

Genetic diversity and geographical peculiarity of Tibetan wild soybean (*Glycine soja*)

Ke-Jing Wang · Xiang-Hua Li

Received: 3 December 2010 / Accepted: 4 April 2011 / Published online: 15 May 2011
© Springer Science+Business Media B.V. 2011

Abstract Tibet is the highest Plateau in China and the world, however wild soybean has been found in its southeastern fringe region adjacent to the northwest of Yunnan Province. Tibetan wild soybean was distributed only in the Gongrigabuqu-River Gorge in southeast Tibet. This regional plant species belong to the flora of Himalayan-Hengduan Mountains, which comprises composite elements of ancient arcto-tertiary flora, tropical and semitropical flora. To date, few studies have been carried out on Tibetan wild soybean. Studying Tibetan wild soybean together with other regional ones helps to understand the history of the origin and dissemination of wild soybean species in China. Here we reported the status of genetic diversity in Tibetan wild soybean and the genetic relationship between Tibetan and other regional wild soybeans revealed by nuclear SSR markers. The results showed that the Tibetan wild soybean sample was significantly differentiated from other regional ones, as characterized by the lowest mean allelic richness ($\hat{r} = 1.40$) and gene diversity ($H_c = 0.130$) and the highest ratios of regionally

unique alleles (63.26%) and fixed alleles (46.94%). These genetic attributes suggested that Tibetan wild soybean may have undergone severe adaptation selection for the plateau climate and ecogeographical conditions, and had less genetic exchange with inland populations. The regional population south of the Changjiang River (Central and South China) showed higher genetic richness. UPGMA cluster analysis revealed two large geographical groups, Tibetan and inland, and revealed closer relationship among the eastern populations, which suggested that the dissemination of this species in the eastern part of China might be rapid.

Keywords Genetic diversity · Geographical differentiation · *Glycine soja* · Tibet · Wild soybean

Introduction

Wild soybean (*G. soja* Sieb. et Zucc.), the progenitor of cultivated soybeans [*Glycine max* (L.) Merr.], is native to eastern Asia, including China, Japan, the Korean peninsula, and the Far East of Russia. In China, wild soybean mainly grows in the eastern half, and no wild soybean has found in the westernmost Xinjiang, western Qinghai, and southernmost Hainan areas, where the climate is unsuitable for this species. Although wild soybean is widely distributed through most parts of this country, it does not grow in some

K.-J. Wang (✉) · X.-H. Li
The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), Institute of Crop Science, The Chinese Academy of Agricultural Sciences, 100081 Beijing, China
e-mail: wangkj@caas.net.cn

particular areas or small-scale ecological environments (habitats) because the local geographical and ecological conditions are unsuitable; e.g. paramos areas, arid or hot regions. In China, generally wild soybean does almost not grow in areas that had less than 300 mm of annual precipitation or no 1 month of monthly mean temperature of 20°C and more in the warmest months, or continuous 7 months and more of more than 20°C monthly mean temperature during a year (Wang et al. 2001).

Over 30 years, Chinese wild soybean resources were been collected all over the limits of its distribution range, contoured by 23°57′–53°29′ N and 97°04′–135° E and extensively studied in many aspects. Geographical genetic diversity of Chinese wild soybean has been evaluated morphologically (Wang et al. 2008) and at the molecular level (Liu et al. 2009). Xu 1995 showed two geographical centres of wild soybean in the Northeast and Huanghe River Valley of China on the basis of the geographical distribution of seed characteristics. Based on eight morphological characters defined in the China Genebank database, Dong et al. (2001) reported three centres of morphological diversity (including semi-wild soybean type, *G. gracilis* Skvortz.) in the northeast, the Yellow River Valley and the southeast coasts of China. Wang et al. (2005) reported that the mean seed sizes of natural populations in the northern and central wild soybean were greater than those in the southern ones in China. An investigation of phenotypic diversity in Chinese semi-wild soybean showed that the Northeast and North were two main centres (Wang et al. 2008). Shimamoto et al. (1998) reported that the Chinese wild soybeans in the Changjiang River valley had the greatest genetic diversity in cytoplasmic DNAs. Wen et al. (2009) analyzed 196 wild soybean accessions using SSR markers and found that Chinese wild soybean had geographical differentiation in three artificial-divided large regions, north, middle and south; the southern China was the major centre of the genetic diversity with highest genetic diversity and largest number of specific-present alleles, and that geographical grouping and genetic clustering of Chinese wild soybean accessions are correlated. Li et al. (2009) found that the northern and southern geographical populations of Chinese wild soybean had the largest genetic differences, and the Central regional population had higher genetic diversity, and there was clear

geographical correlation through nuclear SSR analysis of 375 wild accessions including semi-wild soybean.

Although there were extensive collection and many studies of Chinese wild soybean resource, a particular regional wild soybean–Tibetan population has almost not been known about its genetic information. Tibet is located at the southwest edge of China, which is a plateau of over mean 4,000 m in altitude, where plateau surface topography is sloping gently from the northwest to the southeast, descending from 5,500 m above sea level to 2,300 m on average. The southeast area has many flat and gently sloping patches and gorges of 1,600–2,000 m above sea level, where the microclimate is warm.

