

Y.S. Dong · B.C. Zhuang · L.M. Zhao · H. Sun · M.Y. He

The genetic diversity of annual wild soybeans grown in China

Received: 25 June 2000 / Accepted: 18 October 2000

Abstract Annual wild soybeans (*Glycine soja*), the ancestors of cultivated soybeans (*G. max*), are important sources of major genes for resistance to pests, diseases and environmental stresses. The study of their genetic diversity is invaluable for efficient utilization, conservation and management of germplasm collections. In this paper, the number of accessions, the variation of traits, the genetic diversity indexes (Shannon index) and the coefficient of variation were employed to study the geographical distribution of accessions, genetic diversity of characters and genetic diversity centers of annual wild soybean by statistical analysis of the database from the National Germplasm Evaluation Program of China. Most annual wild soybeans are distributed in Northeast China, and the number of accessions decreases from the Northeast to other directions in China. The genetic diversity indexes (Shannon index) were 0.49, 0.74, 0.02, 0.55, 1.45, 2.41, 1.27 and 1.89 for flower color, sootiness of seed coat, cotyledon color, pubescence color, hilum color, leaf shape, stem type and seed color, respectively. Coefficients of variation were 7.1%, 28.7%, 76.43% and 18.2% for protein content, oil content, 100-seed weight and days to maturity, respectively. Three genetic diversity centers, the Northeast, the Yellow River Valley and the Southeast Coasts of China, are proposed based on the geographical distribution of the number of accessions, genetic diversity and the multivariate variation coefficient. Based on these results and Vavilov's theory of

crop origination, two opposing possible models for the formation of the three centers are proposed, either these centers are independent of each other and the annual wild soybeans in these centers originated separately, or the Northeast center was the primary center for annual wild soybeans in China, while the Yellow River Valley center was derived from this primary center and served as the origin for the Southeast Coast center.

Keywords Annual wild soybean · Genetic diversity · Origin · Germplasm · Diversity center

Introduction

Annual wild soybeans (*Glycine soja*) are distributed throughout East Asia, including Korea, Japan and the far eastern regions of Russia, and are centered in China (Li 1994). The geographical distribution of wild soybeans in China is very wide. The horizontal distribution is from Tahe County, Heilongjiang (latitude: 53°) in the north to Xiangzhou County, Guangxi (latitude: 24°) in the south and from Raohe County, Heilongjiang (longitude: 134°), in the east to Chayu County, Tibet (longitude: 97°), in the west. The vertical distribution is from 1.8 to 2650 meters above sea level. The soybean species grow in very different soils and under diverse climate conditions. Such a broad geographical distribution and diverse growing conditions contribute to the plentiful genetic diversity of annual wild soybeans.

Studies on the genetic diversity of annual wild soybean have been limited by materials. Most investigations have focused on isoenzyme polymorphism analysis using a small set of annual wild soybean germplasms. Chiang (1986) studied 72 accessions of *G. soja* from Japan and the Republic of Korea for 43 isoenzyme loci analysis using gel electrophoresis, and the genetic variation among those accessions showed an average polymorphism of 67.4%. Four *G. soja* populations collected in Japan showed 38.1% polymorphism loci and an average of 1.55 alleles per locus; of the total genetic diversity

Communicated by H.C. Becker

Y.S. Dong (✉) · B.C. Zhuang · L.M. Zhao · H. Sun
Jilin Academy of Agricultural Science,
No. 6, Xixinghua Street, Gongzhuling City,
136100 Jilin Province, Peoples Republic of China
e-mail: dongysh@public.jaas.sp.jl.cn
Fax: +86-434-6237387

Y.S. Dong · M.Y. He
Northeast Normal University, Changchun 130024, China

Present address:

