Natural variations of \textit{FT} family genes in soybean varieties covering a wide range of maturity groups

Bingjun Jiang\textsuperscript{†}, Shouwei Zhang\textsuperscript{†}, Wenwen Song\textsuperscript{†}, Mohammad Abdul Awal Khan, Shi Sun, Chengsheng Zhang, Tingting Wu, Cunxiang Wu and Tianfu Han*  

Abstract
Background: Flowering time and maturity are among the most important adaptive traits in soybean (\textit{Glycine max} (L.) Merill). \textit{Flowering Locus T} (\textit{FT}) family genes function as key flowering integrators, with flowering-promoting members \textit{GmFT2a}/\textit{GmFT5a} and flowering-inhibiting members \textit{GmFT4}/\textit{GmFT1a} antagonistically regulating vegetative and reproductive growth. However, to date, the relations between natural variations of \textit{FT} family genes and the diversity of flowering time and maturity in soybean are not clear. Therefore, we conducted this study to discover natural variations in \textit{FT} family genes in association with flowering time and maturity.

Results: Ten \textit{FT} family genes, \textit{GmFT1a}, \textit{GmFT1b}, \textit{GmFT2a}, \textit{GmFT2b}, \textit{GmFT3a}, \textit{GmFT3b}, \textit{GmFT4}, \textit{GmFT5a}, \textit{GmFT5b} and \textit{GmFT6}, were cloned and sequenced in the 127 varieties evenly covering all 14 known maturity groups (MG0000-MGX). They were diversified at the genome sequence polymorphism level. \textit{GmFT3b} and \textit{GmFT5b} might have experienced breeding selection in the process of soybean domestication and breeding. Haplotype analysis showed that a total of 17 haplotypes had correlative relationships with flowering time and maturity among the 10 \textit{FT} genes, namely, 1a-H3, 1b-H1, 1b-H6, 1b-H7, 2a-H1, 2a-H3, 2a-H9, 2b-H3, 2b-H4, 2b-H6, 2b-H7, 3b-H4, 5a-H1, 5a-H2, 5a-H4 and 5b-H1. Based on the association analysis, 38 polymorphic sites had a significant association with flowering time at the level of p < 0.01.

Conclusions: Some natural variations exist within the 10 \textit{FT} family genes, which might be involved in soybean adaptation to different environments and have an influence on diverse flowering time and maturity. This study will facilitate the understanding of the roles of \textit{FTs} in flowering and maturity.

Keywords: Soybean, \textit{FT} family genes, Flowering time, Maturity group, Haplotype

Background
Soybeans are planted within a wide range of latitudes across the world resulting from the rich diversity of varieties in flowering time and maturity. In North America, soybeans are classified into 13 maturity groups: MG000 to MGX in the ascending order of maturity [1]. Gai divided 256 Chinese soybean landraces into 12 maturity groups without MGX [2]. MG0000, a new maturity group that matures the earliest to date was identified in some super-early varieties from high-latitude cold regions in China and the Far East of Russia [3]. Previous research indicates that maturity diversity in soybean is attributed to the variation and combination of genes responding to photoperiod and temperature [3–6].

Multiple genes control flowering time and maturity in soybean. A total of 10 maturity loci are characterized in soybean, including \textit{E1–E10} and \textit{J} [7–16]. A majority of these loci were reviewed by Xia et al. [17]. They play different roles in flowering and maturity maintenance under diverse photoperiod patterns. To date, six of these loci, \textit{E1} [18], \textit{E2} [19], \textit{E3} [20], \textit{E4} [21], \textit{E9} [22, 23] and \textit{J} [24, 25], have been cloned. Among the six loci, \textit{E9} has been confirmed as \textit{GmFT2a}, an orthologue of the Arabidopsis...
Flowering Locus T (FT) gene, which plays an important role in flowering [22, 23, 26].

FT, which encodes a putative florigen, is a key integrator gene in the regulation of flowering in Arabidopsis [27], and the function is highly conserved in different species [28, 29]. Ten FT-like genes have been identified in soybean among which GmFT2a and GmFT5a are confirmed as flowering promoters; whereas GmFT1a is divergent as a floral and maturity inhibitor [26, 30, 31]. GmFT2a and GmFT5a coordinately control flowering as flowering integrators and can be regulated by E1, E2, E3 and E4 [18, 19, 26, 32]. Ectopic expression experiments in Arabidopsis also demonstrated that other soybean FT family genes

