Artificial selection for determinate growth habit in soybean

Zhixi Tianb,1, Xiaobo Wangb,1, Rian Lee,6 Yinghui Li,3 James E. Specht4, Randall L. Nelson6, Phillip E. McCleanb,2, Lijuan Qui,2 and Jianxin Maa,2

*Department of Agronomy, Purdue University, West Lafayette, IN 47907; †National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China; ‡Department of Plant Sciences, and Genomics and Bioinformatics Program, North Dakota State University, Fargo, ND 58108; ‡Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583; and *Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, US Department of Agriculture–Agricultural Research Service, and Department of Crop Sciences, University of Illinois, Urbana, IL 61801

Edited* by Jeffrey L. Bennetzen, University of Georgia, Athens, GA, and approved April 2, 2010 (received for review January 6, 2010)

Determinacy is an agronomically important trait associated with the domestication in soybean (Glycine max). Most soybean cultivars are classifiable into indeterminate and determinate growth habit, whereas Glycine soja, the wild progenitor of soybean, is indeterminate. Indeterminate (Dt1/Dt1) and determinate (dt1/dt1) genotypes, when mated, produce progeny that segregate in a monogenic pattern. Here, we show evidence that Dt1 is a homolog (designated as GmTFL1) of Arabidopsis terminal flower 1 (TFL1), a regulatory gene encoding a signaling protein of shoot meristems. The transition from indeterminate to determinate phenotypes in soybean is associated with independent human selections of four distinct single-nucleotide substitutions in the GmTFL1 gene, each of which led to a single amino acid change. Genetic diversity of a minicore collection of Chinese soybean landraces assessed by simple sequence repeat (SSR) markers and allelic variation at the GmTFL1 locus suggest that human selection for determinacy took place at early stages of landrace radiation. The GmTFL1 allele introduced into a determinate-type (tf1/tf1) Arabidopsis mutants fully restored the wild-type (TFL1/TFL1) phenotype, but the GmTFL1 allele in tf1/tf1 mutants did not result in apparent phenotypic change. These observations indicate that GmTFL1 complements the functions of TFL1 in Arabidopsis. However, the GmTFL1 homolog, despite its more recent divergence from GmTFL1 than from Arabidopsis TFL1, appears to sub- or neo-functionalized, as revealed by the differential expression of the two genes at multiple plant developmental stages and by allelic analysis at both loci.


The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GU046912–GU047324).

1Z.T. and X.W. contributed equally to this work.

2To whom correspondence may be addressed. E-mail: philipp.mcclean@ndsu.edu, quij_liujian263.com, or m.ai@purdue.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1000088107/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1000088107

PNAS | May 11, 2010 | vol. 107 | no. 19 | 8563–8568

Artificial selection for determinate growth habit in soybean

Soybean (Glycine max L. Merr.) is one of the most economically important leguminous seed crops that provide the majority of plant proteins, and more than a quarter of the world’s food and animal feed (1). It is suggested that soybean was domesticated from its annual wild relative, G. soja Sieb & Zucc., in China approximately 5,000 years ago (2), resulting in a multitude of soybean landraces that were adapted to various climate environments. Currently, 23,587 soybean landraces collected from 29 provinces of China are deposited in the Chinese GenBank, representing the world largest reservoir of soybean genetic diversity (3). Some of the landraces are still planted for production in several southern provinces, and some are used worldwide to develop modern soybean cultivars (2, 3).

Based on the timing of the termination of apical stem growth, most soybean cultivars can be classified into two categories of stem growth habit, commonly known as indeterminate and determinate types (4, 5). The apical meristems at the stem and branch apices in indeterminate cultivars maintain vegetative activity (i.e., produces new nodes with trifoliate leaves) until photosynthate demand by developing seeds causes a cessation in the production of vegetative dry matter. In contrast, the apical meristems in determinate cultivars cease vegetative activity at or soon after photoperiod-induced floral induction, and then the meristems become reproductive inflorescences (6). Because determinacy is nonexisting (or rare) in G. soja (4, 5), determinacy in the cultivated soybean is thought to be a trait associated with soybean domestication (7).

