

## RESEARCH ARTICLE

## Genome-wide association and epistatic interactions of flowering time in soybean cultivar

Kyoung Hyoun Kim<sup>1,2</sup>, Jae-Yoon Kim<sup>1,2</sup>, Won-Jun Lim<sup>1,2</sup>, Seongmun Jeong<sup>1</sup>, Ho-Yeon Lee<sup>1,2</sup>, Youngbum Cho<sup>1,2</sup>, Jung-Kyung Moon<sup>3</sup>, Namshin Kim<sup>1,2\*</sup>

**1** Genome Editing Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Republic of Korea, **2** Department of Bioinformatics, KRIBB School of Bioscience, University of Science and Technology (UST), Daejeon, Republic of Korea, **3** National Institute of Agricultural Sciences, Rural Development Administration, Jeonju, Republic of Korea

☞ These authors contributed equally to this work.

\* [deepreds@kribb.re.kr](mailto:deepreds@kribb.re.kr)



## OPEN ACCESS

**Citation:** Kim KH, Kim J-Y, Lim W-J, Jeong S, Lee H-Yeon, Cho Y, et al. (2020) Genome-wide association and epistatic interactions of flowering time in soybean cultivar. PLoS ONE 15(1): e0228114. <https://doi.org/10.1371/journal.pone.0228114>

**Editor:** Istvan Rajcan, University of Guelph, CANADA

**Received:** August 29, 2019

**Accepted:** January 7, 2020

**Published:** January 22, 2020

**Copyright:** © 2020 Kim et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** SNP Genotype data of 2,662 soybean cultivars are publicly available at [http://github.com/kyounghyoun/Soybean\\_epistasis](http://github.com/kyounghyoun/Soybean_epistasis).

**Funding:** This work is supported by grants from the National Research Foundation of Korea (NRF-2014M3C9A3064552), Rural Development Administration, Republic of Korea (Project No. PJ01313201) and the KRIBB Initiative program. The funders had no role in study design, data

## Abstract

Genome-wide association studies (GWAS) have enabled the discovery of candidate markers that play significant roles in various complex traits in plants. Recently, with increased interest in the search for candidate markers, studies on epistatic interactions between single nucleotide polymorphism (SNP) markers have also increased, thus enabling the identification of more candidate markers along with GWAS on single-variant-additive-effect. Here, we focused on the identification of candidate markers associated with flowering time in soybean (*Glycine max*). A large population of 2,662 cultivated soybean accessions was genotyped using the 180k Axiom<sup>®</sup> SoyaSNP array, and the genomic architecture of these accessions was investigated to confirm the population structure. Then, GWAS was conducted to evaluate the association between SNP markers and flowering time. A total of 93 significant SNP markers were detected within 59 significant genes, including *E1* and *E3*, which are the main determinants of flowering time. Based on the GWAS results, multilocus epistatic interactions were examined between the significant and non-significant SNP markers. Two significant and 16 non-significant SNP markers were discovered as candidate markers affecting flowering time via interactions with each other. These 18 candidate SNP markers mapped to 18 candidate genes including *E1* and *E3*, and the 18 candidate genes were involved in six major flowering pathways. Although further biological validation is needed, our results provide additional information on the existing flowering time markers and present another option to marker-assisted breeding programs for regulating flowering time of soybean.

## Introduction

A genome-wide association study (GWAS) is one of the promising approaches for the identification of genomic variants responsible for specific phenotypes [1]. With the introduction of

collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

high-density marker arrays, GWAS has been actively used in many crop species [2] and has enabled the discovery of single nucleotide polymorphism (SNP) markers associated with numerous agronomic traits [3]: flowering time [4, 5], cold tolerance [6], salt tolerance [7, 8], drought tolerance [9], disease resistance [10, 11], plant height [12, 13], leaf architecture [14, 15], and seed weight [16, 17]. The knowledge of the trait-related SNP markers has served as the genetic basis for the improvement of various traits in crop breeding programs [18].

Days to flowering (DTF) is a crucial agronomic trait that regulates the maximum use of sunlight and temperature [19], and affects the growth and yield potentials [20]. Many GWAS on single-variant-additive-effect thus have carried out to improve yield productivity, and have led to identify DTF markers in many crops. For example, three candidate genes including *Nsn1*, *Fpa*, and *Zmm22* were identified in 942 maize samples (*Zea mays*) [21]; two candidate genes, *CO1* and *BFL*, were identified in using 218 barley samples (*Hordeum vulgare*) [22]; eight candidate genes including *Hd1* were confirmed in 950 rice samples (*Oryza sativa*) [23]; and ten candidate genes including *SOC1*, *AGL6*, and *ELF8* were reported in 309 soybean samples (*Glycine max*) [13]. These findings have presented valuable information to various breeding programs focused on DTF, but have a limitation to further improve DTF, because they are not sufficient for explaining all of the phenotypic variations in DTF such as interaction effects between markers [24].

Epistasis is defined as the interaction between genes or SNP markers that influences a trait [25]. Each SNP marker above a significant level in GWAS has a strong effect on the determination of a trait, but non-significant markers that interact with each other could also have a large influence on the trait [24]. Therefore, considering epistatic interactions for multi-variant non-additive effects, enables to discover more markers associated with traits, together with GWAS on single-variant-additive-effects [26]. For this reason, many GWAS studies have utilized epistatic analysis as a complementary approach, and have reported significant epistasis and GWAS markers associated with various traits: 12 epistatic markers with 12 GWAS markers for sudden death syndrome resistance in soybean [27], ten epistatic markers with 33 GWAS markers for iron deficiency-related chlorosis in soybean [28], nine epistatic markers with 14 GWAS markers for seed weight in soybean [29], and 38 epistatic markers with 113 GWAS markers for DTF in barley [30]. Significant markers controlling many traits have been investigated through these studies; however, soybean's DTF markers related to both epistasis and GWAS have not been investigated in depth as other traits.

Twelve major genes affecting DTF have been identified in soybean, including *E1-E10* [31–41], *J* [42], and *Dt1* [43, 44]. To better understand the DTF-related genetic factors beyond the twelve major genes, an epistasis study with GWAS is needed. Here, we genotyped 2,662 cultivated soybean accessions using the high-density Axiom<sup>®</sup> 180k SoyaSNP array developed by our team in 2016 [45]. Also, we evaluated the DTF trait of all accessions through a phenotypic survey. The aims of this study were: 1) to examine population structures of the 2,662 accessions for GWAS; 2) to identify significant markers associated with DTF through GWAS; 3) to reveal epistatic markers with interactions between significant and non-significant GWAS markers; and 4) to present final candidate markers with the relation to major DTF pathways.

## Materials and methods

### Plant materials and genotyping

A total of 2,872 soybean (*Glycine max*) accessions were collected from the National Agrobiodiversity Center in the Rural Development Administration (RDA, Jeonju, Korea), and were genotyped using the Axiom<sup>®</sup> 180k SoyaSNP. Of these, 210 hybrid accessions were excluded from the analysis. The remaining 2,662 accessions originated from South Korea (2,415), North

Korea (96), the USA (60), China (59), and Japan (32), and comprised 335 improved cultivars (ICs), 2,175 landraces (LRs), and 152 unknown cultivars (UCs) (S1 Dataset). These accessions belonged to the following seven maturity groups: I (2), II (31), III (60), IV (187), V (88), VI (16), and VII (3); however, the maturity group of most of the Korean accessions (2,275) was not reported. After genotyping, 180,961 SNP markers were detected, and haplotype phasing and imputation were conducted using BEAGLE version 3.3.1 [46]. Then, 78,427 SNP markers with a minor allele frequency (MAF) > 0.05 were obtained. The non-MAF filtered SNPs were used to examine the genomic structure and relationship, and the MAF filtered SNPs were used to perform the GWAS. The genotype data of all 2,662 accessions, which were generated in collaboration with the RDA, is available at <http://k-crop.kr> and [https://github.com/kyounghyoun/Soybean\\_epistasis](https://github.com/kyounghyoun/Soybean_epistasis) [47]. Phenotypic evaluation was conducted in the experimental field of the National Institute of Crop Science (NICS, Jeonju, Korea) (35°50'26.7" N, 127°02'42.7" E), and the DTF trait was measured from June to October in 2014. Because of an agreement on limited disclosure with our phenotypic evaluation team, the raw DTF values are provided in the distribution plot and categorical data (Fig 1f, S1 Dataset).

