

# Genetic Diversity of Wild Soybeans from Some Regions of Southern China Based on SSR and SRAP Markers

Bingrui Sun<sup>1,2\*</sup>, Chongyun Fu<sup>2\*</sup>, Cunyi Yang<sup>1</sup>, Qibin Ma<sup>1</sup>, Dajian Pan<sup>2</sup>, Hai Nian<sup>1#</sup>

<sup>1</sup>College of Agriculture, South China Agricultural University, Guangzhou, China; <sup>2</sup>Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China.

Email: <sup>#</sup>hnian@scau.edu.cn

Received January 7<sup>th</sup>, 2013; revised January 29<sup>th</sup>, 2013; accepted February 3<sup>rd</sup>, 2013

## ABSTRACT

There are rich annual wild soybean (*Glycine soja*) resources in Southern China, which are the progenitor of cultivated soybean. To evaluate the genetic diversity and differentiation of *G. soja* in Southern China, we analyzed allelic profiles of 141 annual wild soybean accessions from Southern China and 8 core wild soybean accessions from Northern China by using 41 simple sequence repeat (SSR) markers and 18 Sequence-related amplified polymorphism (SRAP) primer combinations. The 41 SSR markers produced a total of 421 alleles (10.27 per locus) with a mean of gene diversity of 0.825 (Simpson index) and 1.987 (Shannon-weaver index). The 18 SRAP primer combinations detected a total of 90 polymorphism bands (5 per primer combination) with a mean of gene diversity of 0.918 (Shannon-weaver index). SSR and SRAP markers detected 43 and 5 rare alleles in 149 wild soybeans, respectively. The wild soybeans from Fujian province showed the highest genetic diversity with Shannon-weaver index of 1.837 (by SSR) and 0.803 (by SRAP), and the highest allelic richness with an average of 8.8 alleles per locus and the most number of rare alleles of 0.68 per locus based on SSR data. An analysis of Molecular Variance (AMOVA) analysis showed that significant variance did exist among Hunan, Fujian, Guangxi and Northern China subpopulations based on SSR and SRAP data. The unweighted pair-group method of the arithmetic average (UPGMA) cluster analysis indicated that the wild soybeans from Fujian province occurred in different clusters based on both SSR and SRAP data. The above results indicated that Fujian province could be the major center of genetic diversity for annual wild soybean in Southern China. In addition, Mantle test showed there was a weak positive linear correlation ( $r = 0.25$ ) between SSR and SRAP analysis in the study.

**Keywords:** Genetic Diversity; *Glycine soja*; SSR; SRAP; Southern China

## 1. Introduction

Annual wild soybean (*Glycine soja* Sieb. et Zucc.), the progenitor of the cultivated soybean (*G. max* (L.) Merr.), is distributed in China, Korea, Japan and the far eastern regions of Russia. The horizontal distribution in China is from 53° to 24° north latitude and from 135° to 97° east longitude [1]. China has the most abundant germplasm resources of wild soybeans in the world, currently, about 7000 wild soybean accessions had been collected and preserved in National germplasm Bank of China, accounting for more than 90% of the world saving, and still considerable wild soybeans are being collected in various regions [2]. During the long term of evolution, wild soybeans accumulated a wide range of variation to adapt to the geographic, abiotic and biotic environmental conditions. *G. soja* and *G. max* are cross compatible and can produce fertile offspring. Therefore, the wild soybean

germplasm is a rich gene pool for improvement of resistances and tolerances to biotic and abiotic stresses of cultivated soybean and can be used in breeding research of high yield and high protein content.

The genetic diversity of annual wild soybeans has been studied using morphological traits [3-5], isozymes [6], RFLP [7,8], RAPD [9,10], SSR [11], ISSR [12] etc.. *G. soja* has a higher genetic diversity than *G. max*, and many rare variants have been lost throughout the domestication process of *G. soja* [13,14]. Through SSR markers and morph-biological methods, Wen *et al.* [11] found the genetic diversity of *G. soja* was much higher than that of its cultivated counterpart, and after domestication, the genetic diversity of the cultigens decreased, with its 65.5% alleles inherited from the wild soybean, while 34.5% alleles newly emerged by evaluating the genetic diversity and differentiation of 196 wild soybeans and 200 landrace soybeans in China. Hyten *et al.* [14] also indicated that the soybean has lost many rare sequence variants and has undergone numerous allele frequency

\* Authors contributed equally.

#Corresponding author.

changes throughout its history.

With regard to the origin of soybean in China, there were several hypotheses including Northeast China origin [15,16], Yellow River Valleys origin [17,18], Southern China origin [19,20] and multiple center origin [21], according to archaeological and historical investigation, the geographic distribution genetic diversity of annual wild soybean and cultivated soybean, ecological and biological comparison. Dong *et al.* [4] analyzed the genetic diversity of twelve morphological and biological traits of 6172 annual wild soybeans grown in China, and proposed there were three genetic diversity centers in China, namely, the Northeast, the Yellow River Valley and the Southeast Coasts, among which the Northeast center might be the primary center of annual wild soybean in China. Xu *et al.* [22] found the wild soybean subpopulation from southern China showed the highest genetic diversity, the medium in Huang-Huai-Hai region, and the lowest in northern China through analyzed morphological characters, isozymes and RFLPs of cytoplasmic DNA of more than 200 annual wild soybeans from various ecological regions in China, and proposed that South China was the center genetic diversity and might be the center of origin of wild soybean in China. Gai *et al.* [20] suggested that the ancient southern China *G. soja* population might be the common ancestor of all cultivated soybeans, and evaluated progressively to the ancient cultivated soybean, then ancient cultivated soybean in various regions accordingly evaluated to all kinds of cultivated soybean types. Through analyzing genetic diversity and peculiarity of annual wild soybean from various eco-regions in China, Wen *et al.* [11] also found the Southern China subpopulation showed the highest allelic richness, diversity index and largest number of specific-present alleles, thence, proposed Southern China should be the major center of diversity for annual wild soybean.

Tandem simple sequence repeats (SSR) or microsatellites, iterations of 1 - 6 bp nucleotide motifs, are abundant across genomes and show high levels of polymorphism. SSRs are usually considered as evolutionarily neutral DNA markers [23], and constitute a rather large fraction of non-coding DNA and are relatively rare in protein-coding regions. More than 1000 SSR markers have been explored and mapped in *G. soja* involved mapping populations [24]. Sequence-related amplified polymorphism (SRAP) developed by Li and Quiros in 2001 [25] is a simple marker technique for the amplification of open reading frames (ORFs). It combines simplicity, reliability, moderate throughput ratio and facile sequencing of selected bands. SRAP targets coding sequences in the genome and results in a moderate number of co-dominant markers. It is based on two-primer amplification, and twenty percent of the SRAP markers were co-dominant. Li *et al.* successfully tagged the glu-

cosinolate desaturation gene *BoGLS-ALK* with SRAP markers. SRAPs were earliest developed and utilized in *Brassica* [25] and currently also amplified in other crops such as rice, wheat, cotton, maize, soybean, rapeseed (*Brassica napus* L.), potato, orange, eggplant [26-30]. SRAP has been adapted for a variety of purposes, including map construction, gene tagging and genetic diversity studies [31]. In the study, we utilized both SSR and SRAP markers, for a more comprehensive explanation of the genetic diversity from the whole genome level including noncoding DNA and the open reading frames.

In this article, in order to comprehensively evaluate the genetic diversity, genetic relationship and differentiation of 149 annual wild soybeans mainly from 3 provinces of southern China, we made use of SSR mainly locating at the non-coding regions and SRAP markers lying in the coding regions to analyze them, which contributes to reveal the genetic relationship of wild soybean from southern China, and proposes new evidence for explanation of origin of soybean in China.

## 2. Materials and Methods

### 2.1. Plant Materials

One hundred and forty one wild soybean accessions (**Table 1**) used in this study including 58 wild soybeans from Hunan province distributed in twenty three counties and cities, 56 from Fujian province distributed in twenty six counties and cities, 27 from Guangxi province distributed in fourteen counties, and 8 from Northern China (3 from Liaoning, 2 from Hebei, 1 from Gansu, 1 from Shandong and 1 from Shanxi), in which 8 wild soybeans from Xintian of Hunan province are collected by our research group, and another 141 wild soybean germplasm were provided by Prof. Qiu from the Chinese Academy of Agricultural Sciences. For analyzing the genetic relationships between wild soybeans distributed in different ecological regions, we partitioned the 149 wild soybeans into four groups that were Hunan subpopulation, Fujian subpopulation, Guangxi subpopulation and Northern China subpopulation.

