*Pak. J. Bot.*, 41(5): 2229-2242, 2009.

# GENETIC DIVERSITY IN ANNUAL WILD SOYBEAN (*GLYCINE SOJA* SIEB. ET ZUCC.) AND CULTIVATED SOYBEAN (*G. MAX*. MERR.) FROM DIFFERENT LATITUDES IN CHINA

**WEI AN[1], HONGKUN ZHAO[2], YINGSHANDONG[1,2]*, YUMIN WANG[2], QIYUN LI[2] , BINGCHANG ZHUANG[2], LEI GONG[1] AND BAO LIU[1]***

*[1]Key Laboratory of Molecular Epigenetics of MOE, Northeast Normal University, Changchun 130024, China, [2]Jilin Academy of Agricultural Science, Changchun 130124, China. Corresponding author: ysdong@cjaas.com or baoliu@nenu.edu.cn*

**Abstract**

A total of 50 accessions of annual wild soybean (*Glycine soja*) and cultivated soybean (*G. max*) were used to study genetic variations by three genetic fingerprinting systems, RAPD, SSR, and AFLP. Ninety polymorphic RAPD bands were generated by 10 decamer primers, 56 SSR loci were obtained by 12 pairs of primers and 935 bands were resolved by 11 AFLP primer combinations. Based on the marker data, an UPGMA-cluster analysis separated the soybean collections into two groups that corresponded to wild soybean and cultivated soybean respectively. Genetic diversity calculation suggested that the diversity of *G. soja* was higher than that of *G. max*. Specific informative bands for *G. soja* and *G. max* were detected, which provided molecular evidence in support of the clear distinction of *G. soja* and *G. max*. Finally, it was found that environmental factors may play important roles in soybean evolution. A comparison of the three molecular markers suggests that AFLP and SSR are more suitable than RAPD for genetic diversity studies in soybean.

## Introduction

Annual wild soybean (*Glycine soja*) is the direct progenitor of cultivated soybean (*G. max*). Because the two species (or forms) are sexually compatible to an extent, *G. soja* is considered as a valuable source of genetic diversity for soybean breeding. Annual wild soybean is widely distributed in China, which has high protein content and demonstrates a high tolerance to various environmental stresses. In addition to being essential to elucidating the origins and evolution of cultivated soybean, knowledge gained in annual wild soybean may provide genetic information useful for more efficient improvement of soybean cultivars.

Rapid progress in molecular biology and genomics have greatly improved the reliability and usefulness of a variety of DNA-based molecular markers such as random amplified polymorphic DNA (RAPD), simple sequence repeats (SSRs) and amplified fragment length polymorphisms (AFLPs). These DNA-based markers have many advantages over morphological and isozyme-based biochemical markers, particularly in the dramatically increased saturation of genome coverage provides much detailed information on the nature and extent of genetic variation within and between species (Storfar, 1996). Generally, an ideal molecular marker should possess the following characteristics: (1) rich polymorphism, (2) being easy to work with, (3) robustness and (4) cost effectiveness. Thus, the randomly amplified polymorphic DNA or RAPD (Williams & Kubelik, 1990) is a favored method due to its technical advantages over restriction fragment length polymorphism (RFLP), including being non-radioactive, easier to handle, and requiring only a small amount of DNA, in genetic diversity studies, including soybean. Unfortunately, RAPD markers are generally dominant, non-locus-specific, and have low reproducibility. This potential skewing however can be reduced by examining a large number of RAPD loci (>30) and using data analysis techniques that can compensate for the underestimation of the recessive alleles by calculating allele frequencies (Lynch & Milligan, 1994, Gillies *et al*., 1999).


Fig. 1.Geographic localities where the samples were collected (designated by solid circles).

SSRs, or microsatelites, are fast-evolving and ubiquitous in eukaryotic genomes. Akkaya *et al*., (1992) conducted one of the first SSR analyses in soybean, followed by an array of studies in soybean genetic diversity (Yu & Maroof 1994; Jiang *et al*., 1995; Akkaya *et al*., 1995; Maughan *et al*., 1995; Powell *et al*., 1996). The amplified fragment length polymorphism (AFLP) marker, by combining the advantages of RFLP and RAPD, represents the most informative, reliable and robust method (Vos *et al*., 1995), which allows for the simultaneous analysis of a large number of DNA bands per gel-run. Moreover, AFLP markers are largely locus-specific.

In the present investigation, RAPD, SSR and AFLP were simultaneously used to study the genetic diversity in both annual wild and cultivated soybeans. First, 50 accessions collected from different latitudes in China were analyzed by RAPD. Then, 44 accessions were analyzed by RAPD and SSR, and finally, 20 accessions were further analyzed by AFLP. The efficiency of the three molecular markers was evaluated.

## Materials and Methods

**Plant materials:** Fifty accessions or cultivars of wild (*Glycine soja*) and cultivated soybean (*G. max*) from 24 to 52° N latitude in China were selected, inclusive of nearly every 1° of latitude (Fig. 1). All plant lines were maintained at the germplasm collections of the Jilin Academy of Agricultural Sciences in Changchun, Jilin Province, China (Table 1).

**DNA extraction:** Total genomic DNA was extracted from fresh leaves harvested from field-grown plants using the CTAB (cetyltrimethylammonium bromide) methods (Rogers *et al.,* 1988). Extracted DNA was treated with RNase at 37°C overnight and purified by ethanol precipitation. DNA samples were diluted to 100 ng/ul with TE buffer and stored at –20°C.

