© NIAB 2014
ISSN 1479-2621

Plant Genetic Resources: Characterization and Utilization (2014) **12**(S1); S45–S48
doi:10.1017/S1479262114000239

Genetic diversity and population structure of wild soybean (*Glycine soja* Sieb. and Zucc.) accessions in Korea

Kil Hyun Kim[1], Seukki Lee[2], Min-Jung Seo[1], Gi-An Lee[3], Kyung-Ho Ma[3], Soon-Chun Jeong[4], Suk-Ha Lee[5,6], Eui Ho Park[7], Young-Up Kwon[1] and Jung-Kyung Moon[1]*

[1] *National Institute of Crop Science, Rural Development Administration, Suwon, Republic of Korea,* [2] *Technology Cooperation Bureau, Rural Development Administration, Suwon, Republic of Korea,* [3] *National Academy of Agricultural Science, Rural Development Administration, Suwon, Republic of Korea,* [4] *Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Chungbuk, Republic of Korea,* [5] *Plant Genomics and Breeding Institute, Seoul National University, Seoul, Republic of Korea,* [6] *Department of Plant Science and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea and* [7] *School of Biotechnology, Yeungnam University, Gyeongbuk, Republic of Korea*

Abstract

Genetic variation in wild soybean (*Glycine soja* Sieb. and Zucc.) is a valuable resource for crop improvement efforts. Soybean is believed to have originated from China, Korea, and Japan, but little is known about the diversity or evolution of Korean wild soybean. Therefore, in this study, we evaluated the genetic diversity and population structure of 733 *G. soja* accessions collected in Korea using 21 simple sequence repeat (SSR) markers. The SSR loci produced 539 alleles (25.7 per locus) with a mean genetic diversity of 0.882 in these accessions. Rare alleles, those with a frequency of less than 5%, represented 75% of the total number. This collection was divided into two populations based on the principal coordinate analysis. Accessions from population 1 were distributed throughout the country, whereas most of the accessions from population 2 were distributed on the western side of the Taebaek and Sobaek mountains. The Korean *G. soja* collection evaluated in this study should provide useful background information for allele mining approach and breeding programmes to introgress alleles into the cultivated soybean (*G. max* (L). Merr.) from wild soybean.

Keywords: genetic diversity; Korea; population structure; rare alleles; wild soybean

Introduction

Soybean (*Glycine max* (L.) Merr) is an important legume crop, providing oil and protein for human and animal consumption, and its oil is a major source for biodiesel production (Pimentel and Patzek, 2005; Singh *et al.*, 2007). Most of the elite soybean cultivars were developed from a narrow genetic base, derived from a limited number of ancestral lines. Wild soybean (*Glycine soja* Sieb. and Zucc.) is presumed to be the undomesticated progenitor of the closest relative of soybean to *G. max*. Therefore, various genetic accessions in *G. soja* are expected to be a valuable genetic resource for soybean crop improvement. Recently, whole-genome sequencing of *G. max* var. Williams 82 and wild soybean (*G. soja* var. IT182932) has been completed in the USA and Korea, respectively

*Corresponding author. E-mail: moonjk2@korea.kr

https://doi.org/10.1017/S1479262114000239 Published online by Cambridge University Press

(Kim *et al.*, 2010; Schmutz *et al.*, 2010). The consensus sequence of *G. soja* spanned 915.4 Mb, covering 97.65% of the Williams 82 genome sequence. *G. soja* distributed in Eastern Asia including eastern China, Korea, Japan and eastern Russia has important phenotypic characteristics and specific alleles that are not present in *G. max* (Hymowitz and Singh, 1987; Carter *et al.*, 2004). Interestingly, Kuroda *et al.* (2009) reported that a higher proportion of rare alleles are present in wild soybean from Korea than in that from other countries. Several studies have been conducted using electrophoresis or simple sequence repeat (SSR) markers to estimate genetic diversity within the Korean wild soybean (Yu and Kiang, 1993; Choi *et al.*, 1999; Lee *et al.*, 2008). However, a limited number of *G. soja* accessions from Korea were used in the previous studies. In this study, genotypes of 733 *G. soja* accessions collected from Korea were used to estimate genetic diversity and population structure.

