Received: 19 February 2021 | Accepted: 6 July 2021

DOI: 10.1002/tpg2.20146

The Plant Genome

ORIGINAL RESEARCH

Genetic control and allele variation among soybean maturity groups 000 through IX

Gustavo Zimmer[1,2] | **Mark J. Miller**[1] | **Clinton J. Steketee**[1] | **Scott A. Jackson**[1] | **Lilian Vanussa Madruga de Tunes**[2] | **Zenglu Li**[1]

[1] Institute of Plant Breeding, Genetics, and Genomics, and Department of Crop and Soil Sciences, University of Georgia, Athens, GA, 30602, USA

[2] Department of Crop Production, Federal University of Pelotas, Capão do Leão, RS, 96160-000, Brazil

Correspondence
Zenglu Li, Institute of Plant Breeding, Genetics, and Genomics, and Department of Crop and Soil Sciences, University of Georgia, Athens, GA, USA 30602.
Email: zli@uga.edu

Assigned to Associate Editor David Hyten.

Abstract

Soybean [*Glycinemax* (L.) Merr.] maturity determines the growing region of a given soybean variety and is a primary factor in yield and other agronomic traits. The objectives of this research were to identify the quantitative trait loci (QTL) associated with maturity groups (MGs) and determine the genetic control of soybean maturity in each MG. Using data from 16,879 soybean accessions, genome-wide association (GWA) analyses were conducted for each paired MG and across MGs 000 through IX. Genome-wide association analyses were also performed using 184 genotypes (MGs V–IX) with days to flowering (DTF) and maturity (DTM) collected in the field. A total of 58 QTL were identified to be significantly associated with MGs in individual GWAs, which included 12 reported maturity loci and two stem termination genes. Genome-wide associations across MGs 000–IX detected a total of 103 QTL and confirmed 54 QTL identified in the individual GWAs. Of significant loci identified, *qMG*-5.2 had effects on the highest number (9) of MGs, followed by *E2*, *E3*, *Dt2*, *qMG*-15.5, *E1*, *qMG-13.1*, *qMG-7.1*, and *qMG-16.1*, which affected five to seven MGs. A high number of genetic loci (8–25) that affected MGs 0–V were observed. Stem termination genes *Dt1* and *Dt2* mainly had significant allele variation in MGs II–V. Genome-wide associations for DTF, DTM, and reproductive period (RP) in the diversity panel confirmed 15 QTL, of which seven were observed in MGs V–IX. The results generated can help soybean breeders manipulate the maturity loci for genetic improvement of soybean yield.

1 | INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is a photosensitive short-day plant, and soybean varieties differ in geographical adaptability according to photoperiod, temperature, rate of photoperiodic changes, daily irradiance, management practices, and genetics (Cober et al., 2014; Mourtzinis & Conley, 2017). Hypothetical maturity group (MG) zones were first

Abbreviations: Chr, chromosome; DTF, days to flowering; DTM, days to maturity; FarmCPU, fixed and random model circulating probability unification; GWA, genome-wide association; MG, maturity group; NIL, near-isogenic line; QTL, quantitative trait loci; RP, reproductive period; SNP, single nucleotide polymorphism.

defined across the U.S. territory using empirical data (Scott & Aldrich, 1970) and represent a defined area where a cultivar is best adapted and recommended for agricultural use (Figure 1). Maturity groups range from 000 to X (Boerma & Specht, 2004) and are essential in determining planting dates and locations, as late planting can lead to truncated vegetative growth periods, reduced nitrogen uptake, and ultimately lower yields (Ortel et al., 2020).

The genetic loci associated with soybean maturity have been defined as *E* loci with 12 loci reported so far. The first two loci, *E1* and *E2*, were discovered by Bernard (1971) using genotypes from MGs II–V. The *E1* allele delayed flowering by downregulating *GmFT2a* and *GmFT5a* (Xu et al., 2015). Besides the *E1* locus on chromosome (Chr) 6, the soybean genome has two *E1* homologues, *E1La* and *E1Lb*, located on Chr 4, with minor effects on flowering time (Xu et al., 2015). A recent study demonstrated that *E1Lb* delays flowering under long-day conditions independently of *E1* (Zhu, et al., 2019). *E2*, located on Chr 10, is an ortholog of the *Arabidopsis thaliana* (L.) Heynh. *GIGANTEA* (*GI*) gene, which delays flowering by inhibiting expression of *GmFT2a* under long days (Watanabe et al., 2011).

E3 and *E4* encode *PYTHOCHROME A* genes *GmPHYA3* (Watanabe et al., 2009) and *GmPHYA2* (Liu et al., 2008), which suppress the expression of *GmFT2a* and *GmFT5a* (Kong et al., 2010). The *E3* locus, identified by Buzzell (1971) using genotypes from MGs 0–III, is located on Chr 19 (Watanabe et al., 2009), while *E4* was reported by Buzzell and Voldeng (1980) using genotypes from MGs 00–II, which is located on Chr 20 (Liu et al., 2008). The recessive allele of *E4* was identified to confer adaptation to high-latitude environments in Japan (Kanazawa et al., 2009). *E3* and *E4* are also responsible for postflowering photoperiod responses affecting the expression of *Dt1*, the determinate growth habit gene (Xu et al., 2013).

The *E5* locus was detected in the cross of Harosoy (*e5*) × PI 80837 (*E5*) (McBlain & Bernard, 1987), which are genotypes from MGs II and IV, respectively. A subsequent study by Dissanayaka et al. (2016) failed to identify a QTL for *E5* in the original cross, suggesting that a unique *E5* gene may not exist. However, the molecular analysis of the near-isogenic lines (NILs) at the transcriptome level of *E5* indicated that it was a unique locus (Wu et al., 2019).

The delayed flowering under short-day conditions has been defined as a long-juvenile trait (Hartwig & Kiihl, 1979). *E6* is involved in the control of late flowering under short-day conditions and the dominant allele conditions early flowering (Bonato & Vello, 1999). The *E6* locus was identified in natural mutations in the cultivar Parana (MG VI). Li et al. (2017a) mapped the *E6* locus to Chr 4 near the single nucleotide polymorphism (SNP) marker *HRM101* and demonstrated that *E6* had a suppressive effect on *E1*. Additionally, *E6*, the *J* gene located on Chr 4, is also involved in this characteristic where the dominant allele conditioned early flowering and the recessive allele was responsible for the long-juvenile trait (Ray et al., 1995). Genetic mapping demonstrated that *E6* and *J* are tightly linked (Li et al., 2017a), and subsequent research discovered that these are different alleles of *ELF3* (Miranda et al., 2020; Fang et al., 2021; Nissan et al., 2021).

The *E7* locus was detected because of an association between early maturity and tawny pubescence observed in short-season soybean from MGs 000–II (Cober & Voldeng, 2001). Molnar et al. (2003) mapped *E7* to Chr 6 between *Satt100* and *Satt460*. While markers identified by Molnar et al. (2003) for *E1* and *T* loci are located near their known position in the Glyma.Wm82.a2 genome assembly, the position identified for *E7* relative to *E1* and *T* was not consistent with the result from Cober and Voldeng (2001).

Core Ideas

- Soybean maturity is a primary factor that affects yield and agronomic traits
- Understanding maturity genetic control will aid breeders to use germplasm across maturity groups
- The genetic profile of maturity groups were determined using genome-wide association analyses
- Effects of these QTL were calculated
- Our results can help breeders manipulate the maturity loci for genetic improvement of yield

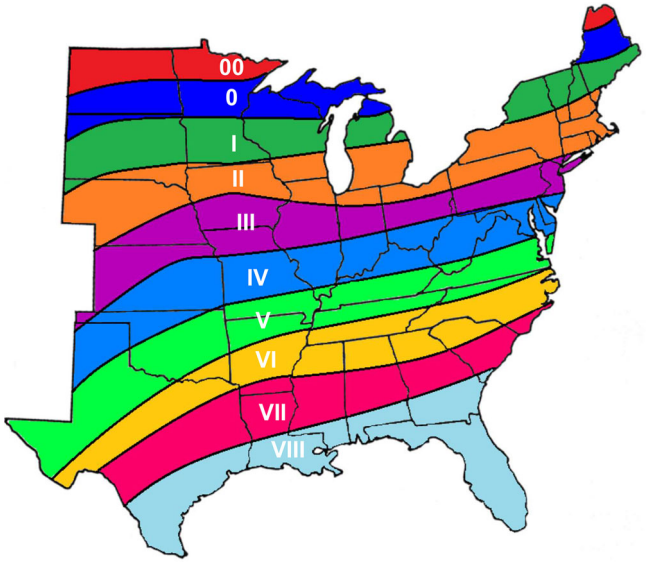

FIGURE 1 Soybean maturity group map depicting approximate growing regions in the United States based on Stowe and Dunphy (2017).

The *E8* locus was mapped to Chr 4 using isolines developed from the crosses of genotypes in MGs 000–II and the dominant *E8* allele resulted in late maturity (Cober et al., 2010). *GmCRY1a* played a predominant role in photoperiodic flowering (Zhang et al., 2008) and is most likely associated with *E8* (Cheng et al., 2011). However, while some markers reported in these studies are located closer to *GmCRY1a*, others are either unknown or located closer to *E1La* and *E1Lb* in the Glyma.Wm82.a2 genome assembly. *E8* was associated with the photoperiod sensitivity, prolonging the reproductive period in normal environments and long day conditions (Cheng et al., 2011).

The *E9* locus was identified on Chr 16 using a cross between early cultivar Tokey 780 and a late wild soybean (*G. soja* Siebold & Zucc.) accession Hidaka 4 (Kong et al., 2014). The dominant allele promotes early flowering under a long-day condition (Kong et al., 2014). The *e9* allele maintains vegetative growth in early flowering genetic backgrounds and can be used to select the long-juvenile trait (Zhao et al., 2016). *E9*, which encodes *GmFT2a* and *GmFT5a*, are *Flowering Locus T* orthologs that have been shown to control photoperiod-regulated flowering redundantly and differentially in soybean, inducing the expression of floral identity genes (Nan et al., 2014). The *E10* locus was mapped to Chr 8 using materials from MGs 000–0 and encodes the florigen gene *GmFT4*, also a *Flowering Locus T* ortholog (Samanfar et al., 2016). *GmFT4* is expressed in parallel with *E1* and upregulated under long days, delaying flowering, and downregulated under short days, resulting in earlier flowering (Zhai et al., 2014).

The latest maturity locus reported was *E11*, which was mapped to Chr 7 in a cross of 'Minsoy' (MG 0) by 'Archer' (MG I). The dominant allele of *E11* promotes earlier flowering and maturity under long day conditions (Wang et al., 2019). Recently, several independent studies also reported the effect of *PSEUDO RESPONSE REGULATOR* (*PRR*) genes, *PRR3* and *PRR7*, which controlled flowering time and maturity in soybean (Li et al., 2019a, 2019b, 2020; Lu et al., 2020; Ogiso-Tanaka et al., 2019). These genes were identified through targeted resequencing of germplasm collections followed by multiple-regression analysis (Ogiso-Tanaka et al., 2019) and genome-wide association (GWA) studies followed by fine mapping with recombinant inbred lines (Li et al., 2019a, 2019b, 2020; Lu et al., 2020). *PRR3* and *PRR7* were associated with ancient flowering time adaptation and domestication resulting from reduced presence of fully functional alleles (Li et al., 2019a, 2020; Lu et al., 2020). The identification of these genes disclosed the involvement of the circadian clock in flowering time regulation in soybean (Li et al., 2019b). Lu et al. (2020) proposed a model for *PRR3* and *PRR7* actions after observing that *E3* and *E4* upregulated *PRR3* and *PRR7*. Quantitative trait loci for *PRR3* and *PRR7* were only detected in recombinant inbred line subpopulations carrying the *E1* allele (Lu et al., 2020), indicating these genes might be of special importance for late MGs where previous evidence suggested that late MG soybean accessions have the *E1* allele (Jiang et al., 2014; Li et al., 2017b).

The determinate growth habit in soybean is controlled by a recessive allele at the *Dt1* locus (Liu et al., 2010). Several studies have shown that *Dt1* can also affect flowering and maturity time (Bernard, 1972; Zhang et al., 2015a; Kong et al., 2018; Ogiso-Tanaka et al., 2019; Yue et al., 2021). The semi-indeterminate phenotype in soybean is conditioned by a dominant epistatic effect of *Dt2* in *Dt1* backgrounds, which represses *Dt1* in the shoot apical meristem promoting the early conversion of shoot apical meristem into reproductive inflorescences (Bernard, 1972; Ping et al., 2014). However, the effect of *Dt2* on flowering time and maturity has not been reported in the literature.

The allelic diversity for major maturity loci has been characterized in different soybean collections (Miladinović et al., 2018; Tsubokura et al., 2014; Ogiso-Tanaka et al., 2019) and mostly was focused on reported maturity genes *E1*, *E2*, *E3*, and *E4*. Langewisch et al. (2017) identified the eight most common genotype combinations for *E1*, *E2*, and *E3* alleles within MGs in the USDA Soybean Germplasm Collection. Additionally, they identified *E1* in ~98% of MG V+ lines and *E3* in a majority of those lines (Langewisch et al., 2017). Li et al. (2017b) reported 12 different allele combinations of *E1*, *E2*, *E3*, and *E4* in a Chinese mini-core collection, in which 99% of lines in MG V+ had a combination of functional *E1* and *E3*. A similar finding was also previously reported by Jiang et al. (2014). These findings suggested that combining the roles of *E* loci could be used to design MGs when selecting soybean parents to breed varieties adapted to distinct cropping systems. Furthermore, these studies demonstrated that *E1* and *E3* alleles are prevalent in MGs V+, while most soybean accessions throughout MGs 000–VI (Li et al., 2017b), and possibly up to MG X (Langewisch et al., 2017), carry an *e2* allele. Soybean accessions containing an *E1/E2/E3/E4* genotype have also been detected in MGs I and II (Liu et al., 2020).

Despite the recent progress in identifying new soybean maturity genes, the relationship between these loci and MG classification is still poorly understood. Soybean breeders tend to use genotypes with similar maturities for crosses, while they try to introgress the desirable traits from different maturity groups (Kumar et al., 2015; Prenger et al., 2019). The *e2* allele, for example, was a confounding factor during the combined introgression of the high oleic–low linolenic acid trait and elevated vitamin E in soybean seed oil (Hagely et al., 2021). Understanding the allelic variation of maturity genes and the genetic control of each MG will help soybean breeders leverage soybean germplasm across MGs in their breeding programs. During the expansion of soybean cultivation to the tropics, the introgression of loss-of-function *j* alleles to temperate cultivars was successfully used in the

development and introduction of elite cultivars in Brazil and Southern China (Destro et al., 2001; Lu et al., 2017; Yue et al., 2017). The objectives of this study were to identify the QTL associated with MGs using GWA analyses and determine the genetic control of maturity in each MG.

2 | MATERIALS AND METHODS

2.1 | Materials

The Germplasm Resources Information Network (GRIN, www.ars-grin.gov) contains MG observations for 18,432 *Glycine* accessions. SoySNP50K data for these accessions were retrieved at SoyBase (www.soybase.org), which contains data from 20,087 *Glycine* genotypes (Song et al., 2015). After filtering, 16,879 soybean accessions having both MG and SNP genotype data were used for the analyses in this study (Supplemental Table S1).