### 2.2. Methods

#### 2.2.1. Molecular Marker Analysis

Genomic DNA samples were extracted from fresh leaves of a single seedling at about the three-leaf stage, using the modified CTAB method. 60 SSR primer pairs (average two loci in every linkage group) reference to Hua Xie [32] and 26 SRAP primer combinations [26] were chosen to show soybean germplasm polymorphism, and finally picked out 41 SSR primer pairs (located in 20 linkage groups) and 18 SRAP primer combinations exhibited high polymorphism and clear bands to be used in calculating allelic number, genetic diversity index, ge-

**Table 1. Wild soybean accessions with their respective geographic origins and assigned clusters based on 41 SSR markers and 18 SRAP primer combinations by the UPGMA clustering.**

Code	Origin	Code number	Cluster <sup>1</sup>	Cluster <sup>2</sup>	Code	Origin	Code number	Cluster <sup>1</sup>	Cluster <sup>2</sup>
H1	Xintian, Hunan		I (1)	A(a)	F18	Jian'ou, Fujian	ZYD04923	I (2)	B(d)
H2	Xintian, Hunan		III	A(a)	F19	Taining, Fujian	ZYD04959	I (2)	B(d)
H3	Xintian, Hunan		I (1)	A(a)	F20	Taining, Fujian	ZYD04969	III	B(e)
H4	Xintian, Hunan		III	A(a)	F21	Xiapu, Fujian	ZYD04981	I (2)	B(d)
H5	Xintian, Hunan		I (2)	A(a)	F22	Xiapu, Fujian	ZYD04995	I (2)	B(d)
H6	Xiangtan, Hunan		III	A(a)	F23	Pingnan, Fujian	ZYD04996	I (2)	B(d)
H7	Xiangtan, Hunan		V	A(a)	F24	Jianning, Fujian	ZYD04997	I (2)	B(d)
H8	Yueyang, Hunan		III	A(a)	F25	Jianning, Fujian	ZYD05002	I (2)	B(d)
H9	Xintian, Hunan		I (1)	A(a)	F26	Jianning, Fujian	ZYD05004	I (7)	B(d)
H10	Xintian, Hunan		I (2)	A(a)	F27	Shunchang, Fujian	ZYD05005	I (2)	B(d)
H11	Xintian, Hunan		IV	A(a)	F28	Jiangle, Fujian	ZYD05010	I (2)	B(d)
H12	Huarong, Hunan	ZYD04662	II	A(a)	F29	Jiangle, Fujian	ZYD05017	II	B(d)
H13	Huarong, Hunan	ZYD04663	II	A(a)	F30	Jiangle, Fujian	ZYD05035	I (2)	B(d)
H14	Huarong, Hunan	ZYD04665	II	A(a)	F31	Longyan, Fujian	ZYD05055	I (7)	B(d)
H15	Huarong, Hunan	ZYD04666	II	A(a)	F32	Longyan, Fujian	ZYD05057	I (7)	B(c)
H16	Huarong, Hunan	ZYD04667	V	A(a)	F33	Gutian, Fujian	ZYD05070	I (2)	B(d)
H17	Huarong, Hunan	ZYD04668	V	A(a)	F34	Sha, Fujian	ZYD05071	II	B(d)
H18	Yueyang, Hunan	ZYD04669	I (3)	A(a)	F35	Sha, Fujian	ZYD05075	II	B(d)
H19	Yueyang, Hunan	ZYD04670	I (3)	A(a)	F36	Mingxi, Fujian	ZYD05076	I (2)	B(d)
H20	Yueyang, Hunan	ZYD04672	II	A(a)	F37	Mingxi, Fujian	ZYD05077	I (2)	B(d)
H21	Changde, Hunan	ZYD04673	II	A(a)	F38	Ninghua, Fujian	ZYD05078	I (2)	B(d)
H22	Changde, Hunan	ZYD04674	III	A(a)	F39	Ninghua, Fujian	ZYD05082	I (5)	B(d)
H23	Changde, Hunan	ZYD04675	V	A(a)	F40	Ninghua, Fujian	ZYD05085	I (5)	B(d)
H24	Xikou, Hunan	ZYD04678	III	A(a)	F41	Youxi, Fujian	ZYD05086	III	B(e)
H25	Longshan, Hunan	ZYD04679	I (3)	A(a)	F42	Qingliu, Fujian	ZYD05087	I (2)	B(e)
H26	Longshan, Hunan	ZYD04680	I (3)	A(a)	F43	Qingliu, Fujian	ZYD05095	I (2)	B(e)
H27	Longshan, Hunan	ZYD04681	V	A(a)	F44	Qingliu, Fujian	ZYD05100	I (6)	B(e)
H28	Xiangyin, Hunan	ZYD04682	I (3)	A(a)	F45	Yong'an, Fujian	ZYD05101	III	B(e)
H29	Taoyuan, Hunan	ZYD04683	I (3)	A(a)	F46	Yong'an, Fujian	ZYD05109	I (2)	B(e)
H30	Taoyuan, Hunan	ZYD04684	I (3)	A(a)	F47	Yong'an, Fujian	ZYD05110	I (2)	B(e)
H31	Yuanling, Hunan	ZYD04685	II	A(a)	F48	Changting, Fujian	ZYD05113	I (2)	B(e)
H32	Changsha, Hunan	ZYD04686	II	A(a)	F49	Liancheng, Fujian	ZYD05116	II	B(e)
H33	Fenghuang, Hunan	ZYD04687	II	A(a)	F50	Liancheng, Fujian	ZYD05122	I (2)	B(e)
H34	Fenghuang, Hunan	ZYD04688	VI	A(a)	F51	Liancheng, Fujian	ZYD05132	I (2)	B(e)
H35	Fenghuang, Hunan	ZYD04689	II	A(a)	F52	Longyan, Fujian	ZYD05155	I (2)	B(e)
H36	Liuyang, Hunan	ZYD04690	I (3)	A(a)	F53	Shanghang, Fujian	ZYD05158	III	B(e)
H37	Liuyang, Hunan	ZYD04691	IV	A(b)	F54	Xiapu, Fujian	ZYD05160	I (5)	B(e)
H38	Qianyang, Hunan	ZYD04692	II	A(b)	F55	Jian'ou, Fujian	ZYD05161	III	B(e)
H39	Qianyang, Hunan	ZYD04693	II	A(b)	F56	Taining, Fujian	ZYD05170	I (2)	B(e)
H40	Qianyang, Hunan	ZYD04694	II	A(b)	G1	Quanzhou, Guangxi	ZYD05194	I (4)	B(e)
H41	Hengshan, Hunan	ZYD04695	I (6)	A(b)	G2	Quanzhou, Guangxi	ZYD05196	I (4)	B(e)
H42	Hengshan, Hunan	ZYD04696	I (8)	A(b)	G3	Quanzhou, Guangxi	ZYD05207	I (6)	B(e)

