this study, we analyzed levels of RFLP variation in the random versus pre-selected subsets to test the first prediction. Next we compared our data with Lorenzen et al. (1995) who analyzed RFLP diversity in 64 soybean lines including the ancestors and milestone cultivars in the USA to test the second prediction.

MATERIALS AND METHODS

Plant materials

Seeds of the accessions used in this study were obtained from the USDA Soybean Germplasm Collection courtesy of Dr. R. Nelson (USDA-ARS, University of Illinois, Urbana). Two subsets of this collection originated from six provinces in central China (Gansu, Hebei, Henan, Jiangsu, Shaanxi, and Shanxi): the first was selected at random and the second was pre-selected based on allozyme diversity (Liao and Palmer, 1977a, b, c, d), root tip fluorescence (Torkelson and Palmer, 1997), and the pubescence tip morphology of the legumes. These subsets consisted of 43 accessions for random selection (Gansu, 8; Hebei, 2; Henan, 8; Jiangsu, 5; Shaanxi, 17; and Shanxi, 3) and 53 pre-selected accessions (Gansu, 18; Hebei, 1; Henan, 3; Jiangsu, 5; Shaanxi, 2; and Shanxi, 24) (Table 1). An additional 11 accessions (all pre-selected) from three other provinces (Anhui, 7; Ningxi, 1; and Shandong, 3) were included in this study (thus a total of 107; Table 1) to compare our accessions with the ancestral lines and milestone cultivars (n = 64) studied by Lorenzen et al. (1995).

DNA preparation and RFLP probe analysis

Seeds were planted in a greenhouse and grown to the second leaf stage. They were then collected and freeze-dried. Total DNA was extracted from the leaves using a chloroform extraction method (Sambrook et al., 1989). The DNA was digested with five restriction endonucleases, *Dra*I, *Eco*RI, *Eco*RV, *Hind*III, and *Taq*I. Restriction endonuclease digestions, electrophoresis, Southern transfer, and DNA hybridizations were performed as described by Keim et al. (1990). Genomic DNA probes were screened against these genotypes with probe-enzyme combinations that were identical to those used in the preparation of the USDA-ARS public soybean *Glycine max* by *G. soja* Sieb. Et Zucc. (A81-356022 X PI 468916) genetic map (Shoemaker and Specht, 1995). As controls, *G. max* breeding line A81-356022 and *G. soja* plant introduction PI 468916 were included for RFLP probe analysis to consistently score banding patterns.

Data analysis

The RFLP banding patterns were scored according to an allele-locus model as suggested by Bruebaker and Wendel (1994). The low copy number of RFLP fragments seen, as often observed in highly homozygous species, facilitates the use of the allele-locus model. Bands (alleles) at a given locus that were difficult to score were recorded as a "no score." A locus was considered to be polymorphic if at least one genotype had a mapped fragment that differed from the remaining genotypes. Loci that were difficult to score or had a high frequency of missing data points were discarded. A total of 46 probes were used for the analysis of the 107 accessions.

We estimated three genetic parameters to determine

Table 1. Soybean accessions from the central provinces of China assayed for RFLP diversity. The accessions were selected at random or were pre-selected based upon isozyme diversity.