**Table 1. The average genetic distance (AGD) and polymorphism of wild and cultivated soybean analyzed by the three markers.**

| | Polymorphism (%) | | AGD | |
|---|---|---|---|---|
| | Wild soybean | Cultivated soybean | Wild soybean | Cultivated soybean |
| RAPD | 79.55 | 73.81 | 0.172 | 0.162 |
| SSR | - | - | 0.176 | 0.150 |
| AFLP | 28.4 | 8.2 | 0.135 | 0.091 |

**Table 2. The average genetic distance of different individuals in different latitude ranges.**

| Latitude (°N) | AGD of wild soybean | | AGD of cultivated soybean | |
|---|---|---|---|---|
| | SSR | AFLP | SSR | AFLP |
| 40-52 | 0.180 | 0.124 | 0.124 | 0.084 |
| 30-39 | 0.229 | 0.131 | 0.170 | 0.085 |
| <30 | 0.256 | 0.146 | 0.187 | 0.114 |

**RAPD:** PCR was performed in 25 µl of a reaction mixture containing 2.5 µl 10x buffer (100 mM Tris- HCl [pH=8.3], 500 mM KCl, 20 mM MgCl2, 0.001% geltatin, 0.1% Np-40), 0.2 mM each of dNTP (Perkin-Elmer, Waltham, Massachusetts, USA), 0.2 µM of each primer (Operon ,Cologne, Germany), 1 U Taq polymerase (Perkin-Elmer) and 50 ng of template DNA. Amplification was performed with a Perkin-Elmer model 480 thermal cycler (PE 480). The PCR conditions were as follows: (1) 94°C for 4 min., followed by 5 cycles of denaturation at 94°C for 1 min., (2) annealing at 37°C for 2 min., and an extension at 72°C for 2 min., (3) 35 cycles of denaturation at 94°C for 30 sec, (4) annealing at 37°C for 1 min., and xtension at 72°C for 1.5 min., and (5) a final extension at 72°C for 10 min. The PCR products were resolved on 1.5% agarose gel. After electrophoresis, gels were stained with Ethidium bromide, destained with distilled water, visualized and photographed with Gel Doc 1000 (Bio Rad). The reproducibility of PCR experiments was confirmed by repeating the procedure at least twice.

**SSR:** PCR was conducted in 25 µl of a reaction mixture containing 2.5 µl 10× buffer (100 mM Tris-HCl [pH=8.3], 500 mM KCl, 20 mM MgCl2, 0.001% geltatin, 0.1 % Np-40), 0.2 mM each of dNTP (Perkin-Elmer), 0.2 µM each of forward primer and reverse primer (Table 2), 1 U Taq polymerase (Perkin-Elmer) and 60 ng of template DNA. Amplification was performed in a PE 480. The PCR conditions were as follows: 1) 94°C for 5 min., followed by 30 cycles of denaturation at 94°C for 30 sec., 2) annealing at 52°C for 30 sec and extension at 68°C for 30 sec., and finally 3) a final extension at 72°C for 10 min. The PCR products were checked on 2.0% agarose gel. After electrophoresis, gels were stained with Ethidium bromide, destained with distilled water, visualized and photographed with Gel Doc 1000 (Bio-Rad). The reproducibility of PCR experiments was confirmed by repeating the procedure at least twice.

**AFLP:** The AFLP method was performed as described by Vos *et al*., (1995). The restriction enzymes *Mse*I (frequent cutter) and *Hin*dIII (rare cutter) were used. Adapters to the digestion reactions for MseI were 5'-GACGATGAGTCCTGAG-3' and 5'-TACTCAGGACTCAT-3'; adaptors for HindIII were 5'-CTCGTAGACTGCGTACC-3'

and 5'-AGCTGGTACGCAGTCTAC-3'. Amplifications were performed in two steps, the first of which (pre-amplification) involved the *Mse*I 5'-GATGAGTCCTGAGTAAC-3' primers with two selective nucleotides. Primers with three selective nucleotides were used for the second step. PCR amplification involved a radio-labeled (α-33P dCTP) *Hin*d III primer in combination with the MseI primer. Pre-warmed 1×TBE buffer was used to pre-run a 6% polyacrylamide sequencing gel system (Sequi-Gen GT electrophoresis cell, Bio Rad). A total of 3 µl of amplified product (each) was loaded when the gel temperature reached 50°C. Electrophoresis was performed at a constant power level (130 W) for about 2.5 h. Gels were dried on Whatman 3MM paper in a gel dryer at 80°C and exposed to X-ray film (Kodak) for 3-4 days at –70°C.

**Cluster analysis:** Results were visually scored and recorded in a data matrix for the presence ("1") or absence ("0") of a given band from different primer (RAPD) or primer combinations (SSR and AFLP). Genetic diversity was partitioned within groups of accession (Nei, 1973; Nei & Li 1979) and the genetic distance (GD) and genetic similarity (GS) were calculated using the POPGENE software (available at URL http://www.ualberta.ca/~fyeh). A dendrogram based on genetic distance (Nei, 1972) was constructed by POPGENE using the unweighted pair group method and the arithmetic average (UPGMA) procedure.

**AMOVA analysis:** To test the genetic variation between the wild soybean accessions and soybean cultivated, the non-parametric analysis of molecular variance (AMOVA) (Excoffier *et al.,* 1992) using Arlequin version 3.01 (Excoffier *et al.,* 2005) was applied. AMOVA was based on the pairwise squared euclidian distances between molecular phenotypes, which are equal to the number of different band states, because band states can only take the value "0" or "1" (Reisch *et al.,* 2005). It thus allowed the calculation of variance components (wild--G. soja accessions; cultivated--G. max accessions) and their significance levels for variation based on data of each of the three markers.

## Results

**RAPD analysis:** Ten decamer RAPD primers were selected for this analysis because of their reliable reproducibility and good polymorphism in soybean lines based on a pilot preliminary investigation. Ninety bands were scored, with 4 to 15 bands by each primer (Fig. 2). Sixteen bands were found conserved in all the soybean lines, leaving 74 informative bands (82.2 %). Among these polymorphic bands, the polymorphism frequency for wild soybean was 79.55%, a little higher than that of cultivated soybean (73.81 %) (Table 1). The calculated average genetic distance (AGD) of wild soybean (0.172) was slightly higher than that of cultivated soybean (0.162), indicating a somewhat higher level of genetic diversity in the former.

**SSR analysis:** Forty pairs of soybean SSR primers that previously showed clear polymorphism (Jiang RW *et al*., 1995) were used in this study. Twelve of them with good polymorphism were analyzed. Fifty-six alleles were detected with an average of 4.7 alleles per locus across the soybean lines (Fig. 3). The AGDs of wild soybean and cultivated soybean were 0.176 and 0.150 respectively, again indicating higher genetic diversity among the wild soybean accessions than that among the soybean cultivars (Table 1). In the relationship dendrogram constructed from the SSR data, two distinctive

groups (I and II) are apparent (Fig. 4), which are inclusive of the wild soybean and cultivated soybean, respectively, except for three lines, "7" (ZYD 00152), "31" (ZYD 03564) and "49" (ZYD 05173) (annotated with "*" in Fig. 4). These three lines might belong to some semi-wild or semi-cultivated types, and thus showed characteristics of both species. Importantly, these results corresponded to our previous morphological observations (Dong *et al.,* 2001, 2004). Notably, although most of the polymorphisms were intraspecific i.e., being exhibited by one or more accessions or cultivars within the wild or cultivated soybean, some loci specific to either the wild or domestic soybeans were observed (Fig. 3).

