GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.109.107904/DC1

Excision of an Active CACTA-Like Transposable Element From *DFR2* Causes Variegated Flowers in Soybean [*Glycine max* (L.) Merr.]

Min Xu, Hargeet K. Brar, Sehiza Grosic, Reid G. Palmer and Madan K. Bhattacharyya

DOI: 10.1534/genetics.109.107904

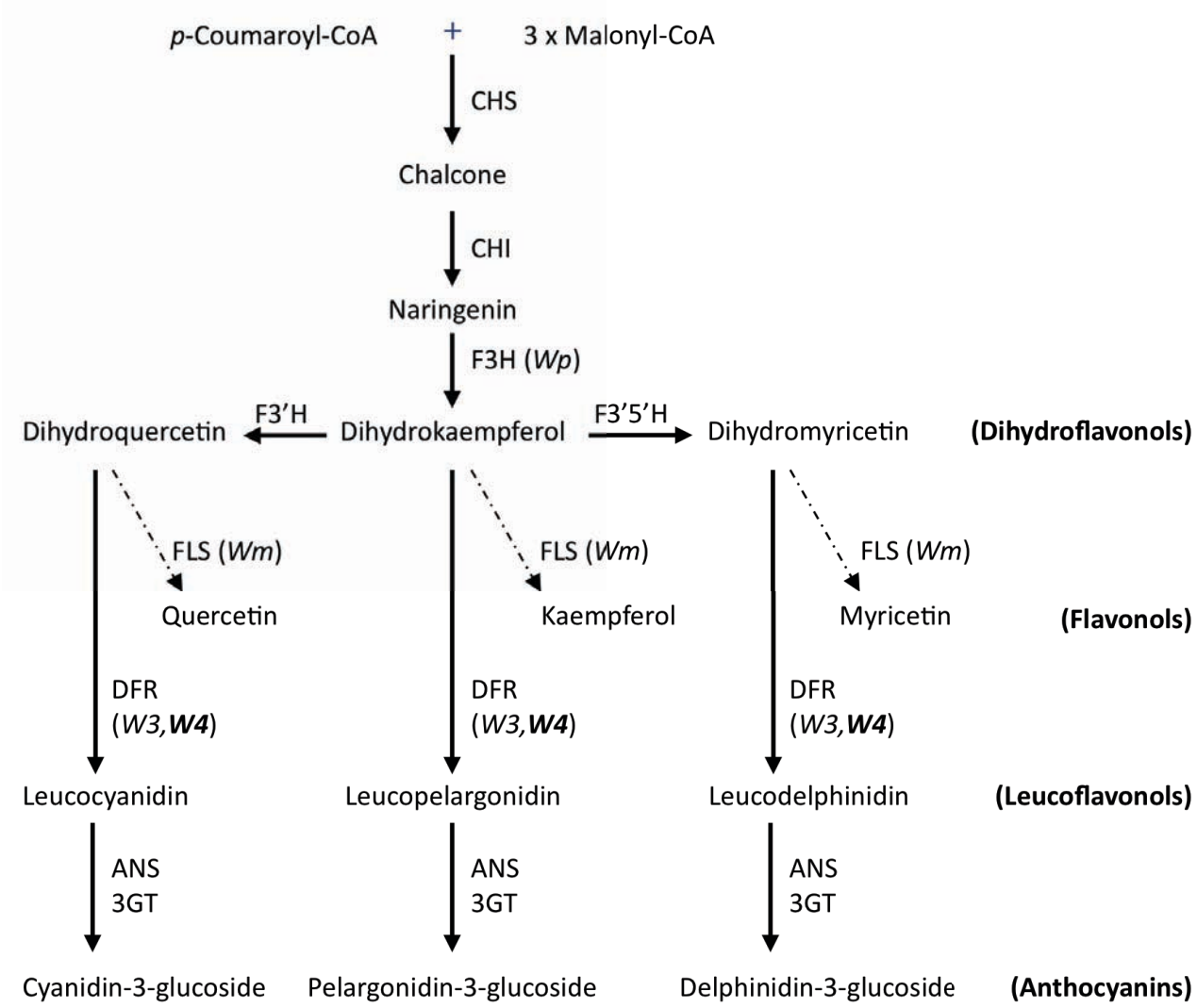

FIGURE S1.—Anthocyanin biosynthesis pathway involved in flower color development.CHS, Chalcone Synthase; CHI, Chalcone Isomerase; F3H, Flavanone 3-Hydroxylase; F3'5'H, Flavonoid 3'5'-Hydroxylase; F3'H, Flavonoid 3'-Hydroxylase; DFR, Dihydroflavonol-4-Reductase; ANS, Anthocyanidin Synthase; 3GT, 3- Glucose Transferase; FLS, Flavonol Synthase

