

RESEARCH

Open Access



# Combining GWAS and comparative genomics to fine map candidate genes for days to flowering in mung bean

Kevin O. Chiteri<sup>1</sup>, Ashlyn Rairdin<sup>1</sup>, Kulbir Sandhu<sup>2</sup>, Sven Redsun<sup>3</sup>, Andrew Farmer<sup>3</sup>, Jamie A. O'Rourke<sup>1,4</sup>, Steven B. Cannon<sup>1,4\*</sup> and Arti Singh<sup>1\*</sup>

## Abstract

**Background** Mung bean (*Vigna radiata* (L.) Wilczek), is an important pulse crop in the global south. Early flowering and maturation are advantageous traits for adaptation to northern and southern latitudes. This study investigates the genetic basis of the Days-to-Flowering trait (DTF) in mung bean, combining genome-wide association studies (GWAS) in mung bean and comparisons with orthologous genes involved with control of DTF responses in soybean (*Glycine max* (L) Merr) and Arabidopsis (*Arabidopsis thaliana*).

**Results** The most significant associations for DTF were on mung bean chromosomes 1, 2, and 4. Only the SNPs on chromosomes 1 and 4 were heavily investigated using downstream analysis. The chromosome 1 DTF association is tightly linked with a cluster of locally duplicated *FERONIA* (*FER*) receptor-like protein kinase genes, and the SNP occurs within one of the *FERONIA* genes. In Arabidopsis, an orthologous *FERONIA* gene (*AT3G51550*), has been reported to regulate the expression of the *FLOWERING LOCUS C* (*FLC*). For the chromosome 4 DTF locus, the strongest candidates are *Vradi04g00002773* and *Vradi04g00002778*, orthologous to the Arabidopsis *PhyA* and *PIF3* genes, encoding phytochrome A (a photoreceptor protein sensitive to red to far-red light) and phytochrome-interacting factor 3, respectively. The soybean *PhyA* orthologs include the classical loci *E3* and *E4* (genes *GmPhyA3*, *Glyma.19G224200*, and *GmPhyA2*, *Glyma.20G090000*). The mung bean *PhyA* ortholog has been previously reported as a candidate for DTF in studies conducted in South Korea.

**Conclusion** The top two identified SNPs accounted for a significant proportion (~65%) of the phenotypic variability in mung bean DTF by the six significant SNPs (39.61%), with a broad-sense heritability of 0.93. The strong associations of DTF with genes that have orthologs with analogous functions in soybean and Arabidopsis provide strong circumstantial evidence that these genes are causal for this trait. The three reported loci and candidate genes provide useful targets for marker-assisted breeding in mung beans.

**Keywords** Mung bean, Pulses, Legumes, Genomics, Marker-assisted breeding, Orthologs, Homologs

\*Correspondence:

Steven B. Cannon  
steven.cannon@usda.gov

Arti Singh  
arti@iastate.edu

Full list of author information is available at the end of the article



This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Days to flowering (DTF), an important agronomic trait, marks the transition from vegetative to reproductive development. Flowering timing has a high positive correlation with crop maturity and hence yield [1]. As such, plants have evolved complex regulatory networks to ensure survival across various environmental conditions. Flowering time is regulated by environmental signals, such as temperature and day length, and endogenous pathways, including vernalization and the autonomous floral-promotion pathway [2, 3]. The Arabidopsis flowering pathway involves over 400 genes, some of which can be explored via the Flowering Interactive Database (FLOR-ID, <http://www.flor-id.org/>, accessed February 23, 2023) [4, 5]. Notable among these genes are the *FLOWERING LOCUS C (FLC)*, *FLOWERING LOCUS T (FT)*, *FRIGIDA (FRI)*, and *SUPPRESSOR OF OVER EXPRESSION OF CO1 (SOC1)*. *FLC* operates by suppressing the expression of *FT* and *SOC1*, which promote flowering [3, 5]. In soybean, classical loci with strong effects on DTF include *E1* through *E10* and the JUVENILE (*J*) locus [6–10]. For most of these, the orthologous counterparts have been identified in Arabidopsis [6, 7, 11].

Mung bean (*Vigna radiata* (L.) Wilczek), a warm-season pulse crop, is a valuable source of high-quality proteins, folates, carbohydrates, and vitamins. It contributes to a bal-

anced diet by complementing cereals in South Asia, Africa, and South America [12, 13]. Originally domesticated in India, mung bean has successfully spread to various regions, including the northern and southern latitudes [14]. Selective breeding has resulted in lines adapted to the prolonged summer days and short nights of higher latitudes. In the Midwestern United States, there are several ongoing studies to support the introduction of mung bean as a cash crop through double or multiple cropping systems [12, 15–18]. Varieties exhibiting early flowering and maturation are preferred in such scenarios, making these important traits for mung bean improvement. The present work aims to facilitate selection for these traits through technologies such as marker-assisted breeding and genomic selection [19–23].

$$Y_{ijk} = \mu + (1|\text{accession})_i + \text{loc\_year}_j + (1|\text{accession} : \text{loc\_year})_{ij} + (1|\text{loc\_year} : \text{bloc})_{k(j)} + e_{ijk}$$

In this study, we report association mapping for DTF and evaluate shortlisted candidate genes relative to genes with established functions related to DTF in soybean and Arabidopsis.

In this study, we report association mapping for DTF and evaluate shortlisted candidate genes relative to genes with established functions related to DTF in soybean and Arabidopsis.

## Materials and methods

### Phenotypic data collection

The Iowa Mung bean Diversity (IMD) panel reported by Sandhu and Singh [12] consists of 482 diverse accessions selected from the USDA germplasm collection (~3000

### Descriptive statistics

accessions) and commercial varieties from the World Vegetable Center (<https://avrdc.org/>). DTF data was collected in 2019 and 2021 near Ames, Iowa [12]. The panel was planted at the Iowa State University Agricultural Engineering and Agronomy (AEA) and Bruner farms near Boone, Iowa. A randomized complete block design (RCBD) was used with two replicates at each location. Each single-row plot was 7 ft. long, with 30 inches row-to-row spacing, and 50 seeds were planted per plot spaced 2 inches apart. The date when the first flower was observed in a plot was recorded, and the DTF was calculated from the planting date. This approach has been used in other species [24, 25] and is important here as the mung bean lines are not advanced and still show variation in DTF. We note that the panel was also planted in 2020, but the data from that year were compromised and discarded because of a derecho that swept across the Midwest on August 10. The derecho acted as a stressor leading to premature flower-drops, followed by a second flowering flush.

Analysis was conducted on 478 accessions planted in both 2019 and 2021. The mixed model below was used for the calculation of the best linear unbiased predictors (BLUPs) within the “inti” R package [26]:

where  $Y_{ijk}$  is the phenotypic value (DTF) of the  $i^{\text{th}}$  genotype in the  $j^{\text{th}}$  environment (location\*year combination) and  $k^{\text{th}}$  block,  $\mu$  is the overall mean,  $(1|\text{accession})_i$  is the random effect due to the  $i^{\text{th}}$  genotype,  $\text{loc\_year}_j$  is the fixed effect due to the  $j^{\text{th}}$  environment,  $(1|\text{Accession} : \text{loc\_year})_{ij}$  is the random effect due to the  $i^{\text{th}}$  genotype \*  $j^{\text{th}}$  environment interaction,  $(1|\text{loc\_year} : \text{bloc})_{k(j)}$  is the random effect due to the  $k^{\text{th}}$  block nested within the  $j^{\text{th}}$  environment and  $e_{ijk}$  is the random error following  $N(0, \sigma e^2)$ . Broad-sense heritability was estimated as described by Cullies et al. [27]. Variance components and BLUPs were estimated. Analysis was conducted in the open-source R statistical computing environment [28].

### Single nucleotide polymorphism (SNP) coordinate update

The recently published mung bean reference genome version 7 for accession VC1973A [29] was used for positional analysis. To take advantage of this new genomic data, we projected the SNPs coordinates onto assembly version 7 from SNPs identified previously using genotyping-by-sequencing (GBS) [12] with coordinates relative to reference genome version 6 [30]. To project the sequences from assembly version

6 to version 7, the flanking 1001 bases around version 6 SNP coordinates were first extracted, (from 500 bases before the SNP to 500 bases after it). Retrieval of the flanking sequences was accomplished by deriving a four-column BED file with the molecule ID, the start and end coordinates, and the marker name. The sequence described by the BED file was extracted using the bedtools “getfasta” utility. The flanking sequences from the assembly version 6 were then used as queries for a blastn search, with parameters “-evalue 1e-10 -outfmt 7 -perc\_identity 99”. This tabular BLAST output was then filtered (using a simple awk script) to select the top match among those BLAST hits with 99% identity—with the further requirement that the match length be at least 990 bases. Of the initial 26,550 SNPs, 23,590 remained for downstream analysis after filtering out those with minor allele frequencies < 0.01 and > 15% missing data. Reference genome version 7, annotations, markers, and variant data can be found at the Legume Information System (LIS) Datastore (<https://data.legumeinfo.org/Vigna/radiata/>).

#### Genome-wide association study (GWAS)

Genome-wide association mapping for DTF was conducted using both single loci mixed-linear model (MLM) [31] and multi-locus FarmCPU [32]. Both methods were implemented within the Genome Analysis and Prediction Integrated Tool (GAPIT) [33]. To control population structure, both the principal components (set PCA.total=3) and the default kinship matrix were generated within GAPIT and used as covariates in the model. From the GWAS results, SNPs on chromosomes were visualized on Manhattan and Q-Q plots using CMplot [34] and qqman [35] packages. A significant threshold of  $-\log_{10}(1e-5)$  with a Bonferroni correction was used.

#### Candidate gene identification

The two most statistically significant SNPs were identified for significant downstream analysis due to the fact their *p*-values were  $\sim 3\times$  those of the next four significant SNPs while accounting for  $\sim 65\%$  of the phenotypic variance explained (PVE) by the significant SNPs (39.61%). We evaluated a 290 kb region on either side of each of the top two SNPs (580 kb total per locus) (Additional file 1 and 2 for scripts used and Additional file 3 and 4 for genes and fasta files). The 290 kb distance was selected as this is the average linkage disequilibrium (LD) decay within the IMD population [12]. A similar method had previously been applied in soybeans to find genes associated with iron deficiency chlorosis [36] as well as in mung bean to scan a 1 Mb region flanking soybean flowering genes [11].

