

Genetic diversity analysis using simple sequence repeat markers in soybean

Zhenbin Hu¹, Guizhen Kan¹, Guozheng Zhang¹, Dan Zhang², Derong Hao³ and Deyue Yu^{1*}

¹National Center for Soybean Improvement, National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, People's Republic of China, ²Department of Agronomy, Henan Agricultural University, Zhengzhou 450002, People's Republic of China and ³Nantong Institute of Agricultural Sciences, Nantong, Jiangsu 226541, People's Republic of China

Abstract

To evaluate the genetic diversity (GD) of wild and cultivated soybeans and determine the genetic relationships between them, in this study, 127 wild soybean accessions and 219 cultivated soybean accessions were genotyped using 74 simple sequence repeat (SSR) markers. The results of the study revealed that the GD of the wild soybeans exceeded that of the cultivated soybeans. In all, 924 alleles were detected in the 346 soybean accessions using 74 SSRs, with an average of 12.49 alleles per locus. In the 219 cultivated soybean accessions, 687 alleles were detected, with an average of 9.28 alleles per locus; in the 127 wild soybean accessions, 835 alleles were detected, with an average of 11.28 alleles per locus. We identified 237 wild-soybean-specific alleles and 89 cultivated-soybean-specific alleles in the 346 soybean accessions, and these alleles accounted for 35.28% of all the alleles in the sample. Principal coordinates analysis and phylogenetic analysis based on Nei's genetic distance indicated that all the accessions could be classified into two major clusters, corresponding to wild and cultivated soybeans. These results will increase our understanding of the genetic differences and relationships between wild and cultivated soybeans and provide information to develop future breeding strategies to improve soybean yield.

Keywords: genetic diversity; principal coordinate analysis; soybean

Introduction

Soybean (*Glycine max* (L.) Merr.), an important cash crop, was domesticated in China approximately 3000 to 5000 years ago and is currently cultivated mainly in China, the USA, Brazil and Argentina. Soybean can provide 69 and 30% of dietary protein and oil, respectively (Lam *et al.*, 2010). Many studies on soybean genetic diversity (GD) have been carried out using different molecular markers and different germplasms, all of which showed wild soybeans to be more genetically

diverse than cultivated soybeans (Gai *et al.*, 2000; Xu and Gai, 2003; Kuroda *et al.*, 2006).

The assessment of GD in cultivated soybeans and their wild relatives will provide useful information for soybean improvement, especially regarding the use of wild soybeans in future soybean breeding programmes. Molecular marker techniques have been widely used to characterize GD in wild and cultivated germplasms (Xu and Gai, 2003; Vigouroux *et al.*, 2005; Li *et al.*, 2010). Simple sequence repeat (SSR) markers possess a number of advantages over other types of molecular markers, such as high polymorphism, co-dominance, locus specificity, good reproducibility, and distribution throughout the genome, and have been widely used for assessing GD, constructing linkage maps and mapping

*Corresponding author. E-mail: dyyu@njau.edu.cn

quantitative trait locus (QTL) in several crops, such as soybean, rice, maize and winter wheat (Vigouroux *et al.*, 2005; Wen *et al.*, 2009; Zhang *et al.*, 2009, 2010). Many SSR markers have been developed for soybean and extensively utilized in genetic map construction, QTL mapping and GD assessment (Song *et al.*, 2004; Hwang *et al.*, 2009; Li *et al.*, 2010; Xu *et al.*, 2010).

In this study, 74 SSR markers were used to evaluate the GD of wild and cultivated soybeans and to determine the genetic relationships between them.

Materials and methods

Plant materials

A total of 346 soybean accessions, including 219 cultivated accessions and 127 wild accessions, were used in this study. The 219 cultivated accessions have been described by Chao *et al.* (2014); 113 of the 127 wild accessions have been described by Hu *et al.* (2014) with another 14 lines from Heilongjiang province in China. All the plant materials were provided by the National Center for Soybean Improvement of China.

Seventy-four SSR markers, selected from published genetic maps (Song *et al.*, 2004; Hwang *et al.*, 2009), were used to genotype the 346 accessions. Genotyping was performed using the method described previously by Wen *et al.* (2009).

