Dt2 Is a Gain-of-Function MADS-Domain Factor Gene That Specifies Semideterminacy in Soybean

Jieqing Ping
Purdue University

Yunfeng Liu
Purdue University

Lianjun Sun
Purdue University

Meixia Zhao
Purdue University

Yinghui Li
Chinese Academy of Agricultural Sciences

Follow this and additional works at: http://digitalcommons.unl.edu/agronomyfacpub

Part of the Agricultural Science Commons, Agriculture Commons, Agronomy and Crop Sciences Commons, Botany Commons, Horticulture Commons, Other Plant Sciences Commons, and the Plant Biology Commons

Ping, Jieqing; Liu, Yunfeng; Sun, Lianjun; Zhao, Meixia; Li, Yinghui; Tang, Zongxiang; Nguyen, Hanh; Tian, Zhixi; Qiu, Lijuan; Nelson, Randall L.; Clemente, Thomas E.; Specht, James; and Ma, Jianxin, "Dt2 Is a Gain-of-Function MADS-Domain Factor Gene That Specifies Semideterminacy in Soybean" (2014). Agronomy & Horticulture -- Faculty Publications. 866.
http://digitalcommons.unl.edu/agronomyfacpub/866

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Dt2 Is a Gain-of-Function MADS-Domain Factor Gene That Specifies Semideterminacy in Soybean

Jieqing Ping, Yunfeng Liu, Lianjun Sun, Meixia Zhao, Yinghui Li, Maoyun She, Yi Sui, Feng Lin, Xiaodong Liu, Zongxiang Tang, Hanh Nguyen, Zhixi Tian, Lijuan Qiu, Randall L. Nelson, Thomas E. Clemente, James E. Specht, and Jianxin Ma

Department of Agronomy, Purdue University, West Lafayette, Indiana 47907
Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
Department of Agronomy and Horticulture/Center for Plant Science Innovation, University of Nebraska, Lincoln, Nebraska 68583
Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, U.S. Department of Agriculture–Agricultural Research Service, Department of Crop Sciences, University of Illinois, Urbana, Illinois 61801

Soybean (Glycine max) stem growth habit is a key adaptation and agronomic trait that directly affects plant height, flowering time and duration, node production, leaf morphology, root architecture, maturity, water use efficiency, abiotic stress tolerance, and, ultimately, soybean yield (Bernard, 1972; Specht et al., 2001; Heatherly and Smith, 2004). Based on the timing of the termination of apical stem growth, most soybean cultivars can be classified into two categories of stem architecture, commonly known as determinate and indeterminate types. A determinate stem arises when apical stem growth abruptly ceases at the onset of floral induction. This generally produces a thick stem because latitudinal growth in stem girth continues after apical growth in stem length has ceased. An indeterminate stem tip continues terminal growth, as does its lateral growth, though both cease at the onset of seed filling, thus producing a stem that is tapered in thickness from base to tip. Despite this simple classification, the abruptness of stem termination varies among soybean accessions in the USDA Soybean Germplasm Collection, with phenotypic scores ranging from 1 (very determinate) to 5 (very indeterminate). Scores of <2.0 are generally classified as determinate, scores equal to or greater than 2.0 and less than 2.5 as semideterminate and scores of 2.5 or greater as indeterminate (http://www.ars-grin.gov/npgs/descriptors/soybean).

In the US and China, most of the soybean cultivars grown in the north are indeterminate types, which allow for more overlap of vegetative growth with reproductive development, providing better adaptation to a shorter growing season. In contrast, most of the cultivars grown in the south are determinate types, which have distinctly separate vegetative and reproductive stages (Heatherly and Elmore, 2004). Semideterminate cultivars are also useful in the north, and while they usually produce fewer stem nodes than indeterminate cultivars, they do not require a dense seeding rate to achieve yields like determinate cultivars. Moreover, the semideterminate cultivars are somewhat shorter than indeterminate cultivars, which provide some degree of lodging resistance (Chang et al., 1982), similar to that achieved by the “green revolution” gene in cereals (Peng et al., 1999). In the past decade, more semideterminate cultivars have been developed for use, particularly in high-yield, lodging-prone environments where short stature is desirable; for example, NE3001 is one such...
semideterminate cultivar that performs extremely well in irrigated production systems (Setiyono et al., 2007). Actually, semideterminate cultivars produce even more pods per plant than indeterminate cultivars if they do not lodge (Setiyono et al., 2007). Hence, it was deemed worthwhile to explore soybean yield potential by modifying genes affecting stem architecture and optimizing management practices.

Classical genetic analyses demonstrated that soybean stem growth habit was regulated by an epistatic interaction between two major genes, Dt1 and Dt2 (Bernard, 1972). In Dt1Dt1 genetic backgrounds, Dt2Dt2 genotypes produce semideterminate phenotypes, whereas dt2dt2 genotypes produce indeterminate phenotypes. However, in dt1dt1 genetic backgrounds, the phenotype is determinate, indicating an epistatic effect of the dt1 allele on the expression of the Dt2/dt2 locus. Because Dt1 is incompletely dominant over dt1, Dt1/dt1 heterozygotes are also semideterminate, whereas Dt2 is completely dominant over dt2; a dihybrid (Dt1d1;Dt2d2) produces progeny with an F2 phenotypic ratio of 1 indeterminate:11 semideterminate:4 determinate. Recent studies showed that Dt1 was a functionally conserved ortholog of Arabidopsis thaliana TERMINAL FLOWER1 (TFL1) (Liu et al., 2010; Tian et al., 2010), a floral suppressor gene primarily expressed in shoot apical meristems (SAMS) (Shannon and Meeks-Wagner, 1991; Bradley et al., 1997), and that the transition from indeterminate phenotype to determinate phenotype was caused by independent artificial selection of four point mutations in the TFL1 gene during soybean domestication (Tian et al., 2010). The Dt1 locus is located on chromosome 19 (Liu et al., 2007; Tian et al., 2010). The Dt2 locus was inferentially localized to the distal end of chromosome 18 because of its linkage to a gene governing the isozyme mannose-6-phosphate isomerase (MPI) that was mapped there (Muehlbauer et al., 1989).

Semideterminate stem termination was also observed and genetically investigated in tomato (Solanum lycopersicum) (Elkind et al., 1991; Pnueli et al., 1998; Friedman et al., 2002) and two legume species, pigeon pea (Cajanus cajan) (Gupta and Kapoor, 1991), and chickpea (Cicer arietinum) (Hedge, 2011). In tomato, the stem growth habit was found to be regulated by two genes, SELF-PRUNING (SP), the TFL1/Dt1 equivalent, and the Sdt/sdt locus responsible for semideterminacy. However, unlike in soybean, semideterminacy (sdsdt) in tomato is recessive, which is suppressed in the Sp- genotypes, leading to a dominant epistasis (i.e., 12 indeterminate:3 determinate:1 semideterminate individuals in F2 progeny derived from a dihybrid [Sp+/sdsdt]). This ratio has also been found in pigeon pea and chickpea. Because semideterminacy is dominant in soybean but recessive in the other three species, it is worthwhile to examine the evolutionary novelty of the genetic mechanism underlying semideterminacy in soybean.

Here, we report the isolation and characterization of the Dt2 gene by an integrated approach that involved linkage mapping, target gene association analysis, interspecific genetic and genomic comparison, profiling of gene expression, and complementation test. The research findings, coupled with the previous elucidation of Dt1, have laid the foundation for further dissection of the molecular mechanisms by which these genes and other factors act to determine soybean stem architecture.

RESULTS

Molecular Mapping of the Dt2 Locus to a Genomic Region Near the End of the Short Arm of Chromosome 18

To map the Dt2 gene, a cross between a semideterminate soybean cultivar NE3001 (Dt2Dt2;Dt1Dt1) and an indeterminate soybean cultivar IA3023 (dt2dt2;Dt1Dtl) (Setiyono et al., 2007) was made to generate an F2 population comprising 681 individual F2 plants. Each of the F2 plants were advanced to the F2.3 progenies, which were then used to deduce the genotypes of individual F2 plants. Based on high-confidence phenotyping data from the 679 F2.3 families, 156 F2 individuals were deduced as semideterminate homozygotes (Dt2Dt2), 350 F2 individuals as semideterminate heterozygotes (Dt2dt2), and 173 F2 individuals as indeterminate homozygotes (dt2dt2). These data confirmed the reported single-gene inheritance pattern of dominant semideterminacy versus recessive indeterminacy (3:1; \( \chi^2 \) test, \( P = 0.77 \)). The observed genotypic segregation pattern also fits the expected 1:2:1 ratio (\( \chi^2 \) test, \( P = 0.47 \)).