To validate the results obtained from the Soybean Germplasm Collection with the data collected from field evaluations such as days to maturity (DTM), days to flowering (DTF), and reproductive period (RP), an assembled diversity panel consisting of 184 diverse soybean genotypes in MGs V–IX from 27 different countries (Steketee et al., 2020) was used (Supplemental Table S2).

2.2 | Genotype data and quality control

All accessions obtained from the USDA Soybean Germplasm Collection and most of the accessions from the diversity panel have high-density SNP data available from the SoySNP50K project (Song et al., 2015). Eight genotypes on the diversity panel were genotyped with SoySNP50K iSelect BeadChips as described by Steketee et al. (2020). Prior to the analyses, datasets were filtered to remove SNPs with minor allele frequencies <5% or with >10% missing data. Therefore, the number of SNPs used varied slightly across analyses, ranging from 31,627 to 35,133. The number of SNPs used in the GWA analysis of the diversity panel was 34,764. The physical positions of Glyma.Wm82.a2.v1 reference genome were used to determine the locations of the SNPs used in the analyses.

2.3 | Experimental design and phenotypic data collection

The diversity panel was evaluated for DTF and DTM in Athens, GA, in both 2015 (GA-15) and 2016 (GA-16). Genotypes were planted in two-row plots in a randomized complete block design with three replications per environment. Experiments were sown on 16 June 2015 (GA-15) and 8 June 2016 (GA-16). The plots were 2.43 m in length and planted with 0.76-m row spacing at a seeding density of 32 seeds m^{-2}. Days to flowering was recorded as the number of days from planting until 50% of the plants in a plot reached the R1 stage. Days to maturity was recorded as the number of days from planting to maturity, when 95% of the pods of plants in a plot exhibited a mature pod color at the R8 stage. Reproductive period was calculated as the difference between DTF and DTM.

2.4 | Statistical analyses

Genotypic data of 16,879 accessions from MGs 000–IX were coded numerically (0, 1, 2). Principal component analysis was performed using GAPIT (Lipka et al., 2012). The first two principal components were plotted for visualization with TIBCO Spotfire (Spotfire).

Maturity groups retrieved from the USDA Soybean Germplasm Collection were converted into integers, ranging from 0 to 11 for MGs 000–IX, respectively, for the subsequent GWA analyses. Phenotypic observations of DTF, DTM, and RP from the diversity panel were used for analysis of variance (ANOVA) using PROC GLM in SAS v9.4 (SAS Institute Inc.). The least-squares means of DTF, DTM, and RP were calculated using PROC MIXED, with genotype treated as a fixed effect and replication and year as random effects. The model for the combined datasets from two environments GA-15 and GA-16 accounted for the interaction between genotypes and years. To determine the correlation among DTF, DTM, and RP, a Pearson correlation analysis was performed using PROC CORR in SAS v9.4.

Genome-wide association analyses were performed using a 'fixed and random model circulating probability unification' (FarmCPU) method (Liu et al., 2016). Individual GWA analysis was performed for the paired MGs from USDA Soybean Germplasm Collection. To control false positives, a GWA analysis was also performed across MGs 000–IX. The population structure was accounted for using the first four principal components, while a Bonferroni threshold of 0.01 per number of markers was used for the significance level. The threshold ranged from $-\log_{10}(P) > 6.50$ to $-\log_{10}(P) > 6.55$ according to the number of markers used in each analysis. To further reduce false positives, significant markers with minor allele frequency ≤0.09 were not included in this paper except for those observed for MGs VIII–IX, where MG IX only accounted for 10% of the observations used in the analysis. Manhattan and quantile–quantile plots were visualized with the 'qqman' R package (Turner, 2018) using *p* values generated from the FarmCPU output.

For the diversity panel, GWAs were performed similarly to the method previously described with minor modifications. The least-squares means of DTF, DTM, and RP were used

as the phenotype values for subsequent GWAs. Two principal components were used to account for population structure. Considering the reduced number of phenotypic observations used in the GWA, a less stringent Bonferroni threshold (0.05 per number of markers) was used for significance level [$-\log_{10}(P) > 5.84$].

2.5 | Identification and naming of QTL associated with MGs

Significant markers observed in an individual analysis were attributed to the same QTL if the distance between them was <3 Mbp. Similarly, these QTL were considered as 'confirmed' if markers in the GWA across MGs 000–IX were identified <3 Mbp from their positions. The distance of 3 Mbp was chosen based on the linkage disequilibrium tests. At the QTL where more than one known maturity genes are located, such as *Dt1* and *E3* on Chr 19, the QTL were then split and SNPs attributed to the closet maturity locus. Quantitative trait loci identified in individual GWA were defined using a prefix '*qMG-*' followed by the number of the chromosome and an identifier associated with the QTL order on the chromosome. The prefix for the QTL identified in GWA across MG 000–IX is '*ECqMG-*', wherein EC stands for 'entire collection', while for QTL only identified from the diversity panel, the prefix is '*qDTF-*', '*qDTM-*' or '*qRP-*' according to the trait where each region was identified.

2.6 | Orthologs of Arabidopsis flowering genes

Except for *E1* and *E1L* genes, all other reported maturity loci, the recently identified *PRR* genes, and stem termination genes are orthologs of genes identified in Arabidopsis. An extensive list of Arabidopsis genes is already identified and characterized. Therefore, previously assembled lists of soybean orthologs of Arabidopsis flowering genes were verified as possible candidates for QTL identified in our GWAs (Jung et al., 2012; Kim et al., 2012; Watanabe et al., 2012). Gene positions were converted from the Glyma.wm82.a1 assembly to Glyma.wm82.a2 using the 'gene model correspondence lookup' available at SoyBase (SoyBase.org). Additionally, using the 'TAIR keyword search and browse' available at TAIR (arabidopsis.org) and selecting the option 'go biological process', we searched for the terms 'flowering', 'photoperiodism', and 'vegetative to reproductive phase transition of the meristem' (Berardini et al., 2015). Using the gene model names downloaded, all annotated genes on the Glyma.wm82.a2 were filtered.

3 | RESULTS

3.1 | GWA analyses on MGs among soybean germplasm

Soybean genetic diversity is unevenly distributed across MGs in the USDA Soybean Germplasm Collection (Figure 2, Table 1). Based on the principal component analysis plot, in general, soybean germplasm can be grouped approximately by MG. While some overlap exists, there is a clear separation between genotypes from MG III and earlier and those from MGs V+. A list of reported loci affecting maturity is presented in Table 2. For the loci where the causal genes have been identified, the physical position and the gene names are also provided in the table. Overall, GWA analyses using paired MGs identified 58 QTL, consisting of 146 SNPs that are significantly associated with differences in paired MGs (Table 3). These QTL included the genomic regions for reported *E* maturity loci, flowering and maturity genes *PRR3* and *PRR7*, and stem termination genes described in Table 2.

3.1.1 | MGs 000–00

Genome-wide association for MGs 000–00 did not reveal significant SNPs near known *E* maturity genes (Supplemental Figure S1a; Table 3). However, two QTL, *qMG*-13.4 and *qMG*-19.2 located on Chrs 13 and 19, showed a significant association with MGs 000–00 (Tables 3 and 4). The SNP at *qMG*-13.4 is only 85 kb from an ortholog of *TARGET OF EARLY ACTIVATION TAGGED (EAT) 1* (*TOE1*) (Supplemental Table S3). *TOE1* has been shown to inhibit early flowering in Arabidopsis by suppressing *Flowering Locus T* genes (Jung et al., 2007). The SNP for *qMG*-19.2 is <530 kb from *GmFT3b* and *GmFT5b*, which are soybean orthologs of *Flowering Locus T* genes and promoted flowering in Arabidopsis when expressed ectopically (Fan et al., 2014). Both QTL identified in this analysis have not been reported (Supplemental Table S4).

3.1.2 | MGs 00–0

The GWA between MGs 00–0 also did not yield significant SNPs near known *E* maturity loci (Table 4; Supplemental Figure S1b). However, four significant QTL, defined as *qMG*-5.2, −10.1, −13.1, and −15.5, were detected (Table 4). The SNPs for the QTL *qMG*-5.2 and −15.5 are both located near orthologs of the genes associated with *FLC* expression (Supplemental Table S3). The SNP for *qMG*-10.1 is near *GmPHYA1* (Supplemental Table S3), a homologous copy of

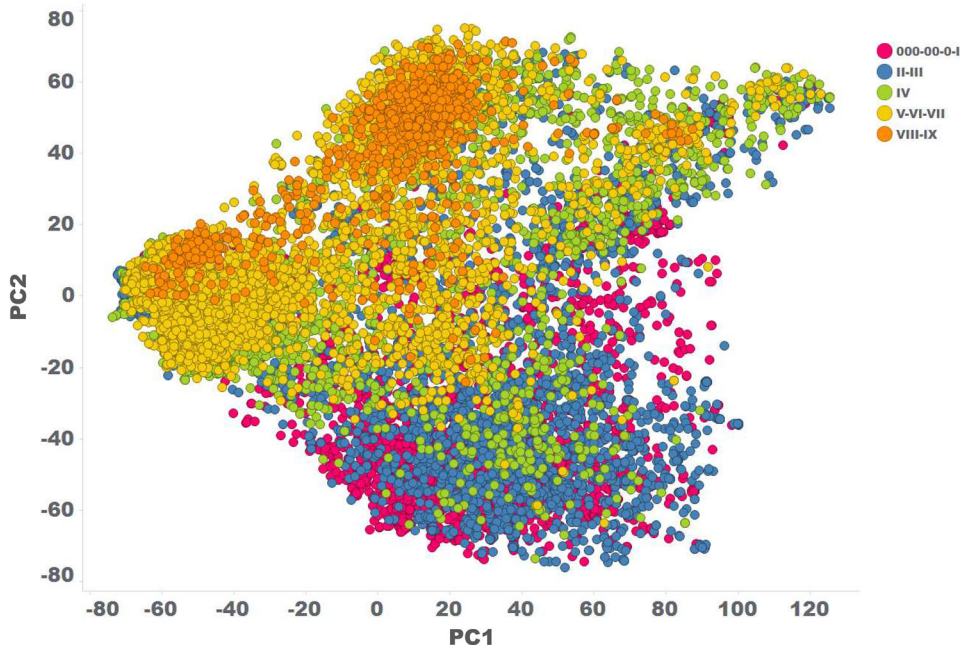

FIGURE 2 Principal component (PC) plot of soybean accessions from the USDA Soybean Germplasm Collection determined by genome-wide single nucleotide polymorphism markers. Accessions are colored by maturity group.

TABLE 1 Number of soybean [*Glycine max* (L.) Merr.] accessions from each maturity group used in this study

Maturity group	000	00	0	I	II	III	IV	V	VI	VII	VIII	IX	X	Total
Total no. of accessions[a]	136	509	1,150	1,719	2,011	1,954	4,037	2,707	1,547	940	913	30	1	17,654
No. of accessions used	136	498	1,119	1,642	1,868	1,823	3,939	2,572	1,476	901	887	18	0	16,879

[a]Number of accessions after filtering out *Glycine soja* accessions and duplicated entries on the raw data downloaded at Germplasm Resources Information Network (GRIN, https://www.ars-grin.gov/).

E4. However, the effect of *E4* was not detected in the individual GWA and only a minor effect was found in the GWA for MGs 000–IX. At *qMG*-13.1, the significant SNP is near an ortholog of *ENHANCED EM LEVEL* (*EEL*). *qMG*-10.1 and −13.1 were previously reported as GWA QTL in SoyBase (Supplemental Table S4). Furthermore, QTL overlapping *qMG*-5.2 and −15.5 were previously reported for the DTF in a study with 86 plant introductions from MGs 000 and 00 (Copley et al., 2018).

3.1.3 | MGs 0–I

Genome-wide association for MGs 0 and I detected 15 significant QTL, of which six QTL were located near the known *E* maturity loci: *E1*, *E1La*, *E2*, *E3*, *J*, and *E10* (Supplemental Figure S1c; Table 4). The most significant SNP in this analysis is located near *E1La*. The SNPs near *E1* and *E1La* presented extensive linkage blocks. The second and third most significant SNPs were located near *E2* and *E3*.

Nine additional QTL were also detected in this GWA analysis (Table 4). The most significant SNP among these regions was observed for *qMG*-15.2, near an ortholog of *SHOOT MERISTEMLESS* (*STM*) (Supplemental Table S3), a well-characterized regulator of shoot apical meristem maintenance in Arabidopsis, which is also involved in flower meristem identity (Roth et al., 2018). Significant SNPs for seven of these QTL (*qMG*-1.1, −2.2, −5.1, −7.1, −13.2, −15.1, and −15.3) are located near orthologs of the genes associated with the vernalization pathway and responses to low temperatures in Arabidopsis, especially genes that affect *FLC* expression (Supplemental Table S3). Of these seven regions, *qMG*-5.1, −13.2, and −15.1 are reported as QTL for maturity related traits (Supplemental Table S4).

TABLE 2 Reported soybean maturity loci or genes and flowering and stem termination genes and associated SNP markers

Name	Gene ID	Description	Chromosome	Gene position[a]	ID of the most significantly associated SNP	SNP position[a]	$-\log_{10}(P)$	MG[b]	Reference
				bp		bp			
Reported E maturity loci									
E1	*Glyma.06G207800*	B3 domain-containing	6	20,207,077–20,207,940	ss715593833	19,913,355	194.6	II–V	Bernard, 1971; Xia, et al., 2012
E1La	*Glyma.04G156400*	B3 domain-containing	4	36,758,125–36,758,770	ss715587784	38,334,609	47.8	I[c]	Xu, et al., 2015
E1Lb	*Glyma.04G143300*	B3 domain-containing	4	26,120,011–26,120,532	–	–	–	I[c], 000–II	Xu, et al., 2015; Zhu et al., 2019
E2	*Glyma.10G221500*	*GIGANTEA* (*GI*)	10	45,294,735–45,316,121	ss715607502	45,550,230	141.4	II–V	Bernard, 1971; Watanabe, et al., 2011
E3	*Glyma.19G224200*	*Phytochrome A* (*PHYA3*)	19	47,633,059–47,641,958	ss715635703	47,632,093	93.4	0–III	Buzzell, 1971; Watanabe, et al., 2009
E4	*Glyma.20G090000*	*Phytochrome A* (*PHYA2*)	20	33,236,018–33,241,692	ss715637690	36,473,914	13.1	00–II	Buzzell & Voldeng, 1980; Liu, et al., 2008
E5	Not Identified	Not Identified	10	–	–	–	–	II, IV	McBlain & Bernard, 1987
E6/J	*Glyma.04G050200*	*Early Flowering 3* (*ELF3*)	4	4,075,901–4,081,260	ss715587817	3,674,817	7.1	VI–VIII	Hartwig & Kiihl, 1979; Bonato & Vello, 1999; Yue, et al., 2017; Fang, et al., 2021; Nissan, et al., 2021
E7	Not Identified	Not Identified	6	–	ss715592723	10,863,803	28.0	000–II	Cober & Voldeng, 2001; Molnar, et al., 2003
E8	*Glyma.04g101500*	*Cryptochrome 1A* (*CRY1a*)	4	9,337,214–9,341,731	ss715589384	8,296,302	10.3	000-II	Cober, et al., 2010; Cheng, et al., 2011
E9	*Glyma.16G150700*	*Flowering Locus T* (*FT2a*)	16	31,109,999–31,114,963	ss715624382	31,208,131	22.4	NA	Kong, et al., 2014; Zhao, et al., 2016
E10	*Glyma.08G363100*	*Flowering Locus T* (*FT4*)	8	47,458,142–47,459,829	ss715602235	44,623,218	21.7	000–0	Samanfar, et al., 2016; Zhai, et al., 2014
E11	*Glyma.07G048500*	*LHY1/CCA1-like* (*LCL4*)	7	4,102,968–4,114,174	ss715598197	4,429,773	26.5	0–I	Wang, et al., 2019

(Continues)

TABLE 2 (Continued)

Name	Gene ID	Description	Chromosome	Gene position[a]	ID of the most significantly associated SNP	SNP position[a]	$-\log_{10}(P)$	MG[b]	Reference
Reported other flowering and maturity genes									
PRR3	*Glyma.U034500*	*Pseudo Response Regulator* 3	S32	197,150–220,019	ss715608786	10,848,358	46.3	NA	Ogiso-Tanaka, et al., 2019; Lu et al., 2020
PRR7	*Glyma.12g073900*	*Pseudo Response Regulator 7*	12	5,508,365–5,522,772	ss715613180	5,502,184	164.7	NA	Ogiso-Tanaka, et al., 2019; Li, et al., 2019; Lu et al., 2020
Stem termination genes									
Dt1	*Glyma.19G194300*	*TERMINAL FLOWER 1*	19	45,183,357–45,185,175	ss715635506	45,557,751	7.6	II–V	Bernard, 1972; Liu, et al., 2010
Dt2	*Glyma.18g273600*	*Agamous-Like 8 (AGL8)*	18	55,638,209–55,646,547	ss715632241	55,784,646	18.3	II–V	Bernard, 1972; Ping, et al., 2014

[a]Positions of the genes and SNPs were based on the Glyma.Wm82.a2.v1 genome assembly and the genes were identified from the SoyBase (http://SoyBase.org).
[b]Maturity group.
[c]Virus-induced gene silencing.