The total number of alleles, GD and polymorphism information content (PIC) were calculated using the PowerMarker version 3.25 software (Liu and Muse, 2005). To evaluate the reduction in GD, the Δ GD parameter was used as described by Vigouroux *et al.* (2005), where Δ GD = $1 - G_C/G_W$ and G_C and G_W are the GD of the cultivated and wild soybeans, respectively. The Kruskal–Wallis (KW) test was carried out using R language to determine the statistical significance of the differences in GD between the cultivated and wild soybean accessions (R Development Core Team, 2010).

Principal coordinate analysis (PCoA) based on Nei's (1972) genetic distance was conducted using the

DARwin version 5.0 software to estimate the genetic relationship among the accessions (Perrier and Jacquemoud-Collet, 2006). The first two principal components were plotted for visual examination of the grouping pattern of the accessions.

A neighbour-joining tree was constructed using a neighbour-joining algorithm based on the natural log transformation of the proportion of shared alleles, and it is available in the DARwin version 5.0 software package (Perrier and Jacquemoud-Collet, 2006).

Results and discussion

Genetic diversity

In all the 346 accessions, 924 alleles were detected using the 74 SSRs, with an average of 12.49 alleles per locus. The number of alleles per locus ranged from two (GMES3693) to 28 (satt210). In the 219 cultivated soybean accessions, 687 alleles were detected, with an average of 9.28 alleles per locus. In the 127 wild soybean accessions, 835 alleles were detected, with an average of 11.28 alleles per locus (Table 1). There were 237 wild-soybean-specific alleles and 89 cultivated-soybean-specific alleles, which indicated that many alleles had been lost during domestication and a few had been acquired or produced. These changes may have occurred as only a limited number of accessions were used during domestication (Doebley *et al.*, 2006; Caicedo *et al.*, 2007). These altered alleles accounted for 35.28% of all the alleles in the sample. It is relatively common for domesticated specimens to exhibit less GD than their wild relatives. In general, they exhibit approximately two-thirds of the GD of their wild relatives (Buckler *et al.*, 2001).

In the whole sample, GD ranged from 0.02 (GMES3693) to 0.93 (satt614), with an average of 0.71, and the PIC ranged from 0.02 (GMES3693) to 0.93 (satt614), with an average of 0.68. In the cultivated accessions, GD ranged from 0.00 to 0.93, with an average of 0.64, and the PIC ranged from 0.00 to 0.93, with

Table 1. Summary of the statistical analysis of genetic diversity (GD) and of differences between the wild and cultivated soybeans

Index	Total		Cultivated soybeans		Wild soybeans		P value
	Mean	Range	Mean	Range	Mean	Range	
AN	12.49	2–28	9.28	1–27	11.28	2–25	<0.01
GD	0.71	0.02–0.93	0.64	0.00–0.93	0.75	0.06–0.93	<0.01
PIC	0.68	0.02–0.93	0.60	0.00–0.93	0.73	0.06–0.93	<0.01

AN, total allele number; PIC, polymorphism information content.

an average of 0.60. In the wild accessions, GD ranged from 0.06 to 0.93, with an average of 0.75, and the PIC ranged from 0.06 to 0.93, with an average of 0.73 (Table 1).

Comparison of GD between the cultivated and wild soybean populations

The wild soybean accessions had more alleles, more GD and higher PIC values than the cultivated soybean accessions ($P < 0.01$) (Table 1), despite the smaller sample size of the wild soybean accessions than the cultivated soybean accessions (127 compared with 219). The cultivated soybean accessions were found to be considerably less diverse than the wild soybean accessions with a 23.05% deficit in the cultivated soybean accessions relative to the wild soybean accessions, which is less than the deficit detected in an analysis carried out by Hyten *et al.* (2006). Kuroda *et al.* (2006) analysed the microsatellite variation of 616 Japanese wild soybean lines and 53 cultivated soybean lines using 20 SSR markers and found that Nei's diversity value of the cultivated soybean lines was only 57% of that of the wild soybean lines. The KW test carried out in the present study indicated that the wild soybean accessions had more GD, higher PIC values and more alleles per locus than the cultivated soybean accessions.