A previous linkage analysis with 20 F2 plants demonstrated that Dt2 was linked at ~17% recombination units from the gene MPI (Muehlbauer et al., 1989), which is located at 61.7 Mb, a position that is only ~0.6 Mb from the distal end of the short arm of chromosome 18 (Schmutz et al., 2010). Given this information, we then randomly selected simple sequence repeat (SSR) markers (Song et al., 2010) distributed in the 4 Mb (58 to 62 Mb) genomic segment located at the end of chromosome 18 to genotype the 679 F2 individuals and mapped the Dt2 locus to a 1.5-centimorgan region between SSR_18_1791 and SSR_18_1842, which spans 263 kb, according to the reference genome sequence (Figure 1B). Subsequently, polymorphic markers SSR_18_1821, SSR_18_1822, and SSR_18_1825 were used to search for recombinants identifiable between SSR_18_1791 and SSR_18_1842, and among the 47, we discovered from the 679 F2 individuals, 1, 0, and 2 recombination events were detected in a 81-kb region bounded by SSR_18_1821 and SSR_18_1825 (Figure 1C). In an attempt to further narrow the region encompassing the presumptive Dt2 gene, we next developed six single nucleotide polymorphism (SNP) markers within the 81-kb region by sequencing DNA fragments from genes adjacent to the boundaries of the region in the two parents (Figure 1B; Supplemental Table 1). These markers were used to genotype the three recombinants detected by SSR_18_1821 and SSR_18_1825, but no additional recombination events were identified.

Sequence Comparison between Semideterminate and Indeterminate Soybean Lines

According to the Williams 82 reference genome, 10 genes were predicted in the defined 81-kb Dt2 region (Figure 1D; Supplemental Table 2). In an attempt to pinpoint the candidate gene for Dt2, we amplified and sequenced the coding regions of the 10 genes in the two parents NE3001 and IA3023. In each of the three genes, Glyma18g50910, Glyma18g50960, and Glyma18g50980, a single nucleotide variant (SNV) in the predicted coding region was observably different between the two parents, and each of the SNVs altered an amino acid (Figure 1E).
Then, the coding regions of these three genes in the semideterminate near isogenic lines (NILs) of Harosoy L62-364, the semideterminate soybean variety LG90-2550, and the indeterminate Harosoy were sequenced and compared with the coding sequences of these genes from six highly diverged *Glycine soja* (the progenitor species of cultivated soybeans) accessions (Kim et al., 2010; Li et al., 2014). Each of the *G. soja* accessions contained the *Dt1* allele and exhibited an indeterminate phenotype. As shown in Figure 1E, none of the three SNVs detected as differing between the mapping population parents NE3001 and IA3023 were consistently associated with those two stem termination types in other accessions.

**Figure 1.** Map-Based Cloning of the *Dt2* Locus and Target Gene Association Analysis.

(A) Physical location of the *Dt2* regions in the Williams 82 reference genome. The bars indicate two arms of chromosome 18, and the circle indicates approximate position of the centromeric region.

(B) Physical locations of molecular markers defining the *Dt2* region.

(C) Graphical genotypes of recombinants carrying crossovers in the *Dt2* region determined by molecular markers and phenotypes of individual recombinants.

(D) Genes predicted in the defined *Dt2* region according to annotation of the reference genome and an LTR-retrotransposon located 2.7 kb upstream of *Glyma18g50910*.

(E) Comparison of the coding sequences of the three genes in the mapped *Dt2* region between two parental lines NE3001 and IA3023 and among additional semideterminate and indeterminate soybean accessions. In each of the three genes, the trinucleotide differences between semideterminate and indeterminate soybean accessions. In each of the three genes, the trinucleotide differences between semideterminate NE3001 and indeterminate IA3023 that resulted in a single amino acid difference (shown in square brackets) were not consistently associated with those two stem termination types in other accessions.
were able to distinguish the semideterminate accessions from the indeterminate ones examined (Figure 1E).

We further investigated the SNV detected in the coding sequence of Glyma18g50910 in a population including in 20 G. soja accessions and 17 soybean landraces (Hyten et al., 2006), which were phenotyped as “indeterminate” (Tian et al., 2010). Eleven and five were found to have the same nucleotide as NE3001 and the respective remaining ones were found to have the same nucleotide as IA3023 at this SNV site (Supplemental Table 3). Glyma18g50930, contained an ~1334-bp deletion in NE3001 compared with IA3023 and Williams 82 and appeared to be a pseudogene (null mutation) in the former. For the remaining six of the 10 genes in the 81-kb segment, the coding sequences between the two parents were identical. These observations suggest that it was most likely that the allelic difference between the Dt2 and dt2 alleles responsible for the phenotypic difference in stem growth habit could be attributed to the gene’s non-coding sequences or the flanking regulatory elements.

**Prediction of the Dt2 Candidate by Interspecific Comparison of Homologous Genes**

The 10 genes in the 81-kb Dt2 region were next compared with the whole set of genes annotated in the Arabidopsis genome (Arabidopsis Genome Initiative, 2000) by BLAST searches and analysis of interspecific syntenic genomic regions as described previously (Tian et al., 2010). Glyma18g50910 was found to be the only soybean gene in the mapped Dt2 region that had a significant match with the Arabidopsis genes involved in the Arabidopsis flowering networks (Liu et al., 2009; Yant et al., 2009; Formara et al., 2010). The best matches of Glyma18g50910 in Arabidopsis were the three floral homeotic MADS domain factor genes, which were the fruit tissue identity gene FRUITFUL (FUL) (Gu et al., 1998), the floral meristem identity gene APETALA1 (API) (Gustafson-Brown et al., 1994; Liljegren et al., 1999), and the floral regulatory gene CAULIFLOWER (CAL) (Kempin et al., 1995) (Figure 2; Supplemental Figure 1). It has been demonstrated that AP1 and another floral identity gene, LEAFY (LFY; Weigel et al., 1992), antagonize TFL1, the functional ortholog of the soybean Dt1, to regulate the fate of lateral meristems at the inflorescence apex in Arabidopsis (Bradley et al., 1997; Liljegren et al., 1999; Ratcliffe et al., 1999; Liu et al., 2009, 2013). These findings, along with all of our foregoing observations, including the deduced interaction between Dt2 and Dt1 and the mapping of Dt2, suggest that Glyma18g50910 was most likely to be the candidate for the Dt2 locus.

**Extrapolation of the Dt2 Candidate by Analysis of Gene Expression**

Given that none of the nucleotide changes in the 10 genes in the mapped Dt2 genomic region that resulted in amino acid changes could explain the phenotypic difference in stem growth habit between the semideterminate and indeterminate accessions examined (Figure 1E), the development of stem growth habit in soybean is very likely related to differential allelic expression at the Dt2/dt2 locus, and if this is the case, then the expression of Dt1 would be strongly downregulated by Dt2 and not regulated, or upregulated, by dt2.
indeterminate lines. By contrast, the expression level of
Dt2/dt2
pressed at higher level in the semideterminate lines than in the
dt2/dt2
lines upregulates the expression of
Glyma18g50910
or inversely, that recessive
in the semideterminate lines downregulates the expression of
Given transcription levels, suggest that dominant
Dt2/dt2
was the candidate gene for the Dt2/dt2 locus and that the semi-
derminacy was regulated by the transcriptional activity of this
gene. As expected for a MADS domain factor, the protein of this
candidate
Dt2
gene was localized to the nucleus (Supplemental Figure 4).

Figure 3. Expression of 10 Genes in the Mapped Dt2 Region in Apical
Stem Tips of NE3001 and IA3023 at V2 Stage Detected by qRT-PCR.
The y axis indicates the expression levels of individual genes (x axis)
relative to expression of Cons4. Expression levels were shown as
means ± standard errors of the means from four replicates.

