Article

Genome-Wide Association Study for Agronomic Traits in Wild Soybean (*Glycine soja*)

Woon Ji Kim [1], Byeong Hee Kang [1,2], Chang Yeok Moon [1,2], Sehee Kang [1,2], Seoyoung Shin [1], Sreeparna Chowdhury [1], Soon-Chun Jeong [3], Man-Soo Choi [4], Soo-Kwon Park [4], Jung-Kyung Moon [4] and Bo-Keun Ha [1,2,*]

1   Department of Applied Plant Science, Chonnam National University, Gwangju 61186, Republic of Korea
2   BK21 FOUR Center for IT-Bio Convergence System Agriculture, Chonnam National University, Gwangju 61186, Republic of Korea
3   Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju 28116, Republic of Korea
4   National Institute of Crop Science, RDA, Wanju 55365, Republic of Korea
*   Correspondence: bkha@jnu.ac.kr; Tel.: +82-62-530-2055

Abstract: The agronomic traits of soybean are important because they are directly or indirectly related to its yield. Cultivated soybean (*Glycine max* (L.) Merr) has lost genetic diversity during domestication and selective breeding. However, wild soybean (*G. soja*) represents a useful breeding material because it has a diverse gene pool. In this study, a total of 96,432 single-nucleotide polymorphisms (SNPs) across 203 wild soybean accessions from the 180K Axiom® Soya SNP array were employed in the association analysis. Wild soybean accessions were divided into four clusters based on their genetic distance using ADMIXTURE, principal component analysis, and neighbor-joining clusters. The linkage disequilibrium decayed rapidly in wild soybean. A genome-wide association study was conducted for days to flowering (DtF), days to maturity (DtM), the number of pods (NoP), and the 100-seed weight (100SW), which are major agronomic traits for wild soybean accessions. A total of 22 significant SNPs were found to be associated with DtF, DtM, and the 100SW. Based on the detected SNP markers, *Glyma.12g210400*, a gene related to DtF, *Glyma.17g115300*, a gene related to DtM, and *Glyma.14g140200*, a gene related to the 100SW, were selected as candidate genes. The SNP markers related to agronomic traits identified in this study are expected to help improve the quality of soybean cultivars through selective breeding.

Keywords: wild soybean; agronomic trait; genome-wide association study

1. Introduction

Soybean (*Glycine max* L.) is one of the world's most valuable legumes due to its use as a source of food, oil, livestock feed, and industrial materials. Because soybean is a self-pollinated crop, its homozygosity increases with successive generations, resulting in lower genomic variation [1]. Its genetic diversity has also been reduced due to genetic bottlenecks during domestication and selective breeding [2,3]. Therefore, a genetically diverse germplasm is required to improve the agronomic and seed quality characteristics of soybean. Wild soybean (*Glycine soja* Siebold & Zucc.) is the wild ancestor of cultivated soybean (*G. max*) and represents a valuable gene pool for improving the agronomic traits of soybean [4], particularly in terms of retrieving lost alleles. In addition, wild soybean may have unique alleles because they are found across a variety of climates and geographic topographies [5].

The individual growth stages of soybean can affect its seed composition, harvest time, and yield. In particular, flowering, which refers to plant transition from vegetative stage to reproductive stage [6], is related to photoperiod response and vernalization and is the most important agronomic characteristic of soybean in that it determines regional adaptation and yield [7–9]. In addition, the number of pods and the seed weight have a positive

correlation with soybean yields [10]. Soybean is a short-day plant, and its growth and flowering are regulated by the photoperiod [11]. The flowering and maturity of soybean are controlled by multiple genes, with 11 major genes (*E1* through *E11*) having been identified across various soybean chromosomes to date [12–19]. The *E1* gene, which encodes a legume-specific transcription factor, is located on chromosome 6 and has been reported to be involved in the regulation of flowering and maturity in soybean [12]. In addition, the *E1*, *E2*, *E3*, *E4*, *E5*, and *E7* genes are sensitive to long days and/or light quality [13,20,21]. The mechanism of action or function of these genes has been confirmed to be homologous with *Arabidopsis thaliana*, which is widely used as a model for plant research [22–24]. In addition, the seed size, seed weight, and number of pods are controlled by the *GmCYP78A10* gene in soybean, with a 7.2% difference in seed weight and 5.8% difference in the number of pods between the wild soybean allele *GmCYP78A10A* and the cultivated soybean allele *GmCYP78A10B* [25].

Bi-parent QTL analysis generally uses populations derived from both parental lines and is limited to a specific phenotype. On the other hand, a genome-wide association study (GWAS) typically uses natural populations and allows for a comparison of various genotypes and phenotypes [26]. GWASs have been successfully employed to identify QTLs and genes in various crops, such as rice, wheat, maize, and soybean [27–32]. Factors such as selection, migration, local adaptation, geographic isolation, and genetic shifts affect the population structure, leading to differences in allele frequencies, and these differences in the population structure can lead to spurious associations between traits and markers in GWASs [33]. In addition, limited linkage disequilibrium (LD) between pairs of adjacent markers can originate from a set of unrelated genotypes, which requires a fairly large number of markers to resolve [32].

With the development of high-throughput genotyping technology, the use of single-nucleotide polymorphism (SNP) genotyping has become easier. In Korea, a 180K Axiom® Soya SNP array has been developed from more than 4 million high-quality SNPs identified from the resequencing of 16 Korean and 31 Chinese genetic resources [34]. This array has been used in the structural analysis of genetic diversity [35] and in association analysis for various traits, such as the flowering time, ultraviolet B resistance, and the soyasaponin biosynthesis of soybean [36–38]. The purpose of this study was to investigate the genetic diversity of soybean using high-density SNP chip data from 203 wild soybean accessions and to detect candidate genes for major agricultural traits that can affect soybean yield using GWAS analysis.

2. Materials and Methods

2.1. Plant Materials and Experiment Field Management

A total of 203 wild soybean accessions were obtained from the Rural Development Administration (RDA) in South Korea. These accessions were collected from South Korea (167), China (22), Japan (11), and Russia (1), with two from an unknown location (Table S1). They were cultivated in an experimental field at Chonnam National University (Gwangju, 36°17′ N, 126°39′ E, Republic of Korea) in 2015 and 2016. Seeds were sown in late June and harvested from September to November based on the flowering time and maturity. Each accession was planted in a single hill plot at a spacing of 1 × 1 m with two replications. Compound fertilizer composed of nitrogen, phosphorus pentoxide, and potassium oxide at a ratio of 8:8:9 was applied at 40 kg per 1000 square meter before sowing, and pesticides and fungicides were alternately applied every two weeks until one month before harvest.

2.2. Phenotypic Evaluation and Data Analysis

The agronomic traits investigated were days to flowering (DtF), days to maturity (DtM), number of pods (NoP), and 100-seed weight (100SW). The DtF was recorded as the date when the first flower for each individual plant bloomed. The DtM was recorded as the number of days between flowering and the date when approximately 90% of the pods exhibited the color associated with maturity. The NoP was measured as the number of pods

per raceme. For 100SW, mature seeds were dried in a dry oven at 30 °C for 72 h, and then 100 of them were randomly selected and weighed. Descriptive statistics and correlation analysis were performed with SPSS 25.0 software, Microsoft Excel 2016, and Python 3.9 using the phenotypic data for each trait. The broad-sense heritability (h^2) was calculated using QTLmaxV2 software [39] using the following equation:

$$h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_E^2}$$

where σ_u^2 is the variance component for random genetic effects and σ_E^2 is the variance component for environmental effects.

2.3. DNA Extraction and SNP Genotyping

Fresh leaf tissue was collected for DNA extraction at the beginning of growth and ground in a mortar using liquid nitrogen. Genomic DNA was isolated from 20 mg of lyophilized leaf tissue using a DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. The quality and quantity of the extracted total DNA were verified using a Nano-MD UV-Vis spectrophotometer (Scinco, Seoul, Korea). The extracted DNA was stored in a freezer at −80 °C until further use. A total of 203 wild soybean accessions were genotyped using the 180K Axiom® Soya SNP array [34]. SNPs with genotypes observed in less than 95% of the sample were removed from further analysis (CallRate < 0.95). Similarly, one individual had missing genotypes for more than 5% of the SNPs, so it was excluded from the sample. SNPs with a minor allele frequency (MAF) under 5% were also eliminated to remove low-quality SNPs, while duplicated parts of the raw data were eliminated using R software [40]. Missing genotypes were imputed using BEAGLE software 3.0 [41]. As a result, a total of 96,432 SNPs were used for analysis.