Y.S. Dong, Plant Science Department,
The Weizmann Institute of Science, Rehovot 76100, Israel

of the four populations, which were well-differentiated by isoenzyme analysis, 19.8% was interpopulational and 81.2% intrapopulational. Gorman (1984) examined 400 cultivars of *G. max* and more than 100 accessions of *G. soja* of various geographical distributions with an analysis of 15 enzymes and concluded that Northeast China had slightly higher levels of diversity for *G. max*, while Korea had significantly higher diversity levels for *G. soja*. Further studies showed that domesticated soybeans have a reduced genetic diversity, a changed distribution of alleles and, in many cases, a change in allele frequencies. Abe et al. (1992) demonstrated that Korea might be a center of genetic diversity for annual wild soybean based on an analysis of nine enzymes of annual wild soybeans that 383 accessions from Japan and 28 accessions from Korea. Yu and Kiang (1993) checked 17 enzymes and 1 protein from annual wild soybeans that came from six natural populations. The results showed a surprisingly high variation in the South Korean wild soybean as well as the cultivated soybean indicating that South Korea is one of the major soybean gene centers. Kiang et al. (1992) researched four natural populations of *G. soja* with an analysis of 15 enzymes plus one non-enzyme protein and showed that genetic diversity was 0.07 and 0.3 between- and within-populations, respectively, and that the coefficient of differentiation among populations was 0.2. Vaughan et al. (1995) analyzed ten annual wild soybean populations, collected within a 20 km radius of Tsukuba, using randomly amplified polymorphic DNA (RAPD), and the results indicated a relatively high degree of local variation for *G. soja* on the Kanto plain and a mosaic of variation rather than small-scale clonal variation. Xu (1995) assumed that the Northeast and Yellow River Valley of China were the two soybean genetic diversity centers by analyzing the geographical distribution of seed characters for annual wild soybean in China.

Comprehensive collecting of annual wild soybean germplasms was initiated in the late 1970s in China. More than 6000 accessions have been collected, characterized, evaluated and documented with support from the National Soybean Germplasm Evaluation Project of China. A database was established based on this project.

In order to understand overall genetic diversity, distribution of genetic diversity and formation modes of ge-

netic diversity centers, we analyzed the data from the National Annual Wild Soybean Database using several parameters. The hypotheses proposed here will supply a reference point for other studies on genetic diversity of annual wild soybean in China and in the world as well as for a better utilization, conservation and management of annual wild soybean.

Materials and methods

The data used in this study came from the National Soybean Germplasm Database, which was established by characterizing and evaluating 6,172 accessions of annual wild soybean collected from 25 provinces in China. The accessions that were different in same ecological region were collected, characterized and evaluated in different ecological regions of China. The data were generated by characterizing and evaluating the accessions using two 2-m-long rows in the original ecological regions of the accessions. Several groups in China have conducted this work for more than 20 years. Twelve traits, including eight qualitative traits pubescence color, flower color, leaf shape, stem type, bloom color, cotyledon color, hilum color, seed color and four quantitative traits 100-seed weight, days to maturity, protein content and oil content were used to calculate genetic diversity index and synthetic variation coefficient. The distribution areas of annual wild soybean in China were divided into 79 sections based on latitude and longitude. Each section covered two degrees of latitude and three degrees of longitude. The accessions that have the same quality characters were defined as a type, and the 379 types were separated using eight 8 quality characters. The quality characters were quantified (Table 1). The genetic diversity index of types and synthetic variation coefficient of 12 traits were calculated in every section using the following two formulas with DPS software (Data processing system, Tang 1997):

1) The Shannon genetic diversity index H' (Tang 1997):

$$H' = -\sum_{i=1}^S P_i \ln(P_i)$$

where S is the number of types in a section, i is the i th type in a section and P is the proportion of the i th type in a section.

2) Synthetic variation coefficient CV (Dong 1998):

$$CV\% = \frac{1}{m} \sum_{j=1}^m \frac{S_j}{\bar{X}} \times 100$$

where m equals the number of traits in a section, S_j equals the standard error of j th traits in a section and \bar{X} is the mean of j th traits in a section.

Table 1 Quantification of the qualitative traits

Traits	Quantification					
Flower color	White=1	Purple=2				
Seed sootiness	Smooth=1	No=2	Have=3			
Cotyledon color	Yellow=1	Green=2				
Seed color	Yellow=1	Green=2	Brown=3	Black=4	Double color=5	
Pubescence color	Gray=1	Brown=2				
Hilum color	Brown=1	Black=2				
Leaf shape	Line=1	Lanceolate=2	Elipse=3	Round=4	Ovate round=5	Other=6
Stem type	Not obvious=1	Half-obvious=2	Obvious=3			

