Raghuvanshi *et al. BMC Plant Biology* *(2025) 25:1426*
https://doi.org/10.1186/s12870-025-06669-6

BMC Plant Biology

RESEARCH **Open Access**

Dissecting genetic architecture of flowering and maturity traits in soybean using GWAS in Indian environment

Rishiraj Raghuvanshi[1], Giriraj Kumawat[1], Rucha Kavishwar[1], Sanjay Gupta[1], Annapurna Chitikineni[2,3], Subhash Chandra[1], Gyanesh K. Satpute[1], Vennampally Nataraj[1], Rajeev K. Varshney[2,3], Henry Nguyen[4], Vangala Rajesh[1], Shivakumar Maranna[1], Mrinal K. Kuchlan[1], Punam Kuchlan[1], Ajay K. Singh[5], Kunwar H. Singh[1] and Milind B. Ratnaparkhe[1]*

Abstract

Background Soybean (*Glycine max* [L.] Merril) is a photoperiod-sensitive crop, with traits like days to flowering, days to maturity playing crucial roles in its adaptability and yield. These traits are regulated by genetic networks controlling flowering time and environmental adaptation, making their genetic basis as an essential knowledge for breeders aiming to improve yield and adaptability. In this study, a Genome-Wide Association Study (GWAS) was conducted for Days to flowering (DTF) and days to maturity (DTM) by using FarmCPU, BLINK and MLM models on 254 diverse soybean genotypes over four consecutive years (2019–2022) to dissect genetic architecture for flowering and maturity traits in an Indian Environment.

Results In this study, GWAS identified 20 significant loci for days to flowering and maturity, among them 12 are new and 8 were previously reported loci. Among the 12 newly identified loci, a significant locus, *Lee.Gm03-3* on chromosome 03, is associated with days to flowering and linked with SNP markers S3_46108324 and S3_46108342. Key candidate genes for *Lee.Gm03-3*, include *Glyma.03G227300* (circadian rhythm and photomorphogenesis, Phytochrome region), *Glyma.03G225000* (circadian rhythm, gibberellic acid signaling, red/far-red light signaling), *Glyma.03G219100* (cytokinin signaling, embryo sac development), and *Glyma.03G226000* (meristem initiation). These genes are vital for light-response and developmental pathways. In addition, we also validated eight previously known genes *E2*, *E4*, *E9*, *E11*, *E10/FT4*, *PRR7/Tof12*, *Dt1*, and *Dt2* that influence flowering and maturity in Indian environment.

Conclusions This study advances understanding of the genetic basis underlying photoperiod sensitivity related genes for circadian rhythm and photomorphogenesis, gibberellic acid signaling, red/far-red light signaling in soybean and highlights potential targets for genetic improvement of flowering, maturity duration and adaptability of soybean under Indian environment.

Keywords GWAS, Days to flowering, Maturity, Soybean, Adaptation

---

*Correspondence:
Milind B. Ratnaparkhe
milind.ratnaparkhe@gmail.com
Full list of author information is available at the end of the article

Introduction

Soybean (*Glycine max*) is an important global crop, valued for its high protein content in animal feed and as a source of vegetable oil. Soybeans are a facultative short-day plants and have originated in higher latitudes of China, [1–4]. Soybeans flowers readily when the day length falls below the critical day length of a genotype [5]. Genotypes adapted to higher latitudes (long days) enter into reproductive phase with little biomass, when they are introduced into lower latitudes (short days). Conversely soybeans of lower latitude either do not enter into reproductive phase or delay flowering in higher latitudes [6]. In spite of photosensitive response, the soybean is now cultivated widely throughout the globe. Although the adaptation of soybean ranges from 50°N to 35°S [7], its individual genotypes adapt to a narrow latitudinal band. India lies between 8° 4' N to 37° 6'N latitudes and soybean cultivation in India ranges from 15.31° N to 27.02° N Latitudes. In contrast to most of the countries where soybean is grown as the full season crop, it is a short season crop in India with maturity duration ranging from 90 days in lower latitudes to 125 days in higher latitudes. It is probable that hitherto unknown genes are present in gene pool for imparting adaptation to Indian conditions.

Genotypes adapted to higher latitudes have evolved through null or hypo- mutations of genes responsible for sensing photoperiod (photo insensitivity) and helps genotypes to flower under long day conditions of these latitudes. Contrarily, genotypes adapted to lower latitudes have evolved through mutations in which delay in flowering occurs even under short day conditions (long juvenility) [8]. To date, a number of major genetic loci, namely *E1* [9–11]), *E2* [11], *E3* [12], *E4* [13]), *E5* [14]), *E6* [15]), *E7* [16]), *E8* [17]), *E9* [18–20], *E10* [18–20]), *E11* [21], *J* [22] and several QTLs, such as *Time of flowering 5* (*Tof5*) [23]), *Tof8* [24]), *Tof9* [25], *Tof11*/*Gp11* and *Tof12*/*Gp1*/*qFT121* [25], *Tof13* [26], *Tof16* [27], *LJ16.1* and *LJ16.2* [28] and *Tof18* [29] have been identified to be involved in the control of flowering and maturity in soybean. Among these genes *E1* & *E2* [30], *E3* [31], *E4* [32], *E5* [33], *E6* [34], *E7* [35] and *E8* [17] have been identified through classical studies while *E9, E10, E11, Tof5, Tof8, Tof9, Tof11, Tof12, Tof13* and *Tof16* have been identified through genotypic studies. Dominant alleles of *E1, E2, E3, E4, E7, E8,* and *E10* inhibit flowering, whereas dominant alleles at *E6, E9, E11* and *J* promote flowering [25]. Genes *E1, E2, E3, E4, E7* and *E8* are involved in photoperiod sensitivity, especially to different light qualities under artificially induced long days (LDs). Soybean *E1* locus, containing a B3 DNA binding domain gene, is known to have the greatest impact on flowering and maturity periods by suppressing *GmFT2a* and *GmFT5a* [10]. *E2* is a homolog of Arabidopsis *Gigantea* (*GI*), characterized at *FT2* locus [11]. *E4* and *E3* are phytochrome A (PHYA) genes, *GmPHYA2* and *GmPHYA3*, respectively, acting as photoreceptors that perceive light signals to influence downstream genes. The dominant *E9* gene confers early flowering, whereas the recessive *e9* gene confers late flowering [11, 19]. *E9* carries *Flowering locus T* (*FT*) homolog *FT2a* [18, 36]. *Tof5* locus for flowering time in higher latitudes was identified by genome wide association analysis [37, 38]. Soybean *Tof5* encodes a homolog of *Arabidopsis thaliana FRUITFULL (FUL)*, which promotes flowering by directly activating the transcription of two important *FT* genes *FT2a* and *FT5a* [37]. *Tof11* and *Tof12* encode PSUDO-RESPONSE REGULATOR (PRR) proteins PRR3a and PRR3b, respectively [39]. *Tof16* was identified by genome wide association analysis of flowering time under natural short days (SDs) conditions [27]. *Tof18* was identified by genome wide association analysis of flowering time under natural LDs conditions [40]. *Tof18* encodes a *SOC1* homologue which promotes early flowering and in LDs as well as SDs conditions. Inheritance studies identified that three genes, namely, *E6* [34], *E9* [19] and *J* [41], govern the long juvenility trait, an important trait for short day conditions of lower latitudes [8]. Mapping studies identified that *E9* encodes *FT2a*, while *J* encode a homolog of Arabidopsis *ELF3* [18, 42]. *E6* and *J* loci were co-localized at the same position on Gm04 and *J* was the Arabidopsis flowering gene *ELF3* [42, 43]. Relation between *e6* and *j* was elucidated as the alleles of same locus [44].

Though several genome wide association studies have been conducted for flowering and maturity time in soybean identifying new loci specific to gene pool or phenotypic environment, however no such study has been conducted under Indian environmental conditions. In the present study, a diverse set of soybean germplasm with varying flowering and maturity durations was constituted and evaluated for four years at Indore, India (22.71° N). This set was genotyped by using Genotyping by sequencing (GBS) and an association study was conducted to decipher the presence of new flowering/maturity genes.

Results

Phenotypic evaluation

Phenotypic evaluations for flowering and maturity traits, including days to flowering (DTF), and days to maturity (DTM), revealed continuous and significant variation across the 254 soybean accessions (Table 1, Fig. 1). The mean values for DTF, and DTM were 46.0 and 109.6 days, respectively, with observed ranges of 28.8—55.1 days for DTF, and 87.4 –120.1 days for DTM (Table 1, Fig. 1A, 1B) indicating a substantial variability

Table 1 Phenotypic evaluation for flowering and maturity related traits

DTF	Mean and Standard Deviation	46.0 ± 5.4				
	Range (Min–Max)	26.33 (28.82 to 55.1)				
	ANOVA table	SS	DF	MS	F (DFn, DFd)	*P* value
	Genotypes x Years	2888	756	3.82	F (756, 1518) = 1.698	****
	Genotypes	87,361	252	346.7	F (252, 506) = 18.65	****
	Years	7502	3	2501	F (1.000, 506.0) = 1111	****
	Subject	9403	506	18.58	F (506, 1518) = 8.259	****
	Residual	3415	1518	2.25		
DTM	Mean and Standard Deviation	109.6 ± 5.6				
	Range	32.68 (87.42–120.1)				
	ANOVA table	SS	DF	MS	F (DFn, DFd)	*P* value
	Genotypes x Years	4021	756	5.319	F (756, 1518) = 119,196,082,660	****
	Genotypes	95,557	252	379.2	F (252, 506) = 14.97	****
	Years	17,522	3	5841	F (1.000, 506.0) = 1.309e + 014	****
	Subject	12,819	506	25.33	F (506, 1518) = 567,737,524,145	****
	Residual	6.77E-08	1518	4.46E-11		

Two way repeated measured ANOVA was conducted for Days to Flowering (DTF) and Days to Maturity (DTM) for four consecutive years (2019–2022)