Materials and methods

A total of 733 *G. soja* accessions originating from Korea were used in this study (see online supplementary Table S1). Genomic DNA was extracted from young leaves using a modified CTAB procedure (Keim *et al.*, 1988). Based on the soybean genetic map (Song *et al.*, 2004), a total of 21 SSR markers were randomly selected by greater polymorphic information content (PIC). The M13-tail PCR method labelled with one of the fluorescent dyes 6-FAM, NED or HEX (Applied Biosystems, Foster City, CA, USA) was used to measure the size of the PCR products (Schuelke, 2000). SSR alleles were resolved on the ABI 3730 × 1 automatic DNA sequencer (Applied Biosystems). Allelic differences in SSR markers were analysed using the GeneMapper 3.0 software (version 3.7; Applied Biosystems). For overall genetic diversity assessment of the collection, the observed number of alleles and rare alleles, genetic diversity, PIC per locus and Shannon's information index were calculated using POPGENE 1.31 and POWERMARKER 3.25 (Nei, 1973; Yeh *et al.*, 1999; Liu and Muse, 2005; see online supplementary Table S2). To estimate genetic differentiation among the populations, F_{st} statistics with AMOVA was implemented using Arlequin 3.11 (Excoffier *et al.*, 2005; see online supplementary Table S3). For population structure analysis, STRUCTURE 2.34 analysis was carried out using a burn-in period of 10,000 Markov chain Monte Carlo iterations, a run length of 10,000 and a model allowing for admixture and correlated allele frequencies (Pritchard *et al.*, 2000; Fig. 1(A) and (B)). Ten independent runs were performed for each simulated value of the true number of populations (K) from 1 to 15. For each value of K, the estimated log probability of data Pr($X|K$) was used to calculate ΔK (Evanno *et al.*, 2005). Principal coordinate analysis was carried out based on Nei's genetic distance matrix among the populations using GenAlEx 6.5 (Nei, 1978; Peakall and Smouse, 2012; Fig. 1(C) and online supplementary Table S4).

Results and discussion

The genetic diversity and population structure of 733 *G. soja* accessions collected in Korea were evaluated

