An Active CACTA-Family Transposable Element is Responsible for Flower Variegation in Wild Soybean Glycine soja

Ryoji Takahashi,* Yasumasa Morita, Masayoshi Nakayama, Akira Kanazawa, and Jun Abe

Abstract
A plant producing flowers with purple and white variegation was discovered in an accession of Glycine soja Siebold & Zucc. that was introduced from Russia. The mutant line was designated as B00146-m. Lines with white flowers (B00146-w) and purple flowers (B00146-r) were developed from the progeny of B00146-m. The flower color was controlled by the W1 locus encoding a flavonoid 3’5’-hydroxylase (F3’5’H). The allele for variegated flowers was designated as w1-m. The mutant line was approved by the Soybean Genetics Committee. Polymerase chain reaction (PCR) suggested that a DNA fragment with a molecular size of ~3.9 kb was inserted in the first exon of the F3’5’H gene in B00146-m whereas such insertion was not observed in B00146-w and B00146-r. These results suggested that an active mobile element was inserted in the first exon and was responsible for flower variegation. The inserted fragment was identified as a 3883 bp long CACTA-family transposable element and it was designated as Tgs1. Similarity of overall sequence and terminal inverted repeats suggested that Tgs1 and the soybean lectin gene transposable element Tgm1 make up a subgroup. Frequency of germinal reversion was low probably due to the integration into an exon. Tgs1 had a truncated version of the transposase gene and may be a nonautonomous element.

In Soybean [Glycine max (L.) Merr.,] six major genes (W1, W2, W3, W4, Wm, and Wp) control flower color and two genes (T and Td) control pubescence color (reviewed by Palmer et al., 2004; Takahashi et al., 2008). The W1 gene has a pleiotropic effect on flower and hypocotyl color: soybean cultivars with dominant W1 allele have purple flowers and purple hypocotyls whereas cultivars with recessive w1 allele have white flowers and green hypocotyls (Takahashi and Fukuyama, 1919). Chromatographic experiments suggested that the W1 locus is responsible for the formation of flavonoids with a 3’, 4’, 5’ B-ring hydroxylation pattern, indicating that W1 encodes a flavonoid 3’5’-hydroxylase (F3’5’H) (Buzzell et al., 1987). The white flowers contain no anthocyanins, probably because dihydroflavonol reductase of soybean exclusively uses tri-hydroxylated dihydroflavonols as a substrate (Iwashina et al., 2007). Zabala and Vodkin (2007) cloned the F3’5’H gene from soybean near-isogenic lines (NILs) for the W1 locus and found that the gene of white-flowered lines (w1 allele) contained a 65-bp insertion around the C terminus.

Soybean is believed to have been domesticated in northeastern China from its wild relative, Glycine soja (Hymowitz, 2004). Glycine soja is native throughout China and the adjacent areas belonging to Russia, Korea, Japan, and Taiwan (Hymowitz, 2004). Flower color of G. soja is almost exclusively purple; by contrast, 33% of the soybean accessions in the USDA Soybean Germplasm...

Published in The Plant Genome 5:62–70.

Abbreviations: cDNA, complementary DNA; dGTP, deoxyguanosine triphosphate; En, Enhancer; F3’5’H, flavonoid 3’5’-hydroxylase; NIL, near-isogenic line; PCR, polymerase chain reaction; RNA, ribonucleic acid; Spm, Suppressor-Mutator; STR, subterminal region; TIR, terminal inverted repeat.

R. Takahashi, National Institute of Crop Science, Kannondai, Tsukuba, Ibaraki, 305-8518 Japan; M. Nakayama and Y. Morita, National Institute of Floricultural Science, Fujimoto, Tsukuba, Ibaraki, 305-8519 Japan; A. Kanazawa and J. Abe, Research Faculty of Agriculture, Hokkaido University, Sapporo, 060-8589 Japan. Received 6 Nov. 2011. *Corresponding author (masako@affrc.go.jp).