BLASTP [37] was used to query the mung bean coding sequences identified above against all proteins in

Arabidopsis (Araport11, [38]), using an E-value cutoff of  $1e-10$ . Matches were evaluated for functional descriptions associated with flowering time or maturity and publications in the TAIR database related to these traits. Additionally, the conserved domain search function at NCBI was used to identify domains present in Arabidopsis and mung bean proteins. Identified domains were visualized using PROSITE (<https://prosite.expasy.org/>) [39].

#### Comparative genomics candidate gene identification

For comparative genomic analyses, we used the following tools available at the Legume Information System (LIS; legumeinfo.org) [40, 41]: ZZBrowse, Genome Context Viewer (GCV), JBrowse, Gene Family Search, Funannotate, and Phylogenetic tree viewers [42, 43]. The SoyBase (<https://www.soybase.org/>) “Convert Gene Model Names” tool [44, 45] was used to determine correspondences between different versions of soybean gene models.

#### Gene family and phylogenetic analysis

Gene families for the *FERONIA*, *PhyA*, and *PIF3* families were calculated using the Pandagma workflow [46], on the set of all predicted protein sequences for the included species: *Vigna radiata* (VC1973 genome 7 annotation 1), *Vigna unguiculata* (IT97K-499–35 genome 1 annotation 2), *Glycine max* (Williams 82 genome 4 annotation 1), *Medicago truncatula* (A17\_HM341 genome 4 annotation 1), *Lupinus albus* genome 1 annotation 1, *Cicer arietinum* (CDC Frontier genome 3 annotation 1), *Lotus japonicus* (MG20 genome 3 annotation 1), and *Arabidopsis thaliana* (Col0 genome 9 annotation 11). The source for all these annotations was legumeinfo.org (<https://data.legumeinfo.org> and <https://data.legumeinfo.org/annex/>; Redsun et al. 2022); metadata for each is in the respective collection folders in the annotation for each (Supplementary Table 1). The gene family construction approach uses MMSeqs2 [47] to identify pairwise matches between each protein set; then DAGChainer [48] to filter the matches by synteny. For those filtered matches, further filtering is done based on synonymous-site changes (Ks values) that are calculated in in-frame aligned coding sequence, using the PAML package [49]. The filtered results are then clustered with Markov clustering (mcl; [50] to generate provisional gene families. Sequences without family placements to this point are then compared against the provisional gene families and added to the top-matching family if homology thresholds are met (protein identity of at least 30% and alignment coverage of at least 40%). Protein sequences in each gene family were aligned using FAMSA [51]. From these alignments, hidden Markov models (HMMs) were calculated using the HMMER4 package [52]; hmmer.org). The

family sequences were then realigned to the HMMs, and match-state characters were removed prior to calculating gene phylogenies. Finally, phylogenetic trees were calculated from cleaned alignments using FastTree v2.1 [53].

**Results**

**Descriptive statistics and variance components**

DTF in mung bean exhibited an approximately symmetrical distribution of the measured value, with a mean and median of 49 days, a standard deviation of 5 days, and a coefficient of variation of 10% (Fig. 1). As shown in Table 1, the accession effect accounted for the highest variation at 71.4%, followed by the effect of the interaction between accession and location-year at 13.8%, while the location-year/block interaction effect accounted for 1.2% (Table 1). DTF exhibited a high broad-sense heritability of 0.93.

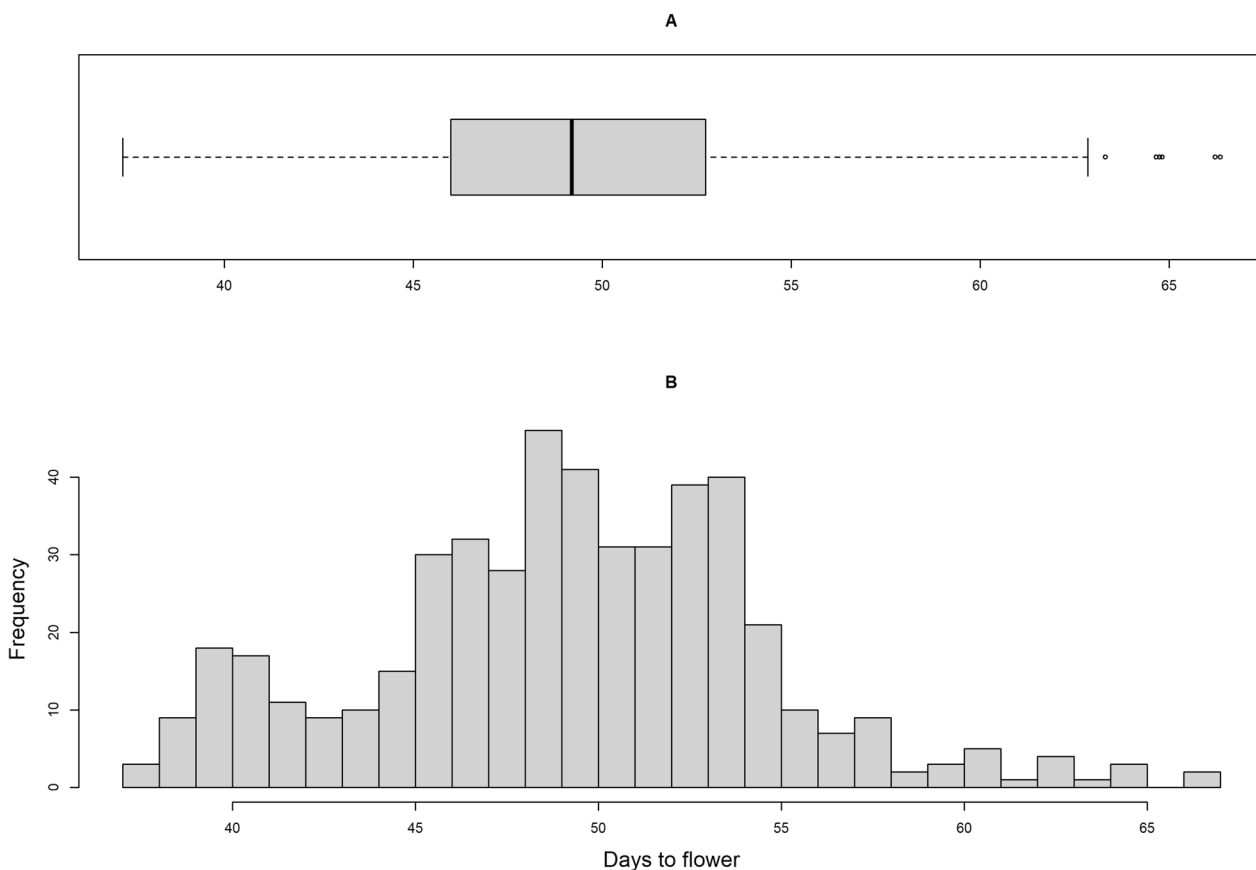
**SNP coordinate projection between assemblies**

Mapping the IMD panel SNPs from genome assembly version 6 to version 7 substantially rearranged the SNP locations, due to large structural changes between

these two assemblies (Fig. 2). For example, from Fig. 2, it is apparent that the lengths of chromosomes 1, 2, 3, 4 and 9 increased in assembly 7, while those of chromosomes 5, 6, 7, 8, 10 and 11 decreased. Ha et al. [29] reported these rearrangements between these two assembly versions.

**GWAS results**

MLM and FarmCPU methods resulted in similar significant SNPs; hence, only the results from FarmCPU are reported here due to the strength of multi-locus models [32]. Six SNPs associated with DTF were observed with *p*-values below 1e-5 (Fig. 3). The six SNPs were 1\_11367629 (Chr 1), 8\_3586220 (Chr 2), 5\_4604047 (Chr 2), 5\_430302 (Chr 2), 2\_10755945 (Chr 4), and 3\_8727529 (Chr 4) (Fig. 3 and Table 2). We note that the SNP names were assigned relative to assembly version 6 and should not be construed to have meaning relative to assembly 7. For example, SNP 1\_11367629 remained on chromosome 1 in assembly 7, while SNPs 2\_10755945 and 3\_8727529 were relocated to chromosomes 4. As shown in Table 2, the SNPs 3\_8727529

