Plant Omics Journal POJ

POJ 7(6):415-423 (2014) ISSN:1836-3644

Genetic and epigenetic diversity of wild and cultivated soybean in local populations in Northern Huang Huai region of China

Xuefei Yan[2]† Xiaodong Liu[1]† Jiandong Li[2]† Hongkun Zhao[1] Qiyun Li[1] Yumin Wang[1] Cuiping Yuan[1] Ling Zhang[1] Yingshan Dong[1*]

[1]Jilin Academy of Agricultural Science, Changchun, Jilin 130033, People's Republic of China
[2]College of Agronomy, Shenyang Agricultural University, Shenyang, Liaoning 110866, People's Republic of China
†These authors contributed equally to this work.

***Corresponding author: Yingshan Dong, ysdong@cjaas.com**

Abstract

This study aimed to characterize genetic/epigenetic structures within and between cultivated and wild soybean at local population level in Northern Huang Huai region of China. We analyzed a total of 124 individuals including 94 wild accessions and 30 cultivars using 10 AFLP primer pairs and 8 MSAP primer pairs. The results revealed that the genetic and epigenetic diversity in cultivated soybean exceeded that of wild accessions at local population level. The reason for greater amount of diversity in cultivars could be due to bulking and mixing of cultivated gene pool from geographically distant populations by crossbreeding under condition of artificial domestication, while, gene flow was difficult in natural habitats between wild populations. Interestingly, the genetic and epigenetic diversity were differed slightly (F-test，P=0.181-0.531) in the cultivars, while, the epigenetic diversity was significantly higher (F-test，P<0.001) than the genetic diversity in the wild populations. A possible explanation was that a fluctuating environment in natural habitats might lead to more epigenetic variation in wild populations. Furthermore, the Structure, AMOVA, and PCA analyses indicated that genetic differentiation has occurred between the wild soybean and cultivars (Φrt=0.12), and among wild populations (0.307<Φst<0.352). Furthermore, existence of epigenetic divergence among the wild populations (0.133<Φst<0.168) was evidenced. This indicated that natural selection might act on epigenetic variation in a similar manner like genetics. Finally, AMOVA and Structure analyses reflected greater epigenetic similarity than genetics in the studied individuals. We suggest that the reason was basically due to adaptive convergence to similar environments.

Keywords: Cultivated soybean; epigenetic diversity; genetic diversity; population; wild soybean.
Abbreviations: AFLP_Amplified fragment length polymorphism**;** MSAP_methylation-sensitive fragment length polymorphism; PPL_the percentage of polymorphic loci; I_Shannon-winner diversity index; H_Nei's gene diversity.

Introduction

"Epigenetics "is broadly defined as alteration of phenotype, morphology or biological molecules, without changes in either the coding sequence of a gene or the upstream promoter region (Rapp and Wendel, 2005). This highlights the view that natural variation can exist not only in the DNA sequence but also at the epigenetic level (Vaughn et al., 2007). The best understood type of epigenetic modification is DNA methylation. The extensive technique studying DNA methylation is MSAP (Angers et al., 2010). Reyna-Lopez et al. (1997) firstly used the MSAP approach to study the DNA methylation patterns on fungi. Ever since, a number of studies have also documented methylation-based epigenetic variation in model and cultivated species (Xiong et al., 1999; Ashikawa, 2001; Cervera et al., 2002; Salmon et al., 2008). Recently, with the increasing interests in understanding the role of epigenetic process in ecology and evolution, which is the primary issues of the new field of "ecological epigenetic" (Bossdorf et al., 2008), several studies have focused on the population epigenetic studies with the MSAP technique. Li et al. (2008) assessed the epigenetic structure in wild barely populations and demonstrated the utility of MSAP for its detection of high level of DNA methylation polymorphism. Herrera et al. (2010) measured the epigenetic variation in 14 populations of the violet and concluded that the observed epigenetic variation might be involved in population differentiation in ecologically important traits. These are considered as essential first step for assessing the ecological and evolutionary relevance of epigenetics (Richards et al., 2010). Richard et al. (2012) found a great deal of epigenetic differentiation in the study on a low-diversity plants of *Fallopia* species, and further indicated that epigenetic effects could contribute to phenotypic variation in genetically depauperate invasive populations. Similar examples were found in the investigation of vertebrates (Massicotte and Angers, 2012; Schrey et al., 2012), which indicated that epigenetic variation may compensate the decreased genetic variation. These studies collectively demonstrated that variation of DNA methylation existed commonly in organism and could contribute to the phenotypic differences, and thus to diversity and evolutionary potential of population.

Cultivated soybean is an economically important leguminous seed crop for feed, oil and soybean products for its rich seed protein and oil (Singh et al., 1999). Wild soybean, as the

putative progenitor of cultivated soybean (Dong et al., 2001), is a unique resource for its possibility of retaining genetic information before artificial selection during soybean domestication. Such crop and its wild relatives; therefore, constitute a reservoir of a variety of genetic material. Levels and patterns of diversity within and between cultivated and wild soybean gene pools have been reported before. These studies have shown greater genetic diversity in wild clusters than that in cultivated accessions (Abe et al., 1999; Li and Nelson, 2002; Wen et al., 2008), clear genetic differentiation within cultivated pool (Li et al., 2008), within wild soybean populations and between them (Xu et al., 2000; Xu and Gai, 2003; Kuroda et al., 2006), and gene flow among wild individuals (Kuroda et al., 2010).

Wild relatives are threatened by the loss of natural habitats due to human population growth (Jin et al., 2003), which lead to more isolated local populations. For more utilitarian conservation, it is necessary to evaluate the local genetic/epigenetic structure in these populations. Unfortunately, there is a little information about the extent and partitioning of genetic and epigenetic diversity in wild soybean populations on a local scale, as well as the genetic/epigenetic relationships between wild soybean populations and local cultivars.