Published July, 2012

THE PLANT GENOME • JULY 2012 • VOL. 5, NO. 2
Collections have white flowers (Dr. R.L. Nelson, personal communication, 2006). One white-flowered plant was found by USDA researchers among the progeny of a purple-flowered G. soja accession that was introduced from South Korea (Chen and Nelson, 2004). Genetic analysis indicated that the white flower was caused by a recessive allele at the W1 locus similar to the white-flowered soybeans (Chen and Nelson, 2004). The mutation probably occurred during propagation at USDA.

Takahashi et al. (2010) reported a G. soja accession B09121 with light purple flowers that was discovered in southern Japan. B09121 may be the first example of a flower color variant of G. soja found in the natural habitat. Genetic analysis revealed that a new allele of the B09121 with light purple flowers that was discovered probably occurred during propagation at USDA. B09121 may be the first example of a flower color variant of G. soja found in the natural habitat. Genetic analysis revealed that a new allele of the W1 locus, w1-lp, was responsible for light purple flowers. Flower petals of B09121 contained lower amounts of four major anthocyanins (malvidin 3,5-di-O-glucoside, petunidin 3,5-di-O-glucoside, delphinidin 3,5-di-O-glucoside, and delphinidin 3-O-glucoside) common in purple flowers and small amounts of the 5′-unsubstituted versions of the above anthocyanins, peonidin 3,5-di-O-glucoside, cyanidin 3,5-di-O-glucoside, and cyanidin 3-O-glucoside, suggesting that F3′5′SH activity was reduced and flavonoid 3′-hydroxylase activity was increased in this variant.

Takahashi discovered a single plant with variegated flowers in 2004 among plants of a purple-flowered G. soja accession B00146 that was introduced from Russia (R. Takahashi, unpublished data, 2004) (Fig. 1). Generally, transposable elements are integrated in genes responsible for flavonoid biosynthesis in plants having variegated flowers. In soybean, 11 transposable elements, Tgm1, Tgm2, Tgm3, Tgm4, Tgm5, Tgm6, Tgm7, Tgm-Express1, Tgm*, Tgm9, and Tgm10, all belonging to the CACTA-family class II transposable elements, have been identified (Vodkin et al., 1983; Rhodes and Vodkin, 1988; Zabala and Vodkin, 2005, 2008; Xu et al., 2010). Characteristic features of the CACTA superfamily include terminal inverted repeats (TIRs) and a 3-bp target site duplication on integration (Gierl et al., 1989). Integration of Tgm-Express1, Tgm*, or Tgm9 into genes responsible for flavonoid biosynthesis is associated with flower or pubescence color variation. A 5.7-kb element, Tgm-Express1, was inserted into the second intron of Wp gene encoding flavanone 3′-hydroxylase (Zabala and Vodkin, 2005). A 20.5-kb element, Tgm*, was integrated into the first intron of T gene encoding a flavonoid 3′-hydroxylase (Zabala and Vodkin, 2008). A 20.5-kb element, Tgm9, was inserted into the second intron of W4 gene encoding dihydroflavonol 4-reductase (Xu et al., 2010). Among these soybean elements, only Tgm9 has been confirmed to be active (Xu et al., 2010). Tgm* and Tgm9 have quite high sequence similarity, suggesting that the active element Tgm9 may be a progenitor of Tgm* (Xu et al., 2010). A 3.5-kb element, Tgm1, was inserted in the exon of a lectin gene (Le1) of a soybean cultivar having a recessive allele of the locus (Vodkin et al., 1983). In contrast to soybean, no transposable elements have been reported in G. soja. This study was conducted to determine the locus and to isolate the genetic element responsible for flower variegation of the G. soja line B00146.

Materials and Methods

Plant Materials

A single mutant plant with variegated flowers was discovered in a purple-flowered accession of wild soybean G. soja introduced from Russia (B00146). The plant was selfed and the line was designated as B00146-m. Plants with white flowers and purple flowers were found in the progeny of B00146-m. They were selfed and the resultant lines were designated as B00146-w (white flowers) and B00146-r (purple flowers).