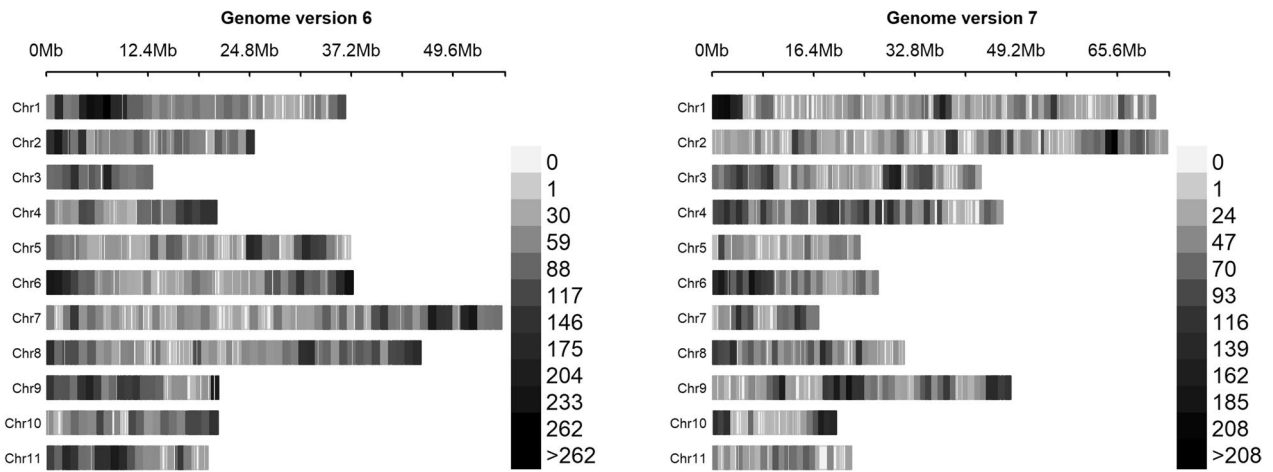


**Fig. 1** Boxplot (A) and histogram (B) illustrating the distribution of mung bean days to flowering (DTF) BLUPs

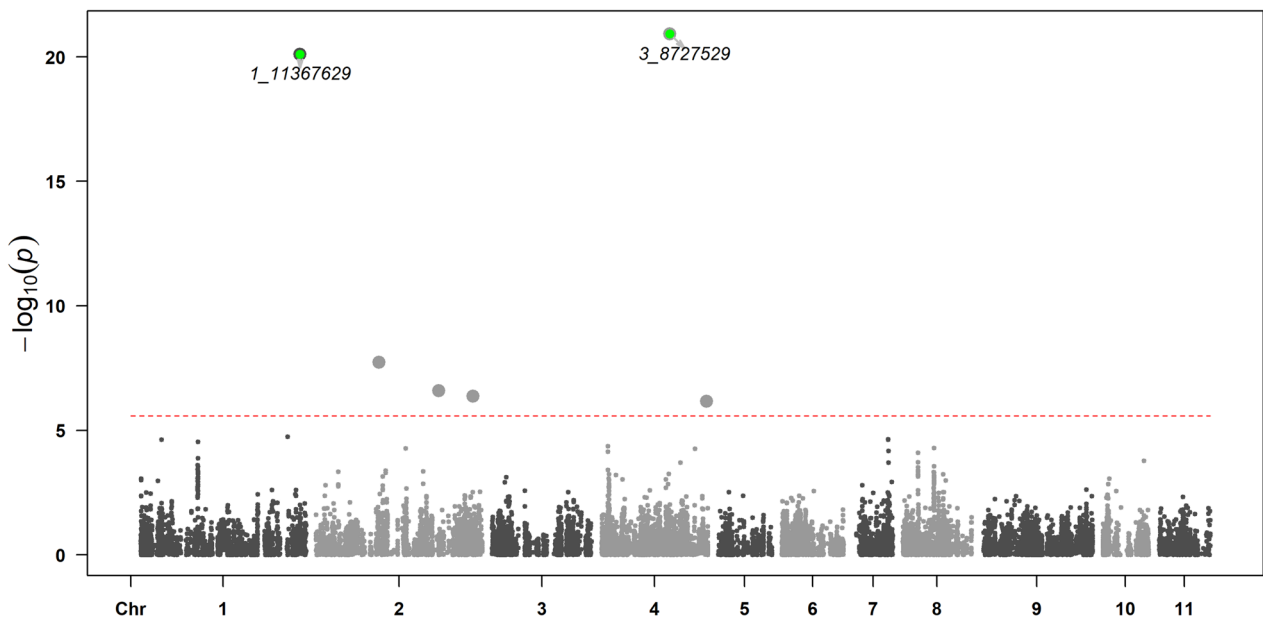
**Table 1** Variance components from the linear mixed model used in this study

Source	Variance	Percent Variation
Accession	30.93	71.4
Accession: location year	5.99	13.8
Location year: bloc	0.5	1.2
Residual	5.92	13.7
Totals	43.34	100.1

and 1\_11367629 accounted for 13.72% and 11.84% of the phenotypic variance, respectively. In total, they accounted for 25.56% of the phenotypic variance explained (PVE) for DTF. The remaining four statistically significant SNPs only accounted for an additional 14.05% of the PVE and were thus not considered in the remaining analyses (Table 2). Our results validate previous reports of significant SNP associated with DTF in mung bean [12], with an additional year of data.



**Fig. 2** Iowa Mung bean Diversity panel SNP density plots after updating the SNP coordinates from the old genome version 6 (left) to the new genome version 7 (right) within a 1 Mb window size. X-axis shows the chromosome length, while the scale indicates SNP densities



**Fig. 3** Manhattan plot showing significant SNPs associated with DTF. SNPs are labeled with the SNP name. SNPs in green are discussed further in the text. Trait-SNP associations were performed using FarmCPU in GAPIT. The horizontal dotted red line represents the Bonferroni correction at  $p=0.05$ . See the accompanying Q-Q plot in Additional Fig. 1

**Table 2** The most significant SNPs from GWAS GAPIT results

SNP	Chromosome (chr)	Position (bp)	P-value	Phenotypic Variance Explained % (PVE)
3_8727529	4	29,758,658	1.21E-21	13.72
1_11367629	1	69,228,460	7.90E-21	11.84
8_3586220	2	53,453,008	2.55E-07	5.27
5_4604047	2	27,377,521	1.83E-08	4.05
5_430302	2	68,312,851	4.27E-07	3.52
2_10755945	4	45,689,185	6.75E-07	1.21
		<b>Total</b>		39.61

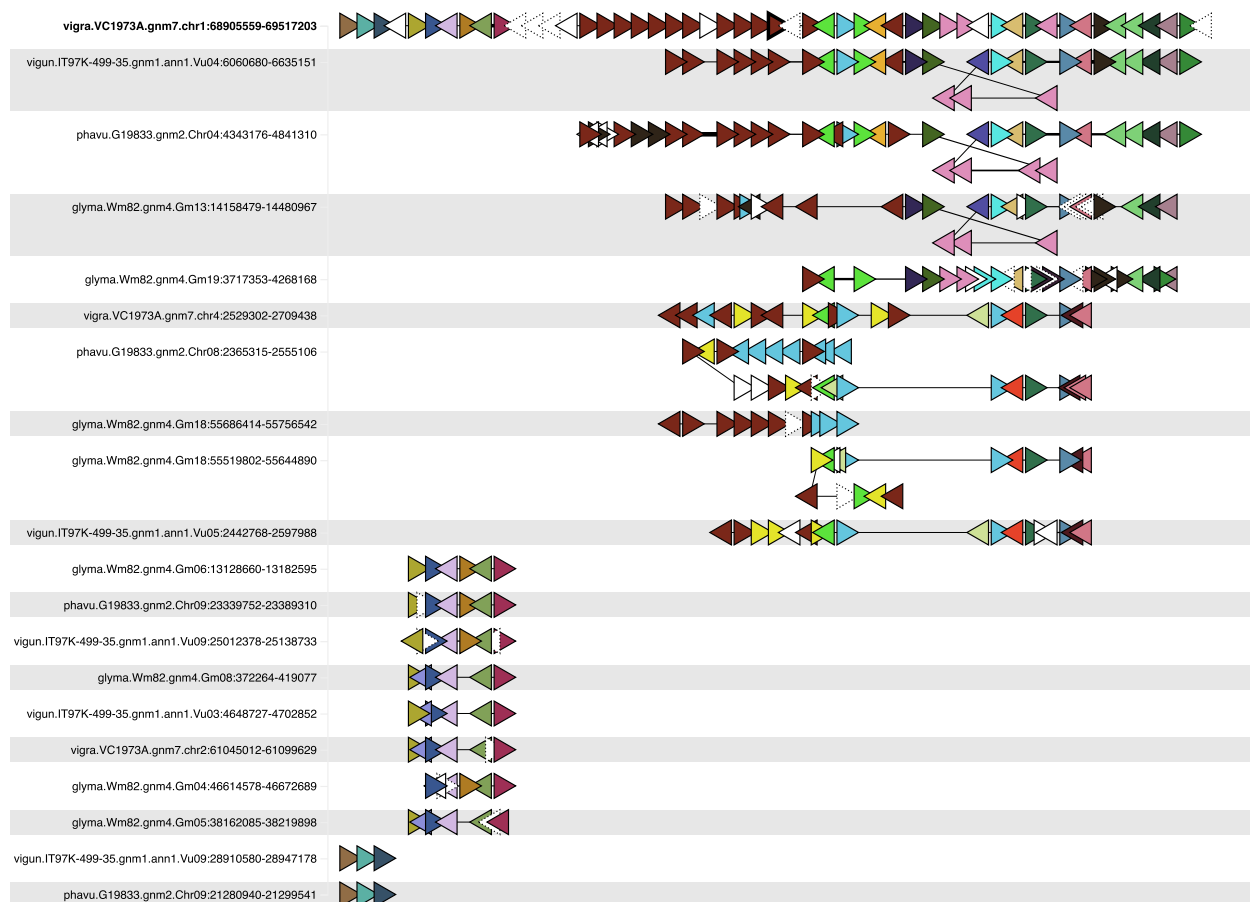
### Candidate gene evaluation

The 580 kb region associated with SNP 1\_11367629 (on chr 1) contained 39 genes, while the 580 kb region associated with SNP 3\_8727529 (on chr 4) contained 65 genes. For each of these 104 genes, the top Arabidopsis match was identified (<https://arabidopsis.org/>, accessed February 23, 2023) and checked for publications associated with flowering in Arabidopsis (Table 3). This search yielded 18 mung bean genes with such homologues (Table 3). Gene family and phylogenetic analyses were also conducted for the strongest candidate genes (Figs. 6 and 8).

For SNP 1\_11367629 on chromosome 1, the SNP occurs within a cluster of 13 orthologs of the Arabidopsis gene *FERONIA* (*FER*) (Fig. 4). These *FER* orthologs were *Vradi01g00003484-Vradi01g00003490* (seven genes), *Vradi01g00003492-Vradi01g00003495* (4 genes), *Vradi01g00003497*, and *Vradi01g00003502*. The SNP occurs within the single exon of *FER* gene *Vradi01g00003495*. The C/T allele in *Vradi01g00003495* is a missense mutation, with the observed C and T variants coding for histidine 82 (CAC) and tyrosine (TAC), respectively. The conserved domain analyses revealed that only proteins encoded by the genes *Vradi01g00003488*, *Vradi01g00003492*, and *Vradi01g00003502* share the same two conserved domains with the Arabidopsis *FER* protein (AT3G51550) (Fig. 5). However, gene family and phylogenetic analyses (Fig. 6), based primarily on sequence from the Malectin domain (IPR024788), show the *FER* gene family to be large and complex, with expansions and contractions of clusters – as evident in the cluster of *FER* homologs on mung bean chromosome 1. Similarly, in Arabidopsis, there are five *FER* paralogs in two clusters – one on chromosome 3 and one on chromosome 5. From the perspective of that domain, all 15 *FER* genes in *V. radiata* are orthologous to the five *FER* genes in *A. thaliana*, likely deriving from a common shared ancestral gene.

