2014

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

See next page for additional authors

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.
https://digitalcommons.unl.edu/agronomyfacpub/866

Authors

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

The Plant Cell, Vol. 26: 2831–2842, July 2014, www.plantcell.org

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

Jieqing Ping,[a,1] Yunfeng Liu,[a,1] Lianjun Sun,[a,1] Meixia Zhao,[a] Yinghui Li,[b] Maoyun She,[a] Yi Sui,[a,2] Feng Lin,[a] Xiaodong Liu,[a] Zongxiang Tang,[a] Hanh Nguyen,[c] Zhixi Tian,[a,3] Lijuan Qiu,[b] Randall L. Nelson,[d] Thomas E. Clemente,[c] James E. Specht,[c] and Jianxin Ma[a,4]

[a] Department of Agronomy, Purdue University, West Lafayette, Indiana 47907
[b] Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[c] Department of Agronomy and Horticulture/Center for Plant Science Innovation, University of Nebraska, Lincoln, Nebraska 68583
[d] 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

Similar to *Arabidopsis thaliana*, the wild soybeans (*Glycine soja*) and many cultivars exhibit indeterminate stem growth specified by the shoot identity gene *Dt1*, the functional counterpart of *Arabidopsis TERMINAL FLOWER1* (*TFL1*). Mutations in *TFL1* and *Dt1* both result in the shoot apical meristem (SAM) switching from vegetative to reproductive state to initiate terminal flowering and thus produce determinate stems. A second soybean gene (*Dt2*) regulating stem growth was identified, which, in the presence of *Dt1*, produces semideterminate plants with terminal racemes similar to those observed in determinate plants. Here, we report positional cloning and characterization of *Dt2*, a dominant MADS domain factor gene classified into the *APETALA1/SQUAMOSA* (*AP1/SQUA*) subfamily that includes floral meristem (FM) identity genes *AP1*, *FUL*, and *CAL* in *Arabidopsis*. Unlike *AP1*, whose expression is limited to FMs in which the expression of *TFL1* is repressed, *Dt2* appears to repress the expression of *Dt1* in the SAMs to promote early conversion of the SAMs into reproductive inflorescences. Given that *Dt2* is not the gene most closely related to *AP1* and that semideterminacy is rarely seen in wild soybeans, *Dt2* appears to be a recent gain-of-function mutation, which has modified the genetic pathways determining the stem growth habit in soybean.

INTRODUCTION

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

---

[1] These authors contributed equally to this work.
[2] Current address: Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China.
[3] Current address: Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China.
[4] Address correspondence to maj@purdue.edu.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Jianxin Ma (maj@purdue.edu).
[C] Some figures in this article are displayed in color online but in black and white in the print edition.
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.114.126938

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 (*Dt1dt1;Dt2dt2*) produces progeny with an F_2 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 *Dt1* 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; Fridman et al., 2002) and two legume species, pigeon pea (*Cajanus cajan*) (Gupta and Kapoor, 1991), and chickpea (*Cicer arietinum*) (Hegde, 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 (*sdtsdt*) 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 F_2 progeny derived from a dihybrid [*Spsp;Sdtsdt*]). 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

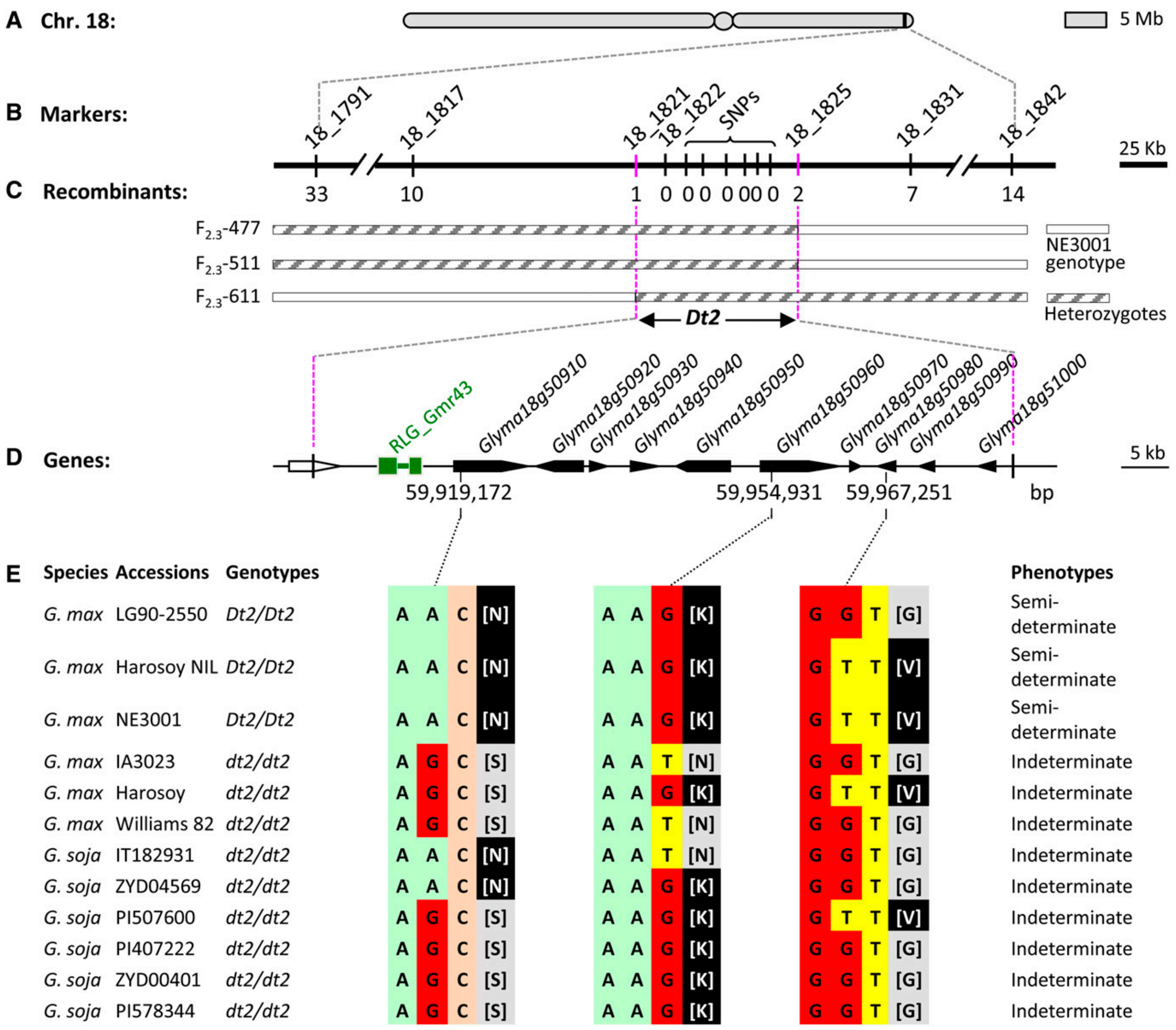
Molecular Mapping of the *Dt2* Locus to a Genomic Region Near the End of the Short Arm of Chromosome 18

To map the *Dt2* gene, a cross between a semideterminate soybean cultivar NE3001 (*Dt2Dt2*;*Dt1Dt1*) and an indeterminate soybean cultivar IA3023 (*dt2dt2*;*Dt1Dt1*) (Setiyono et al., 2007) was made to generate an F_2 population comprising 681 individual F_2 plants. Each of the F_2 plants were advanced to the $F_{2:3}$ progenies, which were then used to deduce the genotypes of individual F_2 plants. Based on high-confidence phenotyping data from the 679 $F_{2:3}$ families, 156 F_2 individuals were deduced as semideterminate homozygotes (*Dt2Dt2*), 350 F_2 individuals as semideterminate heterozygotes (*Dt2dt2*), and 173 F_2 individuals as indeterminate homozygotes (*dt2dt2*). These data confirmed the reported single-gene inheritance pattern of dominant semideterminacy versus recessive indeterminacy (3:1; χ^2 test, P = 0.77). The observed genotypic segregation pattern also fits the expected 1:2:1 ratio (χ^2 test, P = 0.47).