In this study, we used two molecular marker systems AFLP and MSAP to evaluate genetic and epigenetic variation of wild and cultivated soybean in local populations in Northern Huang Huai region of China. With these data, we sought to address the questions. (1) Do the cultivars exhibit a lower diversity than the wild soybeans as expected on a local scale? (2) In local populations, does the genetic differentiation occur within cultivated pool, within wild soybean populations and between them, as in previous studies? (3) What is the extent and structure of epigenetic variation within and among local populations? (4) What role does the genetic and epigenetic variation play in evolution and adaptation to the same environments?

Results

Genetic diversity and structure in wild and cultivated soybean

The ten primer combinations assayed in the AFLP analysis produced a total of 1796 fragments ('loci' hereafter) that could be unambiguously scored for the 124 plants of wild and cultivated soybean from 5 populations sampled. The number of loci detected in the cultivated pool was 1618. Of which, 463 (28.6%) were private. In the wild pool, 178 loci (13.4%) were private out of the 1333 loci detected. These results revealed that the cultivated pool had more loci and private loci. This kind of trend also existed in the diversity parameters. Genetic diversity parameters for the cultivated and wild populations were presented in Table 2. In comparison, the gene diversity in the cultivated pool (PPL=88.6%, I=0.371, H=0.236) exceeded that of the wild group (PPL=73.4%, I=0.270, H=0.171) despite smaller sample sizes (30 vs. 94 accessions) in cultivated accessions. At the population level, CY population (a cultivated population) had the highest number of polymorphic loci (82.0%), the highest value of the Shannon Index of Diversity (0.368) and the highest value of Nei's gene diversity (0.236), whereas, WJ population (a wild population) had the lowest diversity values for all measures (PPL=40.5%,I=0.171, H=0.109).

Hierarchical AMOVA revealed that the largest component of variation (65.8%) was among individuals within population (Table 3). However, there were significant Φ values between species (Φrt=0.12), as well as among populations (Φpr=0.25), which suggested existence of the genetic differentiation. Pairwise Φst amongst the 5 populations ranged from 0.002 to 0.352 with the highest values occurring between the populations of WA and WW (Table 4). Genetic differentiation was significant between all the populations but the comparisons were between the two cultivated populations. Specially, the three wild populations appeared to be significantly distinct from each other ($0.307<\Phi st<0.352$, $P<0.01$) and from the two cultivated populations ($0.273<\Phi st<0.343$, $P<0.01$). In contrast, the comparison between two cultivated populations ($\Phi st=0.002$, $P=0.33$) revealed a strong similarity between the local cultivated populations.

In the PCA, most of variation was explained by the first and second axes (Fig.2) which accounted for 38.39% and 29.02% of the total, respectively. The first axis of cultivated accessions plotted separately from wild individuals. Furthermore, the cultivated accessions were not separated into two clusters as expected, while, the wild populations of AX, WQ and JZ were generally separated from one another, except some accessions scattered in the middle position, indicating that clustering of wild accessions basically corresponded to the geographical origin of the populations.

To further investigate the genetic clustering among the 124 accessions, Structure software was applied. The estimated log probability of data for a given K, that is L(K), kept increasing with increasing of K values, whereas △K showed a clear peak at K=4, suggesting that four was the most likely number of genetic clusters. At k=2, all wild individuals showed an average ancestry of 98.5% in one cluster (Table 5), whereas all cultivars exhibited an average ancestry of 94.1% in the other cluster. At K=3, all the cultivated accessions also clustered together with a high average ancestry (92.7%). However, the wild accessions were split into two clusters; one cluster was constituted by the accessions from WW populations with an average ancestry of 91.1%, while the WA and WJ accessions had average ancestries of 96.7% and 87.7%, respectively, in the other cluster. At K=4, one cluster was predominant across all the cultivated accessions with a high average ancestry (>90%) in this cluster. The wild soybean accessions from different populations tended to be classified into three clusters, corresponding closely to their geographical origins, with high average ancestries (>90%). This pattern revealed that genetic differentiation existed in wild soybean at local population level. At K=5, which was the actual population number of the accessions, the first three clusters (clusters dⅠ, dⅡ and dⅢ) clearly corresponded to the three wild populations respectively, with high average ancestries (>90%). But, the two cultivated populations did not belong to distinct clusters. If we considered that an accession belong to a cluster when its ancestry in the cluster exceeded 80%, the fourth cluster (cluster dⅣ) contained four CY accessions and one CQ accessions, while the fifth cluster (cluster dⅤ) was comprised of seven CY accessions and eight CQ accessions. The result reflected that one-third of cultivated individuals were of mixed genotypes, sharing parts of their genome with other cultivars due to common ancestry or gene flow. Even at higher K, Structure approach did not differentiate the two cultivated populations. However, at K=6, the wild population of WW was split into two clusters, followed by WJ population at K=7.

Epigenetic diversity and structure in wild and cultivated soybean

A total of 1452 reproducible methylated fragments were detected with a set of eight primer pairs based on MSAP in all the samples. Across the 1452 methylated loci, cultivars harbored fewer total loci and private loci as compared with the wild samples. Of the 1197 loci present in the cultivars, 168 were

Table 1. Information on five populations of wild and cultivated soybean analyzed in this study.

Collection site	Location	Sample size	Abbreviations	habitat	Environment variables (mean annual temperature and annual rainfall)
Yongqing,Hebei Province	N39.32° E116.49°	19	CY	wetland	11.5℃ 540mm
Qingxian, Hebei Province	N38.57° E116.83°	11	CQ	wetland	12.1℃ 618mm
Anxin, Hebei Province	N38.57° E115. 57°	32	WA	wetland	12.1℃ 552mm
Jizhou, Hebei Province	N37.33° E115.31°	25	WJ	wetland	13.0℃ 519mm
Wuqing, Tianjin	N39.25° E117.16°	37	WW	wetland	11.6℃ 606mm

Fig. 1 Sampling sites of wild soybean populations and cultivation areas of the cultivars soybean analyzed.