Validation of the Dt2 Candidate by Complementation Test

To validate the candidacy of
Glyma18g50910
for the Dt2 locus, we introduced this candidate gene amplified from NE3001 into
Thome (McBlain et al., 1993), an indeterminate cultivar (dt2dt2;
Dt1Dt1) routinely used in
Agrobacterium tumefaciens–mediated soybean transformation experiments. In this study, two
constructs were made: one harboring a
Glyma18g50910 cassette regulated by the cauliflower mosaic virus 35S promoter, the
Coding sequence (CDS) of
Glyma18g50910
in NE3001, coupled with a 35S terminator (dubbed 35S: CDS-Dt2). The other
Genetic element consisted of the
Glyma18g50910 cassette reg-
ulated by the putative endogenous promoter that resides −2.5 kb
upstream of the CDS and terminated with −1.5 kb downstream of
Glyma18g50910 from NE3001 (dubbed Pro-Dt2: CDS-Dt2).

A total of nine independent events carrying the 35S: CDS-Dt2
expression and six independent events harboring the Pro-Dt2:
CDS-Dt2 transgenic allele were obtained. Progeny (T1) plants
from each event were advanced to T3 in the greenhouse and
subsequent T3 lineages were phenotyped for stem growth habit
under field conditions. As shown in Supplemental Table 4, in all
levels, irrespective of a Dt2Dt2 or dt2dt2 background. Together,
these expression analyses suggest that
Glyma18g50910 was the
candidate gene for the Dt2/dt2 locus and that the semi-
derminacy was regulated by the transcriptional activity of this
gene. As expected for a MADS domain factor, the protein of this
candidate
Dt2
gene was localized to the nucleus (Supplemental Figure 4).

Figure 4. Expression of Dt1 or dt1 and the Dt2/dt2 Candidate Gene
Glyma18g50910 in NE3001 and/or IA3023 Detected by qRT-PCR.
The y axis indicates expression of the Dt2 candidate gene or
Dt1/dt1 relative to expression of Cons4 in apical stem tips collected at four
developmental stages from V0 to V3 (V3, the stage begins when the 2nd
trifoliate leaflets are fully expanded but before the 3rd trifoliate leaflets are
still unrolled). Expression levels were shown as means ± standard errors
of the means from four replicates.

We also monitored the expression patterns of
Glyma18g50910 and
Dt1 in apical stem tips at the V2 stage in Harosoy and the
three Harosoy NILs. As shown in Figure 5, the expression level of
Glyma18g50910 in the Harosoy NIL L62–364 (Dt2Dt2;Dt1Dt1) is
similar to that in NE3001 (Dt2Dt2;Dt1Dt1), the expression level of
Glyma18g50910 in Harosoy (dt2dt2;Dt1Dt1) is similar to that in
IA3023 (dt2dt2;Dt1Dt1), and overall
Glyma18g50910 was ex-
pressed at higher level in the semideterminate lines than in the
indeterminate lines. By contrast, the expression level of
Dt1 in the
Dt2Dt2;Dt1Dt1 semideterminate genotypes was lower than that in the
dt2dt2;Dt1Dt1 indeterminate genotypes. These data, at the
given transcription levels, suggest that dominant
Glyma18g50910 in
the semideterminate lines downregulates the expression of
Dt1, or inversely, that recessive
Glyma18g50910 in the
indeterminate lines upregulates the expression of
Dt1. The two Harosoy NILs homozygous for dt1dt1 were found to be expressed at minimal
levels, irrespective of a Dt2Dt2 or dt2dt2 background. Together,
these expression analyses suggest that
Glyma18g50910 was the
candidate gene for the Dt2/dt2 locus and that the semi-
derminacy was regulated by the transcriptional activity of this
gene. As expected for a MADS domain factor, the protein of this
candidate
Dt2
gene was localized to the nucleus (Supplemental Figure 4).

Validation of the Dt2 Candidate by Complementation Test

To validate the candidacy of
Glyma18g50910 for the Dt2 locus, we introduced this candidate gene amplified from NE3001 into
Thome (McBlain et al., 1993), an indeterminate cultivar (dt2dt2;
Dt1Dt1) routinely used in
Agrobacterium tumefaciens–mediated soybean transformation experiments. In this study, two
constructs were made: one harboring a
Glyma18g50910 cassette regulated by the cauliflower mosaic virus 35S promoter, the
coding sequence (CDS) of
Glyma18g50910
in NE3001, coupled with a 35S terminator (dubbed 35S: CDS-Dt2). The other
genetic element consisted of the
Glyma18g50910 cassette reg-
ulated by the putative endogenous promoter that resides −2.5 kb
upstream of the CDS and terminated with −1.5 kb downstream of
Glyma18g50910 from NE3001 (dubbed Pro-Dt2: CDS-Dt2).

A total of nine independent events carrying the 35S: CDS-Dt2
expression and six independent events harboring the Pro-Dt2:
CDS-Dt2 transgenic allele were obtained. Progeny (T1) plants
from each event were advanced to T3 in the greenhouse and
subsequent T3 lineages were phenotyped for stem growth habit
under field conditions. As shown in Supplemental Table 4, in all
NE3001 was the other evidence described above, indicate that To shed light on potential causative mutation(s) at the
Indeterminate Soybean Lines Downstream Sequences between Semideterminate and
that led to differential allelic expression responsible for the
Dt2
any semideterminate plants. We speculated that it was likely that
observed between the semideterminate
transgenic plants (Figure 6F; Supplemental Figures 5 and 6). Of these SNPs, three
were found in the 2.5-kb sequence upstream of the CDS of the
locus, 23 (62%) in the first intron of the
locus, one in the second intron of the
locus, and one in the 1.5-kb downstream of the CDS of the
locus. However, because there are only a limited number of elite cultivars with clearly defined
semideterminate phenotypes available, and because it is difficult to distinguish semideterminate phenotypes from determinate phenotypes of plants with diverged genetic background, further identification of causative mutations by association analysis in the targeted region may not be very effective, or perhaps completely ineffective.

Figure 5. Expression of Dt1/dt1 and the Dt2/dt2 Candidate Gene in Apical Stem Tips of Different Genotypes Detected by qRT-PCR.
(A) Expression levels of Dt2 or dt2 relative to expression of Cons4.
(B) Expression levels of Dt1 or dt1 relative to expression of Cons4.
1 to 6 are NE3001 (Dt2/Dt2;Dt1/Dt1), IA3023 (dt2/dt2;Dt1/Dt1), L62-364 (Dt2/Dt2;Dt1/Dt1), Harosoy (dt2/dt2;Dt1/Dt1), L67-3256 (Dt2/Dt2;dt1/dt1), and L62-973 (dt2/dt2;dt1/dt1). Expression levels were shown as means ± standard errors of the means from four replicates.

nine 35S:CDS-Dt2 events, transgenic plants with semideterminate stems were observed. In addition, indeterminate plants were also observed in the T3 progenies derived from all nine transgenic events, and this phenotypic segregation was perfectly associated with the presence and absence of the
transgene (Supplemental Table 4). The semideterminacy varied among different events, which was largely reflected by the plants’ height and node numbers of the main stems (Figures 6E and 6F). Generally, the plants with similar expression levels to the
harosoy NIL and
several
were completely ineffective.

Nucleotide Variation in Dt2 and Its Upstream and Downstream Sequences between Semideterminate and Indeterminate Soybean Lines
To shed light on potential causative mutation(s) at the
locus that led to differential allelic expression responsible for the
phytotypic variation in soybean stem growth habit, we compared genomic sequences of the
locus and its flanking intragenic spaces that cover an ~22-kb region from the SSR marker 18-1821 to the 3’ untranslated region of the adjacent gene Glyma18g50920 (Figures 1D and 7A). The NE3001 sequences from this region were generated by PCR amplification and sequenced and were then compared with corresponding sequences from the Williams 82 reference genome. Subsequently, the forms of sequence variations, including SNPs and insertions/deletions (Indels) detected between these two soybean lines, in additional indeterminate soybean lines were determined using the available genome resequencing data and/or de novo genome sequencing data from IA3023 and six
soja accessions (Kim et al., 2010; Li et al., 2014; www.soybase.org). As shown in Figure 7, a total of 37 SNPs that each distinguished NE3001 from the eight indeterminate accessions were identified, and all of these indeterminate lines shared the same nucleotide at each of the 37 SNP sites (Figure 7B). Of these SNPs, three
flowered earlier than Thorne, the recipient line of the transgene (Figures 6A and 6B), similar to that observed between the semideterminate
Harosoy NIL and the indeterminate Harosoy NIL. These observations, together with other evidence described above, indicate that Glyma18g50910 in NE3001 was the
gene.
By contrast, none of the six Pro-Dt2:CDS-Dt2 events produced any semideterminate plants. We speculated that it was likely that some of the regulatory components essential for the expression of
were not included in the native-Dt2 construct. To test this possibility, we designed PCR primers that can specifically amplify the transcripts of the transgene and the Dt2/dt2. As expected, high levels of expression of
in NE3001, 35S:CDS-Dt2 transgene in semideterminate transgenic plants were detected in apical stem tips by qRT-PCR but were not detected in the Pro-Dt2:CDS-Dt2 transgenic plants (Figure 6F; Supplemental Figures 5 and 6).