2.4. Population Structure and Genetic Diversity

The population structure was determined using ADMIXTURE v1.30 [42]. In this analysis, the optimal *K* value was assigned to different admixture models (*K* = 2 to 10) based on the cross-validation error. The box plots were generated using the "sort by *Q*" command based on the best *K* with the lowest cross-validation error. For the proportional member probability (*Q*), the cut-off assignment for each cluster was >0.55. The phylogenetic trees were drawn based on the maximum-likelihood approach using 1000 bootstrap replicates and inter-site ratios for the automatic initial trees (neighbor-joining) in MEGA11 [43]. A three-dimensional principal component analysis (PCA) plot was used to identify the distribution of the genetic variation between geographic locations using PLINK v1.07 [44]. The neighbor-joining tree and PCA used the same color based on the classification from the ADMIXTURE analysis.

2.5. Linkage Disequilibrium Estimation and Candidate Gene Identification

LD estimation was conducted using TASSEL with an LD sliding window of 500 within individual chromosomes [45]. The identification of candidate genes was carried out in the region considering the LD for the SNP detected as a result of the GWAS and in the downstream and upstream regions. Glyma.Wm82.a2.v1, a soybean reference genome from www.soybase.org (accessed on 1 November 2022), was used as the gene model for candidate gene identification.

2.6. Genome-Wide Association Study Analysis

The GWAS was conducted with a linear mixed model using the restricted maximum-likelihood (REML) algorithm and considering the population structure from ADMIXTURE analysis and an identity-by-state (IBS) similarity matrix. The IBS matrix was obtained using the proportions of IBS alleles and used as a kinship matrix in PLINK v1.07. Genomic control was also conducted to correct for genomic inflation, which means that the *p* value

did not follow a chi-square distribution with 1 degree of freedom and that the statistical significance was exaggerated in QTLmaxV2. In order to increase the resolution between environments, association analysis was conducted using a merged phenotype [39]. The $-\log_{10}(p)$ threshold of the Manhattan plot was calculated using the Bonferroni method ($p = \alpha/$n). With 96,432 SNPs employed in this study, at $\alpha = 1$ and $\alpha = 0.05$, the Bonferroni-corrected thresholds for the *p* values were 1.04×10^{-5} ($\alpha = 1$) and 5.19×10^{-7} ($\alpha = 0.05$), with corresponding $-\log_{10}(p)$ values of 4.98 for the suggestive threshold and 6.29 for the significance threshold [46].

3. Results

3.1. Phenotypic Variation and Correlation Analysis

To evaluate the major agronomic traits of soybean, a total of 203 wild soybean accessions were cultivated in 2015 and 2016. The descriptive statistics and heritability (h^2) for each trait are summarized in Table 1.

Table 1. Descriptive statistics for the agronomic traits in 2015 and 2016.

Trait	Year	Min	Max	SD	Mean	CV (%)	Skew	Kur	h^2
DtF	2015	29	80	7.69	64.61	11.90	−2.33	6.39	
	2016	26	81	8.71	61.11	14.26	−1.34	3.11	0.89
	Mean	28	81	7.90	63.20	12.50	−1.91	5.03	
DtM	2015	44	82	6.22	51.98	11.98	2.28	6.04	
	2016	42	90	9.02	60.29	14.96	0.71	0.33	0.55
	Mean	43	83	6.67	56.15	11.87	1.33	2.23	
NoP	2015	2.0	11.4	1.47	5.36	27.51	0.91	1.58	
	2016	1.8	10.2	1.60	4.63	34.55	0.76	0.61	0.39
	Mean	2.3	10.2	1.19	4.99	23.88	0.69	1.53	
100SW	2015	1.30	10.57	1.20	3.15	38.00	2.80	11.01	
	2016	1.15	7.69	1.01	2.88	35.00	2.01	5.85	0.75
	Mean	1.26	7.83	1.02	3.00	33.99	2.33	7.17	

DtF, days to flowering; DtM, days to maturity; NoP, number of pods; 100SW, 100-seed weight; Min, minimum; Max, maximum; SD, standard deviation; CV, coefficient of variation; Skew, skewness; Kur, Kurtosis; h^2, broad sense heritability.

In 2015 and 2016, the average DtF was 28 to 81 d, with an average of 63 d, and DtM was 43 to 83 d, with an average of 56 d. The NoP was 2.3 to 10.2, with an average of 4.99, and the 100SW was 1.26 to 7.83 g, with an average of 3 g. Of the four traits, h^2 was the highest for DtF (0.89) and the lowest for the NoP (0.39), while the coefficient of variation was the highest for the 100SW (33.99) and lowest for DtM (11.87). As shown in Figure 1, these traits were quantitatively inherited with an approximately normal distribution in the two environments.

The correlation coefficients between the agronomic traits in each environment are presented in Figure 2. In 2015 and 2016, DtF and DtM exhibited a negative correlation ($r = −0.729$ ** and $−0.440$ **, respectively). DtF and the 100SW also showed a negative correlation ($r = −0.390$ ** and $−0.345$ **, respectively). On the other hand, DtM and the 100SW had a positive correlation ($r = 0.532$ ** and 0.423 **, respectively). DtF and the NoP, DtM and the NoP, and the 100SW and the NoP had a significant correlation in 2015 but not in 2016.

3.2. Population Structure and Linkage Disequilibrium

To understand the population structure of the 203 wild soybean accessions, various statistical analyses were conducted. First, cross-validation was used to cluster the populations, with $K = 4$ selected for this classification because it had the lowest error (Figure 3a). Analysis with the parametric clustering method using ADMIXTURE also found that the model with four ancestral populations (*Q1*–*Q4*) best described the genetic structure. According to the

membership coefficient ($qi > 0.55$), of the 203 accessions, 178 were assigned to Q1–Q, and 25 accessions were assigned to an admixed group (Figure 3b). In particular, 13 were assigned to Q1, all of which were collected in China, 16 were assigned to Q2 (15 from Korea and 1 unknown), 53 to Q3, (42 from Korea and 11 from Japan), and 96 to Q4 (90 from Korea, 4 from China, 1 from Russia, and 1 unknown). The admixed group contained 25 accessions, consisting of 20 from Korea and 5 from China. In the PCA, the 203 wild soybean accessions were divided into four main components (Figure 3c). The three-dimensional PCA plot explained 59% of the total genetic variance (28% for PC1, 21% for PC2, and 10% for PC3). Neighbor-joining cluster analysis using MEGA11 also confirmed that the 203 accessions were clearly separated into four clusters (Figure 3d). As a result, the ADMIXTURE-based PCA analysis and neighbor-joining cluster analysis strongly indicated that the 203 wild soybean accessions consisted of four well-differentiated genetic sub-populations.
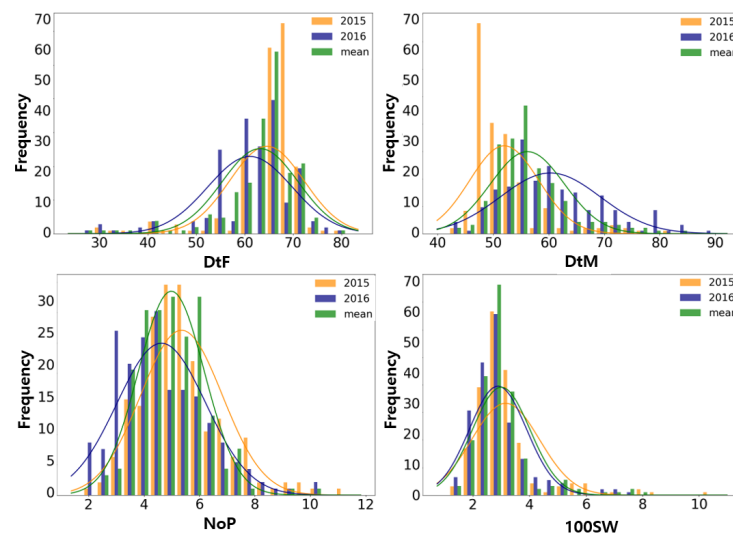