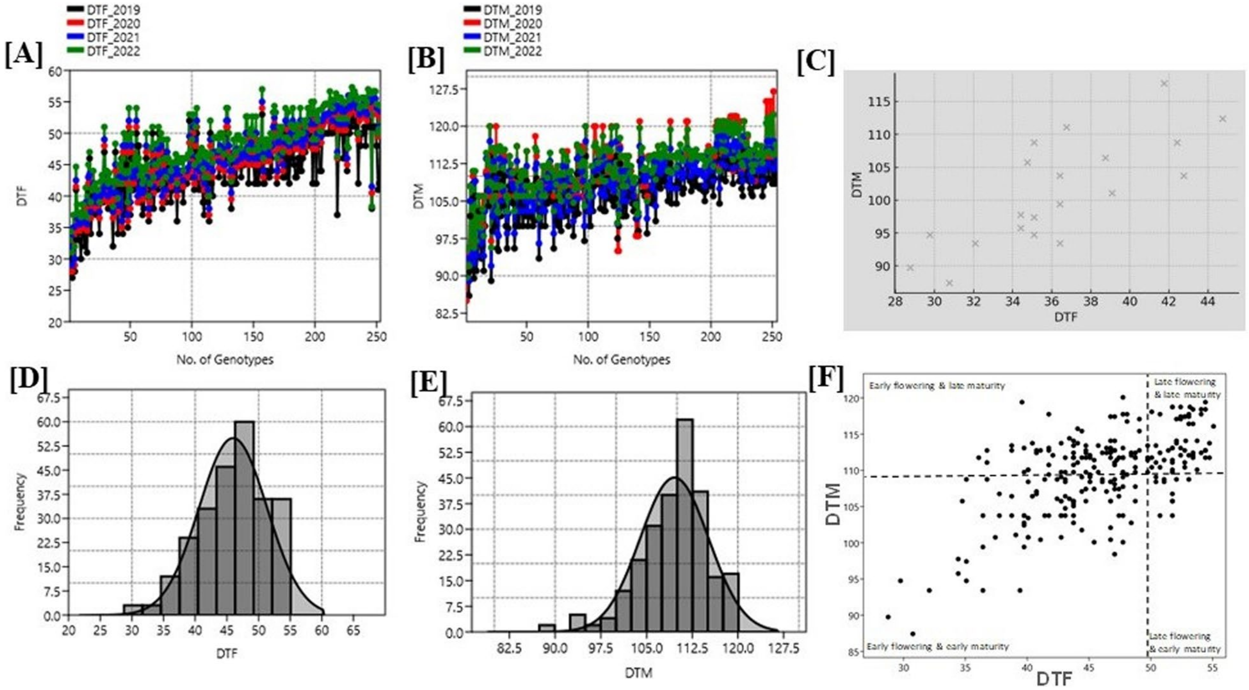


Fig. 1 Phenotypic evaluation of flowering and maturity-related traits in diverse soybean lines. **A** & **B**, Distribution pattern for four consecutive years (2019–2022) for Days to Flowering (DTF) and Days to Maturity (DTM). **C**, Correlation between DTF and DTM. **D** & **E** Frequency distribution for average DTF and DTM, across 254 soybean accessions. **F**, The scattered plot depicting genotypic distribution based on their flowering and maturity groups

in the GWAS panel. The correlation analysis reveals a moderate positive correlation exists between DTF and DTM (r = 0.77), suggesting that early flowering contributes to shorter times to full maturity (Fig. 1C). To analyze the interactions and associations between genotypes and varying environments, ANOVA tests, and principal component analysis (PCA) were conducted. In addition, a two-way ANOVA revealed significant differences in the interaction of genotype and environment (years) for both the traits (Table 1).

Scatter plot for DTF and DTM explained distinct patterns among the genotypes, which indicates a significant amount of genetic diversity (Fig. 1F). Further we identified various groups based on flowering and maturity which include early flowering and early maturity (EC-538828, CAT-47, CAT-146, NRC-12), early flowering late maturity (NRC-2, M-1052, EC-572136, EC-287457), late flowering late maturity (AGS-143, B-1667, AGS-110, B-471), late flowering early maturity (JS-75–46, CAT-290, DS-97–12, EC-251416). This analysis reveals that early flowering does not always lead to early maturity while late flowering does not consistently indicate late maturity. These results emphasize the independent genetic regulation of flowering and maturity traits in soybean.

Analysis of genetic diversity and linkage disequilibrium

After filtration, a total of 66,300 polymorphic SNPs (MAF <0.05) were retained for analysis (Supplementary figure S1). Population structure analysis of 254 soybean accessions revealed that ΔK peaked when K was set to six (Fig. 2), indicating the division of the 254 germplasm accessions into six distinct subpopulations. This stratification was also supported by the neighbor-joining phylogenetic tree, which displayed six clades (Fig. 2A), and was consistent with the clustering observed in the principal component analysis (PCA) (Fig. 2B, (Supplementary figure S2). Additionally, linkage disequilibrium (LD) analysis showed that the average genome-wide LD for the diversity panel was $r^2 = 0.471$. Common population genetic tests were computed to assess natural diversity and selection pressure in the GWAS panel. The average level of silent-site nucleotide diversities per site (π) [45] and population mutation parameter (θ) [46] were observed 0.0000209 and 482.718, respectively. Statistical tests of neutrality, which include Tajima's D, Fu and Li's D* and Fu and Li's F*, yielded values of 3.55, 4.22, and 4.36, respectively (Supplementary Table S1).

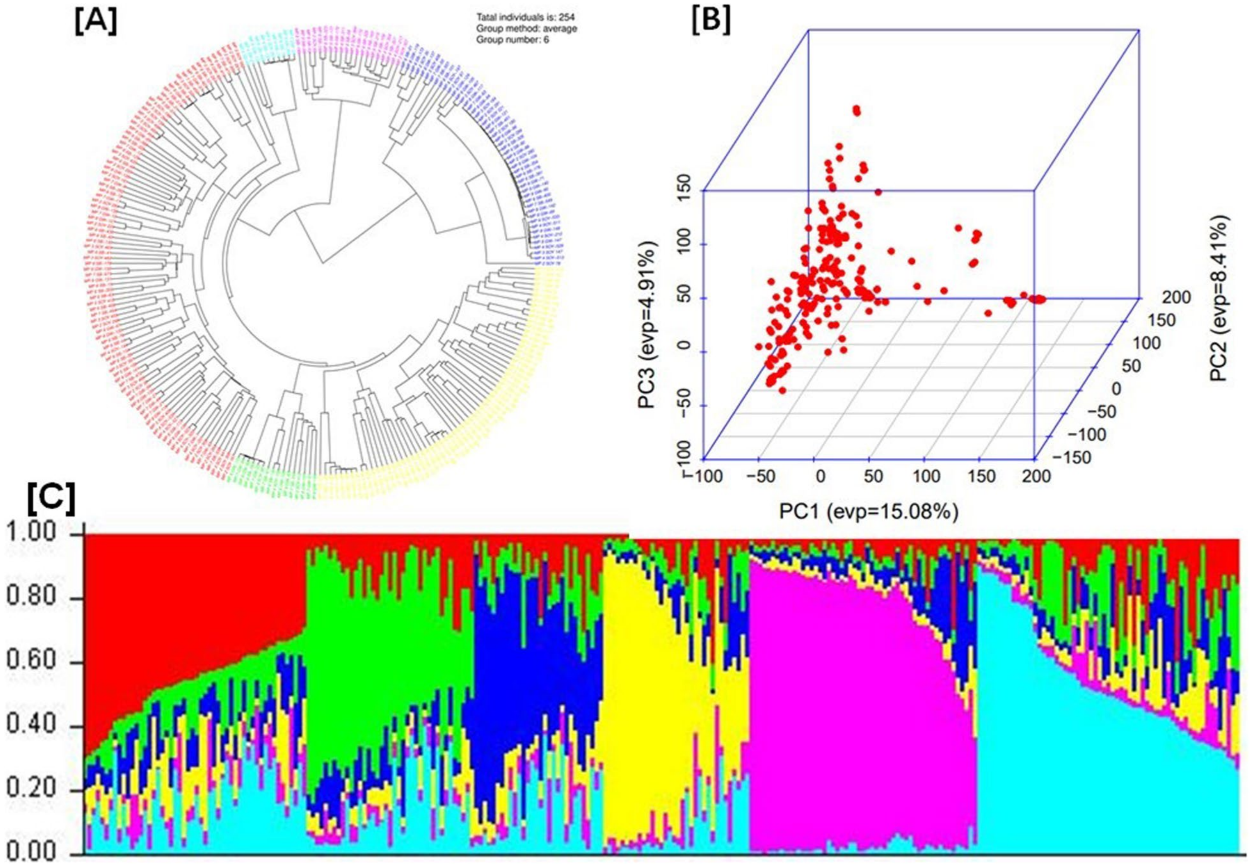


Fig. 2 Diversity analysis using SNPs for the GWAS panel of soybean. A total of 66,300 SNPs derived from a GWAS panel of 254 diverse lines were used to assess genetic diversity. **A**, A Neighbor-joining tree, constructed using the TASSEL software (https://www.maizegenetics.net/tassel), shows six distinct clades in the GWAS panel. **B**, Principal Component Analysis reveals the extent of diversity in the panel. **C**, Population structure developed using STRUCTURE software (https://web.stanford.edu/group/pritchardlab/structure.html), indicates six distinct subpopulations coded by different colors