Genetic structure of the wild and cultivated soybean populations

To further assess the relationships among the accessions, PCoA was conducted using DARwin version 5.0. Two groups of accessions were identified, with axis 1 discriminating between the cultivated and wild soybean accessions. The apparent sharing between groups must be carefully considered because the first two axes represented 7.01% of the global diversity. We were able to determine genetic proximity. The first coordinate explained 4.65% of the variation and the second coordinate explained 2.36% (Fig. 1(a)).

Phylogenetic analysis

The neighbour-joining tree, constructed based on genetic distances, showed two major clusters: *G. max* and *G. soja* (Fig. 1(b)). We found some mix lines, which is consistent with the results of the PCoA. We also found the genetic distances among the wild soybean accessions to be greater than those among the cultivated soybean accessions (Fig. 1(b)).

The results of this study indicated that wild soybeans possess higher GD than cultivated soybeans and we were able to classify all the accessions into two major clusters, corresponding to the wild and cultivated soybeans.

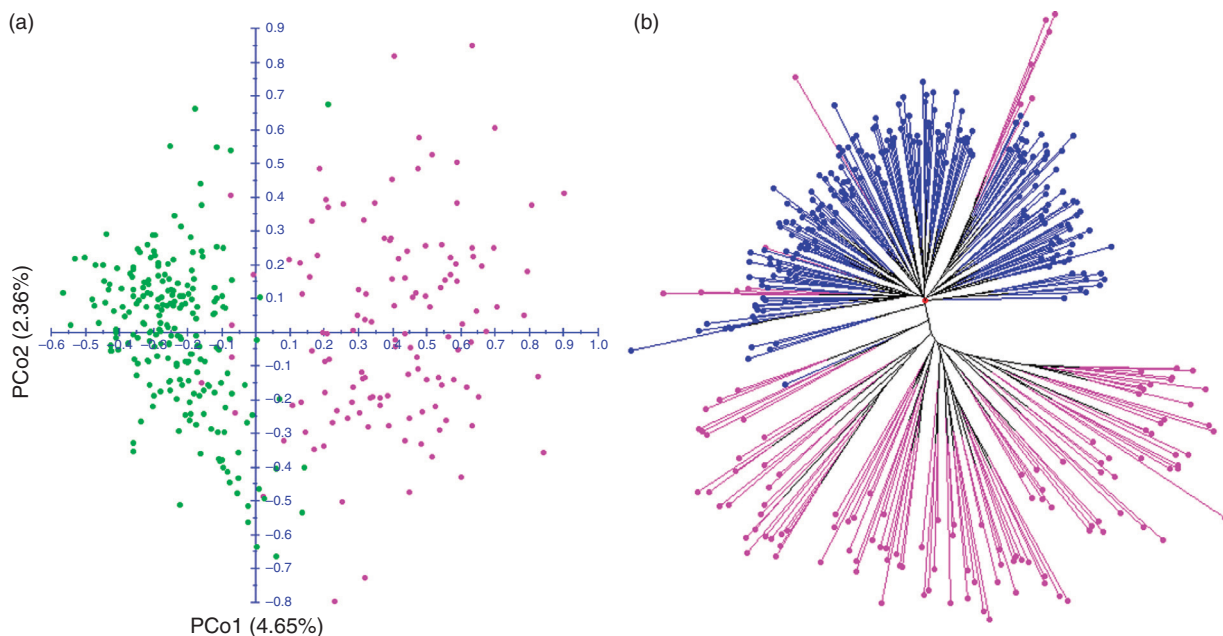


Fig. 1. Principal coordinate analysis (PCoA) and neighbour-joining tree of the 346 accessions. (a) PCoA of the cultivated and wild soybeans. PCo1 and PCo2 are the first and second principal coordinates, which account for 4.65 and 2.36% of the total variation, respectively. Green spots represent cultivated soybeans and red spots represent wild soybeans. (b) Neighbour-joining tree constructed for the 346 accessions using the log-transformed proportion of the shared allele distance. Blue lines represent cultivated soybeans and red lines represent wild soybeans.