Thirty years ago, wild soybean was found in the southeast of Tibet, a warm and humid gorge (Gongrigabuqu-River Gorge), Chayu County (Li 1987). The Gongrigabuqu-River Gorge is 1,600–2,000 m above sea level and located in the northwest of Chayu County. The southeastern edge of Chayu is adjacent to Yunnan Province. The Hengduan Mountains (Meilixueshan mountain of Hengduan mountains is 6,740 m in above sea level) run along the eastern borders of Chayu and greatly obstruct the pervasion of plants from east to west. The Tibetan population of wild soybean is almost completely isolated from other areas. Although genetic diversity has been studied extensively for other regional wild soybeans in China, little is known about that of Tibetan wild soybean. Because only four Tibetan wild accessions have been reserved in the China Genebank, we very little understand the geographical, morphological and molecular genetic diversity for Tibetan wild soybean. Investigating the genetic attributes of Tibetan wild soybean will help to understand the geographical dissemination of wild soybean in China and the history of the community formation of wild soybean in Tibet. In view of the limited information on Tibetan wild soybean, the objective of this study was to analyse (1) the genetic differences and genetic diversity of the Tibetan wild soybean populations isolated from the eastern ones, (2) their geographical relationship between Tibetan and other, geographically distant populations in China; (3) to discuss how Tibetan wild soybean differentiated geographically and genetically from other regional populations.

Materials and methods

Plant sampling

To compare genetic differences between Tibetan and other eastern wild soybeans, wild soybean populations were sampled from various corners of China, including four Tibetan populations (pop. F, G, H, and I) from the Gongrigabuqu-River gorge, Chayu, and seven populations from remote regions in the centre (pop. D and E), north (pop. C), and outermost northeast (pop. A and B), southwest (pop. J) and south (pop. K) (Table 1; Fig. 1). These population sites covered four edges (northeast, southwest, southeast, and south) and two inland areas (centre and north) of the distribution range of this species in China. These are massive variation in the geographical and ecological conditions among the sites.

Each population was randomly sampled by 30–37 individuals at intervals of at least 10 m (exception for the smaller population J). The Tibetan wild soybean populations were sampled directly by field-gathering young leaves from growing plants in natural populations during flowering period (in August, 2006) in Tibet. In other populations, seed samples were collected at maturation period in 2004–2007. These population sites covered four edges (Northeast, Southwest, Southeast, and South) and two inland areas (Centre and North) of the species distribution in China; they embodied vast differences in geographical and ecological conditions.

We also used two groups of Chinese northernmost and southernmost soybean varieties (including landraces and improved varieties) as reference in the data analysis to gain insight into whether the wild populations had some genetic involvement with their surrounding cultivated soybeans. These two groups of cultivated soybeans had once been cultivated within a radius of about 100 km around the population sites A and K (Table 1).

SSR marker test

Total genomic DNA was extracted from young leaves of each plant according to the method of Doyle and Doyle (1990). We used 20 pairs of SSR markers selected from previously reported linkage groups the linkage groups (Cregan et al. 1999). These markers

showed good stability in our laboratory record. Polymerase chain reaction (PCR) amplification was carried out in a 20- μ l reaction mixture containing 100 ng template DNA, 10 pmol each primer, 100 μ M each of the four dNTPs, 1 U Taq polymerase, 2 μ l 10 \times PCR buffer. Cycling conditions were as follows: 30 s denaturation at 94°C, 30 s annealing at 47°C, 30 s extension at 72°C (30 cycles). The amplified products were separated on a 6% gel (SDS-PAGE) and bands were visualized by silver staining.

Data analysis

POPGENE1.31 (Yeh et al. 1999) was used to evaluate the genetic variability as the measures of the observed total number of alleles (N_a), effective allele number (N_e), percentage of polymorphic loci (A_p), Nei gene diversity (H_e , Nei 1973), and Shannon–Weaver index (I , Shannon and Weaver 1949). Geographical differentiation was evaluated by F -statistics values (F_{st}) among the geographical regions using software POPGENE version 1.31. A cluster dendrogram (UPGMA) was constructed to evaluate the genetic relationships for these natural populations based on the average genetic distances using software powerMarker version 3.25 (Liu and Muse 2005). To compare the levels of genetic diversity, we also calculated the amount of allele abundance in a locus by the rarefaction method of Hurlbert (1971) as introduced by El Mousadik and Petit (1996) using FSTAT2.9.3 for the parameter of allelic richness (\hat{r}) (Goudet 2001).

The Bayesian clustering algorithm was applied to assign genetically similar individual plants among these natural populations using STRUCTURE version 2.1 (Pritchard et al. 2000). One to 20 K (number of assumed clusters) was used to infer the number of clusters. Twenty times of the simulation independent were carried out for each K value, using without the prior population information model and the method allowing for the admixture and correlated allele frequencies with 50,000 burn-in period and 50,000 replication number. A consanguinity analysis of individuals was carried out to reveal the genetic interknitting within populations and between cultivated and wild soybeans based on the ancestry analysis method of Oumar et al. (2008).

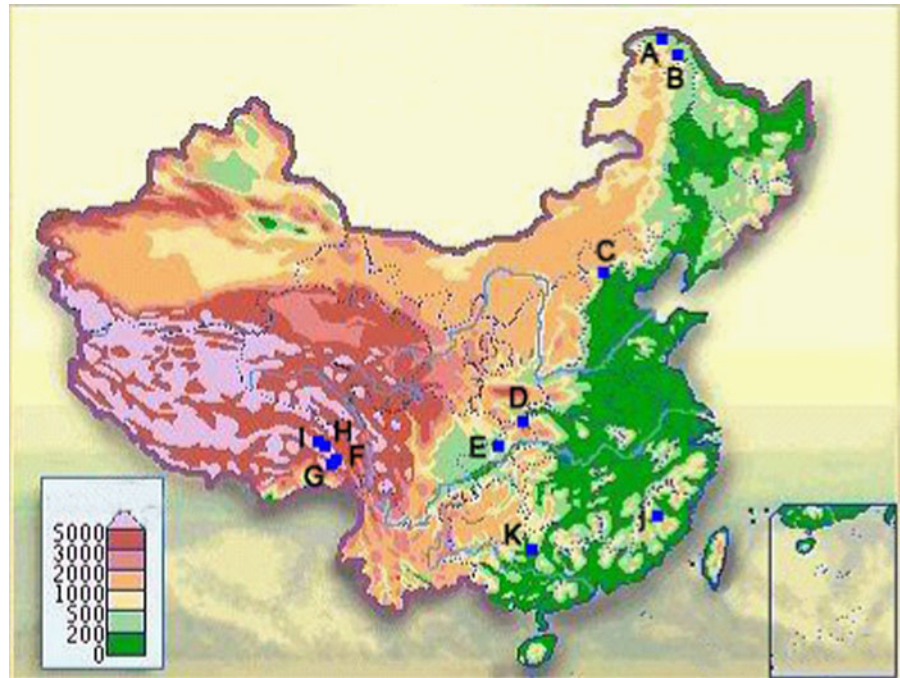
Table 1 Information on eleven wild soybean populations and two groups of soybeans (including landraces and modern varieties)