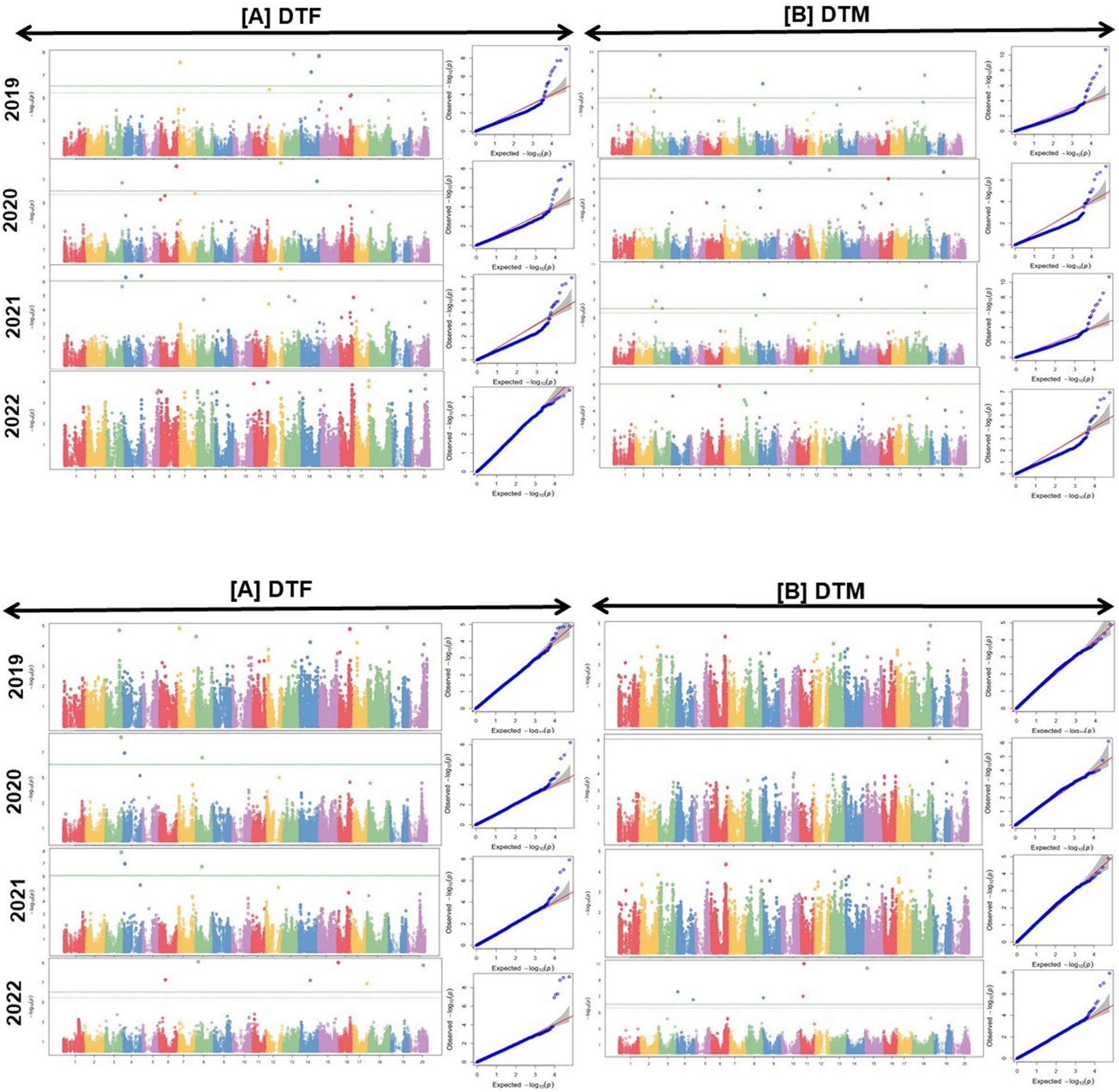


Fig. 3 Association of SNPs with flowering and maturity related traits in soybean. The GWAS analysis involved 254 diverse soybean germplasm lines to study traits associated with flowering and maturity across four years (2019–2022). The analysis was performed using the GAPIT package in R, applying the Fixed and Random Model Circulating Probability Unification (FarmCPU) and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK). Manhattan plot and QQ plot (Quantile–Quantile) showing marker associated with DTF (A) & DTM (B) derived from four consecutive year 2019 to 2022 (Upper panel: FarmCPU, Lower panel: BLINK)

GWAS analysis

The analysis using different association models identified associations between specific loci and DTM and DTF traits across different years (Fig. 3A & 3B, Supplementary figure S3). We identified 20 significant loci for days to flowering and maturity traits, among them 12 are new and 8 were previously known loci (Table 2 & Table 3). The 12 newly identified loci, are distributed across different chromosomes (Chr.) 3, 4, 8, 10, 12, 14, 15, and 18 (Table 2). Among these, Chr. 03 emerged as particularly significant, revealing three important loci i.e. *Lee.Gm03-1, Lee.Gm03-2,* and *Lee.Gm03-3*. Among three loci, *Lee.Gm03-3* stood out as a key finding for DTF across different environments (2019, 2020, 2021), among various models applied (BLINK, FarmCPU and MLM). For DTF in year 2021, the SNP S3_46108324 showed

Table 2 New SNPs identified through GWAS for flowering and maturity traits in soybean

						Position			
S.no	**Locus**	**Chromosome**	**Trait & Environment**	**Model**	**SNP**	**Glyma Lee**	**Wm82.a2**	***P*-values**	**Effect**
1	*Lee.Gm03-1*	3	DTM_2020	FarmCPU	S3_24549807	24,549,807	21,559,727	1.92E-11	−3.84
			DTM_2021	MLM				1.11E-04	−2.34
			DTM_2020	FarmCPU	S3_25992215	25,992,215	22,964,587	8.18E-07	−2.18
2	*Lee.Gm03-2*		DTF_2020	BLINK	S3_40798304	40,798,304	37,356,467	1.72E-05	−1.67
			DTF_2019	MLM	S3_40903281	40,903,281	37,356,467	9.31E-05	−1.8
			DTF_2020	FarmCPU	S3_41286372	41,286,372	37,836,244	8.13E-07	1.44
			DTF_2019	MLM	S3_41452378	41,452,378	37,902,820	5.26E-04	−2.19
3	*Lee.Gm03-3*		DTF_2021	BLINK	S3_46108324	46,108,324	42,575,015	1.22E-08	0.15
				FarmCPU				2.20E-06	1.69
			DTF_2020	BLINK	S3_46108342	46,108,342	42,575,033	6.16E-09	0.15
				FarmCPU				1.76E-07	−1.67
			DTF_2019	MLM				1.37E-04	−3.71
4	*Lee.Gm04-1*	4	DTF_2021	BLINK	S4_6912661	6,912,661	6,809,687	1.09E-07	0.35
			DTF_2022	BLINK				1.00E-07	0.35
			DTF_2022	FarmCPU				4.88E-07	−1.62
5	*Lee.Gm04-2*	4	DTF_2021	BLINK	S4_49785758	49,785,758	47,293,738	6.57E-06	0.28
			DTF_2022	BLINK				4.88E-06	0.28
				FarmCPU				3.77E-07	−1.69
6	*Lee.Gm07-1*	7	DTF_2020	BLINK	S7_5538630	5,538,630	5,471,070	1.39E-05	0.23
				FarmCPU				2.02E-08	1.52
7	*Lee.Gm08-1*	8	DTF_2021	BLINK	S8_20849488	20,849,488	20,430,185	2.49E-07	0.26
			DTF_2022	BLINK				1.70E-07	0.26
				FarmCPU				1.85E-05	1.09
8	*Lee.Gm10-1*	10	DTF_2021	BLINK	S10_40473330	40,473,330	37,138,990	9.79E-05	0.07
			DTM_2021	FarmCPU				5.40E-08	−2.88
			DTM_2019	MLM				1.02E-05	1.8
			DTM_2020					1.05E-05	1.9
9	*Lee.Gm12-1*	12	DTF_2021	BLINK	S12_36598172	36,598,172	13,171,182	9.43E-06	0.07
			DTF_2022					7.94E-06	0.07
			DTF_2021	FarmCPU				3.52E-09	3.11
			DTF_2022					1.15E-07	2.64
10	*Lee.Gm14-1*	14	DTF_2020	FarmCPU	S14_52397292	52,397,292	38,307,979	1.91E-08	2.16
11	*Lee.Gm15-1*	15	DTM_2020	FarmCPU	S15_7613524	7,613,524	7,538,344	8.12E-08	−2.03
			DTM_2021		S15_8057946	8,057,946	7,982,564	9.96E-05	1.61
12	*Lee.Gm18-1*	18	DTF_2020	BLINK	S18_58159788	58,159,788	54,849,471	1.24E-05	0.07
			DTM_2020	FarmCPU				2.88E-09	−2.95
			DTF_2019	MLM				2.5E-05	−2.34

a highly significant *p*-value of 1.22E-08 in the BLINK model and 2.20E-06 in the FarmCPU model. In 2020, the SNP S3_46108342 also exhibits strong significance, with *p*-values of 6.16E-09 in BLINK and 1.76E-07 in FarmCPU for DTF. Identification of significantly associated markers in narrow genomic region across years enhances its reliability and significance as a locus for flowering time across multiple environments (Fig. 3A & 3B). Regarding Locus *Lee.Gm03-1*, significant associations were observed with DTM in multiple years, specifically in 2020 using the FarmCPU and MLM models, and in 2021 with the MLM model. Locus *Lee.Gm03-2* demonstrated associations with DTF, showing significant results in 2020 with both the BLINK and FarmCPU models, and in 2019 with the MLM model, highlighting additional loci of interest for maturation and flowering characteristics. Locus *Lee.Gm03-3* was further investigated to study the effects of alleles and haplotypes on days to flowering (Fig. 4A to