Table 2. Diversity measures of wild and cultivated soybean populations.

Populations[a]	Sample Size	AFLP			MSAP		
		PPL	I	H	PPL	I	H
CY	19	82.02%	0.368	0.236	75.76%	0.357	0.232
CQ	11	65.87%	0.326	0.213	65.84%	0.332	0.218
WA	32	50.33%	0.188	0.117	65.84%	0.286	0.184
WJ	25	40.53%	0.171	0.109	58.54%	0.271	0.177
WW	37	54.90%	0.227	0.146	72.73%	0.301	0.191
Cult. (over all)	30	88.64%	0.371	0.236	82.30%	0.365	0.235
Wild (over all)	94	73.44%	0.270	0.171	88.43%	0.335	0.211
T	124	100%	0.333	0.204	100%	0.368	0.223

[a] Populations abbreviations are defined in Table 1. PPL, the percentage of polymorphic loci; I, Shannon-winner diversity index; H, Nei's gene diversity.

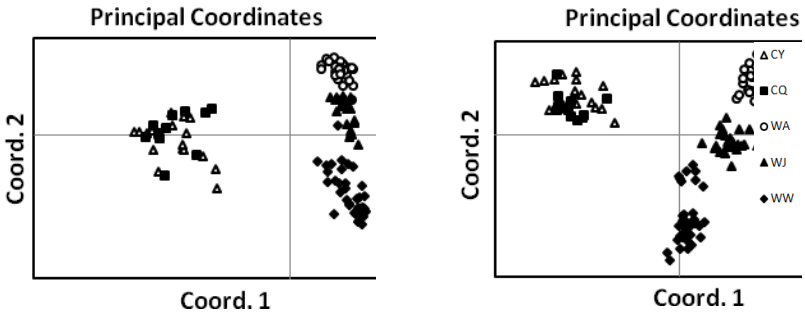

Fig 2. Plots of the first two principal coordinate of 124 soybean accessions from five populations, generating using GenAlEx 6.1 based on AFLP dataset (the left) and MSAP dataset (the right). CY, CQ, WA, WJ, WW Populations abbreviations are defined in Table 1.

private to cultivated pool. In contrast, wild soybean had 255 private loci out of the total 1284 loci. Estimates of epigenetic diversity parameters for cultivated and wild soybean in all populations are presented in Table 1. Overall, wild pool of local populations showed a higher percentage of polymorphic loci (PPL=88.4%) than the cultivated pool (PPL=82.3%); however, epigenetic diversity as measured both by Shannon diversity and by Nei's gene diversity, which was lower in wild pool (I=0.335, H=0.211) than in cultivars (I=0.365, H=0.235). At the population level, all three measures showed the highest (PPL=75.8%, I=0.357, H=0.232) in CY population, whereas the lowest (PPL=58.5%, I=0. 271, H=0.177) was observed in WJ population.

When the AFLP and MSAP analysis were compared on the same set of plants, a major theme revealed. The epigenetic diversity measured by values of I (0.271-0.301) and H (0.177-0.191) were significantly higher than the genetic diversity using the same index (I =0.171-0.227, H =0.109-0.146) in the three wild populations (F-test，P<0.001). However, the genetic and epigenetic diversity were differed slightly in the two cultivated populations (F-test, P= 0.181 -0.531).

In the analysis of hierarchical AMOVA, the proportion of epigenetic variance caused by differences between species and among populations was 7.7% and 11.0% (Table 3), respectively, and both were significant (P<0.01). Pairwise comparisons among all the populations were also significant at epigenetic level but the comparisons between the two cultivated populations (Table 4). The pattern of epigenetic differentiation was similar to the genetic one. However, the levels of epigenetic differentiations based on Φ values between species (Φrt=0.12), among populations (Φpr=0.25) and within populations (Φpt=034) were lower than genetic differentiations (Φrt=0.08, Φpr=0.12, Φpt=0.19). In addition, Pairwise AMOVA of populations also identified much lower differentiation at epigenetic level (Φst =0.002-0.184) than genetic level (Φst =0.002-0.352).

A PCA was conducted on the epigenetic distance matrix based on MSAP data (Fig. 2). Results indicated that the first two principal coordinates accounted for 34.6% and 23.9% of the total variance, respectively. Based on these coordinates, the admixed accessions were separated into five clusters. The first cluster consisted of individuals from two cultivated populations (CY and CQ). The second and third clusters corresponded to WA and WJ populations, respectively. The WW population was separated into two clusters, which were the fourth and fifth clusters.

The Structure analysis of the MSAP data agreed with the AFLP data in some respects (Fig 4). Firstly, the two species were clearly delineated at K=2 on both datasets. Moreover, based on the two molecular marker systems, all the cultivated accessions tended to be clustered together, whereas, wild soybean was differentiated into three clusters, which were largely concordant with their geographical positions. Finally, for both datasets, cultivars were not clearly grouped into two clusters even at higher K.

In addition, The Structure results of the MSAP data varied somewhat from AFLP dataset. Firstly, ΔK showed a clear peak at K=2 in the structure analysis of the MSAP, which indicated that two was the optimal number of epigenetic clusters, whereas Evanno's adhoc DK method determined the K = 4 to be the most likely number of genetic clusters for AFLP data. Secondly, at K=3, WJ accessions were assigned into bⅢ cluster with a high average ancestry (87.7%) in AFLP analysis. However, Structure analysis of the MSAP assigned WJ accessions into BⅢ cluster with a lower average ancestry (59.1%) at K=3, which suggested that more accessions from WJ population were jointly assigned to more than one cluster, probably due to admixture in analysis on MSAP. Furthermore, at K=5, we found that 29 accessions from WW population had an ancestry exceeding 80% in cluster d Ⅰin analysis on MSAP. At this threshold of 80%, 32 WW accessions could be attributed to cluster DⅤ in AFLP analysis. This result reflected more WW accessions tended to be jointly assigned to more than one cluster in analysis on MSAP than that on AFLP. A similar trend was observed at K=4 for comparisons between MSAP and AFLP.