## Continued

H43	Chengbu, Hunan	ZYD04697	II	A(b)	G4	Sanjiang, Guangxi	ZYD05219	I (6)	B(e)
H44	Suining, Hunan	ZYD04699	II	A(b)	G5	Sanjiang, Guangxi	ZYD05220	I (6)	B(e)
H45	Suining, Hunan	ZYD04700	III	A(b)	G6	Xing'an, Guangxi	ZYD05222	I (7)	B(e)
H46	Lingling, Hunan	ZYD04701	III	A(b)	G7	Xing'an, Guangxi	ZYD05230	I (7)	B(e)
H47	Lingling, Hunan	ZYD04702	III	A(b)	G8	Xing'an, Guangxi	ZYD05236	I (4)	B(e)
H48	Lingling, Hunan	ZYD04704	I (2)	A(b)	G9	Guanyang, Guangxi	ZYD05237	II	B(e)
H49	Chenzhou, Hunan	ZYD04705	III	A(b)	G10	Guanyang, Guangxi	ZYD05242	I (4)	B(e)
H50	Chenzhou, Hunan	ZYD04707	IV	A(b)	G11	Guanyang, Guangxi	ZYD05246	I (4)	B(e)
H51	Chenzhou, Hunan	ZYD04708	IV	A(b)	G12	Lingchuan, Guangxi	ZYD05247	I (4)	B(e)
H52	Jiangyong, Hunan	ZYD04709	III	A(b)	G13	Lingchuan, Guangxi	ZYD05250	I (4)	B(e)
H53	Jiangyong, Hunan	ZYD04710	IV	A(b)	G14	Gongcheng, Guangxi	ZYD05251	I (4)	B(e)
H54	Jiangyong, Hunan	ZYD04712	IV	A(b)	G15	Lipu, Guangxi	ZYD05256	I (7)	B(e)
H55	Dao, Hunan	ZYD04713	III	A(b)	G16	He, Guangxi	ZYD05258	I (4)	B(e)
H56	Dao, Hunan	ZYD04714	IV	A(b)	G17	Lingchuan, Guangxi	ZYD05265	I (4)	B(e)
H57	Dao, Hunan	ZYD04715	III	A(b)	G18	Yongfu, Guangxi	ZYD05266	I (4)	B(e)
H58	Dao, Hunan	ZYD04716	IV	A(b)	G19	Yongfu, Guangxi	ZYD05267	I (4)	B(e)
F1	Pucheng, Fujian	ZYD04803	I (3)	B(d)	G20	Rong'an, Guangxi	ZYD05269	I (4)	B(e)
F2	Pucheng, Fujian	ZYD04811	I (2)	B(c)	G21	Rong'an, Guangxi	ZYD05270	I (4)	B(e)
F3	Chong'an, Fujian	ZYD04845	I (2)	B(c)	G22	Luzhai, Guangxi	ZYD05271	I (4)	B(e)
F4	Chong'an, Fujian	ZYD04847	I (2)	B(c)	G23	Fuchuan, Guangxi	ZYD05272	I (4)	B(e)
F5	Guangze, Fujian	ZYD04848	I (2)	B(d)	G24	Nandan, Guangxi	ZYD05274	I (4)	B(e)
F6	Guangze, Fujian	ZYD04850	I (2)	B(d)	G25	Zhaoping, Guangxi	ZYD05279	I (4)	B(e)
F7	Guangze, Fujian	ZYD04855	I (7)	B(d)	G26	Xiangzhou, Guangxi	ZYD05280	I (4)	B(e)
F8	Songxi, Fujian	ZYD04856	II	B(d)	G27	Xiangzhou, Guangxi	ZYD05281	I (4)	B(e)
F9	Songxi, Fujian	ZYD04859	I (2)	B(d)	N1	Kaiyuan, Liaoning	ZYD05589	IV	B(e)
F10	Zhenghe, Fujian	ZYD04860	I (2)	B(d)	N2	Yingkou, Liaoning	ZYD01977	I (2)	B(d)
F11	Zhenghe, Fujian	ZYD04866	I (2)	B(d)	N3	Gai, Liaoning	ZYD02061	IV	B(d)
F12	Jianyang, Fujian	ZYD04867	I (2)	B(d)	N4	Chengde, Hebei	ZYD02739	I (3)	B(d)
F13	Jianyang, Fujian	ZYD04879	I (2)	B(d)	N5	Suning, Hebei	ZYD02755	I (3)	B(d)
F14	Jianyang, Fujian	ZYD04898	I (2)	B(d)	N6	Yangcheng, Shanxi	ZYD03120	I (4)	B(d)
F15	Shaowu, Fujian	ZYD04910	I (2)	B(d)	N7	Lijing, Shandong	ZYD03262	II	B(d)
F16	Shaowu, Fujian	ZYD04916	I (2)	B(d)	N8	Heshui, Gansu	ZYD03298	I (3)	B(c)
F17	Zhouning, Fujian	ZYD04921	I (2)	B(d)					

<sup>1</sup>: UPGMA Clustering by SSR data, <sup>2</sup>: UPGMA Clustering by SRAP data.

netic distance and differentiation, and clustering analyzing.

SSR: The PCR reaction was performed in a 20- $\mu$ l reaction mixture containing 2  $\mu$ l of template DNA (50 ng), 2  $\mu$ l of 1  $\times$  PCR buffer with MgCl<sub>2</sub>, 2  $\mu$ l of 2.5 mM dNTPs, 2  $\mu$ l of 1.25  $\mu$ M forward and reverse primers, and 1U Taq DNA polymerase. The amplification was started with 5 min at 94°C followed by 35 cycles of 30 s at 94°C, 30 s at 52°C, 30 s at 72°C, and 10 min final extension at 72°C. The actual annealing temperature was adjusted by the different primers.

SRAP: The PCR reaction was performed in a 20- $\mu$ l

reaction mixture containing 2  $\mu$ l of template DNA (50 ng), 2  $\mu$ l of 1  $\times$  PCR buffer with MgCl<sub>2</sub>, with MgCl<sub>2</sub>, 2  $\mu$ l of 2.5 mM dNTPs, 2  $\mu$ l of 3  $\mu$ M forward and reverse primers, and 1U Taq DNA Polymerase. The amplification was started with 5 min at 94°C followed by 5 cycles of 1min at 94°C, 1min at 35°C, 1min at 72°C, and then 35 cycles of 1min at 94°C, 1min at 50°C, 1min at 72°C, and 7 min final extension at 72°C.

Amplified products were fractionated by electrophoresis in 6% denaturing polyacrylamide gels and stained with silver staining. The size of the stained band was analyzed based on its migration distance relative to the

50 bp ladder DNA Marker (Dongsheng biotechnology Ltd.).

### 2.2.2. Data Analysis

Banding profiles generated by SSR and SRAP markers were compiled into a data binary matrix based on the presence (1) or absence (0) of the selected band. Every primer pair or combination as a locus and every variant (band) as an allele were counted.

Genetic diversity was calculated using Simpson genetic diversity index ( $H_i$ ) and Shannon-Weaver index ( $I_i$ ).  $H_i = 1 - \sum p_{ij}^2$ ;  $I_i = -\sum p_i \ln p_i$ , where  $p_i$  is the frequency of the  $i$ th allele;  $p_{ij}$  is the frequency of  $j$ th allele at  $i$ th locus. Average diversity index ( $H_i, I_i$ ) was computed with  $H_i = \sum H_i/k$ ;  $I_i = \sum I_i/k$  where  $k$  means the total number of loci in the study. For SSR, number of alleles ( $A$ ) and rare alleles  $\textcircled{R}$  are counted for each locus, where  $A = \sum A_i$ ;  $A_i$  is the number of alleles of the  $i$ th locus in a population;  $A$  is the total number of alleles of all loci in a population. And for SRAP, number of polymorphism bands ( $P$ ) and rare variation  $\textcircled{R}$  for each primer combination are also counted.

The genetic similarity matrix was used in cluster analyses with the unweighted pair-group method of the arithmetic average (UPGMA) under NYSYS-pc 2.1 (Exeter Software, NY, USA) and the resulting clusters were expressed as dendrograms. To examine the genetic relationship among populations, Nei's [33] genetic distance was generated by GenAlEx 6 [34] and a dendrogram was constructed with the unweighted pair-group method of averages (UPGMA). An analysis of Molecular Variance (AMOVA) was used to detect the population differentiation and was also calculated under the GenAlEx6. F-statistic were tested by 1000 permutations and significant differences between populations declared if measured variance is lower than 95% of the variance in the null distribution [35]. A Mantel test [36] was performed to estimate a correlation between the matrices of Nei's [33] genetic distances generated by SSR and SRAP data using GenAlEx 6 (1000 permutations).

## 3. Results

### 3.1. Genetic Diversity and Allelic Richness by SSR and SRAP Markers

Forty one SSR loci produced a total of 421 alleles in 149 wild soybeans with a mean of 10.27 per locus and a range from 5 to 16. The wild soybeans from Fujian province had the highest allelic richness with an average of 8.8 alleles per locus, and that from Guangxi province had the lowest allelic richness with an average of 6.85 alleles per locus in the three Southern wild soybean populations (Table 2). 18 SRAP primer combinations detected a total of 90 polymorphic bands in whole wild

soybean population with a mean of 5 per primer combinations. Hunan subpopulation was amplified the most polymorphic bands ( $P = 86$ ) in 4 subpopulations (Table 3). The two mean genetic diversity index based on SSR data showed the different tendency that Hunan subpopulation had the highest genetic diversity in the Simpson index  $H_i = 0.801$  (Fujian:  $H_i = 0.791$ ); while Fujian subpopulation had the highest genetic diversity in Shannon-Weaver index  $I_i = 1.837$  (Hunan:  $I_i = 1.805$ ), however in comparison, the different degree of  $I_i$  values between Fujian and Hunan subpopulations (0.32) was much higher than that of  $H_i$  (0.01) (Table 2). The mean genetic diversity index based on 18 SRAP markers showed the same tendency as SSR, that was Fujian subpopulation was detected the highest genetic diversity ( $I = 0.803$ ) in whole population (Table 3). In addition, the data of gene diversity index showed there was more genetic diversity detected by SSR markers ( $I_i = 1.987$ ) than by SRAP markers ( $I_i = 0.918$ ), but due to the number of SRAP primer combination was relatively fewer, this result further needed to be verified.