Figure 1. Distribution of agronomic traits (flowering, maturity, number of pods, and seed weight) in 203 wild soybean accessions. DtF, days to flowering; DtM, days to maturity; NoP, number of pods; 100SW, 100-seed weight.
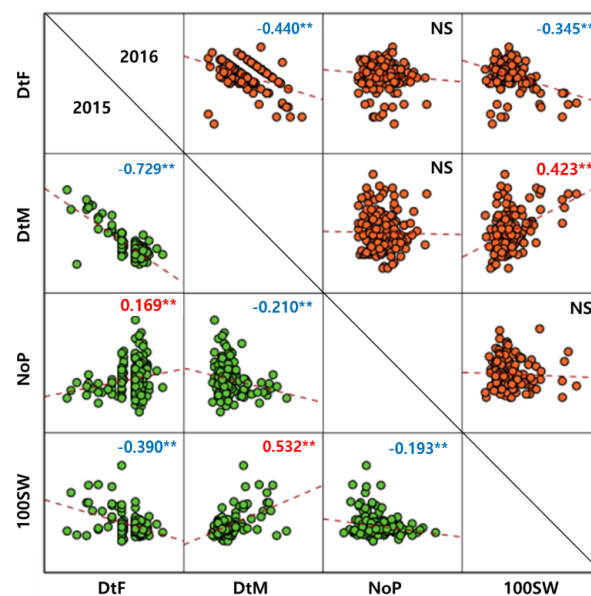

Figure 2. Scatter diagrams for flowering, maturity, number of pods, and seed weight in 203 wild soybean accessions for 2015 (green) and 2016 (orange). DtF, days to flowering; DtM, days to maturity; NoP, number of pods; 100SW, 100-seed weight. **, significant at $p \leq 0.01$.
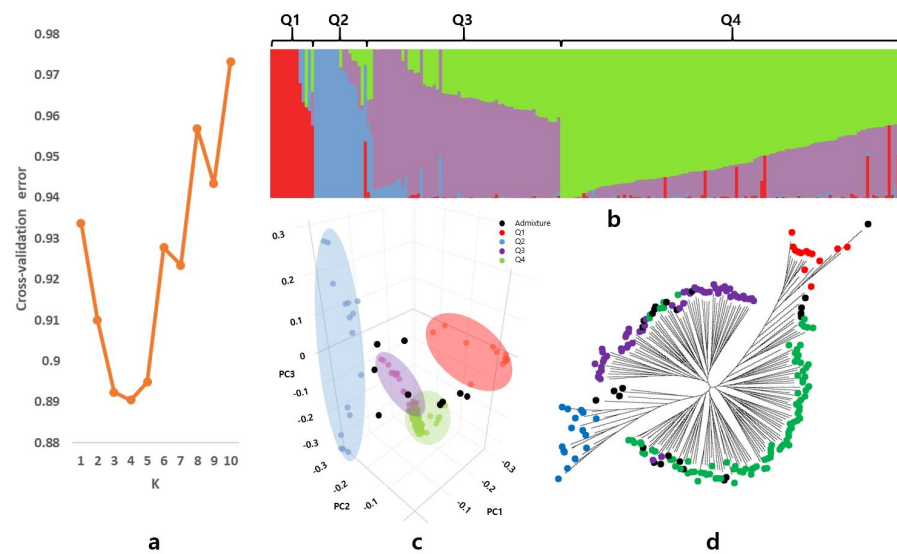
Figure 3. (**a**) Cross-validation error for the assumed number of populations (*K*) in ADMIXTURE analysis. (**b**) Classification of 203 wild soybean accessions into four sub-populations using ADMIXTURE. The colors red, blue, purple, and green represent separate groups with different levels of admixture. (**c**) Scatter diagram for the principal component analysis of 203 accessions calculated using PLINK, with each colored spot representative of one cluster of *Q*1 to *Q*4. (**d**) Neighbor-joining cluster analysis in MEGA11, with each colored spot representative of one cluster of *Q*1 to *Q*4.

By plotting r^2 against the distance (kb) between pairs of SNPs, it was found that the LD decay sharply decreased. Estimating the LD decay for the entire genome resulted in a decrease in the maximum r^2 drop to half at 6 kb (Figure 4), and the average LD decay for each of the 20 chromosomes was 4 to 10 kb (data not shown). Consequently, the size of the LD block was considered to be quite small.
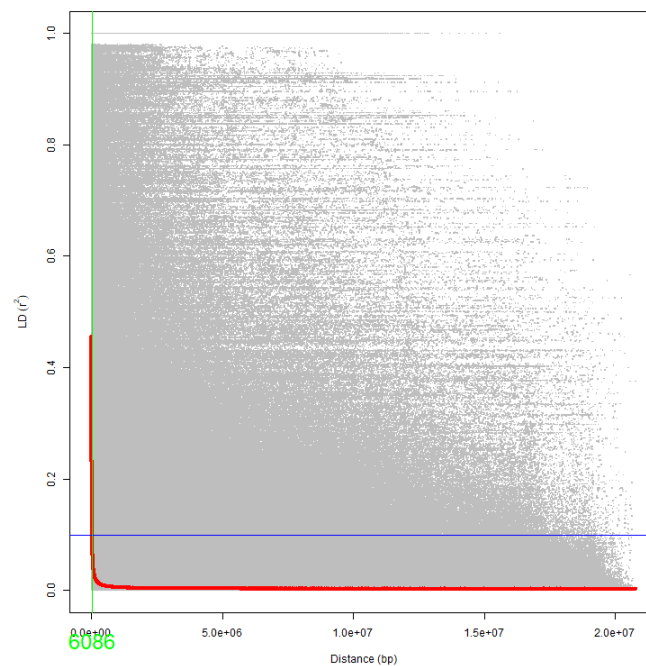

Figure 4. Distribution of the marker density and decay of the linkage disequilibrium (LD) over distance presented as an accumulative distribution. The x-axis indicates the distance (kb) between SNPs, the y-axis indicates the LD (r^2), and the solid blue horizontal line indicates the distance of the LD pattern for all genotypes in wild soybean.