Discussions

Extent of diversity in cultivated and wild soybean

Generally, cultivated crops harbor less genetic diversity in contrast to their wild progenitors because of artificial selection during domestication. Xu et al. (2011) resequenced the genomes of 40 cultivated rice and 10 wild accessions and identified significantly lower diversity in cultivated rices, compared to wild accession. In barely, cultivated accessions exhibited a 50% reduction in diversity when compared with their wild progenitor (Matus and Hayes, 2002). Similarly, the analysis of the nucleotide diversity in wild and cultivated sunflower revealed that the cultivated sunflower had retained only 40–50% of the diversity present in the wild (Liu and Burke, 2006). An SSR survey on maize revealed that the cultivar gene pool had roughly 88% of the gene diversity present in its wild progenitor (Vigouroux et al., 2005). In soybean, Kuroda et al. (2010) found that crop gene pool contained roughly 50% of SSR diversity of wild progenitor. However, a SNP survey provided a slight difference of molecular diversity between the cultivated and wild soybean (Li et al., 2010). Our results indicated that the genetic and epigenetic diversity in the cultivated soybean exceeded that of wild soybean at local population level. In our study, we sampled individuals from local cultivated and wild populations on a small spatial scale. Moreover, soybean is a self-pollinated crop with an outcross rate of 1.8%, in which the gene flow is very difficult in natural habitats between wild populations (Ray et al., 2010). However, under condition of artificial domestication, detectable loci in cultivated soybean might be influenced by cross-breeding, which lead to bulking and mixing of cultivated gene pool. A reason for greater amount of diversity in cultivated accessions could be due to gene flow from geographically distant populations by crossbreeding.

Interestingly, the genetic and epigenetic diversity were differed slightly (F-test，P=0.181, 0.531) in the cultivars, while, the epigenetic diversity were significantly higher (F-test，P<0.001) than the genetic diversity in the wild populations. Epigenetic characterization of this study was achieved by examining patterns of DNA methylation, which was environmentally induced (Crews et al., 2007). A possible explanation was that a fluctuating environment in natural habitats might lead to high level of variation in DNA methylation in wild populations. This supported the view that DNA methylation, unlike genetic modifications, might occur rapidly in response to fluctuating environments and could; therefore, represent a potential way to cope with environmental stress (Rando and Verstrepen, 2007).

Patterns of genetic differentiation in cultivated and wild soybean

Extensive large-scale surveys of genetic variation from broad geographic samples have demonstrated the existence of genetic

Table 3. Analyses of molecular variance (AMOVA) for wild and cultivated soybean populations based on AFLP and MSAP datasets.

	Source of variation	df	% variation	Φ-statistics*	P(Φ)
AFLP	Between species (wild vs.cultivated)	1	11.95	Φrt=0.12	0.010
data	Among populations	3	22.28	Φpr=0.25	0.010
	Within populations	119	65.76	Φpt=034	0.010
MSAP	Between species (wild vs.cultivated)	1	7.73	Φrt=0.08	0.010
data	Among populations	3	10.95	Φpr=0.12	0.010
	Within populations	119	81.31	Φpt=0.19	0.010

* Significant tests of differentiation (1) between species (Φrt), (2) among populations (Φpr) and (3) within populations (Φpt)

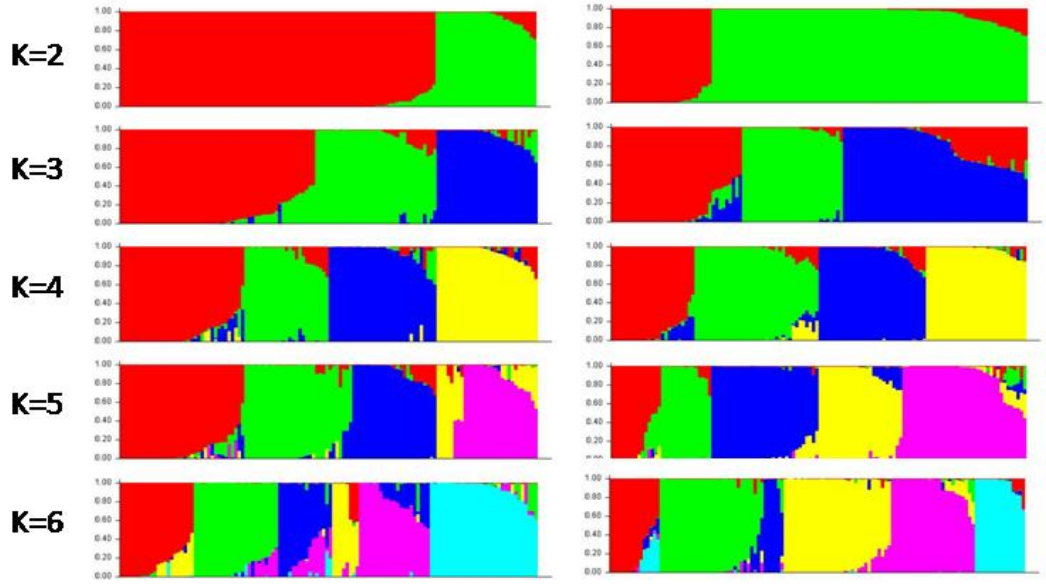

Fig 3. Population structure inferred by Bayesian clustering approaches using Structure based on AFLP (on the left) and on MSAP (on the right). Each sample is shown by a thin vertical line, divided into K (2-7) coloured fragments.Each colour represents one cluster and each fragment represents the membership fraction for one individual in K clusters.

