Genetic diversity and the conservation priority of *Glycine soja* populations from Northern China

Y.L. Niu[1,2], W.Y. Guo[3], L.R. Bai[2] and J.C. Zhao[1]

[1]College of Life Science, Hebei Normal University, Shijiazhuang, China
[2]College of Life Science, Hengshui University, Hengshui, Hebei, China
[3]College of Life Science and Technology, Xinxiang Medical University, Xinxiang Henan, China

Corresponding author: J.C. Zhao
E-mail: zhaojiancheng@mail.hebtu.edu.cn

Genet. Mol. Res. 14 (4): 16608-16615 (2015)
Received April 16, 2015
Accepted July 15, 2015
Published December 11, 2015
DOI http://dx.doi.org/10.4238/2015.December.11.8

ABSTRACT. Knowledge of the spatial patterns of genetic variation in wild populations has significant implications for *in situ* conservation and the determination of conservation order. To study the levels of genetic diversity, spatial genetic structures, and genetic distances in *Glycine soja*, 11 natural populations in northern China were analyzed by estimating genetic coefficients using inter-simple sequence repeat (ISSR) fingerprints via mixed sampling strategies. Sixteen ISSR primers generated 98 reproducible polymorphic amplification banding patterns of 172 scored, accounting for 56.98% of the polymorphisms among the populations. The dendrogram based on Nei's genetic distance showed that distinct genetic differentiation occurred in *G. soja*. The Unweighted Pair-Group Method with Arithmetic Mean cluster analysis indicated two broad groups, and one contained all of the populations except three from Chengde, which formed the smaller second group. The spatial genetic structure evident in the wild soybean populations may be attributed to restricted seed dispersal and the dominant breeding system of this species. The detection of genetic

structures in wild soybean populations could be a significant index for the effective conservation of many wild populations, and it could be exploited by soybean breeding programs to increase production.

Key words: Conservation strategy; Genetic structure; Wild soybean; Inter-simple sequence repeat

INTRODUCTION

The wild soybean (*Glycine soja*), distributed throughout East Asia, has many valuable characteristics, including high protein content, strong disease resistance, and a large reproduction coefficient. It is also valuable to the study of the origin and evolution of the cultivated soybean (Zhuang et al., 1999). It is the direct ancestor of the cultivated soybean, so it serves as a precious resource for soybean improvement (Lu, 2004). Studies of wild and cultivated soybean accessions in seed banks have shown that the wild soybean has greater genetic variation than the cultivated soybean (Chen and Nelson, 2004).

Genetic diversity in wild soybean has been previously evaluated based on seed bank accessions (Chen and Nelson, 2004; Zhao et al., 2006; Dong, 2008) and natural populations. According to an investigation of intra-population and among-population genetic diversity, more than 6570 samples of *G. soja* were collected and preserved in the Chinese national storehouse in addition to the core collection that was gathered in China (Zhao et al., 2005).

To ensure effective utilization and continued availability of genetic resources in the wild soybean gene pool, particularly at present when many wild soybean populations have disappeared or diminished due to habitat changes, it is necessary to consider the strategic conservation of this wild relative through studies of its genetic diversity. Large-scale gene lapses in wild soybean resources were caused by the deteriorating environment. Although *ex situ* conservation of germplasm is a secure and effective protection method, only preservation under the original ecological environmental conditions can give rise to unceasing variation, in which *G. soja* evolves novel genes to adapt to adverse circumstances and conditions. Therefore, the preservation of the original habitat is the most effective method to protect wild soybean resources. However, the wild soybean is widely distributed in China, and although each soybean population has conservation value, there are limited funds and resources. Thus, how to determine the populations with the most value and the highest conservation priority is an important question (Chen et al., 2002; Li et al., 2005). The determination reflects the contribution and the importance of different populations of this species, so it must be effectively protected (Chen et al., 2002). To our knowledge, an investigation of wild soybean conservation has not been previously reported. This study adopted the mixed sample strategy to examine the genetic structure of 11 natural wild soybean populations, providing a specific scientific basis for their protection.