3.3. Genome-Wide Association Study for Agronomic Traits

In this study, association analysis was performed with a linear mixed model for the four major agronomic traits. According to the results of the Manhattan and QQ plots presented in Figure 5, 18 SNPs across 9 chromosomes for DtF, 32 SNPs across 7 chromosomes for DtM, and 41 SNPs across 16 chromosomes for the 100SW were detected at the significance threshold level ($-\log_{10}(p) = 6.29$). However, no suggestive or significant SNPs were detected for the NoP.
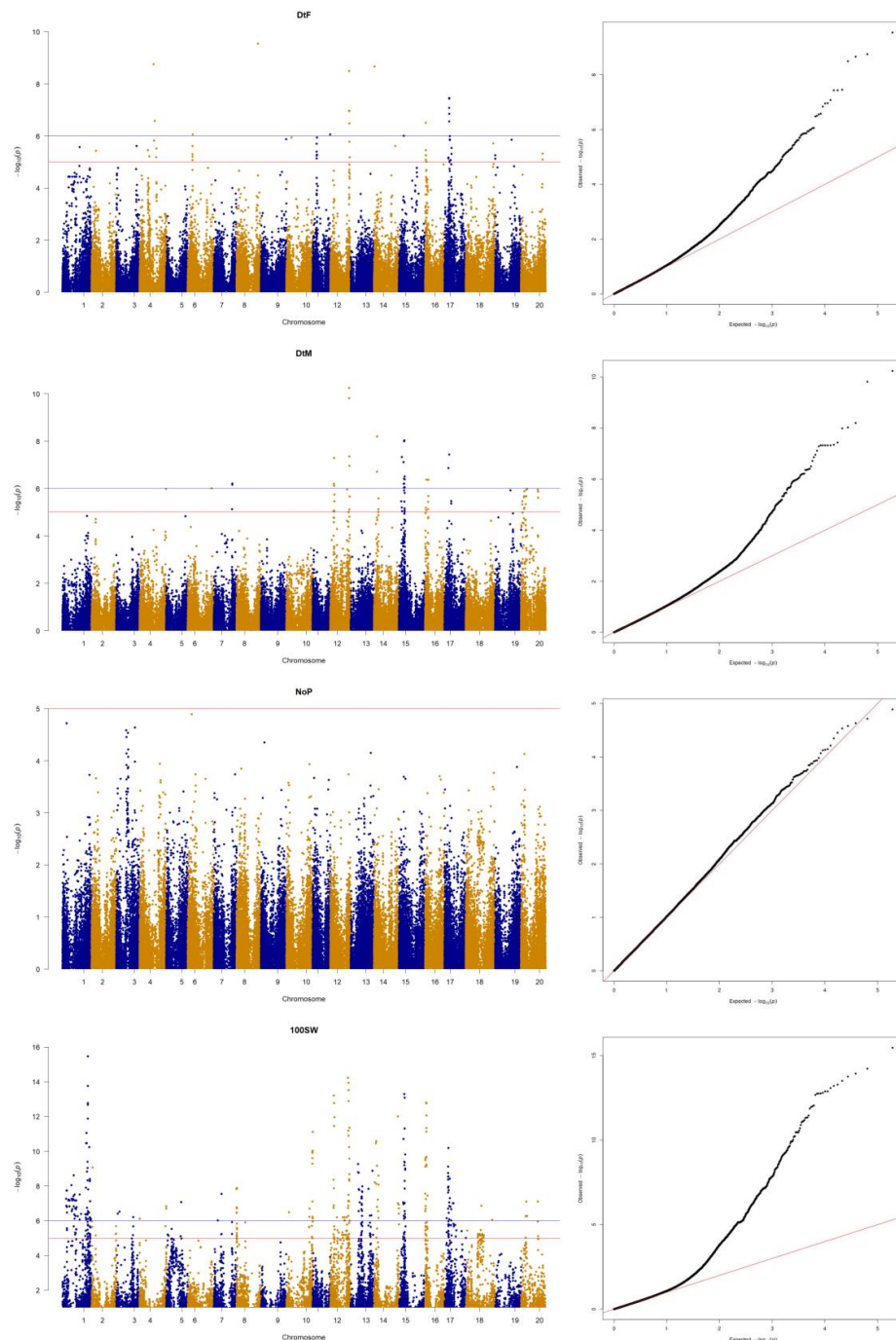

Figure 5. Manhattan plots and quantile–quantile (QQ) plots for four agronomic traits in 203 wild soybean accessions. In the Manhattan plots, the red line indicates the genome-wide threshold $-\log_{10}(p) = 4.98$ and the blue line represents $-\log_{10}(p) = 6.29$, calculated using the Bonferroni method. DtF, days to flowering; DtM, days to maturity; NoP, number of pods; 100SW, 100-seed weight.

Of the detected SNPs, the most significant SNP (i.e., with the lowest *p* value) in a particular associated genetic region was selected as a causal SNP for the trait. For DtF, five SNP markers were identified on chromosomes 6 (AX-90393598), 11 (AX-90507114), 12 (AX-90495922), 16 (AX-90386690), and 17 (AX-90395214) (Table 2).

Table 2. Significant SNP markers related to DtF in the merged phenotype.

SNP	Chr	Position	$-\log_{10}(p)$	Reference [+]	Minor	Major	MAF
AX-90393598	6	9,723,612	6.05	G	T	G	0.09
AX-90507114	11	34,284,790	6.06	G	A	G	0.07
AX-90495922	12	36,991,550	8.49	C	C	T	0.08
AX-90386690	16	1,150,092	6.50	G	G	A	0.16
AX-90395214	17	9,164,086	7.43	T	T	C	0.08

SNP, single-nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency.   [+] gene model: Glyma.Wm82.a2.v1.

For DtM, six SNP markers were identified on chromosomes 12 (AX-90440044 and AX-90495922), 14 (AX-90366397), 15 (AX-90472604), 16 (AX-90351904), and 17 (AX-90338412) (Table 3).

Table 3. Significant SNP markers related to DtM in the merged phenotype.

SNP	Chr	Position	$-\log_{10}(p)$	Reference [+]	Minor	Major	MAF
AX-90440044	12	7,550,988	7.29	T	T	C	0.08
AX-90495922	12	36,991,550	9.82	C	C	T	0.08
AX-90366397	14	5,987,400	8.2	A	A	C	0.06
AX-90472604	15	10,746,176	8.03	G	G	T	0.07
AX-90351904	16	2,017,593	6.37	T	T	C	0.09
AX-90338412	17	9,141,000	7.43	A	A	T	0.07

SNP, single-nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency.   [+] gene model: Glyma.Wm82.a2.v1.

In addition, 12 SNP markers associated with the 100SW were identified on chromosomes 1 (AX-90454597), 8 (AX-90377215), 10 (AX-90374196), 12 (AX-90408186 and AX-90370905), 13 (AX-90318417), 14 (AX-90483232 and AX-90383559), 15 (AX-90472604), 16 (AX-90416982), and 17 (AX-90375042) (Table 4).

Table 4. Significant SNP markers related to 100SW in the merged phenotype.

SNP	Chr	Position	$-\log_{10}(p)$	Reference [+]	Minor	Major	MAF
AX-90454597	1	46,804,555	11.05	C	C	T	0.07
AX-90377215	8	213,783	7.87	T	T	C	0.09
AX-90374196	10	51,266,958	11.10	C	C	A	0.09
AX-90408186	12	6,713,247	13.20	A	A	C	0.06
AX-90370905	12	34,609,739	14.22	T	T	C	0.09
AX-90318417	13	14,296,524	9.26	C	C	T	0.07
AX-90483232	14	3,979,096	10.57	C	C	G	0.06
AX-90383559	14	47,044,439	11.99	A	A	T	0.07
AX-90472604	15	10,746,176	13.28	G	G	T	0.07
AX-90416982	16	1,413,306	12.80	G	G	A	0.05
AX-90375042	17	7,394,535	10.18	G	G	A	0.09

SNP, single-nucleotide polymorphism; Chr, chromosome; MAF, minor allele frequency.   [+] gene model: Glyma.Wm82.a2.v1.

Interestingly, the AX-90495922 SNP marker on chromosome 12 was consistently associated with both DtF and DtM, while AX-90472604 on chromosome 15 was significantly associated with DtM and the 100SW. For each trait, the allelic effect of the SNP marker

with the lowest *p* value was estimated. The DtF-associated SNP marker AX-90495922 on chromosome 12 had the alleles T/C, and the average DtF for the individuals with TT alleles was 64 d, 15 d longer than the average DtF for the individuals with CC alleles (49 d). The DtM-associated SNP marker AX-90366397 on chromosome 14 had the alleles C/A, and the average DtM for the individuals with CC alleles was 55 d, which was 16 d shorter than the average DtM for the individuals with AA alleles (71 d). The 100SW-associated SNP marker AX-90383559 on chromosome 14 had the alleles T/A, and the average 100SW for the individuals with TT alleles was 2.83 g, which was 3.28 g lighter than the average 100SW for individuals with AA alleles (6.11 g) (Figure 6).
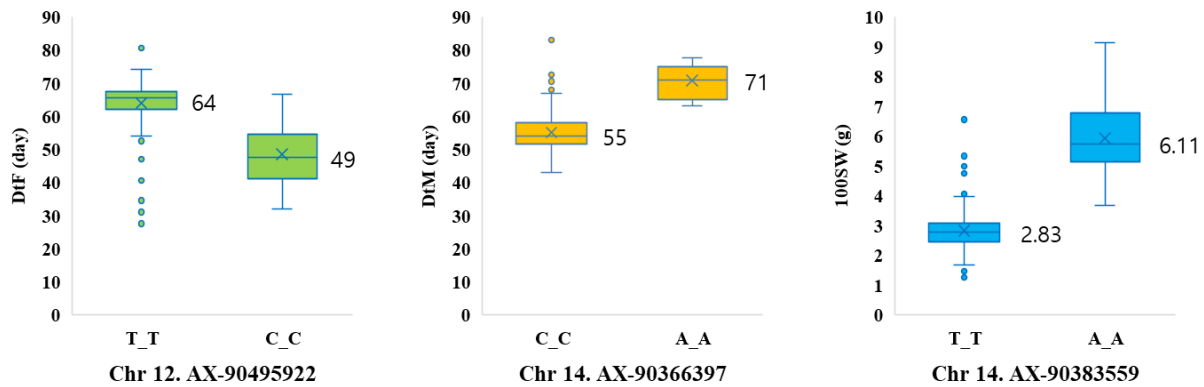

Figure 6. Phenotypic differences between lines carrying different alleles for SNPs associated with DtF, DtM, and the 100SW. DtF, days to flowering; DtM, days to maturity; 100SW, 100-seed weight.

