ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2015) **13**(3); 213–218
doi:10.1017/S1479262114000938

# Diversity of the *W1* gene encoding flavonoid 3′,5′-hydroxylase in white- and purple-flowered soybeans

Gyu Tae Park[1†], Jagadeesh Sundaramoorthy[1†], Jong-Beum Park[1], Jeong Dong Lee[1], Kwang Shik Choi[2], Jeong Hoe Kim[2], Hak Soo Seo[3,4], Soon-Ki Park[1] and Jong Tae Song[1*]

[1]*School of Applied Biosciences, Kyungpook National University, Daegu 702-701, Republic of Korea,* [2]*Department of Biology, Kyungpook National University, Daegu 702-701, Republic of Korea,* [3]*Department of Plant Bioscience, Seoul National University, Seoul 151-742, Republic of Korea and* [4]*Bio-MAX Institute, Seoul National University, Seoul 151-818, Republic of Korea*

**Received 1 August 2014; Accepted 24 September 2014 – First published online 8 December 2014**

## Abstract

Cultivated soybeans [*Glycine max* (L.) Merr.] have various flower colours such as dark purple, purple, light purple, pink, magenta, near white and white. About one-third of the soybean accessions in the United States Department of Agriculture – Germplasm Resource Information Network (USDA-GRIN) Soybean Germplasm Collections have white flowers and are the second dominant accessions after the purple-flowered accessions. Earlier studies have shown that the *w1* recessive allele of the *W1* gene encoding flavonoid 3′,5′-hydroxylase produces white flowers. In the present study, we aimed to understand why the white-flowered accessions have become abundant among the cultivated soybeans and what their genetic and regional origin is. For this purpose, 99 landraces with white flowers and 39 landraces with purple flowers from eight Asian countries and Russia were analysed with regard to the nucleotide sequences of the *W1* locus. We not only found that the *w1* alleles of the 99 white-flowered landraces were identical to those of the white-flowered Williams 82, but also found that these *w1* alleles displayed no polymorphism at all. By carrying out a phylogenetic analysis, we were able to identify a group with *W1* alleles from which the *w1* allele might have diverged.

**Keywords:** phylogenetic analysis; purple flowers; soybean landraces; *W1* gene; white flowers

## Introduction

Soybeans [*Glycine max* (L.) Merr.] have mostly either purple or white flowers, although there are some colour variations such as purple throat, near white, pink and magenta. The flower colour of soybeans is primarily controlled by six genes (*W1*, *W2*, *W3*, *W4*, *Wp* and *Wm*). These genes, with the exception of *W2* (see below), encode enzymes involved in anthocyanin and flavonol biosynthesis and thereby determine the pigmentation of flower petals (Stephens and Nickell, 1992; Palmer *et al*., 2004; Zabala and Vodkin, 2007; Takahashi *et al*., 2010; Yan *et al*., 2014).

Under the *W1* genotype, *W3W4* soybeans have dark-purple flowers, whereas soybeans with *w3W4* and *w3w4* alleles have purple and near-white flowers, respectively (Hartwig and Hinson, 1962). In the anthocyanin and flavonol biosynthetic pathway, both *w3* and *w4* encode dihydroflavonol 4-reductase (DFR)

* Corresponding author. E-mail: jtsong68@knu.ac.kr
† G. T. Park and J. Sundaramoorthy contributed equally to this research.

(Palmer and Groose, 1993; Fasoula *et al.*, 1995; Xu *et al.*, 2010; Yan *et al.*, 2014). DFR is required for the synthesis of three anthocyanin classes (delphinidins, pelargonidins and cyanidins). Soybeans with the *wp* or *wm* genotype have pink or magenta flowers, respectively (Johnson *et al.*, 1998; Takahashi *et al.*, 2007). *Wp* encodes flavonone 3-hydroxylase enzymes required in an earlier step than DFR in the biosynthetic pathway of anthocyanins (Zabala and Vodkin, 2005). *Wm* encodes flavonol synthase, which catalyses the reduction of dihydroflavonols to flavonols. The *w2* allele results in purple-blue colour, and its wild-type gene encodes a myeloblastosis transcription factor required for vacuolar acidification of flower petals (Takahashi *et al.*, 2008, 2011).