Table 4. Pairwise AMOVA 's Φst among the soybean populations.

Populations[a]		Cultivated		Wild		
		CY	CQ	WA	WJ	WW
Cultivated	CY	0.000	0.002	0.184 *	0.171 *	0.147 *
	CQ	0.002	0.000	0.202 *	0.183 *	0.151 *
Wild	WA	0.312 *	0.343 *	0.000	0.138 *	0.165 *
	WJ	0.298 *	0.342 *	0.322 *	0.000	0.133 *
	WW	0.273 *	0.311 *	0.352 *	0.307 *	0.000

[a] Populations abbreviations are defined in Table 1. Results from genetic markers are presented below the diagonal and those from epigenetic markers are presented above the diagonal. An asterisk denotes statistical significance.

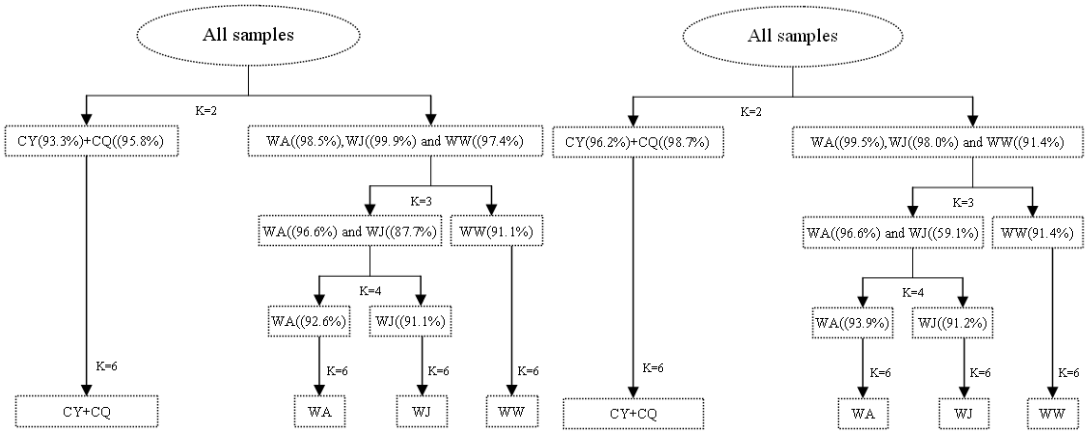

Fig 4. Schematic clustering procedure during inferring population structure using Stucture, based on AFLP (the left) and MSAP (the right) for wild soybean and cultivated soybean. CY, CQ, WA, WJ, WW Populations abbreviations are defined in Table 1.

Table 5. Percentage of average ancestry in a defined cluster for each population in results of Structure analyses based on AFLP and MSAP datasets.

AFLP data Populations[a]	K=2 clusters		K=3 clusters			K=4 clusters					
	a Ⅰ	a Ⅱ	b Ⅰ	b Ⅱ	b Ⅲ	c Ⅰ	c Ⅱ	c Ⅲ	c Ⅳ		
CY	6.73%	93.27%	92.33%	6.38%	1.29%	7.55%	0.35%	1.22%	90.91%		
CQ	4.43%	95.57%	93.37%	2.55%	4.08%	2.57%	0.96%	2.52%	93.95%		
WA	98.50%	1.50%	1.33%	2.01%	96.66%	2.29%	4.30%	92.57%	0.82%		
WJ	99.93%	0.07%	0.17%	12.09%	87.74%	7.57%	91.17%	1.23%	0.01%		
WW	97.41%	2.59%	1.64%	91.13%	7.20%	91.46%	2.66%	4.29%	1.56%		
	K=5 clusters					K=6 clusters					
	d Ⅰ	d Ⅱ	d Ⅲ	d Ⅳ	d Ⅴ	e Ⅰ	e Ⅱ	e Ⅲ	e Ⅳ	e Ⅴ	e Ⅵ
CY	2.78%	0.31%	0.46%	33.26%	63.18%	63.32%	0.38%	0.63%	33.61%	1.89%	0.15%
CQ	0.24%	1.88%	0.82%	21.96%	75.09%	75.31%	0.63%	0.11%	21.93%	0.04%	1.96%
WA	2.23%	92.45%	4.18%	1.09%	0.02%	0.13%	4.20%	0.11%	1.03%	1.87%	92.61%
WJ	7.49%	1.19%	91.21%	0.08%	0.00%	0.04%	91.30%	1.43%	0.03%	6.30%	0.85%
WW	91.26%	3.97%	2.42%	0.87%	1.45%	0.09%	1.97%	42.00%	0.55%	53.51%	1.84%
MSAP data Populations[a]	K=2 clusters		K=3 clusters			K=4 clusters					
	A Ⅰ	A Ⅱ	B Ⅰ	B Ⅱ	B Ⅲ	C Ⅰ	C Ⅱ	C Ⅲ	C Ⅳ		
CY	96.25%	3.75%	2.79%	95.98%	1.23%	1.75%	1.39%	0.78%	96.05%		
CQ	98.66%	1.34%	1.41%	98.01%	0.59%	1.52%	0.75%	0.25%	97.48%		
WA	0.51%	99.49%	2.92%	0.49%	96.60%	4.98%	0.93%	93.89%	0.18%		
WJ	2.06%	97.94%	38.70%	2.21%	59.09%	91.17%	2.61%	6.07%	0.11%		
WW	8.60%	91.40%	91.41%	4.11%	4.48%	4.76%	89.21%	2.17%	3.86%		
	K=5 clusters					K=6 clusters					
	D Ⅰ	D Ⅱ	D Ⅲ	D Ⅳ	D Ⅴ	E Ⅰ	E Ⅱ	E Ⅲ	E Ⅳ	E Ⅴ	E Ⅵ
CY	48.97%	48.26%	0.62%	1.20%	0.94%	48.56%	0.17%	3.32%	0.27%	0.55%	47.08%
CQ	36.25%	62.80%	0.10%	0.63%	0.23%	35.99%	0.32%	0.26%	0.05%	0.63%	62.73%
WA	0.53%	0.31%	93.98%	4.48%	0.71%	0.49%	0.34%	0.46%	93.78%	4.78%	0.14%
WJ	0.28%	0.16%	6.26%	90.80%	2.51%	0.23%	2.24%	0.52%	5.82%	91.02%	0.13%
WW	1.60%	3.06%	1.70%	4.68%	88.96%	0.95%	78.54%	18.28%	0.75%	1.29%	0.15%