**Table 1** Summary of polymorphic sites of the 10 FT family genes in soybean

| Gene    | Sample size | No. of aligned bp | Type of mutation | No. of polymorphic sites | Sequence variation ratio | Nucleotide diversity | Tajima’s D |
|---------|-------------|-------------------|------------------|--------------------------|--------------------------|---------------------|------------|
| GmFT1a  | 119         | 5957              | SNP              | 105                      | 1.79%                    | 0.00434             | 0.92498    |
|         |             |                   | Indel            | 24                       | 0.41%                    | 0.00087             | −0.21635   |
| GmFT1b  | 99          | 2745              | SNP              | 73                       | 2.86%                    | 0.00845             | 1.67398    |
|         |             |                   | Indel            | 20                       | 0.78%                    | 0.00257             | 0.54345    |
| GmFT2a  | 118         | 5365              | SNP              | 40                       | 0.75%                    | 0.0014              | 0.15918    |
|         |             |                   | Indel            | 11                       | 0.21%                    | 0.00072             | 0.72832    |
| GmFT2b  | 118         | 3058              | SNP              | 41                       | 1.34%                    | 0.00385             | 1.59284    |
|         |             |                   | Indel            | 20                       | 0.65%                    | 0.00203             | 1.67091    |
| GmFT3a  | 122         | 2564              | SNP              | 10                       | 0.39%                    | 0.00034             | −3.1769    |
|         |             |                   | Indel            | 1                        | 0.04%                    | 0.00005             | −0.29027   |
| GmFT3b  | 109         | 2241              | SNP              | 20                       | 0.89%                    | 0.003              | 2.20415*   |
|         |             |                   | Indel            | 4                        | 0.18%                    | 0.00099             | 2.84140**  |
| GmFT4   | 127         | 1907              | SNP              | 3                        | 0.16%                    | 0.00009             | −1.17378   |
|         |             |                   | Indel            | 5                        | 0.26%                    | 0.00097             | 0.99565    |
| GmFT5a  | 127         | 2039              | SNP              | 9                        | 0.44%                    | 0.0004              | −1.27444   |
|         |             |                   | Indel            | 5                        | 0.25%                    | 0.0002              | −1.1339    |
| GmFT5b  | 108         | 2942              | SNP              | 21                       | 0.76%                    | 0.00349             | 4.08832*** |
|         |             |                   | Indel            | 7                        | 0.25%                    | 0.00112             | 3.44592*** |
| GmFT6   | 125         | 6545              | SNP              | 28                       | 0.43%                    | 0.00033             | −1.70857   |
|         |             |                   | Indel            | 10                       | 0.15%                    | 0.00013             | −1.48424   |

Notes: n, average nucleotide differences per site between the two sequences; θw, Watterson estimator; Tajima’s D, test for neutral selection (*: significant at p < 0.01; **: significant at p < 0.001; ***: significant at p < 0.0001)
including GmFT2b, GmFT3a, GmFT3b and GmFT5b promoted flowering [33, 34], whereas GmFT4 delayed flowering [35]. In transgenic soybean overexpressing E1, GmFT1a and GmFT4 expression was up regulated, whereas the expression of GmFT2a and GmFT5a was suppressed [31, 35]. Currently, a new GmFT2c was recently found inserted in a putative transposon in the third intron, and GmFT2d was determined structurally rearranged with some remnant in the genome, indicating that GmFT2 subclades (GmFT2a, GmFT2b, GmFT2c and GmFT2d) have different evolutionary trajectories [36].

In plants, multiple regulation pathways regulate FT expression in response to diverse environmental signals, which enable flowering and other developmental responses to be seasonally timed [37]. Different natural variations of FT promoter sequences are confirmed responsible for diverse flowering time in Arabidopsis and also in rice [38, 39]. The promoter of GmFT2a in soybean has high polymorphism, and some SNPs in the promoter region are associated with flowering time and photoperiod sensitivity [34]. GmFT2a has a recessive allele for delayed flowering, which contains a Ty1/copia-like retrotransposon SORE-1 inserted in the first intron that attenuates GmFT2a expression by its allele-specific transcriptional repression [23]. Critical sequence polymorphism across the gene and its flanking regions may reflect the evolutionary process of species adapting to different environments. The FT homologs have functional divergence in soybean [22, 23, 26, 31, 32], implying that FT genes play roles that influence the flowering and maturation time through an interactive balance when receiving different environmental signals. Therefore, whether their natural variations are linked to flowering time and maturity requires investigation.

In this study, 127 soybean varieties covering all known 14 maturity groups (MG) from MG0000 to MGX were selected and grown for phenotypic and genotypic identification. The sequence polymorphisms of 10 soybean FT family genes in all varieties were analyzed. Further analysis of FT haplotypes and their natural variations associated with flowering time and maturity was conducted. The results suggested that some natural variations of the 10 soybean FT genes existed in soybeans of different MGs and that these variations were related to soybean flowering time and maturity. Our findings will elucidate the roles of soybean FT family genes in flowering and maturity.

**Methods**

**Plant materials and investigation of flowering time and maturity**

One hundred and twenty-seven soybean (Glycine max) varieties covering all known 14 maturity groups were collected for the experiment (Additional file 1: Table S1). These varieties were collected from the Institute of Crop
Science, the Chinese Academy of Agricultural Sciences, which are public and available for non-commercial purpose. They were primarily from China and North America, with a few from Brazil, India, Costa Rica and the Far East of Russia. Fifteen soybean seeds were planted in 10-l pots on May 27, 2015, and grown outdoors under natural conditions in Beijing, China (39.95° N, 116.32° E). After Ve (emergence), seedlings with similar size were selected so that each pot contained five uniform plants. Each variety was planted in three replicates pots. We collected data on the developmental stages, including Ve, R1 (beginning bloom), R7 (physiological maturity) and R8 (full maturity), according to the description by Fehr and Carviness [40]. The mean of the three replicates for each variety was used for statistical analysis.