DISCUSSION
Evolutionary Relationship and Novelty of the
Homologs among Soybean and Other Species
Members of MADS domain gene family play essential roles in various aspects of plant development, such as root, flower, seed, and fruit development (Smaczniak et al., 2012). Among the 104 MADS domain genes predicted in the
genome (Martinez-Castilla and Alvarez-Buylla, 2003), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1, AP1, FUL, CAL, and AGAMOUS-LIKE24 have been found to be activators involved in floral induction, a process that transforms the SAM, which forms leaves, into an inflorescence meristem, on which flowers form and develop (Liu et al., 2009; Yant et al., 2009; Fornara et al., 2010). However, despite their close phylogenetic relationships and functional similarities, none of the
, AP1, and CAL regions exhibited syntenic relationships with the
region (Shu et al., 2013). Indeed, Dt2 was not the gene most closely related to these three
genes based on their phylogeny (Figure 2; Supplemental Figure 1 and Supplemental Data Set 1). Hence, it remains unclear which
gene is the functional counterpart of Dt2.
If the established phylogenetic relationships of the MADS domain gene homologs within and between the
and soybean genomes did reflect the orders and timeframes within which these genes were diverged and generated, then Dt2 should
have more functionally diverged from AP1 than the four soybean genes Glyma02g13420, Glyma01g08150, Glyma16g13070, and Glyma08g36380, that were more closely related to AP1 (Figure 2). This speculation appears to be echoed by the observation that the proteins encoded by these four AP1-homologous genes of soybean all contain the conserved eudicot AP1-like (euAP1) motif present in the Arabidopsis AP1 (Rijpkema et al., 2007; Shan et al., 2007) at their C termini, whereas Dt2 contains the conserved paleoAP1 motif at its C terminus. Indeed, a previous study demonstrated that Gm-AP1 (i.e., Glyma16g13070) was most likely to be the functional homolog of the Arabidopsis AP1 in soybean, which is involved in flower development (Chi et al., 2011), although it remains unclear whether additional soybean genes closely related to AP1, such as Glyma02g13420, Glyma01g08150, and Glyma08g36380, also function like AP1 in Arabidopsis.

In Arabidopsis, TFL1 is primarily expressed in the center of SAMs at stem apexes, where the TFL1 protein is produced to repress the expression of AP1 (Shannon and Meeks-Wagner, 1991; Bradley et al., 1997). Such an interaction prevents the conversion of the vegetative SAMs there to a reproductive inflorescence meristem and thus inhibits terminal flowering. If the indeterminate soybean employs a mechanism similar to that

Figure 6. Overexpression of Transgene Dt2 in an Indeterminate Cultivar Resulting in Phenotypic Changes from Indeterminate to Semideterminate Types.

(A) Apical stem tip of immature Thorne showing indeterminate stem growth and late flowering.
(B) Apical stem tip of an immature Thorne 3SS:CDS-Dt2 transgenic plant showing semideterminate stem growth and early flowering.
(C) Apical stem tip of mature Thorne showing indeterminate stem growth.
(D) Apical stem tip of a mature Thorne 3SS:CDS-Dt2 transgenic plant showing semi-indeterminate stem growth.
(E) Photograph of IA3012 (I), NE3001 (N), Thorne (T), and nine T3 Thorne 3SS:CDS-Dt2 transgenic plants derived from nine independent transformation events, which show different degrees of apical stem termination and heights.
(F) Overexpression of transgene Dt2 in nine T3 Thorne 3SS:CDS-Dt2 transgenic plants (1 to 9), each derived from an independent transformation event, relative to expression of dt2 in Thorne in stem tips at the V2 stages (with two sets of unfolded trifoliate leaves). Values were shown as mean ± standard errors of the means from four replicates normalized to expression of dt2 in Thorne, which was set as 1.0. Cons4 was used as an endogenous control for gene expression analysis.
observed in *Arabidopsis*, in maintaining vegetative growth at the stem apexes, then *Dt2* is unlikely to be the functional counterpart of *AP1*. Given that the semideterminate and determinate stem growth habit phenotypes are rarely observed in *G. soja* (Ting, 1946; Nagata, 1950), it would be reasonable to speculate that *Dt2* is a gain-of-function mutation, which occurred after the domestication events of the cultivated soybeans. Because the *Dt2Dt2 Dt1 Dt1* genotype produces semideterminate stems, it is highly likely that the terminal flowering in plants of such a genotype is initiated by the repression of *Dt1* expression in SAMs directly or indirectly by *Dt2*.

In addition to the distinction in the deduced digenic epistatic interactions underlying the stem growth habit and the inverse patterns of dominance-recessiveness of indeterminacy over semideterminacy in other species, the genes interacting with *Dt1* in soybean, *SP* in tomato, and their functional counterparts in chickpea and pigeon pea, as revealed by classical genetic analyses (Bernard, 1972; Elkind et al., 1991; Gupta and Kapoor, 1991; Hegde, 2011), appear to be different. Although *Sdt* (i.e., PW9-2-5) has not been isolated, it was found to cosegregate with *SP-9D* (no recombinants among 4029 gametates), the closest paralog of *SP* (Fridman et al., 2002), suggesting that *Sdt* may be a functionally diverged *SP* homolog. Given such similar inheritance patterns of the stem growth habit among tomato, chickpea, and pigeon pea, it is possible that the functional counterparts of *Sdt1* in these two legume species are also the *TFL1/Dt1*/*SP* homologs. By contrast, *Dt2* is neither a *Dt1* homolog nor located in the *Dt1* paralogous regions (Tian et al., 2010).

It has been documented that two genes, *DETERMINATE (Dt)* and *VEGETATIVE1 (Veg1)*, regulate the stem growth habit in pea (*Pisum sativum*). *Dt* appears to be the functional equivalent of *TFL1/Dt1* (Foucher et al. 2003), while *Veg1* is an *Arabidopsis AGAMOUS-LIKE79 (AGL79)-like MADS box gene and specifies secondary inflorescence meristem identity (Berbel et al., 2012). In the *DetDet* genetic backgrounds, *Veg1*/*Veg1* genotypes produce indeterminate phenotypes, whereas *veg1veg1* genotypes produce plants that never flower. However, in the *detdet* genetic backgrounds, the phenotype is determinate, indicating an epistatic effect of the *det* allele on the expression of the *Veg1*/*Veg1* locus, similar to the effect of *dt1* on the expression of the *Dt2/dt2* locus. Because the pea genome has not been sequenced, whether *Dt2* and *Veg1* are orthologs has not been firmly established by comparison of genome sequences. Nevertheless, comparative mapping in pea and *Medicago truncatula* located the putative *Veg1* ortholog *Medtr7g016630* (i.e., *Mt-FULc*) (Berbel et al., 2012) to a position at the top of *Medicago* chromosome 7 that corresponds to the position of *Veg1* at the bottom of the pea linkage group V (Hecht et al., 2005; Zhu et al., 2005) and the *Medicago* genomic region surrounding *FULc* appears to be orthologous to the *Dt2* (i.e., *Gm-FULc* described in Berbel et al., 2012) genomic region (Supplemental Table 6). As seen in *Dt2*, *Veg1* also contains the paleoAP1 motif at its C terminus. Together, these observations suggest that *Veg1* and *Dt2* may be orthologous genes. However, given the fact that *Dt2*, in the presence of *Dt1*, is responsible for the conversion of apical stems from vegetative growth to reproductive growth to produce semideterminate phenotypes, whereas *Veg1*, in the presence of *Det*, appears to be essential for development of second-order inflorescence (I) in the indeterminate pea plants (Singer et al., 1999; Berbel et al., 2012), the functional divergence between *Dt2* and *Veg1* is expected. Alternatively, because the *Dt1 dt1* genotypes produce semideterminate phenotypes due to incomplete dominance of *Dt1* to *dt1*, similar to those produced by *Dt1Dt1* in the presence of *Dt2*, the lack of semideterminate phenotypes in

Figure 7. Nucleotide Differences in *Dt2* and Its Flanking Intergenic Spaces That Distinguish NE3001 from IA3023 and Additional Indeterminate Varieties Examined.