Previous studies demonstrated that the stem growth habit in soybean was primarily controlled by Dt1 locus and that the indeterminate phenotype controlled by Dt1/Dt1 was dominant or incompletely dominant over the determinate phenotype controlled by dt1/dt1 (8, 9). This gene is a member of classical linkage group (LG) #5 (10), and was mapped to molecular marker linkage group (LG) L (11). Despite the monogenic inheritance pattern for the Dt1 locus (6), a wide range in the abruptness of stem termination among soybean cultivars has also been observed, and a second gene locus, designated as Dt2, was reported (6). The Dt2 allele is nearly dominant to the dt2, and in Dt1/Dt1 genetic backgrounds, Dt2/Dt2 genotypes produce semideterminate phenotypes and dt2/dt2 genotypes produce indeterminate phenotypes. However, in dt1/dt1 genetic backgrounds, the phenotype is determinate, because dt1 is epistatic to Dt2 and dt2 (6). The Dt2 locus was mapped to classical LG #6 (12) and from there to LG G (13). A third allele at the Dt1 locus (dt1-t) has been identified that produces a phenotype that shares some characteristics of both dt1 and Dt2 (14).

It has been documented that it can be difficult to distinguish between indeterminate and determinate stem types under short photoperiod conditions or under adverse growing condition (6). As stem termination has great effects on plant height, flowering period, node production, maturity, water-use efficiency, and soybean yield (6, 15, 16), isolation and characterization of the genes associated with stem growth habit are very important for soybean germplasm assessment and breeding. In addition, characterization and analysis of these genes in soybean landraces and G. soja would allow us to understand the history and nature of human selection for determinacy.

The availability of the genome sequence and various “omics” tools and approaches for the model species such as Arabidopsis has aided the functional analyses of an increasing number of Arabidopsis genes and genetic pathways (17). Although the corresponding genetic pathways in other plant species are generally not known, several studies have identified the genes that are functionally conserved between model species and crops (18, 19). For...
example, the GAI gene in Arabidopsis is functionally orthologous to the “Green Revolution” dwarfing gene in several cereal crops (18). It now seems clearer that the information gained from the model species can aid gene discovery and functional characterization in crops by the candidate gene approach (20), one of the applications for crop improvement that are collectively placed in the category of “plant translational genomics” (21).

Here, we used a combination of genetic linkage analysis, candidate gene association analysis, and heterologous transformation of Arabidopsis determinant (Dt1) mutants to infer the candidacy of a homolog of Arabidopsis TFL1 in soybean for Dt1. In an attempt to track the history of artificial selection for determinacy, we investigated the allelic variation at the GmTfl1 locus and its homolog in G. soja accessions and in G. max cultivars, including a minicore collection of Chinese landraces in the context of their geographical distribution and population structure. This study illustrates how an Arabidopsis mutant was used as a shortcut to the characterization of natural mutations that were artificially selected in soybean.

Results
Identification of Soybean Genes Homologous to Arabidopsis TFL1. The Arabidopsis TERMINAL FLOWER1 (TFL1) gene was previously identified by isolation of the recessive mutations tfl1 in the TFL1 gene by screening a M2 population derived from EMS-mutagenized seeds of ecotype Columbia (22). The recessive mutations resulted in the conversion of the normally indeterminate inflorescence to a determinate inflorescence condition (22–24). By BLAST searching Arabidopsis TFL1 against the soybean (c.v., Williams 82) whole genome sequence (25), we identified four soybean gene models, Glyma03g33250.1, Glyma10g08340.1, Glyma13g22030.1, and Glyma19g37890.1, that are homologous to TFL1 (Fig. S1). Phylogenetic analysis of these genes suggests that Glyma03g33250.1/Glyma10g08340.1, Glyma13g22030.1, and Glyma19g37890.1, are two homeologous pairs, presumably derived from the soybean genome duplication event that occurred ~50 million years ago (MYA) (25, 26), whereas the two members of each pair likely resulted from the more recent soybean genome duplication event (i.e., allotetraploidization) (27) that took place ~13 MYA (Fig. S1) (26).