**Table 3** Mung bean genes orthologous to Arabidopsis genes involved in flowering. Genes within the 580 kb region surrounding the two most significant SNPs with homology (BLASTP, E-value 1e-10) to candidate genes from Araport11. Papers supporting the role of the identified protein in DTF processes are listed in the Supporting Paper column

SNP	Mung bean Gene ID	Araport11	E-value	% ID	Protein	Supporting Paper
	<i>Vradi01g00003484</i>	<i>AT3G51550</i>	1.74E-180	45.97	FERONIA	[54]
	<i>Vradi01g00003485</i>	<i>AT3G51550</i>	7.63E-14	60.00	FERONIA	[54]
	<i>Vradi01g00003486</i>	<i>AT3G51550</i>	1.61E-68	38.36	FERONIA	[54]
	<i>Vradi01g00003487</i>	<i>AT3G51550</i>	4.48E-88	51.17	FERONIA	[54]
	<i>Vradi01g00003488</i>	<i>AT3G51550</i>	0	46.40	FERONIA	[54]
	<i>Vradi01g00003489</i>	<i>AT3G51550</i>	0		FERONIA	[54]
	<i>Vradi01g00003490</i>	<i>AT3G51550</i>	3.05E-159	51.71	FERONIA	[54]
	<i>Vradi01g00003492</i>	<i>AT3G51550</i>	0	43.72	FERONIA	[54]
	<i>Vradi01g00003493</i>	<i>AT3G51550</i>	2.53E-47	41.26	FERONIA	[54]
	<i>Vradi01g00003494</i>	<i>AT3G51550</i>	3.79E-97	39.70	FERONIA	[54]
	<i>Vradi01g00003495</i>	<i>AT3G51550</i>	1.18E-120	57.84	FERONIA	[54]
	<i>Vradi01g00003497</i>	<i>AT3G51550</i>	6.75E-129	43.77	FERONIA	[54]
	<i>Vradi01g00003502</i>	<i>AT3G51550</i>	0	50.30	FERONIA	[54]
	<i>Vradi04g00002773</i>	<i>AT1G09570</i>	0	74.45	PhyA	[55]
	<i>Vradi04g00002778</i>	<i>AT1G09530</i>	1.12E-57	34.60	PIF3	[9]
	<i>Vradi04g00002804</i>	<i>AT1G61040</i>	2.21E-26	65.15	VIP5	[3]
	<i>Vradi04g00002805</i>	<i>AT1G61040</i>	7.46E-175	56.66	VIP5	[3]
	<i>Vradi04g00002812</i>	<i>AT1G23380</i>	4.31E-86	53.64	KNAT6	[56]



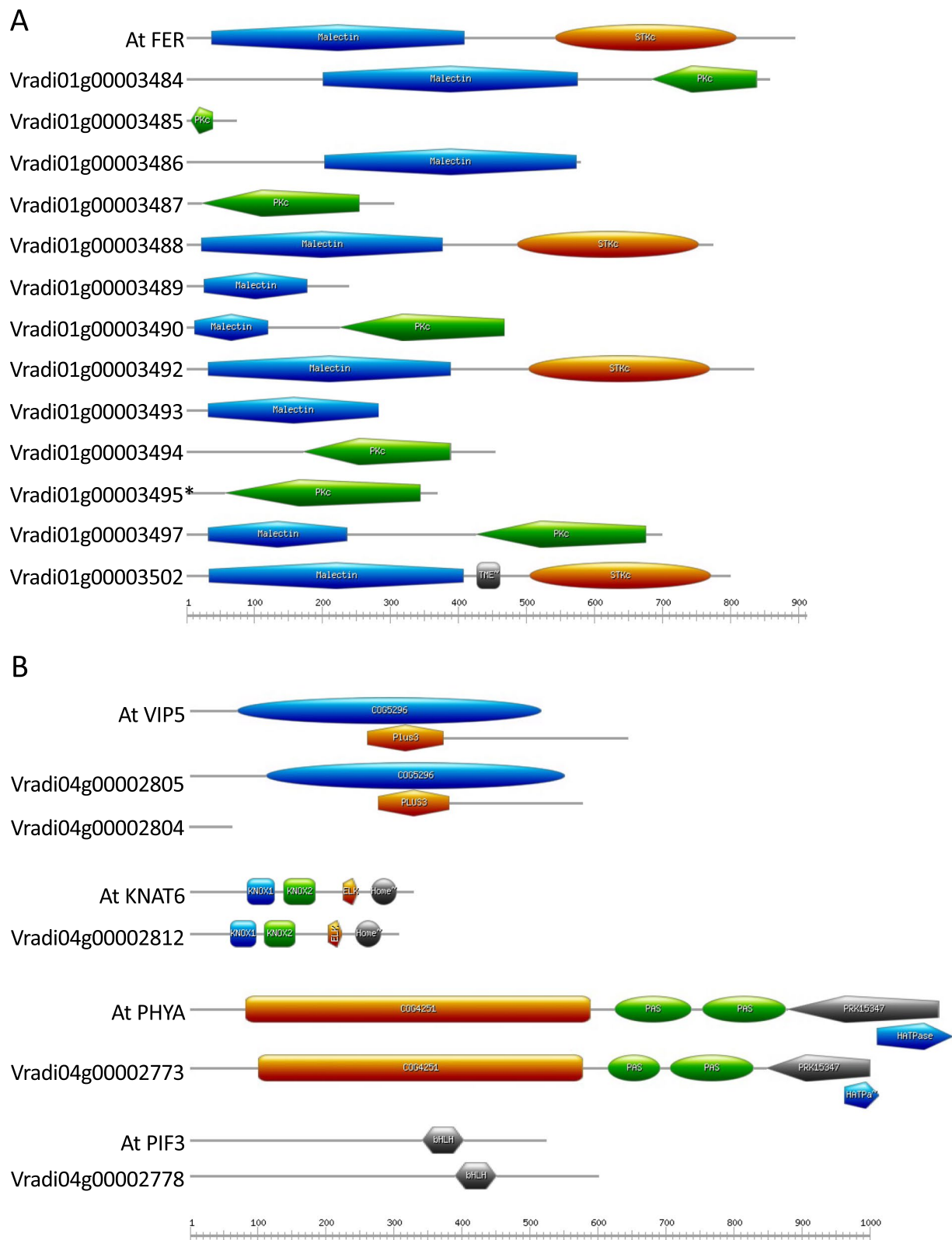
**Fig. 4** The region around SNP 1\_11367629, showing *FERONIA* gene cluster. *FERONIA* genes are shown as brown triangles near the center. The gene containing the SNP with the strongest DTF association (*Vradi01g00003495*) is in the top row, center, with a black border. Genes from other gene families are indicated by other colors (one color per gene family). The triangle direction indicates gene orientation. Singleton genes are shown in white (dotted lines further indicating orphan genes, without gene family orthology). Crossed diagonal lines, for example, in the second row, indicate inversions. The text on the left indicates genus and species: *vigra* = *Vigna radiata*; *vigun* = *Vigna unguiculata*; *phavu* = *Phaseolus vulgaris*; *glyma* = *Glycine max*. Chromosomes and region coordinates are indicated on the left portion of the strings

For SNP 3\_8727529 on chromosome 4, several genes in the 580 kb flanking region have orthologs with roles related to flowering in Arabidopsis, including *Vradi04g00002773*, *Vradi04g00002778*, *Vradi04g00002804*, *Vradi04g00002805*, *Vradi04g00002812* (Fig. 7). These correspond to Arabidopsis genes phytochrome A (*PhyA*), phytochrome-interacting factor 3 (*PIF3*), vernalization independence 5 (*VIP5*, two copies), and knotted1-like from *Arabidopsis thaliana* 6 (*KNAT6*),

respectively. The conserved domain analysis confirmed the mung bean genes perfectly reflected the domains in the Arabidopsis homologs (Fig. 5). Based on proximity to the DTF-associated marker on chromosome 4 and functional studies in Arabidopsis, both *PhyA* (*Vradi04g00002773*) and *PIF3* (*Vradi04g00002778*) are strong candidates for the causal genes. The gene family phylogenies for *PhyA* and *PIF3* are shown in Fig. 8, and synteny among legume genomes in the vicinity of

(See figure on next page.)