The number of rare alleles of Fujian subpopulation based on SSR data was the largest with an average of 0.68 per locus, and that of Guangxi subpopulation was the least (0.20) in three southern China populations. Although the number of wild soybean of Northern China in this article was only 8, there were still found a total of 4 rare alleles in the population. It was worth mentioning that there had 4 (of 6) rare alleles at Satt279 on LG H in wild soybean of Hunan were also detected in cultivated soybean of Hunan in our another research, and at this locus was not any rare allele detected in another 3 wild soybean populations, therefore we speculated some new variations should have arisen only in cultivated and wild soybean of Hunan province (Table 2). Different from SSR results, the number of rare variations detected in Hunan subpopulation was the most in 4 wild soybean populations (Table 3).

## 3.2. Genetic Relationships

### 3.2.1. UPGMA Clustering Analysis of 149 Wild Soybeans by SSR

The SSR profile was used to determine the genetic similarity matrices, which was then used to construct dendrogram by the UPGMA method. The UPGMA-derived dendrogram assigned the 149 wild soybeans into six major clusters designated as: "I", "II", "III", "IV", "V", "VI" at the Similarity Coefficient of 0.8. Cluster "I" consisted of 93 wild soybeans including nearly a quarter of wild soybeans from Hunan province (16) and the vast majority of wild soybeans from Fujian (46) and about all wild soybeans from Guangxi province (26) (Table 1). For further analyzing the relationships of wild soybeans

**Table 2. Number of alleles and genetic diversity index for each SSR primer detected in 4 wild soybean subpopulations.**

Locus	LG	Whole population				Hunan (58)				Fujian (56)				Guangxi(27)				Northern China (8)			
		A	$H_i$	$I_i$		A	R	$H_i$	$I_i$	A	R	$H_i$	$I_i$	A	R	$H_i$	$I_i$	A	R	$H_i$	$I_i$
Satt300	A1	9	0.793	1.872	9	0	0.757	1.759	8	0	0.757	1.730	6	0	0.754	1.672	6	0	0.781	1.667	
Satt236	A1	12	0.861	2.128	8	0	0.847	1.898	9	2	0.813	1.908	7	2	0.693	1.524	4	0	0.719	1.321	
Satt429	A2	7	0.797	1.635	5	0	0.729	1.256	6	1	0.665	1.356	5	0	0.705	1.373	3	0	0.531	0.900	
Satt197	B1	13	0.862	2.280	12	0	0.879	2.249	9	0	0.733	1.716	9	0	0.829	1.952	4	0	0.656	1.213	
Satt415	B1	8	0.808	1.829	7	0	0.736	1.575	7	1	0.829	1.838	7	0	0.785	1.727	4	0	0.656	1.213	
Satt577	B2	9	0.783	1.888	8	0	0.705	1.631	9	1	0.712	1.730	5	0	0.745	1.501	3	0	0.406	0.736	
Satt556	B2	11	0.810	1.995	8	0	0.702	1.708	11	2	0.852	2.150	6	0	0.759	1.476	6	0	0.813	1.733	
Satt194	C1	8	0.831	1.678	7	0	0.824	1.519	8	0	0.830	1.795	7	0	0.691	1.457	3	0	0.828	0.888	
Satt371	C2	8	0.760	1.661	7	1	0.806	1.744	7	1	0.642	1.367	6	0	0.743	1.542	3	0	0.531	0.900	
Satt281	C2	16	0.928	2.546	15	0	0.931	2.250	13	0	0.876	2.279	13	0	0.904	2.538	7	0	0.844	1.906	
Satt184	D1a	11	0.851	2.248	11	1	0.780	2.171	10	0	0.843	2.095	7	0	0.776	1.698	6	0	0.781	1.667	
Satt408	D1a	11	0.849	2.082	10	1	0.835	2.050	9	0	0.837	1.931	6	0	0.678	1.427	5	0	0.750	1.494	
Satt005	D1b	14	0.898	2.343	13	1	0.912	2.265	12	0	0.833	2.119	11	0	0.815	2.045	6	0	0.781	1.667	
Satt216	D1b	10	0.869	2.119	10	1	0.830	1.981	9	0	0.870	2.098	6	0	0.796	1.652	4	0	0.656	1.213	
Satt002	D2	14	0.871	2.280	13	2	0.863	2.253	10	1	0.845	2.019	10	0	0.855	2.164	5	0	0.750	1.494	
Satt226	D2	11	0.844	2.049	9	1	0.871	1.978	10	2	0.805	2.008	5	0	0.668	1.331	4	0	0.719	1.321	
Satt230	E	10	0.833	1.867	8	0	0.844	1.949	9	1	0.786	1.629	8	0	0.739	1.470	4	0	0.688	1.255	
Satt268	E	6	0.766	1.491	5	0	0.799	1.518	5	1	0.720	1.358	4	0	0.664	1.112	5	0	0.750	1.494	
Satt586	F	11	0.893	2.238	11	1	0.813	2.198	9	0	0.878	1.889	7	0	0.845	1.748	4	0	0.563	1.074	
Satt218	F	9	0.743	1.699	8	1	0.686	1.527	7	0	0.593	1.278	5	0	0.691	1.351	3	0	0.531	0.900	
Satt334	F	7	0.846	1.787	7	1	0.795	1.617	6	0	0.798	1.653	5	0	0.833	1.508	5	0	0.781	1.560	
Satt309	G	7	0.750	1.604	4	0	0.626	1.098	7	1	0.763	1.712	3	0	0.491	0.780	4	0	0.688	1.255	
Satt352	G	12	0.788	1.939	8	1	0.688	1.566	9	1	0.834	1.920	8	1	0.689	1.568	3	0	0.813	1.040	
Satt279	H	14	0.880	2.335	14	6	0.885	2.450	8	0	0.777	1.673	8	0	0.731	1.707	5	0	0.797	1.408	
Sat_214	H	12	0.878	2.292	12	0	0.882	2.302	11	0	0.821	2.024	9	0	0.868	2.028	6	0	0.813	1.733	
Satt239	I	10	0.883	2.118	9	0	0.789	1.779	10	1	0.878	2.097	8	0	0.853	1.986	4	0	0.703	1.126	
Sct_189	I	14	0.909	2.157	7	2	0.815	1.280	12	4	0.885	2.205	6	0	0.793	1.681	6	0	0.813	1.733	
Sct_001	J	5	0.747	1.424	4	0	0.685	1.206	5	0	0.735	1.370	5	0	0.794	1.534	5	0	0.750	1.494	
Satt414	J	10	0.808	2.085	9	0	0.787	1.929	9	1	0.805	2.058	7	0	0.660	1.694	4	0	0.719	1.321	
Satt431	J	9	0.842	1.917	8	1	0.837	1.832	7	0	0.779	1.705	7	0	0.815	1.797	5	1	0.750	1.494	
Satt588	K	15	0.916	2.600	15	0	0.917	2.577	14	0	0.896	2.458	10	0	0.853	2.077	7	0	0.844	1.906	
Satt242	K	7	0.386	1.310	6	0	0.679	0.937	7	0	0.423	1.017	5	0	0.715	1.223	3	0	0.656	1.082	
Satt001	K	9	0.892	1.820	7	0	0.897	1.325	9	2	0.870	1.977	4	0	0.486	0.864	4	0	0.719	1.321	
Satt373	L	8	0.813	1.874	7	0	0.818	1.838	7	0	0.789	1.742	6	1	0.695	1.585	6	0	0.781	1.667	
Satt590	M	12	0.869	2.192	9	0	0.839	1.924	10	1	0.850	2.127	8	2	0.823	1.870	6	0	0.813	1.733	
Satt346	M	9	0.785	1.716	7	0	0.756	1.580	8	2	0.734	1.611	6	0	0.654	1.402	5	0	0.750	1.494	
Satt387	N	8	0.853	2.051	8	0	0.818	1.909	8	0	0.811	1.864	8	0	0.830	1.980	4	0	0.719	1.321	
Satt339	N	16	0.893	2.415	12	1	0.880	2.278	12	1	0.866	2.175	9	1	0.808	1.970	6	1	0.813	1.733	
Satt243	O	9	0.842	1.994	7	1	0.812	1.848	7	0	0.843	1.927	5	1	0.749	1.455	5	0	0.781	1.560	
Satt345	O	12	0.882	2.381	10	0	0.831	2.020	11	0	0.867	2.264	9	0	0.833	2.189	5	1	0.672	1.515	
Satt259	O	8	0.714	1.517	5	0	0.667	1.237	7	1	0.676	1.459	5	0	0.362	1.499	5	1	0.750	1.494	
Total		421			359	23			361	28			281	8			192	4	0.723	1.391	
Average		10.27	0.825	1.987	8.76	0.56	0.801	1.805	8.80	0.68	0.791	1.837	6.85	0.20	0.743	1.638	4.68	0.10	0.781	1.667	

A: Number of alleles at each locus; R: Number of rare allele at each locus;  $H_i$ : Simpson diversity index for each locus;  $I_i$ : Shannon-weaver diversity index for each locus.