The Dt1 locus of soybean was recently fine-mapped as a major quantitative trait locus between two simple sequence repeat (SSR) markers, Sat_099 and Satt229, on LG L (7), which is now designated as chromosome 19 (Gm19) (25). We anchored the SSR markers to the Gm19 sequence and found that Glyma19g37890.1 (designated as GmTfl1) was one of the 380 annotated genes physically located between Sat_099 and Satt229 (24) (Fig. 1). This suggests that GmTfl1 may be a candidate gene for Dt1. Because Williams 82 is a typical indeterminate cultivar, it is likely that GmTfl1 is the candidate Dt1 allele.

Fig. 1. Anchoring genetic markers to the genomic sequence to define the candidate Dt1 gene. Vertical bar between Sat_099 and Satt006 on the genetic map and vertical bar on LG and chromosome sequence indicate the candidate Dt1 gene, Glyma19g37890.1. Gene model was predicted and is depicted by the cartoon underneath the “chromosome.”

Allelic Variation of the Candidate Gene in the Wild and Cultivated Soybean Populations. In an attempt to address whether GmTfl1 and the mutations, if any, that may have occurred in this gene are responsible for the conversion from an indeterminate to determinate phenotype observed in many soybean cultivars, we first sequenced the GmTfl1 locus in a wild G. soja population and three soybean cultivars, representing genotypic groups that likely existed before and after genetic bottlenecks (e.g., domestication to produce landraces, introduction of relatively few landraces to North America, and selective breeding) (28). Fourteen unique SNPs and two insertions/deletions (indels) were detected (Table S1). Of the 14 SNPs, 10 were found in noncoding regions and four in exons. Interestingly, each of the four exonic SNPs generated a single amino acid nonsynonymous substitution (Table S1). Not a single individual genotype was found to contain more than one unique amino acid substitutions. Compared with the Williams 82 reference GmTfl1 sequence, these four amino acid substitutions (referred to as GmTfl1) were only detected in the cultivated soybeans, whereas G. soja genotypes were identical to Williams 82.

Association Between Determinacy and Allelic Nonsynonymous Mutations. To elucidate whether GmTfl1 is Dt1, and whether any or all of the mutations caused the transition from indeterminate type to determinate phenotype, we conducted an association analysis between the mutations and phenotypes using the three aforementioned soybean populations. The stem growth habit phenotypes of the soybean cultivars in these populations were obtained from the USDA Soybean Germplasm Collection database at National Plant Germplasm System (NPGS) (http://www.ars-grin.gov/npgs/), and some of them were directly examined in this study. Of the 89 soybean cultivars, 39 are indeterminate, 41 are determinate, and nine are semideterminate. We found that each of the 39 indeterminate cultivars exhibited the same amino acid sequence as encoded by GmTfl1 in Williams 82, whereas none of the 41 determinate cultivars contain the Williams 82 amino acid sequence but instead possess one (or another) of the four amino acid substitutions (Fig. 2A and Table S2).

Two semideterminate cultivars in this study were found to have the GmTfl1 allele. A previous study demonstrated that the genotype of Dt1/Dt1 ordinarily displays an indeterminate phenotype, but in the presence of Dt2, a dominant allele at another locus controlling stem growth habit, the Dt1/Dt1Dt2/Dt2 genotype will display semideterminate phenotype (6); thus, these two semideterminate cultivars were assumed to contain both Dt1 and Dt2 alleles (Fig. 2A and Table S2). The other seven semideterminate cultivars were found to have GmTfl1 allele. Because it is generally difficult to precisely define the semideterminate phenotypes (6), these nine cultivars were not included in the association analysis below.

