Mol Breeding (2022) 42: 38
https://doi.org/10.1007/s11032-022-01310-y

Identification of major genomic regions for soybean seed weight by genome-wide association study

Yongce Cao · Shihao Jia · Liuxing Chen · Shunan Zeng · Tuanjie Zhao · Benjamin Karikari

Received: 15 February 2022 / Accepted: 12 June 2022 / Published online: 28 June 2022

Abstract The hundred-seed weight (HSW) is an important yield component and one of the principal breeding traits in soybean. More than 250 quantitative trait loci (QTL) for soybean HSW have been identified. However, most of them have a large genomic region or are environmentally sensitive, which provide limited information for improving the phenotype in marker-assisted selection (MAS) and identifying the candidate genes. Here, we utilized 281 soybean accessions with 58,112 single nucleotide polymorphisms (SNPs) to dissect the genetic basis of HSW in across years in the northern Shaanxi province of China through one single-locus (SL) and three multi-locus (ML) genome-wide association study (GWAS) models. As a result, one hundred and fifty-four SNPs were detected to be significantly associated with HSW in at least one environment via SL-GWAS model, and 27 of these 154 SNPs were detected in all (three) environments and located within 7 linkage disequilibrium (LD) block regions with the distance of each block ranged from 40 to 610 Kb. A total of 15 quantitative trait nucleotides (QTNs) were identified by three ML-GWAS models. Combined with the results of different GWAS models, the 7 LD block regions associated with HSW detected by SL-GWAS model could be verified directly or indirectly by the results of ML-GWAS models. Eleven candidate genes underlying the stable loci that may regulate seed weight in soybean were predicted. The significantly associated SNPs and the stable loci as well as predicted candidate genes may be of great importance for marker-assisted breeding, polymerization breeding, and gene discovery for HSW in soybean.

Keywords Hundred-seed weight · Association analysis · SNPs · LD block regions · Stable loci · Candidate genes

Yongce Cao and Shihao Jia have equal contributions

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11032-022-01310-y.

Y. Cao · S. Jia · L. Chen · S. Zeng
Shaanxi Key Laboratory of Chinese Jujube,
College of Life Science, Yan'an University, Yan'an,
Shaanxi 716000, China

T. Zhao (✉)
Key Laboratory of Biology and Genetic Improvement
of Soybean, Ministry of Agriculture, National Center
for Soybean Improvement, National Key Laboratory
for Crop Genetics and Germplasm Enhancement, Soybean
Research Institute of Nanjing Agricultural University,
Nanjing 210095, Jiangsu, China
e-mail: tjzhao@njau.edu.cn

B. Karikari (✉)
Department of Crop Science, Faculty of Agriculture, Food
and Consumer Sciences, University for Development
Studies, 00233 Tamale, Ghana
e-mail: bkarikari@uds.edu.gh

Introduction

Soybean (*Glycine max* [L] Merr.) is one of the most important legumes, commercially, and an essential dual-purpose crop with seeds enriched with proteins and oils that provide food and feed for human and livestock consumption, respectively (Gupta et al. 2021). In addition, its roots form a symbiotic association with soil microbes to fix atmosphere nitrogen making it useful in soil improvement programs (Chung and Singh 2008). The domestication of soybean from its wild progenitor (*G. soja* Sieb & Zucc.) started in temperate regions of China between 3000 and 9000 years ago (Hymowitz 1970; Carter et al. 2004; Lee et al. 2011). However, in recent years, soybean has become the most prominent agricultural product with the contradiction between supply and demand in China, and more than 80% of soybean and its products need to be imported every year to meet the domestic demand (FAOSTAT 2019; Sun et al. 2017). This calls for a breeding effort to improve soybean productivity in China to make it self-reliant.

The seed weight (usually expressed as over 100 seeds, designated as HSW) is an important yield component and one of the principal breeding traits in soybean. It is not merely positively related to yield, but also often determines the final utilization of seeds (Hopper et al. 1979; Friedman and Brandon 2001; Clarke and Wiseman 2000). For instance, large-seeded cultivars are boiled as vegetable soybean (*nimame*) which is predominant in China, and demand for it has markedly increased globally over the past two decades (Liu et al. 2022). The HSW is a typical quantitative trait, which is controlled by polygenes with minor effects and highly influenced by the environment and its factors (such as temperature, light, and soil moisture) (Panthee et al. 2005; Liang et al. 2016; Wu et al. 2018). Therefore, dissecting the genes/molecular markers and understanding the genetic basis of HSW would be useful for the development of high-yield soybean cultivars.