PI[a]	Province	SM[b]	PI[a]	Province	SM[b]	PI[a]	Province	SM[b]
567287	Gansu	P	567428	Shanxi	P	567394A	Shaanxi	R
567292	Gansu	P	567436	Shanxi	P	567396C	Shaanxi	R
567294	Gansu	P	567440	Shanxi	P	567403B	Shaanxi	R
567295	Gansu	R	567443	Shanxi	R	567406A	Shaanxi	R
567297	Gansu	P	567446	Shanxi	P	567410B	Shaanxi	R
567304	Gansu	P	567446	Shanxi	P	567660A	Henan	R
567305	Gansu	P	567459	Shanxi	P	567682A	Henan	R
567312	Gansu	R	567462	Shanxi	P	567684B	Henan	R
567314	Gansu	P	567462	Shanxi	P	567765C	Jiangsu	R
567322	Gansu	P	567467	Shanxi	P	567299A	Gansu	P
567327	Gansu	P	567468	Shanxi	R	567299B	Gansu	P
567329	Gansu	R	567478	Shanxi	R	567316A	Gansu	R
567330	Gansu	P	567502	Hebei	P	567336A	Gansu	P
567332	Gansu	P	567509	Hebei	R	567336B	Gansu	P
567340	Gansu	R	567517	Hebei	R	567417A	Shanxi	P
567342	Gansu	P	567565	Shandong	P	567417B	Shanxi	P
567345	Gansu	R	567613	Henan	R	567417C	Shanxi	P
567348	Gansu	P	567635	Henan	P	567419A	Shanxi	P
567353	Gansu	P	567637	Henan	R	567419B	Shanxi	P
567356	Gansu	R	567657	Henan	R	567433A	Shanxi	P
567365	Ningxi	P	567663	Henan	R	567433B	Shanxi	P
567380	Shaanxi	R	567670	Henan	R	567441A	Shanxi	P
567383	Shaanxi	R	567673	Henan	P	567441B	Shanxi	P
567386	Shaanxi	R	567704	Anhui	P	567441C	Shanxi	P
567391	Shaanxi	P	567707	Anhui	P	567466A	Shanxi	P
567393	Shaanxi	R	567715	Anhui	P	567466B	Shanxi	P
567397	Shaanxi	R	567725	Anhui	P	567554A	Shandong	P
567399	Shaanxi	P	567733	Anhui	P	567554B	Shandong	P
567400	Shaanxi	R	567750	Jiangsu	R	567630A	Henan	P
567402	Shaanxi	R	567763	Jiangsu	R	567706A	Anhui	P
567405	Shaanxi	R	567766	Jiangsu	R	567706B	Anhui	P
567408	Shaanxi	R	567773	Jiangsu	R	567756A	Jiangsu	P
567412	Shaanxi	R	567776	Jiangsu	P	567756B	Jiangsu	P
567420	Shanxi	P	567349A	Gansu	R	567780A	Jiangsu	P
567421	Shanxi	P	567376A	Shaanxi	R	567780B	Jiangsu	P
567423	Shanxi	P	567381B	Shaanxi	R			

[a]PI, plant introduction numbers. For comparison between pre-selected and random selection, PIs from Ningxi (n =1), Shandong (3), and Anhui (7) were excluded because they were all pre-selected accessions. For the genetic diversity analysis at the provincial level, seven PIs from three provinces [Hebei (3), Ningxi (1), and Shandong (3)] were excluded due to very small sample sizes.

[b]Selection method: P, pre-selected and R, randomly selected.

levels of RFLP diversity using the programs POPGENE (Yeh et al., 1999) and FSTAT (ver.2.9.3 by Goudet, 2002): mean number of alleles per locus (A), mean effective number of alleles per locus (A_e), and Nei's (Nei, 1978) unbiased gene diversity (H_e). Estimates in 43 random accessions were compared with 53 pre-selected accessions from six provinces. For the genetic diversity analysis at the provincial level, seven accessions from three provinces (Hebei, 3; Ningxi, 1; Shandong, 3) were excluded due to small sample sizes. Thus, we estimated genetic diversity (A and H_e) of soybean accessions in six Chinese provinces (Anhui, Gansu, Henan, Jiangsu, Shaanxi, and Shanxi). Nei's (1973) gene diversity ($H_{eN} = 1 - \Sigma p_i^2$, whereas p_i is the frequency of the ith RFLP allele per locus) also was estimated to compare total diversity in 107 accessions with 64 accessions of the U.S. ancestral lines including milestone cultivars studied by Lorenzen et al. (1995). Lorenzen et al. (1995) estimated gene diversity per locus (H_{eN}) using the Nei's (1973) formula.

We calculated Nei's (1973) G_{ST} (proportion of total genetic diversity partitioned among provinces in China) at each locus, and then averaged them across 46 RFLP loci to determine degree of genetic differentiation among the six provinces. In addition, we tested statistical significance for genetic differentiation among the six provinces using the exact test of Raymond and Rousset (1995). This test is analogous to Fisher's exact test but uses a Markov chain to explore all potential states of an $r \times k$ contingency table based on r groups and k genotypes. This test was conducted using the program ARLEQUIN (Schneider et al., 2000) and 10000 Markov steps.

RESULTS

The 107 accessions were polymorphic for all 46 RFLP loci and a total of 115 alleles were detected across the loci, giving rise to high levels of genetic diversity (mean $A = 2.5$ and mean $H_e = 0.410$) (Table 2). The USDA germplasm collections from central China harbor significantly higher levels of genetic diversity than those for 21 ancestors and derived cultivars in the USA (mean $A = 2.1$ and mean $H_{eN} = 0.339$) (Wilcoxon signed rank test statistic: for A, $z = -3.139$, one-tail probability, $P = 0.001$ and for H_e, $z = -2.305$, $P = 0.010$).