3.1.4 | MGs I–II

A total of 16 significant QTL were identified in GWA for MGs I–II. Of these regions, three were overlapped with *E1*, *E2*, and *E3 loci*, with the most significant SNP near *E2* (Table 4; Supplemental Fig. S1d). All known maturity loci detected in this analysis were also observed in the GWA for MGs 0–I. Additionally, 13 QTL were also detected in this analysis. Significant SNPs for six of these QTL, *qMG-5.2*, *−6.1*, *−13.3*, *−15.2*, *−16.4*, and *−20.1* are located very close to orthologs of the Arabidopsis genes associated with meristem identity (Supplemental Table S3). The SNP for *qMG-15.2* is the second most significant in this analysis, which is close to *Chromatin Remodeling Factor17* (*CHR17*). In Arabidopsis, *CHR17* prevents early activation of the vegetative-to-reproductive transition by regulating several key genes that contribute to flower timing such as *FT*, *SEP1*, *SEP3*, and *FUL* (Li et al., 2012). The third most significant SNP is in *qMG*-3.2, which is near an ortholog *NUCLEAR FACTOR Y SUBUNIT B8* (*NF-YB8*), and the SNP for *qMG*-3.3 close to an ortholog of *NUCLEAR FACTOR Y SUBUNIT C9* (*NF-YC9*). As observed in the previous GWA, significant SNPs for four QTL (*qMG*-3.1, -6.4, −7.1, and -19.5) were located near the genes involved in the regulation of *FLC* expression (Supplemental Table S3). The QTL *qMG*-3.2, −3.3, −6.4, −9.1, −16.4, and −20.1 identified in this analysis were also reported as GWA QTL in SoyBase (Supplemental Table S4).

3.1.5 | MGs II–III

Genome-wide association for MGs II–III detected significant SNPs near *E1* and *E2* and *Dt1* and *Dt2* (Table 4; Supplemental Figure S1e). The most significant SNP is located near *E2* and the second one is close to *E1*. Additionally, four QTL, *qMG*-5.3, −13.3, −15.4, and −16.1, were also detected in this analysis (Table 4). The SNP for *qMG*-16.1 is located near an ortholog of *HLP1*. In Arabidopsis, *HLP1* regulates plant flowering by targeting alternative polyadenylation, which ultimately suppresses *FLC* expression via *FCA* (Zhang et al., 2015b). The QTL *qMG*-13.3 was again detected with a SNP near an ortholog of *SHORT VEGETATIVE PHASE* (*SVP*). Two QTL, *qMG*-5.3 and −15.4, were only detected in this analysis. *qMG*-5.3 is located near *SPA1-related2* (*SPA2*), while *qMG*-15.4 is close to an ortholog of *CYCLING DOF FACTOR 2* (*CDF2*). However, *qMG*-15.4 could not be confirmed in GWA for MGs 000–IX. Of these four additional regions, *qMG*-5.3 and −16.1 have been reported as GWA QTL on SoyBase (Supplemental Table S4).

TABLE 3 Genomic regions detected with genome-wide association (GWA) analyses in paired maturity groups and a diversity panel or across maturity groups (MGs) 000 through IX

		Genomic region identified by individual GWAs in paired MGs and a diversity panel				Genomic region identified by GWA across MGs 000–IX			
Name of QTL	Chromosome	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$
		bp				bp		bp	
qMG-1.1	1	823,278–4,576,595	ss715579545	4,576,595	13.6	683,307–6,613,885	ss715579934	4,934,180	15.6
ECqMG-1.1	1	–	–	–	–	9,410,680–10,015,821	ss715580877	9,736,511	9.1
qMG-1.2	1	17,436,477	ss715578644	17,436,477	7.9	–	–	–	–
ECqMG-1.2	1	–	–	–	–	38,438,736–43,632,505	ss715579323	41,331,273	8.3
qMG-1.3	1	47,064,250–48,078,170	ss715579673	48,078,170	13.7	44,117,573–50,630,943	ss715579488	44,965,732	15.8
qMG-1.4	1	53,074,367–53,136,582	ss715580337	53,074,367	8.3	55,572,527–56,830,220	ss715580723	56,770,865	8.4
qMG-2.1	2	4,363,973–7,233,292	ss715583837	7,233,292	7.2	51,236–7,126,284	ss715583777	6,812,100	34.7
qMG-2.2	2	12,190,975	ss715581049	12,190,975	7.8	9,541,869–15,271,225	ss715581055	12,242,977	36.7
qMG-2.3	2	34,898,716	ss715582072	34,898,716	11.1	–	–	–	–
ECqMG-2.1	2	–	–	–	–	39,244,018–46,461,916	ss715583006	43,440,684	18.6
ECqMG-3.1	3	–	–	–	–	344,529–5,244,122	ss715586167	425,209	9.4
ECqMG-3.2	3	–	–	–	–	21,261,809–23,085,439	ss715584554	21,261,809	20.5
qMG-3.1	3	34,532,040	ss715585469	34,532,040	6.5	33,636,101	ss715585354	33,636,101	14.7
qMG-3.2	3	39,438,564–39,727,905	ss715586078	39,438,564	8.6	38,876,486–39,946,374	ss715585981	38,876,486	7.5
qMG-3.3	3	43,513,124	ss715586428	43,513,124	7.1	41,473,003–41,491,337	ss715586233	41,473,003	7.5
qMG-4.1	4	1,514,139–3,752,035	ss715587850	3,752,035	7	321,728–3,674,987	ss715587817	3,674,987	7.1
qMG-4.2	4	9,358,265	ss715589638	9,358,265	9.2	8,260,240–8,537,216	ss715589384	8,296,302	10.3
ECqMG-4.1	4	–	–	–	–	15,927,855–16,536,652	ss715587172	15,927,855	24.9
ECqMG-4.2	4	–	–	–	–	19,570,043–19,731,043	ss715587282	19,570,043	7.3
ECqMG-4.3	4	–	–	–	–	30,123,651	ss715587362	30,123,651	6.7
qMG-4.3	4	38,334,609–39,977,826	ss715587844	39,977,826	32.9	38,334,609	ss715587784	38,334,609	47.8
ECqMG-4.4	4	–	–	–	–	41,351,742–45,185,501	ss715587995	43,710,656	22.2
ECqMG-4.5	4	–	–	–	–	45,305,251–51,927,929	ss715588448	47,242,528	29.5

(Continues)

TABLE 3 (Continued)

		Genomic region identified by individual GWAs in paired MGs and a diversity panel				Genomic region identified by GWA across MGs 000–IX			
Name of QTL	Chrom-osome	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$
qMG-5.1	5	3,057,232	ss715589900	3,057,232	6.5	1,070,290–5,397,251	ss715592501	1,070,290	25.7
ECqMG-5.1	5	–	–	–	–	7,534,622–8,820,984	ss715592451	8,820,984	12.5
ECqMG-5.2	5	–	–	–	–	19,982,333	ss715589809	19,982,333	8.2
ECqMG-5.3	5	–	–	–	–	24,193,817–28,758,005	ss715590269	26,446,634	12.9
qMG-5.2	5	32,020,107–35,918,853	ss715590998	34,338,769	10.7	33,550,707–36,707,377	ss715591126	35,209,993	15.8
qMG-5.3	5	39,168,144	ss715592097	39,168,144	6.5	38,691,715–42,119,049	ss715591630	41,925,052	44.8
		41,571,382–41,598,296[‡]	ss715591680	41,571,382	7.2				
qMG-6.1	6	608,094	ss715595380	608,094	7.7	894,552	ss715595639	894,552	24.7
qMG-6.2	6	13,462,721–14,363,799	ss715593059	13,509,587	9	10,402,418–14,029,207	ss715592723	10,863,803	28.1
ECqMG-6.1	6	–	–	–	–	15,379,050–17,453,327	ss715593752	17,453,327	25.9
qMG-6.3	6	19,486,921–20,940,014	ss715593832	19,858,251	16.2	19,913,355–22,212,993	ss715593833	19,913,355	194.6
qMG-6.4	6	47,568,246–50,743,212	ss715594829	47,999,849	10.4	47,959,012–51,076,174	ss715595278	51,046,599	20.3
		47,416,685–47,526,846[b]	ss715594695	47,416,685	10.8				
qMG-7.1	7	1,159,635–1,443,929	ss715596226	1,443,929	7.8	1,789,219–2,222,940	ss715596814	2,222,940	6.6
ECqMG-7.1	7	–	–	–	–	4,429,773–6,306,291	ss715598197	4,429,773	26.5
qMG-7.2	7	15,741,518–18,327,575	ss715596697	18,327,575	7.8	–	–	–	–
qMG-7.3	7	35,220,682	ss715597312	35,220,682	7.3	35,613,115–38,770,938	ss715597441	36,523,592	9.9
ECqMG-7.2	7	–	–	–	–	41,776,639–42,909,420	ss715598143	42,909,420	10.6
ECqMG-8.1	8	–	–	–	–	1,632,674–3,810,391	ss715601551	3,604,631	20.6
ECqMG-8.2	8	–	–	–	–	7,470,603–9,909,193	ss715602913	9,909,193	20.3
ECqMG-8.3	8	–	–	–	–	11,259,469–11,843,090	ss715599301	11,571,237	9
ECqMG-8.4	8	–	–	–	–	16,161,353–17,748,776	ss715599780	16,161,353	13.3
ECqMG-8.5	8	–	–	–	–	35,572,498–36,272,305	ss715601530	36,075,677	9.4

(Continues)

TABLE 3 (Continued)

		Genomic region identified by individual GWAs in paired MGs and a diversity panel				Genomic region identified by GWA across MGs 000–IX			
Name of QTL	Chromosome	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$
ECqMG-8.6	8	–	–	–	–	39,599,594–43,023,964	ss715601743	40,648,803	32
qMG-8.1	8	45,603,518–46,696,857	ss715602415	46,696,857	10.5	44,623,218–47,636,307	ss715602235	44,623,218	21.7
		45,939,398[b]	ss715602332	45,939,398	6.7				
ECqMG-9.1	9	6,551,732[‡]	ss715605312	6,551,732	11	722,588–6,256,888	ss715605167	5,557,122	13.2
ECqMG-9.2	9	–	–	–	–	20,477,995–21,780,623	ss715603268	21,494,387	13.2
ECqMG-9.3	9	–	–	–	–	28,590,229–32,388,671	ss715603485	31,230,369	13.1
ECqMG-9.4	9	–	–	–	–	37,706,393–42,366,861	ss715603782	39,137,680	11.5
qMG-9.1	9	43,517,260–4,564,8182	ss715604581	45,416,118	7.3	43,072,567–48,792,461	ss715604897	48,543,094	22.2
ECqMG-10.1	10	–	–	–	–	3,057,808–5,366,082	ss715607708	4,735,320	15.5
ECqMG-10.2	10	7,741,398[b]	ss715608531	7,741,398	6.2	9,092,840–9,707,338	ss715608709	9,707,338	20.1
qDTF-10.1	10	14,457,726[b]	ss715605986	14,457,726	6.5	–	–	–	–
qMG-10.1	10	37,229,632	ss715606418	37,229,632	8	39,017,342–42,522,750	ss715606978	40,636,624	11
qMG-10.2	10	44,990,199–45,826,997	ss715607517	45,826,997	31.9	44,224,500–46,839,076	ss715607502	45,550,230	141.4
qMG-10.3	10	50,364,853	ss715608125	50,364,853	6.7	48,197,057–48,689,540	ss715607788	48,197,057	25.8
ECqMG-11.1	11	–	–	–	–	423,724–7,299,608	ss715611000	6,506,375	16.1
qMG-11.1	11	10,457,269–45,826,997	ss715608812	11,032,954	30.1	8,148,438–11,133,013	ss715608786	10,848,358	46.3
		10,477,945[b]	ss715608757	10,477,945	11.3				
ECqMG-11.2	11	–	–	–	–	15,649,090–15,678,644	ss715609885	15,649,090	11.6
qMG-11.2	11	19,622,661–24,156,840	ss715609781	19,622,661	9.1	20,324,493–26,217,355	ss715609757	20,324,493	26.9
		24,450,418–24,605,877[‡]	ss715609104	24,450,418	10.1				
ECqMG-11.3	11	–	–	–	–	28,292,491–33,838,162	ss715609996	28,333,194	8.4
qMG-12.1	12	2,839,426–7,829,045	ss715613180	5,502,184	19.9	1,464,938–7,545,940	ss715613180	5,502,184	164.7
		3,551,311–6,896,151[‡]	ss715613335	6,896,151	6.6				

(Continues)

TABLE 3 (Continued)