MATERIAL AND METHODS

Sample collection and DNA preparation

In this study, 11 populations were selected from Hebei and Jiangsu Provinces where the

altitude is from 2 to 1372 m, the longitude is from 115°33'E to 119°47'E, and the latitude is from 33°46'N to 41°29'N. The specific information for these 11 populations is listed in Table 1. The mature seeds of all samples were randomly collected from the populations.

Table 1. Geographic distribution information for *Glycine soja* populations used in this study.

Name of population	Distribution location	Ecological environment	Longitude and latitude	Altitude (m)
HCQ	Hebei Chengde Qijia	River bank	41°29.276'N 118°06.715'E	628
HCL	Hebei Chengde Liaoheyuantou	Moist valley	41°17.584'N 118°25.746'E	1372
HCZ	Hebei Chengde Zezhou Reservoir	Seasonally flooded bottomland	41°07.772'N 118°39.564'E	619
HZY	Hebei Zhangjiakou Yuxian Yangjiaping	River bank	39°58.156'N 115°23.016'E	893
HQH	Hebei Qinhuangdao Coast Of Bohai	Both sides of the road	39°41.577'N 119°18.984'E	5
HBA	Hebei Anxin Baiyangdian	Bulrush pond	38°50.101'N 115°55.187'E	7
HCN	Hebei Cangzhou Nandagang	Artificial waterway	38°34.524'N 117°22.271'E	3
HHD	Hebei Hengshui Dazhang	Swamp	37°32.439'N 115°33.306'E	17
JXJ	Jiangsu Xuzhou Jiawang	River bank	34°23.624'N 117°29.997'E	28
JXB	Jiangsu Xuzhou Biantang	Sides of farmland	34°24.186'N 117°39.428'E	24
JYF	Jiangsu Funing County	Tongyu river bank	33°46.286'N 119°47.046'E	2

The mature wild soybean seeds were dipped in sulfuric acid (98.0%) for 10 min, and were then flushed four times with ddH_2O. These seeds were cultured in vermiculite for one week under 1500 lx light conditions. Total DNA was extracted from ~150 mg young leaves using the Plant Genomic DNA Kit (DP305-02, Tiangen Biotech Co., Ltd., Beijing, China). The genomic DNA was quantified on 0.8% agarose gels, and the DNA concentration was determined using a Nanodrop 2000c (Thermo Fisher Scientific Inc.). All DNA samples were stored at -20°C for ISSR analysis.

ISSR amplification and electrophoresis

Eighty ISSR primers (synthesized by Sangong Biotech Co. Ltd., Shanghai, China) were screened using DNA samples from five populations to select primers that were suitable for PCR amplification. In total, 16 ISSR primers were found to produce clear and reproducible DNA bands by electrophoresis (Table 2), and were consequently used to amplify the wild soybean samples. Amplification conditions were optimized using appropriate $MgCl_2$ concentrations and annealing temperatures. To test the reliability and reproducibility of ISSR fragments, all samples were assayed at least three times using the selected 16 ISSR primers, and any detected unreliable fragments were ignored in the consequent scoring.

PCR was performed in a total volume of 25 µL containing 20 ng total DNA, 2.5 µL 10X PCR buffer (without $MgCl_2$), 0.25 mM each dNTP, 6 pM of each primer, and 0.5 U Taq DNA polymerase. The concentration of $MgC1_2$ ranged from 2.00 to 3.25 mM. PCR was performed using a DNA Thermal Cycler (Eppendorf Mastercycler Gradient). PCR conditions for all samples were: 4 min initial denaturation step (94°C); 40 cycles of 40 s at 94°C, 45 s at each specific annealing temperature (the optimum annealing temperature was determined for each primer), and 90 s at

72°C; and 7 min at 72°C. Amplified DNA fragments were separated on 2.2% agarose gels at 90 W for 4 h in 1X TBE buffer (100 mM Tris-borate, pH 8.0, 2 mM EDTA). The gel was stained with ethidium bromide, visualized under UV light, and photographed using the SynGene Genius system. The molecular sizes of the amplified fragments were estimated using Lambda DNA digested by *Eco*RI and *Hin*dIII (TaKaRa Dalian Biotechnology Co., Ltd., China).