Results

Geographical distribution of annual wild soybean in China

The distribution of annual wild soybeans is very broad but not even in China. This distribution is also affected by landscape, soil and climate conditions. A line could be roughly drawn from the Daxinganling Mountains in the northeast to Qingzang Plateau in the southwest of China based on the number of accessions and the location where the collections came from. Most accessions were collected from the eastern part of the line. The 6172 accessions were collected over a wide range of altitude and longitude from Tahe County, Heilongjiang to Xiangzhou County, Guangxi and from Raohe County, Heilongjiang to Chayu County, Xizang (Fig. 1). A large portion of the accessions were distributed throughout the Northeast and centered in seven sections covering latitude 40°–47°N and longitude 119°–130°E. There were 2555 accessions that made up 41.4% of the total in those sections. The next area with a high percentage of collections was located in the Yellow River Valley, covering latitude 36°–40°N and longitude 107°–115°E, where 896 accessions make up 14.5% of the total, concentrated in five sections. The third area with a relatively high number of 228 accessions appeared in the section of 26°–27°N and longitude 116°–118°E.

The vertical distribution of the wild soybeans varied dramatically with a major distribution at lower altitudes. The highest latitude where the wild soybeans were found was at about 1300 m above sea level in Northeast China, 1500–1700 m in the Valleys of the Yellow River and Yangtse River and 2650 m in Ninglang County, Yunnan province (Li 1994).

The genetic diversity of annual wild soybeans

Annual wild soybean genetic diversity was detected by analyzing variation in eight qualitative traits and four quantitative traits. The results are shown in Table 2. Generally, qualitative traits revealed less genetic diversity than quantitative traits. The Shannon indexes for leaf shape, seed coat color, hilum color, stem type, seed sootiness, pubescence color, flower color and cotyledon color were 2.41, 1.89, 1.45, 1.27, 0.74, 0.55, 0.49 and 0.02, respectively. Leaf shape had the highest variation among the eight qualitative traits followed by seed coat color, hilum color, stem type, sootiness, pubescence color, flower color and cotyledon color. The percentage of accessions with round-ovate, ellipse, lanceolate, long-oblong and long-ovate leaves was 31%, 27%, 19%, 12% and 8% respectively; others accounted for 3%. Most accessions had a black seed coat (66%), and quite a few of them had bicolor (18%), brown (4%), green (2%) and other colors (10%). The main hilum colors were black (55%) and brown (39%), and a small number of accessions had a yellow (2%) and black brown (2%) hilum, others 2%. The typical phenotype of the stem was twining. Three different kinds of stem types could be

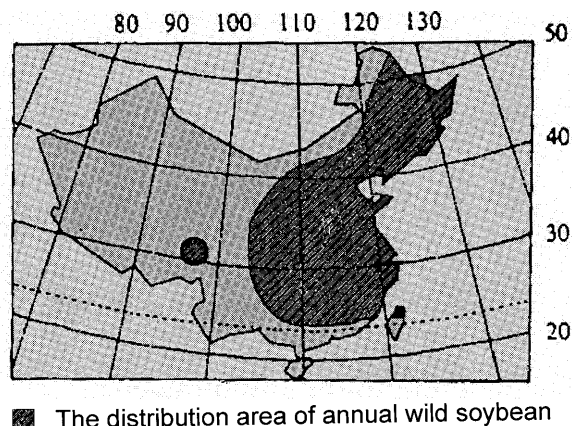


Fig. 1 The geographical distribution of *G. soja* in China

Table 2 The genetic diversity indexes of annual wild soybeans in China

Characters	Number of accessions	Genetic diversity indexes (Shannon)
Pubescence color	6171	0.55
Flower color	6172	0.49
Leaf shape	6169	2.41
Stem type	6171	1.27
Seed sootiness	6171	0.74
Cotyledon color	6172	0.02
Hilum color	6180	1.45
Seed coat color	5969	1.89
100-seed weight	5585	3.68
Maturity	5905	3.42
Protein content	6180	2.15
Oil content	5292	3.23

classified according to the existence of a main stem: 63% of the accessions had an obvious main stem, 28% had no main stem and 9% fell in between. The sootiness of the seed coat is a character special to wild soybean. The percentage of sooty, non-sooty and brilliant seed coat was 84%, 13% and 3%, respectively. Compared to the above-mentioned characters less variation was found for flower color and pubescence color. The purple flower (91%) and brown pubescence (90%) were the main colors of these two characters. However, the diversity for cotyledon color was a unique case in this study. All accessions had yellow cotyledons, with no exceptions.