3.4. Candidate Genes for Trait-Associated SNP Markers

A search for candidate genes for the SNPs identified from the GWAS results was conducted based on whether or not it was included in the gene range, because the LD for the entire genome was quite small, with the search occurring within about 50 kb. Of the total 22 SNP markers associated with DtF, DtM, and the 100SW, 20 were located within a genic region. Of these, four DtF-related SNP markers were located in the *Glyma.06g119400*, *Glyma.11g251500*, *Glyma.12g210400*, and *Glyma.17g116200* genes (Table 5), while DtM-related SNP markers were located in *Glyma.12g091600*, *Glyma.12g210400*, *Glyma.14g071300*, *Glyma.15g133700*, *Glyma.16g021200*, and *Glyma.17g115300* (Table 6), and 100SW-related SNP markers were located in *Glyma.01g140200*, *Glyma.08g002900*, *Glyma.10g295400*, *Glyma.12g184700*, *Glyma.13g047500*, *Glyma.14g050900*, *Glyma.14g205200*, *Glyma.15g133700*, *Glyma.16g016000*, and *Glyma.17g094400* (Table 7).

Table 5. Candidate genes associated with DtF in wild soybean.

SNP	Chr	Position	Gene [+]	Location (bp)	Gene Description
AX-90393598	6	9,723,612	*Glyma.06g119400*	9719172..9726230	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AX-90507114	11	34,284,790	*Glyma.11g251500*	34283168..34289333	squamosa promoter binding protein-like 2
AX-90495922	12	36,991,550	*Glyma.12g210400*	36943077..36946491	14-3-3 protein
AX-90395214	17	9,164,086	*Glyma.17g116200*	9181066..9183991	CCCH-type zinc finger family protein

SNP, single-nucleotide polymorphism; Chr, chromosome; [+] gene model: Glyma.Wm82.a2.v1.

The candidate genes with the strongest SNP variant for each trait were *Glyma.12g210400* for DtF, *Glyma.14g071300* for DtM, and *Glyma.14g205200* for the 100SW, corresponding to the 14-3-3 protein, RING/U-box superfamily protein, and cinnamate-4-hydroxylase, respectively.

Table 6. Candidate genes associated with DtM in wild soybean.

SNP	Chr	Position	Gene [+]	Location (bp)	Gene Description
AX-90440044	12	7,550,988	*Glyma.12g091600*	7494222..7499978	OTUBAIN-LIKE DEUBIQUITINASE 1
AX-90495922	12	36,991,550	*Glyma.12g210400*	36943077..36946491	14-3-3 protein
AX-90366397	14	5,987,400	*Glyma.14g071300*	5986366..5990379	RING/U-box superfamily protein
AX-90472604	15	10,746,176	*Glyma.15g133700*	10744186..10748830	Glycosyl hydrolases family 31 protein
AX-90351904	16	2,017,593	*Glyma.16g021200*	1996084..1998423	Cytochrome P450, Family 78
AX-90338412	17	9,141,000	*Glyma.17g115300*	9122203..9127019	NRT1

SNP, single-nucleotide polymorphism; Chr, chromosome; [+] gene model: Glyma.Wm82.a2.v1.

Table 7. Candidate genes associated with 100SW in wild soybean.

SNP	Chr	Position	Gene [+]	Location (bp)	Gene Description
AX-90454597	1	46,804,555	*Glyma.01g140200*	46825221..46841112	arginase
AX-90377215	8	213,783	*Glyma.08g002900*	211294..221624	cyclin-dependent kinase E;1
AX-90374196	10	51,266,958	*Glyma.10g295400*	51258630..51267380	2-isopropylmalate synthase 1
AX-90370905	12	34,609,739	*Glyma.12g184700*	34608320..34611560	Homeodomain-like superfamily protein
AX-90318417	13	14,296,524	*Glyma.13g047500*	14291927..14296667	NagB/RpiA/CoA transferase-like superfamily protein
AX-90483232	14	3,979,096	*Glyma.14g050900*	3977944..3980627	GDSL-like Lipase/Acylhydrolase superfamily protein
AX-90383559	14	47,044,439	*Glyma.14g205200*	47041931..47046048	cinnamate-4-hydroxylase
AX-90472604	15	10,746,176	*Glyma.15g133700*	10744186..10748830	Glycosyl hydrolases family 31 protein
AX-90416982	16	1,413,306	*Glyma.16g016000*	1400064..1404488	Purple acid phosphatase 29
AX-90375042	17	7,394,535	*Glyma.17g094400*	7393359..7395553	Homeodomain-like superfamily protein

SNP, single-nucleotide polymorphism; Chr, chromosome; [+] gene model: Glyma.Wm82.a2.v1.

4. Discussion

4.1. Genetic Diversity and Origin of Wild Soybean

Wild soybean, which is the ancestor of cultivated soybean, may have a diverse gene pool that has since been lost during domestication and cultivation. The 203 wild soybean accessions used in this study show 99% coverage, with 1135 wild soybean plants as the original population, using the 180K SNP array from the RDA. This collection has high genetic diversity without overlaps. In particular, Korean wild soybean accessions are valuable as a research material because they are reported to contain many rare alleles [47].

For association analysis, it is important to determine the genetic basis of complex traits using population structure analysis. With the ADMIXTURE and PCA analysis strongly supported by neighbor-joining clustering, the 203 wild soybean accessions were clearly clustered according to their origin into four groups: (1) China, (2) Korea, (3) Korea and Japan, and (4) Korea and China. A more detailed classification of the origin was not possible. The wild soybean in Korea was divided into three clusters, which may be related to imports and exports from neighboring countries. Given that Korea and Japan were grouped together in *Q3*, there is a possibility that wild soybean was transferred from Korea to Japan. Similarly, Korea and China were grouped together in *Q4*, with the accessions collected from Heilongjiang and Shandong in China, which are relatively close to Korea. Based on the fact that the Korean accessions were divided into three clusters mixed with neighboring countries, it can be inferred that wild soybeans around the Korean Peninsula contain high genetic diversity.

Cultivated soybean has been bred to increase its vegetative growth period and maturation period and increase the seed weight to ensure high yields. In the present study, it was confirmed that the major alleles for DtF-, DtM-, and 100SW-associated SNPs detected in the wild soybean accessions delayed the flowering period, shortened the maturation period, and reduced the seed weight (Figure 6). When alleles of each trait were compared with the reference genome for cultivated soybean (William 82), DtF was identical, while DtM and

the 100SW differed (Tables 2–4). This suggests that, in the process of breeding wild soybean to produce cultivated soybean, selection for flowering, maturity, and seed weight occurred.

4.2. Correlation of Major Agronomic Traits

The discovery of correlations between agricultural traits is useful for breeding programs. In the present study, traits that exhibited a correlation during both growing years were DtF and DtM, DtF and the 100SW, and DtM and the 100SW. On the other hand, there was a correlation between the NoP and other traits in 2015 but not in 2016 (Figure 2). Therefore, the NoP affected the inter-year correlation because it was greatly influenced by the environment. Liu et al. (2011) reported that the soybean yield had a negative correlation with the flowering time, which is consistent with the results of the present study [7]. In addition, the strong negative correlation between DtF and DtM and the positive correlation between DtM and the 100SW was in agreement with the negative correlation between DtF and the 100SW. On the other hand, it has also been reported that both DtM and the 100SW and DtF and DtM have a positive correlation [48]. These differences from the present results may be due to differences in climatic conditions, such as temperature and day length, because the experimental cultivation in the previous study took place in Indonesia.