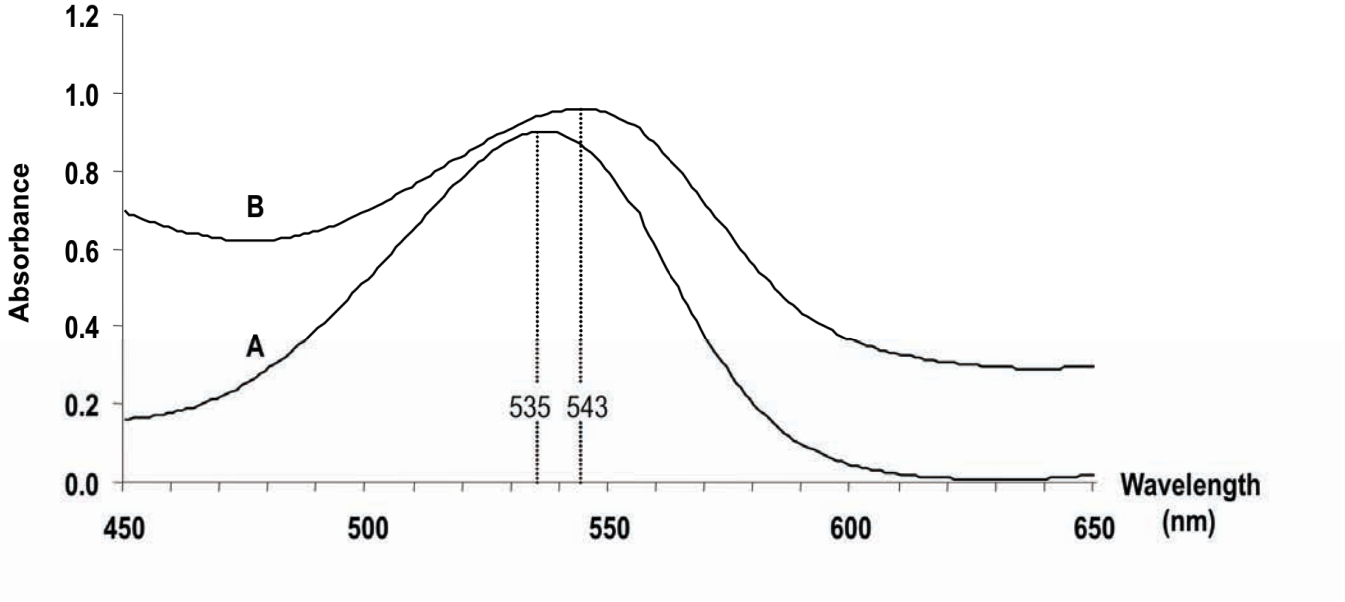

FIGURE S2.—Absorption spectra of anthocyanins extracted from immature flowers of the cultivar Harosoy. (A) Absorption spectrum of anthocyanin extracts in methanol-HCl, the absorption peak is at 535 nm. (B) Absorption spectrum of anthocyanin aglycones hydrolyzed by boiling 30 minutes, the absorption peak shifted 8 nm to 543 nm.

```
DFR2       MGSSSASESVCVTGASGFIGSWLVMRLIERGYTVRATVRDPANMKKVKHLVELPGAKTKL 60
DFR1       MG--SASESVCVTGASGFIGSWLVMRLIERGYTVRATVRDPVNMKKVKHLVELPGAKSKL 58
           **  ************************************.***************:**

DFR2       SLWKADLAQEGSFDEAIKGCTGVFHVATPMDFDSKDPENEVIKPTINGLLDIMKACVKAK 120
DFR1       SLWKADLAEEGSFDEAIKGCTGVFHVATPMDFESKDPENEVIKPTINGVLDIMKACLKAK 118
           ********:***********************:***************:*******:***

DFR2       TVRRLVFTSSAGTVDVTEHPNPVIDENCWSDVDFCTRVKMTGWMYFVSKTLAEQEAWKYA 180
DFR1       TVRRLIFTSSAGTLNVIERQKPVFDDTCWSDVEFCRRVKMTGWMYFVSKTLAEKEAWKFA 178
           *****:*******::***: :**:*:.*****:** *****************:****:*

DFR2       KEHNIDFISVIPPLVVGPFLMPTMPPSLITALSLITGNESHYHIIKQGQFVHLDDLCLGH 240
DFR1       KEQGLDFITIIPPLVVGPFLMPTMPPSLITALSPITGNEDHYSIIKQGQFVHLDDLCLAH 238
           **:.:***::********************** *****.** ***************.*

DFR2       IFVFENPKAEGRYICCSHEATIHDIAKLLNQKYPEYNVLTKFKNIPDELDIIKFSSKKIT 300
DFR1       IFLFEEPEVEGRYICSACDATIHDIAKLINQKYPEYKVPTKFKNIPDQLELVRFSSKKIT 298
           **:**:*:.******.: :*********:********:* ********:*::::*******

DFR2       DLGFKFKYSLEDMFTGAVETCREKGLLPKPEETTVNNELLPKPAETTVNDTMQK 354
DFR1       DLGFKFKYSLEDMYTGAIDTCRDKGLLPKPAEK----GLFTKPGETPVN-AMHK 347
           ************:***::***:******* *.     *:.**.**.** :*:*
```