Unlike the recessive alleles of other flower pigmentation genes, the recessive allele *w1* of the *W1* locus eliminates flower colour completely, because the *W1* locus encodes flavonoid 3′,5′-hydroxylase, which produces the early precursor dihydroflavonols for anthocyanidin and flavonol synthesis (Buzzell *et al.*, 1987; Zabala and Vodkin, 2007). The *w1* allele from the *G. max* accession Williams 82 was created by a small 65 bp insertion with tandem repeats and a 12 bp deletion in the third exon of *W1*, resulting in the premature termination of translation and thus nonfunctional W1 proteins. Takahashi *et al.* (2010) proved that *W1* is also an essential gene in anthocyanin biosynthesis in a wild soybean accession (*G. soja* Sieb. & Zucc).

Of the 19,648 soybean accessions in the United States Department of Agriculture – Germplasm Resource Information Network (USDA-GRIN), 13,132 accessions (67%) have purple flowers and 6344 (32%) have white flowers; only a small fraction of them have flowers of other colours. In contrast, almost all wild soybean accessions have purple flowers. This therefore invites questions as to why and how the white-flowered accessions have become abundant among the cultivated soybeans as well as what the genetic and regional origin of the white-flowered accessions is. To address these questions, we, first, tried to define the cause of white colour production in the white-flowered accessions and found that all the 99 white-flowered landraces randomly selected from the worldwide germplasm collections have the same *w1* allele that Williams 82 has. We also analysed the nucleotide sequences of the *W1* genes of purple-flowered landraces from Korea, China, Japan and other Asian countries and determined the origin of the white-flowered accessions.

## Materials and methods

### *Plant materials*

A total of 153 accessions of *G. max* and *G. soja* were used in this study. Among them, 99 landrace accessions from Korea (64), China (17) and Japan (18) had white flowers and were randomly selected from the National Agrobiodiversity Center of Republic of Korea (IT accessions) and from the USDA-GRIN Soybean Germplasm Collections (PI accessions) (Table S1, available online). These accessions were used for the insertion and deletion (indel) detection analysis to distinguish the *w1* allele-containing soybeans. Williams 82 and Hwangkeumkong, which have white and purple flowers, respectively, were used as controls. A total of 39 purple-flowered landrace accessions from Korea (10), China (10), Japan (7), India (2), Indonesia (2), Myanmar (2), Nepal (2), Vietnam (2) and Russia (2) were selected for phylogenetic analysis (Table S2, available online). In addition, 15 purple-flowered *G. soja* accessions from Korea (5), China (5) and Japan (5) were included in the analysis (Table S3, available online).

### *Indel and phylogenetic analyses*

The genomic DNA of soybean accessions was isolated from trifoliolate leaves using the cetyltrimethylammonium bromide method (Doyle and Doyle, 1987). Polymerase chain reaction (PCR) analysis was performed to identify the 53 bp indel insertion of the *w1* allele using the following profile: 40 cycles of 94°C for 30 sec, 58°C for 30 sec, and 72°C for 1 min. The sequences of upstream and downstream primers were 5'-TGGTGCTGGGAGGAGGATTT-3′ and 5′- CTTGCTGCTTTGGTTACCCC-3′, respectively. The nucleotide sequences of the *W1* locus of the purple-flowered landraces and wild soybeans were determined. The amplified DNAs were about 4.7 kb in length and included a 25 bp upstream region of ATG, 2.9 kb of two introns, 1.5 kb of three exons and a 244 bp downstream region of the stop codon (Zabala and Vodkin, 2007). The GenBank accession numbers of the *W1* genes of soybean accessions (PI483462B, PI458538, PI464939A, PI407290, PI549036, PI366121, PI507624, PI406684, PI507646, PI378701A, PI407280, PI407271, PI407229, PI424120, PI339731, IT182932, IT115634, IT165401, IT141739, IT022876, PI416833, IT154524, PI518716, IT208306 and IT263338) ranged from KJ911862 to KJ911886. Twenty-seven sequences of the *W1* gene were aligned and a neighbour-joining phylogenetic tree was constructed using Mega5.2 with a Kimura two-parameter model and 1000 bootstrap replicates (Tamura *et al.*, 2011). A hierarchical likelihood ratio test was carried out using JModelTest to determine which substitution model best described the evolution of *W1* sequences (Posada, 2008). The Tamura–Nei model was specified for *W1* sequences using the Akaike information criterion (Tamura and Nei, 1993). PhyML 3.0 was used for maximum-likelihood analysis

with the model TPM1uf + I + G (Guindon *et al.*, 2010). In addition, nucleotide diversity, or the average number of nucleotide differences per site, was estimated using DnaSP 5.0 (Rozas *et al.*, 2003).

