Indian J. Genet., 69(1): 28-34 (2009)

Extraction of core sample collection and selection of parents from indigenous and exotic germplasm of soybean [*Glycine max* (L.) Merrill]

A. Anna Durai[1] and B. Subbalakshmi

Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641003

(Received: November 2007; Revised: December 2008; Accepted: January 2009)

Abstract

Five hundred and twelve soybean germplasm accessions were evaluated for 11 agronomical traits and characterized with respect to 18 morphological characters. A well-represented sample collection was derived based on the agronomical data. Through principal component analysis it was inferred that number of clusters per plant, total number of pods per plant, total dry matter accumulation, seed yield per plant and maturity traits contributed maximum towards the genetic diversity of the accessions in base and also in sample collection. With respect to morphological traits, hypocotyl anthocyanin pigmentation, seed shattering, shape of lateral leaflet, flower colour and seed coat colour were highest contributors to genetic divergence. From the sample collections after studying the oil and protein content, TNAU S 55, JS (SH) 89-49, EC 2572, CO 2, EC 799 and JS 92-4 were identified as potential parents to get good segregants.

Key words: *Glycine max,* genetic divergence, agronomic traits, morphological trait

Introduction

Soybean [*Glycine max* (L.) Merrill] a "miracle legume" of the 21st century contains high amount of quality protein (42 per cent) and oil (20 per cent). The theoretical limit of soybean productivity was suggested to be 8 tonnes / ha based on the amount of light energy available in the field [1]. However, world productivity during 2007 was 2.81 tonnes / ha. Even this level has not been achieved in tropical countries like India where productivity is 1067 kg / ha during 2007-08. This relative low productivity is mainly due to short growing period available in subtropical conditions, limited genetic diversity and narrow genetic base of soybean cultivars [2]. In order to widen the genetic base, thought of conserving great number of germplasm accessions came to limelight. The diversity pattern of plant genetic resources having gene(s) of our interest may serve as a valuable guide for finding and incorporating new genes into elite genotypes. For thorough understanding of variability present in gene pool of a crop, critical evaluation and characterization of germplasm collections is imperative.

To identify new and useful genes with the limited resources, a diverse representative and severely limited sample with minimum redundancy *i.e.,* a core collection of the gene pool is required. The economy of size of the core makes it possible to include quality characters in evaluation but that need costly techniques, which can not be employed on large germplasm collection. However, the validity of the core depends on its degree of representativeness of the base collection (BC). This technique of pre-breeding was used in the present study to extract a sample collection (SC) from soybean germplasm. Genetic divergence was also studied in SC to select the parents for further breeding programme.

Materials and methods

A total of 512 indigenous and exotic collections of soybean germplasm maintained in Tamil Nadu Agricultural University, Coimbatore formed as material for the present study on genetic diversity. The experiment was conducted during wet seasons of 2002-2003 and 2003-2004 at Millet Breeding Station, Coimbatore (11°N, 77° E), India. The soil type of the experimental field was sandy loam with low available N (303 kg / ha), medium P (20kg / ha) and high K (1603kg / ha). The soil pH was 7.4 with 0.4 dSm^{-1} EC with 0.50 per cent organic matter. Each accession was sown in a

[1]Present address: Division of Crop Improvement, Sugarcane Breeding Institute, Coimbatore 641 007

ridge of four meters length spaced 45 cm apart with intra row spacing of 20 cm in randomized block design with two replications. Two seeds were hand dibbled on the sides of the ridges and they were thinned to one seedling per hill on 10th day after sowing. At planting, fertilizer was applied at rates equivalent to 20:60:40 kg/ ha, of NPK.