Genetic Analysis

B00146-w was crossed with a NIL of a U.S. soybean cultivar Clark for W1 locus, Clark-w1 (L63-2373, w1w1 W2W2 w3w3 W4W4 WmWm WpWp TT) with white flowers and tawny pubescence in 2006. Flowers of Clark-w1 were emasculated 1 d before opening and pollinated with B00146-w. Hybidity of the F1 plants was indicated by a spindly growth habit. Seeds of L63-2373 were obtained from the USDA Soybean Germplasm Collection. This line was produced by crossing Clark with T139 and back-crossing the progeny to Clark up to BC6 (Bernard et al., 1991). Four F1 seeds were planted in pots filled with soil (low-humic andosols) fertilized with N, P, and K at a rate of 3.0, 4.4, and 8.3 g m^-2, respectively, on 10 Feb. 2007. The pots were placed in a glasshouse set at 25/19°C (day/night). Light was supplemented to extend the daylength to 14 h. A total of 105 F2 seeds were planted in field on 13 June 2007 at the National Institute of Crop Science, Tsukuba, Japan (36°06′ N, 140°05′ E). Nitrogen, P, and K were applied at rates similar to those used in the pots. Flower color of F1 plants and F2 population was scored individually.

Frequency Estimation of Variegated Plants and Germinal Revertants

A total of 256 seeds derived from a single plant of B00146-m with variegated flowers was planted in vermiculite on 26 July and transplanted to the same field on 6 Aug. 2010. Hypocotyl and flower color was individually recorded to estimate the frequency of variegated plants and germinal revertants.

Polymerase Chain Reaction Amplification and Molecular Cloning

Genomic DNA of Clark, B00146, B00146-m, and B00146-r was isolated from trifoliolate leaves by cetyltrimethylammonium bromide (Murray and Thompson, 1980). The first exon was amplified by polymerase chain reaction (PCR) from Clark, B00146, B00146-m, B00146-w, and B00146-r using a pair of PCR primers (first exon primers, Supplemental Table S1) using the genomic DNA as a template. The PCR mixture contained
50 ng of genomic DNA, 10 pmol of each primer, 10 pmol of nucleotides, and 1 unit of Advantage 2 Polymerase Mix in 1x Advantage 2 PCR Buffer supplied by the manufacturer (Clontech) in a total volume of 25 μL. A 1 min denaturation at 95°C was followed by 30 cycles of 30 s denaturation at 95°C and 4 min annealing and extension at 68°C. A final 3 min extension at 68°C completed the program. Polymerase chain reaction products were loaded on a 0.8% agarose gel, stained by ethidium bromide, and visualized under ultraviolet light. The PCR was performed in an Applied Biosystems 9700 thermal cycler. The PCR fragment of about 4.2 kb was extracted from the gel and was cloned into pCR 2.1 vector (Invitrogen) and sequenced. Genomic sequence of F3’5’H gene of B00146 including introns were determined by cloning two PCR fragments (upstream fragment of about 2.3 kb and downstream fragment of about 2.3 kb) overlapping each other. The primer sequences are shown in Supplemental Table S1. The PCR mixture contained 0.5 μg of genomic DNA, 10 pmol of each primer, 5 pmol of deoxyribonucleotide triphosphates, and 1 unit of ExTaq in 1x ExTaq Buffer supplied by the manufacturer (Takara Bio) in a total volume of 25 μL. A 30 s denaturation at 94°C was followed by 30 cycles of 30 s denaturation at 94°C, 1 min annealing at 59°C, and 2 min extension at 72°C. A final 7 min extension at 72°C completed the program.

**RNA Extraction and Complementary DNA Cloning**

Total ribonucleic acid (RNA) was extracted from 200 mg of trifoliolate leaves from B00146, B00146-w, and B00146-r using the TRIZOL Reagent (Invitrogen) according to the manufacturer’s instructions. Full-length complementary DNA (cDNA) of F3’5’H gene was cloned by reverse transcription-polymerase chain reaction using the protocol and PCR primers described in Takahashi et al. (2010).