### Analysis of the genomic structure of and relationship among soybean accessions

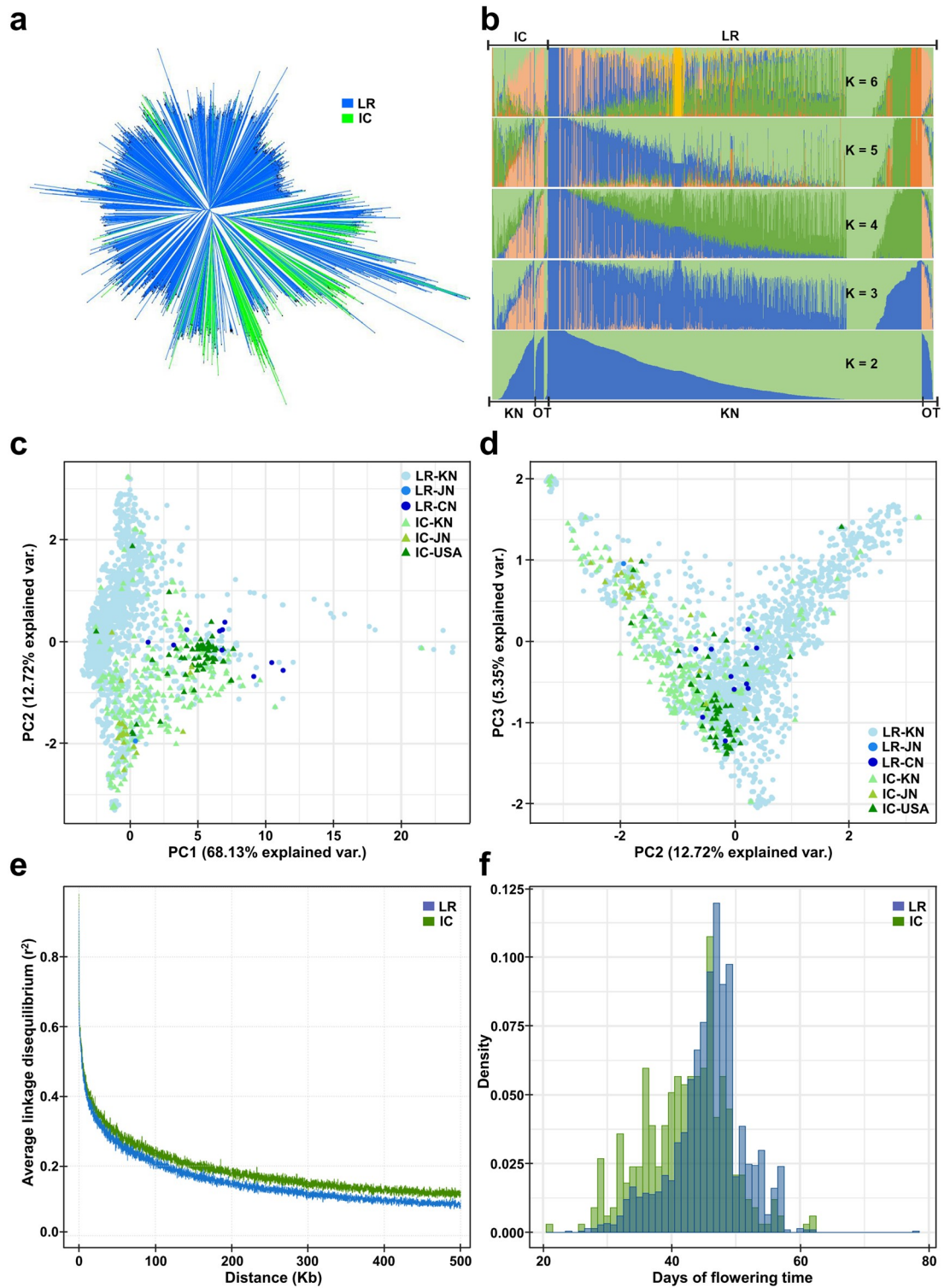
Phylogenetic analysis was conducted based on the identity-by-state coefficient matrix calculated using Plink v1.90b [48]. The phylogenetic tree was reconstructed using the BIO-neighbor-joining algorithm [49] and FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) (Fig 1a). Structure analysis was conducted using fastSTRUCTURE v1.0 [50], based on a variational Bayesian framework (Fig 1b). The number of genetic clusters ( $K$ ) was set at six, considering the error values of 10-fold cross-validation (S1 Fig). The error value was the lowest at  $K = 6$ . Principal component analysis (PCA) was performed by applying singular value decomposition to the distant matrix calculated using the Kimura two-parameter model [51], and then displayed using the PC axes 1, 2, and 3 (Fig 1c and 1d). The linkage disequilibrium (LD) pattern was computed using PopLDdecay v3.2 [52]. The mean value of LD was calculated within 100 and 500 kb regions (S1 Table), and the degree of LD up to 500 kb is displayed in Fig 1e. Nucleotide diversity ( $\pi$ ) was calculated using a 10 kb slide size with a 100 kb window size using VCFtools v4.2 [53], and inbreeding coefficient ( $F$ ) was estimated using the method of moments in VCFtools (S1 Table). Additionally, using the same software, fixation index value ( $F_{st}$ ) was calculated using a 10 kb slide size with a 100 kb window size.

### GWAS

An association study between genomic regions and flowering time was conducted using a compressed mixed linear model within Genomic Association and Prediction Integrated Tool [54]. The MAF-filtered 78,427 SNP markers (MAF > 0.05) was used to consider common markers (S2 Dataset), and a statistically significant cut-off value was adopted to  $-\log(p\text{-value}) = 7$ . To minimize false-positive results, the cut-off was set slightly higher than the Bonferonni-corrected  $-\log(p\text{-value})$  of 6.2. The extent of model fitting was confirmed using a quantile–quantile (Q-Q) plot for the expected and obtained  $p$ -values. After GWAS based on SNP markers, gene annotations were performed using SnpEff v4.3 [55]. All results are summarized in S3 Dataset, and genome-wide plots including the Q-Q plot are shown in Fig 2.

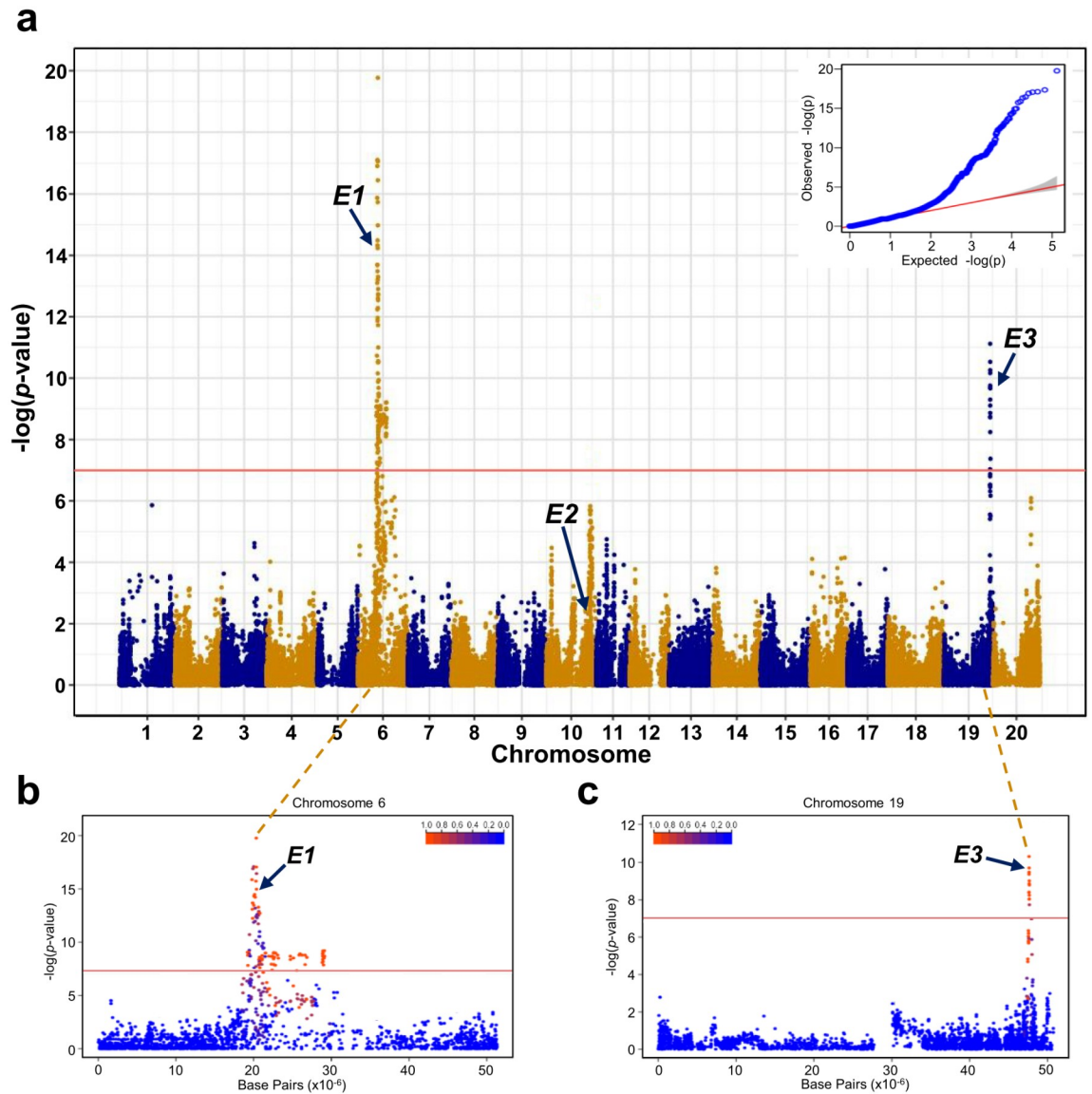
### SNP marker set for epistatic analysis

To construct an SNP marker-set for multilocus epistatic interaction analysis, a biological filtering approach was conducted. First, genes affecting DTF and maturity in *Arabidopsis* were



**Fig 1. Genomic structure of 2,662 cultivated soybean accessions and their relationship among each other.** a. Phylogenetic tree computed using the identical-by-state coefficient. b. Population structure analysis using the number of genetic clusters ( $K$ ) ranging from  $K = 2$  to  $K = 6$ , based on the maximum likelihood-based clustering algorithm. c. Principal component analysis (PCA) plot of PC1 and PC2 derived using the Kimura two-parameter model. d. PCA plot of PC2 and PC3. e. Extent of linkage disequilibrium (LD) decay computed up to 500 kb. f. Distribution of the flowering time of soybean LRs and ICs. Abbreviations, KN, JN, CN, and OT, indicate accessions collected from Korea, Japan, China, and the other regions, respectively.

<https://doi.org/10.1371/journal.pone.0228114.g001>



**Fig 2. Manhattan plot for the DTF trait of soybean accessions.** a. Manhattan plot of all 20 chromosomes. b. Manhattan plot of chromosome 6 harboring the *E1* gene. c. Manhattan plot of chromosome 19 harboring the *E3* gene. Red horizontal lines indicate the statistically significant cut-off of  $-\log(p\text{-value}) = 7$ .

