# The basis of sustainable soybean production: The establishment and application of Chinese soybean core collections revealed by both agronomic characters and SSR markers

Qiu Lijuan, Guan Rongxia, Li Yinghui, Wang Lixia, Guan Yuan, Yan Zhe, Ma Yansong, Piao Rihua, Li Linhai, Ning Xuecheng, Zhu Li, Li Wei, Lin Fanyun, Luan Weijiang, Liu Zhang Xiong and Chang Ruzhen

Key lab of Crop Germplasm and Biotechnology, Chinese Agriculture Ministry; Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences, Beijing 10008, P.R. CHINA Email qiu\_lijuan@263.net

#### Abstract

China owns the most soybean germplasm in the world, and it becomes urgent task to evaluate and utilize the collection efficiently. Based on the phenotypic data of Chinese soybean germplasm, a total of 20 sampling methods were compared and the best one was subjected to establish primary core collection, which was proved to have representative for the whole collection. A core set of SSR loci were identified with 80 autumn soybean accessions and confirmed with 190 random soybean accessions from the whole country. By analyzing the primary core collection with a core set of SSR loci, 1055 cultivated soybean accessions, taking of the 5% of the whole collection. Meanwhile, the good quality sub-collections were set up in order to meet the needs of soybean breeding, and they were compared with the reserved collections for the frequencies of accessions with lacking 28K or high 11S/7S ratio. The results showed that the prior collections had higher ratio for the targeted traits and a novel cultivated soybean lacking  $\beta$  subunit was found. It indicates that the core collection will play a very important role in mining gene, functional genomic study and soybean cultivar improvement.

#### **Media summary**

Chinese soybean core collections were established and found novel favourite traits, which will play an important role in mining genes and breeding.

#### **Key Words**

Soybean; core collection; SSR marker; agronomic trait

#### Introduction

Plant genetic resources played a very important role in crop breeding. So far, more than 6.1 million plant genetic resources had been collected and conserved in worldwide (FAO, 2001). When these abundant resources provided rich genetic base for breeding, they also bring the difficulties for conservation, studies and utilization. The core collection proposed by Frankel et al (1984) and Brown et al (1989a) have been widely applied in more than 30 plant species. Among these core collections, most were established based on the agronomic traits (including quantity traits and quality traits), and a few were applied isozymes and molecular markers, and some of the core collections were estimated for their representative with more than 70%. However, no systematic analysis was conducted for the core collection at DNA level. The purpose of this paper is to establish Chinese soybean core collection by using both agronomic traits and SSR loci, and then estimated its diversity and identifying efficiency referred to the whole collection.

#### Methods

#### Selection of Primary Core Collection based on the Agronomic traits

A total of 23587 soybean accessions collected and conserved in China were used and they included seven planting types across three cultivating regions of China. All accessions are conserved both in the Long-term and Mid-term National Genebank located at the Institute of Crop Germplasm Recourses of

Chinese Academy of Agricultural Sciences in Beijing. The agronomic traits recorded in the Catalogue of Soybean Germplasm Resources in China (Vols. 1-3) were used to establish the primary core collections and subcollection. The quality traits were graded and the quantity traits divided into ten classes by standard error 0.5, the cluster analysis were conducted using S-PLUS software with Euclidean distances between accessions and square deviation of mean by Ward's method between clusters.

## Fingerprinting Chinese soybean germplasm by SSR markers

The seeds of each collection were used to extract genomic DNA with SDS method. A core set of SSR loci were screened by analyzing 80 autumn sowing soybean accessions and 190 random accession across China from the primary core collection. PCR reactions were performed on PE 9600 thermal cycler, and the volume of reaction mixture is 20 ul, containing 1?PCR Buffer, 2mM MgCl<sub>2</sub>, 100uM of each dNTP, 0.4mM of primer, 20ng DNA, and 1U of Taq DNA polymerase. DNA amplification was performance as following: 30s at 94 °C followed by 35 cycles with 30s at 94°C, 30s at 47 °C and 30s at 72 °C, and then holded at 4 °C. The amplification products were first analyzed on 6% SDS-PAGE gel and fragmented on Magabase 1000. The SSR alleles were record as 0-1 data (1 for the presence and 0 for absence) for further analysis. The PIC index means polymorphism information content, and the calculating formula is: PIC=1- $\Sigma P_{i}^2$  Here, P<sub>i</sub> is the frequency of the *i*th SSR alleles (Smith et al. 2000). The similarity coefficient were calculated based on the SSR data referring to Nei-Li(1979). Simpson index and Shannon-weaver were calculated with Foxpro3.0. The cluster analysis was used by unweighted pair group method arithmetic average (UPGMA) employing NTSYSpc2.10t.