Fifty-three pre-selected accessions (mean $A = 2.44$) had significantly more alleles than the 43 randomly selected accessions (mean $A = 2.13$) ($z = -2.982$, $P = 0.001$), but H_e was not significantly different between the two groups (mean $H_e = 0.393$ and 0.394) (Table 2). For pre-selected accessions, mean effective number of alleles per locus (A_e) was only 1.72, which highlights the fact that many of the alleles were present at very low frequencies. This is a reason that the pre-selected accessions maintain nearly the same levels of H_e and A_e as randomly selected accessions (Table 2), though the former had significantly more alleles than the latter.

At the provincial level, RFLP diversity (A and H_e) ranged from 1.64 and 0.263 (Province Anhui) to 2.26 (Province Shanxi) and 0.423 (Province Henan) with means of 2.06 and 0.368 (Table 3). There was a significant difference among the six provinces for H_e (Kruskal-Wallis test statistic: $H = 15.2$, $P = 0.009$), which is primarily due to the low estimate in Anhui. Spearman rank correlation analysis revealed that the mean number of alleles per locus (A) is closely associated with the number of accessions representing each province ($r_S = 0.886$, $P = 0.019$). This suggests that if sample size is increased in Province Anhui, we would expect more alleles per locus. However, no significant correlation between sample size and H_e was detected ($r_S = 0.086$).

There were significant differences in allele frequencies among the six provinces (mean $G_{ST} = 0.133$). Overall, about 87% of the total variation in the samples was common to the six provinces in China. In addition, the exact test of genetic differentiation among the six provinces was highly significant ($P < 0.0001$).

DISCUSSION

Genetic diversity between random accessions vs. pre-selected accessions

Our first prediction was that pre-selected accessions of Chinese soybean germplasm collections would have significantly more alleles than those from randomly selected accessions. This suggests that pre-selection of accessions based on allozyme data may be an effective approach for selecting germplasm collections with more RFLP alleles. However, we failed to detect a significant difference for gene diversity (expected heterozygosity, H_e) between pre-selected and randomly chosen accessions owing to the very low frequencies of several alleles in the pre-selected accessions. Gene diversity (H_e) calculated in this study is a composite measure that summarizes genetic variation at the locus level. The maganitude of H_e is a function of the proportion of polymorphic loci, the number of alleles per polymorphic locus, and the evenness of allele frequencies within populations or accessions. It is the most commonly used index of genetic diversity for codominant data, because it summarizes the fundamental genetic variation of a population in a single statistic (Berg and Hamrick, 1997).

Some propose that decisions for gene conservation of crops and their wild relatives should be based to a greater degree on "allelic richness" (AR) (number of alleles) (e.g., Marshall and Brown, 1975; Lee, 1998; Brown and Brubaker, 2000). In complementary analysis, thus, we calculated RFLP allelic richness between random and pre-selected accessions through rarefaction that accounts for sample size effects (Leberg, 2002) to produce unbiased estimates using the FSTAT (Goudet, 2002). A very similar to the mean number of alleles per locus, estimates of allelic richness revealed that 53 pre-selected accessions (mean AR, allelic richness = 2.26) had significantly more alleles than the 43 randomly selected accessions (mean AR

Table 2. Summary of RFLP diversity for 46 probes (loci) in central Chinese accessions and ancestors and milestone cultivars (n = 64) in the USA.