The crop was raised in irrigated condition. Data were recorded on 11 agronomical, namely, days to 50 % flowering (DFF), number of branches per plant (NBP), days to maturity (DM), plant height (cm) (PHT), number of clusters per plant (NCP), number of pods per cluster (NPC), total number of pods per plant (TPP), total dry matter accumulation (g) (TDMA), number of seeds per pod (NSP), hundred seed weight (g) (HSW) and seed yield per plant (g) (SYP) and 18 morphological characters *viz*., hypocotyl anthocyanin pigmentation colour (HAP): 1= pink , 9 = green; early plant vigour (EPV): 1 = poor, 2 = good, 3 = very good; plant growth type (IGC): 1= determinate, 2= semi-determinate, 4= indeterminate; intensity of green colour (PGT): 3 = light green, 5 = medium green, 7 = dark green; branching pattern (BRP): 1= less branched, 2= highly branched; shape of lateral leaflet (SHL): 3 = round ovate, 5 = pointed ovate, 7 = triangular, 9 = lanceolate; size of the lateral leaflet (SIL): 3 = small, 5 = medium, 7 = large; pubescent colour (PUC): 1 = grey, 2 = tawny; pubescent density (PUD): 1 = glabrous, 3 = sparse, 5 = medium, 7 = normal, 9 = dense; pubescent type (PUT): 1= erect, 2 = semi suppressed, 3 = appressed, 4 = glabrous; flower colour (FLC) 1 = white, 2 = light purple, 3 = purple, 4 = dark purple; pod colour after maturity (PCM): 1 = light brown, 2 = brown, 3 = dark brown, 4 = black, 5 = yellow; lowest pod height (LPH): 3 = low, 4 = slightly low, 5 = intermediate, 6 = slightly high, 7= high; seed coat colour (SCC): 1 = yellowish white, 2 = yellow, 3 = pale green, 4 = green, 5 = reddish brown, 6 = black, 7 = yellowish green, 8 = mottling; seed coat lustre (SCL): 3 = dull, 5 = intermediate 7 = shiny, 9 = blooming; hilum colour (HIC): 1 = brown, 2 = black, 3 = green, 4 = grey, 5 = imperfect black, 6 = buff, 7 = brown and black; lodging resistance (LOR): 1 = 100 per cent lodging, 3 = 75 per cent lodging, 5 = 50 per cent lodging, 7 = 75 per cent lodging, 9 = no lodging; and shattering in the field (SHL): 1 = absent, 9 = present listed in the NBPGR minimal descriptors with addition of some more characters on five randomly selected competitive plants in each of the accession at various phenophases of the crop.

The data on 11 agronomic traits were subjected to initial grouping of accessions to form non-overlapping strata of germplasm through hierarchical cluster analysis using NTSYSpcv2.02i [3]. The coverage of variation in the SC in relation to the BC was analyzed by various parameters like principal component analysis (PCA), range retention and Shannon diversity index (SDI). Fifty one entries in extracted SC were further subjected to genetic divergence analysis [4] after estimating the protein (PC) and oil content (OC).

Results and discussion

Size of the core collection

Lixia *et al*. [5] considered the optimal sampling proportion for Chinese core collection was about 2% retaining 84.4% of SSR marker variation and 91.9% of phenotypic variation of pre-core collection. According to the neutral theory of Brown, a core collection with sample of size of around 10 % of the whole accessions, up to a maximum of 3000 accessions could capture 70 % of alleles with 95 % certainty [6]. For smaller collection, the sampling proportion varies from 6.5 [7] to 29.0 [8]. In chickpea a mini core core set of about 1 per cent was composed and proved to be able to preserve variations in core collection [9] In spring soybean of China, a sampling proportion of 1.4 % can capture 70 % genotypic and over 80 % phenotypic variation of pre-core collection [10]. By comparing the various samples with different proportion of whole collection Qiu *et al*. [11] concluded that the sample size of 9 % was taken as the best proportion for construction of core collection because it could keep the variation with the least varieties in soybean. In the present study the size of SC was fixed at 10 % of BC, a manageable one.

Initial grouping of accession

The collections have to be divided into distinct groups before clustering [12]. The clustering for core collection can be established based on phenotypes [13], pedigree [14] geographical origin [15], Izozymes [16] and DNA markers [17]. Among the different strategies evaluated, the genetic diversity base was found to be the simplest and most efficient. Using hierarchical classification and cluster analysisis Zhao *et al*. [18] selected out 652 accessions accounting for 10.65 % from a total of 6172 accessions of wild soybean (*G. soja*). From the dendrogram (not presented) obtained in the present study through hierarchical cluster analysis of eleven agronomical data, fourteen distinct clusters could be recognized. The accessions having great affinity among them in phenotypic expression of agronomical traits formed a single cluster. The cluster I was found to be the largest with 234 accessions while the cluster IX was the smallest with only one accession (Table 1).