A previous linkage analysis with 20 F_2 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 F_2 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 F_2 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).



Species	Accessions	Genotypes													Phenotypes
G. max	LG90-2550	*Dt2/Dt2*	A	A	C	[N]	A	A	G	[K]	G	G	T	[G]	Semi-determinate
G. max	Harosoy NIL	*Dt2/Dt2*	A	A	C	[N]	A	A	G	[K]	G	T	T	[V]	Semi-determinate
G. max	NE3001	*Dt2/Dt2*	A	A	C	[N]	A	A	G	[K]	G	T	T	[V]	Semi-determinate
G. max	IA3023	*dt2/dt2*	A	G	C	[S]	A	A	T	[N]	G	G	T	[G]	Indeterminate
G. max	Harosoy	*dt2/dt2*	A	G	C	[S]	A	A	G	[K]	G	T	T	[V]	Indeterminate
G. max	Williams 82	*dt2/dt2*	A	G	C	[S]	A	A	T	[N]	G	G	T	[G]	Indeterminate
G. soja	IT182931	*dt2/dt2*	A	A	C	[N]	A	A	T	[N]	G	G	T	[G]	Indeterminate
G. soja	ZYD04569	*dt2/dt2*	A	A	C	[N]	A	A	G	[K]	G	G	T	[G]	Indeterminate
G. soja	PI507600	*dt2/dt2*	A	G	C	[S]	A	A	G	[K]	G	T	T	[V]	Indeterminate
G. soja	PI407222	*dt2/dt2*	A	G	C	[S]	A	A	G	[K]	G	G	T	[G]	Indeterminate
G. soja	ZYD00401	*dt2/dt2*	A	G	C	[S]	A	A	G	[K]	G	G	T	[G]	Indeterminate
G. soja	PI578344	*dt2/dt2*	A	G	C	[S]	A	A	G	[K]	G	G	T	[G]	Indeterminate

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 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.

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 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 noncoding 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; Fornara 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* (*AP1*) (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*.

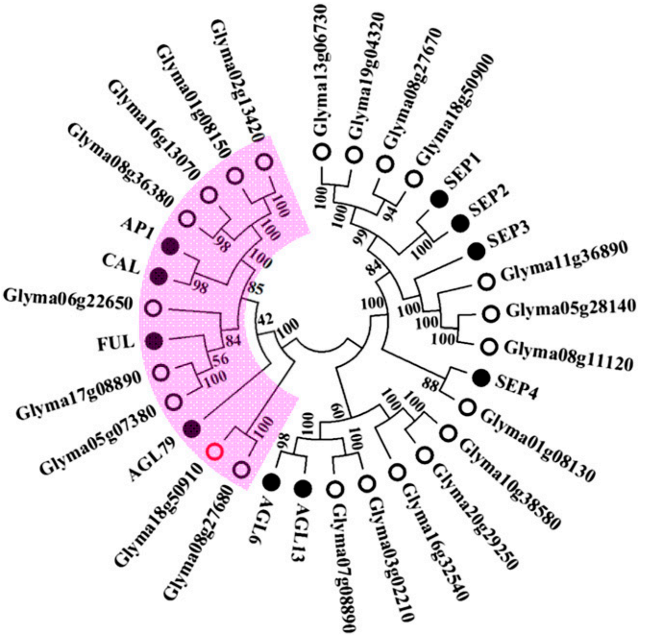


Figure 2. Phylogenetic Relationship of Closely Related Homologs of the *Dt2* Candidate Gene in Soybean and *Arabidopsis*.

The predicted full length of amino acid sequences of the genes was used to construct the neighbor-joining tree. Numbers adjacent to nodes indicating bootstrap values from the test of 1000 replicates. Pink (shaded region) includes all gene homologs identified in the *Arabidopsis* and soybean that belong to the *AP1*/*SQUA* subfamily.
[See online article for color version of this figure.]

To test this postulation, we first examined the expression patterns of the ten genes in the mapped *Dt2* region of NE3001 in various tissues and at various developmental stages before the transition of vegetative growth to reproductive growth of main stem tips by quantitative real-time PCR (qRT-PCR). It was documented that, at the V2 stage, when the first trifoliate leaflets at node 2 are fully expanded but the second trifoliate leaflets at node 3 are not yet unfolded, floral induction occurs in all meristems (apical and lateral), abruptly in the case of the determinants, less abruptly in the case of semideterminants, but not in the terminal apical meristems in indeterminate types (Wilkerson et al., 1989). As expected, the *Dt2* candidate gene *Glyma18g50910* transcripts were found to be the most abundant in apical stem tips collected at the V2 stage (Supplemental Figure 2). Subsequently, we compared the expression patterns of these 10 genes in apical stem tips at this developmental stage between NE3001 and IA3023 and found that *Glyma18g50910* exhibited considerably higher levels of expression in NE3001 than in IA3023 (Figure 3; Supplemental Figure 2). Another gene showing higher level of expression in NE3001 than in IA3023 was the F-box domain gene *Glyma18g51000*, but its expression level was relatively low and did not show obvious difference between the *Dt2* and *dt2* NILs L62-364 and Harosoy (Figure 3; Supplemental Figure 3). None of the other eight genes in the mapped *Dt2* region showed differential expression between NE3001 and IA3023 (Figure 3). These observations suggested *Glyma18g50910* as the

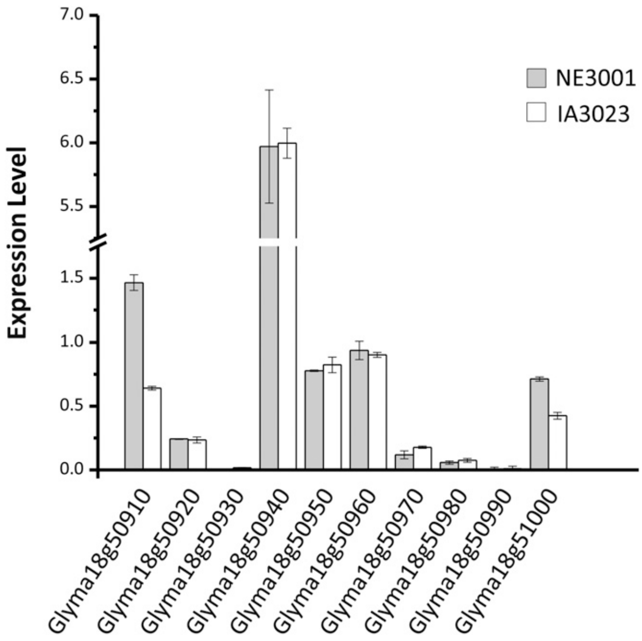


Figure 3. Expression of 10 Genes in the Mapped *Dt2* Region in Apical Stem Tips of NE3001 and IA3023 at V2 Stage Detected by qRT-PCR.

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.

only candidate for *Dt2*. Overall, the expression level of *Glyma18g50910* in NE3001 is higher than that of *dt2* in IA3023 in apical stem tips from the V0 stage (when the cotyledons at node 0 are fully extended but the unifoliate leaflets at node 1 are not yet unrolled) to the V3 stage (when the second trifoliate leaflets are fully expanded but before the third trifoliate leaflets are still unrolled). By contrast, *Dt1* was mainly expressed in main stem tips at the V0 stage (Figure 4). Photoperiod induction is known to begin at the V0 stage (Wilkerson et al., 1989), which converts all existing vegetative meristems to inflorescence meristems except the main stem apex in indeterminate *Dt1Dt1;dt2dt2* genotypes, which remains vegetative.