**DNA isolation, PCR and sequencing**

Genomic DNA was extracted from fresh trifoliate leaves using the standard cetyltrimethyl ammonium bromide (CTAB) method [41]. Seventeen PCR primer pairs were used to amplify 10 soybean FT family genes, and 17 fragments were produced. GmFT1a, GmFT2a and GmFT6 were fragmentally amplified with three primer pairs, GmFT3a with two primer pairs, and GmFT1b, GmFT2b, GmFT3b, GmFT4, GmFT5a and GmFT5b were each amplified with one primer pair. The sequences of these primers are listed in Additional file 2: Table S2. Target regions were amplified with the high-fidelity polymerase of KOD-Plus-Neo and KOD-FX, produced by Toyobo Life Science, Beijing, China. Their reaction conditions are listed in Additional file 3: Table S3. The PCR

![Fig. 3 Phylogenetic relationships of different types of soybean FT proteins using the NJ method. Bootstrap percentage supports are indicated at the branches of the tree. *" represents the protein sequence type that is the same as that of Williams 82](image-url)
products were directly sequenced using the Sanger method at Tsingke Biological Technology Company, Beijing, China.

**Sequence analysis and haplotype definition of the 10 soybean FT family genes**

We used Seqman 1.0 for sequence assembly [42]. The sequences were aligned and adjusted manually using ClusterX 2.0 [43] and Bioedit [44]. The sequences after alignment were input into DnaSP v5 to calculate nucleotide diversity and Tajima's D statistics [45]. The sequence variation ratio was calculated as follows: Sequence variation ratio = (No. of polymorphic sites/gene length, referring to the reference genome) x 100%. The phylogenetic relationships among different types of soybean FT proteins were analyzed using the neighbor-joining (NJ) method with the program MEGA 7 [46]. During the analysis of polymorphic sites, the 10 soybean FT family genes in soybean reference genome in the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) were used as references. We chose polymorphic sites with minor allele frequency above 4 varieties and imported them into Tassel 5 to analyze the linkage disequilibrium (LD) [47]. We conducted haplotype analysis based on sequence alignments and the feature of LD and different sites. After haplotype definition, the primary haplotypes of the 10 FT family genes were analyzed for association with flowering time and maturity. For those polymorphic sites that only appeared in these haplotypes with relation to flowering time and maturity, we defined them as candidate polymorphic sites (Additional file 4: Table S4). Then, we selected the candidate polymorphic sites of each soybean FT family gene and used Tassel 5 to identify polymorphic site-trait associations by generating a general linear model (GLM: trait = marker effect +
residual) [48], and the cutoff of \( p < 0.01 \) is selected as in our previous study on \( GmFT2a \) promoter variations [34]. The polymorphic site information of the 10 soybean \( FT \) family genes is listed in Additional file 5: Table S5.

**Results**

Soybean varieties from different maturity groups varied greatly in days to R1, R7 and R8 from emergence

Except for some late-maturing varieties that did not reach R7 before the end of the experiment, the flowering time (Ve-R1) varied from 22.2 to 120.5 days, with a span of 98.3 days, and the maturity (Ve-R7) varied from 61.4 to 155.7 days, with a span of 94.3 days (Additional file 6: Table S6). Figure 1 shows that the Ve-R1 of the varieties from MG0000 to MG0 was similar under the natural photothermal conditions in Beijing, whereas the Ve-R1 gradually increased from MGI to MGX. Both Ve-R7 and Ve-R8 increased from MG0000 to MGX (Fig. 1). Among the tested varieties, MG0000 variety Hujiao07–2479 flowered the earliest with the Ve-R1 of 22.2 days, and MG0000 variety Star4/75 matured the earliest with the Ve-R7 of 61.4 days (Additional file 6: Table S6). Zigongdongdou of MGVIII flowered the latest with the Ve-R1 of 120.5 days. The varieties exhibited rich diversity in flowering time and maturity.