Constitution of core collection

The right choice of sample size for each cluster was carried out through stratified random sampling following the procedure of logarithmic strategy. When agronomic traits were used for sampling special allele also should be noticed [5]. Accordingly the cluster DC with single accession of highly distint nature was found and had been included in the core without subjecting it to sampling (Table 1).

Validity of the core collection

The degree of representation of BC in SC depends largely on the relative amount of variability present in SC. The total variance accumulated by principal components close to 80 per cent explains satisfactorily the variability manifested between individuals [19].

In the present study, the first three eigen vectors explained 66.1 and 68.6 per cent respectively of the total variance of agronomical traits of BC and SC (Table 2). However, 93.5 per cent of total variation as explained by first three canonical vectors was reported earlier [20]. The first principal component, which alone explained about 40 per cent of gross variability among BC accessions in the present case, had been mainly due to NCP, TPP, TDMA and SYP. Besides above traits, NBP also contributed to the first principal component that explained about 42 per cent of total variance in SC. All the traits that contributed maximum to genetic diversity of BC were also contributed at the same level in SC. However, NBP, which contributed minimum at sixth principal component in BC, contributed maximum in SC at first principal component. There was no difference between BC and SC in the share of maturity traits in the total variation. However, with respect to NSP and HSW, there were due to differential frequency distribution of these traits in these two groups.

The percentage of range retention in SC in relation to BC was found from 57.6 (DFF), 79.4 (HSW), 85.2(NSP), 87.4 (DFF), 91.0 (PHT), 96.3 (NPC), 96.7 (SYP), 97.4 (TPP), 99.4 (NBP), 99.7 (TDMA) to 100.0 per cent (NCP). While comparing the different strategies

Table 1. Constitution of sample collection from base based on logarithmic strategy

Cluster	No. of	Log pi	$\frac{\log p_i}{\Sigma \log p_i}$	Accessions selected for constituting a core
I	234	2.3692	7.58 = 8	AMSS 34, EC 13043, SL 295, MACS 693, PK 1024, NRC 27, 16008 and Himso 1563
II	11	1.0414	3.33 = 3	EC 18266, MACS 694 and EC 95253
III	6	0.7782	2.48 = 2	G 88 and JS 92-4
IV	50	1.6990	5.43 = 5	EC 15996, RAUS 5, LSB 2, NRC 35 and TS 99- 128
V	8	0.9031	2.88 = 3	NRC 36, 13052 and EC 28635
VI	7	0.8451	2.71 =3	EC 799, SL 599 and TNAU 20030
VII	44	1.6435	5.25 = 5	EC 220, Himso 1596, TS 82, RAUS 4 and TNAU 20027
VIII	42	1.6232	5.19 = 5	TNAU 20029, TNAU 20028, KB 165, MAUS68andDS9814
IX	1		-	JS (SH) 89-49
X	69	1.8388	5.88 = 6	EC 110399, MAUS 62, EC 1027, UGM 77, CO 1 and Bragg
XI	24	1.3802	4.42 = 4	MACS 124, DS 9501, Himso 1565 and PK 1014
XII	11	1.0414	3.33=3	151515, CO 2 and EC 16652
XIII	3	0.4771	1.53=2	EC 2572-A and TNAU S 7
XIV	2	0.3010	0.96 = 1	TNAU S 55
	Σ pi = 512	Σ pi = 512	$\Sigma \frac{\log p_i}{\Sigma \log p_i}$	

Table 2. Eigen values and proportion of variation explained by first six principal components for agronomical and morphological traits

