Rapid Excavating a *FLOWERING LOCUS T*-Regulator *NF-YA* Using Genotyping-by-Sequencing

Shichen Li

Northeast Institute of Geography and Agroecology Chinese Academy of Sciences

Tong Su

Northeast Institute of Geography and Agroecology Chinese Academy of Sciences

Lingshuang Wang

Northeast Institute of Geography and Agroecology Chinese Academy of Sciences

Kun Kou

Northeast Institute of Geography and Agroecology Chinese Academy of Sciences

Lingping Kong

Guangzhou University

Fanjiang Kong

Guangzhou University

Sijia Lu

Guangzhou University

Baohui Liu

Guangzhou University

Chao Fang (✉ fangchao@gzhu.edu.cn)

Guangzhou University

Research Article

Abstract

Soybean [*Glycine max* (L.) Merrill] is one of the most important crop plants in the world as an important source of protein for both human consumption and livestock fodder. Soybean flowering time is beneficial to the improvement of soybean yield. Therefore, finding new QTLs and further identifying candidate genes associated with various flowering time are fundamental approaches in enhancing the yield of soybean. In this study, a set of 120 recombinant inbred lines (RILs) which developed from a cross of two soybean cultivars, Suinong4 (SN4) and ZK168, were genotyped by genotyping-by-sequencing (GBS) approach and phenotyped to expand the cognitive of flowering time (R1) by Quantitative Trait Loci (QTL) analysis. Eventually, we detected three stable QTLs related to R1 separately located on chromosome 14, 18, and 19 under long-day conditions. The candidate genes of the three QTLs were predicted, and association analysis of the candidate genes related to flowering time was carried out. Moreover, a transient transfection assay was performed and showed that a candidate gene of the QTL on chromosome 19, *GmNF-YA21* (*Nuclear factor YA21*), might affect flowering by suppressing the expression of *GmFTs*. QTLs detected in this study will provide fundamental resources for finding candidate genes and clarify the mechanisms of flowering which would be helpful for breeding novel high-yield soybean cultivars.

Introduction

Soybean is a major legume crop in the world for its use as edible oil and a main source of high-quality protein for humans (Hoeck et al., 2003). Breeding high-yield soybean cultivar is an ongoing aim of breeders (Yin et al., 2018). An appropriate flowering time is important for increasing yield in soybean, thus identifying flowering related genes are beneficial to increase soybean yield. Time of flowering is also critical to accommodate to different latitudes in plants. In many a crop plants, breeders have explored genetic variations in photoperiod sensitivity and flowering time to adapt crops to a wide range of latitudes (Thakare et al., 2010). Therefore, identification of photoperiod insensitive genes and improvement of cultivars with a flowering response adapted to the different regions is an important subject for broad adaptation of soybean (Zhao et al., 2018).

Soybean is a typical photoperiod sensitive plant, daylength greatly influence its flowering time and reproductively agronomic character (Tasma et al., 2001), soybean can be grown over a wide latitude range, mainly depending on the regulation of photoperiod related genes. Up to now, 14 major genes or loci correlated to flowering time and maturity [*E1* (Bernard, 1971; Xia et al., 2012) and *E2* (Bernard, 1971; Watanabe et al., 2011), *E3* (Buzzell, 1971; Watanabe et al., 2009), *E4* (Buzzell and Voldeng, 1980; Saindon et al., 1989; Liu et al., 2008), *E5* (McBlain and Bernard, 1987), *E6* (Bonato and Vello, 1999; Fang et al., 2020), *E7* (Cober and Voldeng, 2001), *E8* (Cober et al., 2010), *E9* (Kong et al., 2014; Zhao et al., 2016), *E10* (Samanfar et al., 2016), *E11* (Wang et al., 2019), *J* (Ray et al., 1995; Lu et al., 2017), *Tof11* and *Tof12* (Lu et al., 2020)] have been reported in soybean. *E1*, *E2*, *E3*, *E4*, *E8* and *E10* delay flowering while promote early flowering in their recessive alleles (Zhang et al., 2007; Langewisch et al., 2014; Langewisch et al., 2017; Bernard et al., 1971; Xia et al., 2012; Cober et al., 2001; Watanabe et al., 2011; Buzzell et al., 1980;

Watanabe et al., 2009; Xu et al., 2013; Samanfar et al., 2016; Lin et al., 2020). Various allelic combinations of *E1-E4* genes are the foundation of adaptability of soybean in high latitude region (Jiang et al., 2014). Although *E1-E4* have been widely used in breeding at different latitudes, latitude adaptability is not entirely explained and new latitude adaptation genes remain to be explored. Two long juvenile traits, *E6* and *J* has been identified to extend vegetative growth under short day environments (Hartwig et al., 1979; Ray et al., 1995; Lu et al., 2017; Li et al., 2017; Fang et al., 2020;). *Tof11* and *Tof12* suppress *FT* genes expression and inhibit flowering under LD (Long day) conditions. The geographic distribution of *Tof11* and *Tof12* alleles variation enrichment shows compactly associated with increased latitude, suggesting that it may allow for gradual expansion and improvement of the northern limits of early soybean cultivation (Lu et al., 2020).

In addition to these genes, there are many genes also play vital roles in soybean flowering, such as the *FLOWERING LOCUS T* (*FT*), *CONSTANS* (*CO*), *APETALA1* (*AP1*), *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*). Among these genes, *FT* gene family, a major gene family of flowering time control integration has been extensively researched (Takeshima et al 2016; Ogiso-Tanaka et al., 2019; Sun et al., 2019). *GmFT2a* and *GmFT5a* are considered as the major functional members in *FT* family and *GmFT2a* was identified as the causal gene for the *E9* locus (Kong et al., 2010; Kong et al., 2014; Zhao et al., 2016). *NF-Y* is a category of highly conserved transcription factors, which was composed of the *NF-YA*, *NF-YB*, and *NF-YC* subunits (Sinha et al., 1995; Sinha et al., 1996). Original reports indicated that *NF-YA* can negatively regulate flowering due to some *NF-YA* genes overexpression plants caused late flowering (Wenkel et al., 2006; Xu et al., 2014). Subsequently, some studies have been conducted to confirm that *NF-YA* can positively regulate flowering by directly combined to the *FT* promoter in the photoperiod-dependent flowering pathway (Siriwardana et al., 2016). In addition to the factors mentioned above, soybean stem growth habit is also a key adaptation and agronomic trait that directly affects flowering time, node production and ultimately affects soybean yield (Bernard, 1972; Specht et al., 2001; Heatherly and Smith, 2004). It has been confirmed by previous study that *GmTFL1b* is a candidate for *Dt1* by Mapping analysis (Liu et al., 2010). More recently, *Dt1* has been identified to interact with a bZIP transcription factor *FDc1* to repress *AP1* expressions, thereby to regulate flowering (Yue et al., 2021). Although plentiful research has been hurled to understand the molecular mechanism of photoperiod regulation of soybean flowering pathway, soybean flowering mechanism has not been completely understood, many genes related to flowering remain to be uncovered (Kong et al., 2018).

In this study, two soybean cultivars (SN4 and ZK168) with the same *E1*, *E2*, *E3*, *E4* and *Dt1*, *Dt2* genotypes were selected for hybridization to generate population, excluding the influence of known major genes on flowering time. Providing conditions for us to further discover new genes that regulate flowering time. We here used the GBS approach to construct genetic map of the RIL population, and finally three stable QTLs were detected by the ICIM method of flowering time. In addition, we presumed the most possible candidate genes of these QTLs and briefly analyzed the geographic distributions of the candidate genes. Moreover, we considered the candidate gene on chromosome 19 as *GmNF-YA21* (*Nuclear factor YA21*), a homologous gene of Arabidopsis *NF-YA10*. Transient transformation assays were used to detect the

regulation relationship between *GmNF-YA21* and *GmFT2a/GmFT5a*, and we investigated that *GmNF-YA21* might regulate flowering by inhibiting the expression of *GmFTs*.

Materials And Methods

Plant materials

S4-168, the $F_{6:8}$ RIL population, consisting of 120 progenies was used in this study. The S4-168 RIL population was developed by the single-seed descent method and derived from a cross between two soybean cultivars SN4 (*e1-as/e2/e3/E4/Dt1/dt2*) and ZK168 (*e1-as/e2/e3/E4/Dt1/dt2*). SN4 is a breed in north, China and ZK168 is a *e3*-NIL (near isogenic line) of Harosoy.