	Pre-selected			Random			Total sample			NAC[e]	
	A[a]	A_e[b]	H_e[c]	A[a]	A_e[b]	H_e[c]	A[a]	H_e[c]	H_{eN}[d]	A[a]	H_{eN}[d]
A023	3	1.06	0.081	2	1.06	0.057	3	0.070	0.070	2	0.47
A063-1	2	1.90	0.483	2	2.00	0.511	2	0.500	0.494	2	0.28
A063-2	3	2.05	0.519	3	1.97	0.500	3	0.507	0.504	n.a.	n.a.
A085	3	2.39	0.593	2	1.96	0.505	3	0.573	0.567	2	0.50
A086	2	2.00	0.510	2	1.76	0.444	2	0.493	0.488	2	0.36
A095	2	1.90	0.485	2	1.85	0.472	2	0.474	0.469	2	0.40
A186	3	1.46	0.324	2	1.29	0.232	3	0.286	0.283	2	0.20
A257	3	1.98	0.504	2	1.78	0.450	3	0.476	0.471	2	0.45
A333	3	2.15	0.545	2	1.98	0.511	3	0.526	0.520	2	0.13
A374	3	2.04	0.522	2	2.00	0.512	3	0.531	0.526	2	0.10
A381	3	1.13	0.121	2	1.44	0.315	3	0.205	0.201	2	0.38
A398	2	1.97	0.503	2	2.00	0.515	2	0.504	0.498	2	0.31
A401-1	2	1.95	0.495	2	1.94	0.497	2	0.505	0.500	2	0.48
A401-2	2	1.93	0.493	2	1.98	0.508	2	0.504	0.499	n.a.	n.a.
A461-1	2	1.89	0.482	2	1.72	0.431	2	0.458	0.453	2	0.48
A461-2	2	1.08	0.074	2	1.05	0.049	2	0.062	0.062	n.a.	n.a.
A481	3	2.15	0.543	2	1.82	0.463	3	0.512	0.505	2	0.06
A505	2	1.76	0.439	2	1.86	0.477	2	0.499	0.493	2	0.45
A520	2	1.04	0.041	2	1.39	0.286	2	0.153	0.151	2	0.06
A567	2	2.00	0.510	2	1.66	0.409	2	0.490	0.484	2	0.28
A586-1	2	1.19	0.164	2	1.67	0.414	2	0.267	0.263	2	0.47
A586-2	2	1.72	0.429	2	1.85	0.476	2	0.441	0.435	n.a.	n.a.
A668	2	1.59	0.378	2	1.67	0.413	2	0.388	0.384	2	0.33
A681	3	2.05	0.523	3	1.71	0.427	3	0.482	0.476	2	0.43
A691-1	2	1.76	0.439	2	1.40	0.292	2	0.385	0.381	3	0.47
A691-2	2	1.96	0.500	2	1.88	0.481	2	0.488	0.483	n.a.	n.a.
A702	3	1.93	0.491	2	1.83	0.468	4	0.537	0.531	2	0.50
A708	2	1.97	0.503	2	1.69	0.440	2	0.508	0.500	2	0.46
A806	2	1.40	0.294	3	1.71	0.427	3	0.352	0.349	2	0.33
A816	3	2.09	0.531	2	1.36	0.272	3	0.449	0.444	2	0.26
A847	1	1.00	0.000	1	1.00	0.000	2	0.390	0.384	2	0.50
A890	4	2.41	0.597	3	1.80	0.456	4	0.547	0.541	2	0.20
A946-1	3	1.65	0.402	2	1.63	0.398	3	0.396	0.391	2	0.04
A946-2	3	1.60	0.382	3	1.97	0.504	3	0.430	0.426	2	0.15
A963	2	1.32	0.350	2	1.58	0.378	2	0.358	0.354	2	0.40
B039	2	1.99	0.507	2	1.97	0.505	2	0.506	0.500	2	0.40
B122	2	1.32	0.250	2	1.44	0.315	2	0.273	0.269	2	0.50
B164	3	1.09	0.081	2	1.14	0.129	3	0.098	0.098	2	0.44
B166	6	1.63	0.393	4	1.86	0.475	6	0.429	0.424	3	0.33
K002	2	1.63	0.394	2	1.58	0.378	2	0.382	0.378	2	0.44
K003	2	1.89	0.472	2	1.33	0.258	2	0.499	0.493	2	0.40
K007	2	1.48	0.332	2	1.65	0.405	2	0.361	0.357	3	0.57
K069-1	2	1.60	0.384	2	1.98	0.505	2	0.483	0.478	2	0.24
K070-1	2	1.41	0.296	2	1.67	0.412	2	0.351	0.347	2	0.10
K070-2	2	1.26	0.212	2	1.32	0.246	2	0.225	0.223	n.a.	n.a.
R017	2	1.97	0.502	2	1.98	0.508	2	0.560	0.500	2	0.30
Mean	2.44	1.72	0.393	2.13	1.68	0.394	2.50	0.410	0.405	2.08	0.34
1±SE	0.12	0.06	0.024	0.07	0.04	0.020	0.12	0.019	0.019	0.04	0.02

[a]Mean number of alleles per locus or probe. n.a. represents data not available.
[b]Mean effective number of alleles per locus.
[c]Nei's (1978) unbiased gene diversity.
[d]Nei's (1973) biased gene diversity.
[e]NAC, ancestors and milestone cultivars in the USA; H_{eN} estimates from Lorenzen et al. (1995).