**AFLP analysis:** Twenty accessions were further studied by AFLP, including 14 selected from 44 accessions and 6 other accessions from 52° N (ZYD00004, Tiejiaqing), 29° N (ZYD04672, Yueyangbayuebao), and 25° N (ZYD05280, Nandanerzhaohuangdou). Eleven primer combinations (E-TA/M-CTA, E-TA/M-CAT, E-TC/M-CTG, E-TC/M-CAG, E-AG/M-CTG, E-AG/M-CTA, E-TG/M-CTA, E-TG/M-CTG, E-TG/M-CAG, E-TA/M-CTC, E-TA/M-CTG) were selected from 40 tested, which were found to produce easy-to-score profiles, and to produce a high level of polymorphism. From these 11 primer combinations, 935 bands (mean of 85 bands/primer combination) were scored of which 402 bands were polymorphic across the studied lines (Fig. 5). Notably, of these 402 polymorphic bands, 114 (28.4%) exhibited polymorphism within wild soybean and 33 (8.2%) within cultivated soybean, respectively (Table 1). The apparently larger number of amplified polymorphic bands in wild soybean relative to cultivated soybean might be due to the genetic bottlenecks that are known to have associated with soybean domestication (Hyten *et al.,* 2006). Therefore, those genetic variations unique to wild soybean accessions hold great promise if to be introgressed into soybean breeding programs for evaluation and selection.

The AFLP-based AGDs of wild and cultivated soybeans were 0.135 and 0.091 respectively (Table 1). The dendrogram showing relationships among the 20 lines formed two distinctive groups, I and II, and three subgroups in each group (designated as Ia, Ib, Ic and II a, IIb and IIc respectively) could be further divided at GD levels of 0.145 and 0.100. Group I and II were exclusively comprised of wild and cultivated soybeans respectively (Fig. 6). The subgroups of wild soybean and cultivated soybean appeared to correspond to the latitude of origins, although not all data could be classified on a latitudinal basis. For example, ZYD03564 and ZYD03328 were from the same latitude (35° N) but different longitude (114° E, 116° E, respectively), and were divided into two different subgroups. The same was true for Zhengzhou, Dalihuang and Huangdou, which suggests longitude may also have played an important direct or indirect effect on soybean evolution. For example, although being from different latitudes, ZYD03328 grouped with ZYD05191 and ZYD05280, possibly because they originated from approximately the same longitude and high elevation (Table 2). Therefore, as with many other plant species, environmental factors influencing soybean evolution may be connected with latitude and the special conditions of high elevation.

**AMOVA of the three makers:** For the variation between wild accessions and soybean cultivars, 3 independent AMOVA respectively revealed 10.97% and 89.03% variation by RAPD, 19.79% and 80.21% variation by SSR, and 21.56% and 78.44% variation by AFLP (Table 3). All these calculations were statistically significant ($p<0.001$), thus substantiated the above group clustering.

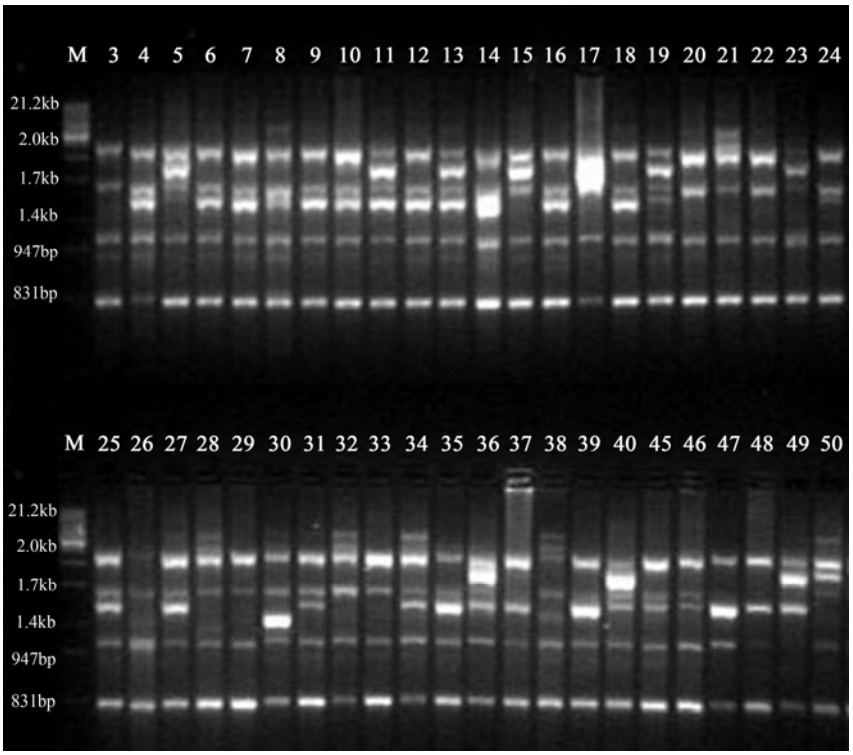


Fig. 2. A RAPD profile of 44 wild sybean (*G. soja*) accessions and soybean (*G. max*) cultivars. Primer OPB-01; M: λ DNA/*Eco*RI+*Hin*dIII Marker; 3,4…40,45…50: No. of the samples, which are the same as in Table1.