		Genomic region identified by individual GWAs in paired MGs and a diversity panel				Genomic region identified by GWA across MGs 000–IX			
Name of QTL	Chrom-osome	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$
qMG-12.2	12	34,178,221	ss715612380	34,178,221	7.2	35,429,338–36,530,315	ss715612674	36,444,348	11
ECqMG-12.1	12	–	–	–	–	39,605,154–39,740,802	ss715612992	39,628,194	9.5
ECqMG-13.1	13	–	–	–	–	437,287–7,767,773	ss715613779	437,287	9
qMG-13.1	13	12,937,337–19,373,536	ss715616972	15,183,485	15.2	10,706,972–20,896,502	ss715613749	10,706,972	27.7
		14,774,149[b]	ss715617053	14,774,149	6.6				
ECqMG-13.2	13	–	–	–	–	23,852,214–25,259,816	ss715614171	25,259,816	25.5
qMG-13.2	13	28,443,623	ss715614630	28,443,623	6.5	27,335,661–28,970,569	ss715614445	27,335,661	10
		28,674,444[b]	ss715614684	28,674,444	6.4				
qMG-13.3	13	36,221,688–37,756,892	ss715615732	36,221,688	7.6	33,204,263–39,607,602	ss715615977	37,756,892	31.0
qMG-13.4	13	42,512,630	ss715616362	42,512,630	8.6	42,643,435–43,322,945	ss715616434	43,322,945	8.9
ECqMG-14.1	14	–	–	–	–	38,001–4,131,877	ss715618715	4,131,877	12.9
ECqMG-14.2	14	8,109,384[b]	ss715619996	8,109,384	10.1	10,170,989	ss715617384	10,170,989	31.5
ECqMG-14.3	14	–	–	–	–	19,438,518	ss715618515	19,438,518	24
qMG-14.1	14	47,621,296	ss715619421	47,621,296	6.7	44,916,140–48,566,026	ss715619207	46,197,333	26.3
qMG-15.1	15	1,969,699–4,334,070	ss715621926	4,334,070	9.3	1,490,854–4,171,376	ss715621693	3,573,258	11.9
qMG-15.2	15	7,632,638–8,829,779	ss715623042	7,632,638	11.3	6,689,402–9,941,854	ss715623029	7,550,103	40.4
qMG-15.3	15	12,816,590–14,876,073	ss715620851	14,876,073	6.9	11,280,882–14,551,414	ss715620628	13,684,555	12.6
ECqMG-15.1	15	–	–	–	–	24,160,945–28,637,600	ss715621754	27,430,182	10.4
ECqMG-15.2	15	–	–	–	–	31,169,128–32,116,867	ss715621511	32,116,867	8.1
qMG-15.4	15	35,067,373	ss715621602	35,067,373	6.6	–	–	–	–
ECqMG-15.3	15	–	–	–	–	43,135,797	ss715621895	43,135,797	6.8
qMG-15.5	15	47,538,275–48,537,095	ss715622238	47,693,099	7.3	46,458,753–49,902,462	ss715622143	47,324,767	28
qMG-16.1	16	204,698–6,042,142	ss715623677	204,698	13.8	137,239–8,037,107	ss715625403	7,229,931	9.2
		114,805–139,928[b]	ss715623373	114,805	13.6				

(Continues)

TABLE 3 (Continued)

		Genomic region identified by individual GWAs in paired MGs and a diversity panel				Genomic region identified by GWA across MGs 000–IX			
Name of QTL	Chrom-osome	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$
qMG-16.2	16	26,791,660	ss715623863	26,791,660	9.9	27,169,242–28,765,102	ss715623971	28,762,445	10
qMG-16.3	16	29,084,941–30,777,681	ss715624158	29,609,620	7.9	29,309,383–31,765,761	ss715624382	31,208,131	22.4
qMG-16.4	16	36,556,952	ss715624868	36,556,952	7.6	33,406,638–37,653,674	ss715624927	37,000,804	10.7
qMG-17.1	17	2,160,065	ss715626414	2,160,065	7.8	3,237,615–3,644,762	ss715626835	3,532,847	19
qMG-17.2	17	7,488,569	ss715628177	7,488,569	9.5	7,913,612–8,285,863	ss715628225	7,913,612	7.6
ECqMG-17.1	17	–	–	–	–	11,379,945–14,188,211	ss715625783	11,404,483	8
ECqMG-17.2	17	–	–	–	–	35,528,444–36,248,107	ss715626948	36,180,745	12.1
qMG-17.3	17	39,528,940	ss715627640	39,528,940	7.6	37,879,637–41,052,244	ss715627754	40,575,507	16.2
ECqMG-18.1	18	–	–	–	–	433,470–9,365,931	ss715630397	3,706,238	19.1
ECqMG-18.2	18	–	–	–	–	18,240,230–22,192,855	ss715629942	22,192,855	14.4
qMG-18.1	18	43,090,752	ss715630715	43,090,752	6.7	41,820,094	ss715630648	41,820,094	8.1
ECqMG-18.3	18	–	–	–	–	49,080,769–50,727,159	ss715631442	49,080,769	12.1
qMG-18.2	18	53,076,934–57,094,471	ss715632205	55,478,846	11.2	54,391,672–57,272,786	ss715632241	55,784,646	18.3
		54,143,499[b]	ss715631995	54,143,499	5.8				
qMG-19.1	19	975,170–1,150,070	ss715633055	1,150,070	8.7	546,020–3,810,489	ss715634669	3,810,489	20
		5,674,894[b]	ss715636141	5,674,894	7.4				
qMG-19.2	19	34,484,934–36,578,790	ss715634147	34,484,934	7.1	33,080,644–37,185,287	ss715634523	36,869,893	13
		38,458,734[b]	ss715634725	38,458,734	7.7				
ECqMG-19.1	19	–	–	–	–	40,022,634–41,094,171	ss715634923	40,022,634	8.6
qMG-19.3	19	45,589,306–45,830,407	ss715635535	45,830,407	7.8	45,557,751–45,643,073	ss715635506	45,557,751	7.6
qMG-19.4	19	46,509,345–48,137,823	ss715635694	47,564,698	33	47,329,907–48,331,596	ss715635703	47,632,093	93.4
qMG-19.5	19	49,327,515–50,325,877	ss715635915	49,327,515	7.8	49,327,515	ss715635915	49,327,515	18.3
qMG-20.1	20	3,413,871	ss715637552	3,413,871	6.6	208,950–3,978,705	ss715639020	678,338	27.6
ECqMG-20.1	20	–	–	–	–	11,811,683–20,229,461	ss715636765	12,236,008	10.8

(Continues)

TABLE 3 (Continued)

Name of QTL	Chrom-osome	Genomic region identified by individual GWAs in paired MGs and a diversity panel				Genomic region identified by GWA across MGs 000–IX			
		Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$	Genomic region[a]	ID of the most significant SNP	SNP position[a]	$-\log_{10}(P)$
ECqMG-20.2	20	–	–	–	–	34,830,489–37,218,045	ss715637690	36,473,914	13.1
ECqMG-20.3	20	–	–	–	–	39,207,633–43,293,034	ss715638198	41,797,319	14.1
qMG-20.2	20	47,638,221	ss715638916	47,638,221	6.5	44,108,146–46,968,588	ss715638617	45,398,904	21

[a]Physical positions are based on Glyma.Wm82.a2.v1 genome assembly. The genomic regions were defined by significant SNPs across GWAs.
[b]Physical positions of SNPs detected with GWA in the diversity panel.

3.1.6 | MGs III–IV

The GWA for MGs III and IV revealed the greatest number of significant SNPs near known flowering and maturity loci, with SNPs being near seven *E* maturity loci and *PRR3* (Table 4; Supplemental Figure S1f). The most significant SNP was associated with *E3*, which was followed by *E1* and *PRR3*. A significant SNP associated with *Dt1* was also detected in this analysis, while *E8* was only detected in this analysis. The greater portion of the most significant SNPs in this analysis are associated with the photoperiodic and circadian clock pathways; however, six QTL (*qMG*-1.1, −6.2 [*E7*], −6.4, −17.2, −18.1, and −19.1) have significant SNPs close to orthologs of genes associated with the control *FLC* expression (Supplemental Table S3). Other QTL detected from this analysis were *qMG*-2.1, −9.1, and −13.1. The QTL *qMG*-2.1, −6.2, and −18.1 have been reported as GWA QTL in SoyBase (Supplemental Table S4).

3.1.7 | MGs IV–V

The GWA between MGs IV and V yielded 25 QTL. Significant SNPs near known *E* maturity loci were observed for *E1*, *E1La*, *E2*, *E7*, and *E10* (Table 4; Supplemental Figure S1g). Additionally, SNPs for *Dt2* and the recently identified *PRR3* and *PRR7* genes were also observed. The SNPs for *PRR3*, *E2*, and *PRR7* were most significant in this analysis, highlighting the importance of circadian-clock-related genes in MGs IV and V. For the other 17 QTL detected, SNPs for three regions (*qMG*-19.5, −13.1, and −16.1) are near orthologs of the genes associated with the circadian clock (Supplemental Table S3). Additionally, 13 QTL (*qMG*-1.4, −2.3, −3.2, −5.2, −8.1 [*E10*], −11.2, −12.2, −15.1, −15.3, −15.5, −16.2, −19.1, and −20.2) contained SNPs close to orthologs of genes associated with responses to environmental temperature (Supplemental Table S3). The SNP for *qMG*-5.2 is only 46 kb from the sole ortholog of *FLC* in the soybean, *Glyma.05g148700*, which has been implicated in long-term, low-temperature-triggered late flowering by inhibiting *FT* expression (Lyu et al., 2020). The other QTL detected from this analysis were *qMG*-7.2 and −14.1. In addition to the QTL previously mentioned in other analyses, *qMG*-1.4, −2.3, and −16.2 have been reported as maturity related GWA QTL in SoyBase (Supplemental Table S4).

3.1.8 | MGs V–VI

For MGs V–VI, significant SNPs near *E2*, *E3*, *E6*, *E9*, *E10*, and *PRR7* were detected (Table 4; Supplemental Figure S1h). The SNPs attributed to *E3*, *E10*, and *PRR7* were near orthologs of other Arabidopsis flowering genes (Supplemental Table S3). Five additional QTL were also identified (Table 4). The most significant SNP in this analysis was observed for *qMG*-16.1, which was closest to an ortholog of *VRN1*. Quantitative trait loci *qMG*-7.3, −17.1, and −17.3 were only detected in this analysis. The SNP for *qMG*-7.3 is closest to an ortholog of *SVP*, *qMG*-17.1 to an ortholog of *EARLY IN SHORT DAYS* 4 (*ESD4*) and *qMG*-17.3 to an ortholog of *MED25-BINDING RING-H2 PROTEIN2* (*MBR2*), while *qMG*-19.2 is closest to an ortholog of *MBR1*. In Arabidopsis, *MBR1* and *MBR2* promote the degradation of *MEDIATOR25* (*MED25*), which is required for the activation of *FT* transcription (Iñigo et al., 2012). Of these five additional regions, only *qMG*-17.3 was also reported as GWA QTL in SoyBase (Supplemental Table S4).

3.1.9 | MGs VI–VII

The GWA analysis in MGs VI–VII detected four QTL (Table 4; Supplemental Figure S1i). The most significant

TABLE 4 Quantitative trait loci (QTL) for maturity groups (MGs) detected by the genome-wide association analyses in paired MGs and across MGs 000 through IX

Name of QTL	Reported locus	Chromosome	000–IX	000–00	00–0	0–I	I–II	II–III	III–IV	IV–V	V–VI	VI–VII	VII–VIII	VIII–IX	No. MGs affected
qMG-1.1	–	1	15.6[a]	–	–	7.3	–	–	13.6	–	–	–	–	–	4
qMG-1.2	–	1	–	–	–	–	–	–	–	–	–	–	–	7.9	2
qMG-1.3	–	1	15.9	–	–	–	–	–	–	–	–	–	7.5	13.7	3
qMG-1.4	–	1	8.4	–	–	–	–	–	–	8.3	–	–	–	–	2
qMG-2.1	–	2	34.7	–	–	–	–	–	7.0	–	–	–	7.2	–	4
qMG-2.2	–	2	36.7	–	–	7.8	–	–	–	–	–	–	–	–	2
qMG-2.3	–	2	–	–	–	–	–	–	–	11.1	–	–	–	–	2
qMG-3.1	–	3	14.7	–	–	–	6.5	–	–	–	–	–	–	–	2
qMG-3.2	–	3	7.5	–	–	–	8.6	–	–	8.1	–	–	–	–	4
qMG-3.3	–	3	7.5	–	–	–	7.1	–	–	–	–	–	–	–	2
qMG-4.1	*E6/J*	4	7.1	–	–	7.0	–	–	–	–	6.5	–	–	–	4
qMG-4.2	*E8*	4	10.3	–	–	–	–	–	9.2	–	–	–	–	–	2
qMG-4.3	*E1La*	4	47.8	–	–	32.9	–	–	10.6	8.4	–	–	–	–	5
qMG-5.1	–	5	25.7	–	–	6.5	–	–	–	–	–	–	–	–	2
qMG-5.2	–	5	15.1	–	8.1	–	8.3	–	–	10.7	–	–	7.0	7.6	9
qMG-5.3	–	5	44.8	–	–	–		6.5	–	–	–	–	–	–	2
qMG-6.1	–	6	24.7	–	–	–	7.7	–	–	–	–	–	–	–	2
qMG-6.2	*E7*	6	28.0	–	–	–	–	–	9.0	8.6	–	–	7.1	7.6	6
qMG-6.3	*E1*	6	194.6	–	–	8.6	7.6	9.8	16.2	7.8	–	–	–	–	6
qMG-6.4	–	6	19.8	–	–	–	6.5	–	6.7	–	–	–	–	10.4	4
ECqMG-7.1	*E11*	7	26.5	–	–	–	–	–	–	–	–	–	–	–	na
qMG-7.1	–	7	6.6	–	–	6.6	7.0	–	–	–	–	–	7.8	–	5
qMG-7.2	–	7	–	–	–	7.8	–	–	–	6.7	–	–	–	–	4
qMG-7.3	–	7	9.9	–	–	–	–	–	–	–	7.3	–	–	–	2
qMG-8.1	*E10*	8	21.7	–	–	10.5	–	–	–	10.0	6.7	–	–	–	5
qMG-9.1	–	9	22.2	–	–	–	6.7	–	7.3	–	–	–	–	–	4
qMG-10.1	–	10	11.0	–	8.0	–	–	–	–	–	–	–	–	–	2
qMG-10.2	*E2*	10	141.4	–	–	27.7	31.9	20.5	7.4	22.2	7.7	13.2	–	–	8
qMG-10.3	–	10	25.8	–	–	–	–	–	–	–	–	6.7	–	–	2
qMG-11.1	*PRR3*	11	46.3	–	–	–	–	–	15.7	30.1	–	8.2	–	–	5
qMG-11.2	–	11	26.9	–	–	–	–	–	–	9.1	–	–	–	–	2
q-MG12.1	*PRR7*	12	164.7	–	–	–	–	–	–	19.9	8.9	9.7	9.5	–	5
qMG-12.2	–	12	11.0	–	–	–	–	–	–	7.2	–	–	–	–	2
qMG-13.1	–	13	27.7	–	8.6	–	–	–	7.5	15.2	–	–	–	–	5
qMG-13.2	–	13	18.3	–	–	6.5	–	–	–	–	–	–	–	–	2
qMG-13.3	–	13	31.0	–	–	–	6.7	7.6	–	–	–	–	–	–	3
qMG-13.4	–	13	8.9	8.6	–	–	–	–	–	–	–	–	–	–	2
qMG-14.1	–	14	26.3	–	–	–	–	–	–	6.7	–	–	–	–	2
qMG-15.1	–	15	11.9	–	–	7.3	–	–	–	9.3	–	–	–	–	4
qMG-15.2	–	15	40.4	–	–	8.3	11.3	–	–	–	–	–	–	–	3
qMG-15.3	–	15	12.6	–	–	6.7	–	–	–	6.9	–	–	–	–	4
qMG-15.4	–	15	–	–	–	–	–	6.6	–	–	–	–	–	–	2
qMG-15.5	–	15	28.0	–	6.7	–	–	–	–	7.3	–	–	6.6	–	6
qMG-16.1	–	16	9.15	–	–	–		7.7	–	13.8	10.8	–	–	–	5