**Sequencing Analysis**

Nucleotide sequences were determined using an ABI3100 Genetic Analyzer (Applied Biosystems). Primers used for sequencing are shown in Supplemental Table S2. Most parts of the transposable element and the surrounding F3’5’H gene fragments were sequenced with BigDye terminator version 3.1 (Applied Biosystems) using a protocol recommended by the manufacturer. The 5’-subterminal region (STR) was sequenced with dGTP (deoxyguanosine triphosphate) BigDye terminator version 1.1 (Applied Biosystems). Nucleotide sequences and amino acid translations were analyzed with the BLAST program (Altschul et al., 1997). Amino acid sequence of F3’5’H gene was deduced from the cDNA sequence and ascertained by the genomic sequence. Survey of the genome sequence of a soybean cultivar Williams 82 was performed with the soybean genome database (Phytozome, http://www.phytozome.net/soybean.php). Sequence alignment was performed with CLUSTAL W (Thompson et al., 1994) using default settings. Gene prediction was performed with the GENSCAN software (Burge and Karlin, 1997). Harr plot analysis was performed using the GENETYX version 8.0 (Genetyx Corporation, 2006).

**Estimation of Transposon Footprint**

The shorter fragment produced by PCR with the first exon primers and genomic DNA of B00146-m was deduced to be generated after excision of the transposable element (Fig. 2). The fragment of about 350 bp was cloned into PCR 2.1 vector by TA cloning and eight clones were sequenced.

**Accession Numbers**

Sequence data from this article have been deposited with the DDBJ Data Libraries under accession numbers AB643637 (Tgs1), AB643638 (cDNA sequence of F3’5’H gene of B00146), and AB703367 (genome sequence of F3’5’H gene of B00146).

**Results**

**Flower Color Variation**

A single plant with variegated flowers was discovered in 2004 in a G. soja accession B00146 with purple flowers that was introduced from Russia. The variegated plant was selfed and the line with variegated flowers was designated as B00146-m. A line producing white flowers and another with purple flowers were developed from the progeny of B00146-m and were designated as B00146-w and B00146-r, respectively (Fig. 1).
was caused by a mutation at the gene. Polymerase chain reaction products using primers 100 plants had white flowers, suggesting that white flower mutants were 25.4 and 0.8%, respectively. The frequencies of variegated plants and germinal revertants were 25.4 and 0.8%, respectively.

To establish the identity of the mutable and the loci, complementation test was performed by crossing between B00146-m and a NIL of the soybean cultivar Clark for the W1 locus, Clark-w1 with white flowers. All four F1 plants had white flowers. One hundred plants grew normally out of 105 F2 seeds planted in field. All the 100 plants had white flowers, suggesting that white flower was caused by a mutation at the W1 locus.

Frequency of Variegated Plants and Germinal Revertants

A total of 256 seeds derived from a single plant of B00146-m with variegated flowers were planted in the field. Of these, 240 plants grew normally. Two plants had purple hypocotyls and purple flowers and 61 plants had purple sectors on hypocotyls and/or variegated flowers. In addition, 177 plants had green hypocotyl and white flowers, and no variegation was observed in hypocotyls and flowers throughout the growing period. Therefore, the frequencies of variegated plants and germinal revertants were 25.4 and 0.8%, respectively.

Genetic Analysis

To establish the identity of the mutable and the W1 loci, complementation test was performed by crossing between B00146-w and a NIL of the soybean cultivar Clark for the W1 locus, Clark-w1 with white flowers. All four F1 plants had white flowers. One hundred plants grew normally out of 105 F2 seeds planted in field. All the 100 plants had white flowers, suggesting that white flower was caused by a mutation at the W1 locus.

Molecular Cloning

Polymerase chain reaction was performed with genomic DNA of a purple-flowered soybean cultivar Clark, B00146, B00146-m, B00146-w, and B00146-r as templates using primers to amplify various parts of the F3’5’H gene. Polymerase chain reaction products using primers

flanking the first exon revealed a fragment length polymorphism (Fig. 2). Polymerase chain reaction with genomic DNA of Clark produced a fragment of 348 bp as expected whereas that of B00146-m produced a longer fragment with about 4.2 kb in addition to the 348 bp band. B00146-w and B00146-r produced fragments identical in length to Clark.