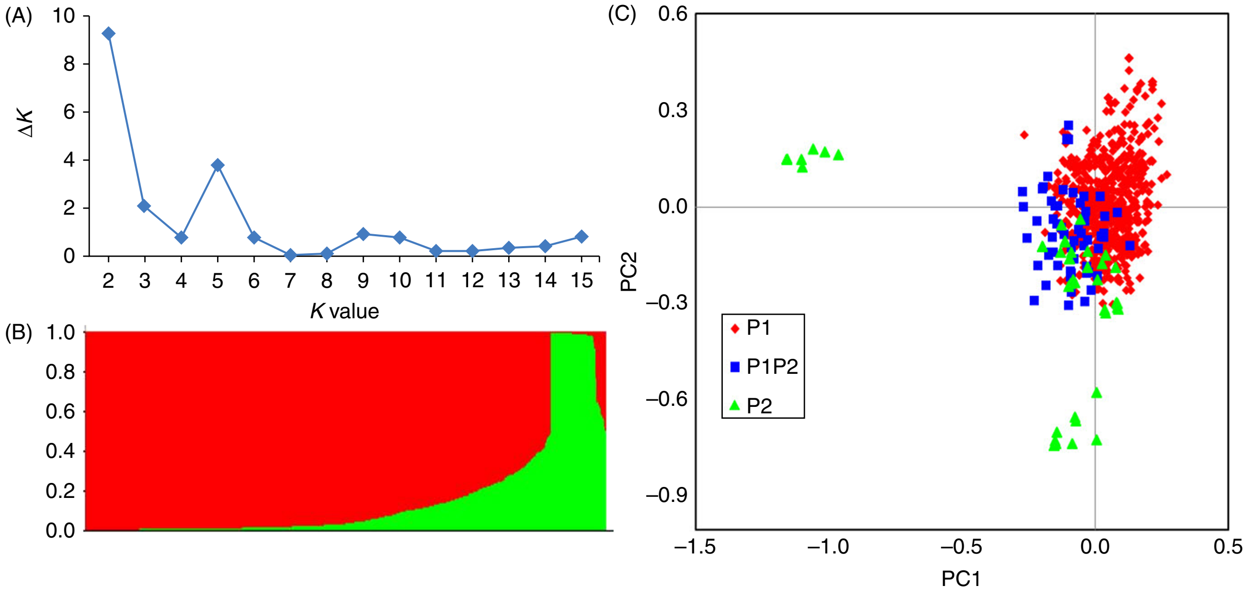

Fig. 1. Model-based populations of *Glycine soja* accessions in Korea. (A) ΔK values for detecting a true K in the STRUCTURE analysis. (B) Population structure of 733 *G. soja* accessions ($K = 2$). (C) Principal coordinate analysis of three populations (P1, P1P2 and P2).

https://doi.org/10.1017/S1479262114000239 Published online by Cambridge University Press

using 21 SSR markers (see online supplementary Tables S1 and S2). A total of 539 alleles were identified, ranging from 18 (Satt070) to 33 (Satt237 and Satt382) per locus with an average allelic richness value of 25.7 (see online supplementary Table S2). Mean values of the estimated Shannon's information index and genetic diversity for all the markers were 2.58 and 0.882, respectively. PIC values ranged from 0.722 (Satt184) to 0.945(Satt373), with an average value of 0.873. In the previous studies using wild soybean from China, Japan, Russia and Korea, the SSR loci produced 364 (18.2 per locus), 405 (20.25 per locus), 462 (23.1 per locus) and 322 (16.1 per locus) alleles, respectively (Kuroda *et al.*, 2009; Wen *et al.*, 2009). Although the average number of alleles in wild soybean from Korea was lower than that in wild soybean from other countries, the number of rare alleles in wild soybean from Korea was much higher (Kuroda *et al.*, 2009). A total of 539 alleles (25.7 per locus) were detected in the Korean wild soybean collection evaluated in the present study. Comparing the previous studies with the present study, it is difficult to determine whether the average number of alleles in wild soybean from Korea is much higher than that in wild soybean from other countries because of the different markers and populations employed. Interestingly, 406 (75% of the total number) rare alleles, present at a frequency of less than 5%, were identified in this study. Among these, 59 (10% of the total number) were unique alleles, among which only one allele was detected in a SSR marker (see online supplementary Table S2), suggesting that many accession-specific alleles were present in the collection. AMOVA of populations P1 and P2 revealed that 7.1% of the variation was due to differences among the populations and 92.9% was due to differences within the populations (see online supplementary Table S3). This indicates that there is significant geographical differentiation in the Korean wild soybean.

Population structure was estimated with STRUCTURE 2.3.4 analysis using 21 SSR markers. The maximum of the *ad hoc* measure ΔK was observed at $K = 2$ (Fig. 1A), indicating that the entire collection was divided into two main populations (namely P1 and P2; Fig. 1B). Of the 733 *G. soja* accessions, 609 were assigned to population P1 and 64 were assigned to population P2 with membership probabilities $>70\%$. The remaining 60 accessions were classified as admixture forms (P1P2). In the principal coordinate analysis, population P2 was divided into three subpopulations based on the cross line (Fig. 1C). As shown in online supplementary Table S4, the main populations and subpopulations reflected the *G. soja* collection sites. Subpopulation 1, consisting of 22 accessions from Gangwon-do and Gyeonggi-do, is positioned at the upper left side in Fig. 1C. Subpopulation 2, consisting of 11 accessions from Eumseong-gun and Goesan-gun in Chungcheongbuk-do, is positioned at the bottom right side. Subpopulation 3, consisting of 31 accessions from diverse provinces, is placed together with P1 and P1P2 populations. Based on its geographical features, the southern part of Korea can be divided into two regions separated by the Taebaek and Sobaek mountains. *G. soja* accessions from populations P1 and P1P2 were distributed throughout the country, while most of the accessions from population P2 were divided into three subgroups, distributed west of the Taebaek and Sobaek mountains.