**Table 3. Genetic diversity of wild soybean populations for 18 SRAP primer combinations.**

Primer	Whole (149)		Hunan (58)			Fujian (56)			Guangxi (27)			Northern China (8)		
	$I_i$	P	$I_i$	P	R	$I_i$	P	R	$I_i$	P	R	$I_i$	P	R
me1-em2	1.407	5	1.290	5	1	1.027	3	0	1.087	4	0	0.974	3	0
me1-em3	1.015	4	1.015	4	0	0.983	3	0	0.587	3	0	1.321	4	0
me1-em4	1.224	5	0.911	4	0	1.180	5	1	0.193	4	0	0.996	4	0
me1-em5	1.282	6	0.502	5	0	0.942	4	0	1.147	5	0	0.347	3	0
me1-em6	0.869	6	1.093	6	0	0.883	6	0	0.671	5	0	0.693	5	0
me1-em7	0.318	2	0.356	2	0	0.353	2	0	0.283	2	0	0.260	2	0
me2-em1	0.765	4	0.602	4	0	0.774	4	0	0.349	3	0	0.822	3	0
me2-em7	1.288	6	1.088	5	0	2.030	6	0	1.366	6	0	1.248	5	0
me4-em4	0.249	2	0.187	2	0	0.274	2	0	0.105	2	0	0.117	2	0
me5-em7	0.490	2	0.466	2	0	0.542	2	0	0.360	1	0	0.476	2	0
me6-em6	1.173	4	0.408	4	0	1.064	4	0	0.985	3	0	1.247	4	0
me7-em7	1.034	4	0.436	4	0	1.224	4	0	0.981	4	0	0.723	4	0
me8-em6	0.857	7	0.613	7	0	0.366	6	0	0.681	7	0	0.368	5	0
me12-em2	0.247	1	0.321	1	0	0.356	1	0	0.327	1	0	0.000	1	0
me15-em5	0.650	8	0.771	7	0	0.392	6	0	0.122	5	0	0.723	6	0
me20-em10	1.109	8	0.789	8	2	0.391	6	0	0.801	6	0	0.736	6	0
me21-em1	1.638	6	0.913	6	0	1.350	6	0	1.406	6	0	1.518	6	0
me22-em1	0.905	10	1.085	10	1	0.330	9	0	0.193	9	0	0.000	8	0
Total		90		86	4		79	1		76	0		73	0
Average	0.918	5	0.714	4.8		0.803	4.4		0.647	4.2		0.698	4	

$I_i$ : Shannon-weaver index for each primer combination; P: Number of polymorphism band for each primer combination; R: Number of rare variation for each primer combination.

from different regions, Cluster ‘ I ’ was further classified into 8 subgroups. Subgroup 1 consisted of 3 wild soybeans from Xintian county of Southern Hunan; Subgroup 2 consisted of 41 wild soybeans including 37 from Fujian, 3 from Southern Hunan and 1 from Liaoning of Northern China; Subgroup 3 consisted of 12 wild soybeans in which 8 from Hunan, 2 and 1 respectively from Hebei and Gansu of Northern China, and 1 from Fujian; Subgroup 4 consisted of 21 wild soybeans in which 20 from Guangxi and 1 from Shanxi of Northern China; Subgroup 5 consisted of 3 wild soybeans from Fujian; Subgroup 6 consisted of 5 wild soybeans including 3 from Guangxi, 1 from Fujian and 1 from Hunan; Subgroup 7 consisted of 7 wild soybeans including 3 from Guangxi, 4 from Fujian; Subgroup 8 consisted of only 1 wild soybean from Hunan. From above analysis, the wild soybean from Southern Hunan gathered in subgroup 1 and 2, and another 8 from Hunan including the northern, western and central region were gathered in subgroup 3, besides having two from central region of Hunan was clustered in subgroup 6 and 8; all wild soybeans from Guangxi except for one from Guanyang were clustered in Cluster ‘ I ’ and most of them gathered in subgroup 4; while in this cluster, the wild soybean germplasm from Fujian

gathered in subgroup 2, 3, 5, 6, 7, indicated the highest distribution.

Cluster ‘ II ’ consisted of 22 wild soybeans including 15 from Hunan, 5 from Fujian, 1 from Guangxi and 1 from Shandong of Northern China in which 5 from Southwestern, 1 from Sha county of Fujian and 1 from Shandong of Northern clustered together; 3 from Western Hunan and 1 from Changsha of Eastern Hunan clustered together.

Cluster ‘ III ’ included 13 wild soybeans from Hunan and 5 from Fujian. In which, 6 from Southern Hunan, 2 from Northeastern Hunan and 1 from Western Hunan clustered in a small group.

Cluster ‘ IV ’ included 8 wild soybeans from Hunan and 2 from Liaoning of Northern China in which 7 from Southern Hunan and 1 from Liuyang of Eastern Hunan.

Cluster ‘ V ’ consisted of 5 wild soybeans all from Hunan including 3 from Northeast region, 1 from Western region and 1 from Easter region. Cluster ‘ VI ’ included only 1 wild soybean from Western Hunan. Through analyzing the dendrogram based on SSR we found the wild soybean from Guangxi mainly gathered in Cluster ‘ I ’; the wild soybean from Hunan mainly gathered in the lower part of the dendrogram including Clus-

ter “II” to “V”; as for the wild soybean population of Fujian, in which 46 gathered in Cluster “I”, 10 gathered in Cluster “II” to “III”, in spite of a few of wild soybeans gathered in the lower part of the dendrogram, their distribution was very widely. Therefore, we thought that Fujian could be a center of genetic diversity of wild soybean in Southern China. And the wild soybean from Fujian and Guangxi had closer genetic distance than that from Hunan, in spite of the geographic distance between Guangxi and Hunan was smaller than that between Guangxi and Fujian. Furthermore, the wild soybeans from Liaoning showed higher genetic distribution than others from Northern China in the study.

### 3.2.2. UPGMA Clustering Analysis of 149 Wild Soybeans by SRAP

The clustering results of UPGMA of 149 wild soybeans by SRAP were very different with that by SSR, in which the wild soybeans from Hunan all clustered together in a Cluster alone while they were distributed in all major Clusters based on SSR. The 149 wild soybeans were distinctly assigned into two major clusters designated as “A” and “B” by the UPGMA-derived dendrogram based on SRAP (Table 1). Cluster “A” consisted of 58 wild soybeans which all from Hunan, and was further classified into two subgroups named as “a” and “b”. Subgroup “a” consisted of 36 wild soybeans including 20 from Northern region of Hunan mainly located on between latitude 29° and 30°, 8 from Xintian of Southern Hunan, 4 from Western Hunan, 4 from Eastern region of Hunan; Subgroup “b” consisted of 22 wild soybeans in which 13 from Southern Hunan, 6 from Southwestern Hunan and 3 from Eastern Hunan. The same with the results based on SSR, in spite of the number of materials was few, the wild soybeans from Eastern Hunan were assigned in two different groups, which suggested the wild soybean from eastern region of Hunan including Changsha and Xiangtan showed higher genetic distribution.

Cluster “B” consisted of wild soybeans respectively from Fujian, Guangxi and Northern China, and could be further classified into three subgroups designated as “c”, “d” and “e” in which 27 wild soybeans from Guangxi all gathered in subgroup “e”; among 8 the wild soybean germplasm from Northern China, 1 from Gansu gathered in subgroup “c”, 6 (2 from Liaoning, 2 from Hebei, 1 from Shanxi and 1 from Shandong) gathered in subgroup “d”, and 1 from Liaoning was clustered in subgroup “e”; the wild soybeans from Fujian were cluster in three subgroups, in which 4 gathered in subgroup “c”, 35 in subgroup “d” and 17 in subgroup “e”.

Similarly, the wild soybeans from Fujian and Guangxi were clustered together in one Cluster, and the genetic distance of wild soybeans of Hunan was far from them.

### 3.3. Genetic Differentiation and Distance among the Four Wild Soybean Populations

The values of Nei Genetic Distance revealed that the genetic distance between Hunan and Fujian wild soybean populations was smallest (0.329), and that between the Northern China and 3 Southern wild soybean populations was farer based on SSR analysis in the study (Table 4).