Principal component	Eigen		Proportion		Characters	
	SC	SC	BC	SC	BC	SC
Agronomical traits						
1	4.457	4.608	40.5	41.9	NCP, TPP, TDMA and SYP	NBP, NCP, TPP, TDMA and SYP
2	1.736	1.881	15.8	17.1	DFF and DM	DFF and DM
3	1.079	1.059	9.8	9.6	NSP	HSW
4	0.978	1.044	8.9	9.5	NPC	NPC
5	0.918	0.879	8.3	8.0	HSW	NSP
6	0.654	0.597	5.9	5.4	PHT and NBP	PHT
Morphological traits						
1	12.432	16.475	21.8	23.9	SHL	HAP and BRP
2	8.348	14.227	14.7	20.7	HAP, SCL	FLC and SHL
3	6.510	8.276	11.4	12.0	-	-
4	4.865	6.949	8.5	10.1	-	PUD
5	4.538	4.868	8.0	7.1	LOR and PUD	LOR, LPH and PCM
6	3.947	4.336	6.9	6.3	-	-

of composing a core collection of annual *Medicago* species, it was concluded that the best one retained in an average between 74 and 81 per cent of ranges of variables [21]. In the present study, the average range retention was 90.2 per cent confirming the effective representation of BC in SC for observed agronomical traits.

The Shannon diversity indices for 22 morphological traits and frequency distribution of their states are presented in the Table 3. The index of SC was higher than that of BC for all the morphological characters except for PUT, PCM, SCC, SCL and HIC. The equal or high value of indices in the selected entries as compared to the whole collection was an indicative of better representation of diversity. The traits, SCC and HIC were not represented fully in SC. Similarly in green gram germplasm accessions, the indices for PCM, SCC and SCL of base were higher than that of core collection. This was due to the presence of rare descriptors in the BC, not represented in SC (22). On the basis of above observations on BC and SC, it was concluded that SC well represented the BC with respect to agronomical and morphological characters of the soybean accessions. Hence, this SC can be studied in detail to select entries for further crop improvement activities.

Genetic divergence study

Among the 17 clusters obtained in SC, cluster 2 was the largest with 20 entries including AMSS 34, TS 99-128, NRC 35, RAUS 4, LSB 2, RAUS 5, KB 165, MAUS 68, TNAU 20028, EC 28635, SL 599, 13052, TNAU 20030, TNAU 20027, Himso 1596, Himso 1565, MAUS 62, TS 82, DS 9814 and NRC 36 followed by cluster 1 with 16 entries *viz*., SL 295, PK1024, NRC 27, MACS 694, EC 15996, CO 1, 16008, UGM 77, EC 1027, G88, Himso 1563/MACS 693, EC 18266, 151515, Bragg and EC 110399. Each one of remaining 15 entries dispersed away to form a separate group (Table 4). The accessions selected from the cluster V of BC were found in the single cluster in SC also. The accessions selected from the cluster X of BC followed the same trend. However accessions, from others cluster of the BC were found either as single accession cluster or in different clusters of SC. While the uniform clustering pattern in BC and SC indicated no variation, differential clustering pattern revealed the divergence with respect to oil and protein content which were estimated only in SC.

The cluster composition of SC revealed that there was no association between eco-geographical distributions of genotypes. Constellation of genotypes from different eco geographic locations into one cluster could be due to unidirectional selection and genetic

Table 3. Shannon diversity indices (SDI) for qualitative descriptors and frequency of their different states in base collection