Plant cultivation

The F6:7 RIL and two parents were sown in the experimental field in Harbin (45°43'N, 126°45'E), China, in May, 2018 while the F7:8 RIL and two parents were sown in the same field in May, 2019. Both two years RIL population and the parental lines seeds were separately sown with the length of 4 m a row, row spacing of 60 cm, and a space of 20 cm between each plant. We sowed about 15 plants on each row. The data on the developmental stages, including Ve (emergence), R1 stage (beginning bloom) (Fehr et al., 1971), were recorded.

DNA extraction

A piece of fresh trifoliate leaf was collected from each parental and RIL individual at the V2 stage and the samples were stored in the -80 degrees refrigerator. Genomic DNA was extracted from these leaf samples using the NuClean PlantGen DNA Kit (CWBIO, Beijing, China). The quality of sample DNA were tested by one percent concentration agarose gel electrophoresis and the concentrations of these DNA were measured using NanoDrop 2000 instrument (ThermoScientific, Wilmington, DE, United States).

Genotyping by high-throughput sequencing

Extracting genomic DNA from each leaf sample for resequencing, and sequencing libraries are based on the method reported in Cheng et al (2015). GBS technology was used for the S4-168 population resequencing. High-coverage sequencing was used to SN4 and ZK168, while low-coverage sequencing was applied to identify the genotypes of SNPs of the RIL population according to the reference polymorphic loci of two parents (Huang et al., 2009; Davey et al., 2013s).

Genetic map construction

The selected polymorphic SNP markers were assigned to different chromosomes by alignment against the reference genome (http://phytozome.jgi.doe.gov). SNPs analysis to detect segregation distortion using the Chi-square (χ2) test. Markers were distributed in 20 linkage groups (or chromosomes) based on

the physical position. The genetic distances between each marker were estimated to construct the high-density genetic map.

QTL analysis

The ICIM mapping method in the QTL IciMapping software (Meng et al., 2015) was used to detect QTLs for each year. Permutation test (PT) was used for the calculation of LOD thresholds at a significance level of $P < 0.05$, N= 1,000. The QTL was considered as a significant QTL when its LOD score was higher than the threshold in two years. In addition, we also recorded QTLs which LOD score values between 2.5 and the LOD thresholds.

Candidate gene identification

Soybean genomic data and the reference genome sequences of Wm82.a2.v1 (Schmutz et al., 2010) were downloaded from the Phytozome website (http://phytozome.jgi.doe.gov). The candidate genes in QTL intervals were categorized using Gene Ontology (GO) analysis. Gene ontology (GO; http://geneontology.org/) databases and WE GO (http://wego.genomics.org.cn) were used to obtain detailed pathway, gene ontology and annotation information. We organized the functional annotation information for all genes within the QTL interval which had been identified in both two years and selected the most likely candidate genes within variations exist in two parents of the GBS analysis consequence.

Haplotype calling and association analysis

Resequencing data of the 1295-accession panel, VCF files and the 424 accessions phenotypic data of flowering time used in this study were obtained from Lu et al (2020). VCF files processing using the VCFtools software (v.0.1.16) (Danecek et al., 2011). The significance of association analysis was calculated by IBM SPSS software 20 (https://www.ibm.com/analytics/spss-statistics-software) with Ducan's multiple range test.

RNA extraction

Ultrapure RNA Kit (CWBIO, China) was used for extracting total RNA from the ternately compound leaf of SN4 and ZK168. PrimeScript RT Reagent Kit (Takara, Japan) was used for the cDNA Synthesis. Concentrations of these cDNA were measured using NanoDrop 2000 instrument (ThermoScientific, Wilmington, DE, United States).

Transient transfection assays

Around 3 kb promoter sequences of *FT2a* and *FT5a* were amplified from the cDNA extracted from leaves of Williams82 and introduced into pGreen0800-LUC/REN to generate the pFT2a-LUC/REN and pFT5a-LUC/REN reporters. *NF-YA21* coding regions were amplified from the cDNA extracted from leaves of SN4 and ZK168. p35S:NFYA21-Flag (*GmNFYA21*-SN4-type and *GmNFYA21*-ZK168-type) and p35S: Flag were used as the effectors. All the reporters and effectors were transformed into A. tumefaciens strains

GV3101, respectively. The transformed effector proteins were mixed with A. tumefaciens containing the reporter structure and co-permeated into leaves. Each combination used at least three leaves from individual N. benthamiana. Renilla Luciferase Assay System (Promega)was used to measure the quantitative values of LUC and REN activities. The primers used for amplification are listed in Supplementary Table 14.

Results

Phenotype variation in parents and the RIL population

The trait of flowering time of each RIL population individual and two parents were all recorded in two years (Harbin in 2018 and 2019) (Supplementary Fig. 1 and Supplementary Table 1). As shown in Supplementary Fig. 1, a transgressive segregation was observed in this population, which indicates the polygenic control of the soybean flowering time. Correlation analysis was conducted on the two-year data of all traits recorded of the RIL population, the result shows that R1 were significantly correlated in two years. The absolute value of skewness of the mean value of R1 trait in the RIL population across each year was <1, indicating an approximately normal distribution. Therefore, the R1 data of these two years were trustingly used to detect QTLs.

Analysis of sequencing data and construction of a genetic linkage map

Both of the parental cultivar SN4 and ZK168 were resequenced at a higher coverage level of 14,7X and 8.1X individually, to detect SNP markers and call variations between parents. For SN4 and ZK168, a total of 16,182,987,000 and 8,878,524,900 bases were identified, individually. The Q30 ratio of SN4 and ZK168 was 94.66% and 91.35% while the GC content of SN4 and ZK168 was 40.41% and 35.85%, respectively (Supplementary Table 2). Finally, a total of 2942 polymorphic SNP markers were obtained and used in linkage map construction of 20 Chromosomes (Supplementary Figure 2). The genetic distance between the markers was estimated in cM using the QTL IciMapping software (Supplementary Table 3).

QTL mapping of flowering time

We proceeded QTL identification with the two-year phenotypic data of the RIL population. The threshold of the LOD scores for evaluating the statistical significance of QTL effects is 3.43 and 3.56 of two year, respectively. Three QTLs were detected by the ICIM method named as *qR1-L*, *qR1-G* and *qR1-B2* (Fig.1 and Supplementary Table 4). Within these QTLs, *qR1-L* was located on chromosome 19, physical length from 45,564,991 to 48,271,467 of the reference genome Wm82.a2.v1. The LOD scores of *qR1-L* were 6.72 and 4.10 for each year, could explain 12.1−17.36% of the observed PV (phenotype contribution). *qR1-G* was located on chromosome 18 with physical length from 56,210,047 to 57,296,740. The LOD scores of *qR1-G* ranged from 2.78 to 3.65, could explain 4.6% −8.52% of the observed PV. *qR1-B2* was located on chromosome 14, physical length from 34,291,536 to 39,844,217. The LOD scores of *qR1-B2* ranged from 3.57 to 4.39. The LOD of *qR1-L* and *qR1-B2* were higher than the LOD threshold in both two year whereas the LOD of *qR1-G* was only greater than the LOD threshold in 2019. The LOD value of *qR1-G* in 2018 was

2.78, not exceeds the threshold, but due to the LOD value was higher than 2.5, we also considered it as a credible QTL. In addition, two major QTLs of R1 trait with the max LOD score of 6.12 and 8.57, respectively, on chromosome 4 were also detected, since these QTLs only could be detected in 2019 but could not detected in 2018, we considered them as unstable QTLs without further analyze the candidate genes of them (Supplementary Table 5).

Candidate Gene Prediction of *qR1-G*

qR1-G within the physical interval of 56,210,047 - 57,296,740 on chromosome 18. The position interval of the QTL involved 83 genes could be functionally annotated by GO analysis (Supplementary Table 6). Among them, 49, 18 and 79 were functionally annotated to the categories of cellular components, molecular functions and biological processes, individually (Supplementary Figure 3). A single gene *Glyma.18G281400* was found to be related with transcription regulation activity. In addition, we found that the interval includes 15 genes with nonsynonymous mutations between parents based on the resequencing data (Supplementary Table 7). All the 15 genes functional annotations were shown on Supplementary Table 8. Especially, *Glyma.18G281400* not only has variation between the parents but also the single gene related to transcription regulation activity.