Genetic diversity estimation and phylogenetic tree construction

Polymorphic bands were recognized by the Gene Tools Analysis Software (Version 3.03.03) and confirmed by visual inspection. Each polymorphic band generated by ISSR was scored for its presence (1) or absence (0). From the binary data, the genetic diversity among the individual populations was calculated according to Nei's unbiased statistic using the Popgene32 software (version 1.32). Nei's gene diversity, Shannon's information index, genetic similarity, genetic distance, and the percentage of polymorphism among the populations were also interpreted using Popgene32. A cluster analysis based on genetic distances was undertaken according to the equations, using the Unweighted Pair-Group Method with Arithmetic Mean method (UPGMA) to generate a dendrogram showing the relationships among populations.

RESULTS

Levels of polymorphism revealed by ISSR-PCR markers

Of the five populations that were prescreened using 80 ISSR primers, 16 markers generated bright amplification products and polymorphisms, and these were used in subsequent analyses (Table 2). A total of 172 reliable fragments were obtained. The number of bands per primer ranged from 7 to 15 with an average of 11.8 bands. Of these, 98 bands were polymorphic, accounting for 56.98% of the variation. The number of polymorphic bands per primer ranged from 2 to 10. The percentage of polymorphic bands (PPB) for each ISSR marker ranged from 28.60 to 83.30% (Table 2).

Table 2. Inter-simple sequence repeat (ISSR) primers used to analyze *Glycine soja* populations and the polymorphisms determined with these primers.

Primer	Sequence (5′-3′)	No. of total loci	No. of polymorphic loci	Percentage of polymorphic loci
UBC807	$(AG)_8T$	10	6	60.0
UBC810	$(GA)_8T$	9	3	33.3
UBC813	$(CT)_8T$	10	6	60.0
UBC818	$(CA)_8G$	12	6	50.0
UBC822	$(TC)_8A$	14	10	71.4
UBC824	$(TC)_8G$	13	9	69.2
UBC827	$(AC)_8G$	9	4	44.4
UBC828	$(TG)_8A$	10	5	50.0
UBC834	$(AG)_8YT$	15	10	66.7
UBC835	$(AG)_8YC$	13	6	46.2
UBC840	$(GA)_8YT$	10	5	50.0
UBC846	$(CA)_8RT$	7	2	28.6
UBC849	$(GT)_8YA$	9	6	66.7
UBC856	$(AC)_8YA$	12	6	50.0
UBC880	$(GGAGA)_3$	7	3	42.9
UBC881	$(GGGTG)_3$	12	10	83.3

Genetic distance and variation

The Shannon's indices (I) ranged from 0.0459 to 0.1140 at the population level. Among the 11 populations, population HQH exhibited the greatest level of variability (PPB = 19.19%; I = 0.1140). The mean observed number of alleles (N_A) ranged from 1.0841 for JXF to a maximum of 1.1919 for HQH (Table 3). Values of the effective number of alleles, N_E, were less than those for N_A for every population, and they ranged from 1.0532 for JYF to 1.1401 for HQH. The mean Nei's gene diversity (H) ranged from 0.0310 for JYF to 0.0783 for HQH. I ranged from 0.0459 to 0.1140 with an average of 0.0894 among populations, which demonstrated that there was 3.92-8.50% heterozygosis within the *G. soja* population. The inter-population genetic distance (*D*) ranged from 0.1012 to 0.3115, which suggested that samples from HBA were distantly related to HZY samples (Table 3). Among the 11 populations, the mean coefficient of gene differentiation (G_{ST}) was 0.7201, accounting for 28.00% of the total genetic diversity within the populations. Based on the G_{ST} value, the mean estimated gene flow number (N_m) between populations was 0.1943 (Table 4), indicating significant genetic diversity of different populations. The results revealed that of the 11 populations, the HQH population exhibited the greatest variability (PPB = 19.19%; I = 11.40%) (Table 3).