G. soja accessions are typically viny, and highly diverged in plant architecture and morphology from G. max; thus, the stem growth habit of the G. soja accessions included in this study was not carefully measured. All of the 20 G. soja accessions were found to contain the same GmTfl1 genotype as Williams 82, which seemingly associates the indeterminacy with G. soja as is generally conjectured.

Thus, we observed a perfect association between the amino acid substitutions and the determinacy when G. soja accessions and the semideterminate cultivars were excluded. This suggests that GmTfl1 is the Dt1 allele, and the four single-point mutations (which could be characterized as functional SNPs) resulted in the four distinct amino acid substitutions are dt1 alleles.

Excluding the four functional SNP variants, the GmTfl1 alleles in the four populations were classified into two distinct types, designated as GmTfl1-a and GmTfl1-b. The four mutations were subclassifiable as GmTfl1-ta, GmTfl1-bb, GmTfl1-tb, and GmTfl1-ab (Fig. 2B). GmTfl1-a and GmTfl1-ta share the same form, and GmTfl1-b, GmTfl1-bb, GmTfl1-tb, and GmTfl1-ab share the other form, suggesting that GmTfl1-ta was derived from GmTfl1-a whereas GmTfl1-bb, GmTfl1-tb, and GmTfl1-ab were derived from GmTfl1-b. Linkage...
dis-equilibrium (LD) analysis showed that the SNPs and indels in the first intron (from +285 to +311) are linked with the two SNPs in the 5' UTR (−499 and −410), but the four functional SNPs did not show LD with the other sites (Fig. S2A). We also sequenced Glyma19g37900.1, a gene flanking GmTfl1, in six landraces that contain different alleles at the GmTfl1 locus, and found that LD exists between the SNPs detected at the Glyma19g37900.1 locus and the nonfunctional polymorphisms (−499, −410, and +285 to +311) at the GmTfl1 locus (Fig. S2C). These observations indicate that the transition from indeterminate type to determinate type was not caused by the linked polymorphisms within the GmTfl1 locus, or between the GmTfl1 locus and its flanking gene, but by the four functional mutations. These observations further strengthen the inference that that GmTfl1 is Df1.

In addition to the four populations analyzed above, we sequenced the GmTfl1 locus in 17 previously described determinate soybean cultivars (Table S3). All of the 17 cultivars were found to be GmTfl1 mutations (two GmTfl1-ta, two GmTfl1-ab, and 13 GmTfl1-ab), a result consistent with the association analysis above. We also sequenced the GmTfl1 locus in three semideterminate isogenic lines that share the Clark (an indeterminate cultivar) genetic background but differ from Clark at the D2 locus (6), and did detect the GmTfl1 allele in all these isogenic lines (Table S3).

GmTfl1 Complements the Functions of TFL1 in Arabidopsis. To validate the function of GmTfl1 for indeterminacy (vs. GmTfl1 for determinacy) we introduced the Williams 82 GmTfl1 allele into the Arabidopsis determinate mutant (ftfl1-1) (24) (Materials and Methods), and obtained two transgenic lines, one of which is shown in Fig. 3C. The absence of Arabidopsis TFL1 allele and the presence of the soybean GmTfl1 allele in the transgenic lines were confirmed by PCR analysis and sequencing of PCR fragments (Fig. 3G and I). The transgenic (GmTfl1) lines (Fig. 3C and G) showed the same phenotypes as the wild-type Arabidopsis (i.e., indeterminate and late flowering). Because the transgene (GmTfl1) is a combination of the Arabidopsis TFL1 promoter and the protein coding sequence (CDS) of the GmTfl1 allele, the conversion of the transgenic lines from the mutant type (determinate and early flowering) to the wild type would be interpreted that the transgene in the Arabidopsis (ftfl1/ftfl1) mutant fully complements the functions of TFL1 observed in the wild-type Arabidopsis.