The correlations between these agricultural traits are important because the characteristics of soybean growth stages can influence changes in seed composition, the harvest period, and the yield. It has been widely reported that, in soybean, the yield increases as both the maturation period and seed weight increase [49,50]. When comparing the genotype of the SNP markers detected for each trait and the genotype of the cultivated soybean reference genome, the cultivated soybean was found to have a genotype that increased DtM and the 100SW. Therefore, this suggests that cultivated soybean has been bred from wild soybean to increase its yield.

4.3. Candidate Genes for Major Agronomic Traits

It has been reported that the reduction in genetic diversity via selection during the domestication and cultivation of crops increases the LD [51]. Depending on the size of the LD, it may be easy to search for a candidate gene for a trait but, if the size of the LD is too large, an excessive number of candidate genes will be detected. In the present study, the LD was estimated using wild soybean, the ancestor of cultivated soybean, and the average LD was found to be 6 kb, which is considerably smaller than that of cultivated soybean, which is about 150 kb [1]. These results are in agreement with a previous study reporting that the LD range for wild soybean is ~27 kb, smaller than the LD of soybean landraces and improved cultivars 83 kb and 133 kb, respectively [52]. These results are also consistent with the LD of ~11 kb reported by Kim et al. (2021) [53], which was obtained from resources similar to those employed in the present study. Therefore, more precise candidate gene detection is possible.

No significant markers were found for the NoP in the GWAS results. As shown in Table 1, the heritability of the NoP was 0.39, which was quite low compared to the other traits, suggesting that it is greatly influenced by the environment. Cultivated soybean has been bred to adapt to a universal environment, while the wild soybean accessions used in the present study were collected from various environments. The gene *Glyma.06g119400* containing the DtF-related SNP marker AX-90393598 was annotated as S-adenosyl-L-methionine-dependent methyltransferase superfamily protein (SAM-Mtase). It has been reported that mutations in SAM-Mtase, an orthologue of trimethylguanosine synthase1, are involved in the regulation of the vernalization response and flowering time in the legume narrow-leafed lupin (*Lupinus angustifolius* L.) [54]. In addition, in the study of the lilium "Siberia", there is a report that SAM-Mtase is related to benzenoid biosynthesis during flower development [55]. In soybean, it has been reported that SAM-Mtase is involved in expressed gene interaction at temperatures of 35 °C and under dark stress during the seedling stage [56]. Soybean is a short-day plant, and its growth and flowering are regulated by the photoperiod [11]. In addition, soybean generally blooms in summer

when the temperature is high. Thus, SAM-Mtase from the *Glyma.06g119400* gene can be considered to affect flowering.

Gene *Glyma.11g251500*, which contains SNP marker AX-90507114, was annotated as squamosa promoter-binding protein-like 2 (SPL2). SPL family plays an important role in the flowering pathway, pollen development, and fertility for soybean [57,58]. The SNP marker AX-90495922 was detected for both DtF and DtM, and the gene *Glyma.12g210400* was selected as a strong candidate gene. This marker had the highest $-\log_{10}(p)$ of 8.49 and 9.82 (Tables 4 and 5) for DtF and DtM, respectively; in particular, for DtF, there was a 15 d difference between the major and minor alleles (Figure 6). *Glyma.12g210900* encodes the 14-3-3 protein, and analysis of the rice transcription factor FLOWERING LOCUS D1 (OsFD1) revealed that 14-3-3 plays an essential role in flowers [59]. In addition, 14-3-3 protein has been found to interact with FLOWERING LOCUS T(FT) in *Arabidopsis thaliana* and with PRUNING(SP) and SINGLE FLOWER TRUSS(SFT) acting as anti-florigen and florigen in tomatoes [60–62]. In the present study, it was found that DtF and DtM were significantly correlated. Additional studies are, thus, needed to determine whether these candidate genes affect soybean by a similar mechanism.

In addition, the *Glyma.17g116200* gene around the SNP marker AX-90395214 was annotated as a CCCH-type zinc finger family protein. CCCH-type zinc finger family proteins are known to regulate adaptation to various stresses, such as drought, salinity, and temperature [63]. Abiotic stress has a significant association with flowering time, which is a key plant survival strategy [64]. In addition, *Glyma.16g021200*, which was related to maturation, is a nitrate transporter and is known to affect early embryonic development in Arabidopsis [65]. In this study, genes located at *E1* to *E11* and *j* [12–19,66], known as soybean flowering-related loci, were not detected. It is presumed that this is not only because the accessions used in the study were different, but also because wild soybean was used.

The 100SW is an important criterion for determining and estimating the yield. The *Glyma.01g140200* gene, located on chromosome 1, is an arginase and has been proven to be related to yield by playing an important role in nitrogen transport and storage in rice [67]. SNP AX-90383559 detected for the 100SW is located in the *Glyma.14g205200* gene and was annotated as cinnamate-4-hydroxylase (C4H). It has been reported that C4H acts as a defense mechanism against various pathogens by catalyzing the synthesis of lignin during seed development [68,69]. During this stage, pathogens can significantly reduce the yield [11]. Therefore, it is suggested that the detected candidate gene *Glyma.14g205200* indirectly affects the 100SW. However, *GmCYP78A10* [25], known to be a gene related to seed weight in wild soybean, was not found in the present study. This could be because the gene has been identified in mixed accessions of wild soybean and cultivated soybean in previous analyses, but the present study only investigated wild soybean. The *Glyma.16g016000* gene, located approximately 10 kb downstream of the marker AX-90416982, encodes purple acid phosphatase 29. It has been reported that an increase in the expression level of purple acid phosphatase, which plays an important role in phosphate acquisition in chickpeas, leads to an increase in seed weight [70]. Additional expression analysis of the candidate genes detected in this study is required, potentially contributing to soybean breeding.

5. Conclusions

In this study, clustering using the genetic diversity of wild soybean was conducted, and candidate genes related to four target agricultural traits were searched for using GWAS analysis. The 203 wild soybean accessions were clustered into four groups according to origin, which included China, Korea, and Japan. The LD of the 203 wild soybean accessions was quite low, which could serve as an advantage or a disadvantage in the detection of candidate genes. *Glyma.12g210400*, *Glyma.17g115300*, and *Glyma.01g140200* were selected as potential candidate genes for DtF, DtM, and the 100SW. The genetic diversity of wild soybean, which can be hybridized with cultivated soybean, thus offers a useful source for soybean breeding programs.