Sequence polymorphisms of the 10 soybean \( FT \) family genes

The results of sequence analysis showed that the 10 soybean \( FT \) family genes diversified in sequence polymorphism, with the polymorphism sites ranging from 8 (\( GmFT4 \)) to 129 (\( GmFT1a \)) (Table 1). The \( \pi \) and \( \theta_w \) of nucleotide diversity parameters extended from 0.00039 (\( GmFT3a \)) to 0.01102 (\( GmFT1b \)) and from 0.00080 (\( GmFT3a \)) to 0.00787 (\( GmFT1b \)), respectively (Table 1). For all the genes except \( GmFT1a \), the variation ratios of introns were higher than those of exons. The variation ratio of introns in \( GmFT1a \) was slightly lower than that of exons (Fig. 2). Notably, variation ratios of exon4 of the 10 \( FT \) family genes were commonly higher than those of other exons, suggesting that the exon4 or its resultant protein domain was less conserved than other exons (Fig. 2).
Among the 10 soybean FT genes, GmFT1a had the most polymorphic sites (129), including 105 SNPs and 24 Indels (Table 1). However, because GmFT1a was one of the longest genes, only shorter than GmFT6, its nucleotide diversity ($\pi$) was 0.00521, lower than that of GmFT1b ($\pi = 0.01102$) and GmFT2b ($\pi = 0.00588$) (Table 1). GmFT1b had 93 polymorphic sites in the 2745 aligned bases, including 73 SNPs and 20 Indels (Table 1). GmFT1b nucleotide diversity ($\pi$) reached the peak of 0.01102 among the 10 soybean FT family genes, whereas GmFT3a had the lowest $\pi$ of 0.00039, with only 11 polymorphic sites (Table 1). For the two polymorphic types, SNPs and Indels contributed differently to nucleotide diversity for each FT family gene with the contribution of the SNPs greater than that of Indels. GmFT1a was such a case, and SNP and Indel polymorphisms accounted for 0.00434 and 0.00087, respectively, with respect to the total nucleotide diversity ($\pi$). GmFT1b had a comparatively high $\pi$ of SNPs of 0.00845; nevertheless, $\pi$ of SNPs of GmFT4 was much lower at 0.00009 (Table 1). GmFT1b also had the highest $\pi$ of Indels, up to 0.00257, and the $\pi$ of GmFT2b was slightly lower than that of GmFT1b, up to 0.00203, compared with the lowest $\pi$ of the Indels of GmFT3a of only 0.00005 (Table 1). Figure 2 shows that GmFT1b had high polymorphisms, followed by GmFT1a and GmFT2b with polymorphisms lower than those of GmFT1b, whereas GmFT3a and GmFT4 were conservative. The coding region of GmFT4 was highly conserved, without any polymorphic sites, and GmFT3a, GmFT5a and GmFT5b were comparatively conservative, with only one polymorphic site in the coding region (Fig. 2 and Additional file 5: Table S5). To study the population selection pressure, we conducted neutral testing using Tajima’s $D$. Tajima’s $D$ value of both GmFT3b and GmFT5b was positive, with their values reaching extremely significant levels ($p < 0.001$ and $p < 0.0001$), except that the SNPs of GmFT3b exhibited a significant level ($p < 0.01$) (Table 1). These data indicated that these regions of the two genes might experience balancing selection or population shrinkage.

In our study, we failed to identify any parsimony informative sites that led to premature stop codons in...
the coding region; however, some synonymous and nonsynonymous polymorphic sites existed. GmFT1a had seven types of protein, and GmFT1b had 5 types of protein, but GmFT3a, GmFT4, GmFT5a and GmFT6 had only one type of protein, indicating the high evolutionary conservation of these genes (Fig. 3).

Compared with the polymorphic sites of the 10 soybean FT family genes in the Phytozome database, new polymorphic sites were found (Additional file 5: Table S5). For example, 94 new polymorphic sites were found in GmFT1a (Additional file 5: Table S5). These data suggested that the population we selected had extensive coverage of polymorphic sites and was particularly useful to research natural variations.

### Haplotype analysis of the 10 soybean FT family genes

Strong linkage disequilibrium (LD) has a large contribution to haplotype analysis and eases the identification of some SNPs and Indels in relation to phenotypes. Soybean FT family genes exhibited different levels of LD. Among the genes, GmFT1a and GmFT1b both had a similar level of polymorphism, and LD was weak in GmFT1a but strong in GmFT1b (Figs. 4 and 5).

GmFT3b presented strong LD across the region from starting site to nearly the middle site, as a haplotype block, and across the entire region, GmFT5b presented some strong LD dispersed (Figs. 9 and 12). The entire gene region of GmFT3a could be defined as a haplotype block because of the strong LD of its polymorphic sites (Fig. 8).