FIGURE S3.—Alignment of DFR1 with DFR2. "*" represents identical residues; ":" means conserved substitutions between similar residues; "." indicates the semi-conserved substitutions between similar residues.

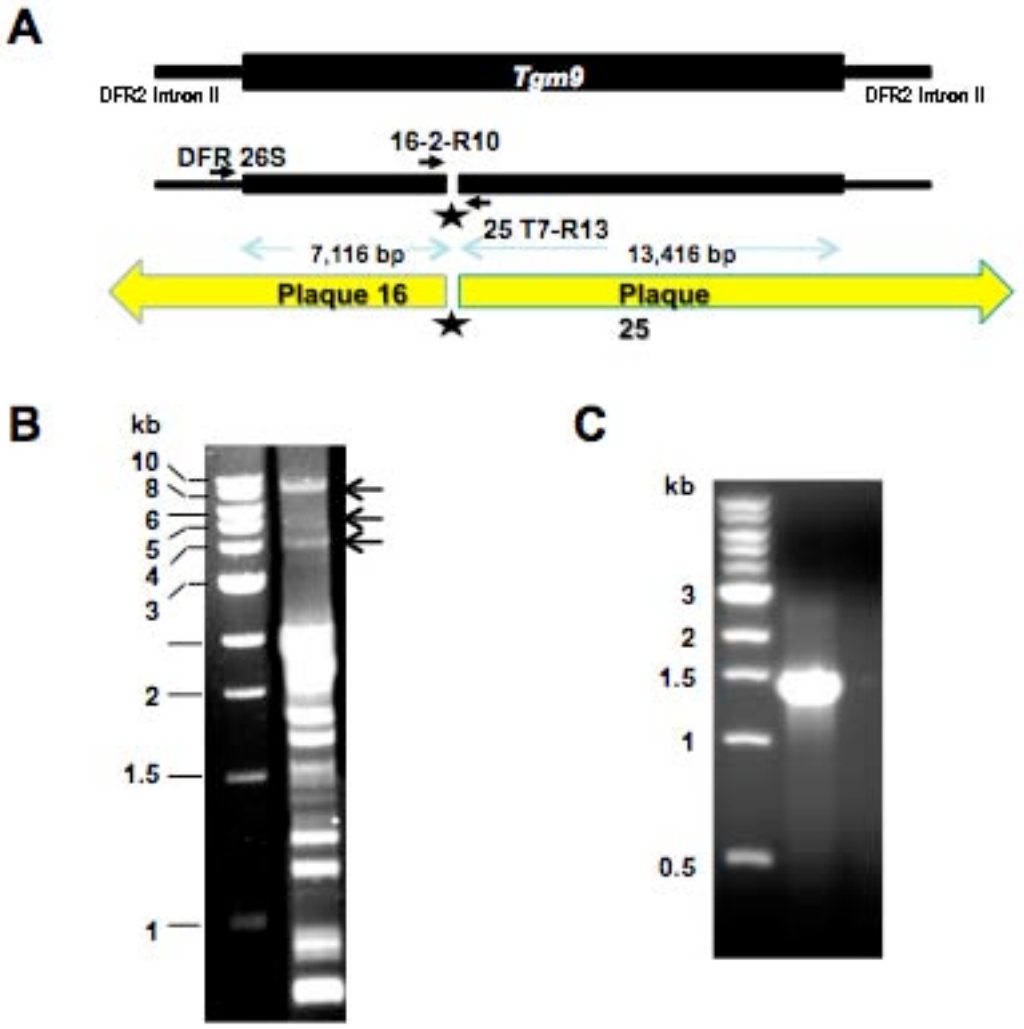

FIGURE S4.—PCR and sequencing strategies in obtaining a missing sequence of *Tgm9*. (A) Diagrammatic presentation of plaques 16 and 25 that were used to obtain most part of the *Tgm9* sequence. (B) Long-range (LR) PCR using primers DFR26S (5'- CAAGGACCCTGAGGTATGT-TGATCAT-3') and 25 T7-R13 (5'- CCCACACAAACGTATTTCTCGAAC-3') was applied to amplify the fragments including *DFR2* intron II and part of *Tgm9* as shown in (A). (C) Sub-PCR of the LR PCR products (shown by arrows in B). Sub-PCR product obtained by using primers 16-2-R10 (5'-GGGTTCTGGCGCTTCCAGTGAAG-3') and 25T7-R13 was used to obtain the sequence of the gap region shown by a black star in (A).

FIGURE S5.—Unique footprints left behind by *Tgm9* during germinal reversion. Germinal revertants from nine families identified in Figure 7a and two intermediate germinal revertants T321 (*w4-dp*) and T369 (*w4-p*) were selected for determining foot prints left behind by *Tgm9* in *DFR2* intron II through PCR by compare to the wild-type *DFR2* (WT) from cv. Williams 82 . Nucleotides representing the target site duplication are underlined. Footprint nucleotides left by *Tgm9* germinal excision are in bold font.