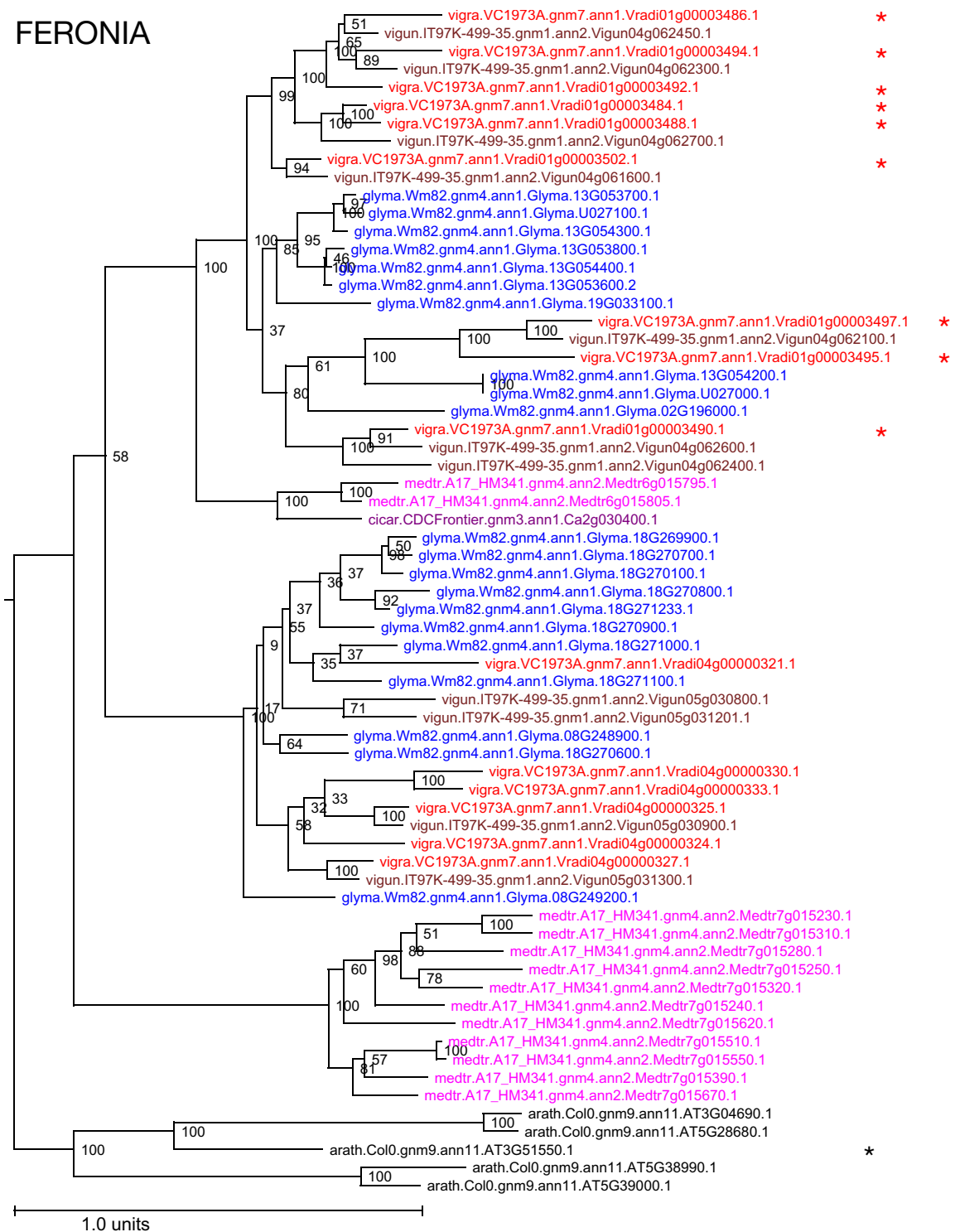
**Fig. 5** Domains of the primary candidate gene families: *FERONIA*, *PhyA*, and *PIF3*. **A:** Conserved domain analysis of Arabidopsis *Feronia* (*AtFER*) and 13 mung bean genes within the QTL that are identified as *FER* using homology and gene family analyses. Conserved domains are color and shape coded. The SNP was identified within the coding region of *Vradi01g00003495*, denoted with an \*. This analysis supports *Vradi01g00003488*, *Vradi01g00003492*, and possibly *Vradi01g00003502* as the most likely candidate genes. **B:** Conserved domain analysis of Arabidopsis candidate genes and the mung bean homologs. Note, no conserved domains were identified in the *Vradi04g00002804* sequence. Other than that gene, all domains identified in Arabidopsis, were identified in the mung bean homologs



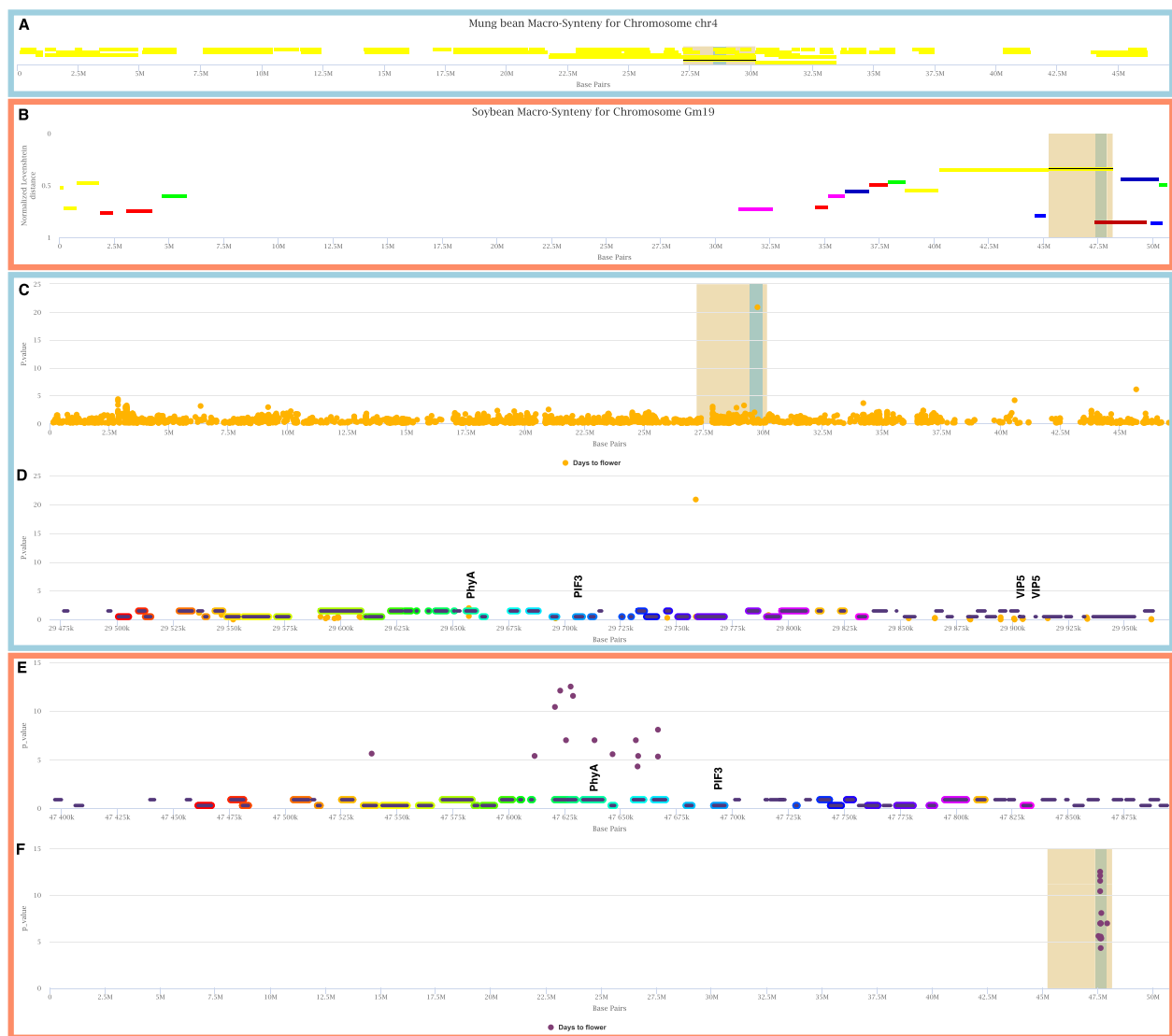
**Fig. 5** (See legend on previous page.)



# FERONIA



**Fig. 6** FERONIA gene family for selected legume species. *Arabidopsis* orthologs are shown in gray at the bottom, including the classical FER gene (AT3G51550), marked with a black asterisk. Each species is represented by a color and is indicated by the leading five-letter prefix in the gene name: arath = *Arabidopsis thaliana*; cicar = *Cicer arietinum*; glyma = *Glycine max*; medtr = *Medicago truncatula*; vigun = *Vigna unguiculata*; vigra = *Vigna radiata*; lupal = *Lupinus albus*. The *V. radiata* paralogs in the cluster on chromosome 1 are marked with red asterisks