Glyma.18G281400 is an AP2-EREBP (ethylene-responsive element-binding protein) transcription factor. EREBPs with the AP2 domain play a role in the regulation of plant development (Zhao et al., 2006; Wang et al., 2014; Kuluev et al., 2015). Previous research has found that two AP2 domain-encoding genes, *SCHLAFMÜTZE* and *SCHNARCHZAPFEN*, repress flowering in Arabidopsis thaliana (Schmid et al., 2003). Therefore, *Glyma.18G281400* was considered as a candidate gene for *qR1-G* and was conducted the correlation analysis of flowering time. At first, we analyze the variation in the *Glyma.18G281400* coding sequence using the 1295-accession panel, then defined seven haplotypes with three disparate nonsynonymous variation alleles. However, the variation existence in two parents of *Glyma.18G281400* was rare in the natural variations (Fig. 2a and Supplementary Table 7), while, H3 (Haplotype 3) with nine bases inserted after the 665th bp and a SNP (Single nucleotide polymorphism) at the 430th bp of coding region was associated with flowering time (Fig. 2). The 424-accession panel flowering time analysis between H3 and H7 showed no significant difference in Guangzhou 2019, while there were significant differences in the other regions (Harbin 2019, Wuhan 2019, Zhengzhou 2019, Zhengzhou 2018, Hefei 2018) (Fig. 2b-g). This suggests that *Glyma.18G281400* may be associated with flowering time in the high latitude and middle latitude rather than the low latitude regions in China.

Candidate Gene Prediction of *qR1-B2*

qR1-B2 within the physical interval of 34,291,536 - 39,844,217 was located on chromosome 14. We compared coding sequence between the parents in the QTL interval and found 23 genes have differences on the exons (Supplementary Table 9). Among them, 20 genes were annotated genes, and GO analysis was carried out of these genes (supplementary table 10). 5 were functionally annotated to the categories of cellular components, 18 were functionally annotated to molecular functions and 14 were functionally

annotated to biological processes (Supplementary Figure 4). Three genes of them were related to transcription regulation activity, namely, *Glyma.14G159400*, *Glyma.14G160600* and *Glyma.14G161900*. Among these genes, *Glyma.14G159400* is a transcription factor encoding a jumonji (jmjC) domain-containing protein. Previous report has elaborated that a jmjc domain-containing protein encoding gene, *Se14*, plays a key role under long-day suppression of flowering in rice (Takayuki et al., 2014). Therefore, we presumed the *Glyma.14G159400* as a candidate gene of *qR1-B2*.

In order to verify whether the gene *Glyma.14G159400* is responsible for *qR1-B2*, we conducted a correlation analysis between the genotypes and flowering time phenotypes. There are nine haplotypes in *Glyma.14G159400* of soybean 1295-accession panel, and we found that the allele of *Glyma.14G159400* from SN4 (3206th bp-T) was also consist in generally natural varieties (Fig. 3a). The origins of haplotypes of *Glyma.14G159400* were analyzed, the most common haplotype H9 (3206th bp-C type) was possibly originated from H6 (3206th bp-T type), and the proportion of wild was greatly reduced in the transform from H6 to H9. Then the wild type fade away in the process of further differentiation from H9 (Fig. 3b). Additionally, in all the six locations, 3206th bp-T type subgroup showed significantly later flowering time than 3206th bp-C type (Fig. 3c-3h). These indicate that the mutation in codon 3206th bp of *Glyma.14G159400* could lead to a variation of flowering time in soybean and implying a key role of *Glyma.14G159400* in control of flowering time across diverse genetic backgrounds and environmental conditions.

Candidate Gene Prediction of *qR1-L*

qR1-L was detected within the physical interval of 45,564,991 - 48,271,467 on chromosome 19. The well-known flowering-related gene *E3* was located in this range, but sequence comparison showed that there was no difference in the *E3* gene between two parents. Meanwhile, the growth period gene *Dt1*, which has been reported to regulate flowering time in soybean recently (Yue et al., 2021) was located closely to the interval, but no sequence difference was found between the parents. There may be a new gene regulates the variation of flowering time for *qR1-L*. Analysis of resequencing data revealed that, there were 131 genes with non-synonymous mutations or frameshift variations between two parental cultivars (Supplementary Table 11). Among them, 87 genes have been annotated by GO (Supplementary Table 12). The result of GO analysis showed that, 24 were functionally annotated to the categories of cellular components, 78 were functionally annotated to molecular functions and 47 were functionally annotated to biological processes (Supplementary Fig. 5). Thereinto, *Glyma.19G200800,* named *GmNF-YA21,* was the only one gene related to transcription regulation activity by GO analysis. *GmNF-YA21* is a homologous gene of Arabidopsis *NF-YA10,* previous reports have demonstrated that *NF-YA* can regulate flowering time in Arabidopsis (Wenkel et al., 2006; Xu et al., 2014; Siriwardana et al., 2016). Consequently, we presumed the *GmNF-YA21* as the candidate gene of *qR1-L*.

Analysis of variation in the *GmNF-YA21* coding sequence using the 1295-accession panel defined eight *GmNF-YA21* haplotypes, including five distinct nonsynonymous variation alleles (Fig. 4a). In the eight haplotypes, the non-synonymous mutation caused by the G-T mutation at the 202nd base was consisted

with the parent variation of the population. Thereafter, we compared the GmNF-YA21 proteins to find it homologous genes in soybean, Arabidopsis thaliana (https://phytozome.jgi.doe.gov/pz/portal.html) as well as several legumes (https://www.legumeinfo.org), and further reproduced the phylogenetic tree (Supplementary Fig. 6). Four *GmNF-YA10* genes were clustered into the clade, which was further subdivided into two groups (Supplementary Fig. 6a). The nonsynonymous variation (the 68th Amino acid changes from Ala to Ser) caused by the 202nd bp mutation was compared in the *GmNF-YA21* homologous genes, and the result shows that the locus was all S (Ser) in the homologous genes of different legumes, soybean and Arabidopsis thaliana while A (Ala) merely in the *GmNF-YA21* (Supplementary Fig. 6b). Combined with the results of protein homologous alignment and haplotypes origin analysis, we found that all the haplotypes originated from the H2 by a SNP and the common H8 was differentiated from H7 (Supplementary Fig. 6b and Fig. 4b). Subsequently, we examined the variations of *GmNF-YA21* associated with flowering time in soybean 424-accession panel at six field sites in different latitudes in China (Lu et al., 2020). Significant associations were identified between the 202nd base allele variation in the coding region in flowering time (Fig. 4d-j). The result shows that *GmNF-YA21* is related to flowering time and the 202nd-G allele leads to early flowering while the 202nd-T allele contributes to late flowering phenotype. Given that *GmNF-YA10* is presumed to be a flowering related gene.

We next examined how the distributions of the major *GmNF-YA10* alleles within the subset of Chinese accessions in different geographic latitudes. Within both landraces and improved cultivars, the proportion of the early flowering allele 202nd-G gradually increased from low latitude to high latitude (Fig. 4c). We further conducted analysis for the 202nd-G type and 202nd-T type in wilds, landraces as well as cultivars using the 1295-accession panel data. The result shows that early flowering allele 202nd-G gradually increased from wild to improve cultivars, indicating strongly favored in landraces and subsequently widely utilized in modern breeding (Fig. 4d). Taken together, these results mainly suggested that *GmNF-YA10* is definitely a latitudinal adaption gene, and the early flowering allele (202nd-G allele) of *GmNF-YA21* is already been used in north region breeding.

Previous study has identified that *NF-YA* can directly bind the distal CCAAT box in the *FT* promoter and are positive regulators of flowering in Arabidopsis thaliana (Siriwardana et al., 2016). In order to further verify the gene function of *GmNF-YA21* in soybean, we performed transient transfection assays to experimental verify the relationship between *GmNF-YA21* and *GmFT2a/GmFT5a*. The result shows that the *LUC* activity driven by *GmFT2a/GmFT5a* promoter was suppressed by both *GmNF-YA21*-SN4 type and *GmNF-YA21*-ZK168 type, indicating that *GmNF-YA21* possesses the ability to repress *GmFT2a/GmFT5a* expression (Fig. 5). Even though both SN4 and ZK168 type can inhibit the expression of *GmFT2a/GmFT5a*, there are significant differences in the inhibition degree of *GmFT2a/GmFT5a* between the two types (Fig 5b and Fig.5d). The inhibition effect of SN4-type was stronger than that of ZK168-type, which was also consistent with 424 accessions flowering time association analysis (Fig. 5 and Fig. 4e-4j). The result suggested that, although a fine-mapping is still need, *GmNF-YA21* is the most

likely candidate gene responsible for *qR1-L*. Furthermore, *GmNF-YA21* affects flowering by inhibiting the expressing of two florigen genes, *GmFT2a* and *GmFT5a*.