A

B

```
T322            ATAGCACATGTTTTCNTATTAATATTTTTAAAATGTTAATTAAAATAGATTTTTCAAATA 60
T369            ATAGCACATGTTTTCATATTAATATTTTTAAAATGTTAANTAAAATAGATTTTTCAAATA 60
T321            ATAGCACATGTTTTCATATTAATATTTTTAAAATGTTAATTAAAATAGATTTTTCAAATA 60
                ************** ********************** ********************

T322            TTTCATTTATCAGTATGAATGTATTTTTCTTAAAAACCTTAATGAGTCATCTGTATGCAT 120
T369            TTTCATTTATCAGTATGAATGTATTTTTCTTAAAAACCTTAATGAGTCATCTGTATGCAN 120
T321            TTTCATTTATCAGTATGAATGTATTTTTCTTAAAAACCTTAATGAGTCATCTGTATGCAT 120
                ************************************************************

T322            GACACCATATTAGTTTTTTATAAAAAAAAAAAATCAATAAGCTAGTATCTTATCCTTTACC 180
T369            GACACCATATTAGTTTTTTATAAAAAAAAAAAATCAATAAGCTAGTATCTTATCCTTTACC 180
T321            GACACCATATTAGTTTTTTATAAAAAAAAAAAATCAATAAGCTAGTATCTTATCCTTTACC 180
                ************************************************************

T322            ATCTACAAGATAATTTTGTTTTATTTAACTAAATCTATAATAAATAAATATTTTTTTAAC 240
T369            ATCTACAAGATAATTTTGTTTTATTTAACTAAATCTATAATAAATAAATATTTTTTTAAC 240
T321            ATCTACAAGATAATTTTGTTTTATTTAACTAAATCTATAATAAATAAATATTTTTTTAAC 240
                ************************************************************

T322            ATATAAATTATGTTATAATTAAAATTTCTAATAAATAAATATTTTTTATATATAAATTAT 300
T369            ATATAAATTATGTTATAATTAAAATTTCTAATAAATAAATATTTTTTATATATAAATTAT 300
T321            ATATAAATTATGTTATAATTAAAATTTCTAATAAATAAATATTTTTTATATATAAATTAT 300
                ************************************************************

T322            ANTGAAATTTATTACAAAAACTATTTCGTGATAATAAAATTACTTTTAATTATTGATACT 360
T369            ATTGAAATTTATTACAAAAACTATTTCGTGATAATAAAATTACTTTTAATTATTGATACT 360
T321            ATTGAAATTTATTACAAAAACTATTTCGTGATAATAAAATTACTTTTAATTATTGATACT 360
                * ************************************************************

T322            CTTTTAAAAAATATAAAAAATTAAATTCAAAAATAAAAATGAAAATAATAGATTTAATAA 420
T369            CTTTTAAAAAATATAAAAAATTAAATTCAAAAATAAAAATGAAAATAATAGATTTAATAA 420
T321            CTTTTAAAAAATATAAAAAATTAAATTCAAAAATAAAAATGAAAATAATAGATTTAATAA 420
                ************************************************************

T322            GATAAGAGGTCAATAAAAATTTCTAAAAACGCTTTCATTGAAAAATAAATTCTAAAAAAA 480
T369            GATAAGAGGTCAATAAAAATTTCTAAAAACGCTTTCATTGAAAAATAAATTCTAAAAAAA 480