References

1. Lam, H.-M.; Xu, X.; Liu, X.; Chen, W.; Yang, G.; Wong, F.-L.; Li, M.-W.; He, W.; Qin, N.; Wang, B.; et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* **2010**, *42*, 1053–1059. [CrossRef] [PubMed]
2. Hyten, D.L.; Song, Q.; Zhu, Y.; Choi, I.-Y.; Nelson, R.L.; Costa, J.M.; Specht, J.E.; Shoemaker, R.C.; Cregan, P.B. Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16666–16671. [CrossRef]
3. Kofsky, J.; Zhang, H.; Song, B.-H. The untapped genetic reservoir: The past, current, and future applications of the wild soybean (*Glycine soja*). *Front. Plant Sci.* **2018**, *9*, 949. [CrossRef] [PubMed]
4. Carter, T.E., Jr.; Nelson, R.L.; Sneller, C.H.; Cui, Z. Genetic diversity in soybean. *Soybeans Improv. Prod. Uses* **2004**, *16*, 303–416.
5. Nawaz, M.A.; Lin, X.; Chan, T.-F.; Ham, J.; Shin, T.-S.; Ercisli, S.; Golokhvast, K.S.; Lam, H.-M.; Chung, G. Korean wild soybeans (*Glycine soja* Sieb & Zucc.): Geographic distribution and germplasm conservation. *Agronomy* **2020**, *10*, 214.
6. Kim, D.-H.; Doyle, M.R.; Sung, S.; Amasino, R.M. Vernalization: Winter and the timing of flowering in plants. *Annu. Rev. Cell Dev.* **2009**, *25*, 277–299. [CrossRef] [PubMed]
7. Liu, W.; Kim, M.Y.; Van, K.; Lee, Y.-H.; Li, H.; Liu, X.; Lee, S.-H. QTL identification of yield-related traits and their association with flowering and maturity in soybean. *J. Crop Sci. Biotechnol.* **2011**, *14*, 65–70. [CrossRef]
8. Fuller, D.Q. Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the Old World. *Ann. Bot.* **2007**, *100*, 903–924. [CrossRef] [PubMed]
9. Cockram, J.; Jones, H.; Leigh, F.; O'Sullivan, D.; Powell, W.; Laurie, D.A.; Greenland, A.J. Control of flowering time in temperate cereals: Genes, domestication, and sustainable productivity. *J. Exp. Bot.* **2007**, *58*, 1231–1244. [CrossRef]
10. Machado, B.; Nogueira, A.; Hamawaki, O.; Rezende, G.; Jorge, G.; Silveira, I.; Medeiros, L.; Hamawaki, R.; Hamawaki, C. Phenotypic and genotypic correlations between soybean agronomic traits and path analysis. *Genet. Mol. Res.* **2017**, *16*, gmr16029696. [CrossRef]
11. Hartman, G.L.; West, E.D.; Herman, T. Crops that feed the World 2. Soybean—Worldwide production, use, and constraints caused by pathogens and pests. *Food Secur.* **2011**, *3*, 5–17. [CrossRef]
12. Bernard, R. Two major genes for time of flowering and maturity in soybeans 1. *Crop Sci.* **1971**, *11*, 242–244. [CrossRef]
13. Cober, E.R.; Voldeng, H.D. A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. *Crop Sci.* **2001**, *41*, 698–701. [CrossRef]
14. Buzzell, R. Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Can. J. Genet. Cytol.* **1971**, *13*, 703–707. [CrossRef]
15. Buzzell, R.; Voldeng, H. Inheritance of insensitivity to long daylength. *Soybean Genet. Newsl.* **1980**, *7*, 26–29.
16. McBlain, B.; Bernard, R. A new gene affecting the time of flowering and maturity in soybeans. *J. Hered.* **1987**, *78*, 160–162. [CrossRef]
17. Bonato, E.R.; Vello, N.A. E6, a dominant gene conditioning early flowering and maturity in soybeans. *Genet. Mol. Biol.* **1999**, *22*, 229–232. [CrossRef]
18. Cober, E.R.; Molnar, S.J.; Charette, M.; Voldeng, H.D. A new locus for early maturity in soybean. *Crop Sci.* **2010**, *50*, 524–527. [CrossRef]
19. Kong, F.; Nan, H.; Cao, D.; Li, Y.; Wu, F.; Wang, J.; Lu, S.; Yuan, X.; Cober, E.R.; Abe, J.; et al. A new dominant gene E9 conditions early flowering and maturity in soybean. *Crop Sci.* **2014**, *54*, 2529–2535. [CrossRef]
20. Cober, E.; Tanner, J.; Voldeng, H. Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. *Crop Sci.* **1996**, *36*, 601–605. [CrossRef]
21. Cober, E.; Tanner, J.; Voldeng, H. Soybean photoperiod-sensitivity loci respond differentially to light quality. *Crop Sci.* **1996**, *36*, 606–610. [CrossRef]
22. Kong, F.; Liu, B.; Xia, Z.; Sato, S.; Kim, B.M.; Watanabe, S.; Yamada, T.; Tabata, S.; Kanazawa, A.; Harada, K.; et al. Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. *Plant Physiol.* **2010**, *154*, 1220–1231. [CrossRef] [PubMed]

23. Thakare, D.; Kumudini, S.; Dinkins, R. The alleles at the E1 locus impact the expression pattern of two soybean FT-like genes shown to induce flowering in Arabidopsis. *Planta* **2011**, *234*, 933–943. [CrossRef]
24. Watanabe, S.; Harada, K.; Abe, J. Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breed. Sci.* **2012**, *61*, 531–543. [CrossRef] [PubMed]
25. Wang, X.; Li, Y.; Zhang, H.; Sun, G.; Zhang, W.; Qiu, L. Evolution and association analysis of GmCYP78A10 gene with seed size/weight and pod number in soybean. *Mol. Biol. Rep.* **2015**, *42*, 489–496. [CrossRef] [PubMed]
26. Josephs, E.B.; Stinchcombe, J.; Wright, S. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytol.* **2017**, *214*, 21–33. [CrossRef] [PubMed]
27. Yu, J.; Zhao, W.; Tong, W.; He, Q.; Yoon, M.-Y.; Li, F.-P.; Choi, B.; Heo, E.-B.; Kim, K.-W.; Park, Y.-J. A genome-wide association study reveals candidate genes related to salt tolerance in rice (*Oryza sativa*) at the germination stage. *Int. J. Mol. Sci.* **2018**, *19*, 3145. [CrossRef]
28. Sheoran, S.; Jaiswal, S.; Kumar, D.; Raghav, N.; Sharma, R.; Pawar, S.; Paul, S.; Iquebal, M.A.; Jaiswar, A.; Sharma, P.; et al. Uncovering genomic regions associated with 36 agro-morphological traits in Indian spring wheat using GWAS. *Front. Plant Sci.* **2019**, *10*, 527. [CrossRef]
29. Wang, M.; Yan, J.; Zhao, J.; Song, W.; Zhang, X.; Xiao, Y.; Zheng, Y. Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant Sci.* **2012**, *196*, 125–131. [CrossRef]
30. Zeng, A.; Chen, P.; Korth, K.; Hancock, F.; Pereira, A.; Brye, K.; Wu, C.; Shi, A. Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. *Mol. Breed.* **2017**, *37*, 30. [CrossRef]
31. Song, J.; Sun, X.; Zhang, K.; Liu, S.; Wang, J.; Yang, C.; Jiang, S.; Siyal, M.; Li, X.; Qi, Z.; et al. Identification of QTL and genes for pod number in soybean by linkage analysis and genome-wide association studies. *Mol. Breed.* **2020**, *40*, 60. [CrossRef]
32. Sonah, H.; O'Donoughue, L.; Cober, E.; Rajcan, I.; Belzile, F. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* **2015**, *13*, 211–221. [CrossRef]
33. Hwang, E.-Y.; Song, Q.; Jia, G.; Specht, J.E.; Hyten, D.L.; Costa, J.; Cregan, P.B. A genome-wide association study of seed protein and oil content in soybean. *BMC Genom.* **2014**, *15*, 1. [CrossRef]
34. Lee, Y.-G.; Jeong, N.; Kim, J.H.; Lee, K.; Kim, K.H.; Pirani, A.; Ha, B.-K.; Kang, S.-T.; Park, B.-S.; Moon, J.-K.; et al. Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant J.* **2015**, *81*, 625–636. [CrossRef] [PubMed]
35. Jeong, N.; Kim, K.-S.; Jeong, S.; Kim, J.-Y.; Park, S.-K.; Lee, J.S.; Jeong, S.-C.; Kang, S.-T.; Ha, B.-K.; Kim, D.-Y.; et al. Korean soybean core collection: Genotypic and phenotypic diversity population structure and genome-wide association study. *PLoS ONE* **2019**, *14*, e0224074. [CrossRef] [PubMed]
36. Kim, K.H.; Kim, J.-Y.; Lim, W.-J.; Jeong, S.; Lee, H.-Y.; Cho, Y.; Moon, J.-K.; Kim, N. Genome-wide association and epistatic interactions of flowering time in soybean cultivar. *PLoS ONE* **2020**, *15*, e0228114. [CrossRef]
37. Lee, T.; Kim, K.; Kim, J.-M.; Shin, I.; Heo, J.; Jung, J.; Lee, J.; Moon, J.-K.; Kang, S. Genome-Wide Association Study for Ultraviolet-B Resistance in Soybean (*Glycine max* L.). *Plants* **2021**, *10*, 1335. [CrossRef]
38. Lee, S.-B.; Lee, K.-S.; Kim, H.-Y.; Kim, D.-Y.; Seo, M.-S.; Jeong, S.-C.; Moon, J.-K.; Park, S.-K.; Choi, M.-S. The discovery of novel SNPs associated with group a soyasaponin biosynthesis from Korea soybean core collection. *Genomics* **2022**, *114*, 110432. [CrossRef]
39. Kim, B.; Dai, X.; Zhang, W.; Zhuang, Z.; Sanchez, D.L.; Lübberstedt, T.; Kang, Y.; Udvardi, M.K.; Beavis, W.D.; Xu, S.; et al. GWASpro: A high-performance genome-wide association analysis server. *Bioinformatics* **2019**, *35*, 2512–2514. [CrossRef]
40. Wickham, H.; François, R.; Henry, L.; Müller, K.; Vaughan, D. dplyr: A Grammar of Data Manipulation. R Package Version 0.7. 6. *Comput. Softw.* 2018. Available online: https://CRAN.R-project.org/package=dplyr (accessed on 19 January 2023).
41. Browning, B.L.; Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **2009**, *84*, 210–223. [CrossRef]
42. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **2011**, *12*, 246. [CrossRef] [PubMed]
43. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef] [PubMed]
44. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef] [PubMed]
45. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef] [PubMed]
46. Guo, Y.; Huang, Y.; Hou, L.; Ma, J.; Chen, C.; Ai, H.; Huang, L.; Ren, J. Genome-wide detection of genetic markers associated with growth and fatness in four pig populations using four approaches. *Genet. Sel. Evol.* **2017**, *49*, 21. [CrossRef]
47. Kuroda, Y.; Tomooka, N.; Kaga, A.; Wanigadeva, S.M.S.W.; Vaughan, D.A. Genetic diversity of wild soybean (*Glycine soja* Sieb. et Zucc.) and Japanese cultivated soybeans [*G. max* (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. *Genet. Resour. Crop Evol.* **2009**, *56*, 1045–1055. [CrossRef]
48. Sulistyo, A.; Sari, K. Correlation, path analysis and heritability estimation for agronomic traits contribute to yield on soybean. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2018.