On the basis of sequence alignment and LD, haplotype analysis proceeded for the 10 soybean FT family genes in the 127 varieties used in this study. Haplotypes of the 10 soybean FT family genes, variation types and polymorphic sites used for composing the haplotype are listed in Additional file 7: Table S7, Additional file 8: Table S8 and Additional file 9: Table S9. These results indicated that 1a-H3, 1b-H1, 1b-H6, 1b-H7, 2a-H1, 2a-H3, 2a-H7, 2b-H3, 2b-H7, 3b-H4, 5a-H1, 5a-H2, 5a-H4 and 5b-H1 showed some association with flowering time and maturity, whereas the haplotypes of GmFT3a, GmFT4 and GmFT6 did not exhibit any apparent associations (Figs. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13). Haplotype combinations of soybean FT family genes showed some associations, i.e., 1a-H3 appeared in the varieties with the haplotype of 1b-H1, and
2a-H1 and 2a-H3 appeared in the varieties with 2b-H6 and 2b-H7, respectively, and 3b-H4 appeared in the varieties with 5b-H1 (Additional file 7: Table S7). A total of 8 haplotypes of GmFT1b, with 64 SNPs and 12 Indels, were categorized in 99 varieties (Fig. 5 and Additional file 9: Table S9), whereas a total of 4 haplotypes of GmFT4 were defined in 127 varieties with 1 SNP and 3 Indels because of its sequence conservation (Fig. 10). 1b-H5 was the most abundant, accounting for 46 varieties; 1b-H1 appeared in varieties in MGI and late-maturing groups, and the Ve-R1 of 1b-H1 varied from 29.2 to 101.1 days; 1b-H7 was distributed in varieties in MGIII and late-maturing groups, and the Ve-R1 of 1b-H7 ranged from 46.6 to 79.4 days, later flowering than that of early-maturing varieties; 1b-H6 appeared in varieties in the maturity groups between MG00 and MGIII, with early flowering, and its flowering time ranged from 25.3 to 34.3 days (Fig. 5, Additional file 6: Table S6 and Additional file 7: Table S7). 1b-P1, a type of GmFT1b protein sequence, appeared in the varieties of 1b-H1 and 1b-H2. The latter was only found in Jindou 39 belonging to MGIV. 1b-P3 appeared in the varieties of 1b-H7 (Fig. 5, Additional file 7: Table S7 and Additional file 10: Table S10). Six haplotypes of GmFT5a were defined in 127 varieties (Fig. 11 and Additional file 9: Table S9). 5a-H3 was the most abundant, accounting for 75 varieties, more than half of the population; 5a-H1 was distributed in varieties with their maturity groups between MG00 and MGV1, and the Ve-R1 of 5a-H1 varied from 24.3 to 49.3 days, except that Diandou 7 belonging to MGV flowered 71.7 days after emergence; 5a-H2 was distributed in varieties with the maturity group MGIII and early-maturing groups, and the Ve-R1 of 5a-H2 varied from 23.1 to 45.6 days; 5a-H4 was distributed in varieties with their maturity groups belonging to MG0000, MG000 and MG00, and the Ve-R1 of 5a-H4 ranged from 22.2 to 27.4 days, flowering comparatively early (Fig. 11, Additional file 6: Table S6 and Additional file 7: Table S7).

By analyzing the LD of polymorphic sites, a small and critical quantity of SNPs and Indels are selected and can define haplotypes, known as tagging haplotype [46]. According to LD analysis and polymorphic site features, some SNPs and Indels were chosen to form tagging haplotype; for example, only 14 polymorphic sites represented the GmFT1b haplotype (Fig. 5), and only five polymorphic sites were used for the GmFT5b tagging haplotype (Fig. 12), whereas the haplotype definition of GmFT1b and GmFT5b contained 76 and 23 polymorphic sites, respectively (Figs. 5 and 12 and Additional file 9: Table S9). Similarly, we only required 16, 13, 2, and 8 polymorphic sites to define tagging haplotype for GmFT2a, GmFT2b, GmFT3a and GmFT3b, respectively, but the haplotype definition of them contained 34, 39, 7, and 17...
Polymorphic sites associated with flowering time

GLM analysis showed that some polymorphic sites were associated with flowering time among GmFT1b, GmFT2a, GmFT2b, GmFT5a, and GmFT5b, which are listed in Additional file 4: Table S4. Fourteen polymorphic sites were in 1b-H1 and 1b-H7 collectively, and they showed a relationship with flowering time at a significance level of \( p < 0.01 \) (Additional file 4: Table S4). Among the sites, an 8 bp fragment inserted in Intron3, namely, Indel2409, was only 4 bp distant from exon4 (Additional file 4: Table S4). s4645 in exon4 appeared in 2a-H9 and showed a relationship with flowering time at a significance level of \( p < 0.01 \) (Additional file 4: Table S4). Eighteen polymorphic sites in 5a-H2 and Indel1914 in 5a-H4 and 5a-H6, as a 49 bp fragment deleted in the 3’UTR, showed a relationship with flowering time at a significance level of \( p < 0.01 \) (Additional file 4: Table S4). The above-mentioned polymorphic sites may have some effect on flowering time and maturity through regulating the transcription of soybean FT family genes or only act as markers associated with those traits.

Discussion

The 10 soybean FT family genes exhibited variations in sequence polymorphism

The flowering time and maturity of the varieties ranged from 22.2 to 120.5 days and from 61.4 to longer than 155.7 days, respectively, indicating a high diversity in these parameters (Additional file 6: Table S6). This finding showed that the soybean population in our study

---

**Fig. 9** GmFT3b linkage disequilibrium analysis, haplotype definition, primary haplotypes distribution in different maturity groups and flowering time (Ve-R1) in the primary haplotypes.

- **a**: Haplotypes of GmFT3b region. The haplotype is shown as a linear combination of alleles. Site combination labeled with gray outline represents tagging haplotype (haplotype composition depends on fewer and critical variation sites).
- **b**: Linkage disequilibrium analysis of GmFT3b.
- **c**: Primary haplotypes distribution of GmFT3b in soybean varieties from different maturity groups.
- **d**: Flowering time (Ve-R1) of the primary haplotypes of GmFT3b. All means in the boxplot include all haplotypes of this gene. The width of the boxplot represents the sample size
was appropriate for the study of genetic factors influencing flowering time and maturity in soybeans.