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 expressed 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 semideterminacy 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 Thorne (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* from NE3001, coupled with a 35S terminator (dubbed *35S*:CDS-*Dt2*). The other genetic element consisted of the *Glyma18g50910* cassette regulated 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

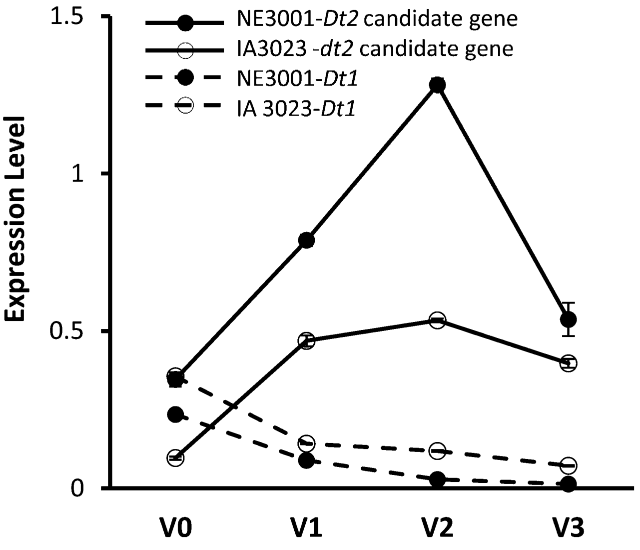


Figure 4. Expression of *Dt1* or *dt1* and the *Dt2*/*dt2* Candidate Gene *Glyma18g50910* in NE3001 and/or IA3023 Detected by qRT-PCR.

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.

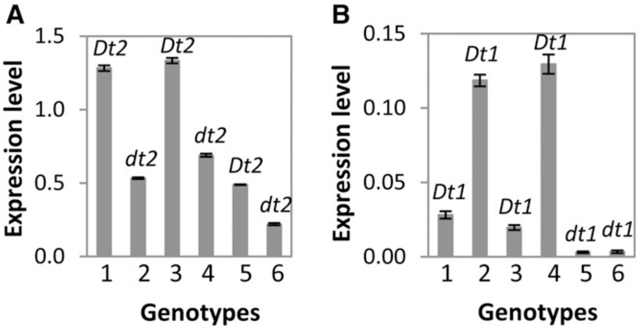


Figure 5. Expression of *Dt1*/*dt1* and the *Dt2*/*dt2* Candidate Gene in Apical Stem Tips of Different Genotypes Detected by qRT-PCR.

(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 *Dt2* 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 *Dt2* in NE3001 showed similar degrees of stem termination (Figure 6E; Supplemental Figure 5). The expression levels of the transgene measured by qRT-PCR were negatively associated with both the node numbers and the heights of the main stems (Supplemental Table 5). Moreover, these transgenic plants flowered earlier than Thorne, the recipient line of the transgene (Figures 6A and 6B), similar to that observed between the semideterminate *Dt2* Harosoy NIL and the indeterminate Harosoy NIL. These observations, together with other evidence described above, indicate that *Glyma18g50910* in NE3001 was the *Dt2* 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 *Dt2* 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 *Dt2* 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).

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 *Dt2* locus that led to differential allelic expression responsible for the phenotypic variation in soybean stem growth habit, we compared genomic sequences of the *Dt2* locus and its flanking intergenic 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 *G. 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 were found in the 2.5-kb sequence upstream of the CDS of the *Dt2* locus, 23 (62%) in the first intron of the *Dt2* locus, one in the second intron of the *Dt2* locus, and one in the 1.5-kb downstream of the CDS of the *Dt2* 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.

DISCUSSION

Evolutionary Relationship and Novelty of the *Dt2* Gene 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 *Arabidopsis* 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 *FUL*, *AP1*, and *CAL* regions exhibited syntenic relationships with the *Dt2* region (Shu et al., 2013). Indeed, *Dt2* was not the gene most closely related to these three *Arabidopsis* genes based on their phylogeny (Figure 2; Supplemental Figure 1 and Supplemental Data Set 1). Hence, it remains unclear which *Arabidopsis* gene is the functional counterpart of *Dt2*.

If the established phylogenetic relationships of the MADS domain gene homologs within and between the *Arabidopsis* and soybean genomes did reflect the orders and timeframes within which these genes were diverged and generated, then *Dt2* should

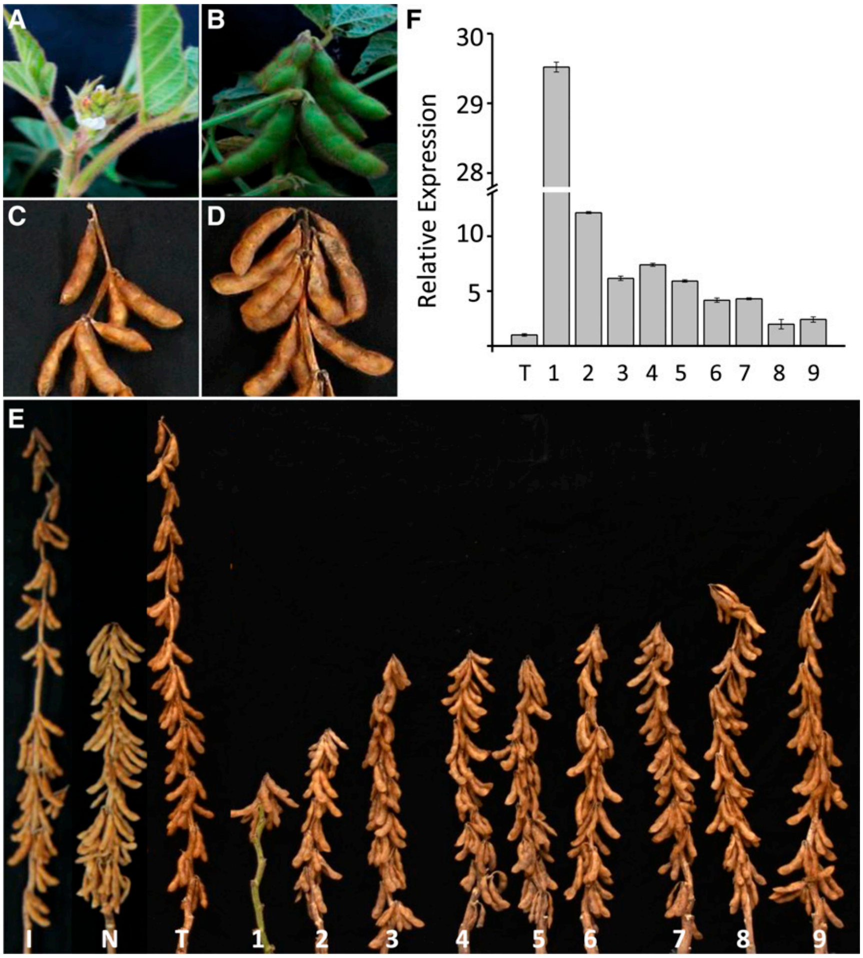


Figure 6. Overexpression of Transgene *Dt2* in an Indeterminate Cultivar Resulting in Phenotypic Changes from Indeterminate to Semideterminate Types.

(A) Apical stem tip of immature Thorne showing indeterminate stem growth and late flowering.

(B) Apical stem tip of an immature Thorne *35S*:CDS-*Dt2* transgenic plant showing semideterminate stem growth and early flowering.

(C) Apical stem tip of mature Thorne showing indeterminate stem growth.