<https://doi.org/10.1371/journal.pone.0228114.g002>

**Table 1. Significant genes related to DTF derived from GWAS, and non-significant genes related to DTF and maturity in *Arabidopsis*.**

DTF	Number of <i>Arabidopsis</i> genes	Corresponding soybean genes	Genes on the 180k SoyaSNP array	SNP markers on the 180k SoyaSNP array
GWAS <sup>a</sup>	55	59	59	93
Related genes <sup>b</sup>	353 (3)	2,639 (4)	2,184 (4)	5,441 (6)
Total	408	2,698	2,243	5,534

<sup>a</sup>Significant genes related to DTF derived from GWAS

<sup>b</sup>Non-significant genes related to DTF and maturity in *Arabidopsis*. Values in parenthesis indicate the number of genes that overlapped with significant genes in GWAS.

<https://doi.org/10.1371/journal.pone.0228114.t001>

examined [56–59], and a total of 356 *Arabidopsis* genes were identified (Table 1, S4 Dataset). Second, nucleotide sequences of these *Arabidopsis* genes were mapped onto the soybean reference genome Wm82.a2.v1 using tBLASTx [60] with matching identity >80% and alignment length >70 bp options, and a total of 2,643 homologous soybean genes were identified. Third, 2,188 genes belonging to the Axiom<sup>®</sup> 180k SoyaSNP array were identified among the 2,643 soybean genes, and 5,447 SNP markers contained in the 2,188 genes were obtained. Finally, the 5,447 SNP markers were combined with 93 significant genic SNP markers derived in the GWAS ( $-\log(p\text{-value}) > 7$ ), and a total of 5,534 SNP markers (excluding 6 overlapping SNP markers) was constructed as the final SNP marker-set for the multilocus epistatic analysis. All results, including the number of SNP markers per gene, are summarized in S4 Dataset, and a brief summary is provided in Table 1.

### Multilocus epistatic interactions

Multilocus interactions among the 5,534 SNP markers related to DTF were examined using Bayesian High-order Interaction Toolkit (BHIT) [61] with 1,000,000 iterations and 990,000 burn-in periods. The relationship between SNPs (X) and trait (Y) in multilocus interactions was inferred as  $P(Y, H|I, H)$  by grouping dependent genotypes (I) and phenotypes (H) [61]. Then, the likelihood was estimated using the following equation:

$$P(Y, H|I, H) = P(Y|X_{\{I=1\}}) \prod_{h=1}^H P(X_{\{I=h\}}|I) = \left( \prod_{m=1}^M (P(Y_{\{m\}}|X_{\{I=1\}} = m)) \right) \left( \prod_{m=1}^H P(X_{\{I=h\}}|I) \right)$$

where,  $X_{\{I=h\}}$  indicates all X in the  $h^{\text{th}}$  group, and  $Y_{\{m\}}$  indicates Y in the  $m^{\text{th}}$  cluster. For  $P(X_{\{I=h\}}|I)$ , Dirichlet prior distribution and multinomial distribution were used as the Bayesian partition model [62]. The threshold for the posterior probabilities on the dependency for multilocus and phenotype was set at 0.5. A Markov Chain Monte Carlo approach was then used to search potential epistatic marker sets, and a Bayesian computational approach was utilized to detect final high-order epistatic marker sets. All results of multilocus epistatic interactions are summarized in Table 2, and displayed in Circular-Manhattan plots in Fig 3. To more complement the multilocus epistasis, two-locus epistatic interactions were investigated to the BHIT results. The two-locus epistatic test was conducted using a fixed linear regression model of the Plink v1.90b [48], and the resulting  $p$ -values are summarized in Table 2.

## Results

### Genomic structure of cultivated soybeans

A large data set of 2,662 cultivated soybean accessions, originated from Korea (2,511), the USA (60), China (59), and Japan (32), were used in this study. These accessions included 335 ICs, 2,175 LRs, and 152 UCs (S1 Dataset). Korean accessions accounted for approximately 94.33% of the entire soybean collection, and covered with most of the soybean accessions in Korea. These 2,662 accessions were genotyped using the Axiom<sup>®</sup> 180k SoyaSNP array developed by our team in 2016 [45], and a total of 180,961 SNP markers were identified. These SNP markers represented 39,674 soybean genes, with an average number of 2.80 SNP markers per gene. Prior to GWAS, the genomic structure of 2,510 accessions (2,175 LRs and 335 ICs, excluding 152 UCs) was examined using the SNP markers (Fig 1). These 2,510 accessions represented a single large population with considerable genomic diversity. A phylogenetic tree was constructed to radial forms rather than branched forms with distinct clades, and LRs and ICs exhibited a mixed aspect with each other (Fig 1a). Structure analysis revealed similar genomic composition and proportion between LRs and ICs at  $K = 2$  (Fig 1b; green and blue). With the

**Table 2. Multilocus epistatic interactions among 5,534 SNP markers related to DTF.**

Group <sup>a</sup>	Chr.	Position (bp)	Soybean gene ID	Gene symbol <sup>b</sup>	<i>p</i> -value (min, max) <sup>c</sup>	Protein	Quantitative trait loci (QTLs)	Related pathway <sup>d</sup>
<b>A. Significant interaction groups including significant DTF-related genes in GWAS</b>								
G1	6	20,208,679	Glyma.06G207800	<i>*E1</i>	(3.69E-03, 3.69E-03)	AP2/B3-like transcriptional factor family protein	First flower 4-1, 5-1, 12-2, 23-1, 26-9, 26-12	P
	16	1,855,885	Glyma.16G020400	<i>GIF3</i>	(3.69E-03, 3.69E-03)	GRF1-interacting factor 3	Pod maturity 9-1	FPI
G2	6	20,207,322	Glyma.06G207800	<i>*E1</i>	(5.36E-03, 8.21E-03)	AP2/B3-like transcriptional factor family protein	First flower 4-1, 5-1, 12-2, 23-1, 26-9	P
	2	45,220,188	Glyma.02G267800	<i>COP1</i>	(8.21E-03, 1.56E-02)	Transducin family protein/WD-40 repeat family protein	-	P
	13	37,658,295	Glyma.13G274900	<i>SPL12</i>	(5.36E-03, 1.56E-02)	Squamosa promoter-binding protein-like 12	Plant height 38-2	FMI
G3	6	20,207,322	Glyma.06G207800	<i>*E1</i>	(3.97E-04, 7.40E-01)	AP2/B3-like transcriptional factor family protein	First flower 4-1, 5-1, 12-2, 23-1, 26-9, 26-12	P
	4	6,893,070	Glyma.04G081800	<i>EMF1</i>	(1.27E-25, 7.25E-02)	Embryonic flower 1	-	V
	6	19,585,253	Glyma.06G205800	<i>FUL</i>	(2.61E-04, 9.17E-02)	AGAMOUS-like 8	First flower 4-1, 5-1, 12-2, 26-9, 26-12	FMI
	8	3,655,383	Glyma.08G046500	<i>FKF1</i>	(1.27E-25, 2.26E-02)	Flavin-binding, kelch repeat, f box 1	-	P
	10	42,450,765	Glyma.10G192000	<i>NF-YB8</i>	(3.97E-04, 9.17E-02)	Nuclear factor Y, subunit B8	Pod maturity 14-2, 15-2	FPI
G4	19	47,638,344	Glyma.19G224200	<i>*E3</i>	(1.84E-03, 1.84E-03)	Phytochrome A	First flower 4-g80	P
	15	8,774,152	Glyma.15G111900	<i>STM</i>	(1.84E-03, 1.84E-03)	KNOX/ELK homeobox transcription factor	First flower 12-3	P
G5	19	47,638,344	Glyma.19G224200	<i>*E3</i>	(2.46E-06, 5.94E-03)	Phytochrome A	First flower 4-g80	P
	3	44,468,848	Glyma.03G248200	<i>FLK</i>	(4.85E-03, 5.94E-03)	RNA-binding KH domain-containing protein	Plant height 26-17	AU
	10	42,457,595	Glyma.10G192000	<i>NF-YB8</i>	(2.46E-06, 4.85E-03)	Nuclear factor Y, subunit B8	Pod maturity 14-2, 15-2	FPI
<b>B. Significant interaction groups among non-significant DTF-related genes</b>								
G6	4	46,807,590	Glyma.04G196200	<i>HAP5A</i>	(3.89E-10, 1.82E-01)	Nuclear factor Y, subunit C1	Plant height 5-4; Pod maturity 6-3, 6-6, 32-1	FPI
	6	26,630,946	Glyma.06G221000	<i>ARP6</i>	(1.64E-03, 2.52E-01)	Actin-related protein 6	First flower 3-1, 8-1, 22-2, 23-1, 26-10, 26-11, 26-13, 26-16	AU
	6	19,586,329	Glyma.06G205800	<i>FUL</i>	(3.89E-10, 2.53E-06)	AGAMOUS-like 8	First flower 4-1, 5-1, 12-2, 26-9, 26-12	FMI
	10	49,716,656	Glyma.10G274300	<i>BBX15</i>	(2.53E-06, 2.52E-01)	B-box type zinc finger protein with CCT domain	-	P
G7	2	1,109,049	Glyma.02G012100	<i>SWN</i>	(1.04E-03, 2.20E-01)	SET domain-containing protein	-	AU, V
	9	6,040,554	Glyma.09G063100	<i>FVE</i>	(2.13E-12, 1.04E-03)	Transducin family protein/WD-40 repeat family protein	-	AM, AU
	15	22,687,645	Glyma.15G196500	<i>PHYE</i>	(1.93E-04, 4.36E-01)	Phytochrome E	Pod maturity 31-2, 37-2	P
	15	40,764,606	Glyma.15G223700	<i>CRY2</i>	(2.13E-12, 4.36E-01)	Cryptochrome 2	Plant height 26-10	P

\*Major DTF-related genes (*E1* and *E3*) identified from GWAS.