Discussion

Soybean plays a key role in global food security and agricultural sustainability. The discovery of new agronomic trait genes is helpful to increase soybean yield (Zhang et al., 2004). Flowering time is an important element affecting the length of the whole growth period of soybean and a key factor for breeding varieties with wider geographical adaptablity in soybean. Researching polymorphism of flowering variation is not only a signature of artificial selection breeding, but also possess great significance for domestication, diversity and improvement of soybean (Zhao et al., 2008).

Molecular Markers, applied to find QTLs and performed numerous molecular biology experiments, are widely used in plant (Sonah et al., 2013; Semagn et al., 2006). Among the various marker types in utilization, single nucleotide polymorphisms (SNPs) are enriched in the genome, hence more appropriate for application (Rafalski, 2002). However, it is relatively lengthy and costly for the high-throughput genotyping of a mass of SNPs (Sonah et al., 2013). To improve the inadequacies, next generation sequencing (NGS) technologies have emerged (Pareek et al., 2011). Among multitudinous NGS technologies, GBS method which provides a greatly simplified library production, is more suitable for sequencing a large number of individuals or lines (Elshire et al., 2011). Because of the advantages of speedy and economical, GBS is now widely used in sequencing the genomes of plants, including soybean. In this study, the effective GBS approach were also used to detect QTLs of flowering time. We detected a total of three QTLs related to R1 named *qR1-L*, *qR1-G* and *qR1-B2* on chromosome 19, 18, 14, respectively. The adjacent regions of these QTLs have been reported (Supplementary Table 13), yet the candidate genes of them have not been identified. Therefore, we predicted the candidate genes responsible for these QTLs and performed briefly functional verification to a candidate gene of *qR1-L*.

In the QTL interval of *qR1-G*, 15 genes were detected with nonsynonymous mutations between parents based on the resequencing data (Supplementary Table 8). Thereinto, *Glyma.18G281400* not only had variation between the parents but also had relation with transcription regulation activity, maybe a candidate gene of *qR1-G*. *Glyma.18G281400* contains an AP2-EREBP domain which belongs to a superfamily of plant specific transcription factors and certain genes have been identified to be related with flowering in Arabidopsis (Weigel, 1995; Okamuro et al., 1997; Schmid et al., 2003). Combined with the results of association analysis of flowering time in 424-accession panel, we further proved the possibility of the candidate gene for *Glyma.18G281400* (Fig. 2). The same analytical methods were applied to *qR1-B2* and *qR1-L*, and two probable candidate genes, *Glyma.14G159400* and *Glyma.19G200800* (*GmNF-YA21*), were detected, respectively.

Nuclear factor Y (*NF-Y*), also known as CCAAT binding factor (CBF), is composed of three subunits: *NF-YA*, *NF-YB* and *NF-YC*, has been reported specifically binding the evolutionarily conserved CCAAT motifs (Hou et al., 2014; Mantovani et al., 1999; Kusnetsov et al., 1999). FT protein, containing a CCAAT box, is a

main mobile hormone that perceives the photoperiod signal in leaves, and then transfers to the shoot apex to promote the floral transition (Lin et al., 2007; Corbesier et al., 2007; Mathieu et al., 2007). NF-YA (NF-YA2) have been proved can directly bind the distal CCAAT box of the FT promoter in complex with NF-YB/NF-YC, and positively regulate flowering in Arabidopsis (Siriwardana et al., 2016). Two homologous genes of *FT*, *GmFT2a* and *GmFT5a* were identified coordinately regulate flowering in soybean (Kong et al., 2010). Here we provide evidence, in the form of transient transfection assays, that *GmNF-YA21* act to inhibit the expression of *GmFT2a* and *GmFT5a* under LD conditions in soybean (Fig. 5). Previously, *NF-YA10* and it homologous genes were researched mainly focused on salinity stress (Ma et al., 2015; Zhang et al., 2020), drought stress (Ma et al., 2015; Yu et al., 2020) and leaf growth (Zhang et al., 2017), the result of our research provides a potential research basis to study the function of *NF-YA10* in soybean flowering.

The ancestors of most flowering plants tended to double their genomes (Masterston, 1994). The frequency of polyploid formation in flowering plants is 1 in 100,000 individuals, and about 2–4% of speciation events are associated with polyploid formation. Therefore, many plants, especially domesticated crop species, are polyploid (Blanc and Wolfe, 2004; Cui et al., 2006). As a flowering plant, soybean developed numerous homologous genes during the progress of genomic replication. However, functional differentiation and weakening may occur among the homologous genes. There is great mass of *NF-Y* homologous genes in soybean. Although the function of these genes in Arabidopsis has been confirmed, the actual function of *GmNF-YA21* (*Glyma.19G200800*) in soybean remains to be studied, and it is still unknown whether it is a functional or silenced gene covered by its homologous genes. In this study, we confirmed that the GmNF-YA21 protein can negatively regulate *GmFT*, and the detection of this QTL also proved that *GmNF-YA21* may be a real functional gene. At the same time, it also provides the foundation for the further study of this gene.

We found that *GmNF-YA21* may be a latitude adaptation gene by geographical latitude analysis. The geographic distribution of the 202nd -G allele of *GmNF-YA21* coding region shows enrichment in high latitudes (Fig. 4b), suggesting that *GmNF-YA21* was selected as a latitude adaptation gene. This character of *GmNF-YA21* can be utilized in high latitude breeding to broaden adaptation and increase yield. In addition, the allele of the early flowering 202nd-G extensively increased from soja to landraces and improved cultivar (Fig. 4b and 4c), indicating that this allele has been selected during the breeding process. Together with the character of its latitude adaptation, the observations provide an opportunity for further breeding of high latitude in soybean.

In conclusion, we detected three QTLs related to flowering time and predicted their candidate genes. In particularly, the candidate gene of *qR1-L, GmNF-YA21*, as a latitude adaptability gene, might regulate soybean flowering by affecting the expression of *FT* genes. These results provide theories basis for further understand soybean flowering regulation network and materials for breeding adaptive cultivars in high-latitude.

Declarations

Author contributions

CF designed the experiments. SL, TS, LW, KK and LK carried out the experiments. SL, BL, FK, SL and CF analyzed the data. LS, BL, SL and CF wrote the paper.

Acknowledgments

Author information

Shichen Li and Tong Su contributed equally to this work.

Affiliations

The Innovative Academy of Seed Design, Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China

Shichen Li, Tong Su, Lingshuang Wang, Kun Kou, Fanjiang Kong, Baohui Liu

Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou, China

Lingping Kong, Fanjiang Kong, Sijia Lu, Baohui Liu, Chao Fang

University of Chinese Academy of Sciences, Beijing, China

Shichen Li, Tong Su, Lingshuang Wang, Kun Kou, Fanjiang Kong, Baohui Liu

Corresponding authors

Correspondence to Chao Fang or Baohui Liu or Sijia Lu or Fanjiang Kong

References

1. Bernard RL (1971) Two major genes for time of flowering and maturity in soybeans. Crop Sci 11:242–247

2. Bernard RL (1972) Two genes affecting stem termination in soybeans. Crop Sci 12:235–239

3. Blanc G, Wolfe KH (2004) Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. Plant Cell 16:1667–1678

4. Bonato ER (1999) E6, a dominant gene conditioning early flowering and maturity in soybeans. Genet Mol Biol 22:229–232

5. Buzzell RI (1971) Inheritance of a soybean flowering response to fluorescent day length conditions. Can J Genet Cytol 13:703–707

6. Buzzell RI, Voldeng HD (1980) Inheritance of insensitivity to long daylength. Soybean Genet Newslett 7:26–29

7. Fang C, Liu J, Zhang T, Su T, Li S, Cheng Q, Kong L, Li X et al (2020) A recent retrotransposon insertion of J caused E6 locus facilitating soybean adaptation into low latitude. J Integr Plant Biol. https://doi.org/10.1111/jipb.13034