The PCR product of about 4.2 kb was cloned into a plasmid vector by TA cloning and four clones were sequenced. Most regions of the transposable element could be sequenced from both strands with ordinary BigDye terminator method. However, the highly structured region close to the 5’ end of the element was difficult to sequence. Especially, sequencing with primers based on sequences inside the element was difficult, so only one strand was sequenced using the dGTP BigDye terminator method.

Frequency of Variegated Plants

A total of 256 seeds derived from a single plant of B00146-m with variegated flowers were planted in the field. Of these, 240 plants grew normally. Two plants had purple hypocotyls and purple flowers and 61 plants had purple sectors on hypocotyls and/or variegated flowers. In addition, 177 plants had green hypocotyl and white flowers, and no variegation was observed in hypocotyls and flowers throughout the growing period. Therefore, the frequencies of variegated plants and germinal revertants were 25.4 and 0.8%, respectively.

Genetic Analysis

To establish the identity of the mutable and the W1 loci, complementation test was performed by crossing between B00146-w and a NIL of the soybean cultivar Clark for the W1 locus, Clark-w1 with white flowers. All four F1 plants had white flowers. One hundred plants grew normally out of 105 F2 seeds planted in field. All the 100 plants had white flowers, suggesting that white flower was caused by a mutation at the W1 locus.

Molecular Cloning

Polymerase chain reaction was performed with genomic DNA of a purple-flowered soybean cultivar Clark, B00146, B00146-m, B00146-w, and B00146-r as templates using primers to amplify various parts of the F3’5’H gene. Polymerase chain reaction products using primers flanking the first exon revealed a fragment length polymorphism (Fig. 2). Polymerase chain reaction with genomic DNA of Clark produced a fragment of 348 bp as expected whereas that of B00146-m produced a longer fragment with about 4.2 kb in addition to the 348 bp band. B00146-w and B00146-r produced fragments identical in length to Clark.

The PCR product of about 4.2 kb was cloned into a plasmid vector by TA cloning and four clones were sequenced. Most regions of the transposable element could be sequenced from both strands with ordinary BigDye terminator method. However, the highly structured region close to the 5’ end of the element was difficult to sequence. Especially, sequencing with primers based on sequences inside the element was difficult, so only one strand was sequenced using the dGTP BigDye terminator method.

Gene Structure

Sequence analysis revealed that Tgs1 consisted of 3883 bp with CACTA motif at both ends and imperfect 30-bp TIRs (Table 1) and STRs containing direct and inverse repetition of short sequence motifs (Table 2), indicating that it belongs to a CACTA-family transposable element (Gierl et al., 1989). In addition, B00146-m had a duplication of three nucleotides (CCA) that is characteristic at the site of integration of a CACTA family transposable element.

GENSCAN analysis (Burge and Karlin, 1997) suggested the existence of an open reading frame consisting of six exons encoding 239 amino acids in the opposite orientation with the F3’5’H gene (Fig. 3). Multiple sequence alignment suggested that the transcript had an amino acid similarity of 34% with GmTNP1, a transposase gene of Tgm9 (Xu et al., 2010). The transcript of Tgs1 lacks the upstream region and had several deletions compared with GmTNP1 (Fig. 4).

BLAST analysis (Altschul et al., 1997) indicated that Tgs1 has a moderate sequence similarity with the soybean lectin gene transposable element Tgm1 (Vodkin et al., 1983). Similarity was slightly higher when the alignment was done in the opposite orientation (63%) compared with alignment in the same orientation (59%). Among CACTA family elements of soybean, TIRs of Tgs1 have the highest similarity with Tgm1 (Table 1). The 5’ TIR of Tgs1 was identical with 3’ TIR of Tgm1. Further, the 3’ TIR of Tgs1 has a higher similarity (86%) with the 5’ TIR of Tgm1 compared with 3’-3’ combination (80%). In addition, a region of Tgs1 from nucleotide number 653 to 1017, which is between the 5’ STR and the truncated transposase, was quite similar with Tgm1 when aligned in the inverted orientation (Fig. 5). These results strongly suggest that the orientation of Tgm1 might actually be inverted. On the other hand, the other regions of the Tgm1 and Tgs1 were not similar.