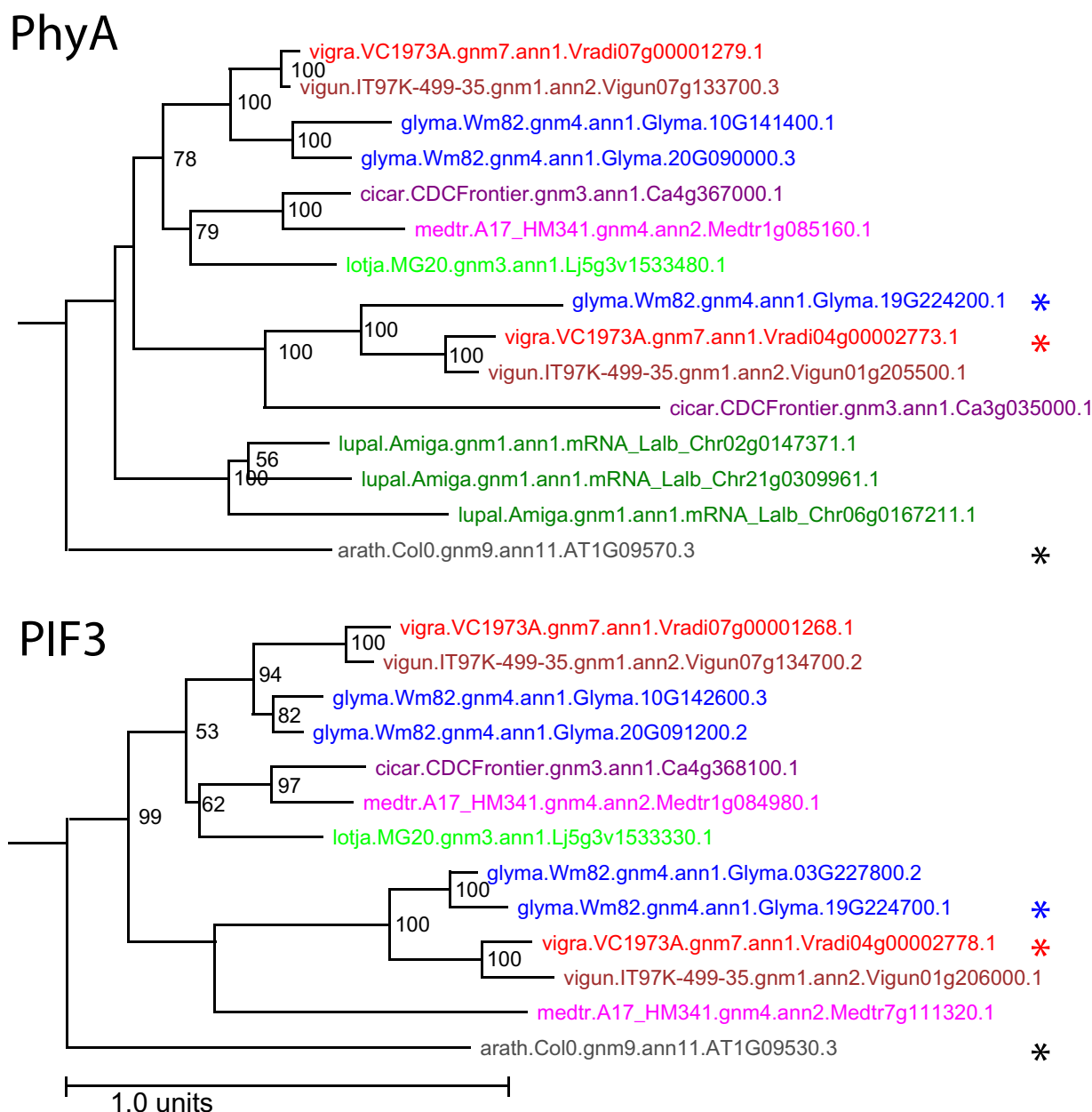
**Fig. 7** Comparison of syntenic regions for Days to Flowering loci in mung bean and soybean. **A.** Mung bean chromosome 4. Regions in synteny analysis are in yellow. A light blue slice near 30 Mb is highlighted in synteny search in soybean in next panel. **B.** Soybean chromosome 19. Colored horizontal regions indicate synteny with mung bean chromosomes. Light blue slice near 47 Mb is orthologous to the slice highlighted in A. **C.** Mung bean markers from GWAS. A blue slice around significant SNP near 30 Mb is shown in the inset in the next panel. **D.** Zoomed region of 250 kb corresponding with a slice in C. Gene models are shown at the bottom of panel. Rainbow colors show orthology with soybean genes in the next panel. Genes with annotations related to flowering are highlighted above each gene: *PhyA*, *PIF3*, *VIP5*. **E.** Soybean region from chromosome 19 corresponds via synteny with regions above from mung bean chromosome 4. The figure is adapted from ZZBrowse [41]

chromosome 4 *PhyA* are shown in Fig. 9. The mung bean *PhyA* and *PIF3* orthologs are located at 29.66 and 29.70 Mb relative to SNP 3\_8727529 at 29.75 Mb, in the 71.9 Mb chromosome. While the two *VIP5* genes and *KNAT6* are further from the SNP at 29.90 Mb (both *VIP5* genes) and 29.97 Mb relative to chromosome 4, they may play a role in mung bean DTF and contribute to the significance of SNP 3\_8727529. In contrast, the ortholog to the gene *GmDt1/ GmTFL1b (Vra-di04g00002442* in mung bean, *Glyma.19G194300* in

soybean), associated with determinate growth in soybean [7], is on the same chromosome as the mung bean chromosome 4 DTF locus but is separated from it by 2.5 Mb, so we consider this an unlikely candidate.

## Discussion

Mung bean shares many of the nutritional and agronomic advantages of other legume crops and is relatively drought- and heat-tolerant, making it an increasingly important crop in the context of climate change [57–59].



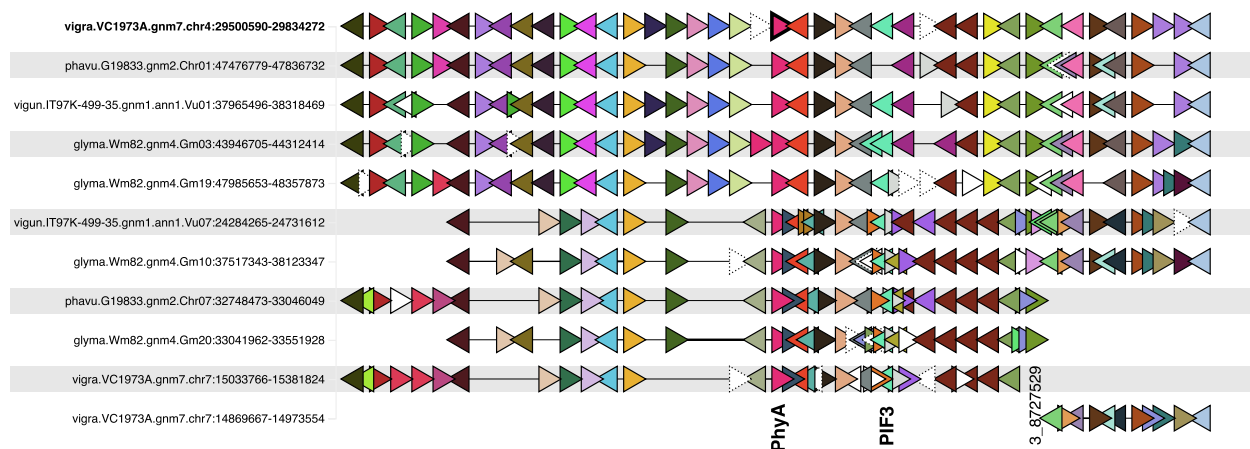
**Fig. 8** *PhyA* and *PIF3* gene families for selected legume species. *Arabidopsis* orthologs are shown in gray at the bottom, including the classical genes marked with a black asterisk: *PhyA*, (AT1G09570; top); and *PIF3* (AT1G09530; bottom). Orthologs in *V. radiata* closely linked to the DTF trait are marked with a red asterisk. Each species is represented by a color and is indicated by the leading five-letter prefix in the gene name: arath = *Arabidopsis thaliana*; cicar = *Cicer arietinum*; glyma = *Glycine max*; lotja = *Lotus japonicus*; medtr = *Medicago truncatula*; vigun = *Vigna unguiculata*; vigra = *Vigna radiata*; lupal = *Lupinus albus*

Interest in mung bean can also be attributed partly to changing consumer habits in the West, associated with reducing meat as a primary protein source [60, 61].

The high heritability score (0.93) of DTF in mung bean makes it a good candidate for genomic-assisted breeding. Combining GWAS results with comparative methods, we have identified strong DTF candidate genes at two loci,

orthologous to genes associated with DTF from soybean and *Arabidopsis* (Table 3).

The previous reference mung bean genome version 6 needed to be structurally corrected, as reported by Ha et al. [29]. Previous work done by our group identified two of the same SNPs (1\_11367629 and 3\_8727529) associated with DTF in mung bean [12], but errors in



**Fig. 9** The region around SNP 3\_8727529, showing candidate genes *PhyA* and *PIF3* and surrounding genes. The *PhyA* orthologs are shown in pink and *PIF3* in turquoise and labeled at the bottom (along with the approximate location of the DTF-associated SNP). The mung bean *PhyA* gene (*Vradi04g00002773*) is in the top row, center, with a black border. Genes from other gene families are indicated by other colors (one color per gene family). The triangle direction indicates gene orientation. Singleton genes are shown in white (dotted lines further indicating orphan genes, without gene family orthology). The text on the left indicates genus and species: *vigna* = *Vigna radiata*; *vigun* = *Vigna unguiculata*; *phavu* = *Phaseolus vulgaris*; *glyma* = *Glycine max*. Chromosomes and region coordinates are indicated on the left portion of the strings

version 6 confounded marker-trait associations. The SNP coordinates have been updated in the genome version 7 sequence, with SNP 3\_8727529 now located on chromosome 4.

For the chromosome 1 DTF locus associated with SNP 1\_11367629, we identify the cluster of 13 mung bean homologs of the Arabidopsis receptor-like protein kinase *FERONIA* (*FER*) gene (*AT3G51550.1*), flanking SNP 1\_11367629, as the best candidates. In version 7 of the genome SNP 1\_11367629 is located within the single exon of *Vradi01g00003495*. It is intriguing that the C/T allele in *Vradi01g00003495* is a missense mutation (histidine-to-tyrosine); however, while it is possible that this mutation is responsible for a changed phenotype, it is also likely that there are structural or copy-number variations at this complex locus. Domain analysis of the 13 *FER* genes identified in mung bean and the *AtFER* reveal only three of the mung bean genes (*Vradi01g00003488*, *Vradi01g00003492* and *Vradi00003502*, but not *Vradi01g00003495*) share the same two conserved protein domains identified in Arabidopsis (Fig. 5). It is notable that *Vradi01g00003502* encodes a third domain, a transmembrane epidermal growth factor receptor-like domain. The two conserved domains are a Malectin domain and a serine/threonine kinase catalytic (STKc) domain. Two of the mung bean *FER* genes encode a protein kinase catalytic (PKc) domain instead of the serine/threonine kinase catalytic (STKc) domain present in Arabidopsis. Both are protein kinase domains, but the STKc domain specifically phosphorylates the hydroxyl (OH) group of serine or threonine while the PKc domain

is not restricted to specific amino acids. The remaining seven genes only encode either the Malectin domain or the PKc domain. *FER* has been associated with various aspects of plant growth, including flowering, root hair development, and hypocotyl and root elongation. In Arabidopsis, *FER* has been shown to regulate flowering time by up-regulating the expression of *FLOWERING LOCUS C* (*FLC*) [12, 54, 62]. Since *FER* exists as a cluster of locally duplicated genes, we speculate that copy number variation in this locus may affect the maturity continuum in mung beans as reflected by the days to flowering. Given the prevalence of the Malectin and PKc domain encoding proteins in this region, we hypothesize these genes may not be major DTF genes like *FER* but may serve to micro-regulate DTF.