Ten FT family genes are in the soybean genome [26, 30]. Only a few of the genes have been studied and their function identified in soybean [22, 23, 26, 30, 31, 35]. Liu et al. found GmFT1a functions divergently to delay flowering and maturing, a function significantly different from that of two known flowering promoters GmFT2a and GmFT5a [31]. Therefore, studying their polymorphisms will help us to learn the evolutionary trends and breeding selection effects on the genes. Based on the Sanger sequencing results, the FT genes exhibited variations in all 127 varieties in the aspect of sequence polymorphism. However, the genes had different levels of polymorphism, and GmFT1b, GmFT1a and GmFT2b had the highest levels of genetic variation, whereas GmFT3a and GmFT4 were highly conserved. Consistently, in the relatively conserved coding region, GmFT1b, GmFT1a and GmFT2b had more polymorphisms than those of GmFT3a and GmFT4. Jiang et al. did not find any variations in the coding region of GmFT2a [34], and only a synonymous site has been found from the 31-Soybean Genome Resequence project [49]. In the current study, two nonsynonymous variations were found in the coding region of GmFT2a suggesting that the selected varieties had extensive genetic variation.

**GmFT3b and GmFT5b may have experienced breeding selection in soybean domestication and breeding**

Zhu et al. found that soybean nucleotide diversity was much lower than that in Arabidopsis and maize, with the value of $\theta_w$ only 0.00097, in studying sequence polymorphism of the soybean genome [50]. The $\theta_w$ values of the 10 soybean FT family genes ranged from 0.00029–0.00569 in the aspect of SNPs in this study. GmFT3a ($\theta_w = 0.00073$), GmFT4 ($\theta_w = 0.00029$), GmFT5a ($\theta_w = 0.00084$) and GmFT6 ($\theta_w = 0.0008$) should be conserved after strong selection. GmFT1a ($\theta_w = 0.00035$), GmFT1b ($\theta_w = 0.00569$), GmFT2a ($\theta_w = 0.00133$), GmFT2b ($\theta_w = 0.00253$), GmFT3b ($\theta_w = 0.00169$) and GmFT1b ($\theta_w = 0.00144$) might have experienced weak selection pressure and had high polymorphism in long-term evolution. This finding also reflected that the selected population had a wide genetic base. Because the ten FT family genes are diverse in the control of flowering and maturity, these genes could play concerto roles to elaborately

---

**Fig. 10** GmFT4 linkage disequilibrium analysis, haplotype definition, primary haplotypes distribution in different maturity groups and flowering time (Ve-R1) in the primary haplotypes. **a** Haplotypes of GmFT4 region. The haplotype is shown as a linear combination of alleles. Site combination labeled with gray outline represents tagging haplotype (haplotype composition depends on fewer and critical variation sites); **b** Linkage disequilibrium analysis of GmFT4; **c** Primary haplotypes distribution of GmFT4 in soybean varieties from different maturity groups; **d** Flowering time (Ve-R1) of the primary haplotypes of GmFT4. All means in the boxplot include all haplotypes of this gene. The width of the boxplot represents the sample size.
regulate flowering and maturing and help soybean to adapt to greatly diverse environments and breeding selections. Neutral test indicated that \( \text{GmFT3b} \) and \( \text{GmFT5b} \) had some selected signals and might have experienced breeding selection. The polymorphic sites of \( \text{GmFT3b} \) and \( \text{GmFT5b} \) had a high proportion of strong LD (Figs. 9 and 12), which suggested that \( \text{GmFT3b} \) and \( \text{GmFT5b} \) underwent relevant selection in adaptation to different environments, consistent with Tajima’s D test.

Haplotypes of distinct soybean \( FT \) family genes and their combinations

Haplotype-based analysis is more informative than SNP-based analysis and is more powerful in analyzing the association with phenotypes [51]. By conducting haplotype analysis of the 10 soybean \( FT \) family genes, we found that seven \( FT \) family genes, including \( \text{GmFT1a} \), \( \text{GmFT1b} \), \( \text{GmFT2a} \), \( \text{GmFT2b} \), \( \text{GmFT3b} \), \( \text{GmFT5a} \), and \( \text{GmFT5b} \), had some haplotypes associated with flowering time and maturity. However, the other genes, \( \text{GmFT3a} \), \( \text{GmFT4} \) and \( \text{GmFT6} \), did not exhibit any relations with flowering time and maturity, which could be because the three genes were relatively conservative. \( 1b-H7 \) appeared in varieties in the MGIII and late-maturity groups, and its protein sequence type was \( 1b-P3 \) (Fig. 5). A 138 bp fragment inserted in Intron2 appeared in \( 1b-H7 \) (Fig. 5).

We observed that \( 2a-H9 \) was distributed in the varieties with the maturity groups of MGIX and MGX, and \( 5a-H4 \) appeared in the varieties with the maturity groups of MG0000, MG000 and MG00 (Figs. 6 and 11). The flowering time of soybean with the haplotypes of \( 2a-H9 \) and \( 5a-H4 \) ranged from 76.9 to 117.6 days and from 22.2 to 27.4 days, respectively (Figs. 6 and 11). These haplotypes will be helpful in studying soybean \( FT \) family genes and can be used as markers associated with flowering time and maturity.

Haplotypes will be helpful in studying soybean \( FT \) family genes and can be used as markers associated with flowering time and maturity.