## Results

### *Polymorphism and indel analysis*

We wondered why the USDA-GRIN database had a higher proportion of white-flowered accessions and which gene was mutated in white-flowered soybeans. To address this question, 99 soybean landraces, not cultivars, with white flowers from Korea (64), China (17) and Japan (18) were randomly selected (Table S1, available online) and were used for PCR analysis with the primer set spanning the indel. The PCR products of all the 99 landraces were compared with those of Williams 82 with white flowers and of Hwangkeumkong with purple flowers, which served as controls, and polymorphism between them was determined. The sizes of the PCR products all the landraces examined were identical to those of Williams 82, as shown in 12 representative examples (Fig. 1), and the sizes were greater than those of Hwangkeumkong due to the excess of 53 bp nucleotides in the third exon (Zabala and Vodkin, 2007). To assess polymorphism among the *w1* alleles examined, 20 landraces were randomly selected from among the 99 landraces (Korean (10), Chinese (5) and Japanese (5)) and about 4.7 kb-long nucleotide sequences of *w1* genes were determined. The nucleotide sequence analysis revealed that the *w1* genes of all the 20 landraces with white flowers had nucleotide sequences that were the same as the *w1* allele of Williams 82 (data not shown), indicating that there is no polymorphism between the *w1* alleles at all.

### *Phylogenetic analysis*

We analysed the nucleotide sequences of the *W1* genes of 39 purple-flowered landraces and 15 wild soybean accessions from eight Asian countries and Russia (Tables S2 and S3 and Fig. S1, available online). The sequences of IT182932 (*G. soja* used in genome sequencing by Kim *et al.*, 2010) and L79-908 (*G. max* used in the characterization of the *W1* gene by Zabala and Vodkin, 2007) were obtained from the GenBank database. A phylogenetic tree was constructed to infer the regional origin of white-flowered soybeans and to evaluate the genetic relationship between *G. max* and *G. soja* along with Williams 82 (white flower) (Fig. 2). These soybean accessions were grouped into two main clusters. The soybean accessions examined in this study are shown in Fig. 2(b). The nucleotide sequence analysis revealed that the *W1* genes of *G. max* accessions belonging to the same branch are identical. Cluster I mostly consisted of wild soybeans, including *G. soja* accession IT182932 (soja16), but not the *G. soja* accession PI549036 (soja15). However, wild soybeans from China, Japan and Korea, even if they originated from the same country, did not group to form a subcluster, but rather scattered across the branches (Fig. 2(a)), suggesting that the wild accessions from these three countries might have genetically intermingled with each other, which is in agreement with the results of previous studies (Hymowitz and Kaizuma, 1981; Han *et al.*, 1999).

Almost all the purple-flowered soybean accessions (*G. max*) were grouped along with Williams 82 (white flower) and *G. max* accession L79-908 (max10) in cluster II. However, two *G. max* accessions, IT208306 (max1) and IT263338 (max2), were placed in cluster I, suggesting that they are more similar to *G. soja* than to *G. max*. On the other hand, one *G. soja* accession PI549036 (soja15) was grouped together with the purple-flowered *G. max* in cluster II, indicating that it is more similar to *G. max* than to *G. soja*.

The nucleotide sequences of the *W1* genes of *G. max* in the branch max6 [IT022876 (China); PI423935, PI227212, PI200458 and PI561377 (Japan); and IT136306 and IT200606 (Nepal)] perfectly matched those of the *w1* gene of Williams 82 when the indel sequence was excluded. Similarly, PI416833 (Japan) of the single-membered branch max7 closely matched the *w1* gene of Williams 82, but with the exception that their first introns differed only in a single nucleotide. Therefore, it is tempting to speculate that the *w1* allele of white-flowered soybeans might have diverged from the branch max6. The maximum-likelihood analysis presented a topology similar to the tree derived from the neighbour-joining analysis (data not shown).