The *FERONIA* gene phylogeny (Fig. 6) and synteny depiction (Fig. 4) both bear indications of actively evolving clusters in the included species. Genes in the *V. radiata* clusters on chromosomes 4 and 1 are interleaved with syntenic clusters on *V. unguiculata* chromosomes 4 and 5 respectively, and *G. max* chromosomes 13, 2, 18, and 8. The interleaving in the phylogeny implies that the clusters on *V. radiata* chromosomes 1 and 4 are at least as old as their common ancestor (the warm-season legume ancestor). On the other hand, one of the two post-papilionoid WGD clusters in *Medicago* has expanded while the other has not.

Finally, for SNP 3\_8727529 on chromosome 4, there were several orthologs to Arabidopsis genes with roles associated with the control of flowering, including *PIF3* (phytochrome interacting factor 3), *VIP5* (vernalization

independence 5), *KNAT6*, which influences inflorescence architecture, and *PhyA* (phytochrome A; *AT1G09570*). Of these, we consider the strongest candidate to be the *Vigna PhyA* ortholog (*Vradi04g00002773*) due to its proximity to the SNP (98.5 kb) and the strong syntenic relationship with the soybean orthologs, *E3/GmPhyA3* and *E4/GmPhyA2*. In soybean, the paralogs *E3* and *E4* (derived from the *Glycine* whole-genome duplication) suppress flowering but do so in response to varying light qualities [55]. In soybean, *E3* shows the strongest GWAS signal for DTF in soybean (Kim et al., 2020) (Fig. 8). This gene also aligns with the conclusions of Hwang et al. [63], and Ha et al., [29]. Both of these publications identified *Vradi04g00002773* as the best candidate for several mung bean QTLs including flower initiation (F14-1), number of nodes (Node4-1), synchronous pod maturity (SPM4-1) and plant height [29]. These findings support our previous findings performed in the United States and those from South Korea. Despite the importance of *PhyA*, we would be remiss not to consider the other candidate genes located within this genomic region. While *PhyA* may be the major gene associated with SNP 3\_8727529, the other genes associated with DTF in this genomic region may also play supporting roles in regulating mung bean DTF. In Arabidopsis, silencing *PIF3* results in early flowering by increasing the levels of *FT* [64]. Similarly, silencing any of the *VIP* genes causes silencing of *FLC*, resulting in early flowering [3, 65]. Given the role of these genes in Arabidopsis, it's plausible that they also contribute to the DTF phenotype measured in this study, despite being further from the SNP of interest. Previous studies in soybean and other species have confirmed that multiple genes may be involved in conferring a QTL, suggesting the same would be true for mung bean.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10156-x>.

Supplementary Material 1.  
Supplementary Material 2.  
Supplementary Material 3.  
Supplementary Material 4.  
Supplementary Material 5.  
Supplementary Material 6.  
Supplementary Material 7.

## Acknowledgements

The authors thank Jacqueline Campbell and Michael Morrisey for their insights into mining gene annotations and Wei Huang for assistance in preparing JBrowse views. We are grateful to Alan Cleary for the Genome Context Viewer tool. The authors thank the statistician, Dr. Somak Dutta, for helping with accurate model selection for calculating BLUPs. We thank the staff and graduate students in the Singh-Singh labs for their support during this project. We are grateful to the anonymous reviewers.

## Plant material identification

AS and KOC identified the plant material used in the study from a collection obtained from the USDA gene bank as described in the methods section. In brief, accessions were obtained from the USDA Plant Genetic Resource Conservation Unit in Griffin, GA in 2017 [12]. Permission and license were obtained using the request protocol outlined by the gene bank (<https://npgsweb.ars-grin.gov/gringlobal/search>). The materials in this study have not been deposited in any public available herbarium.

## Authors' contribution

KOC, SBC, and AS envisioned and designed the study. SR and AF guided the use of the tools at LIS used in this study. Data from AR and KS provided the foundational data for combined analysis and was used in running GWAS. JAO was instrumental in comparative genomics and candidate gene identification. KOC wrote the first manuscript draft. All the authors revised and approved the manuscript.

## Funding

This work was supported with funding from the United States Department of Agriculture–National Institute of Food and Agriculture (USDA-NIFA) Mung bean breeding #2022–67013–37120 (AS &SBC), the RF Baker Center for Plant Breeding (AS), and by the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) CRIS Projects 5030–21000–071–000D (SBC) and 3625–21220–007–000D (JAO). Support from USDA Hatch CRIS project CRIS Project IOW04714 (AS) and USDA-ARS Non-Assistance Cooperative Agreement #58–5030–2–036 (SR & AF) is sincerely appreciated. The USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

## Availability of data and materials

The dataset(s) supporting the conclusions of this article is (are) included within the article (and its additional file(s)).

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors consent to publication.

### Competing interests

The authors declare no competing interests.

## Author details

<sup>1</sup>Department of Agronomy, Iowa State University, Ames, IA, United States.

<sup>2</sup>Agriculture and Agri-Food, Morden, MB, Canada. <sup>3</sup>National Center for Genome Resources, Santa Fe, NM 87505, United States. <sup>4</sup>USDA - Agricultural Research Service, Corn Insects, and Crop Genetics Research Unit, Ames, IA, United States.

Received: 20 July 2023 Accepted: 22 February 2024

Published online: 12 March 2024

## References

1. Dietz N, Chan YO, Scaboo A, Graef G, Hyten D, Happ M, et al. Candidate genes modulating reproductive timing in elite US soybean lines identified in soybean alleles of Arabidopsis flowering orthologs with divergent latitude distribution. *Front Plant Sci.* 2022;13: 889066. <https://doi.org/10.3389/fpls.2022.889066>.
2. Amasino RM, Michaels SD. The timing of flowering. *Plant Physiol.* 2010;154:516–20. <https://doi.org/10.1104/pp.110.161653>.
3. Nasim Z, Susila H, Jin S, Youn G, Ahn JH. Polymerase II-associated factor 1 complex-regulated FLOWERING LOCUS C-clade genes repress flowering in response to chilling. *Front Plant Sci.* 2022;13: 817356. <https://doi.org/10.3389/fpls.2022.817356>.