Haplotype combinations of soybean \( FT \) family genes had some relations. \( 1a-H3 \) and \( 1b-H1 \), \( 2a-H1 \) and \( 2b-H6 \), \( 2a-H3 \) and \( 2b-H7 \), and \( 3b-H4 \) and \( 5b-H1 \) were the combinations. These haplotype combinations were mostly in the same varieties. According to their location in the genome and phylogenetic relationships, \( \text{GmFT1a} \) and \( \text{GmFT1b} \), which are located in Chr18, had the closest evolutionary relationship (1011 kb apart); \( \text{GmFT2a} \) and \( \text{GmFT2b} \), which are located in Chr16, had the closest evolutionary relationship (34 kb apart); and \( \text{GmFT3b} \) and \( \text{GmFT5b} \), which are...
located in Chr19, were only 16 kb apart. In the long process of natural selection and domestication, these short distances might result in linkage disequilibrium. As a result of certain selection effects, some polymorphic sites might have been inherited together in the adaptation to some environments.

**Polymorphisms of soybean FT family genes have relations to flowering time**

We chose polymorphic sites that only existed in the haplotypes with relations to flowering time and maturity. Further association analysis was conducted with GLM, and 38 polymorphic sites showed a significant association with flowering time at the level of $p < 0.01$. These polymorphic sites may be subjected to natural and artificial selection in adaptation to diverse environments and have effects on pre- and postflowering photoperiod responses. An 8 bp fragment insertion in Intron3, close to exon4, appeared in 1b-H1 and 1b-H7 collectively and might have an effect on flowering time. This insertion might regulate GmFT1b expression directly or indirectly through changing the alternative splicing of GmFT1b. GmFT2a and GmFT5a coordinately control flowering and enable the adaptation of soybean to a wide range of photoperiodic environments [26]. A nonsynonymous site in exon4 of GmFT2a, namely, s4645, appeared in 2a-H9 and was distributed in the varieties with their maturity groups belonging to MGIX and MGX. The change at the site resulted in an amino acid substitution from nonpolar G169 to acidic D169. Notably, the flowering inhibitor GmFT1a has an amino acid of D169 in the same location, indicating that this amino acid substitution might have an important role in the function of GmFT2a. A 49 bp fragment deleted in the 3′ UTR appeared in 5a-H4 and 5a-H6 and was distributed in the varieties of MG0000, MG000 and MG00. They may affect the function of GmFT2a and GmFT5a in regulating soybean flowering.

**Conclusions**

Soybean varieties from the earliest maturing MG0000 to the latest maturing MGX showed high diversity in flowering time and maturity. The 10 soybean FT family genes exhibited variation in the aspect of genome sequence polymorphism. GmFT3b and GmFT5b had some selected signals and might have experienced breeding selection in the process of natural selection and
domestication. By analyzing the haplotypes, we found that 1a-H3, 1b-H1, 1b-H6, 1b-H7, 2a-H1, 2a-H3, 2a-H4, 2a-H9, 2b-H3, 2b-H4, 2b-H6, 2b-H7, 3b-H4, 5a-H1, 5a-H2, 5a-H4 and 5b-H1 showed correlative relationships with flowering time and maturity; whereas GmFT3a, GmFT4 and GmFT6 did not have any apparent connections. Thirty-eight polymorphic sites showed a significant association with flowering time at the level of p < 0.01.

Additional files

**Additional file 1**: Table S1. Varieties and their respective maturity group and origin. (DOCX 17 kb)

**Additional file 2**: Table S2. Sequence of primers for the 10 FT family genes in soybean. (DOCX 15 kb)

**Additional file 3**: Table S3. PCR reaction conditions for different templates of the FT family genes in soybean. (DOCX 13 kb)

**Additional file 4**: Table S4. General linear model association of polymorphic sites of soybean FT family genes with relation to flowering time and growth duration. (DOCX 23 kb)

**Additional file 5**: Table S5. Polymorphic sites of the 10 soybean FT family genes from the present study. (DOCX 56 kb)

**Additional file 6**: Table S6. Flowering time (Ve-R1) and growth duration (Ve-R7 and Ve-R8) data in Beijing, China, in 2015. (DOCX 21 kb)

**Additional file 7**: Table S7. Haplotypes of the 10 soybean FT family genes in 127 varieties covering 14 maturity groups. (DOCX 38 kb)

**Additional file 8**: Table S8. Polymorphic sites used for defining the haplotypes of the 10 soybean FT family genes. (DOCX 14 kb)

**Additional file 9**: Table S9. Polymorphic site features used for defining the haplotypes of the 10 soybean FT family genes. (DOCX 16 kb)

**Additional file 10**: Table S10. Different types of soybean FT proteins and their haplotype distribution. (DOCX 18 kb)

**Abbreviations**

CTAB: Cetyl trimethyl ammonium bromide; FT: Flowering Locus T; GLM: General linear model; LD: Linkage disequilibrium; MG: Maturity group; NJ: Neighbor-joining; PCR: Polymerase chain reaction; SNP: Single nucleotide polymorphism

**Acknowledgements**

The authors thank Hongchang Jia (Heihe Branch of Hei-longjiang Academy of Agricultural Sciences) and Tingting Mao (Chinese Academy of Agricultural Sciences) for providing some soybean varieties.