Table 3 Validation of known genes for flowering and maturity traits in soybean

S.N	**Genes**	**Chr**	**Gene position (bp)**	**SNP**	**Identified Loci Positions** Glyma Lee	**Identified Loci Positions** Wm82.a2	***P*-values**	**Effect**	**Trait & Environment**	**Model**	**References**
1	***E11***	7	4,102,968–4,114,174	S7_42711643	42,711,643	38,883,887	3.49E-05	0.17519685	DTF_2021	BLINK	[11, 30]
							3.95E-05	0.17519685	DTF_2022		
				S7_42766230	42,766,230	38,883,887	3.88E-05	0.116141732	DTF_2021		
							4.85E-05	0.116141732	DTF_2022		
2	***E10/FT4***	8	47,458,142–47,459,829	S8_46399480	46,399,480	43,830,824	4.67E-06	−1.081631623	DTM_2020	FarmCPU	[47, 48]
				S8_47131522	47,131,522	44,564,154	3.46E-07	2.062596903	DTF_2020		
3	***E2***	10	45,294,735–45,316,121	S10_47942176	47,942,176	44,538,772	9.21E-05	0.236220472	DTF_2021	BLINK	[11, 30]
				S10_47942176			5.76E-05	0.236220472	DTF_2022		
				S10_47927282	47,927,282	44,523,878	7.99E-05	2.981309437	DTF_2019	MLM	
4	***PRR7***	12	5,508,365–5,522,772	S12_5898961	5,898,961	5,811,502	4.57E-05	1.063003583	DTF_2020	FarmCPU	[39, 49, 50]
5	***E9***	16	31,109,999–31,114,963	S16_29195077	29,195,077	27,706,239	9.60E-06	4.982219442	DTF_2021	MLM	[19, 20]
							9.62E-06	4.98234427	DTF_2020		
				S16_33016027	33,016,027	31,527,189	1.25E-05	2.260087133	DTF_2019		
				S16_26811600	26,811,600	25,322,762	9.02E-07	2.318625057	DTM_2021	FarmCPU	
				S16_29635283	29,635,283	28,079,689	1.69E-05	1.162794485	DTF_2021		
				S16_30814558	30,814,558	29,250,204	1.96E-07	1.788084178	DTF_2020		
				S16_32881071	32,881,071	31,266,178	1.48E-05	0.204724409	DTF_2020	BLINK	
				S16_33917645	33,917,645	32,279,539	1.05E-09	1.629785415	DTF_2020	FarmCPU	
				S16_33931035			2.19E-05	0.38976378	DTF_2021	BLINK	
				S16_33931035			2.00E-05	0.38976378	DTF_2022		
6	***DT2***	18	55,638,209–55,646,547	S18_53487286	53,487,286	50,490,755	8.74E-05	0.352362205	DTF_2020	BLINK	[37, 51]
							7.74E-07	0.352362205	DTF_2021		
							2.53E-06	1.182516985	DTM_2020	FarmCPU	
				S18_54676785	54,676,785	51,651,567	1.23E-05	0.06496063	DTF_2020	BLINK	
				S18_54958192	54,958,192	51,929,135	1.22E-05	−1.369215003	DTF_2020	FarmCPU	
				S18_53487286	53,487,286	50,458,229	3.74E-05	−2.592311118	DTM_2020	MLM	
				S18_54,677,671	54,677,671	51,648,614	8.16E-05	−3.868177112	DTF_2019		
7	***DT1***	19	45,183,357–45,185,175	S19_42283700	42,283,700	39,914,927	1.92E-05	0.200787402	DTF_2021	BLINK	[51, 52]
							2.83E-07	−2.248564364	DTM_2021	FarmCPU	
8	***E4***	20	33,236,018–33,241,692	S20_36021806	36,021,806	33,958,681	7.84E-05	0.265748031	DTF_2021	BLINK	[13, 32]
							8.61E-05	0.265748031	DTF_2022	BLINK	
				S20_36824277	36,824,277	34,746,065	6.46E-06	1.024440493	DTF_2020	FarmCPU	
				S20_36946413	36,946,413	34,866,334	2.51E-05	0.281496063	DTF_2021	BLINK	
							2.44E-05	0.281496063	DTF_2022		

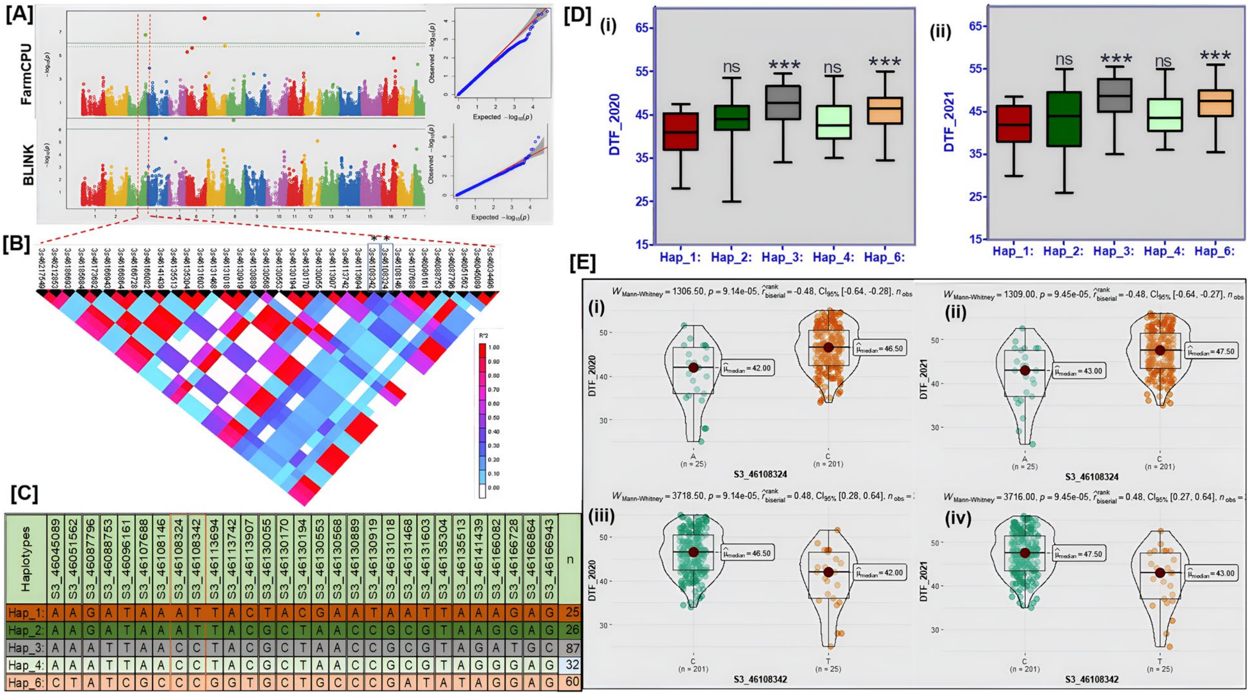


Fig. 4 Analysis of allele and haplotype effects on DTF in Soybean. **A**, Manhattan and QQ plots showing significant markers associated with DTF in 2020 and 2021 on chromosome 3. **B**, Linkage Disequilibrium (LD) region containing significant SNPs on chromosome 3. **C**, Haplotypes identified within the LD region. **D**, Association of haplotypes with DTF across two different environments. Further Dunnett's Multiple Comparison Test was applied to compare haplotype groups using GraphPad Prism 5. **E**, Allele effects on associated phenotypic traits across two environments (2020 and 2021), The Mann–Whitney U test was applied to compare allele groups

4D). The results identified five major haplotypes: Hap1, Hap2, Hap3, Hap4, and Hap6 (Fig. 4C & 4D). Hap1 is associated with early flowering in both environments (2020 and 2021) (Fig. 4D), while Hap3 and Hap6 are associated with comparatively delayed flowering. The allelic effects of two significant markers, S3_46108324 and S3_46108342, showed significant differences between the alleles (Fig. 4E).

In this study, we validated eight previously reported genes through our GWAS panel (Table 3), including *E11* on Chr. 7 (DTF in 2021 and 2022), *E10/FT4* on Chr. 8 (DTM and DTF in 2020), and *E2* on Chr. 10 (DTF in 2021 and 2022). *PRR7/Tof12* on Chr. 12 was associated with DTF in 2020, while *E9* on Chr. 16 was linked to both DTM in 2021 and DTF in 2020 and 2021. *DT2* on Chr. 18 and *DT1* on Chr. 19 showed associations with both DTF and DTM across years, and *E4* on Chr. 20 was linked to DTF in 2021.

Candidate gene identification

All three significant loci on chromosome 03 were explored for candidate gene identification. The functions of the identified candidate genes were predicted based on protein sequence homology in SoyBase (www.soybase.org), suggesting potential roles in flowering and maturity related trait in soybean. The three loci located on chromosome 03 revealed candidate genes related to reproduction, photoperiod, and circadian rhythm. *Lee.Gm03-1* carries a single gene, *Glyma.03G080600*, which plays a critical role in determining bilateral symmetry, meristem initiation, and polarity specification of the adaxial/abaxial axis. *Lee.Gm03-2* stands out as a primary locus for genes involved in reproductive functions (Table 4). It contains multiple genes associated with flower development, floral organ identity and senescence, as well as pollen tube growth and guidance. For example, *Glyma.3g160000* and *Glyma.03 g160700* are annotated for their role in flower development, while *Glyma.03g164200* is annotated with floral organ senescence. Additionally, *Glyma.03g161000*, *Glyma.03g164900* and *Glyma.03 g165700* contribute to pollen tube growth and guidance, while *Glyma.03 g166000* is linked to male meiosis. Together, these genes make *Lee.Gm03-2* a hub for reproduction-related genes. On the other hand, locus *Lee.Gm03-3* predominantly houses genes related to circadian rhythm, photoreception and light responsiveness. This locus consists of several genes that are highly responsive to light, especially red and far-red light signaling pathways, which are crucial for plant adaptation to the environment. For example, genes such as

Glyma.03G227300 and *Glyma.03G227800* are involved in phototropism, photomorphogenesis, and red/far-red light signaling. *Glyma.03G221600* participates in gibberellin catabolic processes responsive to red or far-red light, while *Glyma.03G225000* and *Glyma.03G227800* are involved in red and far-red light signaling pathways. Additionally, locus *Lee.Gm03-3* contains genes controlling circadian rhythm, including *Glyma.03G225000, Glyma.03G227300,* and *Glyma.03G227800,* which regulate processes related to circadian timing, light detection, and photo-morphogenesis.

In summary, *Lee.Gm03-2* is primarily associated with genes governing reproductive processes, while *Lee.Gm03-3* has a greater concentration of genes involved in light responsiveness, photoreception, and circadian rhythm, highlighting each locus's unique contributions to plant development and environmental adaptation.

Expression study for putative candidate genes

We analyzed the tissue specific expression data available at Phytozome database for all identified putative candidate genes (Fig. 5A & 5B) and compared them based on expression data derived from different tissues directly linked to flowering and maturity pathway, including open flowers, unopened flowers, shoot tips, and leaves.