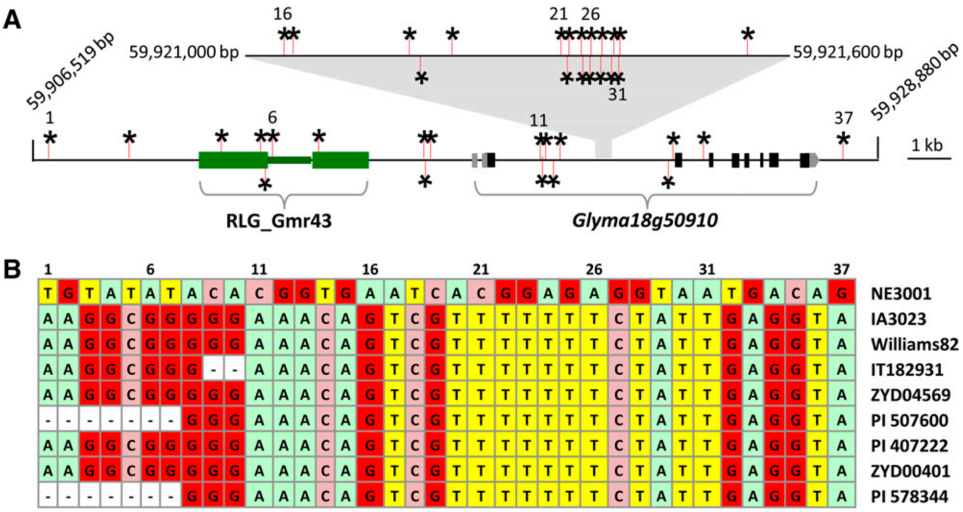
(D) Apical stem tip of a mature Thorne *35S*:CDS-*Dt2* transgenic plant showing semi-indeterminate stem growth.

(E) Photograph of IA3012 (I), NE3001 (N), Thorne (T), and nine T3 Thorne *35S*: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 *35S*: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.

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



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
NE3001	T	G	T	A	T	A	T	A	C	A	C	G	G	T	G	A	A	T	C	A	C	G	G	A	G	A	G	G	T	A	A	T	G	A	C	A	G
IA3023	A	A	G	G	C	G	G	G	G	G	A	A	A	C	A	G	T	C	G	T	T	T	T	T	T	T	C	T	A	T	T	G	A	G	G	T	A
Williams82	A	A	G	G	C	G	G	G	G	G	A	A	A	C	A	G	T	C	G	T	T	T	T	T	T	T	C	T	A	T	T	G	A	G	G	T	A
IT182931	A	A	G	G	C	G	G	G	-	-	A	A	A	C	A	G	T	C	G	T	T	T	T	T	T	T	C	T	A	T	T	G	A	G	G	T	A
ZYD04569	A	A	G	G	C	G	G	G	G	G	A	A	A	C	A	G	T	C	G	T	T	T	T	T	T	T	C	T	A	T	T	G	A	G	G	T	A
PI 507600	-	-	-	-	-	-	-	G	G	G	A	A	A	C	A	G	T	C	G	T	T	T	T	T	T	T	C	T	A	T	T	G	A	G	G	T	A
PI 407222	A	A	G	G	C	G	G	G	G	G	A	A	A	C	A	G	T	C	G	T	T	T	T	T	T	T	C	T	A	T	T	G	A	G	G	T	A
ZYD00401	A	A	G	G	C	G	G	G	G	G	A	A	A	C	A	G	T	C	G	T	T	T	T	T	T	T	C	T	A	T	T	G	A	G	G	T	A
PI 578344	-	-	-	-	-	-	-	G	G	G	A	A	A	C	A	G	T	C	G	T	T	T	T	T	T	T	C	T	A	T	T	G	A	G	G	T	A

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.

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;Dt1Dt1* 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 gametes), 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* (*Det*) and *VEGETATIVE1* (*Veg1*), regulate the stem growth habit in pea (*Pisum sativum*). *Det* 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, *Veg1Veg1* 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 (I2) 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 *Dt1dt1* 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 pea could be the explained by strong, complete dominance of *Det* over *det* (Singer et al., 1999), which may lead *Veg1* to be hypostatic 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 inflorescence meristems of soybean and pea may help to elucidate 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; Bomblies and Doebley, 2005; Rao et al., 2008). A more recent study revealed that the intron sequences, particularly the first introns of *AP1* and *FUL*, were bound by the microRNA-targeted transcription factor SQUAMOSA PROMOTER 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 accepts the thesis that the regulation of *Dt2* in the semideterminate soybean varieties did need such *cis*-regulatory elements, particularly 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 overexpression 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%), arranged in long strings of Ts, As, ATs, 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. Finer-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 surrounding *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 $F_{2:3}$ 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 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.

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 (PI547703). Seed of these four lines were obtained from the USDA Soybean Germplasm Collection. Transgenic soybean lines were phenotyped and advanced 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 sequencing 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 F_2 plants is not always possible. We thus advanced the 681 F_2 population to the F3 generation; subsequently, ~50 F_3 plants grown from each F_2 plant were scored for abruptness of stem termination in a field nursery located in Lincoln, NE, where NE3001 was developed. The stem termination phenotype of each F_3 plant in each $F_{2:3}$ progeny row was examined to determine if the F_3 plants in a given row were all semideterminate, all indeterminate, or segregating in a ratio of 3:1 for semideterminate to indeterminate. This $F_{2:3}$ progeny phenotyping resulted in an accurate deduction of a respective of *dt2dt2*, *Dt2Dt2*, and *Dt2dt2* genotype for nearly all of the F_2 plant progenitors. An equal amount of leaf tissue was collected from the ~15 to 20 F_3 plants tracing to each F_2 plant and was used for molecular marker assays that provided molecular genotypes for each F_2 progenitor plant. Of the 681 F_2 individuals, two produced too few F_3 plants for a reliable inference of the F_2 phenotypes and thus were excluded in linkage analysis. SSR markers for mapping were selected based on their physical locations on chromosome 18 (Song et al., 2010). SNP markers were designed based on genic sequence variation in the mapped regions between two parental lines. Genotyping of recombinants with SNP markers was performed either by direct sequencing of PCR fragments or using the cleaved amplified polymorphic sequence markers (Supplemental Table 1).

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 *Arabidopsis* that include all genes belonging to the *AP1*/*SQUA* subfamily in

Arabidopsis, i.e., *AP1*, *CAL*, *FUL*, and *AGL79* (Rijpkema 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 *Nco*I/*Bam*HI and with *Bam*HI/*Xba*I, 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 *Nco*I/*Xba*I 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 *Hin*dIII/*Xho*I and *Xba*I/*Hin*dIII respectively, to isolate the 2.5- and 1.5-kb fragments. The assembled and verified 35S:CDS-*Dt2* construct (designated pPTN1171) was digested with *Xho*I/*Xba*I to isolate the CDS of the gene. Simultaneously, the pPTN200 vector was digested with *Hin*dIII 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 Throne 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, Beck’s Hybrids, ConAgraFoods, Dow AgroSciences, Indiana Crop Improvement Association, and Pioneer Hi-Bred International.

AUTHOR CONTRIBUTIONS

J.E.S., T.E.C., and J.M. designed the research. J.P., Y.L., L.S., M.Z., Y.L., M.S., Y.S., F.L., X.L., H.N., and Z.T. performed the research. J.P., Y.L., L.S., L.Q., R.L.N., T.E.C., J.E.S., and J.M. analyzed the data. R.L.N., T.E.C., J.E.S., and J.M. wrote the article.

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

REFERENCES

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815.