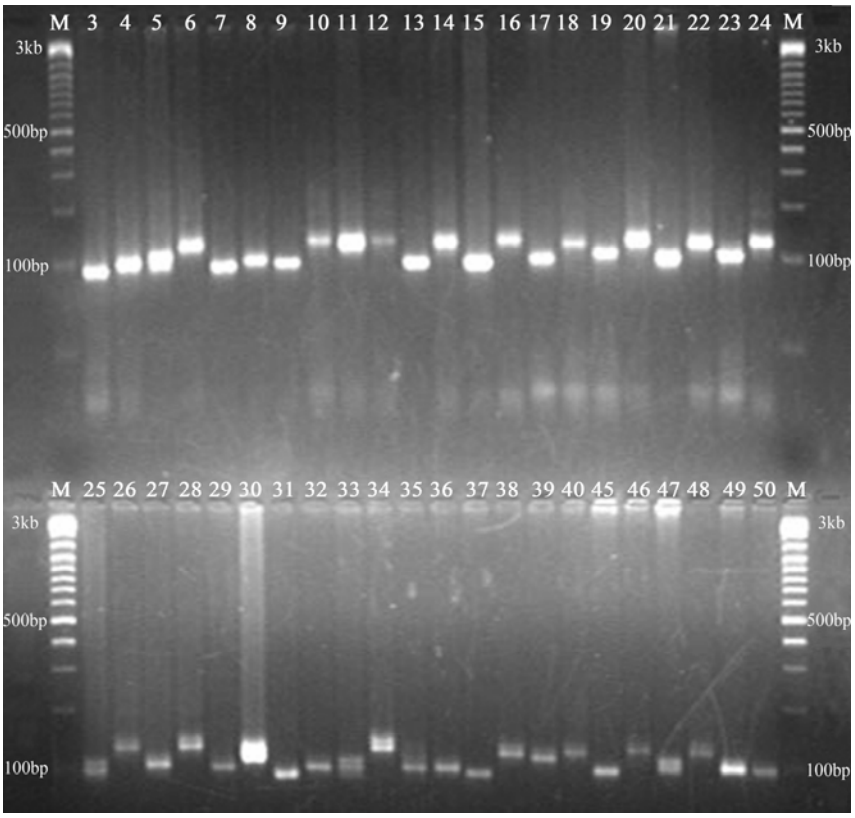


Fig. 3. A SSR profile of 44 wild soybean accessions and soybean cultivars. Primer satt77; M: 100bp ladder; 3,4…40,45…50 : No. of the samples, the same of Table1.

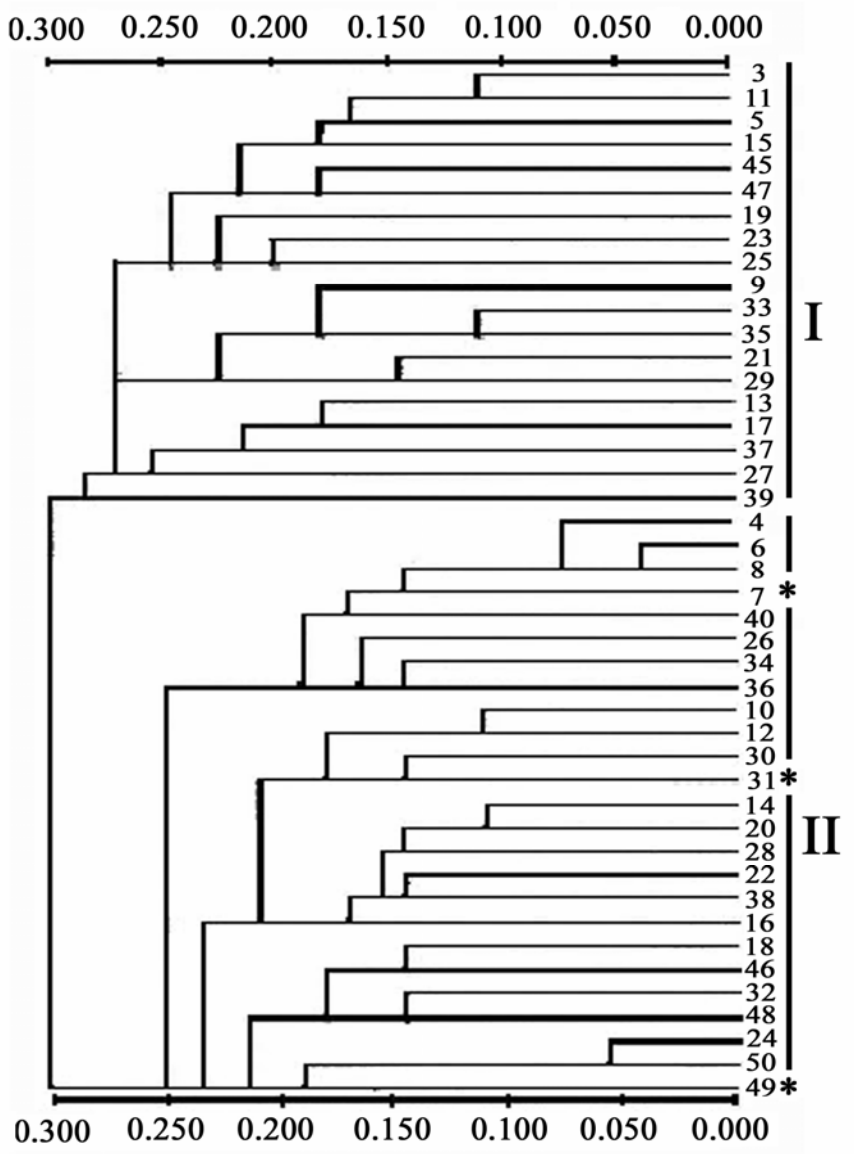


Fig. 4. A dendrogram illustrating the relationships among the of 44 wild soybean accessions and soybean cultivars based on SSR analysis. Beside "7" (ZYD 00152), "31" (ZYD 03564) and "49" (ZYD 05173) (annotated with "*"), the other samples were clustered into groups corresponding to wild or cultivated soybeans, respectively.