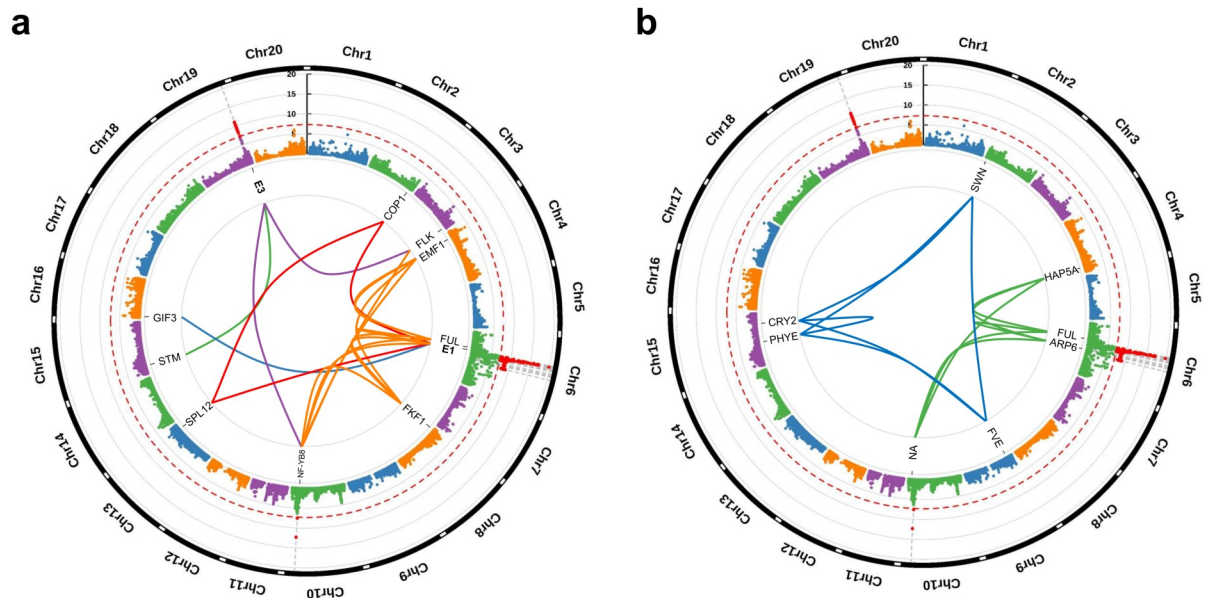
<sup>a</sup>Interaction group derived from multilocus epistatic analysis for DTF.

<sup>b</sup>Representative gene symbol corresponding the soybean gene.

<sup>c</sup>Range of minimum and maximum *p*-values for two-locus interactions, calculated pairwise among SNP markers in an interaction group.

<sup>d</sup>DTF-related pathways: AM, ambient temperature; AU, autonomous; FPI, floral pathway integrator; P, photoperiod; FMI, floral meristem identity; V, vernalization

<https://doi.org/10.1371/journal.pone.0228114.t002>



**Fig 3. Epistatic interactions among DTF-related genes.** a. Epistatic interactions between non-significant genes and significant genes (*E1* and *E3*). Inner curved lines indicate interaction groups, G1 (navy), G2 (red), G3 (orange), G4 (green), and G5 (purple) (Table 2). b. Epistatic interaction between non-significant genes. Inner curved lines indicate interaction groups, G6 (green) and G7 (navy) (Table 2). Red dotted lines in the two plots indicate statistically significant cut-off thresholds of  $-\log(p\text{-value}) = 7$ .

<https://doi.org/10.1371/journal.pone.0228114.g003>

increase in the value of  $K$ , the entire IC group showed a unique genomic composition, which slightly distinguished from the LR group; however, each accession in the IC group exhibited an aspect of inclusion in the LR group, with varying levels of genomic composition. The results of PCA showed that the LR and IC groups formed a single group with large diversity, and the IC group overlapped with the LR group, consistent with the results of structure analysis (Fig 1c and 1d). Average values of LD,  $\pi$ , and DTF showed differences between the IC and LR groups. Due to the impact of conscious selection for improvement, the IC group showed slightly higher average values of LD (0.121583) (Fig 1e, S1 Table), and lower  $\pi$  ( $2.22E-5$ ) and DTF (42.16) (Fig 1f, S1 Dataset) than the LR group (0.112253,  $2.39E-5$ , and 45.86, respectively). Values of inbreeding coefficient were all high (IC,  $F = 0.9710$ ; LR,  $F = 0.9731$ ) because of the characteristics of inbreeding and stringent cleistogamy (S1 Table). Also, a numeric value  $F_{st}$ , supported the results of various population analyses. The  $F_{st}$  between the IC and LR groups was 0.0581 and was small enough to be regarded as a single group. Based on these results, a single dataset consisting of all 2,662 accessions was constructed, and used for subsequent GWAS and epistatic analysis. The results of PCA indicated that 132 UCs overlapped between the LR and IC groups; these accessions were included in GWAS to reflect as much genomic diversity of cultivated accessions as possible.

### GWAS of DTF

Of the 180,961 SNP markers identified in 2,662 accessions, 78,427 SNP markers with  $MAF > 0.05$  were selected as common SNP markers for GWAS. These SNP markers were distributed on all 20 soybean chromosomes and represented 26,162 genes, with an average number of 2.07 SNP markers per gene (S2 Dataset). Then, the association between these SNP markers and DTF was examined using a compressed mixed linear model that adjusts population structure and kinship (Fig 2, S3 Dataset). Statistically significant cut-off was set at  $-\log(p-$



value) = 7, which was higher than the Bonferroni-adjusted  $p$ -value = 6.2, to minimize false-positive results. A total of 132 significant SNP markers including 93 genic SNP markers were detected, and all mapped to chromosomes 6 and 19 (Fig 2a). Among these 93 significant genic SNP markers, 78 SNP markers mapped to 51 soybean genes on chromosome 6 (Fig 2b), and 15 SNP markers mapped to eight genes on chromosome 19 (Fig 2c).

The two major DTF-related genes, *E1* and *E3*, were identified from the 59 significant genes. The *E1* gene has the largest influence on the determination of DTF [37, 63] and affects vegetative development by regulating multiple genes related to plant growth [64]. The *E1* gene was located at the 20,207,322 bp position on chromosome 6 (Glyma.06G207800), and showed a considerably significant  $-\log(p\text{-value})$  of 14.23. The *E3* gene regulates DTF under short-day and long-day conditions, and affects plant maturity [65]. This gene was positioned at 47,638,344 bp on chromosome 19 (Glyma.19G224200) and exhibited a significant  $-\log(p\text{-value})$  of 9.75. Two additional major genes affecting DTF, *E2* and *E4* [66, 67], were not detected as significant genes in this study. The *E2* gene was located at nucleotide positions 45,295,453, 45,295,508, 45,296,750, and 45,300,271 bp on chromosome 10 (Glyma.10G221500), but all showed non-significant  $p$ -values of 2.00, 2.43, 2.31, and 2.43, respectively (Fig 2a). Also the *E4* gene was positioned at 33,236,286, 33,236,286, and 33,241,589 bp on chromosome 20 (Glyma.20G090000), but all were excluded from the GWAS because of low MAF (0.000187, 0.000909, and 0.002441, respectively).

### Epistatic interaction among DTF-related genes

Interaction effects on DTF among significant SNP and non-significant SNP markers were examined to complement GWAS single-variant-additive-effect. Significant SNP markers consisted of 93 genic SNP markers belonging to 59 soybean genes derived from the GWAS, as described above. To focus on interactions among DTF-related genes, 5,441 non-significant SNP markers in 2,184 soybean genes (Table 1) were selected from a set of 356 DTF-related *Arabidopsis* genes on the Axiom<sup>®</sup> 180k SoyaSNP array (S4 Dataset) (see Materials and methods for details). In the 356 *Arabidopsis* genes, a total of 2,188 homologous soybean genes were identified, but four genes (Glyma.06G205700, Glyma.06G207800-*E1*, Glyma.06G221000, and Glyma.19G224200-*E3*) with six SNP markers were excluded from the non-significant SNP marker-set since these markers were already included in the significant marker-set in GWAS. Also, when detecting homologous genes, the non-MAF filtered SNPs were used in order to consider various DTF-related genes which were excluded from the GWAS due to  $MAF < 0.05$ , such as *E4* gene (Glyma.20G090000). Finally, a total of 5,534 SNP markers belonging to 2,243 soybean genes were selected and used for analyzing multilocus epistatic interactions. The results revealed five interaction groups (G1–G5) containing two significant and nine non-significant genes (Fig 3a) and two interaction groups (G6, G7) containing only eight non-significant genes (Fig 3b, Table 2).

The interaction groups G1, G2, and G3 contained two, three, and five genes, respectively, each including the significant DTF-related *E1* gene (Fig 3a, Table 2). The non-significant genes, identified in GWAS, included *GIF3* (belonging to G1 group and quantitative trait locus (QTL) of pod maturity 9–1), which is involved in regulating cell expansion and meristem of leaves [68]; *SPL12* (G2, plant height QTL 38–2), which affects plant growth and development [69]; *COPI* (G2), which regulates photomorphogenesis and skotomorphogenesis, and is related to growth and development [70]; *FUL* (G3, first flower QTLs 4–1, 5–1, 12–2, 26–9, and 26–12), which functions early in controlling flowering time [71, 72]; *EMF1* (G3), which is involved in DTF through the regulation of reproductive development [73, 74]; *FKF1* (G3), involved in DTF by regulating changes in photoperiod and temperature [75, 76]; and *NF-YB8* (G3 and G5,

pod maturity QTLs 14–2 and 15–2), which regulates leaf development and maturity [77]. The interaction groups G4 and G5 possessed the *E3* gene in common, together with one and two other genes, respectively (Fig 3a, Table 2), including *STM* (G4, first flower QTL 12–3), which is related to flower meristem and DTF [78]; *FLK* (G5, plant height QTL 26–17), which regulates DTF through the repression of *FLOWERING LOCUS C* expression and its post-transcriptional modification [76]; and *NF-YB8* (G5), also detected in G3, as described above.