The gene expression data of *Lee.Gm03-2* and *Lee.Gm03-3* loci suggests its roles in plant growth, reproductive development, and environmental responses (Fig. 5A, 5B & 5C). In the *Lee.Gm03-2* locus, *Glyma.03G164000* showed higher expression in the shoot tip (5.307) and low expression in the leaf (0.052), likely aiding growth in the shoot tip and reproductive transition. *Glyma.03G164200*, with high expression in open flowers (1.242) and moderate in unopened flowers (0.417), may contribute to floral organ senescence, potentially regulating flower aging. *Glyma.03G166000* displays expression across tissues, particularly in the shoot tip (5.01), suggesting involvement in reproductive processes, possibly playing a role in male meiosis II.

In the *Lee.Gm03-3* locus, *Glyma.03G219100* is highly expressed in open flowers (5.2) and moderately in unopened flowers (0.806), possibly influencing reproductive development through cytokinin-mediated signaling pathways. *Glyma.03G225000*, with elevated expression in open flowers (4.2) and unopened flowers (2.9), may be involved in circadian rhythm and light response pathways, potentially regulating flowering time in response to light. Strong expression of *Glyma.03G226000* in open flowers (10.8) suggests it could support flower structure formation, possibly through meristem initiation and symmetry determination. Lastly, *Glyma.03G227300* having Phytochrome region; GAF domain; His Kinase A (phospho-acceptor) showing moderate expression across tissues and particularly in unopened flowers (1.078), may contribute to photo-morphogenesis and light response, helping the plant adapt to light conditions through circadian and phototropic responses. This gene expression pattern thus provides insights into how each gene may function in flowering, growth, and environmental adaptation. These genes were further studied with the orthologs from different species (Supplementary figure S4A-S4G). These key genes are also present in important legumes and other species suggesting its crucial role in regulation of flowering and maturity.

Validation of major locus *Lee.Gm03-3* using KASP analysis

The stable locus *Lee.Gm03-3*, containing SNPs S3_46108324 and S3_46108342, was validated using KASP marker analysis in a separate population consisting of 157 Indian soybean cultivars (Fig. 6A & 6B). For SNP S3_46108324, significant associations were found with DTM, exhibiting P-value of 3.23E-04, which explained 8.34% of phenotypic variation (R^2) in 2023 (Fig. 6C). Similarly, SNP S3_46108342 demonstrated significant associations with DTF, and DTM, with p-values of 0.017 and 0.016, respectively, accounting for 4.79% and 4.90%, of phenotypic variation in 2022 (Fig. 6C). Further allelic effect for DTM-2023 and DTM-2022 phenotype also showed significance difference (Fig. 6D & 6E). These findings validate *Lee.Gm03-3* and its SNPs as important genetic markers for flowering and maturity traits.

Discussion

The genetic architectures pertaining to the flowering time and maturity can be the key factors in achieving the designer soybean plants that are adaptable and of a desired type for enhanced productivity in specific environments. Several genetic studies revealed critical role of many single genes such as *E1, E2, E3, E4, E9 FT2a, FT5a, J, Tof4, Tof5, Tof8, Tof9, Tof11, Tof12, Tof13, Tof16* and *Tof18* in photoperiod mediated flowering for adaptation of soybean [25]. In our investigation, for the first time in Indian condition, we dissected the underlying genetic architecture of flowering time, and maturity duration in an Indian environment, along with the key genetic loci and candidate genes that regulate these traits in the soybean. The GWAS panel consisting of diverse 254 germplasm lines, grouped into six distinct subpopulations, indicated by population structure analysis, supports the significance of genetic diversity within the dataset, which is essential for accurately mapping these traits. This stratification, confirmed by both phylogenetic clustering and PCA, indicates a robust genetic basis for the identified loci. In this study, we identified *Lee.Gm03-3* as a key locus for flowering time with markers S3_46108324 and S3_46108342, the two markers linked to the locus in both environments i.e. 2020 and 2021, respectively.

Table 4 Identification of putative candidate genes in soybean

Locus	Gene ID	Start	End	Description	PfamID
Lee.Gm03-1	*Glyma.03G080600*	21,508,100	21,512,393	determination of bilateral symmetry; meristem initiation; polarity specification of adaxial/abaxial axis; regulation of meristem growth; response to light stimulus	FAMILY NOT NAMED
Lee.Gm03-2	*Glyma.03G160000*	37,508,220	37,513,007	positive regulation of flower development;	KH_1;
	Glyma.03G160700	37,584,972	37,586,022	flower development; specification of floral organ identity	zf-C2H2_6;
	Glyma.03G161000	37,612,023	37,615,718	pollen tube growth; protein transport; vesicle-mediated transport;	NA
	Glyma.03G163400	37,813,139	37,813,951	seed maturation;	Cupin_1;
	Glyma.03G164000	37,866,656	37,868,942	post-embryonic plant morphogenesis;	DUF640;
	Glyma.03G164100	37,892,273	37,895,682	photoperiodism, flowering; negative regulation of long-day photoperiodism	JmjN;
	Glyma.03G164200	37,902,284	37,909,534	floral organ senescence;	NAM;
	Glyma.03G164900	37,980,836	37,982,905	pollen tube growth;	PBD;
	Glyma.03G165700	38,045,521	38,050,084	pollen tube guidance;	LRR_8; Pkinase;
	Glyma.03G166000	38,068,482	38,070,661	male meiosis II;	NA
Lee.Gm03-3	*Glyma.03G219100*	42,265,891	42,275,563	cytokinin mediated signaling pathway; embryo sac development	Histidine kinase-, DNA
	Glyma.03G219300	42,285,599	42,286,853	positive regulation of seed maturation	bZIP transcription factor
	Glyma.03G219800	42,312,343	42,319,568	DNA-dependent; vegetative phase change; vernalization response	SET domain
	Glyma.03G219900	42,325,550	42,327,910	floral organ morphogenesis; gibberellic acid mediated signaling pathway	GRAS domain family
	Glyma.03G220100	42,334,845	42,336,847	embryo development ending in seed dormancy;pollen development	WRKY DNA -binding domain
	Glyma.03G221200	42,433,366	42,437,270	embryo sac egg cell differentiation	Helicase conserved C-terminal domain;
	Glyma.03G221600	42,466,464	42,471,412	gibberellin catabolic process; response to red or far red light	2OG-Fe(II) oxygenase superfamily
	Glyma.03G223300	42,601,145	42,604,438	photoperiodism, flowering; protein folding	DnaJ domain
	Glyma.03G224300	42,656,642	42,665,584	vegetative phase change; vernalization response	SET domain
	Glyma.03G225000	42,726,377	42,729,203	circadian rhythm; gibberellic acid mediated signaling pathway;red or far-red light signaling pathway	Helix-loop-helix DNA-binding domain
	Glyma.03G226000	42,818,553	42,823,054	determination of bilateral symmetry; meristem initiation	Cellulase (glycosyl hydrolase family 5)
	Glyma.03G226500	42,874,585	42,876,167	embryo development ending in seed dormancy; vegetative to reproductive phase transition of meristem	NA
	Glyma.03G226600	42,877,084	42,882,874	embryo development ending in seed dormancy	GUCT (NUC152) domain;
	Glyma.03G227300	42,918,771	42,923,401	circadian rhythm; detection of visible light; photomorphogenesis; phototropism;response to continuous far red light stimulus by the high-irradiance response system; response to far red light; response to very low fluence red light stimulus; signal transduction	Phytochrome region; GAF domain; His Kinase A (phospho-acceptor) domain; PAS fold;
	Glyma.03G227800	42,980,227	42,984,347	circadian rhythm; de-etiolation; gibberellic acid mediated signaling pathway; gravitropism; positive regulation of anthocyanin metabolic process; red or far-red light signaling pathway; regulation of transcription, DNA-dependent; response to red or far red light; signal transduction	Helix-loop-helix DNA-binding domain
	Glyma.03G228700	43,069,814	43,073,090	floral organ formation; regulation of flower development; sepal formation; spindle assembly	Domain of unknown function (DUF3635)
	Glyma.03G224400	42,667,804	42,673,022	pollen exine formation; anther wall tapetum development	zf-Sec23_Sec24; Sec23_helical; Sec23_trunk;
	Glyma.03G225000	42,726,377	42,729,203	red, far-red light phototransduction; response to red or far red light; de-etiolation	HLH;

Furthermore, we converted the two significant SNPs of *Lee.Gm03-3* into KASP assays and validated them in a separate soybean population consisting of diverse cultivar, enhancing utility of this locus in soybean breeding. These associations will enable breeders to use this consistent loci to their advantage, further underscoring the importance of *Lee.Gm03-3* in the design of soybean varieties that can grow under different photoperiods. Various genome wide association studies in soybean, revealed several loci controlling flowering time and