Using the 500 bp interval sliding-window analysis, the nucleotide sequences of 16 *G. soja* accessions and 10 *G. max* accessions (one representative accession from each branch of *G. max* in the phylogenetic tree) were analysed to investigate nucleotide diversity (Fig. 3).

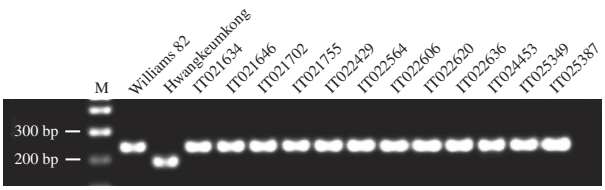
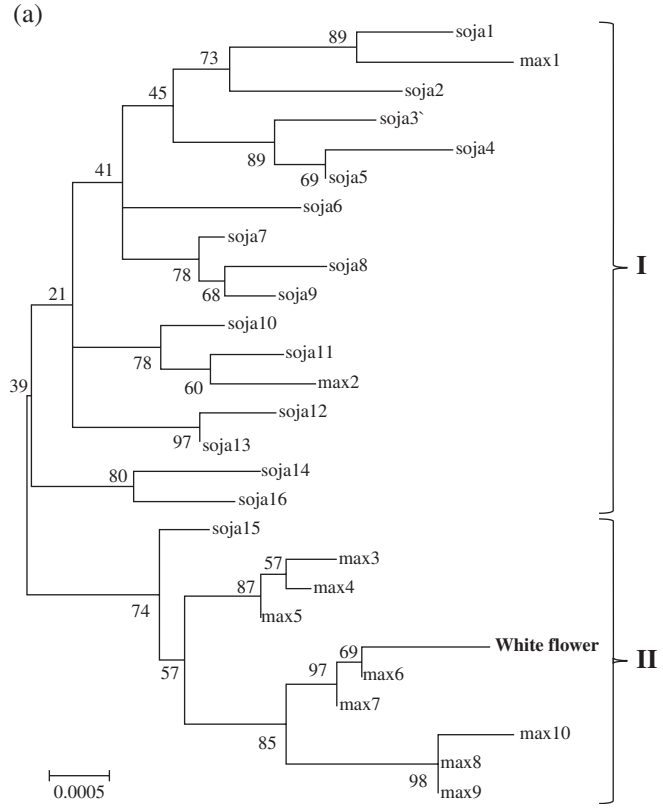



**Fig. 1.** Results of the indel analysis carried out to determine the presence of *w1* alleles in white-flowered landrace accessions. The PCR products of white-flowered Williams 82 and purple-flowered Hwangkeumkong are loaded as controls and their lengths are 236 bp and 183 bp, respectively. M, molecular marker.

https://doi.org/10.1017/S1479262114000938 Published online by Cambridge University Press

(a)




(b)

| | *G. max* accessions |
|---|---|
| max1 | IT208306(I) |
| max2 | IT263338(M) |
| max3 | IT115634(K), PI416793(J), IT161013(IN) |
| max4 | IT165401 (V), IT171172(V) |
| max5 | IT141739(K), IT024353(K), IT186135(K), IT141605(K), IT115853(K), IT141493(K), PI532457(C), I567688A(C), PI612736(C), PI416908(J), IT252795(M), IT165343(IN) |
| max6 | IT022876(C), PI423935(J), PI227212(J), PI200458(J), PI561377(J), IT136306(N), IT200606(N) |
| max7 | PI416833(J) |
| max8 | IT154524(K), IT022491(K), IT154548(K), PI495017A(C), PI201665(C), IT173147(R), IT173148(R), IT263395(I) |
| max9 | PI518716(C), PI549031(C), PI567366A(C), PI567636(C) |
| max10 | L79-908 |

| | *G. soja* accessions |
|---|---|
| soja1 | PI366121(J) |
| soja2 | PI458538(C) |
| soja3 | PI507646(J) |
| soja4 | PI507624(J) |
| soja5 | PI406684(J) |
| soja6 | PI407229(K) |
| soja7 | PI378701A(J) |
| soja8 | PI483462B(C) |
| soja9 | PI464939A(C) |
| soja10 | PI407290(C) |
| soja11 | PI424120(K) |
| soja12 | PI407280(K) |
| soja13 | PI407271(K) |
| soja14 | PI339731(K) |
| soja15 | PI549036(C) |
| soja16 | IT182932 |