SI.No.	Descriptors	SDI of base collection	Frequency of descriptor states
1	HAP	0.1567(0.2554)	1 = 0.12(0.27), 9 = 0.88(0.73)
2	IGC	0.3385 (0.3749)	3 = 0.51(0.55), 5 = 0.47(0.39), 7 = 0.02(0.06)
3	PGT	0.3903 (0.4123)	1 = 0.23(0.29), 2 = 0.63(0.49), 4 = 0.14(0.22)
4	EPV	0.3650 (0.4108)	1 = 0.06(0.10), 2 = 0.58(0.45), 3 = 0.36(0.45)
5	SIL	0.3152(0.3491)	3 = 0.16(0.18), 5 = 0.75(0.70), 7 = 0.09(0.12)
6	SHL	0.1805(0.2655)	3 = 0.88(0.75), 5 = 0.10(0.22), 7 = 0.01(0.02), 9 = 0.01(0.00)
7	PUC	0.2471 (0.2763)	1 = 0.78(0.67), 2 = 0.22(0.33)
8	PUD	0.5521 (0.5790)	1 = 0.02(0.00), 3 = 0.12(0.22), 5 = 0.42(0.33), 7 = 0.11(0.14), 9 = 0.33 (0.31)
9	PUT	0.4939 (0.4704)	1 = 0.29(0.26), 2 = 0.38(0.39), 3 = 0.33(0.35), 4 = 0.09(0.00)
10	BRP	0.1431 (0.1576)	1 = 0.10(0.12), 2 = 0.90(0.88)
11	FLC	0.4735 (0.4907)	1 = 0.12(0.14), 2 = 0.06(0.08), 3 = 0.52(0.57)
12	LOR	0.3336 (0.4676)	1 = 0.00(0.00), 3 = 0.01(0.04), 5 = 0.06(0.02), 7 = 0.20(0.26), 9 = 0.73 (0.68)
13	LPH	0.5785 (0.6088)	3 = 0.25(0.31), 4 = 0.44(0.39), 5 = 0.21(0.18), 6 = 0.09(0.10), 7 = 0.02(0.02)
14	PCM	0.5016(0.4198)	1 = 0.29(0.25), 2 = 0.52(0.63), 3 = 0.15(0.08), 4 = 0.01(0.00), 5 = 0.03 (0.04)
15	SEC	0.3372 (0.3086)	1 = 0.13(0.12), 2 = 0.78(0.80), 3 = 0.02(0.02), 4 = 0.01(0.00), 5 = 0.03 (0.02), 6 = 0.02(0.04), 7 = RD(0.00), 8 = 0.01(0.00)
16	SCL	0.3946 (0.3725)	3 = 0.65(0.67), 5 = 0.21(0.24), 7 = 0.14(0.09), 9 = RD(0.00)
17	HIC	0.4706 (0.3793)	1 = 0.60(0.61), 2 = 0.29(0.33), 3 = 0.02(0.00), 4 = 0.03(0.02), 5 = 0.01 (0.04), 6 = 0.01(0.00), 7 = 0.04(0.00)
18	SHT	0.2489 (0.2763)	1 = 0.26(0.33), 9 = 0.74(0.67)

Values in parentheses represent respective figures of sample collections: RD - rare descriptor

diversity among the genotypes of common geographic origin which may be attributed to genetic architecture of the populations, past history of selection, developmental traits and degree of general combining ability.

By adopting the ranking technique it was understood that TPP showed the highest contribution and DM. None of the 13 characters showed zero or negative effects on genetic divergence. Wide range of mean values among the clusters was noticed for PHT, NBP, NCP, TPP, TDMA and SYP. The perspective donors for different characters of soybean could be taken from the cluster 17 for SYP, PC, and TDMA, cluster 16 for early flowering, NPC and OC, cluster 13 for TPP, cluster 11 for NBP and NCP, cluster 6 for early maturity and bold seeded types and cluster 15 for NSP (Table 4).

The generalized intercluster distance varied from 90.27 (between cluster 17 and 14) to 12.33 (between clusters 5 and 3). Based on the range of D values obtained, it was inferred that all the genotypes within clusters were found with low to moderate level of genetic divergence. From the inter cluster distances among potential clusters of SC, JS (SH) 89-49 was found distantly related with TNAU S 55 (77.96), TNAU S7 (70.00) and JS 92-4 (65.22). Similarly, TNAU S 55 was in great distance from EC 2572 (62.67) and CO 2 (53.56) and EC 2572 from TNAU S 7 (56.83) and JS 92-4 (56.28) Hence hybridization among these entries will give good segregants.