(A) Distribution of 37 SNPs in an ~22-kb genomic region surrounding *Dt2*. Green boxes connected by a green bar indicate a long terminal repeat retrotransposon. Gray boxes and black boxes indicate untranslated regions and exons of Glyma18g50910, respectively. A cluster of high density of SNPs (from 16 to 33) within the first exon of the gene was illustrated on a magnified scale above the gene.

(B) Nucleotides at the 37 SNP positions, as shown in the (A), in the examined soybean accessions. Dashes indicate unknown nucleotides.
pea could be the explained by strong, complete dominance of Det over det (Singer et al., 1999), which may lead Veg1 to be hypo-
static to DetDet or Detdet to produce indeterminate phenotypes. Further examination and comparison of expression patterns of Dt1 versus Det and Dt2 versus Veg1 in the primary (I1) and I2 in-
florescence meristems of soybean and pea may help to eluci-
date functional similarity and/or divergence between Dt2 and Veg1.

Causative Mutation(s) and Differential Allelic Expression

Accumulating evidence has revealed regulatory roles of intronic sequences in gene expression. The introns of LFY are known to be critical for proper expression of LFY in monocots (Prasad et al., 2003; Bombles and Doebley, 2005; Rao et al., 2008). A more recent study revealed that the intron sequences, par-
ticularly the first introns of AP1 and FUL, were bound by the microRNA-targeted transcription factor SQUAMOSA PRO-
MOTER BINDING PROTEIN-LIKE3 (SPL3), and both intron and exon sequences of LFY were bound by SPL3 to activate the expression of these genes (Yamaguchi et al., 2009). If one ac-
cepts the thesis that the regulation of Dt2 in the semideterminate soybean varieties did need such cis-regulatory elements, par-
ticularly the intron sequences, bound by regulatory factors, as AP1 and FUL did in Arabidopsis, the lack of, or reduction in expression of the Pro-Dt2:CDS-Dt2 transgene and, thus, the failure in recovering the expected phenotypes versus over-
expression of the 35S:CDS-Dt2 cassette and the observed phenotypic switch of the transgenic plants from indeterminate type to semideterminate type would be explained by the lack of intron sequences in the Pro-Dt2:CDS-Dt2 cassette.

Unfortunately, the first intron of Glyma18g50910 in NE3001 is composed of a 4483-bp sequence enriched with T/A (72.4%), ar-
ranged in long strings of Ts, As, AsT, and/or TAs, and technical difficulties occurred when amplifying desirable and large genomic fragments from this portion of the gene for cloning. As a result, constructs with the complete genomic sequence of the gene were not successfully made. Finner-scale linkage mapping would be able to define the causative mutation(s) to a smaller region. However, given that 37 SNPs were found within in an ~22-kb region sur-
rounding Dt2, with 24 SNPs in the first intron of the gene (Figure 7), it remains challenging to pinpoint causative mutations, if located with this intron, by fine mapping with a manageable number of F2 families. Further effort that perhaps involves genetic and molecular approaches is needed toward the discovery of the cis-regulatory components responsible for the Dt2 activity.

Although Dt2 was expressed at a significantly higher level than dt2 in the apical stem tip, as measured by qRT-PCR, the ex-
pression of dt2 in both Dt1Dt1 and dt1dt1 genetic backgrounds was still substantial (Figure 5). One might have expected a qual-
titative expression difference (i.e., presence versus absence in Dt2 versus dt2 expression). However, a quantitative difference in ex-
pression was observed, indicating that it was sufficient for the complete dominance of Dt2 over dt2 in downregulating Dt1 ex-
pression or vice versa. Because the proportion of SAM tissue in the apical stem tip is relatively small, differential expression be-
 tween Dt2 and dt2 in SAM may not be fully reflected by the qRT-
PCR experiment. Nevertheless, a lower level of Dt1 expression

was detected in a Dt2Dt2 background than in a dt2dt2 back-
ground, demonstrating the regulatory role that Dt2 has on Dt1. Given that Dt1 is incompletely dominant to dt1 (Woodworth, 1932; Williams, 1950), such an interacting level of differential ex-
pression between Dt2 and dt2, and Dt1 expression may underlie the phenotypic difference between the Dt2Dt2;Dt1Dt1 and dt2dt2; 
Dt1Dt1 genotypes.

METHODS

Plant Materials

The mapping population was derived from the cross between the soybean (Glycine max) cultivars NE3001 (Dt2Dt2;Dt1Dt1) and IA3023 (dt2dt2; 
Dt1Dt1). The three NILs of the recurrent parent Harosoy (PI 548573) were 
L62-364 (PI 547681), L62-973 (PI 547687), and L67-3256 (PI 547703). Seed of these four lines were obtained from the USDA Soybean Germ-
plasm Collection. Transgenic soybean lines were phenotyped and ad-
vanced to T2 in the greenhouse from November 2012 to April 2013 and phenotyped in the field at West Lafayette, IN, in October 2013.

DNA Isolation, PCR, RNA Isolation, RT-PCR, and Sequencing

Genomic DNA isolation, PCR primer design, PCR amplification, PCR fragment purification, total RNA isolation, cDNA synthesis by RT-PCR, sequencing PCR, and RT-PCR fragments were conducted as previously described (Tian et al., 2010). Primers used for PCR, RT-PCR, and se-
quencing are listed in Supplemental Data Set 2.

Molecular Mapping

Because the environment in which soybean plants grow can have large effects on stem growth habit (Bernard, 1972; Specht et al., 2001; Heatherly and Smith, 2004; Setiyono et al., 2007), accurate genotyping of individual F2 plants is not always possible. We thus advanced the 681 F2 population to the F3 generation; subsequently, ~50 F2 plants grown from each F2 plant were scored for abruptness of stem termination in a field nursery located in Lincoln, NE, where NE3001 was developed. The stem termination pheno-
type of each F2 plant in each F3 progeny row was examined to determine if the F2 plants in a given row were all semideterminate, all indeterminate, or segregating in a ratio of 3:1 for semideterminate to indeterminate. This F2 progeny phenotyping resulted in an accurate deduction of a respective of F2 families.

Although Dt2 was expressed at a significantly higher level than dt2 in the apical stem tip, as measured by qRT-PCR, the expression of dt2 in both Dt1Dt1 and dt1dt1 genetic backgrounds was still substantial (Figure 5). One might have expected a qualitative expression difference (i.e., presence versus absence in Dt2 versus dt2 expression). However, a quantitative difference in expression was observed, indicating that it was sufficient for the complete dominance of Dt2 over dt2 in downregulating Dt1 expression or vice versa. Because the proportion of SAM tissue in the apical stem tip is relatively small, differential expression between Dt2 and dt2 in SAM may not be fully reflected by the qRT-PCR experiment. Nevertheless, a lower level of Dt1 expression

was detected in a Dt2Dt2 background than in a dt2dt2 background, demonstrating the regulatory role that Dt2 has on Dt1. Given that Dt1 is incompletely dominant to dt1 (Woodworth, 1932; Williams, 1950), such an interacting level of differential expression between Dt2 and dt2, and Dt1 expression may underlie the phenotypic difference between the Dt2Dt2;Dt1Dt1 and dt2dt2; Dt1Dt1 genotypes.

Sequence Alignments, Comparison, and Phylogenetic Analysis

BLASTp was used to search the soybean Dt2 candidate gene against the Arabidopsis thaliana protein database (www.arabidopsis.org) to identify homologous genes showing high levels of sequence similarity and then the identified Arabidopsis genes were used to identify homologous genes in soybean. A group of homologous genes between soybean and Arab-
idopsis that include all genes belonging to the AP1/SQUA subfamily in
Arabidopsis, i.e., AP1, CAL, FUL, and AGL79 (Ripkema et al., 2007; Shan et al., 2007), and all their homologs in soybean were identified. The full length of predicted protein sequences from these genes was aligned using MUSCLE (V3.6) with default parameters (Supplemental Data Set 1). The phylogenetic tree was generated using neighbor-joining method integrated in MEGA (V6.06) with a bootstrap of 1000 replicates, Poisson model for amino acid substitution, and pairwise deletion of gaps.