Berbel, A., Ferrándiz, C., Hecht, V., Dalmais, M., Lund, O.S., Sussmilch, F.C., Taylor, S.A., Bendahmane, A., Ellis, T.H., Beltrán, J.P., Weller, J.L., and Madueño, F. (2012). VEGETATIVE1 is essential for development of the compound inflorescence in pea. Nat. Commun. **3:** 797.

Bernard, R.L. (1972). Two genes affecting stem termination in soybeans. Crop Sci. **12:** 235–239.

Bomblies, K., and Doebley, J.F. (2005). Molecular evolution of *FLORICAULA/LEAFY* orthologs in the *Andropogoneae* (Poaceae). Mol. Biol. Evol. **22:** 1082–1094.

Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., and Coen, E. (1997). Inflorescence commitment and architecture in *Arabidopsis*. Science **275:** 80–83.

Chang, J.F., Green, D.E., and Shibles, R. (1982). Yield and agronomic performance of semin-determinate and indeterminate soybean stem types. Crop Sci. **22:** 97–101.

Chi, Y., Huang, F., Liu, H., Yang, S., and Yu, D. (2011). An *APETALA1*-like gene of soybean regulates flowering time and specifies floral organs. J. Plant Physiol. **168:** 2251–2259.

Clemente, T., LaValle, B.J., Howe, A.R., Ward, D.C., Rozman, R.J., Hunter, P.E., Broyles, D.L., Kasten, D.S., and Hinchee, M.A. (2000). Progeny analysis of glyphosate selected transgenic soybeans derived from Agrobacterium-mediated transformation. Crop Sci. **40:** 797–803.

Elkind, Y., Gurnick, A., and Kedar, N. (1991). Genetics of semideterminate growth habit in tomato. HortScience **26:** 1074–1075.

Fornara, F., de Montaigu, A., and Coupland, G. (2010). SnapShot: Control of flowering in Arabidopsis. Cell **141:** 550, e1–e2.

Foucher, F., Morin, J., Courtiade, J., Cadioux, S., Ellis, N., Banfield, M. J., and Rameau, C. (2003). DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea. Plant Cell **15:** 2742–2754.

Fridman, E., Liu, Y.S., Carmel-Goren, L., Gur, A., Shoresh, M., Pleban, T., Eshed, Y., and Zamir, D. (2002). Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol. Genet. Genomics **266:** 821–826.

Gu, Q., Ferrándiz, C., Yanofsky, M.F., and Martienssen, R. (1998). The *FRUITFULL* MADS-box gene mediates cell differentiation during Arabidopsis fruit development. Development **125:** 1509–1517.

Gupta, S.C., and Kapoor, R.K. (1991). Inheritance of growth habit in pigeonpea. Crop Sci. **31:** 1456–1459.

Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1994). Regulation of the Arabidopsis floral homeotic gene APETALA1. Cell **76:** 131–143.

Heatherly, L.G., and Smith, J.R. (2004). Effect of soybean stem growth habit on height and node number after beginning bloom in the midsouthern USA. Crop Sci. **44:** 1855–1858.

Heatherly, L.G., and Elmore, R.W. (2004). Managing inputs for peak production. In Soybeans: Improvement, Production, and Uses, 3rd ed, Agronomy Monograph 16, H.R. Boerma and J.E. Specht, eds (Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America), pp. 451–536.

Hecht, V., Foucher, F., Ferrándiz, C., Macknight, R., Navarro, C., Morin, J., Vardy, M.E., Ellis, N., Beltrán, J.P., Rameau, C., and Weller, J.L. (2005). Conservation of Arabidopsis flowering genes in model legumes. Plant Physiol. **137:** 1420–1434.

Hegde, V.S. (2011). Morphology and genetics of a new found determinate genotype in chickpea. Euphytica **182:** 35–42.

Hyten, D.L., Song, Q., Zhu, Y., Choi, I.Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C., and Cregan, P.B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. USA **103:** 16666–16671.

Kempin, S.A., Savidge, B., and Yanofsky, M.F. (1995). Molecular basis of the cauliflower phenotype in Arabidopsis. Science **267:** 522–525.

Kim, M.Y., et al. (2010). Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proc. Natl. Acad. Sci. USA **107:** 22032–22037.

Li, Y.-H., et al. (2014). A pan-genome of wild soybean provides insights into evolution, diversity and domestication of soybean. Nat. Biotechnol., in press.

Libault, M., Thibivilliers, S., Bilgin, D.D., Radwan, O., Benitez, M., Clough, S.J., and Stacey, G. (2008). Identification of four soybean reference genes for gene expression normalization. Plant Genome **1:** 44–54.

Liljegren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S., and Yanofsky, M.F. (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. Plant Cell **11:** 1007–1018.

Liu, B., Fujita, T., Yan, Z.H., Sakamoto, S., Xu, D., and Abe, J. (2007). QTL mapping of domestication-related traits in soybean (*Glycine max*). Ann. Bot. (Lond.) **100:** 1027–1038.

Liu, B., et al. (2010). The soybean stem growth habit gene Dt1 is an ortholog of Arabidopsis TERMINAL FLOWER1. Plant Physiol. **153:** 198–210.

Liu, C., Teo, Z.W., Bi, Y., Song, S., Xi, W., Yang, X., Yin, Z., and Yu, H. (2013). A conserved genetic pathway determines inflorescence architecture in *Arabidopsis* and rice. Dev. Cell **24:** 612–622.

Liu, C., Thong, Z., and Yu, H. (2009). Coming into bloom: the specification of floral meristems. Development **136:** 3379–3391.

Martinez-Castilla, L.P., and Alvarez-Buylla, E.R. (2003). Adaptive evolution in the Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny. Proc. Natl. Acad. Sci. USA **100:** 13407–13412.

McBlain, B.A., Fioritto, R.J., St. Martin, S.K., Calip-DuBois, A., Schmitthenner, A.F., Cooper, R.L., and Martin, R.J. (1993). Registration of 'Thorne' soybean. Crop Sci. **33:** 1406.

Muehlbauer, G.J., Specht, J.E., Staswick, P.E., Graef, G.L., and Thomascompton, M.A. (1989). Application of the near-isogenic line gene-mapping technique to isozyme markers. Crop Sci. **29:** 1548–1553.

Nagata, T. (1950). Studies on the Characteristics of Soybean Varieties. (Tokyo: Japan Soybean Association).

Peng, J., et al. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. Nature **400:** 256–261.

Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganal, M., Zamir, D., and Lifschitz, E. (1998). The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. Development **125:** 1979–1989.

Prasad, K., Kushalappa, K., and Vijayraghavan, U. (2003). Mechanism underlying regulated expression of RFL, a conserved transcription factor, in the developing rice inflorescence. Mech. Dev. **120:** 491–502.

Rao, N.N., Prasad, K., Kumar, P.R., and Vijayraghavan, U. (2008). Distinct regulatory role for RFL, the rice LFY homolog, in determining flowering time and plant architecture. Proc. Natl. Acad. Sci. USA **105:** 3646–3651.

Ratcliffe, O.J., Bradley, D.J., and Coen, E.S. (1999). Separation of shoot and floral identity in Arabidopsis. Development **126:** 1109–1120.

Rijpkema, A.S., Gerats, T., and Vandenbussche, M. (2007). Evolutionary complexity of MADS complexes. Curr. Opin. Plant Biol. **10:** 32–38.

Schmutz, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature **463:** 178–183.

Setiyono, T.D., Weiss, A., Specht, J., Bastidas, A.M., Cassman, K.G., and Dobermann, A. (2007). Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. Field Crops Res. **100:** 257–271.

Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z., and Kong, H. (2007). Patterns of gene duplication and functional diversification during the evolution of the *AP1/SQUA* subfamily of plant MADS-box genes. Mol. Phylogenet. Evol. **44:** 26–41.