Pop.	Area (m ²)	Sample size	Location	Geographic area	Longitude	Latitude	Altitude (m)	Habitat	Mean ^a 100-seed weight (g)	Annual precipitation (mm)	Frost-free season (d) ^b
A	7,000	33	Beijicun, Heilongjiang	Northeast (NE)	122°34.505' E	53°49.772' N	173	Wasteland	1.35 ± 0.21	460.8	85
B	1,000	30	Yixiken Heilongjiang	Northeast (NE)	125°49.954' E	52°34.876' N	105	Spinney	1.43 ± 0.28	470.0	98
C	1,000	27	Liujahe, Beijing	Noth (N)	117°13.135' E	40°11.871' N	89	Drought sandy soil	1.20 ± 0.08	507.7	189
D	5,000	33	Juntiangu, Hubei	Centre (C)	110°41.267' E	32°38.530' N	203	Valley	1.59 ± 0.32	850.0	242
E	2,000	34	Guanmian, Chongqing	Centre (C)	108°40.084' E	31°33.369' N	569	Hillside	1.17 ± 0.17	1,181.4	280
F	1,000	30	Shangchayu 1, Tibet	Southwest (SW)	96°46.287' E	28°43.787' N	1,918	Field side	<2.0	801.0	280
G	800	30	Shangchayu 2, Tibet	Southwest (SW)	96°46.345' E	28°43.539' N	1,912	Field side	<2.0	801.0	280
H	300	30	Xiachayu 1, Tibet	Southwest (SW)	97°00.741' E	28°29.516' N	1,750	Hillside	<2.0	1,000.0	282
I	400	30	Xiachayu 2, Tibet	Southwest (WS)	96°00.991' E	28°43.806' N	1,733	Hillside	<2.0	1,000.0	282
J	150	30	Mingxi, Fujian	Southeast (SE)	117°15.432' E	26°20.924' N	380	Roadside	1.16 ± 0.14	1,737.0	271
K	1,000	36	Lianzhou, Guangdong	South (S)	112°10.168' E	24°58.952' N	178	Riverbank	1.06 ± 0.21	1,612.0	302
Northernmost G.		26	100 km around pop. A		Northeast						
Southernmost G.		27	100 km around pop. K		South						

^a 100-seed weight was the weight of original seeds from wild populations. Seed weight for Tibetan population was empirically estimated from morphology of individual in sampled populations

^b Local latest 20-year mean data

Fig. 1 Geographical distribution of 11 natural populations of wild soybean. These populations grew in fringe and central regions of China, Tibet (F, G, H, and I), the Centre (D and E), North (C), and outermost Northeast (A and B), Southwest (J) and South (K)



Results

SSR polymorphic variation

In the total samples of 343 plants, we observed a total of 202 alleles at the 20 nuclear SSR loci; the allele number (N_a) for each locus varied between 7 and 15, with a mean value of 10.1 (Table 2) and the mean effective allele number was 5.22. Of 202 alleles, 86 (42.57%) were rare, with a frequency less than 0.05 in the whole samples. The number of alleles is one of the most important genetic components for genetic differentiation in populations, types, geographical sites. Besides the N_a , we calculated the allelic richness, which is an objective index independent of sample size. In this sample studied, the allelic richness (\hat{r}) varied from 5.12 to 12.345 per locus, with an average of 8.087. Low frequencies (mean value = 0.0047) for the observed heterozygosity (H_o) per locus were observed as prognosticated for self-plants such as wild soybean. The mean H_c was 0.79 (range: 0.68–0.88 per locus). The correlations between the genetic parameters of genetic diversity per locus were best reflected by N_e , I and H_c in this sample analyzed ($r = 0.94$ – 0.96 , $P < 0.001$).

Genetic diversity and variation in populations

The populations showed genetic differences among the 11 populations. Sixty alleles were fixed among the 11 populations (Table 4), of which higher ratios appeared in all the four Tibetan populations (range: 40–75%), one central population E (75%) and the drought-resistant population C (95%; Table 5). The percentage of polymorphic loci (A_p) ranged from 5 to 95% with a mean of 56.8% between populations, and the mean number of alleles (N_a) per population ranged from 1.05 to 4.00 averaged, averaging 2.01. The observed heterozygosity (H_o) per population was 0.003. The mean allelic richness (\hat{r}) per locus among the populations varied from 1.05 to 3.984, with an average of 2.007. The drought-resistant population C had the lowest values for various genetic parameters (Table 3), which implied that this population had been severely selected by drought stress so that its most loci (95% loci) were fixed (Table 4). The central population D had highest genetic diversity (Table 3), with the highest values for all genetic parameters. However, its nearest population E, about 280 km far away, had lower genetic diversity. Population E was located on a hillside and this population might be formed by the individuals from

Table 2 Genetic parameters for 20 nuclear SSRs used in eleven natural wild populations and two groups of soybeans in China