Plasmid Construction and Transformation

The 2.5-kb upstream sequence from the start codon (ATG) of Glyma18g50910, the 1.5-kb downstream sequence of the gene from the stop codon (TAG), and the CDS of the gene from the semideterminate cultivar NE3001 were obtained by PCR and RT-PCR with primers shown in Supplemental Data Set 2. The PCR fragments were ligated to pCR2.1-TOPO TA vector (Life technologies) and then sequenced. Selected clones with verified inserts by sequencing were used to make two Dt2 constructs: 35S promoter + the CDS of the Dt2 candidate + the 35S terminator (dubbed 35S:CDS-Dt2 construct) and the 2.5-kb upstream sequence + the CDS + the 1.5-kb downstream sequences of the Dt2 candidate gene (dubbed Pro-Dt2: CDS-Dt2 construct).

To make the 35S:CDS-Dt2 construct, a pCR2.1-TOPO clone carrying the CDS of Glyma18g50910 was digested with Ncol/BamHI and with BamHI/XbaI, respectively, to isolate a 348-bp fragment and a 414-bp fragment from the CDS of the gene. Simultaneously, the pRTL2 vector was digested with Ncol/XbaI to generate a linearized plasmid. These three restriction fragments were purified separately and then ligated to form the 35S-Dt2 construct using T4 DNA ligase (Promega), which was then transformed into competent Escherichia coli cells. To make the Pro-Dt2: CDS-Dt2 construct, selected pCR2.1-TOPO clones carrying the verified 2.5-kb upstream sequence and the 1.5-kb downstream sequence of Glyma18g50910 were digested with HindIII/Xhol and XbaI/HindIII respectively, to isolate the 2.5- and 1.5-kb fragments. The assembled and verified 35S:CDS-Dt2 construct (designated pPTN117) was digested with Xhol/XbaI to isolate the CDS of the gene. Simultaneously, the pPTN200 vector was digested with HindIII to generate a linearized plasmid. These four restriction fragments were purified and then ligated to form the Pro-Dt2: CDS-Dt2 construct, designated pPTN1178.

Both the 35S:CDS-Dt2 and Pro-Dt2:CDS-Dt2 constructs were confirmed by digestion with relevant restriction enzymes and by sequencing. Two confirmed constructs were introduced into Agrobacterium tumefaciens separately and subsequently transferred into the indeterminate soybean cultivar Thorne following a protocol as described previously (Clemente et al., 2000). The presence of the constructs in recovered transgenic plants was confirmed by PCR with primers specific to the cloning vectors (Supplemental Data Set 2).

Subcellular Localization

Subcellular localization of Dt2 was performed using coding sequence of a green fluorescent protein fused in-frame to the Dt2 coding sequence. The fusion plasmids were under the control of the cauliflower mosaic virus 35S promoter and introduced into leaf epidermal cells of 3- to 4-week-old Nicotiana benthamiana plants by Agrobacterium infiltration. The transformed leaf cells were observed and photographed through a microscope.

qRT-PCR

qRT-PCR was performed using SYBR Green PCR Master Mix (Life Technologies) as described previously (Tian et al., 2010). The soybean ATP binding cassette transporter gene (Glyma12g02310), dubbed Cons4 (Libault et al., 2008), was used as a control. Three biological replicates were analyzed to quantify the levels of gene expression in NE3001, IA3023, and the four NILs. Three technical replicates were analyzed to measure gene expression in Thorne and transgenic lines. Primers used for qRT-PCR are listed in Supplemental Data Set 2.

Accession Numbers

Sequence data from this article were submitted to the National Center for Biotechnology Information under accession numbers KF908014 to KF908015.

Supplemental Data

The following materials are available in the online version of the article.

Supplemental Figure 1. Alignment of Predicted MADS Box Domains of the Dt2 Candidate Gene Homologs in Soybean and Arabidopsis.

Supplemental Figure 2. Expression of the Dt2 Candidate Genes in the Semideterminate Soybean Cultivar NE3001 Detected by qRT-PCR.

Supplemental Figure 3. Expression of Glyma18g51000 in Apical Stem Tips of NILs L62-364 and Harosoy at V2 Stage Detected by qRT-PCR.

Supplemental Figure 4. Subcellular Localization of the Dt2 Protein in Tobacco Epidermal Cells.

Supplemental Figure 5. Expression of Endogenous Dt2/dt2 and/or the Transgenic Dt2 in Parental and Transgenic Lines Determined by qRT-PCR.

Supplemental Figure 6. Expression of Thorne Endogenous dt2 and the Transgenic Dt2 in the Pro-Dt2:CDS-Dt2 Transgenic Lines Determined by qRT-PCR.

Supplemental Table 1. Molecular Markers Used for Mapping of the Dt2 Gene.

Supplemental Table 2. Genes in the Defined Dt2 Region According to the Williams 82 Reference Genome.

Supplemental Table 3. Polymorphisms of a Single Nucleotide Variant in the Coding Region of the Dt2 Candidate Gene in a Natural Population Previously Described.

Supplemental Table 4. Phenotypic Segregation for Stem Growth Habit of the T3 Progenies from Individual T2 Plants Derived from Nine Independent Transformation Events.

Supplemental Table 5. Correlation between Expression Level of the Transgenes and Phenotypic Variation among Transgenic Plants.

Supplemental Table 6. Genes Surrounding Dt2 in Soybean and Their Putative Orthologs in Medicago truncatula.

Supplemental Data Set 1. Alignment Information Used for Phylogenetic Tree Construction.

Supplemental Data Set 2. Primers Used for Amplification of Gene Fragments by PCR, RT-PCR, qRT-PCR, Gene Cloning and Verification, and Sequencing.

ACKNOWLEDGMENTS

This work was mainly supported by soybean checkoff funds from the United Soybean Board (project numbers 0217 and 1420-532-5617) and partially supported by Purdue Agricultural Research Program, Purdue Research Foundation, and “Partnership for Research & Education in Plant Breeding and Genetics” program funded by the USDA’s National Institute of Food and Agriculture and corporate partners Ag Alumni Seed, AgReliant Genetics, Beek’s Hybrids, ConAgraFoods, Dow AgroSciences, Indiana Crop Improvement Association, and Pioneer Hi-Bred International.
AUTHOR CONTRIBUTIONS


Received April 21, 2014; revised June 3, 2012; accepted June 18, 2014; published July 8, 2014.

REFERENCES


Woodworth, C.M. (1932). Genetics and breeding in the improvement of the soybean.


**Supplemental Figure 1.** Alignment of predicted MADS-box domains of the *Dt2* candidate gene homologs in soybean and *Arabidopsis.*
Supplemental Figure 2. Expression of the Dt2 candidate genes in the semi-determinate soybean cultivar NE3001 detected by qRT-PCR. The y-axis indicates expression of Dt2 relative to expression of Cons4 in different tissues including cotyledon (CT), roots (RT), stems (SM), stem tips (ST), unifoliate leaflets (ULF), trifoliate leaflets (TLF), flowers (FL), and 1-cm immature pod (PD) at different developmental stages including V0 (when the cotyledons at node 0 are fully extended but the unifoliate leaflets at node 1 are not yet unrolled), V1 (unifoliate leaflets at node 1 fully expanded, but 1st trifoliate leaflets at node 2 not yet unrolled), and V2 (the first trifoliate leaflets have fully unrolled but 2nd trifoliate leaflets are still unrolled) stages as shown in x-axis. Expression levels were shown as means ± standard errors of the means from four replicates.
Supplemental Figure 3. Expression of *Glyma18g51000* in apical stem tips of NILs L62-364 and Harosoy at V2 stage detected by qRT-PCR. The y-axis indicates the expression levels of the gene relative to expression of Cos4. Expression levels were shown as means ± standard errors of the means from four replicates.
Supplemental Figure 4. Subcellular localization of the Dt2 protein in tobacco epidermal cells.

(A) Subcellular localization of Dt2-GFP fusion gene under the control of 35S promoter as observed with a dark field for green fluorescence.
(B) The same cell shown in (A) as observed with a bright field for the cell morphology.
(C) The merged image of (A) and (B).
(D) Subcellular localization of GFP protein as a control.
Supplemental Figure 5. Expression of endogenous Dt2/dt2 and/or the transgene Dt2 in parental and transgenic lines determined by qRT-PCR.

(A) Expression of Dt2/dt2 relative to expression of Cons4 in apical stem tips of IA3023 (I), NE3001 (N), Thorne (T), and a T3 Thorne Dt2 transgenic plant as shown in Figure 3 at the V2 stage determined by qRT-PCR. Expression levels were shown as means ± standard errors of the means from four replicates.