More than 250 seed weight quantitative trait loci (QTL) were identified using different bi-parental mapping populations based on linkage analysis in the SoyBase databank (www.soybase.org) and several others being reported in recent years (Hina et al. 2020; Karikari et al. 2019; Beche et al. 2020; Kumawat and Xu 2021). Only a few genes controlling soybean seed weight have been predicted or cloned from QTL mapping. Lu et al. (2017) detected a major QTL for soybean HSW using a recombinant inbred line (RIL) population with high-density genetic maps derived from a cross between a cultivated soybean and a wild soybean. Within the QTL major region, a *phosphatase 2C-1* (*PP2C-1*) gene was found to control HSW in transgenic plants. Nguyen et al. (2021) used forward genetic methods and CRISPR/Cas9 gene editing and identified *GmKIX8-1* as the causative gene for a major QTL of seed weight that regulates the seed size in soybean. Huang et al. (2021) identified a stable QTL for HSW on chromosome 4 and predicted four seed weight–related candidate genes (*Glyma.04G047800*, *Glyma.04G051200*, *Glyma.04G062400*, and *Glyma.04G073900*). However, most of the reported QTL have not been used effectively to improve the phenotype in MAS and to identify the candidate genes for HSW. The main reasons for this limitation include a large genomic region due to the low density of molecular markers in the genetic maps or environmental sensitivity (Han et al. 2012; Niu et al. 2013; Li et al. 2020; Qi et al. 2020). Thus, identification of the major and stable locus within a small region for seed weight is important for facilitating the map-based cloning of genes and understanding the molecular mechanisms of soybean seed weight.

Association mapping is an effective method to analyze the genetic basis of complex traits, which can utilize a number of historic recombination events in a natural population and could significantly improve the resolution and accuracy compared with linkage analysis (Rafalski 2010; Ibrahim et al. 2020). The use of GWAS alternatively called association/disequilibrium mapping for complex traits in crops has gained prominence in recent years due to the development of sequencing technology to obtain the single nucleotide polymorphisms (SNPs) within the whole genome, lesser time for population assemble, and cost (Ibrahim et al. 2020). Recently, there are a few studies on GWAS for soybean seed weight. Zhang et al. (2021) identified a novel candidate gene (*SoyZH13_16G122400*) controlling soybean seed weight by comparative selective signature analysis and high-resolution GWAS, which can provide new information for seed weight breeding. Ikram et al. (2020) detected 43 stable quantitative trait nucleotides (QTNs) and predicted 36 potential candidate genes for HSW. Karikari et al. (2020) reported 39 and

209 SNPs by 2 SL-GWAS and 6 ML-GWAS models, respectively, and predicted 4 hub-genes among the candidate genes by gene co-expression. Zhang et al. (2016) detected 22 HSW QTL based on 309 soybean accessions with 31,045 SNPs, and predicted six important candidate genes. Zhao et al. (2019) based on 185 tested accessions identified 34 environmentally stable QTNs and found sixteen candidate genes were significantly associated with soybean HSW. These studies indicate that GWAS is an effective method for genetic basis analysis and candidate gene mining for complex traits.

In the northern Shaanxi province of China, soybean plays a major role in agricultural production. However, the ecological environment and soybean varieties in this region are different from other major soybean-producing areas in China, and less known about the genetics of HSW in this environment. Therefore in the present study, we utilized 281 diverse soybean accessions evaluated in 3 years (2018, 2019, and 2020), and performed one SL-GWAS model and three ML-GWAS models with 58,112 SNPs to identify major genomic regions for soybean HSW and use stable loci to mine the candidate genes for marker-assisted selection and gene cloning (functional validation), respectively. The results from this study would provide useful information and lay foundation for breeding desirable cultivars with preferred seed weight.

Materials and methods

Soybean materials and field experimental design

In the present study, an association panel consisting of 281 soybean accessions (comprising 4 landraces and 277 improved cultivars) was used to perform the GWAS for HSW. These accessions were collected from diverse soybean production areas of China and have been used in our previously published study (Cao et al. 2021). The association panel was grown from approximately May to October in a 1.5 m single row plot with a row spacing of 50 cm in three environments: the Yan'an Experimental Station (36°72′ N; 109°40′ E) of Yan'an Agricultural Science Institute in Yan'an, Shaanxi, China, in the year 2018, 2019, and 2020 (denoted as E1, E2, and E3, respectively). All accessions were planted in a completely randomized block design with three replications in each year (environment). Field management was conducted under normal conditions. Then, the seeds were harvested at maturity to evaluate the phenotype of HSW.