Shannon, S., and Meeks-Wagner, D.R. (1991). A mutation in the Arabidopsis *TFL1* gene affects inflorescence meristem development. Plant Cell **3:** 877–892.

Shu, Y., Yu, D., Wang, D., Guo, D., and Guo, C. (2013). Genome-wide survey and expression analysis of the MADS-box gene family in soybean. Mol. Biol. Rep. **40:** 3901–3911.

Singer, S., Sollinger, J., Maki, S., Fishbach, J., Short, B., Reinke, C., Fick, J., Cox, L., McCall, A., and Mullen, H. (1999). Inflorescence architecture: a developmental genetics approach. Bot. Rev. **65:** 385–410.

Smaczniak, C., Immink, R.G., Angenent, G.C., and Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development **139:** 3081–3098.

Song, Q.J., Jia, G.F., Zhu, Y.L., Grant, D., Nelson, R.T., Hwang, E.Y., Hyten, D.L., and Cregan, P.B. (2010). Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci. **50:** 1950–1960.

Specht, J.E., Chase, K., Macrander, M., Graef, G.L., Chung, J., Markwell, J.P., Germann, M., Orf, J.H., and Lark, K.G. (2001). Soybean response to water: A QTL analysis of drought tolerance. Crop Sci. **41:** 493–509.

Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J.E., Nelson, R.L., McClean, P.E., Qiu, L., and Ma, J. (2010). Artificial selection for determinate growth habit in soybean. Proc. Natl. Acad. Sci. USA **107:** 8563–8568.

Ting, C.L. (1946). Genetic studies on the wild and cultivated soybeans. J. Am. Soc. Agron. **38:** 381–398.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., and Meyerowitz, E.M. (1992). *LEAFY* controls floral meristem identity in *Arabidopsis.* Cell **69:** 843–859.

Wilkerson, G., Jones, J., Boote, K., and Buol, G. (1989). Photoperiodically sensitive interval in time to flower of soybean. Crop Sci. **29:** 721–726.

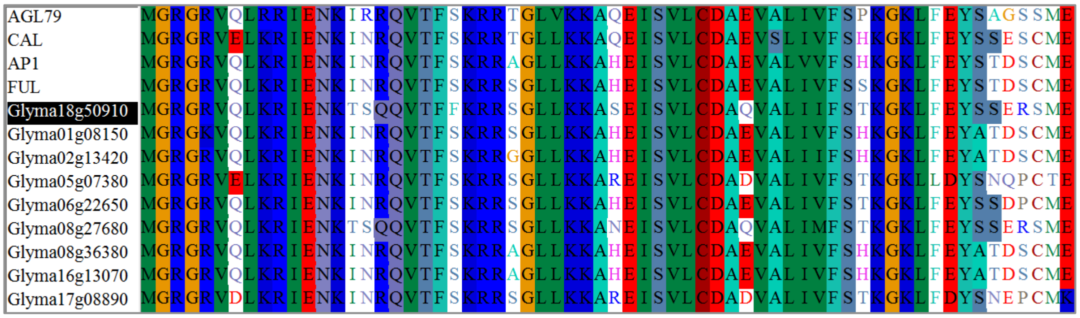
Williams, L. (1950). Structure and genetic characteristics of the soybean. In Soybean and Soybean Products, K.S. Markley, ed (New York: Interscience Publishers), pp. 111–134.

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

Yamaguchi, A., Wu, M.F., Yang, L., Wu, G., Poethig, R.S., and Wagner, D. (2009). The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1. Dev. Cell **17:** 268–278.

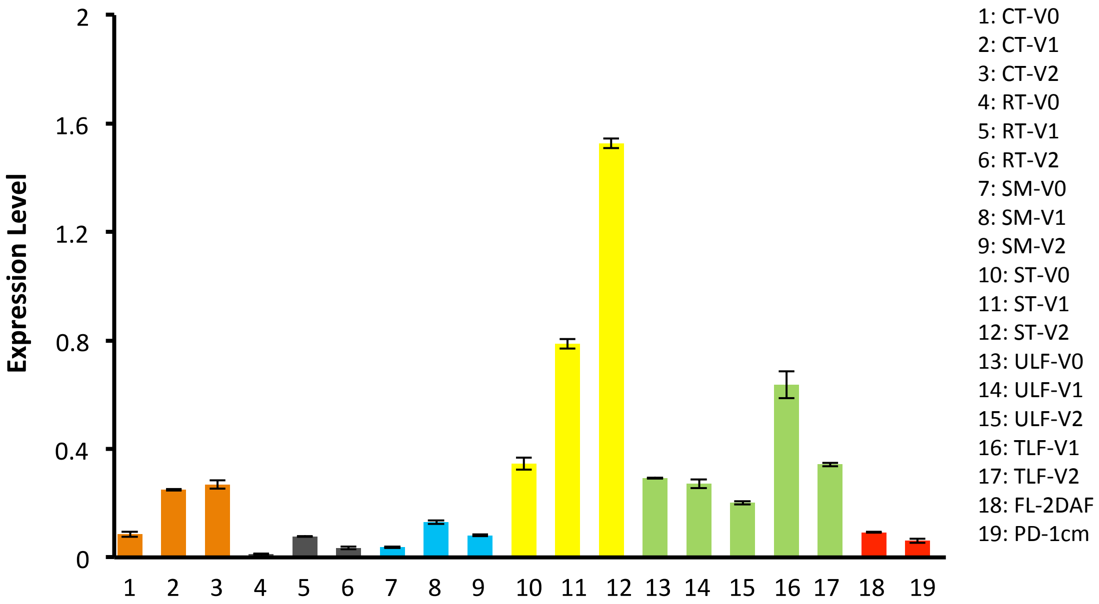
Yant, L., Mathieu, J., and Schmid, M. (2009). Just say no: floral repressors help Arabidopsis bide the time. Curr. Opin. Plant Biol. **12:** 580–586.

Zhu, H.Y., Choi, H.K., Cook, D.R., and Shoemaker, R.C. (2005). Bridging model and crop legumes through comparative genomics. Plant Physiol. **137:** 1189–1196.