The question remained whether the nonsynonymous substitutions (GmTfl1 alleles) detected at the GmTfl1 locus in the cultivated soybean have no or diminished functions relative to the GmTfl1 allele for indeterminacy. To address this question, we introduced the GmTfl1-ab, the predominant allele detected in the cultivated soybean populations (Fig. 2A), into the Arabidopsis ftfl1-1 mutants (Materials and Methods), and obtained eight transgenic (GmTfl1) lines. The absence of the Arabidopsis TFL1 allele and the presence of the soybean GmTfl1-ab allele in the eight transgenic lines were confirmed by PCR analysis and sequencing of PCR fragments (Fig. 3H and J). We found that each of the eight lines showed phenotypes nearly identical to that of the Arabidopsis ftfl1-1 mutant. The phenotypes of one of the eight lines are illustrated in Fig. 3 D and H.

Evolutionary Diversification between GmTfl1 and its Homolog. Since GmTfl1 and Glyma03g35250.1 are thought to be a homologous pair (Fig. 1B), it would be interesting to track the evolutionary difference between GmTfl1 and Glyma03g35250.1. We thus sequenced the Glyma03g35250.1 locus in the same populations used...
for the analysis of the GmTfl1 locus. Five SNPs were detected at the Glyma03g35250.1 locus in the G. soja population, but none were found in the cultivated populations (Table S1 and Fig. S2B). The level of nucleotide diversity at both GmTfl1 and Glyma03g35250.1 loci (Table 1) is lower than the average in the G. soja population estimated based on a set of gene fragments (28). In addition, non-synonymous substitutions were not found at either locus in the G. soja population (Table 1), suggesting that both genes have undergone purifying selection. However, GmTfl1 and Glyma03g35250.1 exhibited a substantial difference in diversity in the cultivated soybean populations. For example, the Glyma03g35250.1 allele was invariant among all of the members of the soybean landrace population, whereas the non-synonymous substitutions at the GmTfl1 locus that resulted in the four GmTfl1 alleles were observed in the same population (Table 1). Together, these observations suggest that the fixation of the four GmTfl1 alleles in cultivated determinate soybean would be the outcome of deliberate human selection during the development of soybean landraces.

**Differential Expression of GmTfl1 and its Homeolog.** To shed light on the functional diversity of GmTfl1 and Glyma03g35250.1, we compared their expression pattern. Quantitative RT-PCR was used to profile the expressions of GmTfl1 and Glyma03g35250.1, in the indeterminate cultivar Williams 82 in different tissues and at different developmental stages. GmTfl1 was mainly expressed in young roots, young leaves and flowers seven day after flowering (7DAF), whereas Glyma03g35250.1 was mainly expressed in young roots, young stems and buds (Fig. S3). Given that Arabidopsis TFL1 is involved in inflorescence meristem development pathway (22, 24), high-level of expression of GmTfl1 in flowers 7DAF is expected. Thus, the lack of expression of Glyma03g35250.1 at this stage may be considered as evidence that the Glyma03g35250.1 was subfunctionalized or neo-functionalized. This inference is echoed by the analysis of allelic variation at both GmTfl1 and Glyma03g35250.1 loci in the soybean populations. It can be deduced that neither Glyma03g35250.1 nor the other pair of homeologous genes (Glyma13g22030.1, Glyma10g08340.1), homologous to TFL1, are potential candidates for the D12 locus, as these three loci are not located on LG G (chromosome 18), where the D12 was mapped.