Overall, the Korean *G. soja* collection had a high mean number of alleles and rare alleles and genetic diversity. Combined with next-generation sequencing technologies, this collection should be a key resource for identifying new genes for soybean improvement with regard to seed yield and resistance to abiotic and biotic stress.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262114000239

Acknowledgements

This study was supported principally by a grant from the Next-Generation BioGreen 21 Program (PJ009091), by the Rural Development Administration, and partly by the National Institute of Crop Science Research Program (PJ009330) as a post-doctoral fellowship.

References

Carter TE Jr, Nelson R, Sneller CH and Cui Z (2004) Genetic diversity in soybean. In: Boerma HR and Specht JE (eds) *Soybeans: Improvement, Production, and Uses*. Madison, WI: American Society of Agronomy, pp. 303–416.

Choi IY, Kang JH, Song HS and Kim NS (1999) Genetic diversity measured by simple sequence repeat variations among the wild soybean, *Glycine soja*, collected along the riverside of five rivers in Korea. *Genes & Genetic Systems* 74: 169–177.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.

Excoffier L, Laval G and Schneider S (2005) Arelequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47–50.

Hymowitz T and Singh RJ (1987) Taxonomy and speciation. In: Wilcox JR (ed.) *Soybeans: Improvement, Production, and Uses*, 2nd edn. Madison, WI: American Society of Agronomy, pp. 23–48.

https://doi.org/10.1017/S1479262114000239 Published online by Cambridge University Press

Keim P, Olson TC and Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. *Soybean Genetics Newsletter* 15: 150–154.

Kim MY, Lee S, Van K, Kim TH, Jeong S-C, Choi I-Y, Kim DS, Lee YS, Park D, Ma J, Kim WY, Kim BC, Park S, Lee KA, Kim D-H, Kim K-H, Shin JH, Jang YE, Kim KD, Liu WX, Chaisan T, Kang YJ, Lee YH, Kim KH, Moon JK, Schmutz J, Jackson SA, Bhak J and Lee S-H (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proceedings of the National Academy of Sciences of the United States of America* 107: 22032–22037.

Kuroda Y, Tomooka N, Kaga A, Wanigadeva SMSW and Vaughan D (2009) Genetic diversity of wild soybean (*Glycine soja* Sieb. and Zucc.) and Japanese cultivated soybeans [*G. max* (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. *Genetic Resources and Crop Evolution* 56: 1045–1055.

Lee JD, Yu JK, Hwang YH, Blake S, So YS, Lee GJ, Nguyen HT and Shannon JG (2008) Genetic diversity of wild soybean (*Glycine soja* Sieb. and Zucc.) accessions from South Korea and other countries. *Crop Science* 48: 606–616.

Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128–2129.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70: 3321–3323.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.

Peakall R and Smouse PE (2012) GenAlEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537–2539.

Pimentel D and Patzek TW (2005) Ethanol production using corn, switchgrass, and wood; biodiesel production using soybean and sunflower. *Natural Resources Research* 14: 65–76.

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

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitors T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Griggs MF, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuranman A, Zhang XC, Shinozake K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC and Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

Singh RJ, Nelson RL and Chung GH (2007) Soybean (*Glycine max* (L.) Merr.). In: Singh RJ (ed.) *Genetic Resources, Chromosome Engineering, and Crop Improvement: Oilseed Crops*, vol. 4. Boca Raton, FL: CRC Press, pp. 13–50.

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.

Wen Z, Diang Y, Zhao T and Gai J (2009) Genetic diversity and peculiarity of annual wild soybean (*G. soja* Sieb. and Zucc.) from various eco-regions in China. *Theoretical and Applied Genetics* 119: 371–381.

Yeh FC, Yang R and Boyle T (1999) POPGENE VERSION 1.31: Microsoft Window-based freeware for population genetic analysis. University of Alberta, Edmonton, AB, Canada.

Yu H and Kiang YT (1993) Genetic variation in South Korean natural populations of wild soybean (*Glycine soja*). *Euphytica* 68: 213–221.

https://doi.org/10.1017/S1479262114000239 Published online by Cambridge University Press