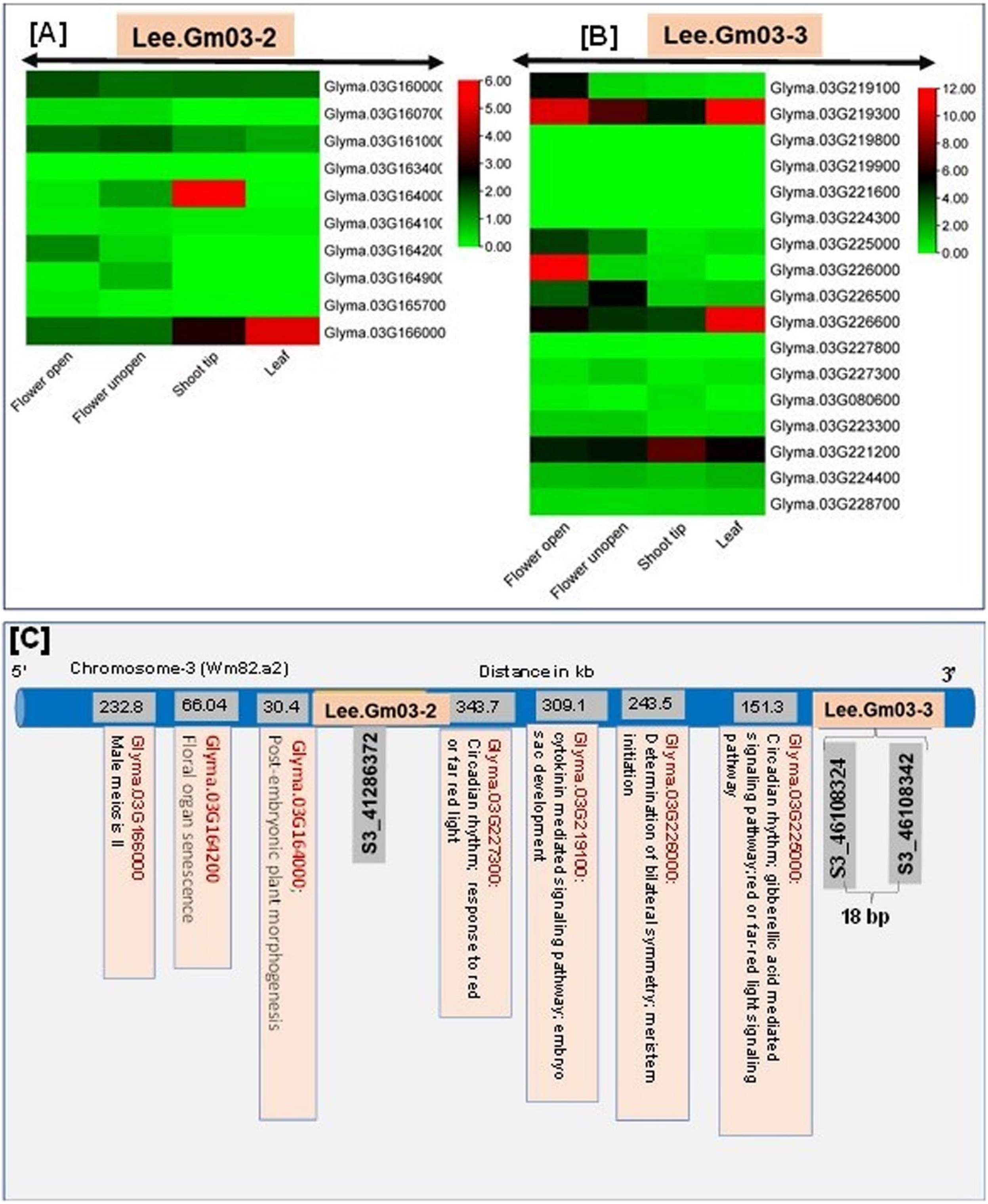
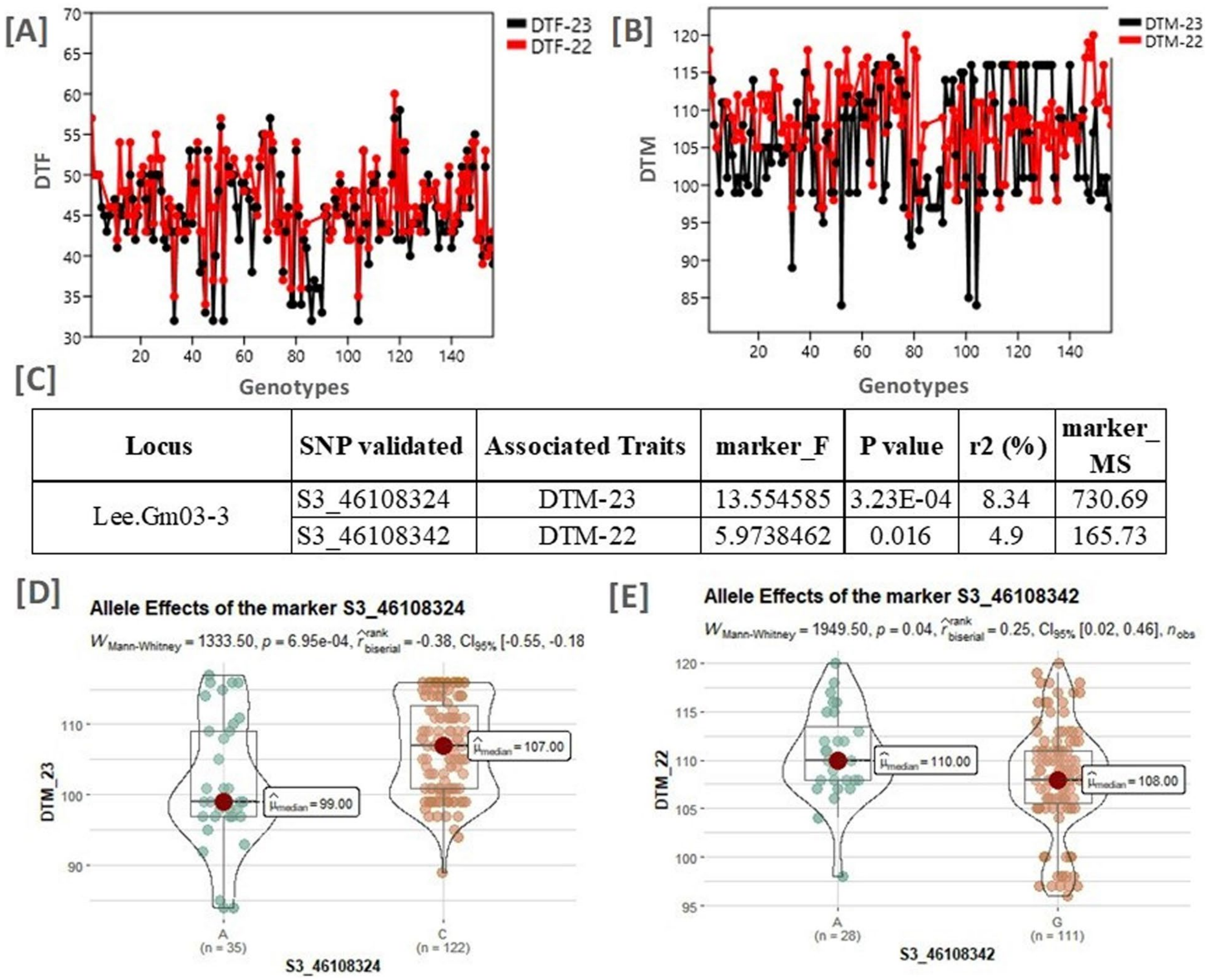


Fig. 5 Expression data analysis for putative candidate genes in Soybean. Expression data for the putative candidate genes identified in *Lee.Gm03-2* (**A**) and *Lee.Gm03-3* (**B**) loci were obtained from the Phytozome database (https://phytozome-next.jgi.doe.gov/) for different stages includes: flower opening, unopened flowers, shoot tip, and leaf tissues. The expression data was converted into a heatmap using the TBtools software. (**C**) Illustrating the genes from both the loci showing differential expression among different tissue



Locus	SNP validated	Associated Traits	marker_F	P value	r2 (%)	marker_MS
Lee.Gm03-3	S3_46108324	DTM-23	13.554585	3.23E-04	8.34	730.69
	S3_46108342	DTM-22	5.9738462	0.016	4.9	165.73

Fig. 6 Validation of *Lee.Gm03-3* loci in soybean cultivars by using KASP marker. Total 157 genotypes consisting of released soybean cultivars of India were phenotyped for DTF and DTM in 2022 to 2023 and further used for association analysis by using GLM (General Liner Model) implemented in TASSEL software (https://www.maizegenetics.net/tassel). (**A**) and (**B**) Phenotypic data distribution, (**C**) Result obtained from GLM model, (**D**) Allelic effect for DTM-2023, (**E**) Allelic effect for DTM-2022

maturity duration, such as *J* [25], *Tof5* [23], *Tof11* and *Tof12* [25], *Tof13* [26], *Tof16* [27]), and *Tof18* [29]. *E2* was also detected through GWAS during the identification of *Tof5*, *Tof13*, and *Tof18* [23, 26, 29].

This GWAS study not only identified several new loci but also validated the presence of previously known loci associated with flowering time and maturity duration, confirming the robustness of our findings. These known loci includes *E2, E4, E11, E10/FT4, PRR7/Tof12, E9, Dt1*, and *Dt2*. Functions of these loci are well studied and their role in regulation of flowering time and maturity in different photoperiod regimes have been reported in several studies [27]. For instance, *E1* is considered as the main contributor of photoperiod sensitivity and the other *E* gene interactions influencing the flowering and maturity response, making it crucial in enhancing soybean adaptability across latitudes [9]. *E1* by interacting with the *FUL* transcription, suppresses *E9* activities through binding the promoter, while *FUL* interacts with *FT2a* (*E9*) and *FT5a* to enhance their expression with the end result of promoting flowering in long day conditions [27]. In this GWAS, according to *p*-value and effect size, *E9* has been identified as the major locus determining the variation of DTF and DTM traits in an Indian environment.

The two candidate genes of *Lee.Gm03-3*, viz. *Glyma.03G227300*, and *Glyma.03G225000*, may be involved in regulating flowering time in soybean by modulating circadian rhythms and light-responsive pathways. Gene *Glyma.03G227300* also reported earlier as *GmPHYA4* [53, 54], contains the Phytochrome region; GAF domain; and His Kinase A (Phospho-acceptor). This gene could be involved in maintaining the circadian rhythm, detecting light sources, the switching on-off a plant's development processes and the directional growth towards or away from the light. The function of this gene in photoreceptor activity is consistent with its genetic architecture since it is also a photoperiod control gene, along with other genes studied in *Arabidopsis thaliana* and *Oryza sativa* in which the role of phytochromes was also essential for flowering with respect the day-length [55].

Moreover, *Glyma.03G225000* is associated with the circadian clock and with the gibberallic acid (GA)-mediated signaling pathways and responses to red light and far-red light. This gene is also associated with flowering time and other GA enhancing growth attributes. GA signaling has been known to have functional crosstalk with photoperiod pathways. Such hypothesis is applicable in this scenario because the gene in question is considered to be responsible for coordinating GA signaling and circadian rhythms that ultimately determine flowering time in response to light. This regulatory pathway has already been reported in other plants whereby flowering is dependent on GA and photoreceptors under different light treated conditions [6, 56]. In combination, *Glyma.03G227300* and *Glyma.03G225000* also represent as key gene candidates in breeding programs that alter flowering and maturity time of soybean plants grown under different photoperiods.