**Timing and Nature of Artificial Selection for the GmTfl1 Alleles.** None of the four GmTfl1 alleles identified in G. max were detected in the G. soja population analyzed in this study. To search for evidence of the history of the human selection with respect to the GmTfl1 alleles, we sequenced the GmTfl1 locus in a minicore collection of 195 soybean landraces, which were selected based on the genetic structure of a core collection of 1,863 landraces that maximally represent the 23,587 Chinese soybean landraces deposited in the Crop GenBank at the Chinese Academy of Agricultural Sciences. We subsequently analyzed the distribution of the four GmTfl1 alleles in the core collection of landraces with respect to their genetic diversity and geographic distribution. The GmTfl1 alleles were seen in all of the major branches of the Neighbor-Joining tree of the 195 landraces constructed based on 59 SSR markers (3) (Fig. 4A). It is noticeable that GmTfl1-ta and GmTfl1-tb were found in a highly diverged group of (seven) landraces that are the most closely related to G. soja, a wild accession used as an outgroup, and six of these seven landraces show “semi-wild” phenotypes, such as viny stems and dark brown seed coat (3). These data indicate that the human selection for determinacy must have occurred before the radiation of all of the lineages of these Chinese landraces, either just after or during the major domestication transition. Although the GmTfl1 landrace alleles were found in all of the three large soybean-growing ecological regions, referred to as Northern eco-region (NR), Huang-Huai eco-region (HR), and Southern eco-region (SR), which were subclassified into NESp and NSp,

**Table 1. Nucleotide diversity per base pair \(\times 10^3\) in GmTfl1 and Glyma03g35250.1**

<table>
<thead>
<tr>
<th>gDNA</th>
<th>(\pi)</th>
<th>(\theta)</th>
<th>CDS</th>
<th>(\pi)</th>
<th>(\theta)</th>
<th>Synonymous</th>
<th>Nonsynonymous</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmTfl1 All</td>
<td>1.86</td>
<td>1.21</td>
<td>1.39</td>
<td>1.46</td>
<td>0.00</td>
<td>0.00</td>
<td>1.83</td>
</tr>
<tr>
<td>Elite cultivars</td>
<td>1.86</td>
<td>0.98</td>
<td>0.98</td>
<td>0.52</td>
<td>0.00</td>
<td>0.00</td>
<td>1.29</td>
</tr>
<tr>
<td>Landraces</td>
<td>1.78</td>
<td>1.05</td>
<td>1.78</td>
<td>1.61</td>
<td>0.00</td>
<td>0.00</td>
<td>2.34</td>
</tr>
<tr>
<td>G. soja</td>
<td>1.65</td>
<td>1.28</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Glyma03g35250.1 All</td>
<td>0.15</td>
<td>0.49</td>
<td>0.20</td>
<td>0.36</td>
<td>0.87</td>
<td>1.57</td>
<td>0.00</td>
</tr>
<tr>
<td>Elite cultivars</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Landraces</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>G. soja</td>
<td>0.66</td>
<td>0.73</td>
<td>0.85</td>
<td>0.54</td>
<td>3.65</td>
<td>2.32</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Discussion

Functional Conservation and Divergence of TFL1 Homologs Within and Among Species. We demonstrated that the soybean GmTFL1 gene is the functional homolog of the Arabidopsis TFL1 gene by a comparative genomics approach. When GmTFL1 was introduced into the Arabidopsis tfl1 mutants, it fully restored the wild-type phenotypes, which are controlled by TFL1 in the wild-type Arabidopsis. The functional homolog of TFL1 has been found in Antirrhinum (29), Solanum lycopersicum (30), and Pisum sativum (31), suggesting that the common mechanism underlies indeterminacy in these species.

Soybean GmTFL1 was found to play the same roles as Arabidopsis TFL1 in determining the inflorescence commitment and architecture (24) in the transgenic Arabidopsis tfl1 mutant, but it does not seem to delay the commitment to inflorescence development in soybean. This is reflected by a general lack of correlation between the flowering time (i.e., late flowering and early flowering) and stem growth habit (i.e., indeterminacy and determinacy) of soybean cultivars. In addition, our expression data and allelic analysis at the GmTFL1 and Glyma03g35250.1 loci indicate that this pair of homeologs has been sub- or neo-functionalized, likely after their duplication through allotetraploidization.