(B) Expression of transgene Dt2 relative to expression of Cons4 in the same samples as shown in (A).
Supplemental Figure 6. Expression of Thorne endogenous dt2 and the transgenic Dt2 in the Pro-Dt2:CDS-Dt2 transgenic lines determined by qRT-PCR. A. RT-PCR products of Thorne native dt2 and transgenic Dt2 amplified from Thorne and all six transgenic lines with a pair of Dt2/dt2 specific primers. The small fragments were PCR products amplified from the native dt2, whereas the larger fragments were amplified from the transgene Dt2 in six transgenic lines from distinct transformation events. Gene Actin11 was used as a control. B. Expression of Thorne dt2 and transgene Dt2 relative to expression of Cons4 detected by qRT-PCR with a pair of Dt2/dt2 specific primers (top plot), and specific expression of the transgene Dt2 relative to expression of Cons4 detected by qRT-PCR with one primer from the coding sequence of Dt2 and the other from pPTN1178 cassette.
**Supplemental Table 1.** Molecular markers used for mapping of the Dt2 gene.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chromosome ID</th>
<th>Forward primer sequences (5’ to 3’)</th>
<th>Reverse primer sequences (5’ to 3’)</th>
<th>Type of markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSR_18_1791</td>
<td>Gm18</td>
<td>TGACCAGTCAATTGTTCATTCTTT</td>
<td>TTTACTCAACCATCTCCGCA</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1807</td>
<td>Gm18</td>
<td>TCATTCTGTAAATGAGTTGTATTC</td>
<td>TTATTTTGCTTTCAACTTACATTCC</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1817</td>
<td>Gm18</td>
<td>GTGAGGCCATCAATCACCCTT</td>
<td>CGCAAGAAGAAAAAGAAAAAGGA</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1821</td>
<td>Gm18</td>
<td>GGTGCCTTTAATTTCTTCGGA</td>
<td>ATTCACACAGATCATGCGCA</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1822</td>
<td>Gm18</td>
<td>AATTGATGCACTTGATAACGA</td>
<td>TGCAAAACACAAAGAATCTACACAA</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1825</td>
<td>Gm18</td>
<td>GAATCCACCACCTACCAAACC</td>
<td>CAATGGCAACCCAGTAAGGT</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1831</td>
<td>Gm18</td>
<td>TGTTTTGGTTAAAATCTTTTTTTG</td>
<td>TGTTATATTTGTGTGTGCTT</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1833</td>
<td>Gm18</td>
<td>GGCTATTGCAACATTTGTGTCCTGTG</td>
<td>TGTTATGTGTGTGTGTGCACTT</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1838</td>
<td>Gm18</td>
<td>TTCTATATTCAAAACTGAACTGAACCTG</td>
<td>AACCTATTATAACGCAATTTATGCTT</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1842</td>
<td>Gm18</td>
<td>TGAAATGGAGGAGAAATGGA</td>
<td>GTCCGGGAAGACTGACC</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1846</td>
<td>Gm18</td>
<td>CTCTTAACGGATGCTGTCGGG</td>
<td>CGTAACACGGCACTGCTTTTGGGG</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1854</td>
<td>Gm18</td>
<td>GCCACCCTACACAACCACACA</td>
<td>TGACCAAAATGGGCTTTCAA</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1858</td>
<td>Gm18</td>
<td>TAGCTTTATAATGAGTGTGATAGAT</td>
<td>GTATGCAAGGATTAATTAAG</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1864</td>
<td>Gm18</td>
<td>TGAATGATATATGTTTTGCGAAGA</td>
<td>CAATAGAGCCGGATGGATGT</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1890</td>
<td>Gm18</td>
<td>TGTTAGTGACGCGTTACAAAATATAA</td>
<td>AAAGTCATGTACATTAGTGAATT</td>
<td>SSR</td>
</tr>
<tr>
<td>SSR_18_1926</td>
<td>Gm18</td>
<td>TTTGGAGATTACTGACAAAAGAGA</td>
<td>TTTTGTCCCTTTAAAAATAACTTCAAC</td>
<td>SSR</td>
</tr>
<tr>
<td>SNP1</td>
<td>Gm18</td>
<td>CTCTGTAATATGCTCAGAGTC</td>
<td>GTAGGTTGCAAGAAACCCCCCC</td>
<td>SNP</td>
</tr>
<tr>
<td>SNP2</td>
<td>Gm18</td>
<td>CAGACATAATCTATGAACAAG</td>
<td>GCAAACAAACCTAAAGGATCACAG</td>
<td>SNP</td>
</tr>
<tr>
<td>SNP3</td>
<td>Gm18</td>
<td>CCATGTACATTAGTATTCAGTAG</td>
<td>AGCAGCTCTGAAATTAGCC</td>
<td>SNP</td>
</tr>
<tr>
<td>SNP4</td>
<td>Gm18</td>
<td>GTGTTTATATTAGTCTTTACC</td>
<td>ACCATGTATAAAATGATAC</td>
<td>SNP</td>
</tr>
<tr>
<td>SNP5</td>
<td>Gm18</td>
<td>CAAGCACTATAGCCTTTAGTC</td>
<td>AGAAGCATTCTTGGAGAGGAAC</td>
<td>SNP</td>
</tr>
<tr>
<td>SNP6</td>
<td>Gm18</td>
<td>TGAAGCGGATCGAGAACAAAACA</td>
<td>AATGATGAACGAGTAGGAACCT</td>
<td>CAPS</td>
</tr>
</tbody>
</table>
**Supplemental Table 2.** Genes in the defined Dt2 region according to the Williams 82 reference genome.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyma18g50910</td>
<td>MADS box transcription factor</td>
</tr>
<tr>
<td>Glyma18g50920</td>
<td>Uncharacterized conserved protein, contains RCC1 domain</td>
</tr>
<tr>
<td>Glyma18g50930</td>
<td>MEKK and related serine/threonine protein kinases</td>
</tr>
<tr>
<td>Glyma18g50940</td>
<td>DSBA-like thioredoxin domain</td>
</tr>
<tr>
<td>Glyma18g50950</td>
<td>Ring finger protein 24-related</td>
</tr>
<tr>
<td>Glyma18g50960</td>
<td>No functional annotations</td>
</tr>
<tr>
<td>Glyma18g50970</td>
<td>Pollen proteins Ole e I like</td>
</tr>
<tr>
<td>Glyma18g50980</td>
<td>Glucosyl Transferases</td>
</tr>
<tr>
<td>Glyma18g50990</td>
<td>F-box domain</td>
</tr>
<tr>
<td>Glyma18g51000</td>
<td>F-box domain</td>
</tr>
</tbody>
</table>
### Supplemental Table 3. Polymorphisms of a single nucleotide variant in the coding region of the *Dt2* candidate gene in a natural population previously described\(^a\)

<table>
<thead>
<tr>
<th>Accessions</th>
<th>Genotype</th>
<th>Phenotype</th>
<th>Type</th>
<th>Region</th>
<th>Country</th>
<th>Maturity Group</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 483464A</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Ningxia</td>
<td>China</td>
<td>III</td>
<td>G</td>
</tr>
<tr>
<td>PI 407301</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Jiangsu</td>
<td>China</td>
<td>V</td>
<td>G</td>
</tr>
<tr>
<td>PI 483465</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Shaanxi</td>
<td>China</td>
<td>V</td>
<td>G</td>
</tr>
<tr>
<td>PI 468400A</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Ningxia</td>
<td>China</td>
<td>IV</td>
<td>G</td>
</tr>
<tr>
<td>PI 407131</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Kumamoto</td>
<td>Japan</td>
<td>VI</td>
<td>G</td>
</tr>
<tr>
<td>PI 447004</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Jilin</td>
<td>China</td>
<td>III</td>
<td>G</td>
</tr>
<tr>
<td>PI 366120</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Akita</td>
<td>Japan</td>
<td>IV</td>
<td>G</td>
</tr>
<tr>
<td>PI 407170</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Kyonggi</td>
<td>Korea, South</td>
<td>V</td>
<td>G</td>
</tr>
<tr>
<td>PI 549046</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Shaanxi</td>
<td>China</td>
<td>III</td>
<td>G</td>
</tr>
<tr>
<td>PI 407140</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Kumamoto</td>
<td>Japan</td>
<td>VII</td>
<td>A</td>
</tr>
<tr>
<td>PI 326582A</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Primorye</td>
<td>Russia</td>
<td>II</td>
<td>A</td>
</tr>
<tr>
<td>PI 464935</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Jiangsu</td>
<td>China</td>
<td>VI</td>
<td>A</td>
</tr>
<tr>
<td>PI 468916</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Liaoning</td>
<td>China</td>
<td>III</td>
<td>A</td>
</tr>
<tr>
<td>PI 339871A</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Cheju</td>
<td>Korea</td>
<td>V</td>
<td>A</td>
</tr>
<tr>
<td>PI 458538</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Heilongjiang</td>
<td>China</td>
<td>OOO</td>
<td>A</td>
</tr>
<tr>
<td>PI 597459D</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Shandong</td>
<td>China</td>
<td>III</td>
<td>A</td>
</tr>
<tr>
<td>PI 393551</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Taiwan</td>
<td>China</td>
<td>X</td>
<td>A</td>
</tr>
<tr>
<td>PI 597461A</td>
<td><em>Dt1</em></td>
<td>indeterminate</td>
<td><em>Glycine soja</em></td>
<td>Shandong</td>
<td>China</td>
<td>IV</td>
<td>A</td>
</tr>
</tbody>
</table>
### Supplemental Data