(Continues)

TABLE 4 (Continued)

Name of QTL	Reported locus	Chromosome	000–IX	000–00	00–0	0–I	I–II	II–III	III–IV	IV–V	V–VI	VI–VII	VII–VIII	VIII–IX	No. MGs affected
qMG-16.2	–	16	10.9	–	–	–	–	–	–	9.9	–	–	–	–	2
qMG-16.3	*E9*	16	22.4	–	–	–	–	–	6.7	–	7.9	–	–	–	4
qMG-16.4	–	16	10.7	–	–	–	7.6	–	–	–	–	–	–	–	2
qMG-17.1	–	17	19.0	–	–	–	–	–	–	–	7.8	–	–	–	2
qMG-17.2	–	17	7.6	–	–	–	–	–	9.5	–	–	–	–	–	2
qMG-17.3	–	17	16.2	–	–	–	–	–	–	–	7.6	–	–	–	2
qMG-18.1	–	18	8.1	–	–	–	–	–	6.7	–	–	–	–	–	2
qMG-18.2	*Dt2*	18	18.3	–	–	–	6.9	7.1	–	11.2	–	–	–	6.9	7
qMG-19.1	–	19	20.0	–	–	–	–	–	7.8	8.7	–	–	–	–	3
qMG-19.2	–	19	13.0	6.7	–	–	–	–	–	–	7.1	–	–	–	4
qMG-19.3	*Dt1*	19	7.6	–	–	–	–	–	7.8	–	–	–	–	–	3
qMG-19.4	*E3*	19	93.4	–	–	12.2	8.0	–	33.0	–	7.3	–	–	–	7
qMG-19.5	–	19	18.3	–	–	–	7.8	–	–	7.3	–	–	–	–	4
ECqMG-20.2	*E4*	20	13.1	–	–	–	–	–	–	–	–	–	–	–	0
qMG-20.1	–	20	27.6	–	–	–	6.6	–	–	–	–	–	–	–	2
qMG-20.2	–	20	21.0	–	–	–	–	–	–	8.4	–	–	–	–	2
Significant threshold	–	–	6.6	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	–
Sum of loci	–	–	56	2	4	15	17	8	17	25	11	4	7	6	–

[a]The values of $-\log_{10}(p)$, indicating significance of SNPs.

SNP was observed for *E2*, while SNPs near *PRR3* and *PRR7* were also detected. Furthermore, a significant SNP for QTL *qMG*-10.3 was only detected in this analysis, which is near orthologs of *JMJC DOMAIN-CONTAINING PROTEIN 27* (*JMJ27*) and *TCP7*, a member of the *TCP* (*TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR*) transcription factor family.

3.1.10 | MGs VII–VIII

The GWA between MGs VII and VIII revealed seven significant QTL (Table 4; Supplemental Figure S1j). The most significant SNP is near *PRR7*. A SNP for the QTL where *E7* might be located is also observed. Furthermore, SNPs for the QTL *qMG*-1.3, −2.1, −5.2, −7.1, and −15.5 were also detected. The QTL *qMG*-1.3 has been reported as GWA QTL for RP in SoyBase (Supplemental Table S4).

3.1.11 | MGs VIII–IX

For MGs VIII–IX, six QTL were identified (Table 4; Supplemental Figure S1k) including SNPs near *E7* and *Dt2* loci. The most significant SNP was observed for QTL *qMG*-1.3, which was also detected in the GWA for MGs VII–VIII. Of three additional QTL, *qMG*-1.2, −5.2, and −6.4, detected, *qMG*-1.2 was only detected in this analysis and was not confirmed in the GWA with MGs 000–IX. There were no known orthologs of Arabidopsis flowering genes close to its position.

3.1.12 | MGs 000–IX

To understand allele variation across MGs 000–IX and confirm the results from each individual analysis, GWA was performed across MGs 000–IX. The analysis identified 103 QTL (Table 3), confirming 54 of the QTL identified in individual GWAs that encompassed 618 significant SNPs. The additional 49 QTL identified in this GWA comprised 508 significant SNPs; 13 have been reported as GWA QTL for maturity related traits in SoyBase, while the other 36 are newly identified GWA QTL (Supplemental Table S4). In this analysis, significant SNPs were identified near all known maturity loci, the stem termination genes, and the recently discovered *PRR* genes (Table 4; Figure 3). The most significant SNPs were observed, in descending order, for *E1*, *PRR7*, *E2*, *E3*, *E1La*, and *PRR3*. Aside from QTL reported near maturity loci and *PRR* genes, four QTL identified in individual GWA and confirmed in this analysis (*qMG*-2.1, −2.2, −5.3, and −15.2) presented highly significant SNPs ($-\log_{10}(P) > 34$) with QTL *qMG*-2.2 and −15.2, representing new, unreported GWA QTL for maturity (Supplemental Table S4). For QTL only detected in this GWA, none reached the stringent threshold of $-\log_{10}$

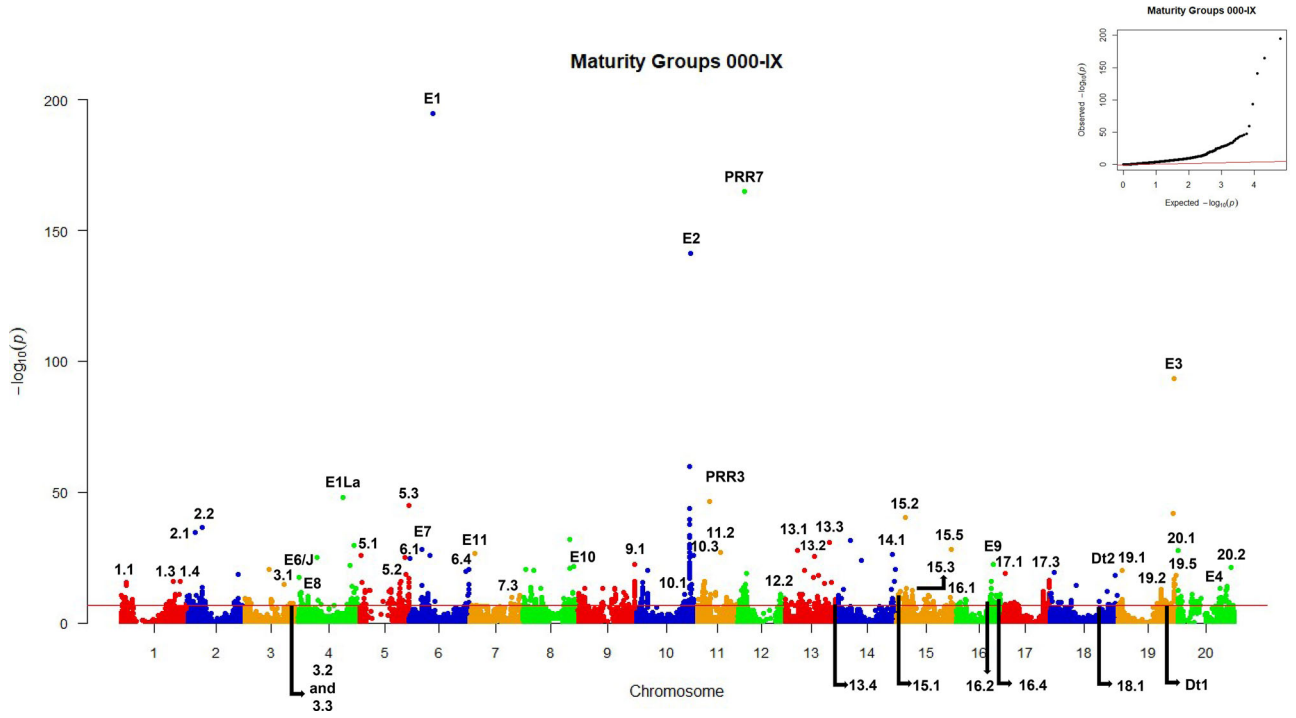

FIGURE 3 Manhattan plot displaying the results of the genome-wide association (GWA) for the accessions across maturity groups 000 to IX. Red threshold line set at $-\log_{10}(2.84 \times 10^{-7}) = \sim 6.55$. The known maturity loci are indicated with most significantly associated single nucleotide polymorphisms on each chromosome. Quantitative trait loci confirmed from individual GWA analyses are presented without the prefix '*qMG*' for ease of visualization. The quantile–quantile plot of observed vs. expected *P* values is presented in the top right corner.

(*P*) > 34; however, seven QTL presented higher significance than reported loci with known effects on maturity such as *Dt1*, *E6*, *E8*, *E9*, *E10*, and *J* (Table 3). Furthermore, five of these seven QTL represent new unreported QTL for maturity (Supplemental Table S4). For QTL detected within individual GWA, a total of 13 also presented higher significance than the abovementioned maturity loci, with eight of these QTL representing new GWA QTL for maturity.

3.2 | Validation using a diversity panel

Soybean maturity is affected by two major periods: from planting to flowering and from flowering to physiological maturity. Increase in DTF or RP periods can extend DTM. Correlation analysis indicated that both DTF and RP had a significant positive correlation with DTM ($P < .01$) with correlation coefficients of 0.85 and 0.53, respectively. No significant correlation was observed between DTF and RP. Besides confirming 15 QTL identified in previous analyses, GWA using the diversity panel also detected one additional QTL for DTF (Tables 3 and 5).

3.2.1 | Days to flowering

Five QTL were identified in a combined DTF dataset from 2 yr (Figure 4, Table 5). Of these five significant regions, four regions have been identified with GWA analyses. The most significant SNP was detected in *ECqMG*-9.1, near an ortholog of *ELF4-Like3* (*EFL3*) (Supplemental Table S3). The region was also observed in the GA-16 environment (Supplemental Fig. S2b). The second most significant SNP is in *qMG*-6.4 near an ortholog of *VOZ1*. The region was detected in both environments (Supplemental Figure S2a and S2b). The third SNP is in *qMG*-5.3, close to an ortholog of *FLAVIN-BINDING*, *KELCH REPEAT*, *F BOX 1* (*FKF1*), and was also detected in GA-16 environment. The QTL *qDTF-10.1* was only observed for the combined DTF dataset and was not observed in the GWAs with MGs 000–IX. A significant SNP for *PRR7* was also detected in the combined dataset. *E10* was observed individually for GA-15 and GA-16 environments but not in the combined dataset. Furthermore, one SNP near *Dt2* was significant in the GA-15 environment only (Table 5).

3.2.2 | Days to maturity

Four QTL, *qMG*-11.2, −13.1, −16.1, and −19.2, were identified to be significantly associated with DTM in GWA analyses using the combined dataset (Figure 4, Table 5). The most significant SNP for the combined dataset is *qMG*-16.1, which is near an ortholog of *ELF4-Like 4* (*EFL4*) (Supplemental Table S3). The second most significant SNP was observed for *qMG*-19.2 and was closest to an ortholog of *AREB3*. For *qMG*-11.2, the SNP was not near any known orthologs of Arabidopsis

TABLE 5 Genomic regions detected for days to flowering (DTF), days to maturity (DTM), and reproductive period (RP) using a genome-wide association analysis in a diverse panel of soybean genotypes

Name of genomic regions	Reported locus	2015			2016			Combined		
		DTF	DTM	RP	DTF	DTM	RP	DTF	DTM	RP
qMG-5.3	–	–	–	–	6.9[a]	–	–	7.2	–	–
qMG-6.4	–	10.8	5.9	–	6.5	–	–	9.6	–	–
qMG-8.1	*E10*	6.7	–	–	5.9	–	–	–	–	–
ECqMG-9.1	–	–	–	–	7.8	–	–	11.0	–	–
ECqMG-10.2	–	–	–	–	–	6.2	–	–	–	–
qDTF-10.1	–	–	–	–	–	–	–	6.5	–	–
qMG-11.1	*PRR3*	–	–	–	–	11.3	–	–	–	–
qMG-11.2	–	–	–	–	–	–	10.1	–	6.7	6.8
qMG-12.1	*PRR7*	–	–	–	–	–	6.4	6.6	–	–
qMG-13.1	–	–	–	–	–	–	–	6.6	–	–
qMG-13.2	–	–	–	–	–	–	–	–	–	6.4
ECqMG-14.2	–	–	–	–	–	–	10.1	–	–	–
qMG-16.1	–	–	–	–	–	10.7	9.4	–	13.6	12.1
qMG-18.2	*Dt2*	5.8	–	–	–	–	–	–	–	–
qMG-19.1	–	–	–	–	–	–	–	–	–	7.4
qMG-19.2	–	–	–	–	–	–	–	–	7.7	

[a]The values of $-\log_{10}(p)$, indicating significance of SNPs.

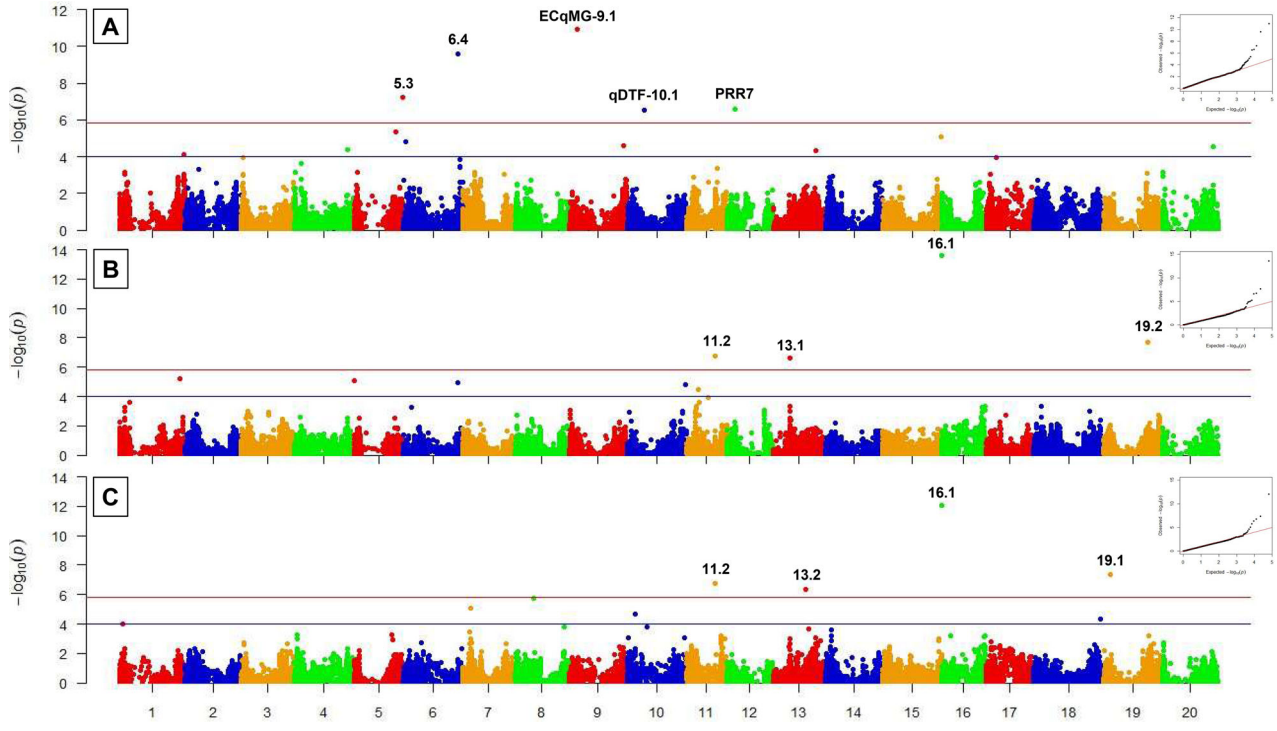

FIGURE 4 Manhattan plots displaying the results of the genome-wide association (GWA) for the diversity panel using a combined dataset. Red threshold line set at $-\log_{10}(1.44 \times 10^{-7}) = \sim 5.84$. The quantitative trait loci overlapping the ones from individual GWA analyses are presented without the prefix '*qMG*.' Quantile–quantile plots of observed vs. expected *P* values are presented in the top right corner of each Manhattan plot. (A) Days to flowering; (B) Days to maturity; (C) Reproductive period.

flowering genes while the SNP for *qMG*-13.1 was close to an ortholog of *COL2*. Only one significant QTL *qMG*-6.4 was observed in GA-15 (Supplemental Figure S2c), while three QTL, *ECqMG-10.2*, *PRR3*, and *qMG*-16.1, were identified in GA-16 environment (Supplemental Figure S2d). Except for *ECqMG-10.2* and *qMG*-13.1, all other regions observed were also found in the individual analyses for MGs V–IX.