8. Cheng W, Liu F, Li M, Hu X, Chen H, Pappoe F (2015) Variation detection based on next-generation sequencing of type Chinese 1 strains of Toxoplasma gondii with different virulence from China. BMC Genom 16:888

9. Cober ER, Molnar SJ, Charette M, Voldeng HD (2010) A new locus for early maturity in soybean. Crop Sci 50:524–527

10. Cober ER, Voldeng HD (2001) A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. Crop Sci 41:698–701

11. Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I et al (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science 316(5827):1030–1033

12. Cui L, Kerr Wall P, James H et al (2006) Widespread genome duplications throughout the history of flowering plants. Genome Res 16:738–749

13. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLOS ONE 6(5):e19379

14. Fehr WR, Caviness CE, Burmood DT, Pennington JS (1971) Stage of development descriptions for soybeans, Glycine max (L.) Merrill. Crop Sci 11:929–931

15. Hartwig EE, Kiihl RA (1979) Identification and utilization of a delayed flowering character in soybeans for short-day conditions. Field Crop Res 2:145–151

16. Heatherly LG, Smith JR (2004) Effect of soybean stem growth habit on height and node number after beginning bloom in the MIDSOUTHERN USA. Crop Sci 44:1855–1858

17. Hoeck JA, Fehr WR, Shoemaker RC, Welke GA, Cianzio SR (2003) Molecular marker analysis of seed size in soybean. Crop Science 43(1)

18. Hou X, Zhou J, Liu C, Liu L, Shen L, Yu H (2014) Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. Nature communications 5:4601

19. Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

20. Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML (2013) Special features of RAD sequencing data: implications for genotyping. Mol Ecol 22:3151–3164

21. Jiang B, Nan H, Gao Y, Tang L, Yue Y, Lu S, Ma L, Cao D, Sun S, Wang J, Wu C, Yuan X, Hou W, Kong F, Han T, Liu B (2014) Allelic combinations of soybean maturity loci *E1, E2, E3* and *E4* result in diversity of maturity and adaptation to different latitudes. Plos One 9: e106042

22. Kong F, Liu B, Xia Z, Sato S, Kim BM et al (2010) Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. Plant physiol 154:1220–1231

23. Kong F, Nan H, Cao D, Li Y, Wu F, Wang J, Lu S, Yuan X, Cober ER, Abe J, Liu B (2014) A new dominant gene coditions early flowering and maturity in soybean. Crop Sci 54:2529–2535

24. Kong L, Lu S, Wang Y, Fang C, Wang F, Nan H et al (2018) Quantitative trait locus mapping of flowering time and maturity in soybean using next-generation sequencing-based analysis. Front Plant Sci 9:995

25. Kuluev B, Avalbaev A, Nurgaleeva E, Knyazev A, Nikonorov Y, Chemeris A (2015) Role of AINTEGUMENTA-like gene NtANTL in the regulation of tobacco organ growth. J Plant Physiol 189:s11–23

26. Kusnetsov V, Landsberger M, Meurer J, Oelmuller R (1999) The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids. J Biol Chem 274(50):36009–36014

27. Langewisch T, Lenis J, Jiang G-L, Wang D, Pantalone V, Bilyeu K (2017) The development and use of a molecular model for soybean maturity groups. BMC Plant Biol 17(1):91

28. Langewisch T, Zhang H, Vincent R, Joshi T, Xu D, Bilyeu K (2014) Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. PLoS One 9(4):E94150

29. Li B, Mao X, Li A, Wang J, Chang X, Hao C et al (2016) Two Novel AP2/EREBP Transcription Factor Genes TaPARG Have Pleiotropic Functions on Plant Architecture and Yield-Related Traits in Common Wheat. Frontiers in Plant Science 7(313545)

30. Li X, Fang C, Xu M, Zhang F, Lu S, Nan H, Su T, Li S, Zhao X, Kong L, Yuan X, Liu B, Abe J, Cober ER, Kong F (2017) Quantitative trait locus mapping of soybean maturity gene E6. Crop Sci 57(5):2547

31. Lin MK, Belanger H, Lee YJ, Varkonyi-Gasic E, Taoka K, Miura E et al (2007) FLOWERING LOCUS T protein may act as the long-distance florigenic signal in the cucurbits. Plant Cell 19(5):1488–1506

32. Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, Abe J (2008) Genetic redundancy in soybean photo responses associated with duplication of the Phytochrome A gene. Genetics 180:995–1007

33. Liu B, Watanabe S, Uchiyama T, Kong F, Kanazawa A, Xia Z et al (2010) The soybean stem growth habit gene dt1 is an ortholog of Arabidopsis TERMINAL FLOWER1. PLANT PHYSIOLOGY 153(1):198–210

34. Lu S, Dong L, Fang C, Liu S, Kong L, Cheng Q, Chen L, Su T, Nan H, Zhang D, Zhang L, Wang Z, Yang Y, Yu D, Liu X, Yang Q, Lin X, Tang T, Zhao X, Yang X, Tian C, Xie Q, Li X, Yuan X, Tian Z, Liu B, Weller J, Kong F (2020) Stepwise selection on homologous PRR genes controlling flowering and maturity during soybean domestication. Nat Genet 52:428–436

35. Lu S, Zhao X, 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 (2017) Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. Nat Genet 49:773–779

36. Ma X, Li C, Wang M (2015) Wheat NF-YA10 functions independently in salinity and drought stress. Bioengineered 6(4):245–247

37. Ma X, Zhu X, Li C et al (2015) Overexpression of wheat NF-YA10 gene regulates the salinity stress response in Arabidopsis thaliana. Plant Physiology Biochemistry 86:34–43

38. Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene 239(1):15–27

39. Masterston J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264:421–423

40. Mathieu J, Warthmann N, Kuttner F, Schmid M (2007) Export of FT protein from phloem companion cells is sufficient for floral induction in Arabidopsis. Current biology CB 17(12):1055–1060

41. McBlain BA, Bemard RL (1987) A new gene affecting the time of flowering-maturity in soybeans. J Hered 178:68–70

42. Meng L, Li H, Zhang L, Wang J (2015) QTL icimapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J 3:269–283

43. Ogiso-Tanaka E, Shimizu T, Hajika M, Kaga A, Ishimoto M (2019) Highly multiplexed AmpliSeq technology identifies novel variation of flowering time-related genes in soybean (Glycine max). DNA Res 26:243–260

44. Okamuro JK, Caster B, Villarroel R, VanMontagu M, Jofuku KD (1997) The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. Proc Natl Acad Sci USA 94:7076–7081

45. Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52(4):413–435

46. Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5(2):94–100

47. Ray JD, Hinson K, Mankono JEB, Malo MF (1995) Genetic control of a long-juvenile trait in soybean. Crop Sci 35:1001–1006

48. Saindon G, Beversdorf WD, Voldeng HD (1989) Adjusting of the soybean phenology using the E4 loci. Crop Sci 29:1361–1365

49. Samanfar B, Molnar SJ, Charette M, Schoenrock A, Dehne F, Golshani A, Belzile F, Cober ER (2016) Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. Theor Appl Genet 130:377–390

50. Schmid M, Uhlenhaut H, Godard F, Demar M, Bressan R, Weigel D, Lohmann J (2003) Dissection of floral induction pathways using global expression analysis. Development 130(24):6001–6012

51. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

52. Semagn K, Bjrnstad, Ndjiondjop MN (2006) An overview of molecular marker methods for plants. AFRICAN JOURNAL OF BIOTECHNOLOGY 525(25):2540–2568

53. Sinha S, Kim IS, Sohn KY, de Crombrugghe B, Maity SN (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. Mol Cell Biol 16:328–337

54. Sinha S, Maity SN, Lu J, Crombrugghe BD (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. Proc Natl Acad Sci USA 92(5):1624–1628

55. Siriwardana CL, Gnesutta N, Kumimoto RW, Jones DS, Iii BFH (2016) Nuclear Factor Y, subunit A (NF-YA) proteins positively regulate flowering and act through flowering locus T. Plos Genetics 12(12)

56. Sonah H, Bastien M, Iquira E et al (2013) An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. Plos One 8

57. Specht JE, Chase K, Macrander M, Graef GL, Chung J, Markwell JP, Germann M, Orf JH, Lark KG (2001) Soybean response to water: A QTL analysis of drought tolerance. Crop Sci 41:493–509