On the other hand, the analysis of molecular variance (AMOVA) was used to partition the SSR and SRAP variation among and within 4 wild soybean populations. Most of variation (SSR 94%, SRAP 73%) was detected within the four populations, while only a smaller but significant part of the variation (SSR 6%, SRAP 27%  $P < 0.001$ ) was attributed to variation among subpopulations (Tables 5 and 6), which indicated that geographic differentiation played an important role in genetic variation. In addition, the AMOVA by SRAP showed much more variation among subpopulations (27%) than that by SSR in the study.

#### UPGMA Clustering Analysis Based on Genetic Distance of the Four Subpopulations

The dendrogram based on the genetic distance by SSR

**Table 4. Genetic Distance (*D*) among four wild soybean populations.**

Pop1	Pop2	<i>D</i> (SSR)	<i>D</i> (SRAP)
Hunan	Fujian	0.329	0.108
Fujian	Guangxi	0.406	0.037
Hunan	Guangxi	0.488	0.134
Fujian	Northern China	0.500	0.036
Hunan	Northern China	0.554	0.117
Guangxi	Northern China	0.675	0.068

*D*: Nei’s Genetic Distance among four subpopulations in the study.

**Table 5. Analysis of molecular variance (AMOVA) based on SSR data.**

Source variation	df	Sum of square	Mean square	% variation	<i>P</i>
Among Populations	3	541.045	180.348	6%	$P < 0.001$
Within Populations	145	8578.955	59.165	94%	$P < 0.001$

**Table 6. Analysis of molecular variance (AMOVA) based on SRAP data.**

Source variation	df	Sum of square	Mean square	% variation	<i>P</i>
Among Populations	3	337.812	112.604	27%	$P < 0.001$
Within Populations	145	1236.007	8.524	73%	$P < 0.001$



and SRAP among the four wild soybean populations illustrated the genetic relationship between the Hunan and Fujian wild populations was nearest, and the wild soybean population from Northern China was finally clustered with another three wild populations from Southern China at the  $D = 0.56$  based on SSR (**Figure 1**). The dendrogram by SRAP showed the very different results that the Fujian, Guangxi and Northern China subpopulations clustered together firstly, then clustered with Hunan wild soybean population at  $D = 0.93$  (**Figure 2**).

### 3.4. Mantel Test between SSR and SRAP

Mantel test between distances matrices of SSR and SRAP indicated that there was a positive linear correlation between the two elements ( $r = 0.25$ ), and which was a weak correlation.

## 4. Discussion

### 4.1. Genetic Diversity of Wild Soybean in China

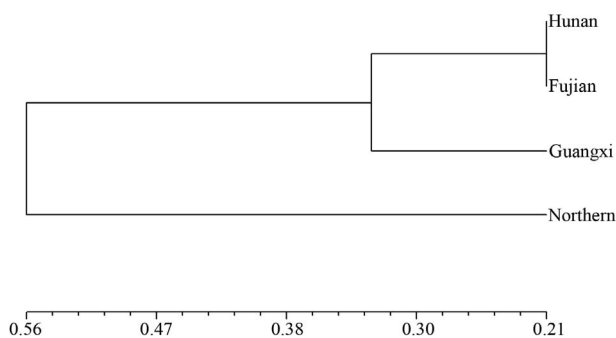
The genetic diversity of wild soybean from some provinces of South China had been analyzed in many studies. Zeng *et al.* [37] analyzed the genetic diversity of 200 wild soybean accessions collected from Guangxi by 23 pairs SSR primers, with an average of Simpson index was 0.689 and Shannon-weaver was 0.683. Liu *et al.* [38]

investigated that the distribution and characters of wild soybean in 41 counties of Hunan province and found the germplasm resources were very rich with vast geographic distribution and many populations. In this study, we analyzed the genetic richness, genetic diversity, and rare variants of annual wild soybean populations mainly from southern China including Hunan, Fujian, Guangxi, and some provinces of Northern China by SSR and SRAP markers. We found that the wild soybean from Fujian province showed the highest genetic richness, the largest genetic diversity index and the most number of rare variants in the four populations based on SSR data in the study, and genetic diversity index by SRAP also showed Fujian subpopulation had the highest genetic diversity in whole population. In the dendrogram, the wild soybean accessions from Fujian showed the highest genetic distribution in different clusters by SSR clustering. In addition, we also found the wild soybeans from Guangxi clustered with those from Fujian in dendrograms based on SSR and SRAP data, although Guangxi is closer to Hunan than to Fujian. Cheng *et al.* [39] analyzed 192 wild soybean genotypes from Jiangxi province in Southern China using 48 SSR markers shared 73% SSR markers with our study, and the results showed the average polymorphism information content (PIC, or Simpson index) was 0.706, which was also lower than genetic diversity of Fujian subpopulation (Simpson index = 0.791) in this study.

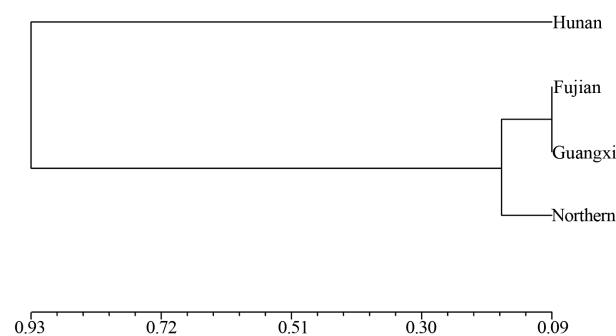
Therefore, based on the above results, we thought that Fujian province especially 25° north latitude region could be one of centers of the genetic diversity of wild soybean in Southern China. Zhuang *et al.* [19] also proposed that the southern China, especially 25° north latitude region was worth noting in the study of origin of soybean because they found that the similarity between the DNA fingerprints of wild soybean and cultivated soybean in southern China was higher than that in northern China. It was also worth mentioning that SRAP markers produced the most polymorphic bands (86) and rare variation site (4) in Hunan subpopulation than another 3 subpopulations, which might be due to the types of phenotype of Hunan wild soybeans was more richer than that of Fujian wild soybeans in the study.

### 4.2. Geographic Differentiation and Distribution

In the long-term evolution of wild soybean, different wild soybean populations were formed to adapt to the special ecological environment. We detected some specific-present alleles in different wild soybean populations from Hunan, Fujian, Guangxi and some regions of Northern China. In population genetics, if an allele is detected in only one subpopulation in whole population, the allele is called rare allele [40]. The results showed the wild soybean population from Fujian was detected the



**Figure 1. Dendrogram of Nei's genetic distance among subpopulations based on SSR data.**



**Figure 2. Dendrogram of Nei's genetic distance among subpopulations based on SRAP data.**

most rare alleles (28) at 41 SSR loci, and even if Northern China population included only 8 core germplasm was detected four rare alleles. We also found there were 4 rare alleles detected at Satt279 on LG H in wild population from Hunan also detected in cultivated soybean population from Hunan in our other research, and not rare allele was found in Fujian, Guangxi and Northern China wild populations at the locus. There were also some rare variations detected in wild soybean populations using 18 SRAP primer combinations, in which Hunan subpopulation had the most number of rare variations with a total of 4 rare variations and Fujian subpopulation was also detected one rare variation. In the light of neutral theory, Kimura [41] thought that polymorphic alleles were predominantly neutral, rare variant alleles could include slightly deleterious and sometimes even definitely deleterious alleles in addition to selectively neutral ones, which indicated that rare variant allele reflect the total mutation rate much more faithfully than polymorphic alleles. In addition, for extremely large populations undergoing sudden environmental change, a rare allele of major effect contribute to adaptation to sustained intense directional selection, and restricted population size may be elimination of advantageous rare alleles [42], which suggests that for the certain population, the number of rare alleles reflects the richness of genetic diversity to some extent and the ability to adaptation in the local ecological environment. Therefore we suggested that some special variations at unique locus might be arisen in specific ecological populations including wild and cultivated soybean for adapting to local environment, and geographic differentiation should have its relevant genetic bases. The variation among different provinces accounted for 6% and 27% of the total variation respectively by SSR and SRAP, which were less than the intra-subpopulations variation. Wen *et al.* [11] found 2.7% variation obtained among subpopulations of annual wild soybean in China based on SSR data, and thought the number of mutant alleles due to geographic differentiation should be relatively small since all the alleles were formed and accumulated during the long systematic development of the species.