The four quantitative traits revealed a greater genetic diversity. The Shannon indexes for 100-seed weight, days to maturity, oil content and protein content were 3.68, 3.42, 3.23 and 2.15, respectively. With respect to 100-seed weight 81% of the accessions fell in the range 0.5–2.5 g and 72.2% fell into the range 1.0–2.0 g. Those accessions with a 100-seed weight in the range 2.51–5.0 g were 10.4% of the total. There were 8.5% of the accessions with 100-seed weight heavier than 5.0 g. The maturity of wild soybean was quite varied: 48.2% of the accessions matured within 120–140 days; a few accessions matured in fewer than 80

days (0.1%) or more than 200 days (1.6%). The oil content of wild soybean was low, 10.5% on average. The percentage of oil content was mostly in the range 7.0–10.9%, which accounted for 62.1% of the total accessions. The lowest oil content was 4.4%. In contrast with oil content the protein content of wild soybean was quite high. The mean value of protein content was 45.4%, with a high of 55.7% and a low of 29%. The accessions with a protein content lower than 36% and higher than 52% accounted for 0.6% and 0.7% of the total, respectively.

Geographical distribution of genetic diversity

The genetic diversity index (Shannon index) with eight qualitative characters of wild soybean accessions collected from each of the 79 sections described above was calculated using the Shannon formula. The value of the Shannon index in each section was different and varied considerably from one to the other (Table 3). The synthetic variation coefficient of 12 characters, including the 8 qualitative characters and four quantitative characters, was also calculated to demonstrate the genetic diversity of the accessions in each section (Table 4). When the

Table 3 Geographical distribution of the diversity index (Shannon index)

Latitude (°N)	Longitude (°E)												
	97	100–103	104–106	107–109	110–112	113–115	116–118	119–121	122–124	125–127	128–130	131–133	134–136
52–53										0.51			
51–50										0.80			
49–48									3.03	1.66	1.80	2.41	
47–46									3.40	4.44	2.57	4.17	2.28
45–44								0.00	2.52	4.39	4.00	2.97	
43–42								2.27	4.95	3.03	2.91		
41–40				1.00	3.21	2.56	3.05	3.52	2.69	2.57			
39–38			0.34	3.20	2.73	4.91	2.28	3.32	2.36				
37–36			2.32	2.64	3.08	3.20	0.25	1.61	0.81				
35–34			2.33	4.77	3.47	4.50	2.79	3.10					
33–32		1.58	2.30	3.00	2.85	2.67	4.10	4.00					
31–30		2.52		3.48	2.35	2.22	2.85	2.87	0.00				
29–28	0.00			3.17	1.62	1.74	2.27	2.47	0.00				
27–26		0.91		2.94	1.86	1.68	2.59	3.22					
25–24				2.50	2.56	0.72	2.13	1.00					

■ Shading indicates those sections with a higher genetic diversity

Table 4 Geographical distribution of the synthetic variation coefficient (%)

Latitude (°N)	Longitude (°E)												
	97	100–103	104–106	107–109	110–112	113–115	116–118	119–121	122–124	125–127	128–130	131–133	134–136
52–53										10.8			
51–50										10.2			
49–48									22.7	15.1	17.9	23.7	
47–46									17.8	31.1	23.1	23.8	10.1
45–44								5.0	28.2	34.2	28.9	19.2	
43–42								16.0	30.8	21.9	26.9		
41–40				10.3	21.6	20.0	29.2	22.0	20.6	19.8			
39–38			10.2	19.6	16.6	29.4	13.9	4.8	22.2				
37–36			17.6	9.7	16.8	20.1	18.2	9.3	4.6				
35–34			12.5	28.4	18.8	27.3	21.4	24.2					
33–32		10.5	19.4	21.9	29.4	23.8	22.2	24.0					
31–30		22.4		21.6	9.2	12.1	15.7	19.5	4.4				
29–28	3.0			22.1	8.0	16.6	15.0	17.7	0.0				
27–26		22.3		22.3	16.1	15.0	13.0	29.2					
25–24				7.5	13.4	7.5	7.5	11.0					

■ Shading indicates those sections with a higher synthetic variation coefficient

two parameters, the Shannon index and the synthetic variation coefficient, were combined two areas where the wild soybean accessions exhibited higher genetic diversity were found in China. The first region was located in 42° – 47° N \times 122° – 133° E. Shannon index values of 4.00–4.95 appeared in five sections and synthetic variation coefficients of 28.2–34.2 in five sections of the region. The second region was located in 34° – 38° N \times 107° – 115° E. Shannon index values of 4.50–4.91 and synthetic variation coefficient values of 27.3–29.4 appeared in three sections of the region. The accessions collected from South and Southeast China lacked genetic diversity. However, a relative high Shannon index value of 3.22 and synthetic variation coefficient value of 29.2 was found in one section located 26° – 27° N \times 119° – 121° E.