Phenotypic investigation and statistical analysis

For the HSW assessment, 100 healthy seeds with approximately 10% moisture content were randomly selected from each block to measure weight via electronic balance. The HSW for each accession in one planting environment was an average of three replications. The descriptive statistics of phenotypic data, such as mean, standard deviation, range, and coefficient of variation (CV %) for HSW in the association mapping population were calculated using the SPSS 20.0 software (SPSS Inc.; Chicago, IL, USA). To evaluate the effects of genotype, environment, and genotype by environment interaction on the HSW, an analysis of variance (ANOVA) for the joint environments was conducted using the SAS PROC mixed linear model program with random factors: genotypes, environments, replications within an environment, and the genotype by environment interaction. The broad-sense heritability (h^2) of HSW in association mapping population was computed using the following equation:

$$h^2 = {\sigma_g}^2/({\sigma_g}^2 + \sigma^2{}_{ge}/n + \sigma^2{}_e/nr)$$

where $\sigma_g{}^2$, $\sigma^2{}_{ge}$, $\sigma^2{}_e$, n, and r represent the genotypic variance, the variance of the genotype by environment interaction, the error variance, the number of environments, and the number of replications within an environment, respectively (Nyquist and Baker 1991).

SNP genotypic and GWAS

All 281 accessions were genotyped for SNP markers using a high-throughput genotyping platform previously described in Cao et al. (2021). Finally, 58,112 SNPs with minor allele frequencies (MAF) > 0.05 unevenly distributed on 20 chromosomes were used for genome association mapping. The details of SNP and population structure information for these 281 accessions are described in our previous study (Cao et al. 2021).

In this study, one SL-GWAS model and three ML-GWAS models were utilized to perform the association analysis. For the SL-GWAS model, the mixed linear model (MLM) with the principal component (PC) and kinship (K) matrix as fixed and random effects was performed using the TASSEL 5.0 software (Bradbury et al. 2007). To obtain more SNPs associated with HSW and examine their stability across the three environments, $-\log_{10}(P) \geq 3$ was used as a threshold to declare a significant association of SNPs with HSW in MLM in this study. In addition, three ML-GWAS models (mrMLM, FASTmrEMMA, and FASTmrMLM) were conducted with *mrMLM.GUI* package in R (https://cran.r-project.org/web/packages/mrMLM/index.html) to complement and validate the loci identified by the SL-GWAS model (Wang et al. 2016; Wen et al. 2018; Tamba and Zhang 2018). The screening criteria for marker-trait significant association was set with a logarithm of odds (LOD) value of more than 3.0 which has been proposed to balance high power and low false-positive rate (Zhang et al. 2019).

Stable locus screening and elite allele analysis

The environmental stability of the QTL/QTN is essential for use in candidate gene cloning and breeding programs. Thus, only the SNP detected in all experimental environments was considered a stable QTN in this study. Then, linkage disequilibrium (LD) block analysis was performed on regions around the stable QTN position by Haploview 4.2 software with default settings (Barrett et al. 2005); the genome-wide association study results with gene structure and linkage disequilibrium matrix were drawn by using *IntAssoPlot* package in R (He et al. 2020). The LD block which contains stable QTNs was considered a major and stable locus for soybean HSW. Based on the effect value and genotype information for the stable QTNs, the elite allele for HSW of the most significantly associated stable QTN and haplotypes at each LD block can be determined. The HSW distribution of the accessions with different alleles and haplotypes was further examined.

Candidate gene prediction within important loci

Potential candidate genes for soybean HSW underlying the environmentally stable QTNs (LD block regions) were predicted according to the functional annotations of genes in soybean and their homologous genes in *Arabidopsis*, the expression of potential candidate genes in different soybean tissues and development stages, and the available literature. The functional annotations and expression datasets of genes (Wm82.a2) were downloaded from SoyBase for downstream analyses.

Results

Phenotypic variation of HSW in the 281 soybean accessions

To evaluate phenotypic variation among the 281 soybean accessions, we investigated the HSW across three environments (E1, E2, and E3). The results of descriptive statistics and phenotypic distribution are shown in Table 1 and Fig. 1. HSW ranged from 8.50 to 36.02 g with the mean ± standard deviation of 20.19 ± 3.18 g across all environments with a CV of 15.77% (Table 1). Whereas the HSW in E1, E2, and E3 environments had a range of 6.75–38.11 g, 6.41–35.07 g, and 12.26–34.87 g, respectively

Table 1 Descriptive statistics of soybean 100-seed weight in 281 accessions in different environments

Environment[a]	Mean (g)	SD[b]	Max (g)	Min (g)	CV (%)[c]	h^2 (%)[d]
E1	20.98	3.60	38.11	6.75	17.16	93.16
E2	19.96	3.63	35.07	6.41	18.17	
E3	19.63	2.91	34.87	12.26	14.85	
Average	20.19	3.18	36.02	8.50	15.77	

[a] E1, E2, and E3 represent the test environment in 2018, 2019, and 2020 respectively