**Fig. 2.** Phylogenetic relationships between *W1* genes. (a) A phylogenetic tree. About 4.7 kb-long nucleotide sequences of the *W1* genes of 39 purple-flowered landraces and 15 wild soybeans were determined. Twenty-seven sequences of the *W1* genes including those of IT182932 (soja16) and L79-908 (max10) and the *w1* gene of a white-flowered soybean (Williams 82) were aligned and a neighbour-joining phylogenetic tree was constructed. The sequences of IT182932 (*Glycine soja*) and L79-908 (*G. max*) were obtained from the GenBank database. The number on each node indicates the bootstrap value and the scale bar indicates nucleotide substitutions per site. (b) List of soybean accessions used in the construction of the phylogenetic tree. The nucleotide sequences of the *W1* genes of purple-flowered accessions belonging to the same branch are identical to each other. Abbreviations in parentheses of IT and PI accessions: C, China; I, India; IN, Indonesia; J, Japan; K, Korea; M, Myanmar; N, Nepal; R, Russia; and V, Vietnam.

The nucleotide sequences and single nucleotide polymorphisms (SNPs) of *G. soja* and *G. max* accessions are shown in Fig. S2 (available online). We found that the distribution of nucleotide diversity was uneven along the whole length of the *W1* gene and the diversity patterns of *G. max* and *G. soja* were similar (Fig. 3). The diversity levels of introns were much higher than those of exons, especially the highest diversity was observed in the approximately 0.4 kb region within the second intron spanning from 2 to 2.4 kb (Fig. 3(c)).

## Discussion

White-flowered *G. max* accessions occupy the second highest position in the USDA-GRIN database after the purple-flowered accessions. We found that the white flower colour of all the soybean landraces examined is due to the presence of the same indel mutation that Williams 82 has. The nucleotide sequence and phylogenetic analyses revealed that the *w1* alleles of the 99 white-flowered landraces were identical to those of the white-flowered Williams 82 and also that the *w1* allele of white-flowered soybeans might have diverged from the branch max6.

Furthermore, the phylogenetic tree revealed that the mixed-up pattern of wild and cultivated soybean accessions has also been reported previously (Xu and Gai, 2003; An *et al.*, 2009). Molecular analyses based on the chloroplast and nuclear simple sequence repeat (SSR) marker variation revealed the occurrence of introgression from cultivated soybean accessions into wild soybeans (Kuroda *et al.*, 2006; Abe *et al.*, 1999; Wang *et al.*, 2010). Similarly, Li *et al.* (2010) and Lam *et al.* (2010) found the occurrence of introgression from wild soybeans into cultivated soybeans, explaining the swapping in the position of some accessions in the phylogenetic clusters. With these exceptional accessions, clusters I

https://doi.org/10.1017/S1479262114000938 Published online by Cambridge University Press