58. Sun F, Xu M, Park C, Dwiyanti MS, Nagano AJ, Zhu J, Watanabe S, Kong F, Liu B, Yamada T, Abe J (2019) Characterization and quantitative trait locus mapping of late-flowering from a Thai soybean cultivar introduced into a photoperiod-insensitive genetic background. PLoS ONE 14:e0226116

59. Takayuki Y, Hiroki S, Yoshihiro Y, Quan X, Takehito A, Takuji T et al (2014) Se14, encoding a jmjc domain-containing protein, plays key roles in long-day suppression of rice flowering through the demethylation of h3k4me3 of rft1. Plos One 9(4):e96064

60. Takeshima R, Hayashi T, Zhu J, Zhao C, Xu M, Yamaguchi N, Sayama T, Ishimoto M, Kong L, Shi X, Liu B, Tian Z, Yamada T, Kong F, Abe J (2016) A soybean quantitative trait locus that promotes flowering under long days is identified as FT5a, a FLOWERING LOCUS T ortholog. J Exp Bot 67:5247–5258

61. Tasma IM, Lorenzen LL, Green DE, Shoemaker RC (2001) Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. Mol Breeding 8(1):25–35

62. Thakare D, Kumudini S, Dinkins RD (2010) Expression of flowering-time genes in soybean E1 near-isogenic lines under short and long day conditions. Planta 231(4):951–963

63. Wang F, Nan H, Chen L, Fang C, Zhang H, Su T, Li S, Cheng Q, Dong L, Liu B, Kong F, Lu S (2019) A new dominant locus, E11, controls early flowering time and maturity in soybean. Mol Breeding 39:70

64. Wang Y, Wang L, Zou Y, Chen L, Cai Z, Zhang S et al (2014) Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate ENOD40 expression and regulate nodule initiation. Plant Cell 26:4782–4801

65. Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y, Yamanaka N, Takahashi R, Ishimoto M, Anai T, Tabata S, Harada K (2009) Map-based cloning of the gene associated with the soybean maturity locus *E3*. Genetics 182:1251–1261

66. Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N et al (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. Genetics188(2):395–407

67. Weigel D (1995) The APETALA2 domain is related to a novel type of DNA-binding domain. Plant Cell 7:388–389

68. Wenkel S, Turck F, Singer K, Gissot L, Gourrierec JL, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. Plant Cell 18(11):2971–2984

69. Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H et al (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proc Natl Acad Sci USA 109(32):E2155–E2164

70. Xiaoya Lin B, Liu JL, Weller J, Abe F, Kong (2021) Molecular mechanisms for the photoperiodic regulation of flowering in soybean. J Integr Plant Biol. https://doi.org/10.1111/jipb.13021

71. Xu M, Xu Z, Liu B, Kong F, Tsubokura Y, Watanabe S, Xia Z, Harada K, Kanazawa A, Yamada T, Abe J (2013) Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. BMC Plant Biol 13:91

72. Xu M, Zhang L, Li W, Hu X, Wang M, Fan Y et al (2004) Stress-induced early flowering is mediated by miR169 in Arabidopsis thaliana. J Exp Bot 65(1):89–101

73. Yin Z, Qi H, Mao X, Wang J, Hu Z, Wu X et al (2018) QTL mapping of soybean node numbers on the main stem and meta-analysis for mining candidate genes. Biotechnology & Biotechnological Equipment 1–8

74. Yu Y, Bai Y, Wang P et al (2020) Soybean nuclear factor YA10 positively regulates drought resistance in transgenic Arabidopsis thaliana. Environ Exp Bot 180:104249

75. Yue L, Li X, Fang C, Chen L, Yang H, Yang J et al (2021) FT5a interferes with the Dt1-AP1 feedback loop to control flowering time and shoot determinacy in soybean. J Integr Plant Biol. https://doi.org/10.1111/jipb.13070

76. Zhang L, Kyei-Boagen S, Zhang J, Zhang M, Freeland T, Watson C (2007) Modifications of optimum adaptation zones for soybean maturity groups in the USA. Crop Manage 6:1

77. Zhang M, Hu X, Zhu M et al (2017) Transcription factors NF-YA2 and NF-YA10 regulate leaf growth via auxin signaling in Arabidopsis. Sci Rep 7(1):1395

78. Zhang Q, Zhang J, Wei H et al (2020) Genome-wide identification of NF-YA gene family in cotton and the positive role of GhNF-YA10 and GhNF-YA23 in salt tolerance. Int J Biol Macromol 165(Pt

B):2103–2115

79. Zhang W, Wang Y, Luo G, Zhang JS, He C, Wu X et al (2004) QTL mapping of ten agronomic traits on the soybean (Glycine max L. Merr.) genetic map and their association with EST markers. Theor Appl Genet 108(6):1131–1139

80. Zhao C, Takeshima R, Zhu J, Xu M, Sato M, Watanabe S, Kanazawa A, Liu B, Kong F, Yamada T, Abe J (2016) A recessive allele for delayed flowering at the soybean maturity LOCUS E9 is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog. BMC Plant Biol 16:20

81. Zhao L, Li M, Xu C, Yang X, Li W (2018) Natural variation in GmGBP1 promoter affects photoperiod control of flowering time and maturity in soybean. The Plant Journal 96(1)

82. Zhao L, Luo Q, Yang C, Han Y, Li W (2018) A RAV-like transcription factor controls photosynthesis and senescence in soybean. Planta 227:1389–1399

83. Zhao L, Xu S, Chai T, Wang T (2006) OsAP2-1, an AP2-like gene from Oryza sativa, is required for flower development and male fertility. Sex Plant Reprod 19:197–206

Supplemental Data

Supplemental Tables and Figures are not available with this version.