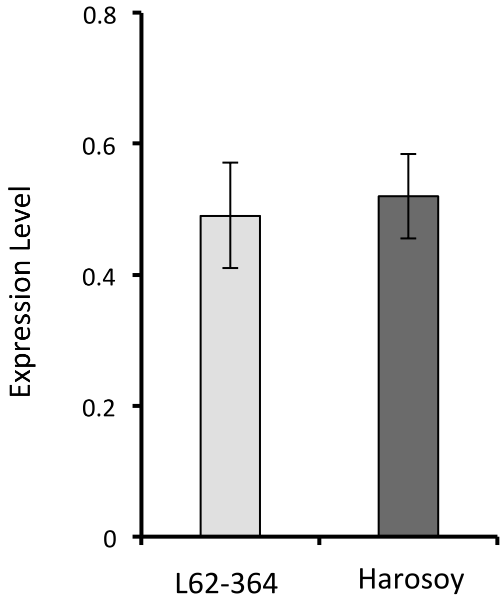


AGL79	MGRGRVQLRRIENKIRRQVTFSKRRTGLVKKAQEISVLCDAEVALIVFSPKGKLFEYSAGSSME
CAL	MGRGRVELKRIENKINRQVTFSKRRTGLLKKAQEISVLCDAEVSLIVFSHKGKLFEYSSESCME
AP1	MGRGRVQLKRIENKINRQVTFSKRRAGLLKKAHEISVLCDAEVALVVFSHKGKLFEYSTDSCME
FUL	MGRGRVQLKRIENKINRQVTFSKRRSGLLKKAHEISVLCDAEVALIVFSSKGKLFEYSTDSCME
Glyma18g50910	MGRGRVQLKRIENKTSQQVTFFKRRSGLLKKASEISVLCDAQVALIIFSTKGKLFEYSSERSME
Glyma01g08150	MGRGKVQLKRIENKINRQVTFSKRRSGLLKKAHEISVLCDAEVALIVFSHKGKLFEYATDSCME
Glyma02g13420	MGRGRVQLKRIENKINRQVTFSKRRGGLLKKAHEISVLCDAEVALIIFSHKGKLFEYATDSCME
Glyma05g07380	MGRGRVELKRIENKINRQVTFSKRRSGLLKKAREISVLCDADVALIVFSTKGKLLDYSNQPCTE
Glyma06g22650	MGRGRVQLKRIENKINRQVTFSKRRSGLLKKAHEISVLCDAEVALIVFSTKGKLFEYSSDPCME
Glyma08g27680	MGRGRVQLKRIENKTSQQVTFSKRRSGLLKKANEISVLCDAQVALIMFSTKGKLFEYSSERSME
Glyma08g36380	MGRGRVQLKRIENKINRQVTFSKRRAGLLKKAHEISVLCDAEVALIVFSHKGKLFEYATDSCME
Glyma16g13070	MGRGRVQLKRIENKINRQVTFSKRRAGLLKKAHEISVLCDAEVALIVFSHKGKLFEYATDSCME
Glyma17g08890	MGRGRVDLKRIENKINRQVTFSKRRSGLLKKAREISVLCDADVALIVFSTKGKLFDYSNEPCMK

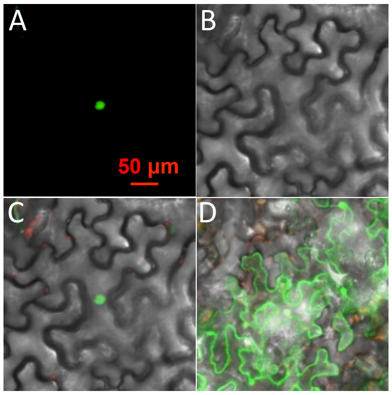
Supplemental Figure 1. Alignment of predicted MADS-box domains of the *Dt2* candidate gene homologs in soybean and *Arabidopsis.*



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 unifolioate 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.

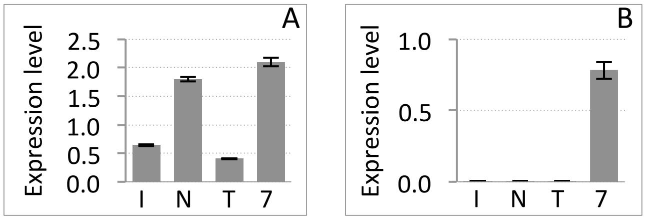


Supplemental Figure 3. Expression of *Glyma18g51000* in apical stem tips of NILs L62-364 and Harosoy at V2 stage detected by qRT-PCR. The y-axis indicates the expression levels of the gene relative to expression of Cos4. Expression levels were shown as means ± standard errors of the means from four replicates.



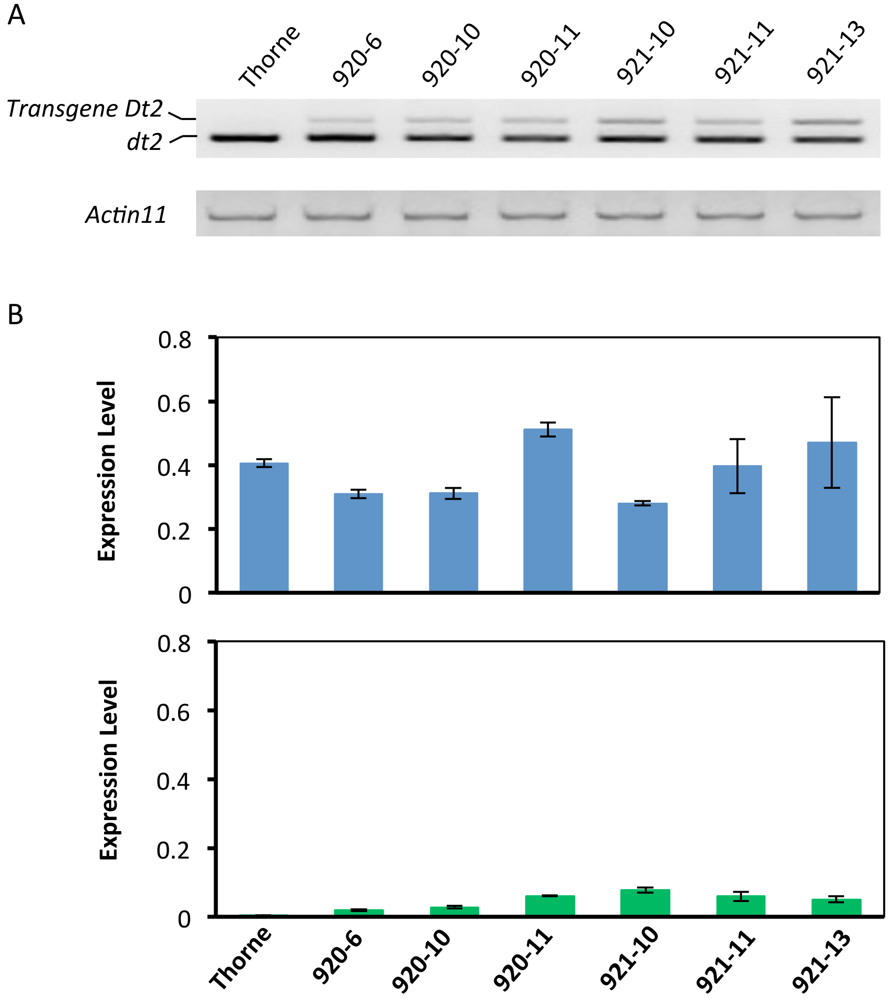
Supplemental Figure 4. Subcellular localization of the Dt2 protein in tobacco epidermal cells.

(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 in . 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.

Marker	Chromosome ID	Forward primer sequences (5' to 3')	Reverse primer sequences (5' to 3')	Type of markers
SSR_18_1791	Gm18	TGACCAGTCAATTGTTCATTCTTT	TTTACTCAACCATCTCCGCA	SSR
SSR_18_1807	Gm18	TCATTCTGTAAAATGAGTTGTGTATTC	TTATTTTGCTTTCAAACTTACAATTC	SSR
SSR_18_1817	Gm18	GTGAGGCCATCAATCACCTT	CGCAAGAAGAAAAGAAAAGGAA	SSR
SSR_18_1821	Gm18	GGTGCCTTTAATTTCTTTGGA	ATTCACCAGATCATGTGCCA	SSR
SSR_18_1822	Gm18	AATTTGATGCACTTGATAACGA	TGACAAACACAAGAACTCACACA	SSR
SSR_18_1825	Gm18	GAATCCACCATCACCAAACC	CAATGGCAACCCAGTAAGGT	SSR
SSR_18_1831	Gm18	TGTTTTTGTTAAATCTTTTGTTTGG	TGTGTATGTTTGTGTGTGCACTT	SSR
SSR_18_1833	Gm18	GGCTATTGCAACATTCGGTT	GAGGAAAGTGTTCATTGCCG	SSR
SSR_18_1838	Gm18	TTCTATATTCAAAACTGAACTGAACTG	AACTTATTATAACGCAATTTTATGCTT	SSR
SSR_18_1842	Gm18	TGAAATGGAGGAGAAAATGGA	GTCCGGGGAAACTGAACC	SSR
SSR_18_1846	Gm18	CTTTTAACGATTGGGTTGGG	CTTCGGCCTTAGACTTTTCG	SSR
SSR_18_1854	Gm18	GCCACCTCTACACCAACACA	TGACCAACAATGGCTTTCAA	SSR
SSR_18_1858	Gm18	TAGCTTTATAATGAGTGTGATAGAT	GTATGCAAGGGATTAATTAAG	SSR
SSR_18_1864	Gm18	TGAATGATATATGTTTTGCGAAGA	CAATAGAGCCGGATGGATGT	SSR