3.2.3 | Reproductive period

Four significant QTL, *qMG*-11.2, −13.2, −16.1, and −19.1, from the diversity panel were identified for RP (Table 5, Figure 4) and these regions have also been detected in previous GWA analyses. Of these four regions, two QTL were also detected in the GWA for DTM. None of these regions are in common with those detected for DTF. The GA-16 environment also confirmed QTL *qMG*-16.1 and -11.2 and detected a SNP for *PRR7* and ECqMG-14.2. No significant SNPs were observed for this trait in GA-15.

4 | DISCUSSION

4.1 | Genetic control of soybean maturity

Soybean maturity is a complex trait controlled by the interaction of many genes, pathways, and the cultivation environment (Cober et al., 1996, 2001, 2014). Even with the application of stringent thresholds, we observed a large number of QTL in this study. This might be attributed to the availability of a high number of phenotypic observations, a common limitation with other GWA studies, and the more efficient detection of true positives by FarmCPU (Liu et al., 2016). While several new, unreported GWA QTL were detected, QTL for all reported maturity loci were observed. From the total of 108 QTL detected on all GWAs, 60 are previously unreported GWA QTL for maturity (Supplemental Table S4). Aside from QTL attributed to reported maturity genes, 45 QTL were identified in individual GWA analyses and 25 of these are newly identified GWA QTL.

All 14 reported maturity loci and two stem termination genes were confirmed in the GWA across MGs 000–IX and GWAs using the paired MGs, except for *E4* and *E11*, which were not detected in the GWAs with the paired MGs. None of the reported maturity loci were detected in MGs 000–0. Of those reported loci, *E1*, *E1L*, *E2*, *E3*, *E7*, *E10*, *PRR3*, and *PRR7* affected five or more MGs throughout the USDA Soybean Germplasm Collection. In contrast, of 45 QTL identified in individual GWA analyses, five QTL, including *qMG*-5.2, −7.1, −13.1, −15.5, and −16.1, impacted five or more MGs. Although *Dt1* and *Dt2* are stem termination genes, they had effects on three and seven MGs, respectively (Table 4).

Individual GWAs demonstrated that the *E1*, *E1La*, *E8*, and *Dt1* loci had effects on MGs V and earlier; in contrast, *E7*, *PRR3*, *and PRR7* loci had impacts on MGs III+ (Table 4). Furthermore, of the 45 QTL identified in individual GWA, 29 QTL were observed that had effects in MGs V and earlier, while there were only six QTL that had effects in MGs V+ (Table 4). Among all the significant loci identified in this study, *qMG*-5.2 had effects on the highest number (9) of MGs from 000 to IX, followed by *E2*, *E3*, *Dt2*, *qMG*-15.5, *E1*, *qMG-13.1*, *qMG-7.1*, and *qMG-16.1*, which affected 5–7 MGs. *E2* presented distributed effects in MGs 0–VII, possessing the most significant SNPs for the GWAs between MGs I–II, II–III, and VI–VII and the second most significant SNP for MGs 0–I and MGs IV–V. Although *qMG-5.2* affected more MGs than *E2*, the effects were small compared with *E2*.

The existence of multiple QTL near each other is a common challenge for QTL detection when working with complex traits (Du et al., 2019). Furthermore, several orthologs of Arabidopsis flowering-related genes are present within the *qMG-5.2* region (Supplemental Table S3). These factors might be responsible for the lower significance observed for this QTL despite its widespread presence in several MGs. There were 22 QTL that affected only two to four MGs in individual GWA but had SNPs with greater significance in the GWA with MG 000–IX compared with *qMG-5.2* (Table 3).

Overall, genes associated with the photoperiodic and circadian clock pathways have the greatest effects on soybean maturity, with *E2* among the reported maturity loci affecting the highest number of MGs (Table 4, Figure 3). The absence of significant SNPs near the *E1* locus on MGs VI–IX and near *E3* on MGs VII–IX supports previous evidence that most genotypes from MGs V+ have *E1* and *E3* alleles (Jiang et al., 2014; Langewisch et al., 2017; Li et al., 2017b; Liu et al., 2020; Miranda et al., 2020). Furthermore, the detection of SNPs near *PRR3* and *PRR7* only on MGs III+ supports previous indications that these genes depend on *E1* to delay maturity and are upregulated by *E3* (Lu, et al., 2020).

Quantitative trait loci for the *E4* locus were not detected in individual GWAs, which might be due to the restricted distribution of the recessive *e4* alleles (Kanazawa et al., 2009). Genotyping of 308 cultivars detected *e4-SORE* and *e4-kes* in only 1.3 and 1.9% of the cultivars, respectively, and these alleles were restricted to MGs 0000 and 000 (Liu et al., 2020). These results suggest that recessive *e4* alleles have a limited relevance for accessions in the USDA Soybean Germplasm Collection. Overall, most SNPs near maturity genes were detected in or near the MGs used for their discovery except for *E7* (Tables 2 and 4). The SNPs significantly associated with *E7*, reported in the crosses involving genotypes from MGs 000–II, were detected within MGs III–IX.

Our results provided evidence for further studies on *E5*, *E7*, and *E8* and suggested some candidate genes. For *E5*, previous results suggested that it might be either an allele of *E2*

or a closely located gene (Dissanayaka et al., 2016) with the molecular phenotype for *E5* NILs, indicating the existence of a unique locus (Wu et al., 2019). Our results indicated a QTL in the vicinity of *E2* for the GWA with MGs VI–VII and MGs 000–IX (*qMG*-10.3, Table 3). The most significant SNP ($P = 1.59 \times 10^{-26}$) for this QTL is located near an ortholog of Arabidopsis flowering genes (Supplemental Table S3), which might be responsible for the *E5* locus.

Two QTL detected in our GWA analysis were associated with *E7*. The first is *qMG*-6.2, where significant SNPs were detected for several paired MG GWAs (Table 4), most frequently near an ortholog of *LEAFY* (*LFY*), a floral identity gene. However, the most significant SNPs for this region on the GWA with MGs 000–IX were close to an ortholog of *PRR5*, which, considering the recent discovery of *PRR3* and *PRR7*, might also affect maturity in soybean. Furthermore, the QTL *ECqMG*-6.1 also fits the order observed by Cober and Voldeng (2001) with a highly significant SNP near an ortholog of *CYCLING DOF FACTOR 3* (*CDF3*) (Supplemental Table S3).

The causal gene for *E8* has been attributed to genes in two different QTL on Chr 4. Significant SNPs near *E1La* (Tables 2 and 3), which might be the true gene for the *E8* locus (Cao et al., 2017), were detected in several GWAs (Table 4). In contrast, a SNP near *GmCRY1a*, suggested by Cheng et al. (2011), was only detected for MGs III–IV and had lower significance than *E1La* in the GWA with MGs 000–IX. The presence of *E1* and *E3* alleles for MGs 000–0 was previously predicted in the USDA Germplasm Collection using en masse prediction (Langewisch et al., 2017). However, significant SNPs near *E1* and *E3* were not detected in GWAs for MGs 000–00 and MGs 00–0. These results might indicate that the effect of these alleles might depend on the gene located in *qMG*-4.3. Nevertheless, resequencing of *E1La* and *E1Lb* from 192 accessions in a soybean mini-core collection did not find any high-effect variants to the gene function in these loci (Ogiso-Tanaka et al., 2019). A recessive allele for *E1Lb*, which eliminated the inhibitory effect of far-red light-enriched long-day conditions, was recently discovered in crosses with far-eastern Russian soybean cultivars (Zhu et al., 2019). Another possible candidate gene located in the region is *Glyma.04G159300*, an ortholog of *FRUITFUL* (*FUL*). The expression of *Glyma.06g205800*, also an ortholog of *FUL*, has been previously associated with maturity in soybean varieties with different photoperiod sensitivities; however, *FUL* homologs and their functions are still poorly understood in soybean (Jia et al., 2015). Furthermore, the semi-indeterminate gene *Dt2* is an ortholog of *FUL* (Ping et al., 2014).

The late MGs in the USDA Soybean Germplasm Collection, in contrast with early MGs, were strongly affected by genes associated with the circadian-clock-associated *PRR* genes (Table 4). However, Lu et al. (2020) observed that *PRR3* and *PRR7* are upregulated by *E3* and *E4*, while *E1* acts both downstream of *E3* and the *PRR* genes being, therefore, simultaneously associated with both pathways. In our analyses, several significant SNPs were identified near orthologs of genes associated with the vernalization pathway, especially genes associated in the regulation of *FLC* expression, the *FLC* ortholog itself and three orthologs of *SVP* (Supplemental Table S3). However, while SNPs near vernalization related orthologs appeared in considerable numbers in some GWAs, such as for MGs 0–I, the significance of the SNPs is usually inferior compared with SNPs associated with the circadian and photoperiodic pathways.

Despite consisting in a key integrator of flowering signals, *FT* orthologs had only a modest effect on soybean maturity. While significant SNPs near *E9* (*GmFT2a*), *E10* (*GmFT4*), and *GmFT5b* were detected in individual GWAs, the significance of SNPs near these genes in the GWA with MGs 000–IX is relatively low. Furthermore, the SNPs attributed to *E10* were usually closer to the orthologs of other Arabidopsis flowering genes than to the actual position for the loci (Supplemental Table S3). The soybean genome has 14 orthologs of *FT* (Zhai et al., 2014), which might be responsible for the reduced effect of allele variants in specific locus; however, this would require that we consider that the number of functional *FT* alleles has reduced effects on the phenotypes.

4.2 | Genome-wide association QTL for DTF, DTM, and RP

In soybean, the period from planting to flowering is genetically controlled by multiple genes, while only a few known maturity genes, such as *E3*, *E4*, and *E8*, also have been reported that affect the reproductive period (Cheng et al., 2011; Xu et al., 2013). Genome-wide association QTL for the reproductive period near *E1La*, *E2*, *E7*, *PRR3*, and *PRR7* have been reported in SoyBase. Correlation coefficients indicated that both DTF and RP had significant positive correlations with DTM ($P < .01$) and there was no correlation between DTF and RP, suggesting that the lengths of DTF and RP in the diversity panel might be controlled by different genes. The GWA QTL detected for these traits support the hypothesis drawn for the correlation analysis.

The QTL detected for DTM were very similar to those identified for RP, which shared the genomics regions *qMG-11.2* and *qMG-16.1* for the combined datasets; however, none of the QTL observed for DTF in the combined dataset were detected for RP. The most significant SNPs for DTM and RP are near an ortholog of *ELF4-Like 4* (*EFL4*) on *qMG*-16.1. The QTL mapping for the long-juvenile trait recently identified a QTL that encompasses *qMG*-16.1 (Fang et al., 2019). This trait is of special importance for soybean varieties cultivated on low latitudes, which are usually from MGs VI–X.

While SNPs near *EFL4* were not significant for DTF, the most significant SNPs for this trait was an ortholog of *ELF4-Like 3* (*EFL3*). The causative gene for the long-juvenile *j-1* allele is an ortholog of *ELF3* (Yue et al., 2017).

Significant SNPs near an ortholog of *VOZ1* on Chr 6 were consistently observed for DTF and once for DTM in GA-15. Additionally, a SNP near another ortholog of this gene was observed for RP in Chr 13 (Table 5; Supplemental Table S3). Accordingly, significant SNPs near orthologs of *CONTANS-Like 2* (*COL2*) on Chr 13 and 19 were detected for DTM and RP, respectively. While no significant markers near *COL* orthologs were detected for DTF, SNPs near an ortholog of *FKF1* in Chr 5 were observed in two instances (Table 5). Overall, most of the QTL detected in the diversity panel overlapped with those identified in individual MG GWAs with MGs V–IX and the GWA with MGs 000–IX. The significant SNPs detected in the diversity panel are generally near the orthologs of genes that interact with *PRR7* and *GI*, which were consistently significant on late MGs.

4.3 | Implications of genetic variation of maturity loci in plant breeding

Based on the GWA analyses using paired MGs, a range of two to 25 QTL were identified for each pair of MGs, which indicated the genetic controls for the MGs might be more complex in some of MGs than others. Genome-wide association across MGs 000–IX identified more QTL (103) than those using paired MGs (58 QTL), where 54 QTL were in common with those identified across MGs 000–IX. These 49 new QTL identified by the GWA across MGs 000–IX might be due to the allele variation across MGs 000–IX. It is possible that some of these QTL have large effects on the MGs that are not detected using the materials from the paired maturity groups. A follow-up analysis will be conducted to investigate the effects of these QTL across MGs that were not detected using the materials from the paired MGs.

Using $-\log_{10}(P) = 34$ as a threshold, 10 QTL were detected by the GWA across MGs 000–IX, which were also identified by individual GWAs including reported loci *E1*, *E2*, *E3*, *E1La*, *PRR3*, *and PRR7* using paired MGs. When we use $-\log_{10}(P) = 15$ as a threshold, GWA across MGs 000–IX detected 38 QTL, of which 37 QTL were also identified by GWAs with paired MGs. This suggested that the majority of high-effect QTL could be detected using a combination of GWAs across MGs and with paired MGs. The number of QTL conditioning each MG from this study could be used to predict maturity groups in a breeding program.

Maturity's large impacts on soybean yield and other agronomic traits make it a significant factor in accelerated breeding pipelines that use genomic selection, limiting the formation of training sets across multiple MGs. Genomic selection models are best able to capture small to moderate marker effects (Heslot et al., 2012; Meuwissen et al., 2001; Stewart-Brown et al., 2019; Whittaker et al., 2000). The large impacts of maturity on low heritability traits, such as grain yield, reduces the model's accuracy. Therefore, maturity must be compensated for within a genomic selection model, typically through the addition of DTM as an environmental effect (Crain et al., 2018; Rutkoski et al., 2016). The use of maturity genes as weighted covariates in genomic selection models presents an opportunity for increased genomic selection accuracy (Smallwood et al., 2019).