The G6 and G7 interaction groups contained only eight non-significant genes identified in GWAS, but were detected as significant interaction groups on DTF in the epistatic analysis (Fig 3b, Table 2). These eight genes included *HAP5A* (G6, one plant height QTL and three pod maturity QTLs), related to earlier flowering [79]; *ARP6* (G6, eight first flower QTLs), involved in plant growth and development [80, 81]; *FUL* (G6), also identified in G3, as described above; *BBX15* (G6), indirectly related to light reaction [82]; *SWN* (G7), involved in the regulation of flowering and development [83, 84]; *FVE* (G7), associated with the regulation and control of DTF [85, 86]; and *PHYE* and *CRY2* (G7, two pod maturity QTLs and one plant height QTL), related to light reaction during flowering [87, 88].

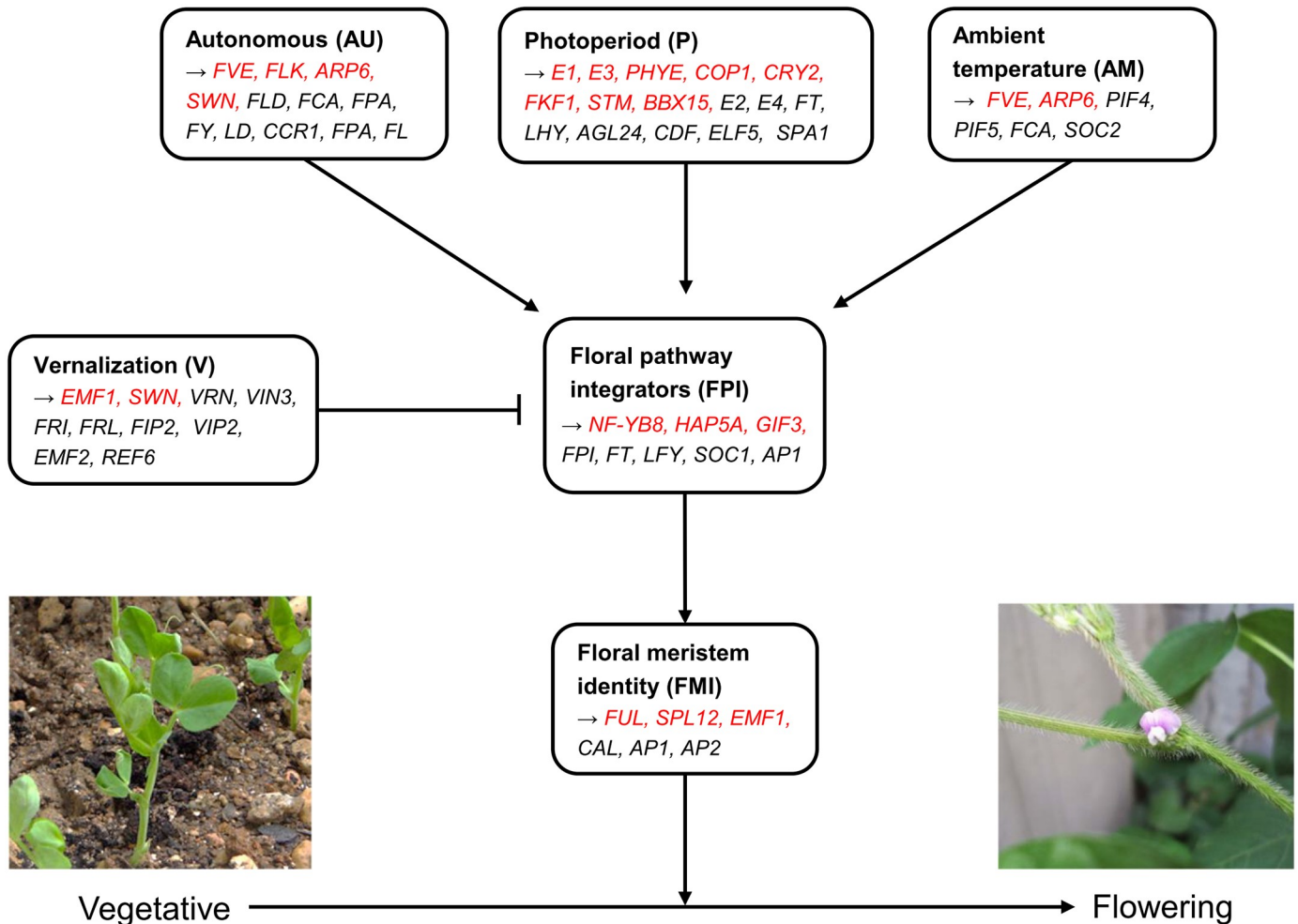
### Pathways of epistatic genes related to DTF

To identify the degree of involvement in DTF, two significant and 16 non-significant genes in all interaction groups were examined on DTF-related pathways. These 18 candidate genes were involved in the following six major DTF-related pathways [56]: ambient temperature (AM), autonomous (AU), flowering pathway integrator (FPI), photoperiod (P), floral meristem identity (FMI), and vernalization (V) (Table 2, Fig 4). The V and P pathways control the overall flowering process [89] and the AM pathway regulates flowering time [90, 91], as external factors. The AU pathway promotes induction of flowering as an internal factor [89]. The FPI pathway regulates the other pathways and triggers the induction of FMI [89], and the FMI pathway induces flowering. The P pathway contained eight candidate genes including the *E1* and *E3* genes; A and AM pathways contained two candidate genes each; AU and FPI pathways contained four candidate genes each; and the FMI pathway contained three candidate genes. These results confirmed that all 18 candidate genes are involved in major DTF-related pathways, and presented that they have a wide effect on DTF while interacting with each other.

### Discussion

The Axiom<sup>®</sup> 180k SoyaSNP array was developed by our team mainly based on Korean soybean accessions, with the aim to perform GWAS using a large number of high-density markers [45]. Of the 2,662 soybean accessions used in this study (S1 Dataset), 94% were of Korean origin and therefore suitable for genotyping with the Axiom<sup>®</sup> 180k SoyaSNP high-density array, generating 180,961 SNP markers. These 2,662 soybean accessions mainly comprised two subgroups, LRs and ICs, and have a tendency to form a single large population with considerable genomic diversity (Fig 1). The *Fst* value, a numeric value indicating population structure, was also sufficiently small (0.0581) for the two subgroups to be considered as a single group (S1 Table). Based on the results of population structure analyses, we conducted GWAS on all 2,662 soybean accessions and identified 59 soybean genes, including *E1* and *E3*, as candidates associated with DTF (Fig 2, S2 and S3 Datasets).

Genes *E1–E10* [31–41], *J* [42], and *Dt1* [43, 44] are known as major DTF-related genes in soybean. Among these, genes *E1–E4* act as the major determinants of DTF, as these genes account for approximately 62–66% of variation in DTF [40, 57]. The *E1* gene has the largest effect on DTF and photoperiod sensitivity [37, 63] through regulations of *GmFT2a*, *GmFT5a*,



**Fig 4. Genes in major pathways regulating plant growth from the vegetative stage to the flowering stage.** Representative genes are displayed in corresponding pathways. Candidate genes identified from interaction groups are shown in red (“Related pathway” column in Table 2).

<https://doi.org/10.1371/journal.pone.0228114.g004>

and *GmFT1a* genes [92, 93]. The *E2* gene enhances photoperiod response and contributes to early flowering by regulating *GmFT2a* [66, 94]. Genes *E3* and *E4* are involved in the regulation of *E1* and affect DTF through the regulation of the far-red light ratio [65, 67]. Above all, *E1* functions as a key regulator of DTF in soybean [95]. In our GWAS, *E1* and *E3* genes showed significant *p*-values, whereas *E2* and *E4* genes were not detected (Fig 2). The *E2* gene showed a non-significant *p*-value and the *E4* gene was excluded from the GWAS because of considerably low MAF of SNPs. According to previous studies, the *E1–E4* genes in soybean are considerably affected by habitat ecology and latitude [57, 96–98]. The soybean population used in this study consists mostly of Korean accessions and therefore reflects the Korean environment. Korean soybeans are also reported to possess high genomic diversity, unlike soybean accessions from other regions [99–101]. Considering the genomic characteristic of our population, the GWAS result suggests that *E1* and *E3* genes play a significant role in the regulation of DTF trait in Korean soybean population.

In addition to GWAS results, the interaction effects between 59 significant genes and the remaining non-significant genes on DTF were examined (Table 1). Identification of epistatic interactions could help in the detection of a greater number of candidate genes, since the

complex DTF trait is affected by interactions among multiple genes [29, 102]. In our epistatic analysis, non-significant genes were utilized as 2,188 soybean genes which are homologous genes to 356 DTF-related *Arabidopsis* genes (S4 Dataset). This approach of marker reduction, called biological filtering, has been reported as one of the effective approaches for epistatic analysis [103, 104], as it allows the overcoming of computational limitations caused by a considerable number of marker combinations [105]. Therefore, we selected 356 *Arabidopsis* genes based on previous reports on DTF and maturity (S4 Dataset), and used to analyze the interaction effects of various DTF-related genes on DTF. The results of epistatic analysis revealed seven interaction groups and 18 candidate genes with significant influence on DTF (Table 2); *E1* and *E3* genes, which were detected in GWAS, also played a major role in the interaction groups (Fig 3).

DTF is regulated by complex networks of biological processes [106, 107], and is characterized to six representative pathways from vegetative to flowering stages in *Arabidopsis* [108, 109]: P (response to day length and light), V (cold exposure), AU (flowering promotion), FPI (floral transition), AM (time of floral transition), and FMI (floral development) (Fig 4). To better understand the genetic effects of candidate genes on DTF, it is necessary to identify the extent to which these candidate genes are involved in the six major pathways. Among the 18 candidate genes identified in this study, eight genes including *E1* (interaction groups G1–G3) are involved in P, FPI, FMI, and V pathways; four genes including *E3* (G4 and G5) are associated with P, AU, and FPI pathways; and eight genes (G6 and G7) are related to P, V, AU, AM, FMI, and FPI pathways (Table 2, Fig 4). Thus, all of these 18 candidate genes were evenly distributed in the six major pathways and affected DTF while interacting with each other in the corresponding interaction groups. Overall, our results of GWAS and epistatic analysis suggest that these 18 candidate genes play a significant role in the regulation of DTF in soybean, at least in the Korean population.