**Table 3. Summary of analysis of molecular variance (AMOV) based on the scored AFLP, RAPD and SSR bands.**

| Marker | Level of variation | SS | V.P. | % | P* |
|---|---|---|---|---|---|
| RAPD | Between wild and cultivated soybean accessions | 37.432 | 1.24292 | 10.97 | <0.001 |
| | Within accessions | 423.682 | 10.08766 | 89.03 | <0.001 |
| SSR | Between wild and cultivated soybean accessions | 42.705 | 1.63912 | 19.79 | <0.001 |
| | Within accessions | 279.045 | 6.64394 | 80.21 | <0.001 |
| AFLP | Between wild and cultivated soybean accessions | 132.100 | 9.68667 | 21.56 | <0.001 |
| | Within accessions | 634.200 | 35.2333 | 78.44 | <0.001 |

Levels of significance test were based on 1023 permutations (SS: sums of squares, V.P.: variance components, %: proportion of genetic variability, P: level of significance)

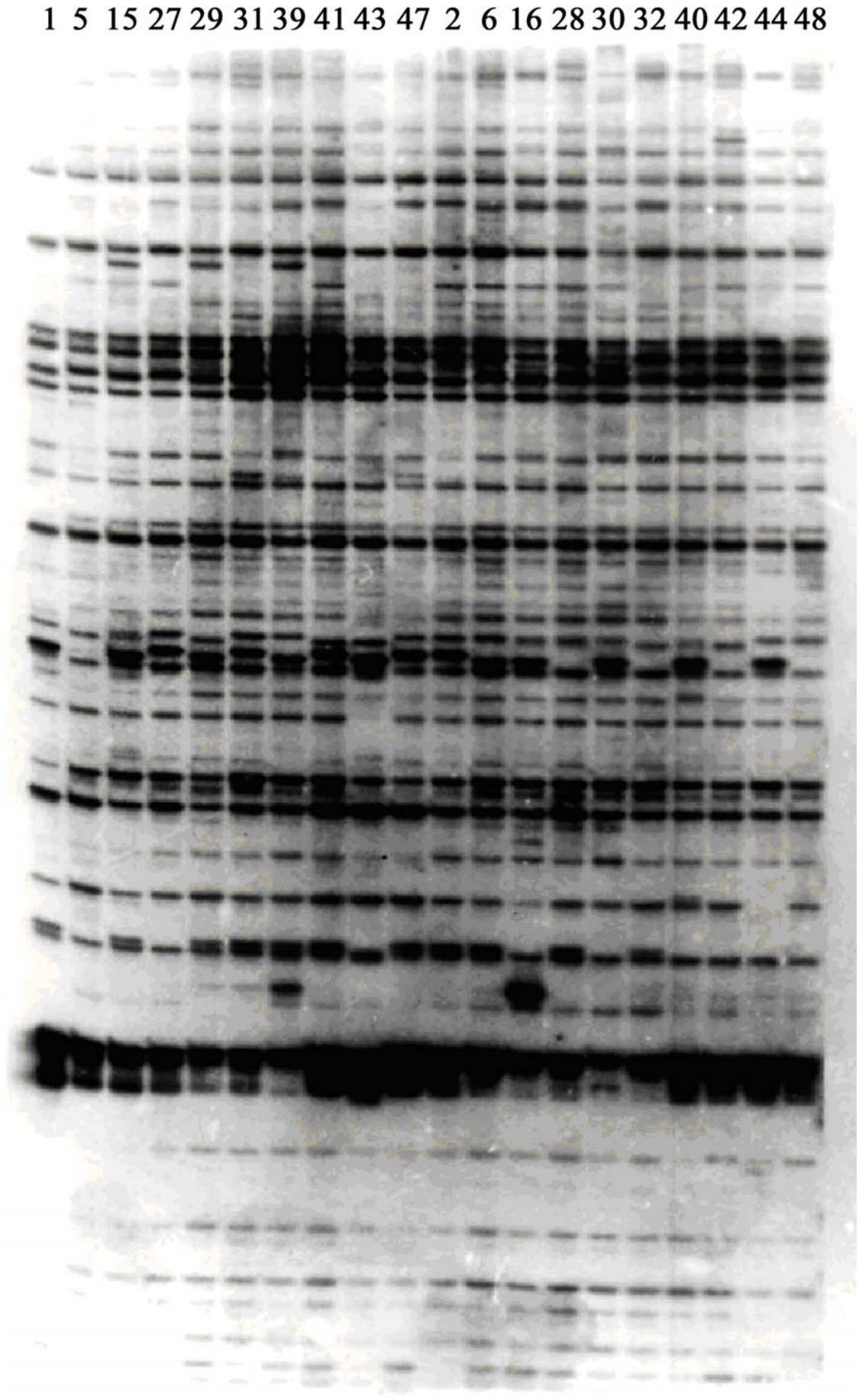


Fig. 5. An example of AFLP profile of 20 wild soybean accessions and soybean cultivars.

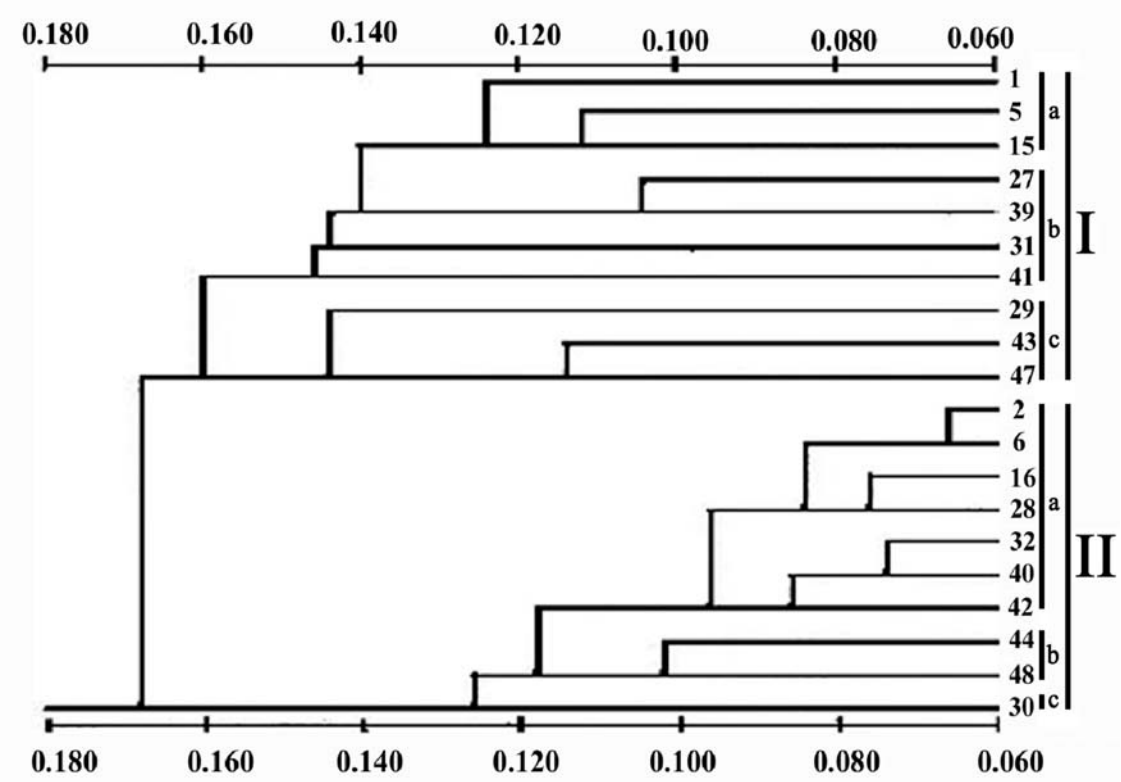


Fig. 6. A dendrogram, built according to the AFLP data, illustrating the relationships among the 20 wild soybean accessions and soybean cultivars.