Table 3. Summary of RFLP diversity for 46 loci in six provinces in central China.

Province[a]	N[b]	*A* (SE)[c]	H_e (SE)[d]
Anhui	7	1.64 (0.08)	0.263 (0.035)
Gansu	26	2.15 (0.09)	0.405 (0.026)
Henan	11	2.04 (0.07)	0.423 (0.029)
Jiangsu	10	2.07 (0.07)	0.409 (0.026)
Shaanxi	19	2.21(0.07)	0.353 (0.025)
Shanxi	27	2.26 (0.09)	0.355 (0.028)
Mean	16.7	2.06 (0.09)	0.368 (0.024)

[a]Seven accessions from three provinces (Hebei, 3, Ningxi, 1, and Shandong, 3) were not included due to small sample sizes.
[b]Sample size.
[c]Mean number of alleles per locus or probe. SE, standard error.
[d]Nei's (1978) unbiased gene diversity.

= 2.10) (z = -2.812, P = 0.002). If several low frequency alleles that are not observed in randomly selected accessions are closely linked with other agronomically important traits, our approach to pre-selection would be a useful tool for developing core collections. In other words, low frequency alleles may not effect allelic evenness measures (e.g., expected heterozygosity or H_e), but could provide significant affects when incorporated into a plant breeding program.

The 107 soybean accessions from central China harbor comparable levels of genetic diversity (mean A = 2.5 and mean H_e = 0.410) relative to other crops. For example, similarly high levels of RFLP genetic diversity for 35 probe-restriction enzyme combinations were observed in ten maize populations of European and northern U.S. germplasm. Dubreuil and Charcosset (1998) estimated 3.5 of the mean number of alleles per locus and 0.470 of Nei's (1978) unbiased genetic diversity in the ten maize populations.

Genetic diversity between central China vs. ancestors and derived cultivars in the USA

Our second prediction was that the accessions from central China harbor significantly higher levels of genetic diversity than those for 21 ancestors and derived cultivars in the USA. Considering the fact that the 43 non-ancestral U.S. cultivars studied by Lorenzen et al. (1995) trace back to just 19 of the earliest plant introductions used in breeding programs in the USA, and that among the 19 ancestral lines, only 12 account for 88% of the germplasm in the derived 43 cultivars, our results are not surprising. Keim et al. (1992) analyzed 38 soybean accessions in the USA (18 ancestral lines and 20 adapted lines) with 132 RFLP probes and found 69% polymorphism in the probes and a very similar estimate (H_{eN} = 0.30) to Lorenzen et al. (1995). Skorupska et al. (1993) identified 29 RFLP probes with gene diversity estimate ≈ 0.30 in elite southern U.S. cultivars and ancestral lines of those cultivars. All these studies suggest that the USDA Soybean Germplasm Collection from China harbors significantly higher levels of genetic diversity than those for the U.S. ancestral lines and their derived cultivars.

More recently, three of the authors analyzed allozyme variation in 1777 accessions in the USDA Soybean Germplasm Collection taken from 19 provinces in China and estimated high levels of genetic diversity (H_e = 0.240, Chung et al., unpublished data). There was no significant difference for H_e between the 83 soybean ancestors both in the U.S. and China and the 1777 accessions from China, suggesting that the ancestral lines in the U.S. and China also harbor high levels of allozyme diversity. However, our RFLP data revealed a significant difference between accessions from central China and the U.S. ancestors and their derived cultivars, suggesting that data from the RFLPs are more informative than those from allozyme data for evaluation of the USDA Chinese Germplasm Collection. Again, previous studies by simple sequence repeats (SSRs) revealed high levels of genetic diversity in Asian soybean accessions. Abe et al. (2003) analyzed 20 SSR loci in 131 Asian soybean accessions from 14 Asian countries and found high levels of genetic diversity (mean H_e = 0.782) in the accessions. To determine if the Asian accessions harbor significantly higher levels of genetic diversity than those of the U.S. ancestors (Diwan and Cregan, 1997) and their derived elite lines (Narvel et al., 2000), we selected the same SSR loci (five loci for each comparison) for pairwise comparisons between studies of Abe et al. (2003) and Diwan and Cregan (1977) and Narvel et al. (2000). The 131 Asian accessions harbored significantly larger number of alleles per locus and higher levels of genetic diversity than those of the 35 ancestors in the U.S. (Diwan and Cregan, 1997) (mean A = 12.6 vs. 8.8, Wilcoxon signed rank test: z = -2.023, P = 0.022, one tail probability; H_e = 0.842 vs. 0.772, z = -1.753, P = 0.040). Similarly, the Asian accessions maintained significantly higher levels of genetic diversity than those of 39 U.S. elite soybeans (Narvel et al., 2000) (mean A = 12.8 vs. 3.0, Wilcoxon signed rank test: z = -2.023, P = 0.026, one tail probability; H_e = 0.841 vs. 0.426, z = -2.023, P = 0.022). Thus, the present RFLP and previous SSR studies consistently revealed that Chinese soybean accessions and other Asian accessions harbor significantly higher levels of genetic diversity than those of the U.S. ancestors and their derived cultivars.