Locus	Linkage group	N_a		N_e		I		H_o		H_e		\hat{r}	
		S	G	S	G	S	G	S	G	S	G	S	G
Satt022	N	10	4	4.33	3.38	1.768	1.286	0		0.770	0.704	7.885	3.999
Sac_112	E	8	3	4.01	1.76	1.642	0.683	0		0.752	0.432	6.757	2.722
Satt168	B2	10	3	4.61	1.76	1.778	0.737	0.009		0.784	0.432	8.095	2.981
Satt173	O	15	5	8.48	2.80	2.350	1.288	0.009		0.884	0.643	12.345	4.990
Satt180	C1	7	4	3.55	2.44	1.539	1.065	0.003		0.720	0.591	6.570	3.926
Satt216	D1a + W	15	6	4.77	3.74	1.964	1.443	0		0.792	0.733	10.204	5.444
Satt236	A1	12	3	6.55	1.40	2.059	0.515	0		0.848	0.288	9.352	2.722
Satt267	D1a + W	8	5	4.69	3.08	1.704	1.289	0.029		0.788	0.676	6.613	4.703
Satt281	C2	13	7	4.30	5.34	1.783	1.784	0.009	0.019	0.768	0.813	9.007	6.952
Satt352	G	8	3	4.09	1.41	1.609	0.546	0		0.757	0.292	6.394	2.926
Satt373	L	10	4	5.98	2.61	1.956	1.130	0		0.874	0.617	8.580	3.981
Satt386	D2	7	2	4.59	1.68	1.662	0.596	0		0.783	0.406	6.132	2.000
Satt414	J	9	3	4.24	2.07	1.664	0.834	0.003		0.767	0.517	7.420	2.981
Satt429	A2	13	3	8.28	1.45	2.234	0.852	0.003		0.881	0.534	10.275	2.981
Satt431	J	6	7	3.57	3.45	1.422	1.567	0		0.721	0.710	5.120	6.900
Satt434	H	13	5	8.03	3.96	2.263	1.470	0.006		0.877	0.748	11.237	4.981
Satt453	B1	9	4	3.57	2.20	1.525	1.042	0.006		0.721	0.545	6.880	3.998
Satt571	I	10	4	6.73	1.36	2.029	0.535	0		0.823	0.266	8.550	3.494
Satt586	F	8	6	3.07	2.43	1.278	1.224	0.017		0.675	0.588	4.875	5.827
Satt590	M	11	3	7.06	2.43	2.139	0.971	0		0.870	0.588	9.458	3.000
Mean		10.1	4.2	5.22	2.53	1.818	0.754	0.005	0.002	0.793	0.556	8.087	4.273

S, *G. soja*; G, *G. max*; N_a , allele number; N_e , effective allele number; I , Shannon–Weaver index; H_o , observed heterozygosity; H_e , Nei gene diversity; \hat{r} , allelic richness

the superior border. The northern and southern populations (A, B, J, and K) showed rather high genetic diversity. All four Tibet-Plateau populations (F, G, H, and I) all were characterized by low genetic richness (1.25–1.55; Table 4).

Intrapopulation status in genetic structure

The individual assignment test showed that all plants were completely assigned to their original populations (Fig. 2). No dissemination over large geographically areas was observed among these populations, even between relatively near populations D and E, which were 280 km apart. We analysed the genetic structure within the populations using STRUCTURE version 2.1. The Tibetan and other populations were statistically near-homologous. Only two northeastern populations A and B showed obvious structures; population A had two distinct genetic component

groups and population B had three genetic component groups (Fig. 2).

Population A occupied a larger area, where the heterogenous components might have originated from the migration of surrounding congeners or from the introgression of soybean farmlands. To clarify reason what caused the heterogenous components in the population A, we conducted ancestry analysis of individuals determine whether gene flow had introgressed from soybeans into this wild population. For this test, we used 26 soybean varieties that grew within approximately 100 km of site A. The result displayed that there had been gene-flow from cultivated soybeans into this wild population in the past (Fig. 3a). At least four wild individual plants were showed consanguinity with soybeans. The heterogenous plants originated from natural cross between wild and cultivated soybeans a long time ago. Under natural selection in the ecosystem, the offspring segregated towards small

Table 3 Estimates of genetic diversity among eleven natural wild populations and two groups of soybean in China

Pop.	Area	Genetic parameter						
		A_p	N_a	N_e	I	H_o	H_e	$\hat{\pi}$
A	NE	90 (18)	2.85	1.67	0.599	0.003	0.350	2.830
B	NE	80 (16)	2.25	1.45	0.409	0.002	0.239	2.240
C	N	5 (1)	1.05	1.00	0.008	0	0.004	1.050
D	CC	95 (19)	4.00	2.71	1.020	0.003	0.554	3.984
E	CC	25 (5)	1.25	1.04	0.067	0	0.036	1.247
F	SW (Tibet)	25 (5)	1.25	1.10	0.108	0	0.069	1.250
G	SW (Tibet)	45 (9)	1.45	1.22	0.197	0	0.131	1.449
H	SW (Tibet)	55 (11)	1.55	1.44	0.342	0.007	0.242	1.550
I	SW (Tibet)	30 (6)	1.35	1.11	0.133	0	0.078	1.350
J	SE	95 (19)	2.45	1.69	0.544	0.008	0.334	2.440
K	S	80 (16)	2.70	1.72	0.596	0.013	0.342	2.687
Northernmost <i>G. max</i> group		95 (19)	2.95	2.02	0.754	0.004	0.436	3.000
Southernmost <i>G. max</i> group		100 (20)	3.55	2.34	0.908	0	0.507	3.543
Mean for <i>G. Soja</i>		56.8	2.01	1.47	0.366	0.003	0.216	2.007
Mean for <i>G. max</i>		97.5	3.25	2.18	0.831	0.002	0.472	3.272

A_p , percentage of polymorphic loci. Parenthesis is the number of polymorphic loci

seed size, so that these offspring plants completely became wild soybean morph. We also analysed the southern population K and 27 soybeans within a 100 km radius of site K, and it appeared that there was no natural hybridization between population K and the surrounding soybeans (Fig. 3b). This was in accordant with the results shown in Fig. 2, where indicated that population K consisted of genetically homogeneous individual plants.