Figures

a
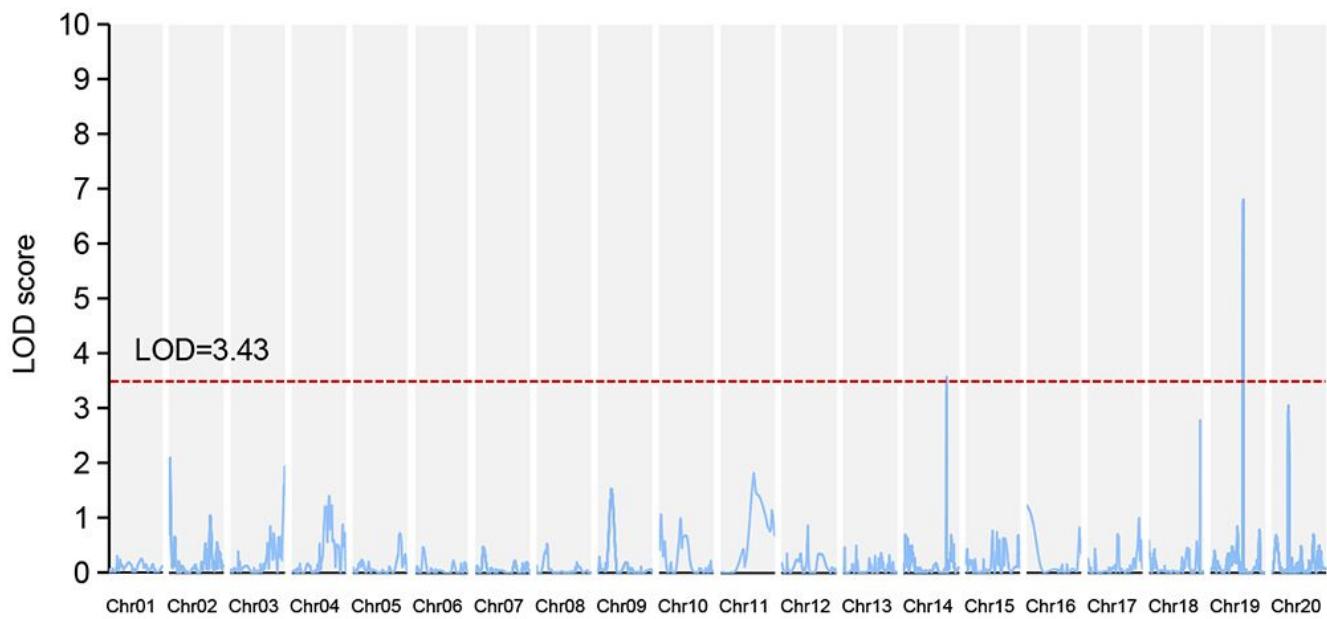
b
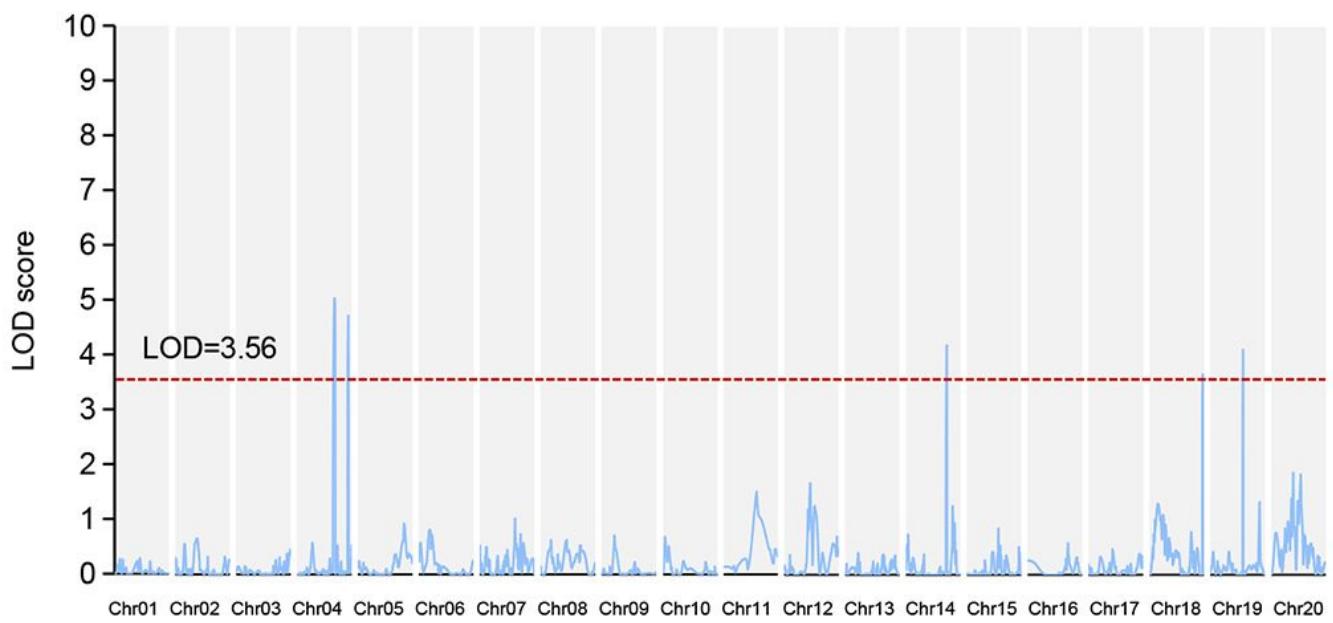
Figure 1

Quantitative trait locus (QTL) mapping by ICIM. a, QTLs of flowering time (R1) in 2018. b, QTLs of flowering time (R1) in 2019. Chr, chromosome. The dotted read lines indicate the threshold of each year.
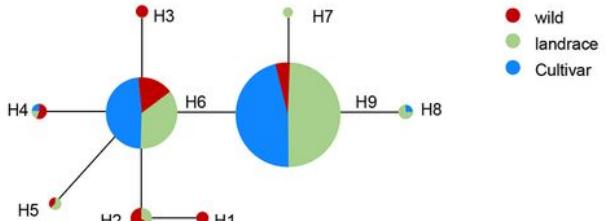
Figure 2

Haplotypes and correlation analysis of Glyma.18G281400. a, Haplotypes of Glyma.18G281400. b-g, Flowering time of Glyma.18G281400 H3 type and H10 type in landraces and improved cultivars in 424 sub-accessions. Flowering time in different regions: b, Harbin, 2019 c, Zhengzhou, 2019 d, Wuhan, 2019 e, Guangzhou, 2019 f, Zhengzhou, 2018 g, Hefei, 2018. Haplotypes was extracted from the 1295 panel of

146 wild soybeans, 575 landraces and 574 improved cultivars. A student's t-test was used to generate the P values.

a

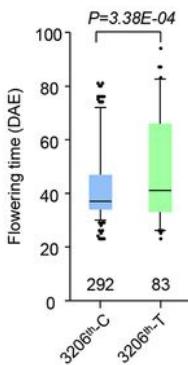|	Haplotype													Number
	H1	T	C	A	A	A	T	G	C	A	C	C	T	3
	H2	T	C	A	A	A	C	G	C	A	C	C	T	22
	H3	C	C	C	A	G	C	G	C	A	C	C	T	3
	H4	C	C	A	T	G	C	G	C	G	A	C	T	5
W82	H5	C	C	A	A	G	C	C	C	A	C	C	T	3
	H6	C	C	A	A	G	C	G	C	A	C	C	T	315
	H7	C	T	A	A	G	C	G	T	A	C	C	C	2
	H8	C	T	A	A	G	C	G	C	A	C	T	C	4
	H9	C	T	A	A	G	C	G	C	A	C	C	C	792
CDS nt # in W82		20	76	223	899	922	1660	2087	2219	2806	2955	3026	3206	
AA identified in W82		7	26	75	300	308	554	696	740	936	985	1009	1069	
AA # in W82		S	L	K	Q	V	H	G	S	I	F	P	S	
AA channge		L	L	Q	L	I	Y	A	L	V	L	L	L	
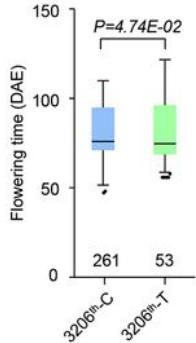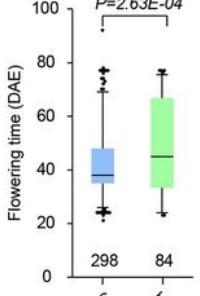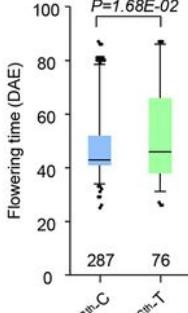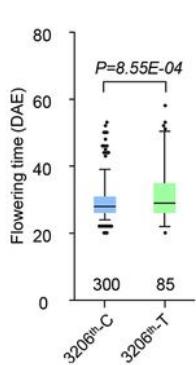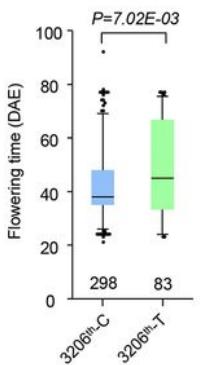
Figure 3

Haplotypes, origins and correlation analysis of Glyma.14G159400. a, Haplotypes of Glyma.14G159400. b, Haplotype origins of Glyma.14G159400. c-h, Flowering time of Glyma.14G159400 3206th-C type and 3206th-T type in landraces and improved cultivars in 424 sub-accessions. Flowering time in different