## Discussion

**Genetic diversity in annual wild soybean and cultivated soybean:** This paper has evaluated the usefulness of three molecular markers, RAPD, SSR and AFLP, for revealing the level of genetic diversity, the distribution of genetic variation, and the genetic relatedness among annual wild soybean accessions and among soybean cultivars collected across China. All three markers revealed similar and consistant results, which indicated that wild soybean possessed a higher level of genetic diversity than cultivated soybean. Based on the constructed dendrograms, the wild soybean could be divided into more subgroups than the cultivated soybean at a given GD level. For example, in SSR analysis, the wild soybean could be divided into 3 groups at a GD level of 0.250, whereby all cultivated soybeans remained within a single group (Fig. 4). AFLP analysis suggested the same conclusion in that, at a GD level of 0.160, whereas 2 subgroups were apparent for the wild soybean accessions the soybean cultivars formed only one group. The distinct seperation of annual wild soybean and cultivated soybean was further subtantiated by the AMOVA results.

The results of this paper is consistant with previous genomic studies in soybean showing that during domestication several severe bottlenecks were experienced and as a result, many or even most of the rare alleles in wild soybean were lost from modern soybean cultivars (Hyten *et al*., 2006). This information has provided a solid molecular genetic basis for the potentiality of using annual wild soybean for soybean improvement.

Distribution of genetic variation and origin of soybean in China: Althouth it has been established that soybean was directly domesticated from annual wild soybean in China, the exact locality of domestication remains to be determined (Dong *et al*., 2001, 2004). Nevertheless, four candidate regions were proposed with regard to the probable

geographical origins of soybean in China, viz., the Northeast (Fukuda, 1933), the lager North area (Hymowitz, 1970), the Yangtze River Valley and southern-wards (Wang, 1985), the middle of the Yellow River Valley and downstream thereof (Xu, 1986), and the multi-center theory (Lu, 1978). Logically, the geographical distribution of the accessions and the genetic diversity data should be able to provide the most deterministic information on the origin of cultivated soybean. Results of this study showed that the GD among different soybean accessions or cultivars was associated with latitude and the AGD of wild soybean was greater than that of cultivated soybean within the same range of latitude. Based on both SSR and AFLP analysis, the AGD of lower latitudes accessions was greater than that of higher latitudes, that is, increasing latitude appeared to translate to lower GD between different soybean lines. This suggests that the southern areas in China should be most promising as likely centers of soybean origin, a view also consistent with Vavilov (1973) classic theory on genetic diversity and species origin, which states that an organism species has a specific genetic diversity center that was the original center of distribution; the species subsequently dispersed from this center to other larger areas and the geographical distribution was successive.

Our data also showed that latitude alone may not be the only or even the major factor influencing genetic variations. Two special cases were revealed in the AFLP analysis: ZYD03564 and ZYD03328A were categorized into two different subgroups even though they were from the same latitude, but from different longitudes. This was also true for Zhengzhou Dalihuang and Huangdou. The grouping of ZYD03328 with ZYD05191 and ZYD05280, though they were from different latitudes, may have been strongly influenced by similar longitudinal origins and high elevation. Thus, genetic differentiation in soybean is not only associated with the latitude but also with specific climatic factors associated with elevation.

Taxonomical status of annual wild soybean and cultivated soybean: Although annual wild soybean and cultivated soybean have been considered, from a technical perspective, as two species, *G. soja* and *G. max*, the biological validity of this classification has been questioned. This is largely because the two forms, regardless of their morphological distinctness, are still sexually compatible, and hence, incongruent with the biological species concept. Thus, an alternative opinion holds that both forms should be considered as one biological species which nonetheless can be divided into three types i.e., wild, semi-wild and cultivated. Indeed, some molecular polygenetic studies are actually in support of this opinion. For example, Hui *et al*., (1996) studied the ITS-I sequences of 24 accessions of the *Glycine* subgenus including both annual wild and cultivated soybeans and found no divergence between the two forms. On the other hand, the opinion insisting that the two forms should be treated as two species argues that the distinctness in both morphological and physiological traits of the two forms are to be intraspecific. This latter view is especially held by soybean breeders and genetists. Indeed, as demonstrated by the present study, lineage-specific genetic differences are abounded between the two forms at the DNA level. In addition, the general genetic differentiation is well supported by the AMOVA based on on all three makers. Therefore, like many other crops e.g., rice and wheat, domestication has clearly shaped the cultivated soyben to an extent that enabled its significant divergence from its direct wild progenitor, annual wild soybean; as such, a distinct species status should be granted. Future large-scale genomic studies between the two forms will undoubtedly shed new light into the issue.