These results increase our understanding of the genetic differences and relationships between wild and cultivated soybeans and provide information that can be used for the development of future breeding strategies to improve soybean yields.

Acknowledgements

This study was supported by the National Basic Research Program of China (973 Program) (2010CB125906) and the National Natural Science Foundation of China (31171573, 31271749, and 31301341).

References

- Buckler ES, Thornsberry JM and Kresovich S (2001) Molecular diversity, structure and domestication of grasses. *Genetical Research* 77: 213–218.
- Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, Bustamante CD and Purugganan MD (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* 3: 1745–1756.
- Chao M, Yin Z, Hao D, Zhang J, Song H, Ning A, Xu X and Yu D (2014) Variation in Rubisco activase (*RCAB*) gene promoters and expression in soybean [*Glycine max* (L.) Merr.]. *Journal of Experimental Botany* 65: 47–59.
- Doebley JF, Gaut BS and Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127: 1309–1321.
- Gai JY, Xu DH, Gao Z, Shimamoto Y, Abe J, Fukushi H and Kitajima S (2000) Studies on the evolutionary relationship among eco-types of *G. max* and *G. soja* in China. *Acta Agronomica Sinica* 26: 513–520.
- Hu Z, Zhang D, Zhang G, Kan G, Hong D and Yu D (2014) Association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). *Breeding science* 63: 1–9.
- Hwang TY, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki H, Hisano H, Sasamoto S, Sato S, Tabata S, Kono I, Hoshi M, Hanawa M, Yano C, Xia ZJ, Harada K, Kitamura K and Ishimoto M (2009) High-density integrated linkage map based on SSR markers in soybean. *DNA Research* 16: 213–225.
- Hyten D, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC and Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences USA* 103: 16666–16671.
- Kuroda Y, Kaga A, Tomooka N and Vaughan AD (2006) Population genetic structure of Japanese wild soybean (*Glycine soja*) based on microsatellite variation. *Molecular Ecology* 15: 959–974.
- Lam HM, Xu X, Liu X, Chen WB, Yang GH, Wong FL, Li MW, He WM, Qin N, Wang B, Li J, Jian M, Wang J, Shao GH, Wang J, Sun SSM and Zhang GY (2010) Re-sequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics* 42: 1053–1059.
- Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS and Qiu L (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytologist* 188: 242–253.
- Liu K and Muse S (2005) Powermarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128–2129.
- Nei M (1972) Genetic distance between populations. *Amer. Naturalist* 106: 283–292.
- Perrier X and Jacquemoud-Collet JP (2006) DARwin software. Available at <http://darwin.cirad.fr/>
- R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE and Cregan PB (2004) A new integrated genetic linkage map of the soybean. *Theoretical and Applied Genetics* 109: 122–128.
- Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSS, Jaqueth J, Smith OS and Doebley JF (2005) An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 169: 1617–1630.
- Wen ZX, Ding YL, Zhao YJ and Gai JY (2009) Genetic diversity and peculiarity of annual wild soybean (*Glycine soja* Sieb. & Zucc.) from various eco-regions in China. *Theoretical and Applied Genetics* 119: 371–381.
- Xu DH and Gai JY (2003) Genetic diversity of wild and cultivated soybeans growing in China revealed by RAPD analysis. *Plant Breeding* 122: 503–506.
- Xu Y, Li H, Li G, Wang X, Cheng L and Zhang Y (2010) Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.). *Theoretical and Applied Genetics* 122: 581–594.
- Zhang D, Zhang H, Wang M, Sun J, Qi Y, Wang F, Wei X, Han L, Wang X and Li Z (2009) Genetic structure and differentiation of *Oryza sativa* L. in China revealed by microsatellites. *Theoretical and Applied Genetics* 119: 1105–1117.
- Zhang D, Bai G, Zhu C, Yu J and Carver B (2010) Genetic diversity, population structure, and linkage disequilibrium in U.S. elite winter wheat. *Plant Genome* 3: 117–127.