However, population B was different from the situation of population A, which showed eco-genetic differentiation that resulted from two microenvironments; osier field and a drought, sandy revetment-dike of the osier field. Fourteen individuals (i.e. group 3) growing on the revetment-dike differentiated from other 16 individual samples from the osier field (groups 1 and 2). Population B generated intrapopulation heterogenous hybrid individuals among the three different components (groups 1, 2 and 3), as shown Fig. 4.

Geographic differentiation

Geographically these populations distinctly showed regional features of genetic differentiation. The largest sample, 120 Tibetan individuals produced only 49 alleles; 23 (46.94%) of which were fixed. This

proportion of fixed alleles was obviously higher than that found in the other regional samples, for example, 23/111 (20.72%) in N-NE, 16/89 (17.98%) in region C, and 5/59 (8.25%) in S-SE (Table 5). The Tibetan wild soybean and southern samples generated more region-unique alleles. The proportion of regionally unique alleles was 63.26% for Tibetan wild soybean sample, 61.02% the southern sample (S-SE), 44.94% for central sample C, and 27.92% for the northern sample (N-NE; Table 4). However, in the Tibetan sample, 42.86% (21 alleles) of the fixed alleles, which was a higher ratio, were region-unique alleles; whereas in the southern sample (S-SE) only 5.05% of fixed alleles were regionally unique. In the C and N-NE regional samples, the regionally unique alleles of fixed alleles were 10.11 and 13.51%, respectively. In all the Tibetan samples 20% of loci were fixed single alleles (Table 5). Samples from regions C and S-SE had the highest allelic richness values ($\hat{\pi}$; 2.616 and 2.564), followed by N-NE (2.041), whereas the Tibetan sample showed the lowest value (1.40; Table 5). Together, these results suggested that Tibetan wild soybean populations had less genetic exchange with other regional populations.

The cluster dendrogram (UPGMA) confirmed that there was regional genetic differentiation among these regional populations (Fig. 5), revealed two

Table 4 The number and distribution status of alleles at each locus among the wild populations

Locus	Four regions and eleven populations											No. of fixed alleles in locus		
	North-Northeast (N-NE) region					Centre (C) region			Tibet region				South-Southeast (S-SE) region	
	Pop. A	Pop. B	Pop. C	Pop. D	Pop. E	Pop. F	Pop. G	Pop. H	Pop. I	Pop. J	Pop. K			
Satt022	2	1-L	1-(A) ^a	4 (g,k,n,q)	2 (d)	2	2	1-L	2	2 (t)	2	2		
Sac_112	1-(B) ^a	2	1-(C) ^a	2 (a)	2 (k)	1-(E)	1-(E)	1-(E)	1-(E)	2 (n)	2	3		
Satt168	3	5 (h)	1-(D) ^a	5	1-(E) ^a	1-(C)	1-(L)	1-(C)	1-(L)	3	2 (b,f)	3		
Satt173	2	2 (m)	1-T	4 (d)	1-(K) ^a	1-(Q)	2 (q)	2 (p)	1-T	4 (h)	5 (i,d,x,aa)	3		
Satt180	2	2 (d)	1-(F) ^a	2 (b)	2	1-(C)	1-(C)	2 (j)	1-(C)	3 (a,i)	1-C	2		
Satt216	5 (k)	3 (d)	1-A	4 (c,j,q)	1-(H) ^a	1-A	2 (y)	2 (u)	1-A	5 (e)	2 (f)	2		
Satt236	2	2 (k)	1-(F) ^a	2 (b)	1-(H) ^a	2	2 (i)	2	2	7 (a,j,n)	2 (l)	2		
Satt267	3	2	1-(E) ^a	3	1-(L) ^a	1-(F) ^a	2 (j)	1-(J)	2 (j)	3 (b)	2 (a)	4		
Satt281	5 (e,f)	2	1-G	7 (a,q,u)	1-G	1-(R) ^a	2	2 (o)	1-(M) ^a	4 (p)	4 (t)	3		
Satt352	3	3	1-G	6 (d,o)	1-G	2	1-G	2	1-(J)	1-(F) ^a	3 (b)	3		
Satt373	4	1-O	1-(E) ^a	5	1-(L) ^a	1-(Q)	2 (f,q)	2 (f)	3 (f,q)	1-O	3 (k)	5		
Satt386	3	2	1-(H) ^a	3 (e)	1-(B) ^a	1-(C)	1-(C)	1-(C)	1-(C)	1-(F) ^a	3	4		
Satt414	4	2	2 (m)	6 (l)	1-(G) ^a	2	1-(D)	2	1-(D)	3 (n)	2	2		
Satt429	4	3 (b,t)	1-(A) ^a	2	2 (j)	1-(N)	1-(N)	1-(M)	1-(M)	2 (o)	2 (d,u)	3		
Satt431	3 (i)	1-(E) ^a	1-(H) ^a	4 (b)	2	1-(C)	1-(C)	1-(C)	1-(C)	3	3	3		
Satt434	3 (c)	1-(F) ^a	1-D	5 (g,l,n)	1-D	2 (m)	1-(J) ^a	2 (m)	1-(M)	1-(E) ^a	2	5		
Satt453	3	2	1-D	6 (a,c,j)	1-D	1-(F)	1-(F)	2	1-(F)	3 (g)	2	2		
Satt571	2 (b)	2 (b)	1-(B)	5 (m)	1-(G) ^a	1-(J) ^a	2	1-(H) ^a	2	2	3	3		
Satt586	2	5 (d,k)	1-(E) ^a	4 (h,j)	1-A	1-A	1-A	1-A	1-A	2 (g)	2	2		
Satt590	1-A	1-I	1-(D) ^a	1-A	1-I	1-(E)	2 (e)	2 (e)	2 (h)	2 (g,n)	2 (k,m)	4		
No. alleles in pop.	57	45	21	80	25	25	29	31	27	54	49	60		
NOP (%)	2 (10.0)	5 (25.0)	19 (95.0)	1 (5.0)	15 (75.0)	15 (75.0)	11 (55.0)	8 (40.0)	14 (70.0)	4 (20.0)	1 (5.0)			
Total alleles in region	111			89		49				59				
NOR (%)	31 (27.92)			40 (44.94)		31 (63.26)				36 (61.02)				