In Smallwood et al. (2019), plant materials within the training and validation sets segregated for maturity, most notably at the *E1* locus. By including *E1* and *Dt1* as covariates separate from the other markers used in cross-validation, grain yield prediction reached an accuracy of 0.51, as indicated by Pearson correlation. These results indicated that genomic selection across segregating maturities is possible if the segregating maturity genes are accounted for as covariates (Smallwood et al., 2019). This would make the prediction of yield and other agronomic traits possible when crossing distant maturity groups with greater accuracy.

Our results demonstrated that the GWA QTL with the greatest effects identified from paired MGs are mostly associated with characterized *E* maturity loci such as *E1*, *E2*, *E3*, and the *PRR* genes *PRR3* and *PRR7*. The available evidence suggests that functional *E1* and *E3* alleles are necessary to obtain MGs V+ (Jiang et al., 2014; Li et al., 2017b) and that *PRR* genes *PRR3* and *PRR7*, consistently detected on late MGs in our study, might depend on *E1* to delay flowering (Lu et al., 2020). These results indicate that manipulating a few genes might be sufficient to significantly change the MG from an accession, as it was the case during the successful introgression of the long-juvenile trait in temperate cultivars to develop elite cultivars in Brazil and southern China (Lu et al., 2017). Besides conventional breeding strategies, recent studies were also able to use the CRISPR/CAS9 technology to develop earlier and later flowering germplasm by mutating or overexpressing maturity related genes (Han et al., 2019; Lu et al., 2020; Wang et al., 2020).

Additionally, for some paired MGs, the number of significant QTL appear too small, such as for MGs 000–0, where only a total of six QTL were detected. In comparison, a total of 28 QTL were yielded for MGs V–IX, compared with only 25 QTL for MGs IV–V alone. Principal component analyses demonstrated a greater genetic diversity for MGs II–IV (Figure 2), which somewhat agree with our results for the number of QTL. Nevertheless, these results might be partially a consequence of a reduced number of genotypes for testing in early (000–0) and late (V–IX) MGs when compared with the GWA for MGs IV–V (Table 1). Identification and characterization of these new GWA QTL could enable the development of significantly better models

especially for use in genomic prediction to improve optimal cross-predictions.

ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)–Finance Code 001 by providing a fellowship to the first author. Support of this research was also provided by the University of Georgia Research Foundation and Georgia Agricultural Experiment Stations. We thank Tatyana Nienow at the University of Georgia for her technical support.

DATA AVAILABILITY STATEMENT

The phenotypes of the accessions used in this study can be accessed through the Germplasm Resources Information Network (GRIN) (www.ars.grin.gov). High-density SNP data for the accessions used in genome-wide association analyses are publicly available and can be retrieved at SoyBase (www.soybase.org). Phenotypes, DTF, DTM, and RP of the diversity panel are listed in the Supplemental Table S2. All other datasets generated for this study are available upon request.

CONFLICT OF INTEREST DISCLOSURE

The authors declare that there is no conflict of interest.

ORCID

Gustavo Zimmer https://orcid.org/0000-0001-7164-889X
Clinton J. Steketee https://orcid.org/0000-0003-3868-2396
Scott A. Jackson https://orcid.org/0000-0002-3172-1607
Zenglu Li https://orcid.org/0000-0003-4114-9509

REFERENCES

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis Information Resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, *53*, 474–485. https://doi.org/10.1002/dvg.22877

Bernard, R. L. (1971). Two major genes for time of flowering and maturity in soybeans. *Crop Science*, *11*, 242–244. https://doi.org/10.2135/cropsci1971.0011183X001100020022x

Bernard, R. L. (1972). Two genes affecting stem termination in soybeans. *Crop Science*, *12*, 235–239. https://doi.org/10.2135/cropsci1972.0011183X001200020028x

Boerma, H.R, & Specht, J.E. (2004). Soybeans: Improvement, production, and uses (3rd Vol.). ASA, CSSA, and SSSA, Madison, Wisconsin.

Bonato, E. R., & Vello, N. A. (1999). E6, A dominant gene conditioning early flowering and maturity in soybeans. *Genetics and Molecular Biology*, *22*, 229–232. https://doi.org/10.1590/S1415-47571999000200016

Buzzell, R. I. (1971). Inheritance of a soybean flowering response to fluorescent daylength conditions. *Canadian Journal of Genetics and Cytology*, *13*, 703–707. https://doi.org/10.1139/g71-100

Buzzell, R., & Voldeng, H. (1980). Inheritance of insensitivity to long daylength. *Soybean Genetics Newsletter*, *7*, 26–29.

Cao, D., Takeshima, R., Zhao, C., Liu, B., Jun, A., & Kong, F. (2017). Molecular mechanisms of flowering under long days and stem growth habit in soybean. *Journal of Experimental Botany*, *68*, 1873–1884. https://doi.org/10.1093/jxb/erw394

Cheng, L., Wang, Y., Zhang, C., Wu, C., Xu, J., Zhu, H., Leng, J., Bai, Y., Guan, R., Hou, W., Zhang, L., & Han, T. (2011). Genetic analysis and QTL detection of reproductive period and post-flowering photoperiod responses in soybean. *Theoretical and Applied Genetics*, *123*, 421–429. https://doi.org/10.1007/s00122-011-1594-8

Cober, E., Curtis, D., Stewart, D., & Morrison, M. (2014). Quantifying the effects of photoperiod, temperature and daily irradiance on flowering time of soybean isolines. *Plants*, *3*, 476–497. https://doi.org/10.3390/plants3040476

Cober, E. R., Molnar, S. J., Charette, M., & Voldeng, H. D. (2010). A new locus for early maturity in soybean. *Crop Science*, *50*, 524–527. https://doi.org/10.2135/cropsci2009.04.0174

Cober, E. R., Stewart, D. W., & Voldeng, H. D. (2001). Photoperiod and temperature responses in early-maturing, near-isogenic soybean lines. *Crop Science*, *41*, 721–727. https://doi.org/10.2135/cropsci2001.413721x

Cober, E. R., Tanner, J. W., & Voldeng, H. D. (1996). Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. *Crop Science*, *36*, 601–605. https://doi.org/10.2135/cropsci1996.0011183X003600030013x

Cober, E. R., & Voldeng, H. D. (2001). A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. *Crop Science*, *41*, 698–701. https://doi.org/10.2135/cropsci2001.413698x

Copley, T. R., Duceppe, M. O., & O'donoughue, L. S. (2018). Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. *BMC Genomics*, *19*, 1–12. https://doi.org/10.1186/s12864-018-4558-4

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., & Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *The Plant Genome*, *11*, 170043. https://doi.org/10.3835/plantgenome2017.05.0043

Destro, D., Carpentieri-Pípolo, V., Kiihl, R. A. S., & Almeida, L. A. (2001). Photoperiodism and genetic control of the long juvenile period in soybean: A review. *Crop Breeding and Applied Biotechnology*, *1*, 72–92. https://doi.org/10.13082/1984-7033.v01n01a10

Dissanayaka, A., Rodriguez, T. O., Di, S., Yan, F., Githiri, S. M., Rodas, F. R., Abe, J., & Takahashi, R. (2016). Quantitative trait locus mapping of soybean maturity gene *E5*. *Breeding Science*, *66*, 407–415. https://doi.org/10.1270/jsbbs.15160

Du, B., Wang, Q., Sun, G., Ren, X., Cheng, Y., Wang, Y., Gao, S., Li, C., & Sun, D. (2019). Mapping dynamic QTL dissects the genetic architecture of grain size and grain filling rate at different grain-filling stages in barley. *Scientific Reports*, *9*, 18823. https://doi.org/10.1038/s41598-019-53620-5

Fan, C., Hu, R., Zhang, X., Wang, Xu, Zhang, W., Zhang, Q., Ma, J., & Fu, Y. F. (2014). Conserved CO–FT regulons contribute to the photoperiod flowering control in soybean. *BMC Plant Biology*, *14*, 9. https://doi.org/10.1186/1471-2229-14-9

Fang, C., Chen, L., Nan, H., Kong, L., Li, Y., Zhang, H., Li, H., Li, T., Tang, Y., Hou, Z., Dong, L., Cheng, Q., Lin, X., Zhao, X., Yuan, X., Liu, B., Kong, F., & Lu, S. (2019). Rapid identification of consistent novel QTLs underlying long-juvenile trait in soybean by

multiple genetic populations and genotyping-by-sequencing. *Molecular Breeding*, *39*, 80. https://doi.org/10.1007/s11032-019-0979-2

Fang, C., Liu, J., Zhang, T., Su, T., Li, S., Cheng, Q., Kong, L., Li, X., Bu, T., Li, H., Dong, L., Lu, S., Kong, F., & Liu, B. (2021). A recent retrotransposon insertion of *J* caused *E6* locus facilitating soybean adaptation into low latitude. *Journal of Integrative Plant Biology*, *63*, 995–1003 . https://doi.org/10.1111/jipb.13034

Hagely, K., Konda, A. R., Kim, J. H., Cahoon, E. B., & Bilyeu, K. (2021). Molecular-assisted breeding for soybean with high oleic/low linolenic acid and elevated vitamin E in the seed oil. *Molecular Breeding*, *41*, 3. https://doi.org/10.1007/s11032-020-01184-y

Han, J., Guo, B., Guo, Y., Zhang, Bo, Wang, X., & Qiu, L. J. (2019). Creation of early flowering germplasm of soybean by CRISPR/Cas9 technology. *Frontiers in Plant Science*, *10*, 1446. https://doi.org/10.3389/fpls.2019.01446

Hartwig, E. E., & Kiihl, R. A. S. (1979). Identification and utilization of a delayed flowering character in soybeans for short-day conditions. *Field Crops Research*, *2*, 145–151. https://doi.org/10.1016/0378-4290(79)90017-0

Heslot, N., Yang, H. P., Sorrells, M. E., & Jannink, J. L. (2012). Genomic selection in plant breeding: A comparison of models. *Crop Science*, *52*, 146–160. https://doi.org/10.2135/cropsci2011.06.0297

Iñigo, S., Giraldez, A. N., Chory, J., & Cerdán, P. D. (2012). Proteasome-mediated turnover of Arabidopsis MED25 is coupled to the activation of *FLOWERING LOCUS T* transcription. *Plant Physiology*, *160*, 1662–1673. https://doi.org/10.1104/pp.112.205500

Jia, Z., Jiang, B., Gao, X., Yue, Y., Fei, Z., Sun, H., Wu, C., Sun, S., Hou, W., & Han, T. (2015). *GmFULa*, a *FRUITFULL* homolog, functions in the flowering and maturation of soybean. *Plant Cell Reports*, *34*, 121–132. https://doi.org/10.1007/s00299-014-1693-5

Jiang, B., Nan, H., Gao, Y., Tang, L., Yue, Y., Lu, S., Ma, L., Cao, D., Sun, S., Wang, J., Wu, C., Yuan, X., Hou, W., Kong, F., Han, T., & Liu, B. (2014). Allelic combinations of soybean maturity loci *E1*, *E2*, *E3* and *E4* result in diversity of maturity and adaptation to different latitudes. *PLoS ONE*, *9*, e106042. https://doi.org/10.1371/journal.pone.0106042

Jung, C. H., Wong, C. E., Singh, M. B., & Bhalla, P. L. (2012). Comparative genomic analysis of soybean flowering genes. *PLoS ONE*, *7*, e38250. https://doi.org/10.1371/journal.pone.0038250

Jung, J. H., Seo, Y. H., Seo, P. J., Reyes, J. L., Yun, J., Chua, N. H., & Park, C. M. (2007). The *GIGANTEA*-regulated microRNA172 mediates photoperiodic flowering independent of *CONSTANS* in *Arabidopsis*. *Plant Cell*, *19*, 2736–2748. https://doi.org/10.1105/tpc.107.054528

Kanazawa, A., Liu, B., Kong, F., Arase, S., & Abe, J. (2009). Adaptive evolution involving gene duplication and insertion of a novel *Ty1/copia*-like retrotransposon in soybean. *Journal of Molecular Evolution*, *69*, 164–175. https://doi.org/10.1007/s00239-009-9262-1

Kim, M. Y., Shin, J. H., Kang, Y. J., Shim, S. R., & Lee, S. H. (2012). Divergence of flowering genes in soybean. *Journal of Biosciences*, *37*, 857–870. https://doi.org/10.1007/s12038-012-9252-0

Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B. M., Watanabe, S., Yamada, T., Tabata, S., Kanazawa, A., Harada, K., & Abe, J. (2010). Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. *Plant Physiology*, *154*, 1220–1231. https://doi.org/10.1104/pp.110.160796

Kong, F., Nan, H., Cao, D., Li, Y., Wu, F., Wang, J., Lu, S., Yuan, X., Cober, E. R., Abe, J., & Liu, B. (2014). A new dominant gene *E9* conditions early flowering and maturity in soybean. *Crop Science*, *54*, 2529–2535. https://doi.org/10.2135/cropsci2014.03.0228

Kong, L., Lu, S., Wang, Y., Fang, C., Wang, F., Nan, H., Su, T., Li, S., Zhang, F., Li, X., Zhao, X., Yuan, X., Liu, B., & Kong, F. (2018). Quantitative trait locus mapping of flowering time and maturity in soybean using next-generation sequencing-based analysis. *Frontiers in Plant Science*, *9*, 995. https://doi.org/10.3389/fpls.2018.00995

Kumar, V., Rani, A., Rawal, R., & Mourya, V. (2015). Marker assisted accelerated introgression of null allele of kunitz trypsin inhibitor in soybean. *Breeding Science*, *65*, 447–452. https://doi.org/10.1270/jsbbs.65.447

Langewisch, T., Lenis, J., Jiang, G. L., Wang, D., Pantalone, V., & Bilyeu, K. (2017). The development and use of a molecular model for soybean maturity groups. *BMC Plant Biology*, *17*, 91. https://doi.org/10.1186/s12870-017-1040-4

Li, C., Li, Y. H., Li, Y., Lu, H., Hong, H., Tian, Yu, Li, H., Zhao, T., Zhou, X., Liu, J., Zhou, X., Jackson, S. A., Liu, B., & Qiu, L. J. (2020). A domestication-associated gene *GmPRR3b* regulates the circadian clock and flowering time in soybean. *Molecular Plant*, *13*, 745–759. https://doi.org/10.1016/j.molp.2020.01.014

Li, G., Zhang, J., Li, J., Yang, Z., Huang, H., & Xu, L. (2012). Imitation switch chromatin remodeling factors and their interacting RINGLET proteins act together in controlling the plant vegetative phase in Arabidopsis. *Plant Journal*, *72*, 261–270. https://doi.org/10.1111/j.1365-313X.2012.05074.x

Li, J., Wang, X., Song, W., Huang, X., Zhou, J., Zeng, H., Sun, S., Jia, H., Li, W., Zhou, X., Li, S., Chen, P., Wu, C., Guo, Y., Han, T., & Qiu, L. (2017b). Genetic variation of maturity groups and four E genes in the Chinese soybean mini core collection. *PLoS ONE*, *12*, e0172106 https://doi.org/10.1371/journal.pone.0172106

Li, M. W., Liu, W., Lam, H. M., & Gendron, J. M. (2019a). Characterization of two growth period QTLs reveals modification of *PRR3* genes during soybean domestication. *Plant & Cell Physiology*, *60*, 407–420. https://doi.org/10.1093/pcp/pcy215