We noted four points resulting from this study that might be interesting to soybean breeders in both the USA and China. First, our results are consistent with previous studies that found levels of the genetic variation were greater for RFLPs than for allozymes in a number of crop plants (Messmer et al., 1991; McGrath and Quiros, 1992; Zhang et al., 1993; Dubreuil and Charcosset, 1998). Second, comparing our results with Lorenzen et al. (1995), the probes used in this study were clearly effective at detecting overall diversity in the Chinese germplasm

collections. Third, as pre-selected accessions of Chinese soybean germplasm collections have significantly more alleles than those from randomly selected accessions, allozymes have proven to be an effective tool for identifying an RFLP diverse (in terms of the number of alleles across loci) germplasm collection, though less diversity was detected with allozymes. Finally, we found a significant difference in levels of genetic diversity among the six provinces in central China and also detected significant differences in allele frequencies among the six provinces. Collectively, all these results suggest that Chinese germplasm collections from various regions or provinces are important to enhancing the genetic diversity of the current soybean cultivars in the US.

In summary, pre-selected accessions based on allozyme diversity have significantly more alleles than randomly selected accessions. Our results suggest that the use of prior data to select genetically diverse germplasm in terms of the number of alleles would be effective. This may be a helpful tool for breeders interested in working with germplasm collections to select for unique traits. It is our hope that methods to identify small but diverse sets of the USDA Soybean Germplasm Collection will encourage breeders in the future to use these accessions in their breeding programs.

Acknowledgements. This research was in part supported by a grant (USB Project # 9211: Identification and utilization of exotic germplasm to improve soybean productivity) from the United Soybean Board to the USDA. Eric Myers read earlier versions of the manuscript and made helpful suggestions. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the USDA or by Iowa State University and does not imply its approval to the exclusion of other products or vendors that also may be suitable. M. G. Chung gratefully acknowledges the exchange program sponsored by the USDA/FAS/ICD/ Research and Scientific Exchange Division from June 2001 to September 2002.

LITERATURE CITED

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

Berg, E.E. and J.L. Hamrick. 1997. Quantification of genetic diversity at allozyme loci. Can. J. For. Res. **27:** 415-424.

Bretting, P.K. and M.P. Widrlechner. 1995. Genetic markers and plant genetic resource management. *In* J. Janick (ed.), Plant Breeding Reviews, Vol. 13, John Wiley and Sons, Indianapolis, pp. 11-86.

Brown, A.H.D. 1989. Core collections: a practical approach to genetic resources management. Genome **31:** 818-824.

Brown, A.H.D. and C.L. Brubaker. 2000. Genetics and the conservation and use of Australian wild relatives of crops. Aust. J. Bot. **48:** 297-303.

Brown-Guedira, G.L., J.A. Thompson, R.L. Nelson, and M.L. Warburton. 2000. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Sci. **40:** 815-823.

Bruebaker, C.L. and J.F. Wendel. 1994. Reevaluation of the origin of domesticated cotton (*Gossypium hirsutum*: Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Amer. J. Bot. **81:** 1309-1326.

Crossa, Z., C.M. Hernandez, P. Bretting, S.A. Eberhart, and S. Taba. 1993. Statistical genetic considerations for maintaining germplasm collections. Theor. Appl. Genet. **86:** 673-678.

Cui, Z., T.E. Carter Jr., J. Cai, J. Qui, and R.L. Nelson. 1999. Origin, description, and pedigree of Chinese soybean cultivars released from 1923 to 1995. Washington: U.S. Department of Agriculture, Agricultural Research Service, Tech. Bull. No. 1871.

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.

Dubreuil, P. and A. Charcosset. 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theor. Appl. Genet. **96:** 577-587.

Gizlice, Z., T.E. Carter Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. **34:** 1143-1147.