However, our findings have several limitations. First, the DTF-related candidate markers identified in this study do not represent the worldwide soybean population. It is possible that these candidate markers are specific to Korean soybean accessions because 94% of the soybean population was of Korean origin. Second, the epistatic analysis was restricted to only 356 DTF-related *Arabidopsis* genes. The 356 genes were used to reduce computational overload; however, markers excluded in the epistatic analysis could also have a significant effect on DTF while interacting with each other. Finally, our candidate markers have not been validated by biological experiments. To minimize this limitation, we examined the genomic characteristics of the soybean population before GWAS and used a conservative cut-off in GWAS, which was higher than the Bonferroni-adjusted *p*-value. Despite these efforts, our candidate markers need further experimental validation, but they are expected to present valuable information to soybean breeding programs aimed at improving the DTF trait.

## Conclusion

DTF is an important agronomic trait relevant to plant growth, development, and productivity in soybean. This trait is regulated by complex biological processes and is affected by endogenous genetic factors as well as environmental cues. Many researches have demonstrated the effect of genotype  $\times$  environment interaction on DTF; however, studies on epistatic interactions among genetic factors have rarely been reported. Although our study is restricted by the origin of soybean accessions and also requires further biological validation, candidate markers identified in this study provides not only additional information for understanding the DTF trait in soybean but also a valuable genetic basis for soybean breeding programs aimed at improving DTF.

## Supporting information

**S1 Dataset. Sample information of origin, maturity group, and flowering time.**  
(XLSX)

**S2 Dataset. List of soybean genes and number of SNP markers (MAF > 0.05) per gene.**  
(XLSX)

**S3 Dataset. Results of GWAS on flowering time in 2,662 soybean accessions.**  
(XLSX)

**S4 Dataset. List of DTF-related genes reported in *Arabidopsis thaliana*, and their homologous genes to soybean.**  
(XLSX)

**S1 Fig. Ten-times cross-validation error values for population structure analysis from  $K = 2$  to  $K = 8$ .**  
(TIF)

**S1 Table. Average  $\pi$ , F, and LD values among soybean accessions.**  
(DOCX)

## Acknowledgments

We appreciate the Rural Development Administration and the reviewers who suggested valuable comments for this manuscript.

## Author Contributions

**Conceptualization:** Namshin Kim.

**Data curation:** Jae-Yoon Kim.

**Formal analysis:** Kyoung Hyoun Kim.

**Funding acquisition:** Namshin Kim.

**Methodology:** Jae-Yoon Kim.

**Project administration:** Namshin Kim.

**Resources:** Jung-Kyung Moon.

**Supervision:** Namshin Kim.

**Validation:** Won-Jun Lim, Seongmun Jeong, Ho-Yeon Lee, Youngbum Cho.

**Visualization:** Won-Jun Lim, Seongmun Jeong, Ho-Yeon Lee, Youngbum Cho.

**Writing – original draft:** Kyoung Hyoun Kim.

**Writing – review & editing:** Kyoung Hyoun Kim, Jae-Yoon Kim, Jung-Kyung Moon.

## References

1. Contreras-Soto RI, Mora F, de Oliveira MAR, Higashi W, Scapim CA, Schuster I. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS One*. 2017; 12(2):e0171105. <https://doi.org/10.1371/journal.pone.0171105> PMID: 28152092
2. Chen H, Xie W, He H, Yu H, Chen W, Li J, et al. A high-density SNP genotyping array for rice biology and molecular breeding. *Mol Plant*. 2014; 7(3):541–53. <https://doi.org/10.1093/mp/sst135> PMID: 24121292

3. Hu X, Ren J, Ren X, Huang S, Sabiel SA, Luo M, et al. Association of agronomic traits with SNP markers in durum wheat (*Triticum turgidum* L. durum (Desf.)). *PLoS One*. 2015; 10(6):e0130854. <https://doi.org/10.1371/journal.pone.0130854> PMID: 26110423
4. Navarro JAR, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet*. 2017; 49(3):476. <https://doi.org/10.1038/ng.3784> PMID: 28166212
5. Nascimento M, Nascimento ACC, Silva FFe, Barili LD, do Vale NM, Carneiro JE, et al. Quantile regression for genome-wide association study of flowering time-related traits in common bean. *PLoS One*. 2018; 13(1):e0190303. <https://doi.org/10.1371/journal.pone.0190303> PMID: 29300788
6. Sales E, Viruel J, Domingo C, Marqués L. Genome wide association analysis of cold tolerance at germination in temperate japonica rice (*Oryza sativa* L.) varieties. *PLoS One*. 2017; 12(8):e0183416. <https://doi.org/10.1371/journal.pone.0183416> PMID: 28817683
7. Fan Y, Zhou G, Shabala S, Chen Z-H, Cai S, Li C, et al. Genome-wide association study reveals a new QTL for salinity tolerance in barley (*Hordeum vulgare* L.). *Front Plant Sci*. 2016; 7:946. <https://doi.org/10.3389/fpls.2016.00946> PMID: 27446173
8. Patil G, Do T, Vuong TD, Valliyodan B, Lee J-D, Chaudhary J, et al. Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. *Sci Rep*. 2016; 6:19199. <https://doi.org/10.1038/srep19199> PMID: 26781337
9. Thabet SG, Moursi YS, Karam MA, Graner A, Alqudah AM. Genetic basis of drought tolerance during seed germination in barley. *PLoS One*. 2018; 13(11):e0206682. <https://doi.org/10.1371/journal.pone.0206682> PMID: 30388157
10. Kim S-M, Reinke RF. A novel resistance gene for bacterial blight in rice, Xa43(t) identified by GWAS, confirmed by QTL mapping using a bi-parental population. *PLoS One*. 2019; 14(2):e0211775. <https://doi.org/10.1371/journal.pone.0211775> PMID: 30753229
11. Chen G, Wang X, Hao J, Yan J, Ding J. Genome-wide association implicates candidate genes conferring resistance to maize rough dwarf disease in maize. *PLoS One*. 2015; 10(11):e0142001. <https://doi.org/10.1371/journal.pone.0142001> PMID: 26529245
12. Begum H, Spindel JE, Lalusin A, Borromeo T, Gregorio G, Hernandez J, et al. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS One*. 2015; 10(3):e0119873. <https://doi.org/10.1371/journal.pone.0119873> PMID: 25785447
13. Zhang J, Song Q, Cregan PB, Nelson RL, Wang X, Wu J, et al. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics*. 2015; 16(1):217. <https://doi.org/10.1186/s12864-015-1441-4> PMID: 25887991
14. Hoang GT, Gantet P, Nguyen KH, Phung NTP, Ha LT, Nguyen TT, et al. Genome-wide association mapping of leaf mass traits in a Vietnamese rice landrace panel. *PLoS One*. 2019; 14(7). <https://doi.org/10.1371/journal.pone.0219274> PMID: 31283792
15. Lu S, Zhang M, Zhang Z, Wang Z, Wu N, Song Y, et al. Screening and verification of genes associated with leaf angle and leaf orientation value in inbred maize lines. *PLoS One*. 2018; 13(12):e0208386. <https://doi.org/10.1371/journal.pone.0208386> PMID: 30532152
16. Zhao X, Dong H, Chang H, Zhao J, Teng W, Qiu L, et al. Genome wide association mapping and candidate gene analysis for hundred seed weight in soybean [*Glycine max* (L.) Merrill]. *BMC Genomics*. 2019; 20(1):648. <https://doi.org/10.1186/s12864-019-6009-2> PMID: 31412769
17. Sehgal D, Mondal S, Guzman C, Barrios GG, Franco C, Singh RP, et al. Validation of candidate gene-based markers and identification of novel loci for thousand-grain weight in spring bread wheat. *Front Plant Sci*. 2019; 10:1189. <https://doi.org/10.3389/fpls.2019.01189> PMID: 31616457
18. Battenfield SD, Sheridan JL, Silva LD, Miclaus KJ, Dreisigacker S, Wolfinger RD, et al. Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLoS One*. 2018; 13(11):e0204757. <https://doi.org/10.1371/journal.pone.0204757> PMID: 30496187
19. Wang B, Li J. Understanding the molecular bases of agronomic trait improvement in rice. *Plant Cell*. 2019; 31(7): 1416–7. <https://doi.org/10.1105/tpc.19.00343> PMID: 31068452
20. Langer SM, Longin CFH, Würschum T. Flowering time control in European winter wheat. *Front Plant Sci*. 2014; 5:537. <https://doi.org/10.3389/fpls.2014.00537> PMID: 25346745
21. Mazaheri M, Heckwolf M, Vaillancourt B, Gage JL, Burdo B, Heckwolf S, et al. Genome-wide association analysis of stalk biomass and anatomical traits in maize. *BMC Plant Biol*. 2019; 19(1):45. <https://doi.org/10.1186/s12870-019-1653-x> PMID: 30704393
22. Alqudah AM, Sharma R, Pasam RK, Graner A, Kilian B, Schnurbusch T. Genetic dissection of photoperiod response based on GWAS of pre-anthesis phase duration in spring barley. *PLoS One*. 2014; 9(11):e113120. <https://doi.org/10.1371/journal.pone.0113120> PMID: 25420105