Table 4. Cluster mean for and percentage of contribution of quantitative characters to genetic diversity in sample collection of soybean

Cluster No.		OFF	PHT	NBP	NCP	NPC	TPP	NSP	HSW	DM	TDMA	SYP	OC	PC
1	(16 accessions)	43.38	46.00	4.39	20.42	2.58	66.35	2.03	11.33	90.63	37.05	14.69	21.54	40.48
2	(20 accessions)	40.10	33.12	2.32	10.72	2.26	33.58	2.08	11.66	86.70	21.89	8.28	21.56	40.83
3	(DS 9501)	34.00	38.80	4.70	25.00	3.15	89.00	*2.00*	10.44	85.50	51.13	21.04	22.28	40.60
4	(EC 16652)	42.00	30.30	6.30	23.00	3.30	91.00	*2.00*	10.52	92.00	34.25	17.81	21.35	40.84
5	(MACS 124)	42.00	47.50	6.75	26.70	3.05	91.00	*2.00*	9.14	92.50	44.80	19.50	21.98	41.15
6	(EC 2572)	34.00	38.40	3.05	17.30	2.05	46.15	*2.00*	**17.84**	84.00	36.20	15.69	21.51	41.60
7	(PK 1014)	44.00	54.85	5.65	28.30	2.70	104.25	2.15	11.45	91.00	48.65	22.73	21.60	40.48
8	(EC 220)	*49.00*	16.40	4.00	14.55	3.05	45.55	2.15	12.34	90.50	27.98	12.82	20.55	40.36
9	(EC 13043)	46.50	51.30	4.70	13.50	2.00	41.50	*2.00*	14.50	92.50	27.15	11.53	*20.08*	41.57
10	(EC 95253)	42.00	**60.45**	4.65	18.00	2.00	66.70	*2.00*	11.60	91.50	42.52	14.66	22.98	39.27
11	(JS 92-4)	43.00	52.00	**8.00**	**42.30**	2.00	104.30	*2.00*	*7.80*	**93.50**	41.52	16.36	20.56	40.45
12	(EC 799)	42.00	35.70	2.00	8.30	2.70	20.75	3.00	11.32	**93.50**	16.62	4.52	22.09	40.14
13	(TNAUS 7)	37.50	56.05	3.95	41.00	2.00	**133.70**	2.05	11.04	90.00	45.23	19.53	21.51	40.45
14	(TNAU 20029)	35.00	*14.80*	0.00	*5.45*	2.05	*14.25*	*2.00*	14.75	85.00	*13.08*	3.72	20.54	40.21
15	(JS(SH)-89-49	*43.00*	32.90	4.00	27.15	2.15	34.75	**3.15**	12.57	90.00	26.08	13.08	21.60	40.70
16	(CO 2)	**32.00**	23.00	5.30	13.50	**4.60**	69.30	*2.00*	10.88	85.00	35.20	14.08	**23.04**	*39.17*
17	(TNAU S 55)	34.50	51.20	4.00	28.45	3.05	140.50	*2.00*	11.12	85.50	**65.49**	**26.93**	20.62	**42.27**
Percentage of contribution to diversity		0.24	4.31	1.96	0.16	2.98	36.16	*5.10*	10.12	1.33	**23.29**	0.39	9.88	4.08