**Supplementary Table 1. Plant materials used in this study.**

| Sample No. | Code in seed pool of local name | Latitude (°N) | Longitude (°E) | Region in China | W/C | Sample No. | Code in seed pool of local name | Latitude (°N) | Longitude (°E) | Region in China | W/C |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1[A] | ZYD00004 | 52 | 127 | | W* | 25[S、R] | ZYD02819 | 39 | 112 | | W* |
| 2[A] 3[S、R] | Tiejiaqing | 52 | 127 | HEI**-Huma | C* | 26[S、R] | Huangdou | 39 | 112 | JIN**-Youyu | C* |
| 4[S、R] | YD00088 | 50 | 127 | | W* | 27[A、S、R] | ZYD02903 | 38 | 111 | | W* |
| 5[A S、R] | Aihuibendizhong | 50 | 127 | HEI**-Aihui | C* W* | 28[A、S、R] | Baiyuandaheidou | 38 | 111 | JIN**-Linxian | C* |
| 6[A S、R] | ZYD00069 | 49 | 128 | | C* | 29[A、S、R] | ZYD03328 | 35 | 106 | | W* |
| 7[S、R] | Xunkebayuemang | 49 | 128 | HEI**-Xunke | W* | 30[A、S、R] | Huangdou | 35 | 106 | GAN**-Qingshui | C* |
| 8[S、R] | ZYD00152 | 48 | 126 | | C* | 31[A、S、R] | ZYD03564 | 35 | 114 | | W* |
| 9[S、R] | Kesanhuigan | 48 | 126 | HEI**-Kesan | W* C* | 32[A、S、R] | Zhengzhoudalihuang | 35 | 114 | YU**-Zhengzhou | C* |
| 10[S、R] | ZYD00471 | 46 | 129 | | W* | 33[S、R] | ZYD03485 | 34 | 113 | | W* |
| 11[S、R] | Duludou | 46 | 129 | HEI**-tonghe | C* | 34[S、R] | Baofengyangyandou | 34 | 113 | YU**-Baofeng | C* |
| 12[S、R] | ZYD00235 | 47 | 127 | | W* | 35[S、R] | ZYD03510 | 33 | 114 | | W* |
| 13[S、R] | Tiecuozi | 47 | 127 | HEI**-shuiling | C* | 36[S、R] | Miyangmiumaohuang | 33 | 114 | YU**-Miyang | C* |
| 14[S、R] | ZYD00810 | 45 | 126 | | W* | 37[S、R] | ZYD04361 | 30 | 109 | | W* |
| 15[A、S、R] | Silihuang | 45 | 126 | JI**-Yushu | C* W* | 38[S、R] | Enshihuangkedou | 30 | 109 | E**-Enshi | C* |
| 16[A、S、R] | ZYD00968 | 44 | 120 | | C* | 39[A、S、R] | ZYD04550 | 30 | 114 | | W* |
| 17[S、R] | Xiaojinhuang | 44 | 120 | JI**-Gongzhuling | W* | 40[A、S、R] | Kaisanbai | 30 | 114 | ZHE**-Shangyu | C* |
| 18[S、R] | ZYD01618 | 43 | 124 | | C* | 41[A、、] | ZYD4672 | 29 | 113 | | W* |
| 19[S、R] | Tianedan | 43 | 124 | LIAO**-Changtu | W* C* | 42[A、、] | Yueyangbayuebao | 29 | 113 | XIANG**-Yueyang | C* |
| 20[S、R] | ZYD01915 | 42 | 123 | | W* | 43[A、] | ZYD05280 | 25 | 107 | | W* |
| 21[S、R] | Youhulu | 42 | 123 | LIAO**-Xinmin | C* | 44[A] | Nandanerzaohuangdou | 25 | 107 | GUI**-Nandan | C* |
| 22[S、R] | ZYD02540 | 41 | 119 | | | 45[S、R] | ZYD04958 | 26 | 117 | | W* |
| 23[S、R] | Tiejia | 41 | 119 | LIAO**-Kazuo | | 46[S、R] | Tainingqingpi | 26 | 117 | MIN**-Taining | C* |
| 24[S、R] | ZYD02726 | 40 | 116 | | | 47[A、S、R] | ZYD05191 | 24 | 113 | | W* |
| | Changpingqidou | 40 | 116 | Beijing | | 48[A、S、R] | Yingdezhongqing | 24 | 113 | YUE**-Yingde | C* |
| | | | | | | 49[S、R] | ZYD05173 | 27 | 101 | | W* |
| | | | | | | 50[S、R] | Huangdou | 27 | 101 | DIAN**-Ningliang | C* |

* W: wild Soybean; C: cultivate Soybean

** HEI-Heilongjiang, JI-Jilin, LIAO-Liaoning, JIN-Shanxi, GAN-Ganshu, YU-Hhenan, E-Hubei, ZHE-Zhejiang, MIN-Jiangxi,YUE-Guangdong, DIAN-Yunna province.

A: Samples used in AFLP

S: Samples used in SSR

R: Samples used in RAPD

**Supplementary Table 2. Selected primers in the SSR analysis.**

| Locus | Core motif | Sense primer (5'-3') | Antisense primer (5'-3') |
|---|---|---|---|
| 1(SoyGPATR ) | $(CTT)_4CCCT(CTT)_7$ | GGA AGA AAG TAT TGG TCT GT | AGG AGA GAG TGG AGA GAT TA |
| 2(BARC-Sat22) | $(AT)_{26}$ | GCC TTT TCT GAC TGT TAA | CAG TGA CTA AAA CTT ACT AT |
| 3(BARC-Sat39) | $(AT)_{17}$ | CAA GAA TAA TCT AAA GGT ACA AC | AGT TAA AAA ACC CAC ACA AC |
| 4(BARC-Sat26) | $(AG)_{14}$ | CGA AAC GCA AAA TCT C | AAA ACG TAT CTG AAG TAG TGG |
| 5(BARC-Satt5) | $(TAA)_{21}$ | TAT CCT AGA GAA GAA CTA AAA AA | GTC GAT TAG GCT TGA AAT A |
| 6(BARC-Satt9) | $(AAT)_{14}$ | CCA ACT TGA AAT TAC TAG AGA AA | CTT ACT AGC GTA TTA ACC CTT |
| 7(BARC-Satt20) | $(ATT)_{16}$ | GAG AAA GAA ATG TGT TAG TGT AA | CTT TTC CTT CTT ATT CTT TGA |
| 8(BARC-Satt22) | $(TAA)_{22}$ | TGT ATT TTA CCT TAC CTT TGA | AAC TGC CAC CAA TGA C |
| 9(BARC-Satt30) | $(TAA)_{18}$ | AAA AAG TGT AAC CAA GCC | TCT TAA ATC TTA TGT TGA TGC |
| 10(BARC-Satt31) | $(ATT)_{12}$ | TTC CAC TTT GTA TCA CTT TC | TGA CTG TAA AAG AAC AGA TAA A |
| 11(BARC-Satt45) | $(ATT)_{17}$ | TGG TTT CTA CTT TCT ATA ATT ATT T | ATG CCT CTC CCT CCT |
| 12(BARC-Satt77) | $(ATT)_{13}$ | GAT CTA AAG TCT GAT ATT TTT AAC TA | AAA AGG AGA AGG GAG TTG AT |

## Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (30490250), the 973 project (2004CB117203-3), and by the Program for Changjiang Scholars and Innovative Research Team (PCSIRT) in University (#IRT0519).