[a] Populations abbreviations are defined in Table 1.

differentiation in wild and cultivated soybean populations (Wang and Takahata, 2007; Wen et al., 2009; Li et al., 2010). In some researches on small-scale populations of wild soybean, the significant genetic differentiation has been also found (Fu et al., 2002; Yan et al., 2008; Wang et al., 2012) except for some special populations such as the populations along the river ecosystems because of the river effect (Kiang et al., 1992; Fujita et al., 1997). In the present study, the Structure, AMOVA, and PCA analyses all revealed the presence of two genetically distinct species of wild and cultivated soybean, greater level of genetic differentiation among local wild populations ($0.307<\Phi st<0.352$) and lower level of differentiation between cultivated populations ($\Phi st=0.02$) on a small spatial scale.

The gene flow, random genetic drift, and natural or artificial selection are primary evolutionary forces that cause population differentiation. Wild soybean is a selfing species and its gene flow between populations is confined (Wang and Takahata, 2007), which should contribute significantly to contemporary genetic structure among wild populations even on a small scale in the present study. In contrast, poorly differentiated structure in cultivated populations can be explained by crossbreeding, which results in gene flow within and among populations and lessens their genetic differences to some extent.

Natural selection by ecological factors will result in development of ecological adaptation and divergence, and the selection could act on both genetic and epigenetic variation (Richards et al., 2010). This view was supported by the existence of epigenetic divergence between wild and cultivated species and among the wild soybean populations in the present study. More specially, we found the lower level of epigenetic differentiation than genetic differentiation in both AMOVA and Pairwise AMOVA. In addition, The Structure analysis of MSAP suggested more admixed accessions (accessions assigned to more than one cluster) in wild populations than AFLP analysis. Considered together, these results may reflect that the individuals we studied had greater epigenetic similarity than genetic one. Epigenetic variation, unlike genetic variation, may be altered directly by ecological interactions; which therefore, provides an accelerated pathway for evolutionary change (Bossdorf et al., 2008). The environmental factors including the temperature and rainfall in our five studied populations we were similar, and the habitats were all wetlands (Table 1). Moreover, the seeds were germinated under the same greenhouse conditions in a common environment. We thought the reason for such epigenetic similarity in our study was basically due to the adaptive convergence to similar environments because some epi-loci were prone to convergent selection in a common environment.

Materials and Methods

Plant materials

94 individuals of wild soybean representing three isolated populations were sampled from its natural habitats in Northern Huang Huai region of China, which all located in the region of

N37-39°×E115-119°. 26-38 accessions were collected from each wild population. For comparison to local cultivars, 30 local and typical cultivated accessions were selected in the immediate neighborhood of the three populations, including 11 individuals from Yongqing County and 19 from Qingxian County. The cultivated accessions were all obtained from Chinese National Soybean GeneBank (CNSGB). A map representing the locations of three wild populations and the two cultivated populations is provided in Fig.1. The information of the populations is listed in Table 1.

DNA extraction, AFLP and MSAP protocols

Because DNA methylation was environmentally labile, seeds from each accession of cultivated and wild soybean were germinated in potted plastic trays under the same greenhouse conditions. After two weeks, the seedling leaves were harvested. Subsequently, genomic DNA was extracted from fresh leaves of a single seedling per accession using the common CTAB method.

We screened a total of 124 individuals on the five populations for genetic variation using AFLP protocol, which employed *Eco*RI as rare cutter and *Mse*I as frequent cutter (Vos et al., 1995). The same set of plants was used to analyze the epigenetic variation, which was assessed through cytosine DNA methylation polymorphism using methylation-sensitive amplified polymorphism (MSAP). MSAP (Reyna-Lopez et al., 1997) is a modified version of the standard amplified fragment length polymorphism (AFLP) technique, by replacing the *Mse*I enzyme with the enzyme either *Hpa*II or *Msp*I, both of which recognize the same restriction site (5'-CCGG) but have different sensitivity to methylation at the inner and outer cytosine. So, for the MS-AFLP, we ran two separate protocols using essentially the same AFLP protocol and four different types of variation occurred: (1) the restriction site was cut by both enzymes, indicating no methylation; (2) *Msp*I did not cut and *Hpa*II cut, indicating a hemi-methylated external cytosine; (3) *Msp*I cut and *Hpa*II did not cut, indicating the restriction site has a full-methylated internal cytosine or hemi- methylated internal cytosine; (4) neither enzymes could cut, caused by either fragment absence or hypermethylation.

The experimental procedures for MSAP and AFLP were as described (Dong et al., 2006). The primer information of MSAP and AFLP was given in Supplementary Tables 1 and 2. DNA fragments were separated on an automated DNA sequencer (ABI Prism-3100 Genetic Analyzer; Applied Biosystems). The size and intensity of each DNA fragment displayed in the electropherogram was interpreted with GeneMapper 4.0 software (Applied Biosystems). Only DNA fragments between 70 basepairs (bp) and 500 bp were scored for each accession with scoring done by the same person.