Number denotes number of alleles. Capital letter denotes fixed alleles in populations. Parenthesis indicates regionally-unique alleles

NOR number of regionally-unique alleles (% in alleles of a region)

^a Denotes fixed regionally-unique alleles. NOP: number of fixed alleles in populations (% in 20 loci)

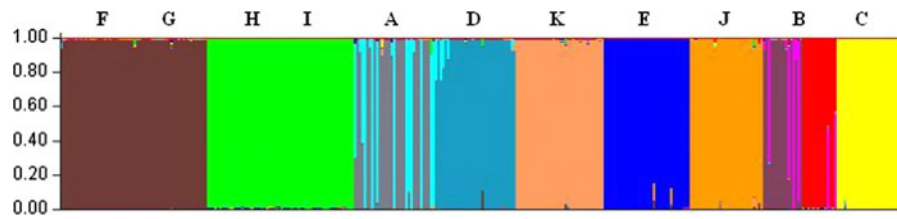


Fig. 2 Individual assignment in the 11 populations (A–K) of wild soybean using the without prior population information model when $k = 12$ [(with the highest likelihood value (In *PrXIK*)]. Two northeastern populations A and B had heterogeneous components in genetic structure. All plants were

completely assigned to their original populations and no dissemination over distant geographical areas was found among these remote populations, even between D and E, which were only 280 km apart

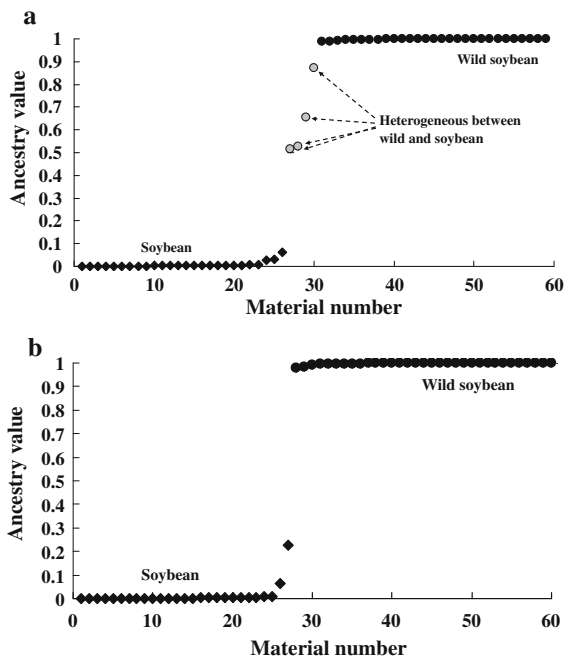


Fig. 3 Possibility of introgression from soybeans into wild populations. Ancestry analysis identified four introgressive plants consanguineous with soybeans in the northernmost population A (a) and no acceptable introgressive plants in southernmost population K (b). As shown in Fig. 2, heterogeneous components were present in population A but not in population K

geographical groups, Tibetan group and inland group. The Tibetan populations were independently differentiated from other regional populations in China. Population C from northern Beijing was a drought and sandy soil population and had specialized adaptation. The two northernmost and southernmost groups of soybeans first clustered and then jointed with the southern population (S). Basically, the populations located close to each other first

converged as geographical groups, with the exception of population J from southeastern Fujian. Population J (SE) first clustered with central population D north but not with transverse population K(S). Genetically, southern K appeared to have closer connections with central E and northernmost A and B along with latitude. This suggested that there may be a latitudinal species-dispersal effect or gene flow, which was consistent with the topography in this part of China, i.e., the mountainous terrain and plateau obstructs dispersal of wild soybean in western China. All the four regional populations (Tibet, N-NE, Centre, and S-SE) showed significant geographical differentiation from one another (Table 6).

Discussion

This is the first study assessing genetic diversity and geographical peculiarity for Tibetan and other remote border populations of wild soybean in China. The Tibetan wild soybean was characterized by low level of genetic diversity as indicated by higher ratio of fixed alleles, higher frequency of regionally-unique alleles (Table 5), and lower genetic richness (Tables 3 and 5). The higher ratio of allele fixation and higher frequency of regionally-unique alleles suggest that Tibetan wild soybean had less genetic exchange from other inland populations because of its geographical isolation. Some results of previous studies in genetic diversity of wild soybean were always not coincident, and the various studies reported different morphological and genetic diversity centres for Chinese wild soybean (Xu 1995; Shimamoto et al. 1998; Dong et al. 2001; Li et al. 2009; Wen et al. 2009; Guo et al. 2010). In the present data of the natural populations, the higher