Discussion

The annual wild soybean is a self-pollinated plant and spreads very widely in China. Most plants grow in East China where it is warm and moist. The available moisture of the growing environment is very important for annual wild soybean (Li 1994). There is a clear line where annual wild soybean is located from the Northeast Daxinganling Mountains to the Qingzang Plateau in the south of China (Fig. 1). Almost all of annual wild soybean grows east of this line and the most of the plants are centered in three centers, only a few of which are located west of the line.

The data employed in this research were obtained from the National Database of wild soybeans, including all the 6,172 accessions collected in China. The genetic diversity of different traits and the geographical distribution of genetic diversity were various. Genetic diversity varying from large to little was found for 100-seed weight, days to maturity, oil content, leaf shape, protein content, seed color, hilum color, stem type, bloom color, pubescence color, flower color and cotyledon. The geographical distributions of accessions, the genetic diversity indexes, which were calculated using the Shannon formula, and synthetic variation coefficient of 12 traits, were analyzed. The results showed that there was abundant genetic diversity and a large number of accessions of annual wild soybean in Northeast China, especially in sections 42° – 43° N \times 122° – 124° E and 44° – 47° N \times 125° – 127° E. The highest number of accessions, genetic diversity indexes and variation coefficients in these sections where a distinct genetic diversity center was formed. The sections where the number of accessions, genetic diversity indexes and synthetic variation coefficients were higher in 34° – 39° N \times 107° – 115° E were located in 38° – 39° N \times 113° – 115° E, 34° – 35° N \times 113° – 115° E and 34° – 35° N \times 107° – 109° E, where another genetic diversity center might be formed. The sections 26° – 27° N \times 119° – 121° E, in which the genetic diversity indexes and synthetic variation coefficients were higher but the number of accessions was not, may define the third genetic diversity center. The three centers of genetic diversity for annual wild

soybean in China are proposed Vavilov (1973) base on theory.

Several hypotheses have been proposed on the origin of cultivated soybeans: Northeast (Fukuda 1933), North (Hymowitz 1970), the Yangtze River Valley and its South (Wang 1973), middle and down the Yellow River Valley (Wang 1985, Xu 1986, 1993) and a multi-center theory (Lu 1978, Dong 1998). Evidence for the cultivated soybean origin included cultivated history, literature records, archaeological data, geographical distribution of accessions and genetic diversity for wild soybean and cultivated soybean (Zhou 1998). In comparison, a similar amount of evidence has not been previously available for annual wild soybean.

The geographical distribution of accessions and genetic diversity were the most important evidence for the origin of annual wild soybean. That three centers of genetic diversity for annual wild soybean exist in China can be proposed from our results. The center in Northeast China was genetically diverse and contained abundant germplasm. The Yellow River Valley and Southeast coast were the two other diversity centers. The distributions of genetic diversity and accessions were continuous. The genetic diversity from high to low was Northeast, Yellow River Valley and Southeast coast. Two opposite possible formation modes of genetic diversity centers are proposed on the basis of this study and Vavilov's theory. Vavilov (1973) concluded that each plant has a special genetic diversity center, which was the original center of this plant; this center spreads out to other larger primary areas, and the geographical distribution is successive. The two formation modes are as follows. (1) the center in the Northeast was the primary original center where annual wild soybeans spread out in other directions. To the East, Southeast and North, this primary center helped the distribution in Korea, Japan and far east of Russia, To the West and South, it helped the formation of the secondary center and the tertiary center in China. (2) In an opposite way, there were three independent centers for annual wild soybeans in China, from which annual wild soybean distributed throughout China.