Goudet, J. 2002. FSTAT, a program to estimate and test gene diversities and fixation indices (ver. 2.9.3.2). Dorigny, Switzerland: Université de Lausanne. Available at http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Grabau, E.A., W.H. Davis, and B.G. Gengenbach. 1989. Restriction fragment length polymorphism in a subclass of the "Mandarin" soybean cytoplasm. Crop Sci. **29:** 1554-1559.

Griffin, J.D. and R.G. Palmer. 1995. Variability of thirteen isozyme loci in the USDA soybean germplasm collections. Crop Sci. **35:** 897-904.

Hamrick, J.L. and M.J.W. Godt. 1997. Allozyme diversity in cultivated crops. Crop Sci. **37:** 26-30.

Helentjaris, T., G. King, M. Slocum, C. Siedenstrang, and S. Wegman. 1985. Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol. Biol. **5:** 109-118.

Keim, P., R.C. Shoemaker, and R.G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. **77:** 786-792.

Keim, P., B.W. Diers, T.C. Olson, and R.C. Shoemaker. 1990. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics **126:** 735-742.

Keim, P., W. Beavis, J. Schupp, and R. Freestone. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. Theor. Appl. Genet. **85:** 205-212.

Kiang, Y.T., Y.C. Chiang, J.H.Y. Doong, and M.B. Gorman. 1987. Genetic variation of soybean germplasm. *In* S. C. Hsieh (ed.), Crop exploration and utilization of genetic resources, Taichung District Agricultural Improvement Station Publisher, Changhua, Taiwan, pp. 93-99.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. **37:** 1317-1325.

Leberg, P. L. 2002. Estimating allelic richness: Effects of sample size and bottlenecks. Mol. Ecol. **11:** 2445-2449.

Lee, M. 1998. Genome projects and gene pools: New germplasm for plant breeding? Proc. Natl. Acad. Sci. USA. **95:** 2001-2004.

Li, Z., L. Qiu, J.A. Thompson, M.M. Welsh, and R.L. Nelson. 2001. Molecular genetic analysis of U.S. and Chinese soybean ancestral lines. Crop Sci. **41:** 1330-1336.

Liao, W. and R.G. Palmer. 1997a. Genetic study of a diaphorase 2 null mutant. Soybean Genet. Newsl. **24:** 157-159.

Liao, W. and R.G. Palmer. 1997b. Inheritance and linkage studies of an aconitase 1 null in soybean. Soybean Genet. Newsl. **24:** 160-163.

Liao, W. and R.G. Palmer. 1997c. Inheritance and linkage studies of 6-phosphogluconate dehydrogenase in soybean. Soybean Genet. Newsl. **24:** 164-167.

Liao, W. and R.G. Palmer. 1997d. A new variant of phosphoglucose isomerase in soybean. Soybean Genet. Newsl. **24:** 179-182.

Lorenzen, L.L., S. Boutin, N. Young, J.E. Specht, and R.C. Shoemaker. 1995. Soybean pedigree analysis using map-based molecular markers: I. tracking RFLP markers in cultivars. Crop Sci. **35:** 1326-1336.

Lorenzen, L.L. and R.C. Shoemaker. 1996. Genetic relationships within soybean cultivar groups. Crop Sci. **36:** 743-752.

Marshall, D.R. and A.H.D. Brown. 1975. Optimum sampling strategies in genetic conservation. *In* O. H. Frankel and J. G. Hawkes (eds.), Crop genetic resources for today and tomorrow. Cambridge University Press, London, pp. 53-80.

McGrath, J.M. and C.F. Quiros. 1992. Genetic diversity in isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. Theor. Appl. Genet. **83:** 783-790.

Messmer, M.M., A.E. Melchinger, M. Lee, W.L. Woodman, E.A. Lee, and K.R. Lamkey. 1991. Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor. Appl. Genet. **83:** 97-107.

Narvel, J.M., W.R. Fehr, W.C. Chu, D. Grant, and R.C. Shoemaker. 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:** 3321-3323.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89:** 583-590.

Perry, M.C., M.S. McIntosh, and A.K. Stoner. 1991. Geographical patterns of variation in the USDA soybean germplasm collection: II. Allozyme frequencies. Crop Sci. **31:** 1356-1360.

Raymond, M. and F. Rousset. 1995. Genepop (V. 1.2): Population genetics software for exact tests and ecumenicism. J. Hered. **86:** 248-249.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor.