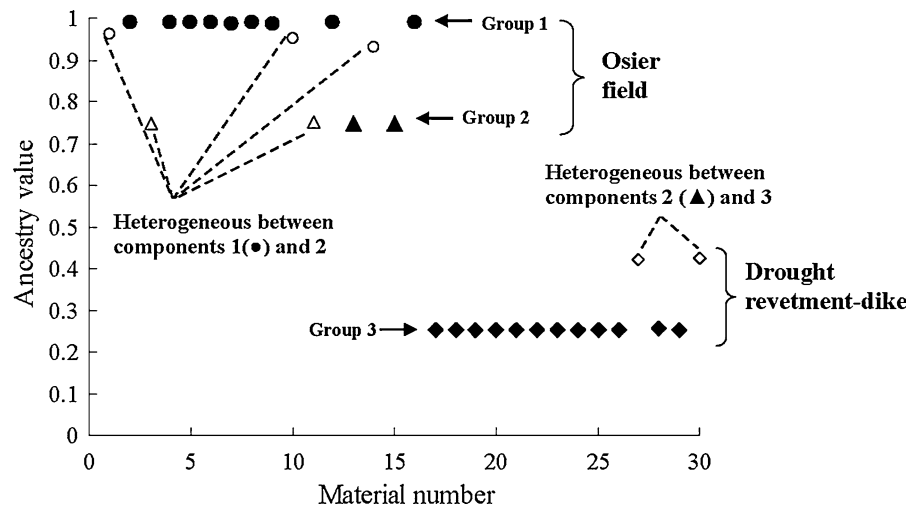


Fig. 4 Heterogeneity of genetic backgrounds in population B. There three genetic components (groups 1, 2 and 3) and two kinds of heterogeneous individuals between components 1 and 2 and between 2 and 3. Some genetic differentiation seemed to

occur between the drought and normal microenvironments; 14 drought plants were genetically separated from other plants growing in the osier field

Table 5 Status of region-unique alleles in wild soybean of four geographical areas

Region	No. of plants	Total alleles	No. of loci with unique alleles (%)	Regionally-unique alleles fixed in pop. (%)	No. of loci with single alleles (%)	No. of fixed alleles (%)	Gene diversity (H_e)	Allelic richness (\hat{r})
S-SE	66	59	15 (75.0)	3 (5.08)	0	5 (8.45)	0.338	2.564
Tibet	120	49	11 (55.0)	21 (42.86)	4 (20.0)	23 (46.94)	0.130	1.400
C	67	89	16 (80.0)	9 (10.11)	0	16 (17.98)	0.295	2.616
N-NE	90	111	17 (85.0)	15 (13.51)	0	23 (20.72)	0.197	2.041

Data summarized from Table 4

levels of allelic richness were observed in populations from central China and the southern region (Table 5).

The UPGMA cluster revealed two large geographical groups, Tibet and inland. Tibetan wild soybean showed large genetic differences from other populations growing in central and various fringe regions of China (Table 4). The southern K appeared to be more closely connected to central E and northernmost A and B, which implied that there may have been more rapid spread of wild soybean between the north and south of eastern China in the dissemination history of this species. The faster dissemination along the south-north route was consistent with the topography of China, as the eastern parts are plain and lower hills and the western parts are plateau and has many north-south longitudinal mountains. However, the ancestry of Tibetan wild soybean remains unclear, and is requires further research.

Wild soybean was maybe a relatively age-old species of eastern Asia and established in subtropical belt area of China, 5 million years ago (Singh and Nelson 2009); this area is regarded as a centre of world's angiosperm (Wu and Wu 1996; Wu 1988). The Japanese islands were joined with Asian continents of until late Miocene-Pliocene and Pleistocene (18,000 years ago), and Taiwan island was conjoint with the Asian continent before the Tertiary period. Japanese wild soybean is likely to be a descendant of the species at that time, which migrated with the motion of the Pacific plate. A study showed the genetic differentiation between Chinese and Japanese wild soybeans (Wang and Takahata 2007). Our present study showed that Tibetan wild soybean, with lower genetic diversity (Tables 3 and 5), differed genetically from other regional wild soybean populations (Table 6 and Fig. 4). The results

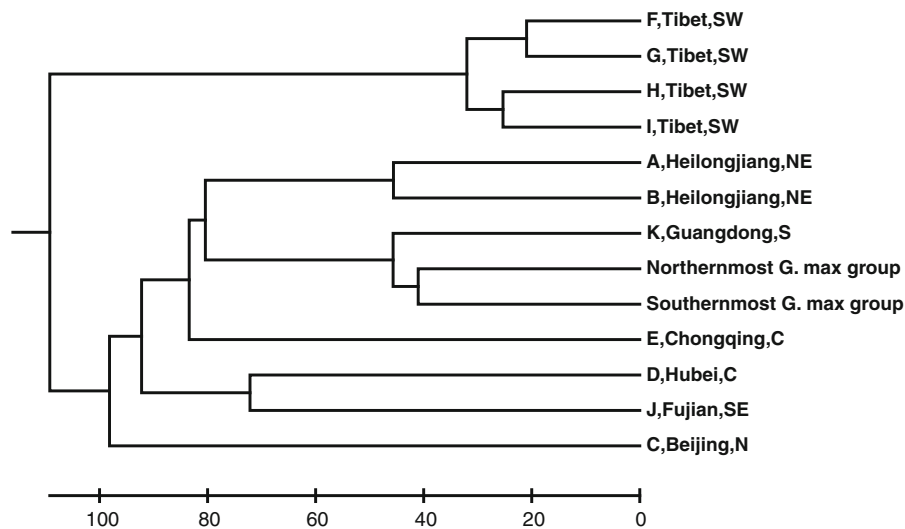


Fig. 5 UPGMA cluster based allele frequencies and Nei (1973)'s distance with 1,000 bootstrap replicates (0.95 confidence interval) for eleven wild soybean populations and two soybean groups. The cluster showed two large geographical groups, Tibetan and inland groups. The Tibetan populations were independently differentiated from other regional populations in China. Population C was a drought sandy soil

population and had specialized adaptation to its environment. Two northernmost and southernmost groups of soybeans first clustered and then joined with southern population (S). Population J (SE) first clustered with central population D north but not with population K (S) transverse. Southern K was more closely genetically related to central E and northernmost populations A and B

Table 6 Differentiation estimated by *F*-statistics values between areas

Area	Tibet	N-NE	CC
N-NE	0.429*		
C	0.472*	0.321*	
S-SE	0.443*	0.305*	0.286*

* $P < 0.05$

suggested that Tibetan wild soybean in the Gongrigabuqu-River gorge likely underwent severe natural selection during its adaptation to the plateau climate and ecogeographic conditions in the gorge. Another possible explanation for the lower genetic diversity in the Tibetan wild soybean populations may implicate the late migration of wild soybean since the this species was established, that is, it could be that Tibetan wild soybean did not arise until after this species had already spread throughout the eastern parts of China and other eastern Asia (Japan, Taiwan, the Korean peninsula, and the Far East of Russia) via land routes, and profiting from the favorable climate. However, the history of the naissance of Tibetan wild soybean remains unclear. The occurrence and distribution of Tibetan wild soybean are likely to be

related to the vicissitudes in climate and flora development of southeastern Tibet during and after the formation of Himalayan-Hengduan Mountains (Boufford and Van Dyck 2000; Sun 2002).