T321            AATAAGAGGTCAATAAAAATTTCTAAAAACGCTTTCTTTGAAAAATAAATTCTAAAAAAA 480
                 ********************************** ***********************

T322            ACACATTCAAAAAGAAATTCTAAAAGAAAAGAGAAATAGGTATTATATATTTAATTTACT 540
T369            ACACATTCAAAAAGAAATTCTAAAAGAAAAGAGAAATAGGTATTATATATTTAATTTACT 540
T321            ACACATTCAAAAAGAAATTCTAAAAGAAAAGAGAAATAGGTATTATATATTTAATTTACT 540
                ************************************************************

T322            ATCAAAAAATCTTTTGAAAATTTAAACTCTTGTTATATTAGTCTTTACAGAAGGCTAATG 600
T369            ATCAAAAAATCTTTTGAAAATTTAAACTCTTGTTATATTAGTCTTTACAGAAGGCTAATG 600
T321            ATCAAAAAATCTTTTGAAAATTTAAACTCTTGTTATATTAGTCTTTACAGAAGGCTAATG 600
                ************************************************************

T322            TAATGCTATAATTGTAATATATTAAATTACTGACAAAAAATTTAAAATATTAGTATCATA 660
T369            TAATGCTATAATTGTAATATATTAAATTACTGACAAAAAATTTAAAATATTAGTATCATA 660
T321            TAATGCTATAATTGTAATATATTAAATTACTGACAAAAAATTTAAAATATTAGTATCATA 660
                ************************************************************

T322            CACTTAAAGAGACATATTTAAAATCAAATGAATTTAGTTTGATTAGTTAAACATATAATG 720
T369            CACTTAAAGAGACATATTTAAAATCAAATGAATTTAGTTTGATTAGTTAAACATATAATG 720
T321            CACTTAAAGAGACATATTTAAAATCAAATGAATTTAGTTTGATTAGTTAAACATATAATG 720
                ************************************************************

T322            TATATAATTTAAGAGTATTATAAATCTTTTAATAATGTGTTGAATTCTTACCAATAAACA 780
T369            TATATAATTTAAGAGTATTATAAATCTTTTAATAATGTGTTGAATTCTTACCAATAAACA 780
T321            TATATAATTTAAGAGTATTATAAATCTTTTAATAATGTGTTGAATTCTTACCAATAANCA 780
                ************************************************************ **

T322            ATAAAAAGGATATTAAAAGCTCA 803
T369            ATAAAAAGGATATTAAAAGCTCA 803
T321            ATAAAAAGGATATTAAAAGCTCA 803
                ***********************
```

FIGURE S6. —Promoters of *w4-dp* (T321) and *w4-p* (T369) alleles were intact following insertion of *Tgm9*. (A) Schematic representation of the promoter region amplified by PCR is shown. Inverted triangle showed the *Tgm9* insertion site. (B) PCR amplified promoter sequences of parental line T322 (*w4*-m) and two stable mutants *w4-dp* (T321) and *w4-p* (T369) are compared. Primer DFRF, CCTATGCCATGTGAGAATAAAGCAG; Primer DFRR, CCGTATGAAGTGGGTGCTTTTATAG.