Schneider, S., D. Roessli, and L. Excoffier. 2000. ARLEQUIN ver. 2000: A software for population genetic data analysis. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva. Available at http://anthro.unige.ch/arlequin.

Shoemaker, R.C. and J.F. Specht. 1995. Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. **35:** 436-446.

Skorupska, H.T., R.C. Shoemaker, A. Warner, E.R. Shipe, and W.E. Bridges. 1993. Restriction fragment length polymorphism in soybean germplasm of the southern U.S.A. Crop Sci. **33:** 1169-1176.

Smartt, J. and T. Hymowitz. 1985. Domestication and evolution of grain legumes. *In* R. J. Summerfield and E. H. Roberts (eds.), Grain Legumes, Collins, London, pp. 37-72.

Torkelson, J. and R.G. Palmer. 1997. Genetic analysis of root fluorescence of central Chinese germplasm. Soybean Genet. Newsl. **24:** 147-149.

Wang, R.W., Z.L. Yang, H.S. Kim, and C. Yen. 1998. Comparative analysis of RFLP diversity in landraces of *Triticum aestivum* and collections of *T. tauschii* from China and southwest Asia. Theor. Appl. Genet. **96:** 312-318.

Yeh, F.C., R.C. Yang, and T.B.J. Boyle. 1999. POPGENE Version 1.31, Microsoft window-based free ware for population genetic analysis. Alberta, Canada: Centre for International Forestry Research, University of Alberta. Available at http://www.ualberta.ca/~fyeh/index.htm.

Yeeh, Y. S.S. Kang, and M.G. Chung. 1996. Evaluations of the natural monument populations of *Camellia japonica* (Theaceae) in Korea based on allozyme studies. Bot. Bull. Acad. Sin. **37:** 141-146.

Zhang, Q., M.A. Saghai Maroof, and A. Kleinhofs. 1993. Comparative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. *spontaneum*. Genetics **134:** 909-916.

源自華中地區之美國農業部大豆種源庫之核酸限制酶切割長度多形性分析

Myong Gi CHUNG[1], Mi Yoon CHUNG[1], April D. Clikeman JOHNSON[2], and Reid G. PALMER[3]

[1]Department of Biology, Gyeongsang National University
Jinju 660-701, Republic of Korea
[2]Department of Agronomy, Iowa State University, Ames, IA 50011, USA
[3]USDA-ARS, Corn Insects and Crop Genetics Research Unit, Project 3769
Department of Agronomy, Iowa State University, Ames, IA 50011, USA

為評估源自中國中部之美國農業部大豆種源庫之基因變異度，共 107 登錄樣品之 46 基因座 (loci) 以核酸限制酶切割長度多形性 (RFLP) 分析之。我們比較逢機取樣和預先選拔兩組試驗材料之基因變異度；比較項目有：根尖螢光，成熟期之形態，以及共 10 種酵素之同功酶型。我們也同時評估此 107 源自華中之登錄樣品之基因變異度，其方法為：和在美國先前已研究過的且具劃時代意義之栽培品種 (NAC, *n* = 64)。最後，我們估算六個中國省份（安徽、甘肅、河南、江蘇、陝西、和山西）之品種的基因歧異分化度。如以每一基因座 (locus) 所含之平均的相對基因 (allele) 數 (*A*, 2.44 vs. 2.13)，及相對基因之豐富度 (2.23 vs. 2.10) 兩參數為憑，則上述預先選拔組和逢機取樣組兩者之間有顯著差異。但是，前者 (*He* = 0.393) 維持基因差異度（或稱期望之歧異度，*He*）相似於後者 (*He* = 0.394)。此乃因：事實上預先選拔組所發現之很多相對基因是以低頻存在的（平均有效相對基因數，*Ae* = 1.72）。在預先選拔組所發現之分佈較廣之相對基因暗示：以同功酶為根據所選擇出之預先選拔組可能適用於篩選具較多數目之 RFLP 相對基因的基因庫收集樣品。源自華中之登 錄樣品 (*n* = 107) 維持較高水平之基因變異度（*He* = 0.405, A= 2.50）當以 NAC (*He* = 0.339, *A* = 2.08) 為比較基準時。我們檢測出上述六個省份樣品間有顯著的基因分化（平均 G_{ST} = 0.133）。這些結果暗示：來自中國各地區或省份之美國農業部種源庫之登錄種源可用來強化美國大豆品種之基因變異度。

關鍵詞：基因變異；*Glycine max*；預選；核酸限制酶切割長度多形性；美國農業部大豆種源庫。