regions: c, Harbin, 2019 d, Zhengzhou, 2019 e, Wuhan, 2019 f, Guangzhou, 2019 g, Zhengzhou, 2018 h, Hefei, 2018. Haplotypes was extracted from the 1295 panel of 146 wild soybeans, 575 landraces and 574 improved cultivars. Red color represented the wild soybeans, green color represented the landraces, blue color represented the improved cultivars in haplotype origins analysis. A student's t-test was used to generate the P values.
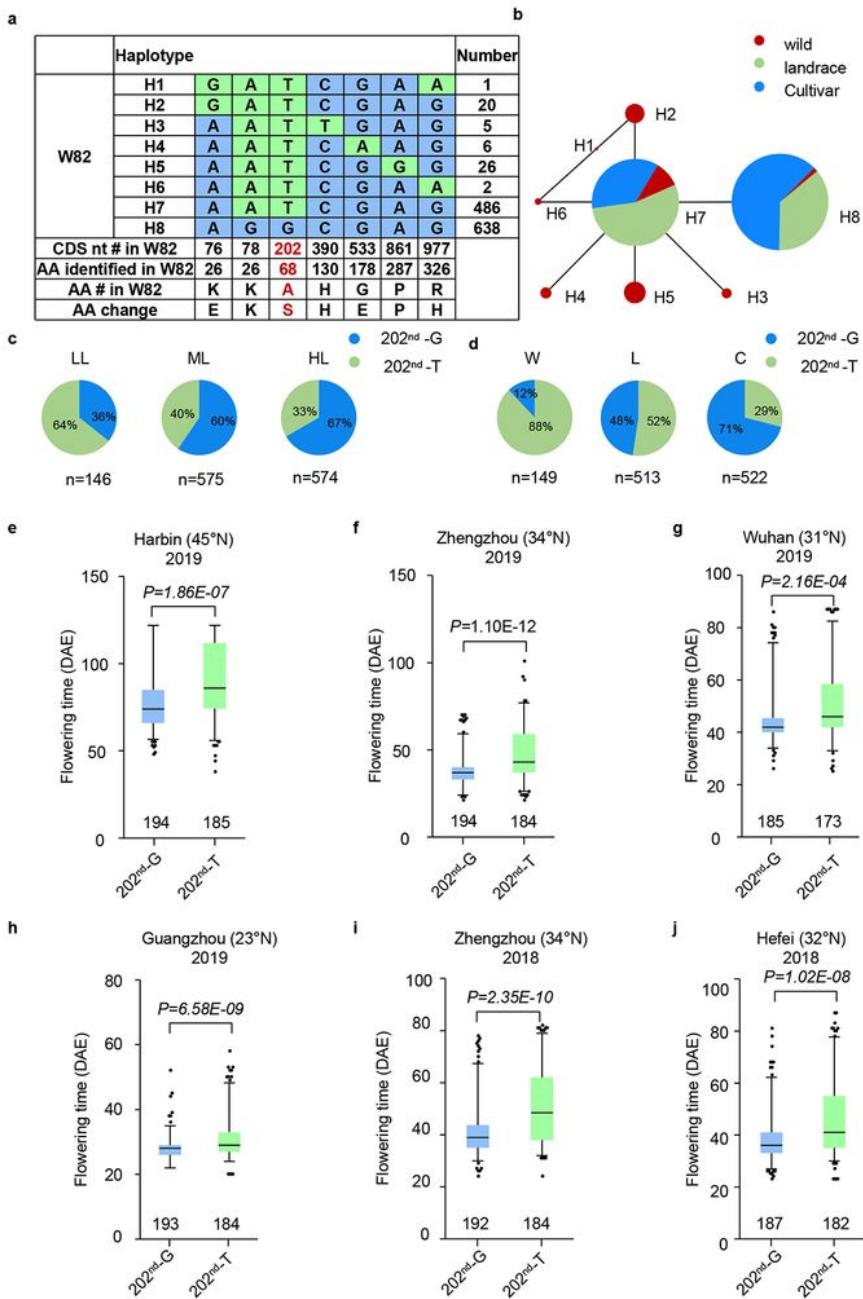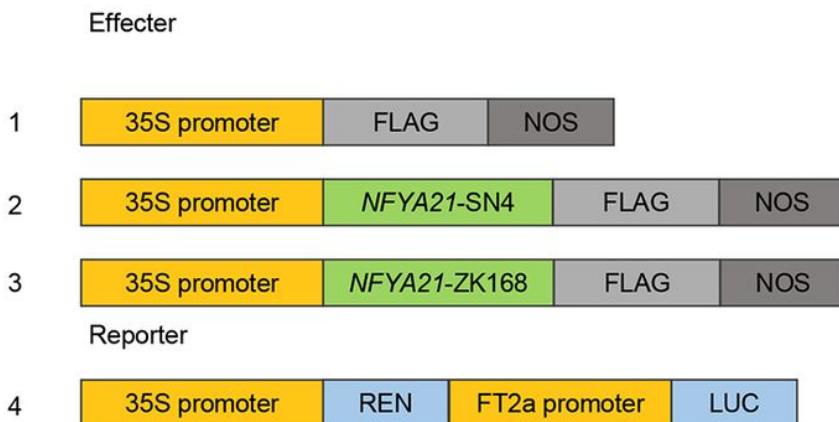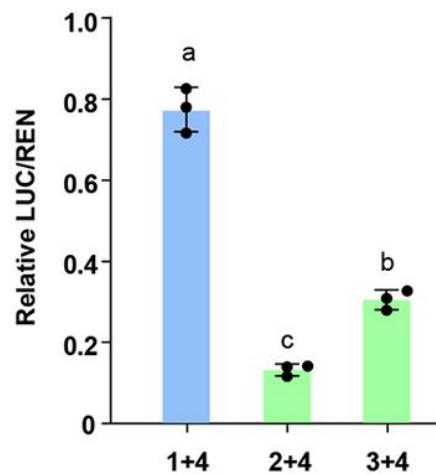

Figure 4

Haplotypes, origins, latitude distribution and correlation analysis of GmNFYA21. a, Haplotypes of GmNFYA21. b, Haplotype origins of GmNFYA21. c, Geographic distributions of GmNFYA21 202nd-G type and GmNFYA21 202nd-T type in landraces and improved cultivars of Chinese landraces and cultivars according to region of origin. d, Allelic distributions of GmNFYA21 in subsets of wilds, landraces and cultivars. e-j, Flowering time of GmNFYA21 202-G type and GmNFYA21 202-T type in landraces and improved cultivars in 424 sub-accessions. Flowering time in different regions: e, Harbin, 2019 f, Zhengzhou, 2019 g, Wuhan, 2019 h, Guangzhou, 2019 i, Zhengzhou, 2018 j, Hefei, 2018. Haplotypes was extracted from the 1295 panel of 146 wild soybeans, 575 landraces and 574 improved cultivars. Red color represented the wild soybeans, green color represented the landraces, blue color represented the improved cultivars in haplotype origins analysis. Geographic distributions were extracted from the 1295 panel of 146 wild soybeans, 575 landraces and 574 improved cultivars. The horizontal line indicates the median value. A student's t-test was used to generate the P values. LL, Low latitude; ML, Middle latitude; HL, High latitude. W, wild accessions in the 1295 panel; L, landrace accessions in the 1295 panel; C, improved cultivar accessions in the 1295 panel.
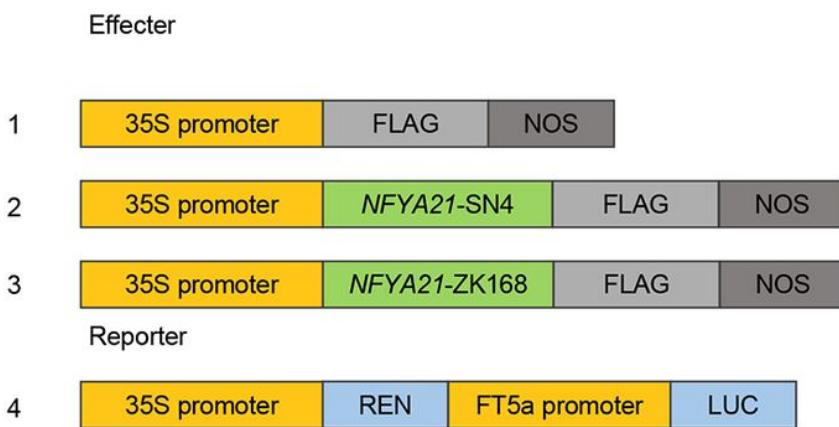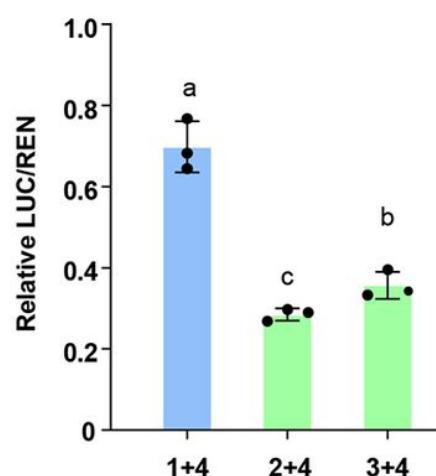
Figure 5

GmNFYA-21 inhibit the expression of GmFT2a/GmFT5a. a, Constructs used for the transient transfection assay of GmNF-YA21 and GmFT2a. b, Luciferase activity under control of GmFT2a promoter regulated by different GmNF-YA21 alleles. c, Constructs used for the transient transfection assay of GmNF-YA21 and GmFT5a. d, Luciferase activity under control of GmFT5a promoter regulated by different GmNF-YA21 alleles. Mean values (+SD) were obtained from three independent replications and the value of each replication was represented by a dot. Duncan's test was used for test of significance of the difference between mean values. Different alphabets indicate significant differences between means.