Data analysis

The AFLP or MSAP bands were scored as binary characters for absence (0) or presence (1). For MSAP data, we transformed the raw data matrix resulting from the *Eco*RI /*Hpa*II and *Eco*RI /*Msp*I profiles into a binary data matrix for statistical analyses and computation. A Methylation scoring approach (Schulz et al., 2013) was used to extract the binary epigenetic information, which only considered the methylated fragments as relevant and scored these as '1' (type 2 and 3 above), and unmethylated fragments and absence fragments were both scored as '0' (type 1 and 4 above).

The resulting data of AFLP and MSAP were processed using Excel 2007. GenAlEx version 6.1 (Peakall and Smouse，2006) software package was used to calculate genetic and epigenetic diversity parameters (the percentage of polymorphic loci, PPL; Shannon-winner diversity index, I; and Nei's gene diversity, H) per locus per population. The significance of differences in diversity parameters (I and H) between cultivated and wild soybean, and between genetic and epigenetic variation in one population was tested using FSTAT.

The GenAlEx was also used to calculate estimates of population differentiation using an analysis of molecular variance (AMOVA) framework. We performed hierarchical AMOVA, which partitioned the total variation into three levels and yielded independent estimates of differentiation between the wild and cultivated groups (Φrt), among populations (Φpr) and within populations (Φpt). We then performed pairwise AMOVA among populations to assess the proportion of total variance that was partitioned between two populations, which resulted in values of Φst. In both cases, statistical significance was tested using 1000 permutations. To graphically represent genetic/epigenetic relationships among soybean individuals, principal coordinate analysis (PCA) was conducted on Nei's genetic/epigenetic distance matrixes based on AFLP and MSAP, respectively. The analysis was carried out in GenAlEx 6.1

The population structure in the whole cultivated and wild accessions was examined with a Bayesian Markov Chain Monte Carlo method, which was implemented in the Structure 2.1 software (Pritchard et al., 2000). We used an admixture and independent allele frequency model, which assumed that the genome of each individual was a mixture of genes originating from K unknown ancestral populations. The approach was applied to AFLP and MSAP datasets separately. The program was run by a number of clusters (K) from 2 to 7 with a burn-in period of 100,000 followed by 100,000 iterations, and this was repeated five times for each value of K. The most likely number of populations (K) was estimated using the maximal value of L(K), an estimate of the posterior probability of the data for a given K (Zeisset and Beebee,2001), and a peak value of△K ,an ad hoc quantity based on the second order rate of change of the likelihood function with respect to K (Evanno et al., 2005).

Conclusions

These results indicate that genetic differentiation occurs between the cultivated soybean and its wild progenitor, and among wild populations on a small spatial scale. In addition, the presence of epigenetic divergence has also been evidenced. So, it could be inferred that natural selection might act on epigenetic variation in a similar manner on genetic one. However, AMOVA reflects lower level of epigenetic differentiation than genetic differentiation. The Structure analysis of MSAP suggests more admixed accessions than analysis of AFLP; both reflect greater epigenetic similarity than genetic in the studied individuals. The epigenetic similarity is basically due to faster adaptive convergence to similar environments.

Given higher diversity in cultivated soybean than its wild progenitor at local population level, it is plausible that gene flow is common in cultivated gene pool from geographically distant populations by crossbreeding. The cultivated accessions might not originated independently from local wild soybean. Furthermore, the epigenetic diversity was significantly higher than the genetic diversity in the wild populations, while the genetic and epigenetic diversity were differed slightly in the cultivars. It thus seems that epigenetic variation may occur more rapidly in natural environments than genetic one.

Acknowledgements

The study was supported by the Special Fund for Agro-scientific Research in the Public Interest (201003021),the Natural Science Foundation of China (31000141), the National "863"Program(2012AA101106-3) and the Science and Technology Development Program of Jilin Province (No.20130102050JC)

References

Abe J, Hasegawa A, Fukushi H, Mikami T, Ohara M, Shimamoto Y (1999) Introgression between wild and cultivated soybeans of Japan revealed by RFLP analysis for chloroplast DNAs. Econ Bot. 53(3): 285-291.

Angers B, Castonguay E, Massicotte R (2010) Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. Mol Ecol. 19(7): 1283-1295.

Ashikawa I (2001) Surveying CpG methylation at 5′-CCGG in the genomes of rice cultivars. Plant Mol Biol. 45(1): 31-39.

Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. Ecol Lett. 11(2): 106-115.

Cervera MT, Ruiz-Garcia L, Martinez-Zapater J (2002) Analysis of DNA methylation in Arabidopsis thaliana based on methylation-sensitive AFLP markers. Mol Genet Genomics. 268(4): 543-552.

Crews D, Gore AC, Hsu TS, Dangleben NL, Spinetta M, Schallert T, Skinner MK (2007) Transgenerational epigenetic imprints on mate preference. Proc Natl Acad Sci USA. 104(14): 5942-5946.

Dong YS, Zhuang BC, Zhao LM, Sun H, He MY (2001) The genetic diversity of annual wild soybeans grown in China. Theor Appl Genet. 103(1): 98-103.

Dong Z Y, Wang YM, Zhang ZJ, Shen Y, Lin XY, Ou XF, Liu B (2006) Extent and pattern of DNA methylation alteration in rice lines derived from introgressive hybridization of rice and Zizania latifolia Griseb. Theor Appl Genet. 113(2): 196-205.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14(8): 2611-2620.

Fu YL, Qian J, Ma YH, Li J, Zheng SZ (2002) Genetic differentiation research on populations of wild soybeans in different scales. Acta Ecologica Sinica. 2: 005.

Fujita R, Ohara M, Okazaki K, Shimamoto Y (1997) The extent of natural cross-pollination in wild soybean (*Glycine soja*). J Hered. 88(2), 124-128.

Herrera CM, Bazaga P (2010) Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. New Phytol. 187(3): 867-876.