49. Balla, M.Y.; Ibrahim, S.E. Genotypic correlation and path coefficient analysis of soybean [*Glycine max* (L.) Merr.] for yield and its components. *Agric. Res. Technol.* **2017**, *7*, 5557.

50. Li, M.; Liu, Y.; Wang, C.; Yang, X.; Li, D.; Zhang, X.; Xu, C.; Zhang, Y.; Li, W.; Zhao, L. Identification of traits contributing to high and stable yields in different soybean varieties across three Chinese latitudes. *Front. Plant Sci.* **2020**, *10*, 1642. [CrossRef]

51. Zhang, J.; Song, Q.; Cregan, P.B.; Nelson, R.L.; Wang, X.; Wu, J.; Jiang, G.-L. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genom.* **2015**, *16*, 217. [CrossRef]

52. Zhou, Z.; Jiang, Y.; Wang, Z.; Gou, Z.; Lyu, J.; Li, W.; Yu, Y.; Shu, L.; Zhao, Y.; Ma, Y.; et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **2015**, *33*, 408–414. [CrossRef]

53. Kim, M.-S.; Lozano, R.; Kim, J.H.; Bae, D.N.; Kim, S.-T.; Park, J.-H.; Choi, M.S.; Kim, J.; Ok, H.-C.; Park, S.-K.; et al. The patterns of deleterious mutations during the domestication of soybean. *Nat. Commun.* **2021**, *12*, 97. [CrossRef]

54. Taylor, C.M.; Garg, G.; Berger, J.D.; Ribalta, F.M.; Croser, J.S.; Singh, K.B.; Cowling, W.A.; Kamphuis, L.G.; Nelson, M.N. A Trimethylguanosine Synthase1-like (TGS1) homologue is implicated in vernalisation and flowering time control. *Theor. Appl. Genet.* **2021**, *134*, 3411–3426. [CrossRef]

55. Shi, S.; Duan, G.; Li, D.; Wu, J.; Liu, X.; Hong, B.; Yi, M.; Zhang, Z. Two-dimensional analysis provides molecular insight into flower scent of Lilium 'Siberia'. *Sci. Rep.* **2018**, *8*, 5352. [CrossRef] [PubMed]

56. Campobenedetto, C.; Mannino, G.; Agliassa, C.; Acquadro, A.; Contartese, V.; Garabello, C.; Bertea, C.M. Transcriptome analyses and antioxidant activity profiling reveal the role of a lignin-derived biostimulant seed treatment in enhancing heat stress tolerance in soybean. *Plants* **2020**, *9*, 1308. [CrossRef] [PubMed]

57. Spanudakis, E.; Jackson, S. The role of microRNAs in the control of flowering time. *J. Exp. Bot.* **2014**, *65*, 365–380. [CrossRef] [PubMed]

58. Ding, X.; Ruan, H.; Yu, L.; Li, Q.; Song, Q.; Yang, S.; Gai, J. miR156b from soybean CMS line modulates floral organ development. *J. Plant Biol.* **2020**, *63*, 141–153. [CrossRef]

59. Jaspert, N.; Throm, C.; Oecking, C. Arabidopsis 14-3-3 proteins: Fascinating and less fascinating aspects. *Front. Plant Sci.* **2011**, *2*, 96. [CrossRef] [PubMed]

60. Pnueli, L.; Gutfinger, T.; Hareven, D.; Ben-Naim, O.; Ron, N.; Adir, N.; Lifschitz, E. Tomato SP-interacting proteins define a conserved signaling system that regulates shoot architecture and flowering. *Plant Cell* **2001**, *13*, 2687–2702. [CrossRef] [PubMed]

61. Lifschitz, E.; Eviatar, T.; Rozman, A.; Shalit, A.; Goldshmidt, A.; Amsellem, Z.; Alvarez, J.P.; Eshed, Y. The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6398–6403. [CrossRef]

62. Purwestri, Y.A.; Ogaki, Y.; Tamaki, S.; Tsuji, H.; Shimamoto, K. The 14-3-3 protein GF14c acts as a negative regulator of flowering in rice by interacting with the florigen Hd3a. *Plant Cell Physiol.* **2009**, *50*, 429–438. [CrossRef]

63. Han, G.; Qiao, Z.; Li, Y.; Wang, C.; Wang, B. The roles of CCCH zinc-finger proteins in plant abiotic stress tolerance. *Int. J. Mol. Sci.* **2021**, *22*, 8327. [CrossRef]

64. Kazan, K.; Lyons, R. The link between flowering time and stress tolerance. *J. Exp. Bot.* **2016**, *67*, 47–60. [CrossRef] [PubMed]

65. Almagro, A.; Lin, S.; Tsay, Y. Characterization of the Arabidopsis nitrate transporter NRT1. 6 reveals a role of nitrate in early embryo development. *Plant Cell* **2008**, *20*, 3289–3299. [CrossRef] [PubMed]

66. Samanfar, B.; Molnar, S.J.; Charette, M.; Schoenrock, A.; Dehne, F.; Golshani, A.; Belzile, F.; Cober, E.R. Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theor. Appl. Genet.* **2017**, *130*, 377–390. [CrossRef] [PubMed]

67. Ma, X.; Cheng, Z.; Qin, R.; Qiu, Y.; Heng, Y.; Yang, H.; Ren, Y.; Wang, X.; Bi, J.; Ma, X.; et al. O s ARG encodes an arginase that plays critical roles in panicle development and grain production in rice. *Plant J.* **2013**, *73*, 190–200. [CrossRef] [PubMed]

68. Baldoni, A.; Von Pinho, E.; Fernandes, J.; Abreu, V.; Carvalho, M. Gene expression in the lignin biosynthesis pathway during soybean seed development. *Genet. Mol. Res.* **2013**, *12*, 2618–2624. [CrossRef]

69. Yan, Q.; Si, J.; Cui, X.; Peng, H.; Chen, X.; Xing, H.; Dou, D. The soybean cinnamate 4-hydroxylase gene GmC4H1 contributes positively to plant defense via increasing lignin content. *Plant Growth Regul.* **2019**, *88*, 139–149. [CrossRef]

70. Bhadouria, J.; Singh, A.P.; Mehra, P.; Verma, L.; Srivastawa, R.; Parida, S.K.; Giri, J. Identification of purple acid phosphatases in chickpea and potential roles of CaPAP7 in seed phytate accumulation. *Sci. Rep.* **2017**, *7*, 11012. [CrossRef]