## Results

## Selection of Primary Core Collection based on the Agronomic traits

More than 23000 soybean accessions are stored in Chinese National Genebank, but some of them, however, are entitled the same name. By analyzing six major groups including 'Man cang jin (Yan et al. 2003), the obvious differences were observed between all homonymy accessions at both phenotypic and molecular levels. Relatively, more differences were observed at the DNA level compared to the phenotypic level. The results indicated that these accessions with the same name might have less redundancy and more genetic diversity, and it is valuable to store all of them in National Genebank, which each of accessions can be taken as different individuals and used to establishing core collection. In order to select primary properly, 20 sampling methods were compared, which 18 of them consisted of 3 types of stratifications, 3 types of sampling number determinations and 2 types of individuals selecting methods, and the rest 2 methods were checks. The criteria against whole collection included variety classification, coincidence of 14 characters, mean of 5 quantitative characters, and variance of genetic diversity, mean of variety distance. The results indicated that variety classification was better than both one (cultivation region) and two (cultivation region + province) stratifications, the proportional sample determination was better than methods of square root or genetic diversity, and clustering selection of sample was better than random selection Therefore, the optimal strategy was selected to compare various samples with different proportion of whole collection, and the sample size of 9% were taken as the best proportion for construction most primary core collection because it could keep the variation with the least varieties (Qiu et al 2003). In addition, a few accessions with extreme agronomic traits were added to the primary core collection so that the primary core collection would have relative high representative of the whole collection. The primary core collection could representative the whole collection when it was tested at either agronomic traits (Cui et al 2004), or SSR loci by sampling the reserved collection for Huanghuai summer soybeans (Cui et al 2003).

## Selection of core SSR loci for fingerprinting Chinese soybean germplasm

A total of 80 Chinese autumn soybeans were analyzed with 96 SSR loci and defined a set of 60 representative SSR loci (Xie et al. 2003). The core set SSR loci had following characteristics (A) distributed on 20 integrated linkage groups integrated by Cregan (1999) with an average of 50 cM genetic distances between two adjacent loci ; (B) defined the relationship among Chinese autumn soybeans

were greatly significant correlation with that using 96 loci (r=0.910\*); (C) showed a higher level of polymorphism with an average of 8.8 alleles and 0.773 value of polymorphism informative content (PIC) per locus. Coincidently, 190 random accessions from Chinese soybean primary core collection appeared to have less than 0.05 standard error of similarity matrices with increasing up to more than 55 SSR loci. The results indicated that the genetic relationships of Chinese soybean cultivars could be defined by using over 570 alleles (Wang et al 2003).





## Establishment of Chinese Soybean Core Collection

The core set of SSR loci were used to analyze diversity of primary core collection and to establish core collection. The differentiation among various planting soybean types was detected by using SSR markers (Qiu et al. 2002). About 80 accessions were selected from five planting types of three cultivation regions and the differentiate coefficient of genetic diversity ranged from 0.039-0.353 with average of 0.154, which means 15.4% variation existed among different sowing types. The  $X^2$  tests for the distribution of alleles at each locus among five types were significant (Xie 2002). Since there are differentiations among 7 sowing types of three cultivation regions, the core collection were established by clustering each of sowing types separately. The accessions taking of the 5% of the total accessions, was selected as core collection, which represent over 85% genetic diversity of the whole collection at both phenotypic and molecular levels.

Using the core collection, 7 sowing types from 3 cultivation regions separated intro tree clusters based on the genetic distances of 45 SSR loci. The southern autumn sowing soybean (pop6) was totally different from the other 5 sowing types. Then huanghuai spring sowing soybean (pop2) was the second cluster. The last cluster included the rest sowing types, which both northern spring sowing soybean and summer sowing soybean grouped together and separated from Changjiang spring sowing soybean (pop4), southern summer sowing soybean (pop7) and southern spring sowing soybean (pop 5). The results indicated that accessions originated in southern region (pop4, pop 7, pop5) were different from the accessions of northern region (pop1) and middle China (pop3) (Figure 2). The same relationship was also defined by using genetic identities (Table 1).