The relationships among subpopulations by UPGMA clustering showed difference results between SSR and SRAP, in the SSR clustering, the wild soybean population from Hunan was clustered with that from Fujian firstly, then with Guangxi subpopulation, and finally with Northern China subpopulation; while in the SRAP clustering, Fujian subpopulation was clustered with Guangxi subpopulation firstly, secondly with Northern China population, finally with Hunan subpopulation. But from UPGMA clustering dendrograms of all 149 wild soybean accessions in the study, we found whether in SSR or SRAP clustering results, the wild soybeans from Hunan

showed farer genetic relationship with Fujian and Guangxi wild soybeans, and the wild soybeans from Guangxi all clustered with wild soybeans from Fujian. The present results of genetic relationship among different provinces in southern China showed the genetic distance between Fujian and other two provinces (Hunan and Guangxi) in southern China was closer. Many researchers thought South China was the center of origin of wild and cultivated soybean [20,22,11], while the soybean of various regions in South China had different genetic diversity, through our study the wild soybean from Fujian province showed highest genetic diversity and rare variations, at the same time, the wild soybeans from Fujian showed highest genetic distribution in clustering dendrograms, and had nearer genetic distance to Hunan and Guangxi, which suggested that Fujian province region could be one of center of diversity of annual wild soybean in southern China and could be a center of origin of wild soybeans, but because of the wild soybean materials in the study only from a few of region in southern China, so we should do further research about the origin of wild soybean of southern China. Liu *et al.* [43] also proposed that the photoperiod and temperature reaction of wild soybean in Fujian was as same as that of the autumn cultivated soybean, and speculated northwest of Fujian might be the origin of cultivated soybean.

#### 4.3. Comparison between SSR and SRAP Markers

SSRs (Simple Sequence Repeats), are uniformly distributed on the eukaryotic genome, and are often found at frequencies much higher than would be predicted purely on the grounds of base composition. SSRs are inherited by codominant Mendelian pattern, and mainly located in non-coding region of genome. While SRAP (Sequence-related amplified polymorphism) are explored for amplification of open reading frames (ORFs), which are a portion of a gene sequence and contain a base sequence can be encoded protein. SRAP targets coding sequences in the genome and results in a moderate number of codominant markers [44]. Due to SSRs and SRAPs amplified distinct region in the genome, we used the two molecular markers and analyzed the genetic diversity and relationship of wild soybean accessions in different provinces through more comprehensive point of view.

Mantle test indicated that there were weak correlation between SSR and SRAP ( $r = 0.25$ ). Gai *et al.* [30] found, genetic distance estimated based on SSR and SRAP data were significantly correlated with each other, and similar clustering results were achieved through investigating 19 maize varieties, and SRAP markers displayed more genetic variation, diversity and higher discriminative power in varieties identification in Maize. In the study, we also found that there were some similarities between SSR and

SRAP result, for instance, the genetic diversity index all showed the genetic diversity of Fujian subpopulation was highest in the four subpopulations in the study; and the clustering results of SSR and SRAP data all showed the gene relationship of wild soybeans from Hunan was far from the other two subpopulations in Southern China including Fujian and Guangxi; moreover the two semi-wild soybeans from Hunan (H46 and H47) which 100-seed weight = 3 g clustered together whether in SSR clustering or in SRAP clustering. In addition there were some difference in the clustering results that Hunan subpopulation clustered alone in a group had farthest genetic distance from the other wild soybean subpopulation from Southern China, and the wild soybeans from Northern China were clustered with Fujian wild soybeans in a group in SRAP clustering; while in SSR clustering course, Hunan subpopulation was clustered with Fujian subpopulation, and Northern China subpopulation showed the farthest genetic relationship with the another three subpopulations from Southern China. And genetic diversity index ( $I = 0.918$ ) by using SRAP was lower than that by using SSR ( $I = 1.995$ ), which could be caused by the relatively small number of SRAP primer combinations

Because of the amplification section was almost different between SSR and SRAP, the discrepancy of clustering should exist between SSR and SRAP. Gai *et al.* [30] proposed the hybrid purity identified by SRAP marker was more vicinal to field test than that by SSR marker. We should do further investigation to verify which marker was closer to the phenotypes using more SRAP primer combinations. At present, we suggested that using SSR and SRAP markers could reflect the whole genome variation more comprehensively, and the analysis results were more reliable.

## REFERENCES

- [1] F. S. Li, "Geographical Distribution and Ecological Diffusing of Wild Soybean Germplasm in China," In: F. S. Li, Ed., *Researching Progress on Wild Soybean of China*, Chinese Agricultural Press, Beijing, 1994, pp. 40-48.
- [2] Y. L. Ding, T. J. Zhao and J. Y. Gai, "Genetic Diversity and Ecological Differentiation of Chinese Annual Wild Soybean (*Glycine soja*)," *Biodiversity Science*, Vol. 16, No. 2, 2008, pp. 133-142.
- [3] B. C. Zhuang, H. Xu and Y. M. Wang, "Polymorphism and Geographical Distribution of the Stem and Leaf Characters of Wild Soybean (*Glycine soja*) in China," *Acta Agronomica Sinica*, Vol. 22, No. 5, 1996, pp. 583-586.
- [4] Y. S. Dong, B. C. Zhuang, L. M. Zhao, H. Sun and M. Y. He, "The Genetic Diversity of Annual Wild Soybeans Grown in China," *Theoretical and Applied Genetics*, Vol. 103, No. 1, 2001, pp. 98-103. [doi:10.1007/s001220000522](https://doi.org/10.1007/s001220000522)
- [5] K. J. Wang, X. H. Li and F. S. Li, "Phenotypic Diversity of the Big Seed Type Subcollection of Wild Soybean (*Glycine soja* Sieb. et Zucc.) in China," *Genetic Resources and Crop Evolution*, Vol. 55, No. 8, 2008, pp. 1335-1346. [doi:10.1007/s10722-008-9332-z](https://doi.org/10.1007/s10722-008-9332-z)
- [6] J. Li, Y. Tao, S. Z. Zheng and J. L. Zhou, "Isozymatic Differentiation in Local Population of *Glycine soja* Sieb & Zucc.," *Acta Botanica Sinica*, Vol. 37, 1995, pp. 669-676.
- [7] Z. A. Hu, M. Zhong, W. Wei and H. X. Wang, "Detection of DNA Diversity of Wild Soybean (*Glycine soja*) in Natural Populations by New Procedure of RAPD and RFLP," *Soybean Genetics Newsletter*, Vol. 24, 1997, pp. 37-38.
- [8] J. Abe, A. Hasegawa, H. Fukushi, T. Mikami, M. Ohara and Y. Shimamoto, "Introgression between Wild and Cultivated Soybeans of Japan Revealed by RFLP Analysis of Chloroplast DNAs," *Economic Botany*, Vol. 53, No. 3, 1999, pp. 285-291. [doi:10.1007/BF02866640](https://doi.org/10.1007/BF02866640)
- [9] Z. L. Li and R. L. Nelson, "RAPD Marker Diversity among Cultivated and Wild Soybean Accessions from Four Chinese Provinces," *Crop Science*, Vol. 42, No. 5, 2002, pp. 1737-1744. [doi:10.2135/cropsci2002.1737](https://doi.org/10.2135/cropsci2002.1737)
- [10] Y. W. Chen and R. L. Nelson, "Genetic Variation and Relationships among Cultivated, Wild, and Semiwild Soybean," *Crop Sciences*, Vol. 44, No. 1, 2004, pp. 316-325. [doi:10.2135/cropsci2004.0316](https://doi.org/10.2135/cropsci2004.0316)
- [11] Z. X. Wen, Y. L. Ding, T. J. Zhao and J. Y. Gai, "Genetic Diversity and Peculiarity of Annual Wild Soybean (*G. soja* Sieb. et Zucc.) from Various Eco-Regions in China," *Theoretical and Applied Genetics*, Vol. 119, No. 2, 2009, pp. 371-381. [doi:10.1007/s00122-009-1045-y](https://doi.org/10.1007/s00122-009-1045-y)
- [12] J. Yan, T. H. He and B. R. Lu, "Genetic Spatial Clustering: Significant Implications for Conservation of Wild Soybean (*Glycine soja*: Fabaceae)," *Genetica*, Vol. 128, No. 1-3, 2006, pp. 41-49. [doi:10.1007/s10709-005-4823-8](https://doi.org/10.1007/s10709-005-4823-8)
- [13] D. H. Xu and J. Y. Gai, "Genetic Diversity of Wild and Cultivated Soybeans Growing in China Revealed by RAPD Analysis," *Plant Breeding*, Vol. 122, No. 6, 2003, pp. 503-506. [doi:10.1046/j.0179-9541.2003.00911.x](https://doi.org/10.1046/j.0179-9541.2003.00911.x)
- [14] D. L. Hyten, Q. J. Song, Y. L. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht, R. C. Shoemaker and P. B. Cregan, "Impacts of Genetic Bottlenecks on Soybean Genome Diversity," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 103, No. 45, 2006, pp. 16666-16671. [doi:10.1073/pnas.0604379103](https://doi.org/10.1073/pnas.0604379103)
- [15] Y. Fukuda, "Cytogenetical Studies on the Wild and Cultivated Manchurian Soybeans (*Glycine* L.)," *Japanese Journal of Botany*, Vol. 6, 1933, pp. 489-506.
- [16] F. S. Li, "Study on Origin and Evolution of Soybean," *Soybean Science*, Vol. 13, No. 1, 1994, pp. 61-66.
- [17] T. Hymowitz and C. A. Newell, "Taxonomy of Genus *Glycine*, Domestication and Uses of Soybeans," *Economic Botany*, Vol. 35, No. 3, 1981, pp. 272-288. [doi:10.1007/BF02859119](https://doi.org/10.1007/BF02859119)
- [18] B. Xu, H. Xu, B. C. Zhuang, Q. H. Lu, Y. M. Wang and F. S. Li, "The Genetic Diversity and Geographic Distribu-