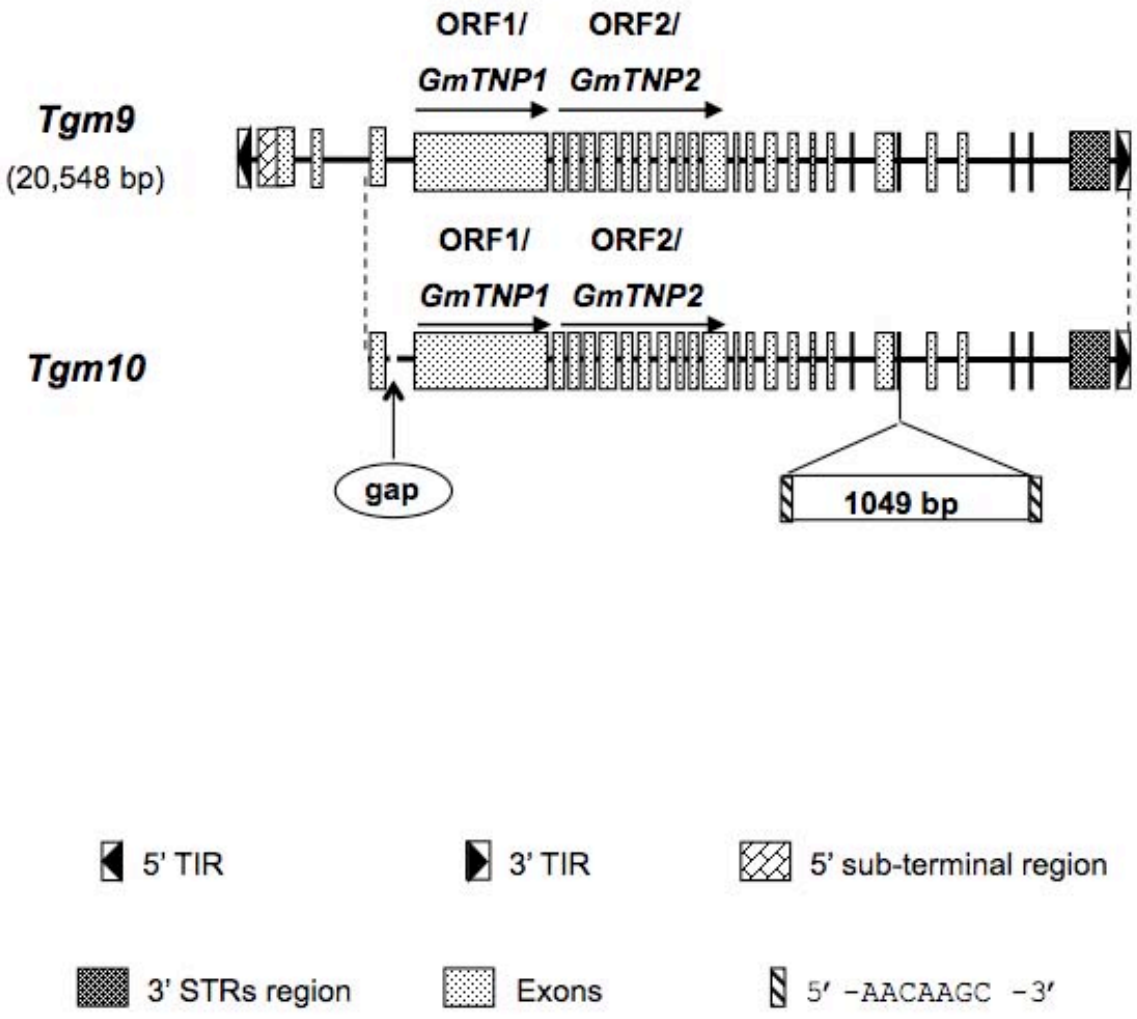

FIGURE S7.—Schematic representation of *Tgm10*. *Tgm10* is located at the 3' end of Scaffold_57. Except for a ~4100bp deletion in 5' end, a gap in 5'end, and a 1049 bp insertion flanked with a 7 bp direct repeat in exon XXIII, *Tgm10* is 99% identical to *Tgm9* element.