Figure 2. The dendrogram of 7 sowing types based on genetic distances

Table 1. Nei's genetic identity (above diagonal) and genetic distance (below diagonal) among 7 sowing types

рор	1	2	3	4	5	6	7
1	****	0.6677	0.836	0.7201	0.6598	0.5941	0.6952
2	0.4039	****	0.7495	0.6437	0.7099	0.517	0.6497
3	0.1791	0.2884	****	0.7907	0.7269	0.6662	0.8349
4	0.3283	0.4406	0.2349	****	0.77	0.6888	0.8314
5	0.4158	0.3426	0.319	0.2614	***	0.6521	0.7953
6	0.5208	0.6597	0.4061	0.3728	0.4276	***	0.7427
7	0.3636	0.4313	0.1804	0.1846	0.229	0.2975	****

When the sample size decreased to 220 accessions (1% of the total cultivated soybeans), the average genetic similarity among accessions was about 0.28, which could represent over 70% genetic diversity of the total collection in cultivated soybean (Figure 3).

#### Efficient utilization of Sub-working Collection

109 accessions from the establishing high-protein and high-oil subcollection and 66 accessions randomly sampled from reserved collections were identified for protein subunits and allergy protein 28K in order to test availability of the good quality core collection and provide the information for establishment of sub-collection of the targeted trait. The ratio of accessions with good quality for either lacking 28K or high 11S/7S in good quality collection

was higher than that in reserved collection. The first lacking  $\beta$  accession was found in cultivated soybean of the good quality core collection. Compared to high-protein core collection and high-oil core collection among different ecotype regions, the different changing trends of good quality accession with lacking 28K or 11S/7S ≥2.5 were observed (Table 2). Therefore, good quality core collection is the important fundament ion of raising identification rate and finding novel gene.



Figur 3. Reservation ratio for different size of core collection

Table 2. The ratio of 11S/7S and lacking 28 K≥ 2.5 within reserved and sub core collection

Source	Total Accessions	11S/7S ≥ 2.5			Lacking 28 K		
		Range %	Ratio %	Mean	Total number	Ratio %	
Reserved collection	66	0.9~2.5	4.5	1.76	21	31.8	
Sub-core collection	109	0.8~4.7	9.2	1.85	54	49.5	
High-protein core collection	64	1.1~3.1	9.4	1.90	40	62.5	
High-fat core collection	45	0.8~4.7	8.9	1.77	14	31.1	
Total	175	0.8~4.7	7.4	1.81	74	41.11	

## Conclusion

Chinese soybean core collection was established, which could represent over 85% diversity with less than 5% accession of the whole collection. It had higher diversity with lower accessions compared to the concept proposed by Frankel et al (1984) and Brown et al (1989a).

#### References

Qiu L, Xie H, Chang R, Li W, Wang W, Zhang B, Zhang M, Feng Z (2002). Utilization of genetic diversity on establishing chinese soyubean (G. max) core collection. Journal of the Chinese Cereals And Oils Association Special: 15-21.

Qiu L, Cao Y, Chang R, Zhou X, Wang G, Xie H, Zhang B, Li X, Sun J, Xu Z, Liu L (2003). Establishment of Chinese soybean (*G. max*) Core Collection. I. Sampling strategy. Chinese Agricultural Sinica 36(2), 1442-1449.

Yan Z, Chang R, Guan R, Liu Z, Qiu L (2003) Analysis of similarity and difference of various collections of soybean variety Mancangjin by using agronomic traits and SSR markers. Journal of Plant Genetic Resources 4(2):128-133

Xie H, Qiu L, Chang R et al., (2003) Selection of core SSR loci by using Chinese autumn soybean (*Glycine max* (L.) Merr) Chinese Agricultural Sinica 36(4), 360-366.

Wang B, Chang R, Yan L, Tao L, Guan R, Zhang M, Feng Z, Qiu L (2003) Identification of SSR Primer numbers for Analyzing Genetic diversity of Chinese soybean cultivated soybean. Molecular Breeding 1(1), 82-88.

Cui Y, Qiu L, Chang R and Lv, W (2003) Examination of representiveness of the primary core collection in Huanghuai summer sowing soybean (Glycine max) using SSR Journal of Plant Genetic Resources 4(2):9-15.

Cui Y, Qiu L, Chang R and Lv, W (2004) Representative Test for Primary Core Collection of Summer Sowing Soybeans in Huanghuai Region of China. Acta Agronomica Sinica 1(3), 284-288.

Lin F, Qiu L and Chang R and He, B (2003) Genetic diversity of landrace and bred varieties of soybean in Shanxi Chinese Journal of Oil Crop Sciences 25(3), 24, 29.