Similar to other CACTA family elements, Tgs1 had subterminal repeats close to both ends. In 5’ and 3’ STRs, there were repetitions capable of forming stem-loop structures. Motifs in STRs were characterized by conserved
Tgs1 fragments similar to Williams 82 suggested the existence of 18 DNA elements in soybean (Table 3). Length of 373 bp and a 99.7% similarity with the other four clones. These results indicate that, similar to B00146 and B00146-r whereas short nucleotide deletions, a 2-bp insertion, and two 4-bp insertions in the first intron. In addition, it had two 1-bp deletions, a 2-bp insertion, and two 4-bp insertions in the second intron.

Alteration of F3’S’H Gene Sequence by Excision

The F3’S’H gene of B00146 had three exons and two introns similar to soybean (Fig. 3). The exons were quite similar to soybean and the deduced polypeptide consisted of 509 amino acids. In contrast, introns of B00146 had seven indels. Compared with the genome sequence of Williams 82, L79-908 (GenBank accession no. EF174665), and Williams 43 (GenBank accession no. EF174666), B00146 had a 181 bp deletion and a 2-bp insertion in the first intron. In addition, it had two 1-bp deletions, a 2-bp insertion, and two 4-bp insertions in the second intron.

The F3’S’H gene of B00146-w had a CA insertion corresponding to the transposon footprint at the site of transposon integration (Fig. 3). The insertion changed the subsequent reading frame and generated a truncated polypeptide consisting of only 59 amino acids. On the other hand, the cDNA sequence of F3’S’H of B00146-r was identical with B00146.

Analysis of Footprint

The shorter PCR fragment of about 350 bp produced from B00146-m with the first exon primers exhibited in Fig. 2 was deduced to be generated by excision of the transposable element. The fragment was cloned into a plasmid vector by TA cloning, and eight clones (clone numbers 2, 4, 6, 8, 9, 10, 13, and 15) were sequenced (Fig. 3). All of these clones were fragments of the first exon of the F3’S’H gene. CA was added to the two clones (clones 2 and 10) similar to B00146-w, two clones (clones 8 and 15) were unaffected similar to B00146 and B00146-r whereas short nucleotide fragments consisting of 1 to 10 bases were deleted in the other four clones. These results indicate that, similar to

<table>
<thead>
<tr>
<th>Transposon</th>
<th>5’ terminal inverted repeat (5′–3’)</th>
<th>Identity to Tgs1</th>
<th>3’ terminal inverted repeat (5′–3’)</th>
<th>Identity to Tgs1</th>
</tr>
</thead>
<tbody>
<tr>
<td>†Tgm1</td>
<td>CACTTAACAAAAAAAGTTTTTAACATCGG 63%</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
<td>Tgm-express 1</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
</tr>
<tr>
<td>Tgm6</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 70%</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
<td>Tgm6</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
</tr>
<tr>
<td>Tgm9</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
<td>Tgm9</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
</tr>
<tr>
<td>Tgm-express</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
<td>Tgm-express</td>
<td>CACTACTACAAAAAAAGTTTTTAACATCGA 76%</td>
</tr>
</tbody>
</table>

†Nucleotide sequence of stems of the 5′ terminal inverted repeat was interchanged.

Table 2. Subterminal repeats of Tgs1 in wild soybean Glycine soja line B00146-m.