<table>
<thead>
<tr>
<th>Code</th>
<th>Type</th>
<th>Origin</th>
<th>Species</th>
<th>Location</th>
<th>Stage</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 562559</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>Glycine soja</td>
<td>Cholla Puk</td>
<td>Korea, South</td>
<td>V</td>
</tr>
<tr>
<td>PI 407282</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>Glycine soja</td>
<td>Cheju</td>
<td>Korea, South</td>
<td>VI</td>
</tr>
<tr>
<td>PI 548603</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Indiana</td>
<td>USA</td>
<td>IV</td>
</tr>
<tr>
<td>PI 548488</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Missouri</td>
<td>USA</td>
<td>V</td>
</tr>
<tr>
<td>PI 548311</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Ontario</td>
<td>Canada</td>
<td>O</td>
</tr>
<tr>
<td>PI 548379</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Heilongjiang</td>
<td>China</td>
<td>O</td>
</tr>
<tr>
<td>PI 548298</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Unknown</td>
<td>China</td>
<td>III</td>
</tr>
<tr>
<td>PI 548318</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Jilin</td>
<td>China</td>
<td>III</td>
</tr>
<tr>
<td>PI 548348</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Unknown</td>
<td>China</td>
<td>III</td>
</tr>
<tr>
<td>PI 548362</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Unknown</td>
<td>Unknown</td>
<td>III</td>
</tr>
<tr>
<td>PI 548391</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Liaoning</td>
<td>China</td>
<td>II</td>
</tr>
<tr>
<td>FC 33243</td>
<td>Dt1</td>
<td>indeterminate</td>
<td>N. Am. Ancestor</td>
<td>Liaoning</td>
<td>Unknown</td>
<td>IV</td>
</tr>
<tr>
<td>PI 548406</td>
<td>Dt1</td>
<td>semideterminate</td>
<td>N. Am. Ancestor</td>
<td>Jilin</td>
<td>China</td>
<td>II</td>
</tr>
<tr>
<td>PI 548382</td>
<td>dt1</td>
<td>semideterminate</td>
<td>N. Am. Ancestor</td>
<td>Liaoning</td>
<td>China</td>
<td>OO</td>
</tr>
<tr>
<td>PI 548485</td>
<td>dt1</td>
<td>determinate</td>
<td>N. Am. Ancestor</td>
<td>Jiangsu</td>
<td>China</td>
<td>VII</td>
</tr>
<tr>
<td>PI 548477</td>
<td>dt1</td>
<td>determinate</td>
<td>N. Am. Ancestor</td>
<td>Tennessee</td>
<td>USA</td>
<td>VI</td>
</tr>
<tr>
<td>PI 548657</td>
<td>dt1</td>
<td>determinate</td>
<td>N. Am. Ancestor</td>
<td>North Carolina</td>
<td>USA</td>
<td>VII</td>
</tr>
<tr>
<td>PI 548445</td>
<td>dt1</td>
<td>determinate</td>
<td>N. Am. Ancestor</td>
<td>Jiangsu</td>
<td>China</td>
<td>VII</td>
</tr>
<tr>
<td>PI 548456</td>
<td>dt1</td>
<td>determinate</td>
<td>N. Am. Ancestor</td>
<td>Pyongyang</td>
<td>Korea, North</td>
<td>VI</td>
</tr>
</tbody>
</table>

**Supplemental Table 4.** Phenotypic segregation for stem growth habit of the T3 progenies from individual T2 plants derived from nine independent transformation events

<table>
<thead>
<tr>
<th>No. in Fig. 3</th>
<th>Transformation event</th>
<th>No. of positive T2 plants planted in greenhouse</th>
<th>No. of semi-determinate T3 plants in the field</th>
<th>No. of indeterminate T3 plants in the field</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>917-70</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>917-49</td>
<td>2</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>913-15</td>
<td>4</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>917-46</td>
<td>4</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>917-56</td>
<td>1</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>917-55</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>917-66</td>
<td>2</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>917-24</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>917-65</td>
<td>1</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>
**Supplemental Table 5.** Correlation between expression level of the transgenes and phenotypic variation among transgenic plants

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Thorne</th>
<th>913-15</th>
<th>917-49</th>
<th>917-46</th>
<th>917-24</th>
<th>917-70</th>
<th>917-56</th>
<th>917-55</th>
<th>917-65</th>
<th>917-66</th>
<th>$r^a$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node number</td>
<td>21</td>
<td>10</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>20</td>
<td>21</td>
<td>-0.842$^{**,b}$</td>
<td>0.004$^b$</td>
</tr>
<tr>
<td>Expression Level</td>
<td>1</td>
<td>29.5</td>
<td>12.2</td>
<td>6.2</td>
<td>7.4</td>
<td>5.9</td>
<td>4.2</td>
<td>4.3</td>
<td>2</td>
<td>2.4</td>
<td>-0.815$^{**,c}$</td>
<td>0.007$^c$</td>
</tr>
<tr>
<td>Plant height (cm)</td>
<td>82.4</td>
<td>26.1</td>
<td>33.6</td>
<td>39.6</td>
<td>40.1</td>
<td>41.9</td>
<td>42.1</td>
<td>42.5</td>
<td>55.2</td>
<td>57.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Pearson’s correlation coefficients were calculated using the SPSS statistics package

$^b$Correlation between node numbers and expression levels of the transgene

$^c$Correlation between plant heights and expression levels of the transgene

$^{**,}$Correlation is significant at the 0.01 level (2-tailed)
### Supplemental Table 6. Genes surrounding *Dt2* in soybean and their putative orthologs in *Medicago truncatula*

<table>
<thead>
<tr>
<th>Query genes in soybean&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Putative orthologs in <em>Medicago</em></th>
<th>BLASTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyma18g50900</td>
<td>Medtr7g016600</td>
<td>82.03</td>
</tr>
<tr>
<td>Glyma18g50910 (<em>Dt2</em>)</td>
<td>Medtr7g016630 (Mt-FULc)</td>
<td>72.59</td>
</tr>
<tr>
<td>Glyma18g50920</td>
<td>Medtr7g016640</td>
<td>88.45</td>
</tr>
<tr>
<td>Glyma18g50940</td>
<td>Medtr7g016650</td>
<td>85.38</td>
</tr>
<tr>
<td>Glyma18g50950</td>
<td>Medtr7g016840</td>
<td>76.99</td>
</tr>
<tr>
<td>Glyma18g50960</td>
<td>Medtr7g016900</td>
<td>82.78</td>
</tr>
<tr>
<td>Glyma18g50970</td>
<td>Medtr7g016950</td>
<td>63.46</td>
</tr>
<tr>
<td>Glyma18g51040</td>
<td>Medtr7g016960</td>
<td>77.88</td>
</tr>
<tr>
<td>Glyma18g51050</td>
<td>Medtr7g016970</td>
<td>70.07</td>
</tr>
<tr>
<td>Glyma18g51060</td>
<td>Medtr7g017100</td>
<td>78.02</td>
</tr>
</tbody>
</table>

<sup>a</sup>Protein sequences of 40 genes flanking *Dt2* (20 upstream of *Dt2* and 20 downstream of *Dt2*) in the soybean reference genome were used to search against the protein sequences of all genes annotated in the *Medicago truncatula* genome by BLASTP.