TABLE S1

Primers used in this study

Primer[a]	Sequence
ANS1S	5'-ATGCACTTGGTGAACCATGG-3'
DFR1S	5'-ATCTTGCTGAAGAGGGAAGC-3'
DFR2S	5'-ACAACGAGAGAGAGAACATG-3'
DFR3S	5'-AGCCTACAATAAACGGATTG-3'
DFR4S	5'-CCCGTTCTTCTATCTTTTTCTG-3'
Tn3'1S	5'-GGTTGCAGGAGAAAACCGTCTTAGTATGTC-3'
Tn3'2S	5'-CGGTTTTCGTAACAATCGTC-3'
P1	5'CCCAATCCAAGCTGCGACCTTCAAAG-3'
P3	5'-GGACGAGCCTTTCATCATGGCAGCAC-3'
ANS1R	5'-AATGCTCTTGTACTTGCCGTTG-3'
DFR1R	5'-TGTACATGTCCTCTAAGCTG-3'
DFR2R	5'-TCCCTCTTGAGCAAGATCAG-3'
DFR3R	5'-ATGATATGGTAATGGGACTC-3'
DFR4R	5'-GGACAAAGACAATGCAGGTCCACATCGAAG-3'
Tn3'1R	5'-GACATTCATGTATTCACTAGTACAAATAAAG-3'
P2	5'-CGCCAGAACCCACTTTGTAGCGTGAC-3'
P4	5'-GCTCCCATGTTTGCTGCTCCATACCA-3'
P5	5'-CTGATGCAAGTATGCTCCTAAGTAC-3'

[a] Primers used for sequencing the element are not listed here. ANS1F and ANS1R were designed from a partial coding sequence of an *ANS* gene identified from soybean seed coats (AF325853). DFR1F and DFR1R were designed according to the consensus sequence of three legume *DFR* genes (AF167556 from *G. max*; AF117263 from *Lotus corniculatus*; and AY389346 from *Medicago truncatula*).

TABLE S2

Probes used in this study

Probe [a]	Description
ANS partial cDNA	cDNA fragment amplified from purple petals of T322 using ANS1F and ANS1R primers.
F3H	An *F3H* EST clone (BM093886) provided by Dr. R. C. Shoemaker (Ames, IA)
DFR2 partial cDNA	cDNA fragment amplified from petals of T322 using DFR1F and DFR1R primers.
DFR 5'	cDNA fragment amplified from petals of T322 using DFR2S and DFR2R primers.
DFR3'	cDNA fragment amplified from petals of T322 using DFR3S and DFR3R primers.
Tgm9 3' end	PCR fragment amplified using primers TN3'2S and TN3'1R from a lambda clone containing the *w4-m* allele, isolated from the 1st T322 lambda genomic library.

[a] All the probes were labeled with α-^{32}P-dATP using Primer-it II randomly labeling kit (Stratagene, La Jolla, CA).

TABLE S3

Polymorphic sequences between *Tgm9* and *Tgmt**

Nucleotide position in *Tgm9*		Nucleotide in *Tgm9*	Nucleotide in *Tgmt**[a]
293	(5' STR Exon 1)	T	Y
746	(Intron 1)	C	Y
920	(Intron 2)	C	Y
1645	(Intron 2)	T	C
2043	(Intron 2)	T	C
2192	(Intron 2)	T	A
2146	(Intron 2)	T	C
2923	(Intron 2)	T	-
4969	(Intron 3)	A	G
5129	(Intron 3)	T	C
5272	(Intron 3)	T	-
5715	(Intron 3)	T	C
5851	(Intron 3)	T	-
5863	(Intron 3)	T	-
5999	(Intron 3)	T	C
6651	(Exon 4)	A	G
6819	(Exon 4)	C	Y
8010	(Exon 4)	G	R
8019	(Exon 4)	C	Y
8078	(Exon 4)	C	Y
13359	(Intron 16)	T	C
13524	(Exon 17)	A	G
13616	(Exon 17)	A	G
13904	(Exon 18)	G	A
15519	(Intron 21)	T	C
17169	(Exon25)	A	G

[a] "-" represents nucleotide missing

TABLE S4

Comparison of 3' terminal inverted repeats among soybean transposable elements

Transposon	3' TIR [a]	Identity to *Tgm9*
Tgm9	5'-**CACTACTACAAATAAAGCTTTTTAAGTCGG-**3'	100%
*Tgmt**	5'-**CACTACTACAAATAAAGCTTTTTAAGTCGG-**3'	100%
Tgm-express1	5'-**CACTACTACAAA**AG**A**G**G**T**TTTTTAAGTCGG**-3'	87%
Tgm1	5'-**CACTATTACAAAAA**GTAG**TTTT**AA**C**A**TCGG**-3'	70%
Tgm6	5'-**CACTACTACAAAAA**GCAG**TTTT**AA**C**A**TCG**A-3'	70%

[a] Nucleotides in bold are identical to the ones in *Tgm9* 3'TIR.