4. Bouché F, Lobet G, Tocquin P, Périlleux C. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res.* 2016;44:D1167–71. <https://doi.org/10.1093/nar/gkv1054>.
5. Fornara F, de Montaigu A, Coupland G. SnapShot: Control of flowering in *Arabidopsis*. *Cell.* 2010;141(550):550.e1–2. <https://doi.org/10.1016/j.cell.2010.04.024>.
6. Copley TR, Duceppe M-O, O'Donoghue LS. Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. *BMC Genomics.* 2018;19:167. <https://doi.org/10.1186/s12864-018-4558-4>.
7. Sedivy EJ, Wu F, Hanzawa Y. Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytol.* 2017;214:539–53. <https://doi.org/10.1111/nph.14418>.
8. Shook JM, Zhang J, Jones SE, Singh A, Diers BW, Singh AK. Meta-GWAS for quantitative trait loci identification in soybean. *G3 (Bethesda).* 2021;11(7):jkab117. <https://doi.org/10.1093/g3journal/jkab117>.
9. Jung C-H, Wong CE, Singh MB, Bhalla PL. Comparative genomic analysis of soybean flowering genes. *PLoS ONE.* 2012;7:e38250. <https://doi.org/10.1371/journal.pone.0038250>.
10. Liu W, Jiang B, Ma L, Zhang S, Zhai H, Xu X, et al. Functional diversification of Flowering Locus T homologs in soybean: GmFT1a and GmFT2a/5a have opposite roles in controlling flowering and maturation. *New Phytol.* 2018;217:1335–45. <https://doi.org/10.1111/nph.14884>.
11. Kim SK, Lee T, Kang YJ, Hwang WJ, Kim KH, Moon J-K, et al. Genome-wide comparative analysis of flowering genes between *Arabidopsis* and mungbean. *Genes Genomics.* 2014;36:799–808. <https://doi.org/10.1007/s13258-014-0215-8>.
12. Sandhu K, Singh A. Strategies for the utilization of the USDA mung bean germplasm collection for breeding outcomes. *Crop Sci.* 2021;61:422–42. <https://doi.org/10.1002/csc2.20322>.
13. Nair R, Schreinemachers P. Global status and economic importance of mungbean. In: Nair RM, Schaffleitner R, Lee S-H, editors. *The Mungbean Genome*. Cham: Springer International Publishing; 2020. p. 1–8. [https://doi.org/10.1007/978-3-030-20008-4\\_1](https://doi.org/10.1007/978-3-030-20008-4_1).
14. Isemura T, Kaga A, Tabata S, Soma P, Srinives P, Shimizu T, et al. Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS ONE.* 2012;7:e41304. <https://doi.org/10.1371/journal.pone.0041304>.
15. Chiteri KO, Jubery TZ, Dutta S, Ganapathysubramanian B, Cannon S, Singh A. Dissecting the root phenotypic and genotypic variability of the low mung bean diversity panel. *Front Plant Sci* 2022;12. <https://doi.org/10.3389/fpls.2021.808001>.
16. Chiteri KO, Chiranjeevi S, Jubery TZ, Rairdin A, Dutta S, Ganapathysubramanian B, et al. Dissecting the genetic architecture of leaf morphology traits in mungbean (*Vigna radiata* (L.) Wiczek) using genome-wide association study. *The Plant Phenome Journal* 2023;6. <https://doi.org/10.1002/ppj2.20062>.
17. Batzer JC, Singh A, Rairdin A, Chiteri K, Mueller DS. Mungbean: A preview of disease management challenges for an alternative U.S. cash crop. *J Integr Pest Manag* 2022;13. <https://doi.org/10.1093/jipm/pmab044>.
18. Chiteri KO. Diversity and comparative genomics studies in mung bean [*Vigna radiata* (L.) Wilczek] [Doctoral dissertation, Iowa State University, 2022]. <https://dr.lib.iastate.edu/handle/20.500.12876/9z0KLEJr>.
19. Kumawat G, Kumawat CK, Chandra K, Pandey S, Chand S, Mishra UN, et al. Insights into marker assisted selection and its applications in plant breeding. In: Abdurakhmonov IY, editor. *Plant Breeding*. Rijeka: IntechOpen; 2021. <https://doi.org/10.5772/intechopen.95004>.
20. Kumar J, Choudhary AK, Solanki RK, Pratap A. Towards marker-assisted selection in pulses: a review. *Plant Breed.* 2011;130:297–313. <https://doi.org/10.1111/j.1439-0523.2011.01851.x>.
21. Collard BCY, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci.* 2008;363:557–72. <https://doi.org/10.1098/rstb.2007.2170>.
22. Heffner EL, Sorrells ME, Jannink J-L. Genomic selection for crop improvement. *Crop Sci.* 2009;49:1–12. <https://doi.org/10.2135/cropsci2008.08.0512>.
23. Singh AK, Singh A, Sarkar S, Ganapathysubramanian B, Schapaugh W, Miguez FE, et al. High-Throughput Phenotyping in Soybean. In: Zhou J, Nguyen HT, editors. *High-Throughput Crop Phenotyping*. Cham: Springer International Publishing; 2021. p. 129–63. [https://doi.org/10.1007/978-3-030-73734-4\\_7](https://doi.org/10.1007/978-3-030-73734-4_7).
24. Jeong N, Kim K-S, Jeong S, Kim J-Y, Park S-K, Lee JS, et al. Korean soybean core collection: genotypic and phenotypic diversity population structure and genome-wide association study. *PLoS ONE.* 2019;14:e0224074. <https://doi.org/10.1371/journal.pone.0224074>.
25. Han X, Wang D, Song G-Q. Expression of a maize SOC1 gene enhances soybean yield potential through modulating plant growth and flowering. *Sci Rep.* 2021;11:12758. <https://doi.org/10.1038/s41598-021-92215-x>.
26. Lozano-Isla F. *Inti: tools and statistical procedures in plant science*. Zenodo; 2021. <https://doi.org/10.5281/zenodo.4578771>.
27. Cullis BR, Smith AB, Coombes NE. On the design of early generation variety trials with correlated data. *J Agric Biol Environ Stat.* 2006;11:381. <https://doi.org/10.1198/108571106X154443>.
28. The R Project for Statistical Computing 2022. <https://www.r-project.org/> (accessed 27 Feb 2023).
29. Ha J, Satyawan D, Jeong H, Lee E, Cho K-H, Kim MY, et al. A near-complete genome sequence of mungbean (*Vigna radiata* L.) provides key insights into the modern breeding program. *Plant Genome.* 2021;14(3):e20121. <https://doi.org/10.1002/tpg2.20121>.
30. Kang YJ, Kim SK, Kim MY, Lestari P, Kim KH, Ha B-K, et al. Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat Commun.* 2014;5:5443. <https://doi.org/10.1038/ncomms6443>.
31. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006;38:203–8. <https://doi.org/10.1038/ng1702>.
32. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 2016;12:e1005767. <https://doi.org/10.1371/journal.pgen.1005767>.
33. Wang J, Zhang Z. GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics.* 2021;19:629–40. <https://doi.org/10.1016/j.gpb.2021.08.005>.
34. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinformatics.* 2021. <https://doi.org/10.1016/j.gpb.2020.10.007>.
35. D. Turner S. qqman: an R package for visualizing GWAS results using q-q and manhattan plots. *J Open Source Softw* 2018;3:731. <https://doi.org/10.21105/joss.00731>.
36. Assefa T, Zhang J, Chowda-Reddy RV, Moran Lauter AN, Singh A, O'Rourke JA, et al. Deconstructing the genetic architecture of iron deficiency chlorosis in soybean using genome-wide approaches. *BMC Plant Biol.* 2020;20:42. <https://doi.org/10.1186/s12870-020-2237-5>.
37. Wheeler D, Medha B. BLAST QuickStart: example-driven web-based BLAST tutorial. *Methods Mol Biol (Clifton, N.J.).* 2007;395:149–76.
38. Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 2017;89:789–804. <https://doi.org/10.1111/tbj.13415>.
39. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic Acids Res.* 2013;41:D344–7. <https://doi.org/10.1093/nar/gks1067>.
40. Dash S, Campbell JD, Cannon EKS, Cleary AM, Huang W, Kalberer SR, et al. Legume information system (Legumeinfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* 2016;44(D1):D1181–8. <https://doi.org/10.1093/nar/gkv1159>.
41. Redsun S, Hokin S, Cameron CT, Cleary AM, Berendzen J, Dash S, et al. Doing genetic and genomic biology using the legume information system and associated resources. In: Edwards D, editor. *Plant Bioinformatics: Methods and Protocols*. New York, NY: Springer US; 2022. p. 81–100. [https://doi.org/10.1007/978-1-0716-2067-0\\_4](https://doi.org/10.1007/978-1-0716-2067-0_4).
42. Berendzen J, Brown AV, Cameron CT, Campbell JD, Cleary AM, Dash S, et al. The legume information system and associated online genomic resources. *Legume Science* 2021;3. <https://doi.org/10.1002/leg3.74>.
43. Brown AV, Grant D, Nelson RT. Using crop databases to explore phenotypes: From QTL to candidate genes. *Plants* 2021;10. <https://doi.org/10.3390/plants10112494>.
44. Brown AV, Connors SI, Huang W, Wilkey AP, Grant D, Weeks NT, et al. A new decade and new data at SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 2021;49:D1496–501. <https://doi.org/10.1093/nar/gkaa1107>.

45. Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 2010;38:D843–6. <https://doi.org/10.1093/nar/gkp798>.
46. Cannon S. Pandagma software package, 2023. Github. <https://github.com/legumeinfo/pandagma>.
47. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol.* 2017;35:1026–8. <https://doi.org/10.1038/nbt.3988>.
48. Haas BJ, Delcher AL, Wortman JR, Salzberg SL. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics.* 2004;20:3643–6. <https://doi.org/10.1093/bioinformatics/bth397>.
49. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91. <https://doi.org/10.1093/molbev/msm088>.
50. Van Dongen S. Graph clustering via a discrete uncoupling process. *SIAM J Matrix Anal Appl.* 2008;30:121–41. <https://doi.org/10.1137/040608635>.
51. Deorowicz S, Debudaj-Grabysz A, Gudyś A. FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Sci Rep.* 2016;6:33964. <https://doi.org/10.1038/srep33964>.
52. Richardson MO, Eddy SR. ORFeus: A computational method to detect programmed ribosomal frameshifts and other non-canonical translation events. *BioRxiv* 2023:2023.04.24.538127. <https://doi.org/10.1101/2023.04.24.538127>.
53. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 2010;5: e9490. <https://doi.org/10.1371/journal.pone.0009490>.
54. Wang L, Yang T, Lin Q, Wang B, Li X, Luan S, et al. Receptor kinase FERONIA regulates flowering time in Arabidopsis. *BMC Plant Biol.* 2020;20:26. <https://doi.org/10.1186/s12870-019-2223-y>.
55. Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y, et al. Map-based cloning of the gene associated with the soybean maturity locus E3. *Genetics.* 2009;182:1251–62. <https://doi.org/10.1534/genetics.108.098772>.
56. Zhao M, Yang S, Chen C-Y, Li C, Shan W, Lu W, et al. Arabidopsis BREVI-PEDICELLUS interacts with the SWI2/SNF2 chromatin remodeling ATPase BRAHMA to regulate KNAT2 and KNAT6 expression in control of inflorescence architecture. *PLoS Genet.* 2015;11: e1005125. <https://doi.org/10.1371/journal.pgen.1005125>.
57. Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, et al. Legumes as a model plant family. Genomics for food and feed report of the Cross-Legume Advances through genomics conference. *Plant Physiol.* 2005;137:1228–35.
58. Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kudapa H, Thudi M, et al. Emerging genomic tools for legume breeding: Current status and future prospects. *Front Plant Sci.* 2016;7:455. <https://doi.org/10.3389/fpls.2016.00455>.
59. Vasconcelos MW, Grusak MA, Pinto E, Gomes A, Ferreira H, Balázs B, et al. The biology of legumes and their agronomic, economic, and social impact. In: Hasanuzzaman M, Araújo S, Gill SS, editors. *The plant family Fabaceae: Biology and physiological responses to environmental stresses*. Singapore: Springer Singapore; 2020, p. 3–25. [https://doi.org/10.1007/978-981-15-4752-2\\_1](https://doi.org/10.1007/978-981-15-4752-2_1).
60. Henchion M, Hayes M, Mullen AM, Fenelon M, Tiwari B. Future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. *Foods* 2017;6. <https://doi.org/10.3390/foods6070053>.
61. Miller V, Reedy J, Cudhea F, Zhang J, Shi P, Erndt-Marino J, et al. Global, regional, and national consumption of animal-source foods between 1990 and 2018: findings from the Global Dietary Database. *Lancet Planet Health.* 2022;6:e243–56. [https://doi.org/10.1016/S2542-5196\(21\)00352-1](https://doi.org/10.1016/S2542-5196(21)00352-1).
62. Somta P, Laosatit K, Yuan X, Chen X. Thirty years of mungbean genome research: Where do we stand and what have we learned? *Front Plant Sci* 2022;13. <https://doi.org/10.3389/fpls.2022.944721>.
63. Hwang WJ, Ha J, Lee T, Jeong H, Kim MY, Kim SK, et al. A candidate flowering gene in mungbean is homologous to a soybean Phytochrome A gene. *Euphytica.* 2017;213:79. <https://doi.org/10.1007/s10681-017-1866-8>.
64. Oda A, Fujiwara S, Kamada H, Coupland G, Mizoguchi T. Antisense suppression of the Arabidopsis PIF3 gene does not affect circadian rhythms but causes early flowering and increases FT expression. *FEBS Lett.* 2004;557:259–64. [https://doi.org/10.1016/s0014-5793\(03\)01470-4](https://doi.org/10.1016/s0014-5793(03)01470-4).
65. Lu C, Tian Y, Wang S, Su Y, Mao T, Huang T, et al. Phosphorylation of SPT5 by CDK2 is required for VIP5 recruitment and normal flowering in Arabidopsis thaliana. *Plant Cell.* 2017;29:277–91. <https://doi.org/10.1105/tpc.16.00568>.

## Publisher's Note

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