Natural Selection vs. Artificial Selection. Despite their functional divergence, GmTFL1 and Glyma03g35250.1 both appear to have undergone purifying selection, which is partly reflected by the lack of nonsynonymous substitutions at either GmTFL1 or Glyma03g35250.1 loci in the natural population of G. soja (Table 1). Our data revealed a total of four unique nonsynonymous substitutions in the domesticated soybean landraces, each of which led to the conversion of soybean stem habit from indeterminacy to determinacy. By contrast, no mutations present in G. soja at the Glyma03g35250.1 locus were detected in the cultivated soybean populations. Given that more than 80% rare alleles presented in the G. soja population were eliminated through the bottleneck of soybean domestication (28), the appearance and maintenance of the GmTFL1 alleles at such a high frequency in the soybean populations, which are currently absent in the G. soja population, must be assumed to be the outcome of deliberate artificial selection. Because only several semideterminate cultivars were identified in the populations investigated, D12 was unlikely an allele associated with soybean domestication.

Artificial Selection, Linkage Disequilibrium, and Genetic Bottleneck. Although 50% of the G. soja genetic diversity was reduced through the bottleneck of soybean domestication (28), it appears that selection for the GmTFL1 mutations did not cause apparent erosion of diversity. This was inferred by the observation that both indeterminate and determinate landraces in the minicore collection exhibited the similar levels of genetic diversity (Fig. S4). Instead, four GmTFL1 alleles were observed among cultivated soybean, whereas G. soja only contained the GmTFL1 allele. Genetic bottle necks are thought to reduce genetic diversity and increase LD (28). We found that LD in the GmTFL1 locus and the flanking regions (Fig. S2C) is extremely high, but LD was decayed at GmTFL1 alleles. This suggests that the selection for the GmTFL1 alleles has had little effects on the genes linked to the GmTFL1 locus. We found that GmTFL1-a and GmTFL1-b were absent in North American Ancestors, and GmTFL1-bb and GmTFL1-a were further eliminated from the Elite Cultivars developed in the USA, reflecting the effects of genetic bottlenecks created by soybean germplasm introduction and modern breeding (28).

Radiation and Adaptation of GmTFL1/GmTFL1 Alleles to Local Eco-Regions. The domestication of soybean is hypothesized to have occurred in China, but there is no consensus about where within China it might have occurred. An early study proposed that soybean was domesticated in the Northeastern (NE) subregion within NR (32). However, a recent analysis of genetic structure and diversity of a core collection of Chinese soybean landraces demonstrated that the landraces collected from the region between 32.0 and 40.5°N, and 105.4 and 122.2°E along the central and downstream parts of the Yellow River (HSu subregion within HR) display the highest genetic diversity. This molecular data were used as evidence for the hypothesis that the cultivated soybean originated in the Yellow River region (3). Our observations are generally consistent with the latter hypothesis for a few reasons. First, GmTFL1-b was found to be the predominant allele in the G. soja accessions from the NESp subregion, but not a single GmTFL1 allele derived from GmTFL1-b (i.e., GmTFL1-bb, GmTFL1-bb, and GmTFL1-bb) were detected in the landraces from this subregion. Second, because indeterminate cultivars were highly desirable in the NESp subregion, the determinate alleles were unlikely to be deliberately selected by humans and from there widely spread to other eco-regions, at least during or after the domestication event. Next, compared with all other subregions, the HSu subregion contains landraces that display the highest level of allelic variation at the GmTFL1 locus (Fig. S4).