Table 3. Genetic diversity analysis of 11 *Glycine soja* populations.

Population	No. of mixed sample	Total loci	Polymorphic loci	Proportion of polymorphic loci (%)	Shannon's index (I)	Nei's index (H)	Observed number of alleles (N_A)	Effective number of alleles (N_E)
HCQ	3	172	22	12.79	0.0752	0.0515	1.1279	1.0914
HCL	3	172	22	12.79	0.0705	0.0471	1.1279	1.0792
HCZ	3	172	29	16.86	0.1030	0.0713	1.1686	1.1302
HZY	3	172	24	13.95	0.0808	0.0550	1.1395	1.0964
HQH	3	172	33	19.19	0.1140	0.0783	1.1919	1.1401
HBA	3	172	32	18.60	0.1065	0.0722	1.1860	1.1254
HCN	3	172	30	17.44	0.1010	0.0687	1.1744	1.1204
HHD	3	172	26	15.12	0.0923	0.0639	1.1512	1.1166
JXJ	3	172	32	18.60	0.1113	0.0766	1.1860	1.1376
JXB	3	172	24	13.95	0.0832	0.0571	1.1395	1.1025
JYF	3	172	14	8.14	0.0459	0.0310	1.0814	1.0532
					0.0894			

Cluster analysis

A cluster analysis revealed two major groups of *G. soja* populations. One cluster contained the HCQ, HCL, and HCZ populations, and the other cluster contained the remaining populations (Figure 1). The correlation between Nei's standard genetic distance matrix (Table 4, Table 5) and the cophenetic (ultrametric) dissimilarity values generated from the UPGMA tree (= 0.94) was high, validating the observed clusters.

DISCUSSION

The improvement of soybean germplasm may be limited by genetic bottlenecks during its development. The primary gene pool used to improve soybean germplasm is *G. soja*, the wild ancestor of the cultivated soybean (Hymowitz and Singh, 1998). Research indicates that, in addition to increased tolerance to biotic and abiotic stresses, introducing genetic diversity from the exotic gene pool of wild soybeans to broaden the genetic background of soybean cultivars may

also facilitate an increase in soybean yield. China is the main distribution region of *G. soja*, but the growth environment of *G. soja* has become limited due to economic development. Its distribution region and population numbers have decreased. Some rare varieties of *G. soja*, such as white flower, are at risk of extinction. Therefore, it is necessary to protect wild soybean resources and their genetic diversity. To preserve the genetic richness of these landraces, efforts must be directed towards the implementation of reliable collection and conservation strategies. The determination of conservation priorities not only reflects the importance of different populations for the species, but also represents those that can be effectively conserved (Chen et al., 2002). To determine conservation priorities, the evolutionary potential of a species and the protection of its genetic variability should be considered first (Sushma and Rana, 2011). Furthermore, species conservation and recovery measures should be based on the genetic structure of the species.

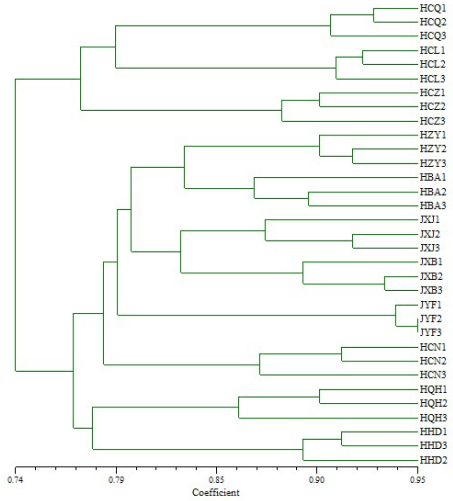

Figure 1. Dendrogram for the 11 wild soybean populations based on ISSR markers.