Cultivated soybean was domesticated somewhere in China 5,000 years ago (Hyten et al. 2006). Nevertheless the results of this study could provide little knowledge regards to the origin of cultivated soybean, because only a small number of natural populations were selected to focus on the genetic characteristics of Tibetan wild soybean. To determine the origin of cultivated soybean, more diverse geographical and evolutionary-level type accessions of Chinese wild and cultivated soybeans should be analyzed. Our results showed that the two soybean groups from the small areas in northernmost and southernmost parts of China were more closely related to the southern population K (Guangdong).

Our results clearly demonstrated that genetic differentiation of wild soybean has occurred not only among natural populations and geographical regions, but also even between heterogeneous microenvironments within a small natural population, and that outcrossing would occur between plants within populations (Fig. 4). As revealed by the ancestry

analysis in Fig. 3a, introgression from cultivated soybean into its wild progenitor species had occurred. It is perhaps that this has continued for 5,000 years in China since the naissance of cultivated soybeans.

References

- Boufford DE, Van Dyck PP (2000) South-Central China. In: Mittermeier RA, Myers N, Mittermeier CG (eds) Hotspots: earth's biologically richest and most endangered terrestrial ecoregions. CEMEX, Mexico, pp 338–351
- Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, Vantoai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464–1490
- Dong YS, Zhuang BC, Zhao LM, Sun H, He MY (2001) The genetic diversity of annual wild soybean grown in China. *Theor Appl Genet* 103:98–103
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theor Appl Genet* 92:832–839
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Lausanne University, Switzerland
- Guo J, Wang YS, Song C, Zhou JF, Qiu LJ, Huang HW, Wang Y (2010) A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Ann Bot* 106:505–514
- Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577–586
- Hyten DL, Song QJ, Zhu YL, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci USA* 103:16666–16671
- Li FS (1987) Report on the survey of soybean resources in Tibet. *Soybean Sci* 6:79–83 (in Chinese with English abstract)
- Li XH, Wang KJ, Jia JZ (2009) Genetic diversity and differentiation of Chinese wild soybean germplasm (*G. soja* Sieb. & Zucc.) in geographical scale revealed by SSR markers. *Plant Breed* 128:658–664
- Liu K, Muse SV (2005) Powermarker: integrated analysis environment for genetic Markerdate. *Bioinformatics* 21: 2128–2129
- Liu YN, Li XH, Wang KJ (2009) Analysis of the genetic variability for the mini-core collection of Chinese wild soybean (*Glycine soja*) collection in the national gene bank based on SSR markers. *J Plant Resour* 10:211–217 (in Chinese with English abstract)
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321–3323
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana
- Shimamoto Y, Fukushi H, Abe J, Kanazawa A, Gai JY, Gao Z, Xue DH (1998) RFLPs of chloroplast and mitochondrial DNA in wild soybean, *Glycine soja*, growing in China. *Genet Resour Crop Evol* 45:433–439
- Singh RJ, Nelson RL (2009) The utilization of soybean wild relatives: how can it be effective? The abstracts of proceedings. The Eight World Soybean Research Conference, Beijing, P. R. China, pp 211–212
- Sun H (2002) Evolution of Arctic-Tertiary flora in Himalayan-Hengduan Mountains. *Acta Bot Yunnan* 24:671–688 (in Chinese with English abstract)
- Wang KJ, Takahata Y (2007) A preliminary comparative evaluation of genetic diversity between Chinese and Japanese wild soybean (*Glycine soja*) germplasm pools using SSR markers. *Genet Resour Crop Evol* 54:157–165
- Wang KJ, Li FS, Cheem KA (2001) Studies on the distribution of wild soybean (*Glycine soja*) in China. *Palistan J Biol Sci* 4:149–155
- Wang KJ, Li XH, Zhang ZW, Li FS, Chao YS (2005) Components and distribution of 100-seed weight in natural populations of wild soybean. *Soybean Sci* 24:243–248
- Wang KJ, Li XH, Li FS (2008) Phenotypic diversity of the big seed type subcollection of wild soybean (*Glycine soja* Sieb. et Zucc.) in China. *Genet Resour Crop Evol* 55:1335–1346
- Wen ZX, Ding YL, Zhao TJ, Gai JY (2009) Genetic diversity and peculiarity of annual wild soybean (*G. soja* Sieb. et Zucc.) from various eco-regions in China. *Theor Appl Genet* 119:371–381
- Wu CY (1988) Hengduan Mountains flora and her significance. *J Jpn Bot* 63:297–311
- Wu ZY, Wu SG (1996) Proposal for a new floristic kingdom (Realm) – The E. Asiatic kingdom, its delineation and characteristics. In: Zhang AL, Wu SG (eds) Floristic characteristics and diversity of east Asian plants. China High Education Press, Beijing, pp 3–42
- Xu B (1995) Geographical distribution and genetic diversity of seed traits of wild soybean in China. *Acta Agron Sin* 21:733–739 (in Chinese with English abstract)
- Yeh FC, Yang RC, Boyle T (1999) POPGENE software package version 1.31 for population genetic analysis, University of Alberta