<table>
<thead>
<tr>
<th>Location</th>
<th>Subterminal repeats (5′–3′)</th>
<th>Nucleotide position (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’ subterminal region</td>
<td>TTAACATCGGTTTT</td>
<td>35–66</td>
</tr>
<tr>
<td>3’ subterminal region</td>
<td>AACATCGGTTTT</td>
<td>2961–2988</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consensus sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTAACATCGGTTTT</td>
</tr>
<tr>
<td>AACATCGGTTTT</td>
</tr>
</tbody>
</table>

1Nucleotide sequence of stems of the 5′ side. Nucleotides mismatched with the consensus sequence of subterminal regions (STRs) are shown in bold font. Dash represents a gap.

The Tgm9 (Xu et al., 2010), the size and composition of the residual footprint are variable.

Discussion

One plant with variegated flowers was discovered in a G. soja accession introduced from Russia. The plant was selfed and a line with variegated flowers was developed and designated as B00146-m. Lines with white flowers
and purple flowers (B00146-r) were developed from the progeny of B00146-m. Complementation test conducted by crossing B00146-w and a soybean NIL with white flowers (w1 allele) suggested that a transposable element was inserted in the \(W1\) gene encoding F3\(^{'5}'\)\(5\)H of B00146-m. The new allele for variegated flowers was designated as \(w1-m\). The gene symbol was approved by the Soybean Genetics Committee.

Polymerase chain reaction using primers flanking the first exon suggested that an element of about 3.9 kb was inserted in the first exon of B00146-m and that the element was absent in B00146-w and B00246-r. Therefore, the flower color phenotype cosegregated with the polymorphic amplicon. This suggested that the insertion of the 3.9 kb element was responsible for flower variegation and that excision of the element was responsible for flower color restoration and alteration to white. These results suggested that the element was active.

Polymerase chain reaction cloning and sequencing analysis revealed that Tgs1 was 3883 bp long, and it had characteristics of the CACTA-family transposable elements: CACTA motif at both ends, imperfect 30-bp TIRs, STRs containing direct and inverse repetition of short sequence motifs, and a duplication of three nucleotides at the site of integration (Gierl et al., 1989). The CACTA-family transposable element includes a maize (Zea mays L.) Enhancer (En) or
Suppressor-Mutator (Spm), and soybean elements Tgm1 to Tgm7, Tgm9 to Tgm10, Tgm-t*, and Tgm-EXPR ESS1.

Tgs1 may encode a defective transposable protein consisting of only 239 amino acids. It had an amino acid similarity of 34% with GmTNP1 encoding a transposase of Tgm9 and it lacks an upstream region and had several deletions compared with GmTNP1. These results suggested the existence of a two-component system in G. soja in which Tgs1 is a nonautonomous element, a deletion derivative of an intact autonomous element that encodes all functions necessary for transposition. Tgs1 may be the first example of an active nonautonomous transposable element in legumes. Tgs1 may be capable of transposing in the presence of active intact element(s) elsewhere in the genome similar to the En or Spm system of maize (Pereira et al., 1986). The En or Spm system consists of two components: the autonomous En or Spm element that possesses complete ability to transpose and defective nonautonomous elements that transpose only when active En or Spm elements are present elsewhere in the genome.

Tgs1 has a nucleotide length comparable with Tgm1. Tgs1 has moderate sequence similarity (63%) with Tgm1.

The 5' and 3' TIRs as well as STR motifs are similar in these two elements. Zabala and Vodkin (2008) classified CACTA elements of soybean into two classes based on STR motifs in which Tgm9 is in a class of its own and the other members including Tgm9, Tgm-t*, and Tgm-EXPR ESS1 constitute another class. The present results strongly suggest that Tgs1 and Tgm1 constitute one distinct group. Tgs1 and Tgm1 are both inserted in the exon in contrast to Tgm-EXPR ESS1, Tgm-t*, and Tgm9 that are inserted into introns. It is uncertain whether preference of integration site exists in the CACTA family element of different groups. Although many features are similar between the two elements, there are distinct differences between the two elements. Tgs1 has a truncated version of transposase in contrast to Tgm1 that is devoid of transposase gene. Further, there is no evidence suggesting that Tgm1 is active in contrast to Tgs1.