Table 4. Genetic differentiation among five *Glycine soja* populations.

Primer	No. of mixed samples	Total gene diversity (H_T)	Within population diversity (H_S)	Genetic differentiation among populations (G_{ST})	Gene flow (N_m)
UBC807	33	0.17148	0.06990	0.47480	0.55308
UBC810	33	0.09430	0.07577	0.16770	2.48152
UBC813	33	0.23337	0.12516	0.45333	0.60294
UBC818	33	0.22757	0.10288	0.53147	0.44079
UBC822	33	0.28577	0.04286	0.85147	0.08722
UBC824	33	0.27273	0.03359	0.88518	0.06486
UBC827	33	0.19687	0.07048	0.63485	0.28759
UBC828	33	0.18797	0.02218	0.79360	0.13004
UBC834	33	0.24630	0.03865	0.85574	0.08429
UBC835	33	0.19153	0.04880	0.76467	0.15388
UBC840	33	0.21855	0.07503	0.64908	0.27032
UBC846	33	0.09201	0.00779	0.94310	0.03017
UBC849	33	0.22992	0.02499	0.80140	0.12391
UBC856	33	0.19500	0.03068	0.83590	0.09816
UBC880	33	0.16484	0.07160	0.55783	0.39633
UBC881	33	0.41747	0.16440	0.62494	0.30008
Mean		0.21850	0.06110	0.72010	0.19430
SD		0.0440	0.0077		

Table 5. Nei's genetic identity (above diagonal) and genetic distance (below diagonal) for *Glycine soja* populations.

Population	HCQ	HCL	HCZ	HZY	HQH	HBA	HCN	HHD	JXJ	JXB	JYF
HCQ	****	0.8414	0.8125	0.7557	0.7835	0.8089	0.7324	0.8153	0.8058	0.7681	0.7601
HCL	0.1727	****	0.8504	0.8041	0.7792	0.8261	0.7954	0.7811	0.8021	0.8178	0.8170
HCZ	0.2076	0.1621	****	0.7894	0.7867	0.8078	0.8195	0.8102	0.8372	0.8059	0.8106
HZY	0.2801	0.2181	0.2365	****	0.8578	0.9038	0.8551	0.8053	0.8494	0.8704	0.8758
HQH	0.2440	0.2495	0.2399	0.1533	****	0.8750	0.8365	0.8510	0.8376	0.8160	0.8156
HBA	0.2121	0.1910	0.2134	0.1012	0.1336	****	0.8535	0.8229	0.8993	0.8613	0.8303
HCN	0.3115	0.2289	0.1991	0.1565	0.1785	0.1584	****	0.8458	0.8702	0.8398	0.8424
HHD	0.2042	0.2471	0.2105	0.2166	0.1613	0.1950	0.1674	****	0.8576	0.8233	0.8380
JXJ	0.2160	0.2206	0.1777	0.1633	0.1772	0.1061	0.1391	0.1536	****	0.9013	0.8403
JXB	0.2639	0.2011	0.2158	0.1388	0.2034	0.1494	0.1746	0.1945	0.1040	****	0.8641
JYF	0.2744	0.2021	0.2100	0.1327	0.2038	0.1859	0.1715	0.1768	0.1740	0.1461	****