Ultimately, the detected loci might provide practical markers for improvement of the timing of flowering and maturity traits of soybean that may lead to increase adoption of new variety in India. The findings of present study are not only consistent with previously identified loci but also show the potential of using GWAS as a tool to discover new and stable markers for important traits like DTF and DTM [7]. The mapping of these loci or genes opens new avenues for crop improvement with genetic designing that could be initiated for altering genotype and phenotype of soybean genotypes.

Additionally, understanding photoperiod sensitivity and genetic control of flowering time will help in developing breeding strategies to balance early flowering with adequate plant growth. This study provides a valuable foundation for future genetic and molecular research aimed at increasing soybean yield and adaptation. For better insight, functional validation and further characterization of these genes will be helpful for improvement of soybean varieties. Similar studies need to be conducted at different altitude to get robust marker and genes for regulation of flowering and maturity time.

Conclusions

This study enhances the current understanding of the genetic mechanisms influencing photoperiod sensitivity and flowering in soybean. By identifying 20 significant loci, including 12 novel loci, and validating 8 previously known reported genes, it provides a comprehensive framework for genetic improvement. The discovery of the 4 key candidate genes *Glyma.03G227300, Glyma.03G225000, Glyma.03G219100* and *Glyma.03G226000* associated with the significant locus *Lee.Gm03-3* on chromosome 03 offers valuable insights into the regulatory pathways of flowering and maturity, including circadian rhythm, hormone signaling, and light-response pathways. SNP markers identified in this study will help in the molecular breeding programme for developing early maturing soybean cultivars. These findings open new avenues for breeding soybean varieties with optimized flowering duration and better adaptability to Indian environmental conditions, contributing to improved agricultural productivity.

Material and methods

Plant materials and phenotypic evaluation

Phenotyping was conducted on 254 diverse soybean germplasm accessions during the summer season (Mid-June to Mid-October) over four consecutive years, from 2019 to 2022, using an augmented design at the ICAR-Indian Institute of Soybean Research (22.7196° N, 75.8577° E), Indore, India. Morphological traits related to flowering and maturity, such as Days to Flowering (DTF) recorded at the R1 stage (the day when 50% of the plants in a plot have an open flower on one of the top four nodes that bears a fully expanded leaf) and Days to Maturity (DTM) were measured in the field for all four years.

Phenotypic analysis included normality distribution and descriptive statistics such as mean, standard deviation (SD), maximum and minimum trait values, and the coefficient of variation (CV%). In addition, correlation, PCA analysis and two-way Analysis of Variance (ANOVA) were performed using GraphPad Prism version 9 (www.graphpad.com) to evaluate the effects of genotype (G), environment (E), and genotype-by-environment interaction (G × E) on both the traits.

Genotyping

The genomic DNA from the leaves derived form 254 soybean diverse germplasm accessions was extracted using CTAB method [57]. Genotyped by sequencing (GBS) was conducted following the methods and recommendations outlined by [45, 58, 59]. The GBS library was created with *Ape*KI restriction enzyme digestion [46]. A 158 million single-end reads were generated with an Ion Torrent Proton System (Thermo Fisher Scientific Inc., USA) by ICRISAT (International Crops Research Institute for the Semi-Arid 'Tropics), Hyderabad, India. These were processed using the Fast-GBS.v2 pipeline [47]. FASTQ files were demultiplexed, trimmed, and then mapped against the soybean reference genome (Glyma.Lee_v2.0). Imputations were performed in TASSEL software to fill missing data and further SNP data were filtered by minor allele frequency (MAF) < 0.05 and missing rate > 10% and finally a total of 66,300 SNPs distributed all over 20 chromosomes were used for further study.

Genetic Diversity analysis

Level of silent-site nucleotide diversities per site (π) and population mutation n parameter (θ) were estimated. Statistical tests of neutrality such as Tajima's D, Fu and Li's D* and F were also calculated to examine the selection pressure at SNPs in our GWAS panel by using DnaSP software version 5.10 [60–62] (http://www.ub.edu/dnasp/index_v5.html). A Neighbour-joining tree was constructed using the TASSEL software [63] (https://www.maizegenetics.net/tassel). Principal Component Analysis (PCA) and LD decay plot were generated using the GAPIT package (https://www.maizegenetics.net/gapit) implemented in R. Population structure was developed using STRUCTURE software (https://web.stanford.edu/group/pritchardlab/structure.html).

Genome-wide association study and candidate gene identification

The analysis involved 254 diverse soybean accessions to study SNPs associated with flowering and maturity across four years (2019–2022). The association analysis was performed using three analysis models. Mixed Linear Model (MLM, using PCA (fixed-effect factor) + K (random-effect factor)) was implemented in TASSEL v5.0. The Fixed and Random Model Circulating Probability Unification (FarmCPU) [64] and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) [65] models were applied through GAPIT package in R [56]. The first two principal components were included as covariates in both models. A Bonferroni correction (α/N) was used to set the significance threshold, where N represents the number of tested SNPs and $\alpha =$ 1. Manhattan plots illustrated significant markers, while quantile–quantile (Q-Q) plots compared expected versus observed p-value distributions (on a -log10 scale).

For candidate gene identification, genomic regions 250 kb upstream and downstream of significant SNPs (totaling 500 kb) were analyzed, based on the average linkage disequilibrium decay in soybean [53, 54]. Genes within these regions were identified using the Lee reference genome and Wm82.a2 genome assembly, with data obtained from SoyBase (www.soybase.org). The functions of the identified candidate genes were predicted based on protein sequence homology in SoyBase (www.soybase.org). Further haplotypes were analyzed with in the LD region by using DnaSP software version 5.10 (http://www.ub.edu/dnasp/index_v5.html).

Expression analysis of putative candidate genes

Expression data for the putative candidate genes identified in various loci were obtained from the Phytozome database (https://phytozome-next.jgi.doe.gov/). The mRNA expression data covered different growth stages related to flowering and maturity, including flower opening, unopened flowers, shoot tip, and leaf tissues. To effectively visualize the expression patterns and relationships among these candidate genes, the expression data were converted into a heatmap using the TBtools software [66].

Validation of SNPs with KASP marker

The two SNPs of major locus *Lee.Gm03-3* were converted to Kompetitive allele specific PCR (KASP) assays and validated using KASP analysis in a new set of 157 soybean genotypes. These 157 soybean genotypes consisting of released soybean cultivars of India, were phenotyped for DTF and DTM in 2022 and 2023. Generalized linear model (GLM) was used for trait association in TASSEL v5.0 [63].

Abbreviations

BLINK	Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway
DTF	Days to Flowering
DTM	Days to Maturity
FarmCPU	Fixed and Random Model Circulating Probability Unification
GBS	Genotyping-by-sequencing
GWAS	Genome-Wide Association Study
KASP	Kompetitive Allele-Specific PCR
LD	Linkage Disequilibrium
MAF	Minor Allele Frequencies
MLM	Mixed Linear Model
PCA	Principal Component Analysis
QTL	Quantitative Trait Loci
SNP	Single Nucleotide Polymorphisms

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06669-6.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

The authors acknowledge ICAR-National Agriculture Science Fund (NASF) for financial support.

Authors' contributions

RR, GK, RK, AC, SC, GKS, VN, VR, SM, MKK, PK conducted field and laboratory experiments, RR, GKS, GK, AKS, SG, MBR, AC conducted data analysis and prepared figures, AKS, SG, HN, KHS, RKV, MBR wrote the main manuscript text. All authors reviewed the manuscript.

Data availability

Data is provided within the manuscript or supplementary information. Sequence data will be freely available on request.

Declarations

Competing interests

The authors declare no competing interests.

Author details

[1]ICAR-National Soybean Research Institute, Indore, India. [2]Murdoch University, Murdoch, Australia. [3]International Crops Research Institute for Semi-Arid Tropics, Hyderabad, India. [4]University of Missouri Columbia, Columbia, MO, USA. [5]ICAR- National Institute of Abiotic Stress Management, Baramati, India.