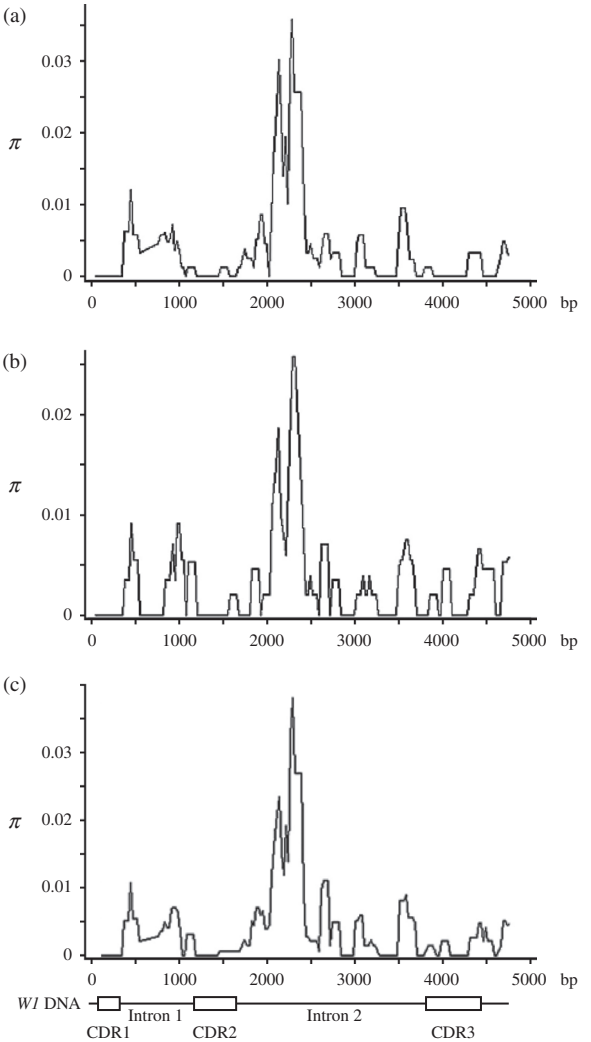



**Fig. 3.** Nucleotide diversity along the whole *W1* locus. The nucleotide diversity ($\pi$) distributions of the *W1* genes of (a) 16 *Glycine soja* accessions and (b) 10 *G. max* accessions were plotted using the 500 bp interval sliding-window analysis and (c) then combined as a pooled diagram. Numbers in bp correspond to the nucleotide position of *W1* gene. CDR, coding region.

and II are indicative of a distinct genetic difference between wild and cultivated soybeans, and similar results have been obtained previously through other analyses, such as random amplified polymorphic DNA (Xu and Gai, 2003), SSR (Powell *et al.*, 1996; Wen *et al.*, 2009), amplified fragment length polymorphism (Maughan *et al.*, 1996), SNP (Li *et al.*, 2010) and microsatellite variation (Kuroda *et al.*, 2006).

However, the high occupancy ratio of white-flowered landraces throughout the world is still unclear. At present, we consider three possibilities. First, white flowers may be simply attractive to the farming community when compared with purple flowers. However, chances of pollination are very low because of the small size of soybean flowers. Second, white flowers may be closely linked to agronomic traits, such as seed yield and stress resistance. Ortiz-Perez *et al.* (2006) reported that white-flowered accessions have more seed set than purple-flowered accessions. In addition, Severson (1983) reported that there is a considerable difference between white-flowered and purple-flowered soybeans with regard to fructose and glucose content, nectar volume and total carbohydrate content per flower. Finally, even if white flowers are not linked to any important agronomic trait, there is the possibility that white-flowered soybeans might have been superior to other cultivated soybeans at the time when seed dispersal of the white-flowered soybean took place in the past.

In spite of these considerations, however, no strong evidence has been found for the high ratio of white-flowered landraces among the cultivated soybeans. Therefore, questions regarding the precise origin and abundance of white-flowered soybeans remain to be answered in the future.

## Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262114000938

## Acknowledgements

This research was supported by a grant from the Next-Generation Biogreen 21 Program (Plant Molecular Breeding Center No. PJ008137), Rural Development Administration, Republic of Korea.

## References

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

An W, Zhao H, Dong YS, Wang Y, Li Q, Zhuang BC, Gong L and Liu B (2009) Genetic diversity in annual wild soybean (*Glycine soja* Sieb. et Zucc.) and cultivated soybean (*G. max*. Merr.) from different latitudes in China. *Pakistan Journal of Botany* 41: 2229–2242.

Buzzell RI, Buttery BR and MacTavish DC (1987) Biochemical genetics of black pigmentation of soybean seed. *Journal of Heredity* 78: 53–54.

Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Fasoula DA, Stephens PA, Nickell CD and Vodkin LO (1995) Cosegregation of purple-throat flower with dihydroflavonol-reductase polymorphism in soybean. *Crop Science* 35: 1028–1031.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307–321.

Han O, Abe J and Shimamoto Y (1999) Genetic diversity of soybean landraces in Korea. *Korean Journal of Crop Science* 44: 256–262.

Hartwig EE and Hinson K (1962) Inheritance of flower color of soybeans. *Crop Science* 2: 152–153.

Hymowitz T and Kaizuma N (1981) Soybean seed protein electrophoresis profiles from 15 Asian countries or regions: hypotheses on paths of dissemination of soybeans from China. *Economic Botany* 35: 10–23.

Johnson EOC, Stephens PA, Fasoula DA, Nickell CD and Vodkin LO (1998) Instability of a novel multicolored flower trait in inbred and outcrossed soybean lines. *Journal of Heredity* 89: 508–515.

Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee YS, Park D, Ma J, Kim WY, Kim BC, Park S, Lee KA, Kim DH, Kim KH, 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 SH (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proceedings of the National Academy of Science, USA* 107: 22032–22037.