**Funding**

This work was funded by the State Key R&D Project of China (2017YFD0101400), China Agriculture Research System (CARS-04), and the CAAS Agricultural Science and Technology Innovation Project. The funding sources had no influence on the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

**Availability of data and materials**

All data supporting the conclusions of this article are included in this published article and its supplementary information files.

**Authors’ contributions**

TH designed the experiments and provided financial supports; SS, TW, CW participated in the design of the study; WS participated in selecting varieties and providing the information of maturity group classification; SZ, BJ, MAAK...
and CZ performed phenotypic data collection; SZ performed genotypic data collection; SZ and BJ analyzed the data; BJ, SZ, and TH wrote the manuscript; BJ and TH revised the manuscript. All authors reviewed and approved the final manuscript.

Ethics approval and consent to participate
All the plant materials used in the current study were collected from the Institute of Crop Sciences, the Chinese Academy of Agricultural Sciences, which are public and available for non-commercial purpose. No specific permits were required for the field studies described here. The study area is not privately owned or protected in any way, and the field studies did not involve endangered or protected species. Experimental researches on this study comply with institutional, national and international guidelines.

Consent for publication
Not applicable.

Competing interests
The authors declared that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 November 2018 Accepted: 28 February 2019
Published online: 20 March 2019

References
1. Hartwig EE. Growth and reproduction characteristics of soybean grown under short-day conditions. Crop Sci. 1970;10:47–53.
2. Gai J, Wang Y, Zhang M, Wang J, Chang R. Studies on the classification of maturity groups of soybeans in China. Acta Agron Sin. 2001;27(3):285–92.
3. Jia H, Jiang B, Wu C, Lu W, Hou W, Sun S, Yan H, Han T. Maturity group classification and maturity locus genotyping of early-maturing soybean varieties from high-latitude cold regions. PLoS One. 2014;9(4):e94139.
4. Jiang B, Nan H, Gao Y, Tang L, Yue Y, Li S, Ma L, Cao D, Sun S, Wang J, Wu C, Yuan X, Hou W, Kong F, Pan H, Xia Z, Qiao X, Ma Q, Han T, Nian H. A single nucleotide deletion in J encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. Mol Plant. 2017;10(4):656–8.
5. Lu S, Zhao H, Hu Y, Liu S, Nan H, Li X, Fang C, Cao D, Shi X, Kong L, Su T, Zhang F, Li S, Wang Z, Yuan X, Cober ER, Weller JL, Liu B, Hou X, Tian Z, Kong F. Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. Nat Genet. 2017;49(11):1559–61.
6. Song H, Jia Z, Cao D, Jiang B, Wu C, Hou W, Tabata S, Kitamura K, Harada K. A map-based cloning strategy employing a residual heterozygous line reveals that the GmFT2a gene is involved in soybean maturity and flowering. Genetics. 2011;188(2):405–7.
7. Watanabe S, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K, Harada K. A map-based cloning strategy employing a residual heterozygous line reveals that the GmFT2a gene is involved in soybean maturity and flowering. Proc Natl Acad Sci U S A. 2012;109(32):2155–64.
37. Wigge PA. FT, a mobile developmental signal in plants. Curr Biol. 2011;21(9):374–8.
38. Schwartz C, Balasubramanian S, Wathmann N, Michael TP, Lempe J, Sureshkumar S, Kobayashi Y, Malofeev JN, Brovelli JD, Chory J, Weigel D. Cis-regulatory changes at FLOWERING LOCUS T mediate natural variation in flowering responses of Arabidopsis thaliana. Genetics. 2009;183(2):223–32.
39. Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto K. Variations in H1d proteins, H1d promoters, and H1dt expression levels contribute to diversity of flowering time in cultivated rice. Proc Natl Acad Sci U S A. 2009;106(11):4555–60.
40. Fehr WR, Caviness CE. Stages of Soybean Development. Special Report 80 1977, Cooperative extension service, agriculture and home economic Experiment Station. Ames: Iowa State University, p. 1–11.
41. Chen Q, Liu C, Lü D, He J. The basic principle of DNA extraction from soybean. J Northeast Agric Univ. 2004;35(2):254–6.
42. Swindell SR, Plasterer TN. SEQMAN. Contig assembly. In: Swindell SR, editor. Sequence Data Analysis Guidebook. Methods In Molecular Medicine, vol. 70; 1997. p. 75–89.
43. Larkin MA, Blackshields G, Brown NF, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947–8.
44. Hall T. BioEdit, version 7.0.9. Carlsbad: Computer program and documentation, Ibis Biosciences; 2007. http://www.mbio.ncsu.edu/BioEdit/ bioedit.html
45. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.
46. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
47. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5.
48. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. Nat Genet. 2001;29:233–7.
49. Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, Li MW, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Wang J, Sun SS, Zhang G. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet. 2010;42(12):1053–9.
50. Zhu YL, Song QJ, Hyten DL, Yan Tassell CP, Matsumall BK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB. Single-nucleotide polymorphisms in soybean. Genetics. 2003;163(3):1123–34.
51. Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol. 2002;5(2):94–100.