Jin Y, He T, Lu BR (2003) Fine scale genetic structure in a wild soybean (*Glycine soja*) population and the implications for conservation. New Phytol. 159(2): 513-519.

Kiang YT, Chiang YC, Kaizuma N (1992) Genetic diversity in natural populations of wild soybean in Iwate Prefecture, Japan. J Hered. 83(5): 325-329.

Kuroda Y, Kaga A, Tomooka N, Vaughan DA (2006) Population genetic structure of Japanese wild soybean (*Glycine soja*) based on microsatellite variation. Mol Ecol. 15(4): 959-974.

Kuroda Y, Kaga A, Tomooka N, Vaughan D (2010) The origin and fate of morphological intermediates between wild and cultivated soybeans in their natural habitats in Japan. Mol Ecol. 19(11): 2346-2360.

Li Y, Guan R, Liu Z, Ma Y, Wang L, Li L, Qiu L (2008) Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. Theor Appl Genet, 117(6): 857-871.

Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single‐nucleotide polymorphism loci. New Phytol.188(1): 242-253.

Li Y, Shan X, Liu X, Hu L, Guo W, Liu B (2008) Utility of the methylation-sensitive amplified polymorphism (MSAP) marker for detection of DNA methylation polymorphism and epigenetic population structure in a wild barley species (*Hordeum brevisubulatum*). Ecol Res. 23(5): 927-930.

Li Z, Nelson RL (2002) RAPD marker diversity among cultivated and wild soybean accessions from four Chinese provinces. Crop Sci. 42(5): 1737-1744.

Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics. 173(1): 321-330.

Massicotte R, Angers B (2011) General-purpose genotype or how epigenetics extend the flexibility of a genotype. Genetics Research International, 2012.317175

Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome. 45(6): 1095-1106.

Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 6(1): 288-295.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics. 155(2): 945-959.

Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. Cell. 128(4): 655-668.

Rapp RA, Wendel JF (2005) Epigenetics and plant evolution. New Phytol. 168(1): 81-91.

Ray JD, Kilen TC, Abel CA, Paris RL (2010) Soybean natural cross-pollination rates under field conditions. Environ Biosafety Res. 2(2): 133.

Reyna-Lopez GE, Simpson J, Ruiz-Herrera J (1997) Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. Mol Gen Genet. 253(6): 703-710.

Richards CL, Bossdorf O, Pigliucci M (2010) What role does heritable epigenetic variation play in phenotypic evolution?. BioScience. 60(3): 232-237.

Richards CL, Bossdorf O, Verhoeven KJ (2010) Understanding natural epigenetic variation. New Phytol. 187(3): 562-564.

Richards CL, Schrey AW, Pigliucci M (2012) Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation. Ecol Lett. 15(9): 1016-1025.

Salmon A, Clotault J, Jenczewski E, Chable V, Manzanares-Dauleux MJ (2008) *Brassica oleracea* displays a high level of DNA methylation polymorphism. Plant Sci. 174(1), 61-70.

Schrey AW, Coon CA, Grispo MT, Awad M, Imboma T, McCoy ED, Martin LB (2012) Epigenetic variation may compensate for decreased genetic variation with introductions: a case study using house sparrows (Passer domesticus) on two continents. Genetics Res Int. 2012. 979751

Schulz B, Eckstein RL, Durka W (2013) Scoring and analysis of methylation‐sensitive amplification polymorphisms for epigenetic population studies. Mol Ecol Resour. 13(4): 642-653.

Singh RJ, Hymowitz T (1999) Soybean genetic resources and crop improvement. Genome. 42(4): 605-616.

Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Martienssen RA (2007) Epigenetic natural variation in *Arabidopsis thaliana*. Plos Biology. 5(7): e174.

Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Doebley J (2005) An analysis of genetic diversity across the maize genome using microsatellites. Genetics. 169(3): 1617-1630.

Vos P, Hogers R, Bleeker M, Reijans M, van De Lee T, Hornes M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23(21): 4407-4414.

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 Ev. 54(1): 157-165.

Wang KJ, Li XH, Liu Y (2012) Fine-scale phylogenetic structure and major events in the History of the current wild soybean (Glycine soja) and taxonomic assignment of semi-wild type (*Glycine gracilis* Skvortz.) within the Chinese subgenus Soja. J Hered. 103(1): 13-27.

Wen Z, Ding Y, Zhao T, Gai J (2009) Genetic diversity and peculiarity of annual wild soybean (*G. soja* Sieb. et Zucc.) from various eco-regions in China. Theor Appl Genet. 119(2): 371-381.

Wen ZX, Zhao TJ, Zheng YZ, Liu SH, Wang CE, Wang F, Gai JY (2008) Association analysis of agronomic and quality traits with SSR markers in *Glycine max* and *Glycine soja* in China: I. Population structure and associated markers. Acta Agronomica Sinica. 34(7): 1169-1178.

Xiong LZ, Xu CG, Maroof MS, Zhang Q (1999) Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. Mol Gen Genet. 261(3): 439-446.

Xu DH, Abe J, Sakai M, Kanazawa A, Shimamoto Y (2000) Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. Theor Appl Genet. 101(5-6): 724-732.

Xu DH, Gai JY (2003) Genetic diversity of wild and cultivated soybeans growing in China revealed by RAPD analysis. Plant Breeding. 122(6): 503-506.

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Wang W (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol. 30(1): 105-111.

Yan MF, Li XH, Wang KJ (2008) Evaluation of genetic diversity by SSR markers for natural populations of wild soybean (*Glycine soja*) growing in the region of Beijing. China J Plant Ecol (China). 32: 938-950.

Zeisset I, Beebee TJ (2001) Determination of biogeographical range: an application of molecular phylogeography to the European pool frog Rana lessonae. P R Soc London. Series B: Biological Sciences. 268(1470): 933-938.