Kuroda Y, Kaga A, Tomooka N and Vaughan DA (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 W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Wang J, Sun SS and Zhang G (2010) Resequencing 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 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 Phytologist* 188: 242–253.

Maughan PJ, Saghai-Maroof MA, Buss GR and Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, near-isogenic line analysis. *Theoretical and Applied Genetics* 93: 392–401.

Ortiz-Perez E, Horner HT, Hanlin SJ and Palmer RG (2006) Evaluation of Insect-mediated seed set among soybean lines segregating for male sterility at the *ms6* locus. *Field Crops Research* 97: 353–362.

Palmer RG and Groose RW (1993) A new allele at the *w4* locus derived from the *w4-m* mutable allele in soybean. *Journal of Heredity* 84: 297–300.

Palmer RG, Pfeiffer TW, Buss GR and Kilen TC (2004) Qualitative genetics. In: Specht JE and Boerma HR (eds) *Soybean: Improvement, Production, and Uses*. Agronomy Monograph 16. 3rd edn. Madison, Wisconsin: American Society of Agronomy Inc, pp. 137–234.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256.

Powell W, Morgante M, Doyle JJ, McNicol JW, Tingey SV and Rafalski AJ (1996) Genepool variation in genus *Glycine* subgenus *soja* revealed by polymorphic nuclear and chloroplast microsatellites. *Genetics* 144: 793–803.

Rozas J, Sanchez-DelBarrio JC, Messeguer X and Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.

Severson DW (1983) Honey bees and soybeans: analyses of floral chemistry relating to foraging preferences. PhD Thesis, University of Wisconsin-Madison

Stephens PA and Nickell CD (1992) Inheritance of pink flower in soybean. *Crop Science* 32: 1131–1132.

Takahashi R, Githiri SM, Hatayama K, Dubouzet EG, Shimada N, Aoki T, Ayabe S, Iwashina T, Toda K and Matsumura H (2007) A single-base deletion in soybean flavonol synthase gene is associated with magenta flower color. *Plant Molecular Biology* 63: 125–135.

Takahashi R, Matsumura H, Oyoo ME and Khan NA (2008) Genetic and linkage analysis of purple-blue flower in soybean. *Journal of Heredity* 99: 593–597.

Takahashi R, Dubouzet JG, Matsumura H, Yasuda K and Iwashina T (2010) A new allele of flower color gene *W1* encoding flavonoid 3′5′-hydroxylase is responsible for light purple flowers in wild soybean *Glycine soja*. *BMC Plant Biology* 10: 155.

Takahashi R, Benitez ER, Oyoo ME, Khan NA and Komatsu S (2011) Nonsense mutation of an MYB transcription factor is associated with purple-blue flower color in soybean. *Journal of Heredity* 102: 458–463.

Tamura K and Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.

Wang KJ, Li XH, Zhang JJ, Chen H, Zhang ZL and Yu GD (2010) Natural introgression from cultivated soybean (*Glycine max*) into wild soybean (*Glycine soja*) with the implications for origin of populations of semi-wild type and for biosafety of wild species in China. *Genetic Resources and Crop Evolution* 57: 747–761.

Wen Z, Ding Y and Zhao T (2009) Genetic diversity and peculiarity of annual wild soybean (*G. soja* Sieb. et 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 M, Brar HK, Grosic S, Palmer RG and Bhattacharyya MK (2010) Excision of an active CACTA-like transposable element *DFR2* causes variegated flowers in soybean (*Glycine max* (L.) Merr.). *Genetics* 184: 53–63.

Yan F, Di S, Rojas Rodas F, Rodriguez Torrico T, Murai Y, Iwashina T, Anai T and Takahashi R (2014) Allelic variation of soybean flower color gene *W4* encoding dihydroflavonol 4-reductase 2. *BMC Plant Biology* 14: 58.

Zabala G and Vodkin LO (2005) The *up* mutation of *Glycine max* carries a Gene-Fragment-Rich transposon of the CACTA superfamily. *Plant Cell* 17: 2619–2632.

Zabala G and Vodkin LO (2007) A rearrangement resulting in small tandem repeats in the F3′5′H gene of white flower genotypes is associated with the soybean *W1* locus. *Plant Genome (A supplement to Crop Science)* 47: S113–S124.