Genetic diversity of *G. soja* populations

The goal of *G. soja* conservation is the protection of its genetic diversity. Research on the genetic diversity of *G. soja* was conducted over the past decade (Qian and Chen, 1998; Chen and Nelson, 2004), and many suggestions for its conservation have been put forward (Lu, 2004). However, a study of *G. soja* population conservation has not been previously reported. ISSR markers are more efficient than other molecular marker systems in the detection of polymorphisms, and they have been extensively used to assess the genetic diversity of *G. soja* (Lu, 2004) and other species (Zhang and Dai, 2010). Hence, we used ISSR markers to determine the populations that should be conserved. In the present study, ISSR markers were used to estimate the genetic diversity within and among different *G. soja* populations. The mixed sampling strategy, which eliminates the genetic diversity among different populations, enables the samples to best represent and more effectively detect the genetic intra-population diversity. Sixteen of the 80 ISSR primers used yielded banding patterns that were clear, which could be scored with confidence (Table 2). Among the 172 loci, 98 bands were polymorphic, accounting for 56.98% of the variation. The number of polymorphic bands per primer ranged from 2 to 10. The PPB (of each ISSR marker ranged from 28.60 to 83.30% (Table 2). The difference among populations was not obvious because of the mixed sampling strategy (Table 3). Their value order was as follows: HQH > HBA = JXJ > HCN > HCZ > HHD > JXB = HZY > HCL = HCQ > JYF. I was calculated according to King and Schaal (1989), and a higher I value indicates higher genetic diversity. The order of these 11 populations was HQH > JXJ > HBA > HCZ > HCN > HHD > JXB > HZY > HCQ > HCL > JYF. The order of H was the same as for I, and the results of the PPB were identical.

Genetic differentiation among populations

Wright (1996) hypothesized that when N_m was greater than one, gene flow existed between populations. Thus, since the average value of N_m was less than one in our study, the results implied that no or very slow gene flow occurred among these 11 populations. The total gene diversity (H_T) was 0.2185, and within-population diversity (H_S) was 0.0611. The G_{ST} ranged from 0.1677 to 0.9431. These results suggested that more genetic mutations occurred among populations and fewer existed within populations. The mixed sampling strategy adopted by the study might account for these results.

In conclusion, based on the genetic diversity analysis of 11 *G. soja* populations, the HQH and HBA populations should be given priority in conservation measures.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Natural Science Foundation of Hebei Province (Grant #C2006000147), the Foundation of Pilot for Comprehensive Reform of the Biological Science Specialties in Hebei Province, and the Key Disciplines of Hengshui University. We also wish to thank Dr. Xian-Jun PENG, who assisted with the preparation and proofreading of data and the manuscript.

REFERENCES

Chen XY, Lu HP, Shen L and Li YY (2002). Identifying populations for priority conservation of important species. *Biodivers. Sci.* 10: 332-38.

Chen YW and Nelson RL (2004). Genetic variation and relationships among cultivated, wild, and semiwild soybean. *Crop Sci*. 44: 316-325.

Dong YS (2008). Advances of research on wild soybean in China. *J. Jilin Agric. Univ*. 30: 394-400.

Hymowitz T, Singh RJ and Kollipara KP (1998). The genomes of the *Glycine*. *Plant Breed. Rev.* 16: 289-318.

Lu BR (2004). Conserving biodiversity of soybean gene pool in the biotechnology era. *Plant Species Biol*. 19: 115-125.

Qian J and Chen Y (1998). Genetic variation of wild soybean among populations from different latitude. *J. Fudan Univ*. 37: 208-212.

Sushma V and Rana TS (2011). Genetic diversity within and among the wild populations of *Murraya koenigii* (L.) Spreng as revealed by ISSR analysis. *Biochem. Syst. Ecol.* 39: 139-144.

Zhang JP, Sheng SK, Gao H and Wang YH (2010). Research advancement on the conservation genetics of wild soybean in China. *Seed* 29: 57-60.

Zhao LM, Dong YS, Liu B, Hao S, et al. (2005). The reconstruction of core resources of *Glycine soja* in China. *Chin. Sci. Bull.* 50: 992-999.

Zhao R, Cheng Z, Lu WF and Lu BR (2006). Estimating genetic diversity and sampling strategy for a wild soybean (*Glycine soja*) population based on different molecular markers. *Chin. Sci. Bull.* 51: 1219-1217.

Zhuang BC (1999). Researches on wild soybean (*Glycine soja)* in china for twenty years. *J. Jilin Agric. Sci*. 24: 3-10.