References

1. **Specht J. E., Hume D. J. and Kumidini S. V.** 1999. Soybean yield potential - A genetic and physiological perspective. Crop Sci., **39**: 1560-1570.
2. **Singh R. J. and Hymowitz T.** 2001. Exploitation of wild perennial glycine species for improving the soybean. *In*: Bhatnagar P.S. (ed.), Proceedings of India Soy forum, 2001, pp 5 8-61.
3. **Rohlf F. J.** 1992. NTSYS-pc: Numerical taxonomy and multivariate analysis system. Version. 1.8. State University of New York, Stony Brook, New York.
4. **Rao C. R.** 1952. Advanced statistical methods in biometric research. John Wiley and sons, Inc., New York.
5. **Lixia W., Yuan G., Rongxia G., Yinghui Li., Yangsong M., Zhimin D., Xian L., Haiyan Z., Yueqiang Z., Zhongxiong L., Ruzhen C., Haiming X., Linhai L., Fanyun L., Weijiang L., Zhe Y., Xuecheng N., Li Z., Yanhua C., Rinhua P., Yan L., Pengying C. and Lijuan Q.** 2006. Establishment of Chinese soybean core collection with agronomic traits and SSR markers. Euphytica, **151**: 215-223.
6. **Brown A. H. D.** 1989. Core collection: A practical approach to genetic resources management. Genome, **30**: 818-824.
7. **Knupfferr H., V. Th Hintum J. L.** 1995. The barley core collection: An international effort. *In*: Hodgkin T., Brown A. H. D., Van Th Hintum J. L., Morales E. A.V. (eds.). Core collection of plant genetic resources, John Wiley and Sons, West Sussex, England.
8. **Cordeiro C. M. T., Morales E. V., Ferreira P., Rocha D. M. S., Costa I. R. S., Valois A. C. C. and Silva S.** 1992. Towards a Brazillian core collection of cassava, Proceedings of an IBPGR/CENARGEN/CGN workshop on core collection, Brasilia, August.
9. **Upadhayaya H. D. and Ortiz R.** 2001. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. Theor. Appl. Genet., **102**: 1292-1298.
10. **Wang L. X., Li Y. H., Li W., Zhu L., Guan Y., Ning X. C., Liu Z. X., Guan R. X., Chang R. Z. and Qiu L. J.** 2004. Establishment of a core collection of Changjiang spring sowing soybean. Biodiversity Sci., **12:** 578-585 (In Chinese with English abstract).
11. **Qiu L. J., Cao Y. S., Chang R. Z., Zhou X. A., Wang G. X., Sun J. Y., Xie H., Zhang B., Li X. H., Xu Z. Y. and Liu.** 2003. Establishment of Chinese Soybean (*Glycine max*) core collection I: Sampling strategy. Sci. Agri Sin., **36**: 1442-1449.
12. **Hintum V. Th. J. L., Bothmer R. V. and Visser D. L.** 1995. Sampling strategies for composing core collection of cultivated barley (*Hardeum vulgare*) collection in China. Hereditas, **122**: 7-17.
13. **Huaman Z., Aguliar C. and Ortiz R.** 1999. Selecting a Peruvian sweeet potato core collection on the basis of morphological ecogeographical and disease and pest reaction data. Theor. Appl. Genet., **98**: 840-844.
14. **Martynov S. P., Dobrotvosskaia T. V., Dotlacil L., Stehno Z., Faberova I. and Bares I.** 2003. Genealogical approach to the formation of the winter wheat core collection. Rus. J. Genet., **39**: 917-923.
15. **Tai P. Y. P. and Miller J. D.** 2001. A core collection of *Saccharum spontaneum* L from the world collection of sugarcane. Crop Sci., **41**: 879-885.
16. **Chandra S., Huaman Z., Krishna H. S. and Ortiz R.** 2002. Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data-a simulation study. Theor. Appl. Genet., **104**: 1325-1334.
17. **McKhann H., Camilleri, Berard A., Bataillon T., David J. I., Reboud X., Corre V. L., Caloustia C., Gut I. G. and Brunei D.** 2004. Nested core collection maximizing genetic diversity in *Arabidopsis thaliana.* The Plant J., **38**: 193-202.
18. **Zhao L., Dong Y., Liu B., Wang K. and Li X.** 2005 Establishment of core collection for the Chinese annual wild soybean (*Glycine sojd*). Chinese Science Bulletin 10.
19. **Mardia K. V., Kent J. T. and Bibby J. M.** 1979. Multivariate analysis. Academic press, London.
20. **Gawande P. P., Shanthi Patil, Mayakumari and Bailkiwad V. J.** 2002. Selection of superior germplasm lines of soybean [*Glycine max* (L.) Merrill] for breeding. J. Soils and Crops, **12**: 236-240.
21. **Diwan N. Melntosh M. S. and Bauchan G. R.** 1995. Methods of developing a core collection of annual *Medicago* species. Theor. Appl. Genet., **90**: 755-761.
22. **Bisht I. S., Mahajan R. K. and Patel D. P.** 1998. The use of characterization of data to establish the mungbean core collection and assessment of genetic diversity. Genet. Resour. Crop Evol., **45**: 127-133.