Received: 9 January 2025 Accepted: 5 May 2025

References

1. Broich SL, Palmer RG. Evolutionary studies of the soybean: the frequency and distribution of alleles among collections of Glycine max and G. soja of Various origin. Euphytica. 1981;30:55–64.
2. Hymowitz T, Newell CA. Taxonomy of the genus *Glycine*, domestication and uses of soybeans. Econ Bot. 1981;35:272–88.
3. Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H, Wang Y. A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. Ann Bot. 2010;106(3):505–14.
4. Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. New Phytol. 2010;188(1):242–53.
5. Garner WW, Allard HA. Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. Mon Weather Rev. 1920;48(7):415–415.
6. Lin X, Liu B, Weller JL, Abe J, Kong F. Molecular mechanisms for the photoperiodic regulation of flowering in soybean. J Integr Plant Biol. 2021;63(6):981–94.
7. Zhang LX, Wei LI, Tsegaw M, Xin XU, Qi YP, Sapey E, Liu LP, Wu TT, Shi SU, Han TF. Principles and practices of the photo-therkongmal adaptability improvement in soybean. J Integr Agric. 2020;19(2):295–310.
8. Gupta S, Agrawal N, Tripathi R, Kumawat G, Nataraj V, Maranna S, Jain M. Long juvenility trait: A vehicle for commercial utilization of soybean (Glycine max) in lower latitudes. Plant Breeding. 2021;140(4):543–60.
9. Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Harada K. Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proc Natl Acad Sci. 2012;109(32):E2155–64.
10. Xu M, Yamagishi N, Zhao C, Takeshima R, Kasai M, Watanabe S, Abe J. The soybean-specific maturity gene E1 family of floral repressors controls night-break responses through down-regulation of FLOWERING LOCUS T orthologs. Plant Physiol. 2015;168(4):1735–46.
11. Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Harada K. A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. Genetics. 2011;188(2):395–407.
12. Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y, Harada K. Map-based cloning of the gene associated with the soybean maturity locus E3. Genetics. 2009;182(4):1251–62.
13. Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, Abe J. Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. Genetics. 2008;180:995–1007.
14. Dissanayaka A, Rodriguez TO, Di S, Yan F, Githiri SM, Rodas FR, Abe J, Takahashi R. Quantitative trait locus mapping of soybean maturity gene E5. Breed Sci. 2016;66:407–15.
15. Cober ER. Long juvenile soybean flowering responses under very short photoperiods. Crop Sci. 2011;51:140–5.
16. Cober ER, Voldeng HD. A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. Crop Sci. 2001;41:698–701.
17. Cober ER, Molnar SJ, Charette M, Voldeng HD. A new locus for early maturity in soybean. Crop Sci. 2010;50:524–7.
18. Kong F, Liu B, Xia Z, Sato S, Kim BM, Watanabe S, Yamada T, Tabata S, Kanazawa A, Harada K, Abe J. Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. Plant Physiol. 2010;154(3):1220–31.
19. Kong F, Nan H, Cao D, Li Y, Wu F, Wang J, Lu S, Yuan X, Cober ER, Abe J, Liu B. A new dominant gene E9 conditions early flowering and maturity in soybean. Crop Sci. 2014;54(6):2529–35.
20. Zhao C, Takeshima R, Zhu J, Xu M, Sato M, Watanabe S, Kanazawa A, Liu B, Kong F, Yamada T, Abe J. 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. 2016;16:20.
21. Wang F, Nan H, Chen L, Fang C, Zhang H, Su T, Li S, Cheng Q, Dong L, Liu B, Kong F, Lu S. A new dominant locus, E11, controls early flowering time and maturity in soybean. Mol. Breed. 2019: 39
22. Fan C, Hu R, Zhang X, Wang X, Zhang W, Zhang Q, Ma J, Fu YF. Conserved CO-FT regulons contribute to the photoperiod flowering control in soybean. BMC Plant Biol. 2014;14:9.
23. Dong L, Cheng Q, Fang C, Kong L, Yang H, Hou Z, Yongli Li Y, Nan H, Zhang Y, Chen Q, Zhang C, Kou K, Su T, Wang L, Li S, Li H, Lin X, Tang Y, Zhao X, Lu S, Liu B, Kong F. Parallel selection of distinct Tof5 alleles drove the adaptation of cultivated and wild soybean to high latitudes. Mol Plant. 2022;15(2):308–21.
24. Li H, Du H, He M, Wang J, Wang F, Yuan W, Huang Z, Cheng Q, Gou C, Chen Z, Liu B. Natural variation of FKF1 controls flowering and adaptation during soybean domestication and improvement. New Phytol. 2023;238(4):1671–84.
25. Wang F, Liu S, Li H, Fang C, Fang S, Wang J, Li S, Liu H, Du H, Wang L, Pei X. Artificial selection of two antagonistic E3 ubiquitin ligases finetunes soybean photoperiod adaptation and grain yield. Proc Natl Acad Sci. 2024;121(45):e2321473121.
26. Li H, Du H, Huang Z, He M, Kong L, Fang C, Chen L, Yang H, Zhang Y, Liu B, Kong F. The AP2/ERF transcription factor TOE4b regulates photoperiodic flowering and grain yield per plant in soybean. Plant Biotechnol J. 2023;21(8):1682–94.
27. Dong L, Fang C, Cheng Q, Su T, Kou K, Kong L, Liu B. Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. Nat Commun. 2021;12(1):5445.
28. Fang C, Chen L, Nan H, Kong L, Li Y, Zhang H, Lu S. Rapid identification of consistent novel QTLs underlying long-juvenile trait in soybean by multiple genetic populations and genotyping-by-sequencing. Mol Breeding. 2019;39:1–11.
29. Kou K, Yang H, Li H, Fang C, Chen L, Yue L, Nan H, Kong L, Li X, Wang F, Wang J, Du H, Yang Z, Bi Y, Lai Y, Dong L, Cheng Q, Su T, Wang L, Li S, Hou Z, Lu S, Zhang Y, Che Z, Yu D, Zhao X, Liu B, Kong F. A functionally divergent SOC1 homolog improves soybean yield and latitudinal adaptation. Curr Biol. 2022;32(8):1728–42.
30. Bernard RL. Two major genes for time of flowering and maturity in soybeans 1. Crop Sci. 1971;11(2):242–4.
31. Buzzell RI. Inheritance of a soybean flowering response to fluorescent-daylength conditions. Can J Genet Cytol. 1971;13(4):703–7.
32. Buzzell R, Voldeng H. Inheritance of insensitivity to long daylength.1980
33. McBlain BA, Bernard RL. A new gene affecting the time of flowering and maturity in soybeans. J Hered. 1987;78(3):160–2.
34. Bonato ER, Vello NA. E6, a dominant gene conditioning early flowering and maturity in soybeans. Genet Mol Biol. 1999;22:229–32.
35. Cober ER, Voldeng HD. Low R:FR light quality delays flowering of E7E7 soybean lines. Crop Sci. 2011;41:1823–6.
36. Tripathi R, Agrawal N, Kumawat G, Gupta S, Kuchlan M, Maranna S, Chand S. Novel role of photo insensitive alleles in adaptation of soybean [Glycine max (L.) Merr.] to rainfed short growing seasons of lower latitudes. Genetic Resources and Crop Evolution. 2021;68:2455–67.
37. Dong L, Cheng Q, Fang C, Kong L, Yang H, Hou Z, Kong F. Parallel selection of distinct Tof5 alleles drove the adaptation of cultivated and wild soybean to high latitudes. Mol Plant. 2022;15(2):308–21.
38. Dong L, Li S, Wang L, Su T, Zhang C, Bi Y, Liu B. The genetic basis of high-latitude adaptation in wild soybean. Curr Biol. 2023;33(2):252–62.
39. Lu S, Dong L, Fang C, Liu S, Kong L, Cheng Q, Chen L, Su T, Nan H, Zhang D, Zhang L. Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. Nat Genet. 2020;52(4):428–36.

40. Kou K, Yang H, Li H, Fang C, Chen L, Yue L, Kong F. A functionally divergent SOC1 homolog improves soybean yield and latitudinal adaptation. Curr Biol. 2022;32(8):1728–42.
41. Ray JD, Hinson K, Mankono JEB, Malo MF. Genetic control of a long-juvenile trait in soybean. Crop Sci. 1995;35(4):1001–6.
42. Lu S, Zhao X, Hu Y, Liu S, Nan H, Li X, Fang C, Cao D, Shi X, Kong L, Su T. Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. Nat Genet. 2017;49(5):773–9.
43. Yue Y, Liu N, Jiang B, Li M, Wang H, Jiang Z, 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.
44. Fang C, Liu J, Zhang T, Su T, Li S, Cheng Q, Liu B. A recent retrotransposon insertion of J caused E6 locus facilitating soybean adaptation into low latitude. J Integr Plant Biol. 2021;63(6):995–1003.
45. Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Belzile F. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE. 2013;8(1):e54603.
46. Torkamaneh D, Boyle B, St-Cyr J, Légaré G, Pomerleau S, Belzile F. NanoGBS: a miniaturized procedure for GBS library preparation. Front Genet. 2020;11:67.
47. Zhai H, Lu S, Liang S, Wu H, Zhang X, Liu B, Xia Z. GmFT4, a homolog of FLOWERING LOCUS T, is positively regulated by E1 and functions as a flowering repressor in soybean. PLoS ONE. 2014;9(2):e89030.
48. Samanfar B, Molnar SJ, Charette M, Schoenrock A, Dehne F, Golshani A, Belzile F, Cober ER. Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. Theor Appl Genet. 2017;130:377–90.
49. Ogiso-Tanaka E, Shimizu T, Hajika M, Kaga A, Ishimoto M. Highly multiplexed AmpliSeq technology identifies novel variation of flowering time-related genes in soybean (*Glycine max*). DNA Res. 2019;26(3):243–60.
50. Li MW, Liu W, Lam HM, Gendron JM. Characterization of two growth period QTLs reveals modification of PRR3 genes during soybean domestication. Plant Cell Physiol. 2019;60(2):407–20.
51. Bernard RL. Two genes affecting stem termination in soybeans 1. Crop Sci. 1972;12(2):235–9.
52. Liu B, Watanabe S, Uchiyama T, Kong F, Kanazawa A, Xia Z, Nagamatsu A, Arai M, Yamada T, Kitamura K, Masuta C. The soybean stem growth habit gene Dt1 is an ortholog of *Arabidopsis* TERMINAL FLOWER1. Plant Physiol. 2010;153(1):198–210.
53. Han X, Xu ZR, Zhou L, Han CY, Zhang YM. Identification of QTNs and their candidate genes for flowering time and plant height in soybean using multi-locus genome-wide association studies. Mol Breeding. 2021;41(6):39.
54. Zhao F, Lyu X, Ji R, Liu J, Zhao T, Li H, Pei Y. CRISPR/Cas9-engineered mutation to identify the roles of phytochromes in regulating photomorphogenesis and flowering time in soybean. The Crop Journal. 2022;10(6):1654–64.
55. Liu Z, Li H, Fan X, Huang W, Yang J, Li C, Qiu LJ. Phenotypic characterization and genetic dissection of growth period traits in soybean (*Glycine max*) using association mapping. PLoS ONE. 2016;11(7):e0158602.
56. Wang J, Zhang Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. Genomics Proteomics Bioinformatics. 2021;19(4):629–40.
57. Minas K, McEwan NR, Newbold CJ, Scott KP. Optimization of a high-throughput CTAB-based protocol for the extraction of qPCR-grade DNA from rumen fluid, plant and bacterial pure cultures. FEMS Microbiol Lett. 2011;325(2):162–9.
58. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011;6(5):e19379.
59. Yao X, Pachner M, Rittler L, Hahn V, Leiser W, Riedel C, Vollmann J. Genetic adaptation of phenological stages in Chinese and European elite soybeans (Glycine max [L.] Merr.) across latitudes in Central Europe. *Plant Breeding*. 2024.
60. Nei M. Molecular evolutionary genetics. Columbia university press 1987.
61. Watterson GA. The genetic divergence of two populations. Theor Popul Biol. 1985;27(3):298–317.
62. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585–95.
63. Breseghello F, Sorrells ME. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics. 2006;172(2):1165–77.
64. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet. 2016;12(2): e1005767. pmid:26828793.
65. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience. 2019;8(2):giy154.
66. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202.

Publisher's Note