The database used in this study was established by character and evaluated phenotypic traits of annual wild soybean; the analysis data came from morphology and biochemical characters. As we know, phenotypic traits are controlled by genes and affected by environment, but large numbers of accessions can adapt to environments. The phenotypic data has more polymorphism in genetic diversity and reveal genetic variation indirectly. The molecular data reveal genetic variation directly, but fewer markers has less polymorphism. It is very difficult to obtain molecular data for a large number of accessions that has enough polymorphism to show the genetic diversity of germplasm. So, for the genetic diversity researches on large numbers of accessions were base on phenotypic traits (e.g. Perry and McIntosh 1991; Xu 1993; Dong 1998), and only small numbers of accessions have been used for molecular research on genetic diversity of germplasm (e.g. Keim et al. 1989; Skorupska et al. 1993;

Doldi et al. 1997). The traits used in most genetic diversity studies of annual wild soybean have been isoenzymes, morphology, biochemical or molecular characters.

Acknowledgements This study was supported by the National Natural Foundation (grant no. 39730330) and "973" Grant (no. G1998010200) of China.

Reference

- Abe J, Ohara M, Shimamoto Y (1992) New electrophoretic mobility variations observed in wild soybean (*G. soja*) distributed in Japan and Korea. *Soybean Genet Newsl* 9:63–72
- Chiang YC (1986) Genetic and quantitative variation in wild soybean (*Glycine soja*) populations. *Diss Abstr Int B* 47:2
- Doldi ML, Vollmann J, Lelley T (1997) Genetic diversity in soybean as determined by RAPD and microsatellite analysis. *Plant Breed* 116:331–335
- Dong YS (1998) Exploration on genetic diversity center for cultivated soybean in China. *Chin Crops J* 1:18–19
- Fukuda Y (1933) Cytogenetical studies on the wild and cultivated Manchurian soybean (*Glycine L.*). *Jpn J Bot* 6:489–506
- Gorman MB (1984) An electrophoretic analysis of the genetic variation in the wild and cultivated soybean germplasm. *Diss Abstr Int B* 44:11
- Hymowitz T (1970) On the domestication of the soybean. *Econ Bot* 23:408–421
- Keim P, Beavis W, Schupp J, Freestone R (1992) Evaluation of soybean RFLP marker diversity in adapted germ plasm. *Theor Appl Genet* 85: 205–212
- Kiang YT, Chiang YC, Kaizuma N (1992) Genetic diversity in natural populations of wild soybean in Iwate prefecture, Japan. *J Hered* 83:325–329
- Li FS (1994) Geographical distribution and ecological diffusing of wild soybean germplasm in China. *Researching progress on wild soybean of China*. Chinese Agricultural Press, Beijing
- Lu SL (1978) Discuss about origin of cultivated soybean in China. *Chin Agric Sci* 4:90–94
- Perry MC, McIntosh MS (1991) Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. *Crop Sci* 31:1350–1355
- Skorupska HT, Shoemaker RC, Warner A, Shipe ER, Bridges WC (1993) Restriction fragment length polymorphism in soybean germplasm of the southern USA. *Crop Sci* 33:1169–1176
- Tang QY (1997) Data processing system. Chinese Agricultural Press, Beijing
- Vaughan DA, Fukuoka S, Okuno K (1995) Population genetic diversity of wild soybeans on the Kanto plain, Japan. *Soybean Genet Newsl* 22:145–146
- Vavilov NI (1973) The origin, variation, immunity and breeding of cultivated plants, vol 13. Ronald Press, New York
- Wang JL (1973) The analysis on light ecological types for wild soybean in China. *Genet Newsl* 3:1–8
- Wang LZ (1985) Origin and diffusion of soybean. *Soybean Sci.* 4:1–7
- Xu B (1986) Three new evidences for origin of soybean. *Soybean Sci* 5: 123–130
- Xu B (1993) Polymorphism and geographical distribution of protein content of wild soybean(*Glycine soja*) in China. *FAO/IBPGR Plant Genet Resources Newsl.* 94/95:1–3
- Xu B (1995) Geographical distribution and genetic diversity of seed traits of wild soybean in China. *Acta Agron Sin* 21: 733–739
- Yu HG, Kiang YT (1993) Genetic variation in South Korean natural populations of wild soybean (*Glycine soja*). *Euphytica* 68:213–221
- Zhou XA (1998) The first discussion about center of genetic diversity of cultivated soybean in China. *Chin Agric Sci* 31:37–43