Regardless of the origin of the cultivated soybean, it is clear that the GmTFL1 and GmTFL1 alleles spread rapidly, fixed, and adapted to local eco-regions or subregions. The GmTFL1 allele was favored

Fig. 4. Allelic mutations at the GmTFL1 locus in the context of genetic diversity and eco-geographic distribution of a core collection of soybean landraces. (A) Phylogenetic relationship of the landraces assessed by 59 SSR markers and the types of alleles (GmTFL1 or GmTFL1) detected in individual landraces. (B) Geographical distribution of the landraces in the soybean growing eco-regions or subregions in China.
in the NR region, whereas Gmflf1 alleles were favored in the SR, and thus formed a middle region (i.e., HR) with Gmflf1 and Gmflf1 alleles fairly evenly distributed (Fig. 4 and Fig S3). Under the assumption that each landrace is homozygous at the Gmflf1 locus, which is highly supported by the high quality of nucleotides at the mutation sites, it is estimated that the frequencies of Gmflf1 and Gmflf1 in the landraces collected from the three major ecoregions, NR, HR, and SR, are 0.18 and 0.82, 0.50 and 0.50, and 0.81 and 0.19, respectively. We still do not know whether the Gmflf1 mutations were selected after the domestication event or integral to the process of domestication, but it is obvious the artificial selection of the natural Gmflf1 mutations played a central role in shaping the radiation of initially developed landraces. Because the determinant phenotype is shorter and thus more lodging-resistant in fertile production areas, its appearance during or after domestication probably resulted in an ancient “green revolution” in soybean cultivation in the southern parts of ancient China.

Materials and Methods

Plant Materials. The G. soja population and the three soybean populations previously described by Hyten et al. (28), and the 17 determine cultivars and the D22 isogenic lines listed in Table S3 were obtained from United States Department of Agriculture Soybean Germplasm Collection. The collection of Chinese soybean landraces previously described by Li et al. (3) were obtained from the Department of Agriculture Soybean Germplasm Collection. The collection of G. soja accession as an outgroup, and visualized by MEGA (39).

DNA Isolation, PCR, and Sequencing. Genomic DNA isolation, PCR primer design, PCR amplification, PCR fragment purification, and sequencing of PCR fragments were conducted as described (33). Primers used for PCR amplification of Gmflf1, and Glyma03g35250.1 were listed in Table S5.

Sequence Alignments, Genetic Structure, Linkage Disequilibrium, and Phylogenetic Analysis. The alignments of nucleotide and amino acid sequences were performed using MUSCLE (35). The observed nucleotide diversity (θ) was calculated using DnaSP (3). The SNP data (28) and SSR marker data (3) were used to analyze the genetic structures of G. soja and G. max populations using the software package STRUCTURE (36). LD was evaluated using TASSEL (37).

Plasmid Construction and Transformation. The promoter region of Arabidopsis TFL1 (the same as that of flf1) was fused with the CDS of Gmflf1 (amplified from an indeterminate soybean cultivar Williams 82) or Gmflf1-ab (amplified from a determinate soybean cultivar Young), and inserted to pCAMBIA1391 vector (CAMBIA). Then the constructs were introduced into the Arabidopsis flf1-1 mutants by the floral dip procedure (40). The absence of the Arabidopsis TFL1 allele and the presence of the Gmflf1 and Gmflf1-ab constructs were confirmed by PCR and sequencing of PCR fragments. Primers used are listed in Table S5. All Arabidopsis plants were grown at 24 °C under the condition of 16 h of 120 μE m−2 s−1 light and 8 h of dark.

RNA Extraction and Expression Analysis. Total RNA isolation, cDNA synthesis, and RT and quantitative PCR were conducted as previously described (41). The soybean Actin1 gene was used as control. Primers used are listed in Table S5.

Acknowledgments. We thank Dr. Enrico Coen for providing the wide-type Arabidopsis and the flf1-1 mutant seeds, Dr. Michael Zanis for insightful comments, and the anonymous reviewers for their constructive suggestions.