### References

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

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

Dong, Y.S., B.C. Zhuang, L.M. Zhao, H. Sun and M.Y. He. 2001. The genetic diversity of annual wild soybeans grown in China. *Theor. Appl. Genet*., 103: 98-103.

Dong, Y.S., L.M. Zhao, B. Liu, Z.W. Wang, Z.Q. Jin and H. Sun. 2004. The genetic diversity of cultivated soybean grown in China. *Theor. Appl. Genet*., 108**:** 931-936.

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

Excoffier, L., G. Laval and S. Schneider. 2005. Arlequin ver.3.0: An intergrated software package for population genetics data analysis. *Evol. Bioinfo. Oneline*, 1: 47-50.

Fukuda, Y. 1933. Cytogenetical studies on the wild and cultivated Manchrian soybean (*Glycine* L.). *Jpn. J. Bot*., 6: 489-506.

Gillies, A.C.M., C. Navarro, A.J. Lowe, A.C. Newton, M. Hernandez, J. Wilson and J.P. Cornelius. 1999. Genetic diversity in Mesoamerican populations of mahogany (*Swietenia macrophylla*), assessed using RAPDs. *Heredity*, 83: 722-723.

Haley, S.D., L.K. Afanador, P.N. Miklas, J.R. Stavely and J.D. Kelly. 1994. Heterogeneous inbred populations are useful as source of near-isogenic lines for RAPD marker localization. *Theor. Appl. Genet*, 88: 337-342.

Hui, D.W., B.C. Zhuang and S.Y. Chen. 1996. Phylogeny of Genus *Glycine* Reconstructed by RAPD Fingerprinting. *Acta Genetic Sinica*, 23(6): 460-468.

Hymowitz, T. 1970. On the domestication of the soybean. *Econ Bot*., 23: 408-421.

Hyten, D.L., Q. Song, Y. Zhu, I.Y. Choi, R.L. Nelson, J.M. Costa, J.E. Specht, R.C. Shoemaker and P.B. Cregan. 2006. Impact of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci USA* .,103: 16666-16671.

Ilan Paran, K. Kichard, R. Michelmore.1991. Identification of RFLP and RAPD markers linked to downy mildew resistance gene in lettuce, using near-isogenic lines. *Genome,* 34: 1021-1026.

Jiang, RW., M.S. Akkaya, A.A. Bhagwat, U. Lavi and P.B. Cregan. 1995. The use of microsatellite DNA markers for soybean genotype identification. *Theor. Appl. Genet*, 90: 43-48.

Liu, F., S.Y. Chen and B.C. Zhuang. 1999. Application of SSR in soybean genetic mapping. *High Technique Communication*, 6: 8-11.

Lu, S.L. 1978. Discuss about origin of cultivated soybean in China. *Chin Agricu Sci*., 4: 90-94.

Lynch, M and B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. *Mol. Ecol*., 3: 91-99.

Maughan, P.J, M.A. Saghai-Maroof and G.R. Buss. 1995. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. *Genome*, 38(4): 715-723.

Maughan, P.J., M.A. Saghai Maroof, G.R. Buss and G.M. Hueskis. 1996. Amplified fragment length polymorphism(AFLP)in Soybean: Species diversity, inheritance and near-isogenic line analysis, *Theor. Appl. Genetics*, 93(3): 392-401.

Mohan, M., S. Nair, J.S. Bentur, U.P. Rao and J. Bennett. 1994. RFLP and RAPD mapping of the rice Gm2 gene that confers resistance to biotype 1 of gall midge (Orseolia oryzae). *Theor. Appl. Genet*., 87: 782-788.

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

Nei, M and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonuclease. *Proc. Natl. Acad. Sci. USA*, 74: 5267-5273.

Nei, M. 1972. Genetic distance between populations. *Am Nat*., 106: 283-292.

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

Powell, W., M. Morgante, J.J. Doyle, J.W. McNicol, S.V. Tingey and A.J. Rafalske. 1996. Gene pool variation in genus *Glycine* subgenus soja revealed by polymorphic nuclear and chloroplast micro-satellites. *Genetics*, 144: 793-803.

Qian, H.R. and K.L. Zheng. 1998. DNA markers and molecular breeding. *China Journal of Progress in Bioengineer*, 18(3): 12-18.

Stiles, J.I., G. Lemme, S. Sondur, M.B. Morshidi and B. Manshardt. 1993. Using RAPD for evaluating genetic relationships among papaya cultivars. *Theor. Appl. Genet*., 85: 697-701.

Storfar, A. 1996. Quantitative genetics: a promising approach for the assessment of genetic variation in endangered species. *Trends Ecol. Evol*., 11: 343-348.

Vavilov, N.I. 1973. *The origin, variation, immunity and breeding of cultivated plants*, Vol. 13. Ronald Press, New York.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van De Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nueleic Acids Research*, 23: 4407-4414.

Wang, L.Z. 1985. Origin and diffusion of soybean. *Soybean Sci*., 4: 1-7.

Welsh, J and M. McClelland. 1991. Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers. *Nucleic Acids Research*, 19(2): 303-306.

Williams, J.G.K. and A.R. Kubelik. 1990. DNA polymorphisms amplified by arbitary primers are useful as genetic markers. *Nueleic Acids Research*, 21(11): 2697-2702.

Xu, B. 1986. Three New evidences for origin of soybean. *Soybean Sci*., 5: 123-130.

Yi, H., W.M. Yuan and M. Gu. 1997. Total DNA extraction from different tissues of soybean. *China Journal of Biochemistry*, 13(6): 719-720.

Yu and Maroof. 1994. RFLP and Microsatellite mapping of a gene for soybean mosaic virus resistance. *Genetics*, 84: 59-64.

Zhang Erquan. 1998. Gene structure and molecular evolution structure of rbcS gene of *G. Gracilis*. *Journal of Fudan University* 37(2): 151-157.

Zhuang, B.C., D.W. Hui, Y.M. Wang, J. Gu, B. Xu and S.Y. Chen. 1994. RAPD analysis of soybean from different latitude in China. *Chines Science Bulletin*, 39(23): 2118-2180.

(Received for publication 18 October 2008)