Li, X., Fang, C., Xu, M., Zhang, F., Lu, S., Nan, H., Su, T., Li, S., Zhao, X., Kong, L., Yuan, X., Liu, B., Abe, J., Cober, E R., & Kong, F. (2017a). Quantitative trait locus mapping of soybean maturity gene *E6*. *Crop Science*, *57*, 2547–2554. https://doi.org/10.2135/cropsci2017.02.0106

Li, Y., Dong, Y., Wu, H., Hu, Bo, Zhai, H., Yang, J., & Xia, Z. (2019b). Positional cloning of the flowering time QTL *qFT12-1* reveals the link between the clock related *PRR* homolog with photoperiodic response in soybeans. *Frontiers in Plant Science*, *10*, 1303. https://doi.org/10.3389/fpls.2019.01303

Lipka, A E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P J., Gore, M A., Buckler, E S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, *28*, 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., & Abe, J. (2008). Genetic redundancy in soybean photoresponses associated with duplication of the Phytochrome A gene. *Genetics*, *180*, 995–1007. https://doi.org/10.1534/genetics.108.092742

Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., Nagamatsu, A., Arai, M., Yamada, T., Kitamura, K., Masuta, C., Harada, K., & Abe, J. (2010). The soybean stem growth habit gene *Dt1* is an ortholog of Arabidopsis *TERMINAL FLOWER1*. *Plant Physiology*, *153*, 198–210. https://doi.org/10.1104/pp.109.150607

Liu, L., Song, W., Wang, L., Sun, X., Qi, Y., Wu, T., Sun, S., Jiang, B., Wu, C., Hou, W., Ni, Z., & Han, T. (2020). Allele

combinations of maturity genes *E1–E4* affect adaptation of soybean to diverse geographic regions and farming systems in China. *PLoS ONE*, *15*, e0235397. https://doi.org/10.1371/journal.pone.0235397

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, *12*, e1005957. https://doi.org/10.1371/journal.pgen.1005767

Lu, S., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., Chen, L., Su, T., Nan, H., Zhang, D., Zhang, L., Wang, Z., Yang, Y., Yu, D., Liu, X., Yang, Q., Lin, X., Tang, Y., Zhao, X., … Kong, F. (2020). Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. *Nature Genetics*, *52*, 428–436. https://doi.org/10.1038/s41588-020-0604-7

Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., Fang, C., Cao, D., Shi, X., Kong, L., Su, T., Zhang, F., Li, S., Wang, Z., Yuan, X., Cober, E. R., Weller, J. L., Liu, B., Hou, X., … Kong, F. (2017). Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nature Genetics*, *49*, 773–779. https://doi.org/10.1038/ng.3819

Lyu, J., Cai, Z., Li, Y., Suo, H., Yi, R., Zhang, S., & Nian, H. (2020). The floral repressor *GmFLC-like* is involved in regulating flowering time mediated by low temperature in soybean. *International Journal of Molecular Sciences*, *21*, 1322. https://doi.org/10.3390/ijms21041322

McBlain, B. A., & Bernard, R. L. (1987). A new gene affecting the time of flowering and maturity in soybeans. *Journal of Heredity*, *78*, 160–162. https://doi.org/10.1093/oxfordjournals.jhered.a110349

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E., (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*, 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Miladinović, J., Ćeran, M., Đorđević, V., Balešević-Tubić, S., Petrović, K., Đukić, V., & Miladinović, D. (2018). Allelic variation and distribution of the major maturity genes in different soybean collections. *Frontiers in Plant Science*, *9*, 1286. https://doi.org/10.3389/fpls.2018.01286

Miranda, C., Scaboo, A., Cober, E., Denwar, N., & Bilyeu, K. (2020). The effects and interaction of soybean maturity gene alleles controlling flowering time, maturity, and adaptation in tropical environments. *BMC Plant Biology*, *20*, 65. https://doi.org/10.1186/s12870-020-2276-y

Molnar, S. J., Rai, S., Charette, M., & Cober, E. R. (2003). Simple sequence repeat (SSR) markers linked to *E1*, *E3*, *E4*, and *E7* maturity genes in soybean. *Genome*, *46*, 1024–1036. https://doi.org/10.1139/g03-079

Mourtzinis, S., & Conley, S. P. (2017). Delineating soybean maturity groups across the United States. *Agronomy Journal*, *109*, 1397–1403. https://doi.org/10.2134/agronj2016.10.0581

Nan, H., Cao, D., Zhang, D., Li, Y., Lu, S., Tang, L., Yuan, X., Liu, B., & Kong, F. (2014). GmFT2a and GmFT5a redundantly and differentially regulate flowering through interaction with and upregulation of the bZIP transcription factor GmFDL19 in soybean. *PLoS ONE*, *9*, e97669. https://doi.org/10.1371/journal.pone.0097669

Nissan, N., Cober, E. R., Sadowski, M., Charette, M., Golshani, A., & Samanfar, B. (2021). Identifying new variation at the *J* locus, previously identified as *e6*, in long juvenile 'Paranagoiana' soybean. *Theoretical and Applied Genetics*, *134*, 1007–1014. https://doi.org/10.1007/s00122-020-03746-2

Ogiso-Tanaka, E., Shimizu, T., Hajika, M., Kaga, A., & Ishimoto, M. (2019). Highly multiplexed AmpliSeq technology identifies novel variation of flowering time-related genes in soybean (*Glycine max*). *DNA Research*, *26*, 243–260. https://doi.org/10.1093/dnares/dsz005

Ortel, C. C., Roberts, T. L., Hoegenauer, K. A., Purcell, L. C., Slaton, N. A., & Gbur, E. E. (2020). Soybean maturity group and planting date influence grain yield and nitrogen dynamics. *Agrosystems, Geoscience & Environment*, *3*, e20077. https://doi.org/10.1002/agg2.20077

Ping, J., Liu, Y., Sun, L., Zhao, M., Li, Y., She, M., Sui, Yi, Lin, F., Liu, X., Tang, Z., Nguyen, H., Tian, Z., Qiu, L., Nelson, R L., Clemente, T E., Specht, J E., & Ma, J. (2014). *Dt2* is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. *The Plant Cell*, *26*, 2831–2842. https://doi.org/10.1105/tpc.114.126938

Prenger, E M., Yates, J., Mian, M. A. R, Buckley, B., Boerma, H. R., & Li, Z. (2019). Introgression of a high protein allele into an elite soybean cultivar results in a high-protein near-isogenic line with yield parity. *Crop Science*, *59*, 2498–2508. https://doi.org/10.2135/cropsci2018.12.0767

Ray, J D., Hinson, K., Mankono, J. E B, & Malo, M F. (1995). Genetic control of a long-juvenile trait in soybean. *Crop Science*, *35*, 1001–1006. https://doi.org/10.2135/cropsci1995.0011183X003500040012x

Roth, O., Alvarez, J P., Levy, M., Bowman, J L., Ori, N., & Shani, E. (2018). The KNOXI transcription factor SHOOT MERISTEMLESS regulates floral fate in Arabidopsis. *Plant Cell*, *30*, 1309–1321. https://doi.org/10.1105/tpc.18.00222

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., Reynolds, M., & Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes|Genomes|Genetics*, *6*, 2799–2808. https://doi.org/10.1534/g3.116.032888

Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., Belzile, F., & Cober, E. R. (2016). Mapping and identification of a potential candidate gene for a novel maturity locus, *E10*, in soybean. *Theoretical and Applied Genetics*, *130*, 377–390. https://doi.org/10.1007/s00122-016-2819-7

Scott, W. O., & Aldrich, S. R. (1970). *Modern soybean production* (1st ed.). S & A Publ. Inc.

Smallwood, C. J., Saxton, A. M., Gillman, J. D., Bhandari, H. S., Wadl, P. A., Fallen, B. D., Hyten, D. L., Song, Q., & Pantalone, V. R. (2019). Context-specific genomic selection strategies outperform phenotypic selection for soybean quantitative traits in the progeny row stage. *Crop Science*, *59*, 54–67. https://doi.org/10.2135/cropsci2018.03.0197

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. (2015). Fingerprinting soybean germplasm and its utility in genomic research. *G3 Genes|Genomes|Genetics*, *5*, 1999–2006. https://doi.org/10.1534/g3.115.019000

Steketee, C. J., Schapaugh, W. T., Carter, T. E., & Li, Z. (2020). Genome-wide association analyses reveal genomic regions controlling canopy wilting in soybean. *G3 Genes|Genomes|Genetics*, *10*, 1413–1425. https://doi.org/10.1534/g3.119.401016

Stewart-Brown, B. B., Song, Q., Vaughn, J. N., & Li, Z. (2019). Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3 Genes|Genomes|Genetics*, *10*, 2253–2265. https://doi.org/10.1534/g3.118.200917

Stowe, K. D., & Dunphy, J.. (2017). The soybean plant. In *North Carolina soybean production guide* (pp. 3–10). AG-835. North Carolina State Extension.

Tsubokura, Y., Watanabe, S., Xia, Z., Kanamori, H., Yamagata, H., Kaga, A., Katayose, Y., Abe, J., Ishimoto, M., & Harada, K. (2014). Natural variation in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. *Annals of Botany*, *113*, 429–441. https://doi.org/10.1093/aob/mct269

Turner, D. S. (2018). qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. *The Journal of Open Source Software*, *3*, 731. https://doi.org/10.21105/joss.00731

Wang, L., Sun, S., Wu, T., Liu, L., Sun, X., Cai, Y., Li, J., Jia, H., Yuan, S., Chen, Li, Jiang, B., Wu, C., Hou, W., & Han, T. (2020). Natural variation and CRISPR/Cas9-mediated mutation in *GmPRR37* affect photoperiodic flowering and contribute to regional adaptation of soybean. *Plant Biotechnology Journal*, *18*, 1869–1881. https://doi.org/10.1111/pbi.13346

Wang, F., Nan, H., Chen, L., Fang, C., Zhang, H., Su, T., Li, S., Cheng, Q., Dong, L., Liu, B., Kong, F., & Lu, S. (2019). A new dominant locus, *E11*, controls early flowering time and maturity in soybean. *Molecular Breeding*, *39*, 70. https://doi.org/10.1007/s11032-019-0978-3

Watanabe, S., Harada, K., & Abe, J. (2012). Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breeding Science*, *61*, 531–543. https://doi.org/10.1270/jsbbs.61.531

Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., Sato, S., Nakamoto, Y., Yamanaka, N., Takahashi, R., Ishimoto, M., Anai, T., Tabata, S., & Harada, K. (2009). Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics*, *182*, 1251–1262. https://doi.org/10.1534/genetics.108.098772

Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., Takahashi, R., Anai, T., Tabata, S., Kitamura, K., & Harada, K. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics*, *188*, 395–407. https://doi.org/10.1534/genetics.110.125062

Whittaker, J., Thompson, R., & Denham, M. (2000). Marker-assisted selection using ridge regression. *Genetic Research*, *75*, 249–252. https://doi.org/10.1017/S0016672399004462

Wu, F., Kang, X., Wang, M., Haider, W., Price, W. B., Hajek, B., & Hanzawa, Y. (2019). Transcriptome-enabled network inference revealed the *GmCOL1* feed-forward loop and its roles in photoperiodic flowering of soybean. *Frontiers in Plant Science*, *10*, 1221. https://doi.org/10.3389/fpls.2019.01221

Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., Anai, T., Sato, S., Yamazaki, T., Lu, S., Wu, H., Tabata, S., & Harada, K. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proceedings of the National Academy of Sciences*, *109*, E2155–E2164. https://doi.org/10.1073/pnas.1117982109

Xu, M., Xu, Z., Liu, B., Kong, F., Tsubokura, Y., Watanabe, S., Xia, Z., Harada, K., Kanazawa, A., Yamada, T., & Abe, J. (2013). Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. *BMC Plant Biology*, *13*, 91. https://doi.org/10.1186/1471-2229-13-91

Xu, M., Yamagishi, N., Zhao, C., Takeshima, R., Kasai, M., Watanabe, S., Kanazawa, A., Yoshikawa, N., Liu, B., Yamada, T., & Abe, J. (2015). The soybean-specific maturity gene *E1* family of floral repressors controls night-break responses through down-regulation of *FLOWERING LOCUS* T orthologs. *Plant Physiology*, *168*, 1735–1746. https://doi.org/10.1104/pp.15.00763

Yue, L., Li, X., Fang, C., Chen, L., Yang, H., Yang, J., Chen, Z., Nan, H., Chen, L., Zhang, Y., Li, H., Hou, X., Dong, Z., Weller, J. L., Abe, J., Liu, B., & Kong, F. (2021). FT5a interferes with the Dt1-AP1 feedback loop to control flowering time and shoot determinacy in soybean. *Journal of Integrated Plant Biology*, *63*, 1004–1020. https://doi.org/10.1111/jipb.13070

Yue, Y., Liu, N., Jiang, B., Li, Mu, Wang, H., Jiang, Ze, Pan, H., Xia, Q., Ma, Q., Han, T., & Nian, H. (2017). A single nucleotide deletion in *J* encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. *Molecular Plant*, *10*, 656–658. https://doi.org/10.1016/j.molp.2016.12.004

Zhai, H., Lü, S., Liang, S., Wu, H., Zhang, X., Liu, B., Kong, F., Yuan, X., Li, J., & Xia, Z. (2014). *GmFT4*, a homolog of *FLOWERING LOCUS T*, is positively regulated by E1 and functions as a flowering repressor in soybean. *PLoS ONE*, *9*, e89030. https://doi.org/10.1371/journal.pone.0089030

Zhang, J., Song, Q., Cregan, P., Nelson, R., Wang, X., Wu, J., & Jiang, G. (2015a). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics*, *16*, 217. https://doi.org/10.1186/s12864-015-1441-4

Zhang, Q., Li, H., Li, R., Hu, R., Fan, C., Chen, F., Wang, Z., Liu, X., Fu, Y., & Lin, C. (2008). Association of the circadian rhythmic expression of GmCRY1a with a latitudinal cline in photoperiodic flowering of soybean. *Proceedings of the National Academy of Sciences*, *105*, 21028–21033. https://doi.org/10.1073/pnas.0810585105

Zhang, Y., Gu, L., Hou, Y., Wang, L., Deng, X., Hang, R., Chen, D., Zhang, X., Zhang, Yi, Liu, C., & Cao, X. (2015b). Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein, regulates plant flowering by targeting alternative polyadenylation. *Cell Research*, *25*, 864–876. https://doi.org/10.1038/cr.2015.77

Zhao, C., Takeshima, R., Zhu, J., Xu, M., Sato, M., Watanabe, S., Kanazawa, A., Liu, B., Kong, F., Yamada, T., & Abe, J. (2016). A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of *FT2a*, a *FLOWERING LOCUS T* ortholog. *BMC Plant Biology*, *16*, 20. https://doi.org/10.1186/s12870-016-0704-9

Zhu, J., Takeshima, R., Harigai, K., Xu, M., Kong, F., Liu, B., Kanazawa, A., Yamada, T., & Abe, J. (2019). Loss of function of the *E1-Like-b* gene associates with early flowering under long-day conditions in soybean. *Frontiers in Plant Science*, *9*, 1867. https://doi.org/10.3389/fpls.2018.01867

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Zimmer G, Miller MJ, Steketee CJ, Jackson SA, de Tunes LVM, Li Z. Genetic control and allele variation among soybean maturity groups 000 through IX. *Plant Genome*, 2021;*14:*e220146. https://doi.org/10.1002/tpg2.20146