The F3'S'H gene of B00146-w contained a CA insertion at the site of transposon integration corresponding to a transposon footprint. The footprint resulted in a truncated polypeptide consisting of only 59 amino acids, suggesting a complete loss of function. This is in contrast...
to the white-flowered soybean cultivars in which a 65-bp fragment was inserted around the C terminus of the \( F3'5'H \) gene (Zabala and Vodkin, 2007). There are two possibilities in transposon-mediated dysfunction of the \( F3'5'H \) gene. One is the genetic change described above and the other is the epigenetic transposon silencing (Lisch, 2009). To deactivate possibly harmful mobile elements, species have developed mechanisms to recognize and silence repetitive DNA by DNA methylation. B00146 is an example of the former mechanism whereas some other white-flowered descendants have possibly been generated by transposon silencing.

Frequency of variegated plants and germinal revertants was 25.4 and 0.8%, respectively. Frequency of variegated plants was comparable to that of \( Tgm9 \) (25%) whereas the frequency of germinal revertants was substantially lower than that of \( Tgm9 \) (4%) (Xu et al., 2010). Analysis of the footprint revealed that \( Tgs1 \) leaves a footprint in most cases similar to \( Tgm9 \) (Xu et al., 2010). Because \( Tgs1 \) was inserted in the first exon, in the upstream region of the \( F3'5'H \) gene, its excision altered the open reading frame resulting in a truncated polypeptide lacking the \( F3'5'H \) function in most cases. The affected genes may continue to function only and the transposon does not leave any footprint, or the length of the footprint

---

**Table 3. Alignment of subterminal repeat motifs among CACTA family transposable elements in *Glycine soja* and soybean.**

<table>
<thead>
<tr>
<th>Element</th>
<th>Subterminal repeat motif (5’–3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Tgs1 ) (5’)</td>
<td>TTAACATCGGTTTT</td>
</tr>
<tr>
<td>( Tgs1 ) (3’)</td>
<td>AACATCAGTTTTTT</td>
</tr>
<tr>
<td>( Tgm1 )</td>
<td>TTAACATCGGTT</td>
</tr>
<tr>
<td>( Tgm9 )</td>
<td>TCTAAGACGGTT</td>
</tr>
<tr>
<td>( Tgmt )</td>
<td>TCTAAGACGGTT</td>
</tr>
<tr>
<td>( Tgm-EXPRESS 1 )</td>
<td>TCTAAGACGGTT</td>
</tr>
</tbody>
</table>

---

**Figure 5. Harp plot analysis of a soybean lectin transposable element \( Tgm1 \) vs. the *Glycine soja* transposable element \( Tgs1 \).** Regions corresponding to terminal inverted repeats and subterminal regions are indicated by red arrows. Exons of \( Tgs1 \) are shown by black boxes. Start and end of the \( Tgs1 \) region having high similarity with \( Tgm1 \) are shown by black arrows. Numbers indicate nucleotide positions.
is in multiples of three and the transposed sequence does not lead to the generation of a stop codon. This is in contrast to the Tgm9 that was inserted in the second intron (Xu et al., 2010). In the latter case, frame shift mutation does not occur and many of the affected genes may still produce functional dihydroflavonol 4-reductases. This may be the cause of low frequency of germlinal reversion of Tgs1. By taking the frequent occurrence of frame-shift mutations into account, the transposition activity of Tgs1 may be comparable with the highly active element Tgm9.

A database survey revealed the existence of DNA fragments quite similar to Tgs1 in the genome of a soybean cultivar Williams 82. A progenitor of Tgs1 may have been active in soybean ancestors that may have been inactivated over time probably due to genetic surveillance mechanisms that detect and deactivate potentially harmful mobile elements. Tgs1 may have survived in the G. soja line along with an intact element. The intact element should be cloned to elucidate the transposition mechanism of Tgs1.

Considering its high transposition activity, Tgs1 could be a useful tagging tool in G. soja as well as in soybean. B00146 originated from a high latitude region and is too early maturing in central and southern Japan, producing a white-flowered wild soybean plant. Crop Sci. 44:339–342.


References