SSR_18_1890	Gm18	TGTTAGTGTACGCGTTACAAAATATAA	AAAGTGCATGTACATTAGTGAATTTTA	SSR
SSR_18_1926	Gm18	TTTGGAGATTACTGACAAAAGAGA	TTTTGTCCCTTAAAATAACTTCAAC	SSR
SNP1	Gm18	CTCTGTAATATGCTCAGAGTC	GTAGGTGGCAAGAAACCCCCC	SNP
SNP2	Gm18	CAGACATAATCTATGAACAAG	GCAAACAACCTAAAGGATCACAG	SNP
SNP3	Gm18	CCATGTACATTAGTATTCAGTAG	AGCAGCTCTGAAATTAGCC	SNP
SNP4	Gm18	GTGTTTATATTAGTTCTTTACCC	ACCATGTATAAATGATAC	SNP
SNP5	Gm18	CAAGCACTATAGCCTTTAGTC	AGAAGCATTCTTTGAAGAGGAAAC	SNP
SNP6	Gm18	TGAAGCGGATCGAGAACAAAACA	AATGATGAACGAGTAGGAACCT	CAPS

Supplemental Table 2. Genes in the defined *Dt2* region according to the Williams 82 reference genome.

Genes	**Annotation**
Glyma18g50910	MADS box transcription factor
Glyma18g50920	Uncharacterized conserved protein, contains RCC1 domain
Glyma18g50930	MEKK and related serine/threonine protein kinases
Glyma18g50940	DSBA-like thioredoxin domain
Glyma18g50950	Ring finger protein 24-related
Glyma18g50960	No functional annotations
Glyma18g50970	Pollen proteins Ole e I like
Glyma18g50980	Glucosyl Transferases
Glyma18g50990	F-box domain
Glyma18g51000	F-box domain

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]

Accessions	**Genotype**	**Phenotype**	**Type**	**Region**	**Country**	**Maturity Group**	**SNP**
PI483464A	*Dt1*	indeterminate	*Glycine soja*	Ningxia	China	III	G
PI 407301	*Dt1*	indeterminate	*Glycine soja*	Jiangsu	China	V	G
PI 483465	*Dt1*	indeterminate	*Glycine soja*	Shaanxi	China	V	G
PI468400A	*Dt1*	indeterminate	*Glycine soja*	Ningxia	China	IV	G
PI 407131	*Dt1*	indeterminate	*Glycine soja*	Kumamoto	Japan	VI	G
PI 447004	*Dt1*	indeterminate	*Glycine soja*	Jilin	China	III	G
PI 366120	*Dt1*	indeterminate	*Glycine soja*	Akita	Japan	IV	G
PI 407170	*Dt1*	indeterminate	*Glycine soja*	Kyonggi	Korea, South	V	G
PI 549046	*Dt1*	indeterminate	*Glycine soja*	Shaanxi	China	III	G
PI 407140	*Dt1*	indeterminate	*Glycine soja*	Kumamoto	Japan	VII	A
PI326582A	*Dt1*	indeterminate	*Glycine soja*	Primorye	Russia	II	A
PI 464935	*Dt1*	indeterminate	*Glycine soja*	Jiangsu	China	VI	A
PI 468916	*Dt1*	indeterminate	*Glycine soja*	Liaoning	China	III	A
PI339871A	*Dt1*	indeterminate	*Glycine soja*	Cheju	Korea	V	A
PI 458538	*Dt1*	indeterminate	*Glycine soja*	Heilongjiang	China	OOO	A
PI597459D	*Dt1*	indeterminate	*Glycine soja*	Shandong	China	III	A
PI 393551	*Dt1*	indeterminate	*Glycine soja*	Taiwan	China	X	A
PI597461A	*Dt1*	indeterminate	*Glycine soja*	Shandong	China	IV	A

PI 562559	*Dt1*	indeterminate	*Glycine soja*	Cholla Puk	Korea, South	V	A
PI 407282	*Dt1*	indeterminate	*Glycine soja*	Cheju	Korea, South	VI	A
PI 548603	*Dt1*	indeterminate	N. Am. Ancestor	Indiana	USA	IV	G
PI 548488	*Dt1*	indeterminate	N. Am. Ancestor	Missouri	USA	V	G
PI 548311	*Dt1*	indeterminate	N. Am. Ancestor	Ontario	Canada	O	G
PI 548379	*Dt1*	indeterminate	N. Am. Ancestor	Heilongjiang	China	O	G
PI 548298	*Dt1*	indeterminate	N. Am. Ancestor	Unknown	China	III	G
PI 548318	*Dt1*	indeterminate	N. Am. Ancestor	Jilin	China	III	G
PI 548348	*Dt1*	indeterminate	N. Am. Ancestor	Unknown	China	III	G
PI 548362	*Dt1*	indeterminate	N. Am. Ancestor	Unknown	Unknown	III	G
PI 548391	*Dt1*	indeterminate	N. Am. Ancestor	Liaoning	China	II	A
FC 33243	*Dt1*	indeterminate	N. Am. Ancestor	Unknown	Unknown	IV	A
PI 548406	*Dt1*	semideterminate	N. Am. Ancestor	Jilin	China	II	G
PI 548382	*dt1*	semideterminate	N. Am. Ancestor	Liaoning	China	OO	G
PI 548485	*dt1*	determinate	N. Am. Ancestor	Jiangsu	China	VII	A
PI 548477	*dt1*	determinate	N. Am. Ancestor	Tennessee	USA	VI	A
PI 548657	*dt1*	determinate	N. Am. Ancestor	North Carolina	USA	VII	G
PI 548445	*dt1*	determinate	N. Am. Ancestor	Jiangsu	China	VII	G
PI 548456	*dt1*	determinate	N. Am. Ancestor	Pyongyang	Korea, North	VI	A

---

[a] Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. (2006). Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA. 103:16666-16671.

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

No. in Fig. 3	**Transformation event**	**No. of positive T2 plants planted in greenhouse**	**No. of semi-determinate T3 plants in the field**	**No. of indeterminate T3 plants in the field**
1	917-70	2	4	1
2	917-49	2	13	1
3	913-15	4	19	8
4	917-46	4	17	3
5	917-56	1	6	2
6	917-55	1	3	1
7	917-66	2	12	3
8	917-24	1	5	2
9	917-65	1	6	3

Supplemental Table 5. Correlation between expression level of the transgenes and phenotypic variation among transgenic plants

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

[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*

		BLASTP		
Query genes in soybean[a]	Putative orthologs in *Medicago*	Identity (%)	Alignment length	Expect value
Glyma18g50900	*Medtr7g016600*	82.03	256	4.00E-118
Glyma18g50910 (Dt2)	*Medtr7g016630 (*Mt-*FULc)*	72.59	259	1.00E-97
Glyma18g50920	*Medtr7g016640*	88.45	476	0
Glyma18g50940	*Medtr7g016650*	85.38	212	7.00E-108
Glyma18g50950	*Medtr7g016840*	76.99	226	3.00E-102
Glyma18g50960	*Medtr7g016900*	82.78	790	0
Glyma18g50970	*Medtr7g016950*	63.46	301	1.00E-75
Glyma18g51040	*Medtr7g016960*	77.88	660	0
Glyma18g51050	*Medtr7g016970*	70.07	441	0
Glyma18g51060	*Medtr7g017100*	78.02	2384	0

[a] 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