- tion of Kernel Traits of *G. soja* in China,” *Acta Agronomica Sinica*, Vol. 21, No. 6, 1995, pp.733-739.
- [19] B. C. Zhuang, D. W. Hui, Y. M. Wang, J. Gu, B. Xu and S. Y. Chen, “RAPD Analysis of Different Evolutionary Types in Different Latitude in China,” *Chinese Science Bulletin*, Vol. 39, No. 23, 1994, pp. 2178-2180.
- [20] J. Y. Gai, D. H. Xu and Z. Gao, “Studies on the Evolutionary Relationship among Eco-Types of *G. max* and *G. soja* in China,” *Acta Agronomica Sinica*, Vol. 26, No. 5, 2000, pp. 513-520.
- [21] S. L. Lv, “Discussion on the Original Region of Cultivated Soybean in China,” *Scientia Agricultura Sinica*, No. 4, 1978, pp. 90-94.
- [22] D. H. Xu, Z. Gao, Q. Z. Tian, J. Y. Gai, H. Fukushi, S. Kitajima, J. Abe and Y. Shimamoto, “Genetic Diversity of the Annual Wild Soybean (*Glycine soja*) in China,” *Chinese Journal of Applied & Environmental Biology*, Vol. 5, No. 5, 1999, pp. 439-443.
- [23] P. Awadalla and K. Ritland, “Microsatellite Variation and Evolution in the *Mimulus guttatus* Species Complex with Contrasting Mating Systems,” *Molecular Biology and Evolution*, Vol. 14, No. 10, 1997, pp. 1023-1034. doi:10.1093/oxfordjournals.molbev.a025708
- [24] Q. J. Song, L. F. Marek, R. C. Shoemaker, K. G. Lark, V. C. Concibido, X. Delannay, J. E. Specht and P. B. Cregan, “A New Integrated Genetic Linkage Map of the Soybean,” *Theoretical and Applied Genetics*, Vol. 109, No. 1, 2004, pp. 122-128. doi:10.1007/s00122-004-1602-3
- [25] G. Li and C. F. Quiros, “Sequence-Related Amplified Polymorphism (SRAP), a New Marker System Based on a Simple PCR Reaction: Its Application to Mapping and Gene Tagging in *Brassica*,” *Theoretical and Applied Genetics*, Vol. 103, No. 2-3, 2001, pp. 455-461. doi:10.1007/s001220100570
- [26] Y. C. Wen, H. Z. Wang, J. X. Shen, G. H. Liu and S. F. Zhang, “Analysis of Genetic Diversity and Genetic Basis of Chinese Rapeseed Cultivars (*Brassica napus* L.) by Sequence-Related Amplified Polymorphism Markers,” *Scientia Agricultura Sinica*, Vol. 39, No. 2, 2006, pp. 246-256.
- [27] C. Li, X. X. Lu, Z. E. Zhang, X. L. Chen, F. Lin, X. Y. Zhang and C. J. Liu, “Genetic Integrity Analysis of Cotton (*Gossypium hirsutum* Linn.) Accessions Using SRAP and SSR Markers,” *Journal of Plant Genetic Resources*, Vol. 8, No. 1, 2007, pp. 21-25.
- [28] F. Zhao, Z. J. Cai, T. Z. Hu, H. G. Yao, L. Wang, N. Dong, B. Wang, Z. G. Ru and W. X. Zhai, “Genetic Analysis and Molecular Mapping of a Novel Gene Confering Resistance to Rice Stripe Virus,” *Plant Molecular Biology Reporter*, Vol. 28, No. 3, 2010, pp. 512-518. doi:10.1007/s11105-009-0178-0
- [29] N. Mutlu, F. H. Boyacı, M. Göçmen and K. Abak, “Development of SRAP, SRAP-RGA, RAPD and SCAR Markers Linked with a Fusarium Wilt Resistance Gene in Eggplant,” *Theoretical and Applied Genetics*, Vol. 117, No. 8, 2008, pp. 1303-1312. doi:10.1007/s00122-008-0864-6
- [30] S. P. Gai, W. L. Gai and J. Y. Huang, “Comparison of SSR and SRAP Marker for Varieties Identification in Maize (*Zea mays* L.),” *Journal of Plant Genetic Resources*, Vol. 12, No. 3, 2011, pp. 468-472.
- [31] O. Gulsen, S. Karagul and K. Abak, “Diversity and Relationships among Turkish Okra Germplasm by SRAP and Phenotypic Marker Polymorphism,” *Biologia*, Vol. 62, No. 1, 2006, pp. 41-45. doi:10.2478/s11756-007-0010-y
- [32] H. Xie, “Genetic Diversity on Representative Samples from Primary Core Collection of Soybean (*G. max*) in China,” Ph.D. Thesis, Chinese Academy of Agricultural Sciences, Beijing, 2002.
- [33] M. Nei, “Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals,” *Genetics*, Vol. 89, 1978, pp. 583-590.
- [34] R. Peakall and P. E. Smouse, “GENALEX 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research,” *Molecular Ecology Notes*, Vol. 6, No. 1, 2006, pp. 288-295. doi:10.1111/j.1471-8286.2005.01155.x
- [35] L. Excoffier, P. Smouse and J. M. Quattro, “Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data,” *Genetics*, Vol. 131, No. 2, 1992, pp. 479-491.
- [36] N. A. Mantel, “The Detection of Disease Clustering and a Generalized Regression Approach,” *Cancer Research*, Vol. 27, 1967, pp. 209-220.
- [37] W. Y. Zeng, J. Liang, Y. Chen, Q. Y. Wei, F. Y. Tang, K. Z. Zhong and W. J. Chen, “Genetic Diversity Analysis of New Wild Soybean Collection in Guangxi,” *Soybean Science*, Vol. 30, No. 3, 2011, pp. 379-383.
- [38] H. Y. Liu, H. Zhang, X. Zhou, K. J. Wang and X. H. Li, “Preliminary Report on Distribution Status and Character Identification of Wild Soybean in Hunan,” *Hunan Agricultural Sciences*, No. 5, 2008, pp. 20-21.
- [39] C. M. Cheng, C. Y. Yang, Q. B. Ma and H. Nian, “Genetic Diversity Analysis of Wild Soybean Resources in Jiangxi,” *Journal of Plant Genetic Resources*, Vol. 12, No. 6, 2011, pp. 928-933.
- [40] G. B. Xu, “Plant Population Genetics,” 2th Edition, Science Press, Beijing, 2009.
- [41] M. Kimura, “Rare Variant Alleles in the Light of the Neutral Theory,” *Molecular Biology and Evolution*, Vol. 1, No. 1, 1984, pp. 84-93.
- [42] R. Lande, “Adaptation to an Extraordinary Environment by Evolution of Phenotypic Plasticity and Genetic Assimilation,” *Journal Evolutionary Biology*, Vol. 22, No. 7, 2009, pp. 1435-1446. doi:10.1111/j.1420-9101.2009.01754.x
- [43] D. J. Liu and S. C. Xu, “The Ecological Distribution and Classification on Wild Soybean in Fujian,” *Journal of Fujian Academy of Agricultural Sciences*, Vol. 6, No. 2, 1991, pp. 18-24.
- [44] M. Agarwal, N. Shrivastava and H. Padh, “Advances in Molecular Marker Techniques and Their Applications in Plant Sciences,” *Plant Cell Reports*, Vol. 27, No. 4, 2008, pp. 617-631. doi:10.1007/s00299-008-0507-z