23. Huang X, Zhao Y, Li C, Wang A, Zhao Q, Li W, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet.* 2012; 44(1):32–9. <https://doi.org/10.1038/ng.1018> PMID: 22138690
24. Ahsan A, Monir M, Meng X, Rahaman M, Chen H, Chen M. Identification of epistasis loci underlying rice flowering time by controlling population stratification and polygenic effect. *DNA Res.* 2018; 26(2):119–30. <https://doi.org/10.1093/dnares/dsy043> PMID: 30590457
25. Koo CL, Liew MJ, Mohamad MS, Salleh AHM, Deris S, Ibrahim Z, et al. Software for detecting gene-gene interactions in genome wide association studies. *Biotechnol Bioproc E.* 2015; 20(4):662–76. <https://doi.org/10.1007/s12257-015-0064-6>
26. Moellers TC, Singh A, Zhang J, Brungardt J, Kabbage M, Mueller DS, et al. Main and epistatic loci studies in soybean for *Sclerotinia sclerotiorum* resistance reveal multiple modes of resistance in multi-environments. *Sci Rep.* 2017; 7(1):3554. <https://doi.org/10.1038/s41598-017-03695-9> PMID: 28620159
27. Zhang J, Singh A, Mueller DS, Singh AK. Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. *Plant J.* 2015; 84(6):1124–36. <https://doi.org/10.1111/tj.13069> PMID: 26561232
28. Mamidi S, Lee RK, Goos JR, McClean PE. Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). *PLoS One.* 2014; 9(9): e107469. <https://doi.org/10.1371/journal.pone.0107469> PMID: 25225893
29. Assefa T, Otyama PI, Brown AV, Kalberer SR, Kulkarni RS, Cannon SB. Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC Genomics.* 2019; 20(1):527. <https://doi.org/10.1186/s12864-019-5907-7> PMID: 31242867
30. He T, Hill CB, Angessa TT, Zhang X-Q, Chen K, Moody D, et al. Gene-set association and epistatic analyses reveal complex gene interaction networks affecting flowering time in a worldwide barley collection. *J Exp Bot.* 2019; 70(20):5603–16. <https://doi.org/10.1093/jxb/erz332> PMID: 31504706
31. Bernard R. Two major genes for time of flowering and maturity in soybeans. *Crop Sci.* 1971; 11(2):242–4. <https://doi.org/10.2135/cropsci1971.0011183X001100020022x>
32. Buzzell R. Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Can J Genet Cytol.* 1971; 13(4):703–7. <https://doi.org/10.1139/g71-100>
33. Buzzell R, Voldeng H. Inheritance of insensitivity to long daylength. *Soyb Genet Newsl.* 1980; 7(1):13. <https://doi.org/10.1093/jhered/esp113>
34. Saindon G, Voldeng H, Beversdorf W, Buzzell R. Genetic control of long daylength response in soybean. *Crop Sci.* 1989; 29(6):1436–9. <https://doi.org/10.2135/cropsci1989.0011183X002900060021x>
35. McBlain B, Bernard R. A new gene affecting the time of flowering and maturity in soybeans. *J Hered.* 1987; 78(3):160–2. <https://doi.org/10.1093/oxfordjournals.jhered.a110349>
36. Bonato ER, Vello NA. E6, a dominant gene conditioning early flowering and maturity in soybeans. *Genet Mol Biol.* 1999; 22(2):229–32. <https://doi.org/10.1590/S1415-4757199900200016>
37. Cober ER, Voldeng HD. A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. *Crop Sci.* 2001; 41(3):698–701. <https://doi.org/10.2135/cropsci2001.413698x>
38. Cober ER, Molnar SJ, Charette M, Voldeng HD. A new locus for early maturity in soybean. *Crop Sci.* 2010; 50(2):524–7. <https://doi.org/10.2135/cropsci2009.04.0174>
39. Kong F, Nan H, Cao D, Li Y, Wu F, Wang J, et al. A new dominant gene E9 conditions early flowering and maturity in soybean. *Crop Sci.* 2014; 54(6):2529–35. <https://doi.org/10.2135/cropsci2014.03.0228>
40. Zhao C, Takeshima R, Zhu J, Xu M, Sato M, Watanabe S, et al. 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(1):20. <https://doi.org/10.1186/s12870-016-0704-9> PMID: 26786479
41. Samanfar B, Molnar SJ, Charette M, Schoenrock A, Dehne F, Golshani A, et al. Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theor Appl Genet.* 2017; 130(2):377–90. <https://doi.org/10.1007/s00122-016-2819-7> PMID: 27832313
42. Ray JD, Hinson K, Mankono J, Malo MF. Genetic control of a long-juvenile trait in soybean. *Crop Sci.* 1995; 35(4):1001–6. <https://doi.org/10.2135/cropsci1995.0011183X003500040012x>
43. Liu B, Watanabe S, Uchiyama T, Kong F, Kanazawa A, Xia Z, et al. The soybean stem growth habit gene Dt1 is an ortholog of Arabidopsis TERMINAL FLOWER1. *Plant Physiol.* 2010; 153(1):198–210. <https://doi.org/10.1104/pp.109.150607> PMID: 20219831
44. Tian Z, Wang X, Lee R, Li Y, Specht JE, Nelson RL, et al. Artificial selection for determinate growth habit in soybean. *Proc Natl Acad Sci U S A.* 2010; 107(19):8563–8. <https://doi.org/10.1073/pnas.1000088107> PMID: 20421496

45. Lee YG, Jeong N, Kim JH, Lee K, Kim KH, Pirani A, et al. Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant J*. 2015; 81(4):625–36. <https://doi.org/10.1111/tpl.12755> PMID: 25641104
46. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet*. 2018; 103(3):338–48. <https://doi.org/10.1016/j.ajhg.2018.07.015> PMID: 30100085
47. Jeong S-C, Moon J-K, Park S-K, Kim M-S, Lee K, Lee SR, et al. Genetic diversity patterns and domestication origin of soybean. *Theor Appl Genet*. 2019; 132(4):1179–93. <https://doi.org/10.1007/s00122-018-3271-7> PMID: 30588539
48. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. <https://doi.org/10.1086/519795> PMID: 17701901
49. Gascuel O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*. 1997; 14(7):685–95. <https://doi.org/10.1093/oxfordjournals.molbev.a025808> PMID: 9254330
50. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*. 2014; 197(2):573–89. <https://doi.org/10.1534/genetics.114.164350> PMID: 24700103
51. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980; 16(2):111–20. <https://doi.org/10.1007/bf01731581> PMID: 7463489
52. Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. 2019; 35(10):1786–8. <https://doi.org/10.1093/bioinformatics/bty875> PMID: 30321304
53. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27(15):2156–8. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
54. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012; 28(18):2397–9. <https://doi.org/10.1093/bioinformatics/bts444> PMID: 22796960
55. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012; 6(2):80–92. <https://doi.org/10.4161/fly.19695> PMID: 22728672
56. Jung C-H, Wong CE, Singh MB, Bhalla PL. Comparative genomic analysis of soybean flowering genes. *PLoS One*. 2012; 7(6):e38250. <https://doi.org/10.1371/journal.pone.0038250> PMID: 22679494
57. Tsubokura Y, Watanabe S, Xia Z, Kanamori H, Yamagata H, Kaga A, et al. Natural variation in the genes responsible for maturity loci E1, E2, E3 and E4 in soybean. *Ann Bot*. 2014; 113(3):429–41. <https://doi.org/10.1093/aob/mct269> PMID: 24284817
58. Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio—the analysis and navigation of molecular networks. *Bioinformatics*. 2003; 19(16):2155–7. <https://doi.org/10.1093/bioinformatics/btg290> PMID: 14594725
59. Watanabe S, Harada K, Abe J. Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breed Sci*. 2012; 61(5):531–43. <https://doi.org/10.1270/jsbbs.61.531> PMID: 23136492
60. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10(1):421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
61. Wang J, Joshi T, Valliyodan B, Shi H, Liang Y, Nguyen HT, et al. A Bayesian model for detection of high-order interactions among genetic variants in genome-wide association studies. *BMC Genomics*. 2015; 16(1):1011. <https://doi.org/10.1186/s12864-015-2217-6> PMID: 26607428
62. Zhang J, Hou T, Wang W, Liu JS. Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. *Proc Natl Acad Sci U S A*. 2010; 107(4):1321–6. <https://doi.org/10.1073/pnas.0907304107> PMID: 20080674
63. Stewart DW, Cober ER, Bernard RL. Modeling genetic effects on the photothermal response of soybean phenological development. *Agron J*. 2003; 95(1):65–70.
64. Zhang X, Zhai H, Wang Y, Tian X, Zhang Y, Wu H, et al. Functional conservation and diversification of the soybean maturity gene E1 and its homologs in legumes. *Sci Rep*. 2016; 6(1):29548. <https://doi.org/10.1038/srep29548> PMID: 27405888



65. Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y, et al. Map-based cloning of the gene associated with the soybean maturity locus E3. *Genetics*. 2009; 182(4):1251–62. <https://doi.org/10.1534/genetics.108.098772> PMID: 19474204
66. Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, et al. 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. <https://doi.org/10.1534/genetics.110.125062> PMID: 21406680
67. 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(2):995–1007. <https://doi.org/10.1534/genetics.108.092742> PMID: 18780733
68. Lee S-J, Lee BH, Jung J-H, Park SK, Song JT, Kim JH. GROWTH-REGULATING FACTOR and GRF-INTERACTING FACTOR specify meristematic cells of gynoecia and anthers. *Plant Physiol*. 2018; 176(1):717–29. <https://doi.org/10.1104/pp.17.00960> PMID: 29114079
69. Preston JC, Hileman L. Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Front Plant Sci*. 2013; 4:80. <https://doi.org/10.3389/fpls.2013.00080> PMID: 23577017
70. Xu D, Zhu D, Deng XW. The role of COP1 in repression of photoperiodic flowering. *F1000Res*. 2016; 5. <https://doi.org/10.12688/f1000research.7346.1> PMID: 26949521
71. Zeng X, Liu H, Du H, Wang S, Yang W, Chi Y, et al. Soybean MADS-box gene GmAGL1 promotes flowering via the photoperiod pathway. *BMC Genomics*. 2018; 19(1):51. <https://doi.org/10.1186/s12864-017-4402-2> PMID: 29338682
72. Jaudal M, Zhang L, Che C, Putterill J. Three *Medicago* MtFUL genes have distinct and overlapping expression patterns during vegetative and reproductive development and 35S:MtFULb accelerates flowering and causes a terminal flower phenotype in *Arabidopsis*. *Front Genet*. 2015; 6:50. <https://doi.org/10.3389/fgene.2015.00050> PMID: 25745430
73. Ren C, Zhang Z, Wang Y, Li S, Liang Z. Genome-wide identification and characterization of the NF-Y gene family in grape (*Vitis vinifera* L.). *BMC Genomics*. 2016; 17(1):605. <https://doi.org/10.1186/s12864-016-2989-3> PMID: 27516172
74. Sánchez R, Kim MY, Calonje M, Moon Y-H, Sung ZR. Temporal and spatial requirement of EMF1 activity for *Arabidopsis* vegetative and reproductive development. *Mol Plant*. 2009; 2(4):643–53. <https://doi.org/10.1093/mp/ssp004> PMID: 19825645
75. Kim SY, Lee J, Eshed-Williams L, Zilberman D, Sung ZR. EMF1 and PRC2 cooperate to repress key regulators of *Arabidopsis* development. *PLoS Genet*. 2012; 8(3):e1002512. <https://doi.org/10.1371/journal.pgen.1002512> PMID: 22457632
76. Song YH, Estrada DA, Johnson RS, Kim SK, Lee SY, MacCoss MJ, et al. Distinct roles of FKF1, GIGANTEA, and ZEITLUPE proteins in the regulation of CONSTANS stability in *Arabidopsis* photoperiodic flowering. *Proc Natl Acad Sci U S A*. 2014; 111(49):17672–7. <https://doi.org/10.1073/pnas.1415375111> PMID: 25422419
77. Seaton DD, Smith RW, Song YH, MacGregor DR, Stewart K, Steel G, et al. Linked circadian outputs control elongation growth and flowering in response to photoperiod and temperature. *Mol Syst Biol*. 2015; 11(1):776. <https://doi.org/10.15252/msb.20145766> PMID: 25600997
78. Agliassa C, Narayana R, Berteaux CM, Rodgers CT, Maffei ME. Reduction of the geomagnetic field delays *Arabidopsis thaliana* flowering time through downregulation of flowering-related genes. *Bioelectromagnetics*. 2018; 39(5):361–74. <https://doi.org/10.1002/bem.22123> PMID: 29709075
79. Cai X, Ballif J, Endo S, Davis E, Liang M, Chen D, et al. A putative CCAAT-binding transcription factor is a regulator of flowering timing in *Arabidopsis*. *Plant Physiol*. 2007; 145(1):98–105. <https://doi.org/10.1104/pp.107.102079> PMID: 17631525
80. Capovilla G, Schmid M, Posé D. Control of flowering by ambient temperature. *J Exp Bot*. 2015; 66(1):59–69. <https://doi.org/10.1093/jxb/eru416> PMID: 25326628
81. Kang H, Zhang C, An Z, Shen W-H, Zhu Y. AtINO80 and AtARP5 physically interact and play common as well as distinct roles in regulating plant growth and development. *New Phytol*. 2019; 223(1):336–53. <https://doi.org/10.1111/nph.15780> PMID: 30843208
82. Jiao Y, Yang H, Ma L, Sun N, Yu H, Liu T, et al. A genome-wide analysis of blue-light regulation of *Arabidopsis* transcription factor gene expression during seedling development. *Plant Physiol*. 2003; 133(4):1480–93. <https://doi.org/10.1104/pp.103.029439> PMID: 14605227
83. Helliwell CA, Robertson M, Finnegan EJ, Buzas DM, Dennis ES. Vernalization-repression of *Arabidopsis* FLC requires promoter sequences but not antisense transcripts. *PLoS One*. 2011; 6(6):e21513. <https://doi.org/10.1371/journal.pone.0021513> PMID: 21713009

84. Hohenstatt ML, Mikulski P, Komarynets O, Klose C, Kycia I, Jeltsch A, et al. PWWP-DOMAIN INTER-ACTOR OF POLYCOMBS1 interacts with Polycomb-group proteins and histones and regulates Arabidopsis flowering and development. *Plant Cell*. 2018; 30(1):117–33. <https://doi.org/10.1105/tpc.17.00117> PMID: 29330200
85. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis*. 2015; 53(8):474–85. <https://doi.org/10.1002/dvg.22877> PMID: 26201819
86. Cheng J-Z, Zhou Y-P, Lv T-X, Xie C-P, Tian C-E. Research progress on the autonomous flowering time pathway in Arabidopsis. *Physiol Mol Biol Plants*. 2017; 23(3):477–85. <https://doi.org/10.1007/s12298-017-0458-3> PMID: 28878488
87. Liu Q, Wang Q, Deng W, Wang X, Piao M, Cai D, et al. Molecular basis for blue light-dependent phosphorylation of Arabidopsis cryptochrome 2. *Nat Commun*. 2017; 8(1):15234. <https://doi.org/10.1038/ncomms15234> PMID: 28492234
88. Yu X, Liu H, Klejnot J, Lin C. The cryptochrome blue light receptors. *Arabidopsis Book*. 2010; 2010(8):e0135. <https://doi.org/10.1199/tab.0135> PMID: 21841916
89. Kim MY, Kang YJ, Lee T, Lee S-H. Divergence of flowering-related genes in three legume species. *Plant Genome*. 2013; 6(3). <https://doi.org/10.3835/plantgenome2013.03.0008>
90. Amasino RM, Michaels SD. The timing of flowering. *Plant Physiol*. 2010; 154(2):516–20. <https://doi.org/10.1104/pp.110.161653> PMID: 20921176
91. Boss PK, Bastow RM, Mylne JS, Dean C. Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell*. 2004; 16(suppl 1):S18–S31. <https://doi.org/10.1105/tpc.015958> PMID: 15037730
92. Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, et al. Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. *Proc Natl Acad Sci U S A*. 2012; 109(32):E2155–E2164. <https://doi.org/10.1073/pnas.1117982109> PMID: 22619331
93. Liu W, Jiang B, Ma L, Zhang S, Zhai H, Xu X, et al. Functional diversification of Flowering Locus T homologs in soybean: GmFT1a and GmFT2a/5a have opposite roles in controlling flowering and maturation. *New Phytol*. 2018; 217(3):1335–45. <https://doi.org/10.1111/nph.14884> PMID: 29120038
94. Jiang B, Nan H, Gao Y, Tang L, Yue Y, Lu S, et al. Allelic combinations of soybean maturity loci E1, E2, E3 and E4 result in diversity of maturity and adaptation to different latitudes. *PLoS One*. 2014; 9(8):e106042. <https://doi.org/10.1371/journal.pone.0106042> PMID: 25162675
95. Zhai H, Lü S, Liang S, Wu H, Zhang X, Liu B, et al. 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. <https://doi.org/10.1371/journal.pone.0089030> PMID: 24586488
96. Xu M, Xu Z, Liu B, Kong F, Tsubokura Y, Watanabe S, et al. Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. *BMC Plant Biol*. 2013; 13(1):91. <https://doi.org/10.1186/1471-2229-13-91> PMID: 23799885
97. Tsubokura Y, Matsumura H, Xu M, Liu B, Nakashima H, Anai T, et al. Genetic variation in soybean at the maturity locus E4 is involved in adaptation to long days at high latitudes. *Agron J*. 2013; 3(1):117–34. <https://doi.org/10.3390/agronomy3010117>
98. Zhai H, Lü S, Wang Y, Chen X, Ren H, Yang J, et al. Allelic variations at four major maturity E genes and transcriptional abundance of the E1 gene are associated with flowering time and maturity of soybean cultivars. *PLoS One*. 2014; 9(5):e97636. <https://doi.org/10.1371/journal.pone.0097636> PMID: 24830458
99. Li Z, Nelson RL. Genetic diversity among soybean accessions from three countries measured by RAPDs. *Crop Sci*. 2001; 41(4):1337–47. <https://doi.org/10.2135/cropsci2001.4141337x>
100. Lee G-A, Choi Y-M, Yi J-Y, Chung J-W, Lee M-C, Ma K-H, et al. Genetic diversity and population structure of Korean soybean collection using 75 microsatellite markers. *Korean J Crop Sci*. 2014; 59(4):492–7. <https://doi.org/10.7740/kjcs.2014.59.4.492>
101. Sedivy EJ, Wu F, Hanzawa Y. Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytol*. 2017; 214(2):539–53. <https://doi.org/10.1111/nph.14418> PMID: 28134435
102. Ritchie MD, Steen KV. The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. *Ann Transl Med*. 2018; 6(8):157. <https://doi.org/10.21037/atm.2018.04.05> PMID: 29862246
103. Sun X, Lu Q, Mukherjee S, Crane PK, Elston R, Ritchie MD. Analysis pipeline for the epistasis search—statistical versus biological filtering. *Front Genet*. 2014; 5:106. <https://doi.org/10.3389/fgene.2014.00106> PMID: 24817878

104. Niel C, Sinoquet C, Dina C, Rocheleau G. A survey about methods dedicated to epistasis detection. *Front Genet.* 2015; 6:285. <https://doi.org/10.3389/fgene.2015.00285> PMID: 26442103
105. Yi N. Statistical analysis of genetic interactions. *Genet Res.* 2010; 92(5–6):443–59. <https://doi.org/10.1017/S0016672310000595> PMID: 21429274
106. Bernier G, Périlleux C. A physiological overview of the genetics of flowering time control. *Plant Biotechnol J.* 2005; 3(1):3–16. <https://doi.org/10.1111/j.1467-7652.2004.00114.x> PMID: 17168895
107. Rouse DT, Sheldon CC, Bagnall DJ, Peacock WJ, Dennis ES. FLC, a repressor of flowering, is regulated by genes in different inductive pathways. *Plant J.* 2002; 29(2):183–91. <https://doi.org/10.1046/j.0960-7412.2001.01210.x> PMID: 11851919
108. Henderson IR, Dean C. Control of Arabidopsis flowering: the chill before the bloom. *Development.* 2004; 131(16):3829–38. <https://doi.org/10.1242/dev.01294> PMID: 15289433
109. Corbesier L, Coupland G. The quest for florigen: a review of recent progress. *J Exp Bot.* 2006; 57(13):3395–403. PMID: 17030536