[b] SD represents standard deviation

[c] CV represents coefficient of variation

[d] $h.^2$ represents broad-sense heritability

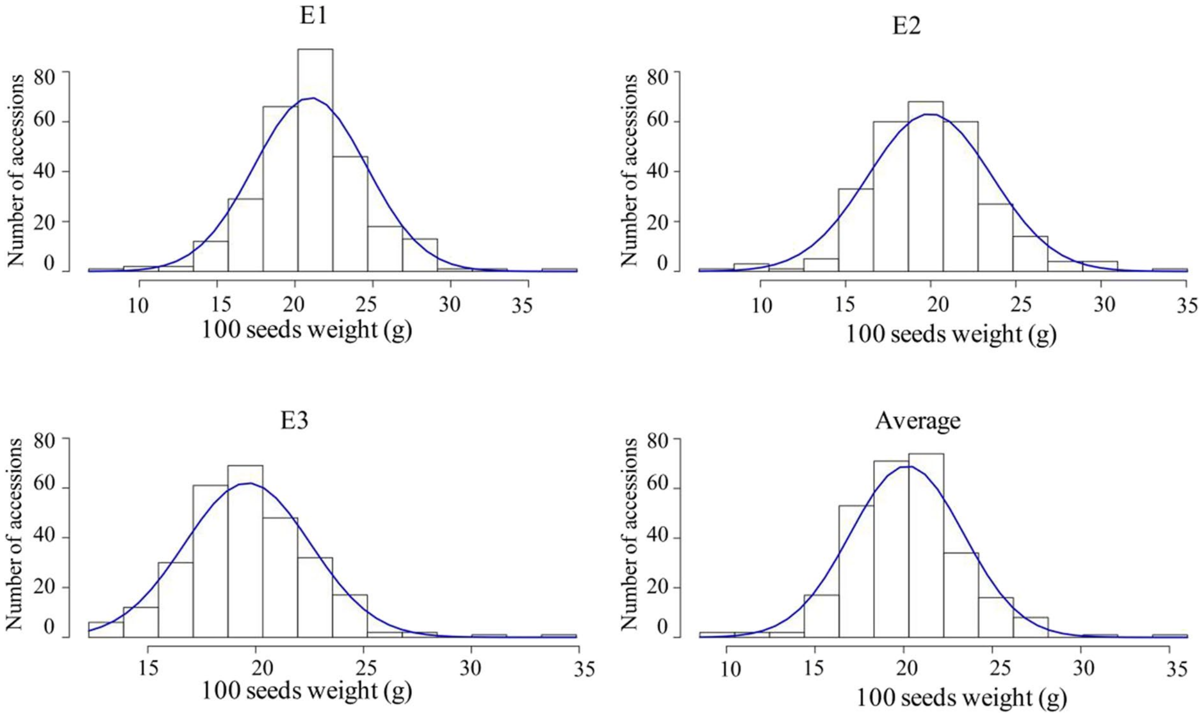

Fig. 1 Frequency distribution of HSW in three environments and average of three environments. Where E1, E2, E3, and average represent the test environment in 2018, 2019, 2020, and the average of three environments respectively

(Table 1). CV % ranged from 14.85 to 18.17% in the three environments. Across all environments (E1, E2, E3, and mean), the phenotypic frequency of HSW had a continuous and relatively normal distribution. These results pinpoint that the HSW varied widely in this population and accorded with the typical of quantitative traits (Fig. 1). The h^2 of HSW in 281 soybean accessions was high (93.16%) (Table 1), suggesting that HSW in this population is mainly controlled by genetic factors and less influenced by the environment, and its interaction with genotype (Table S1).

SNP-trait association mapping by SL- and ML-GWAS models

In this study, 281 soybean accessions with 58,112 high-quality SNPs (MAF > 0.05) were used to perform the association analyses. The PCA and K for this panel were calculated as previously described (Cao et al. 2021). In order to obtain accurate and stable QTNs for HSW, two contrasting types of GWAS, SL-GWAS, and ML-GWAS models were used. According to the quantile–quantile (Q-Q) plots from SL-GWAS results, when $-\log_{10}(P) \geq 3$, the observed value is greater than the expected value, indicating that these SNPs may be associated with HSW (Fig. 2).

The SL-GWAS model (MLM) with a threshold of $-\log_{10}(P) \geq 3$ detected a total of 75, 87, and 98 SNPs associated with HSW in the E1, E2, and E3, respectively (Table 2, Table S2, Fig. 2, Fig. 3). These SNPs were distributed on 14 chromosomes with exception chromosome (Chr.) 1 (Chr.01), Chr.03, Chr.06, Chr.08, and Chr.12 (Table 2, Fig. 2). The phenotypic variance explained (PVE/R^2) and allelic effect by a single SNP marker was between 3.99 to 8.81% and −4.24 to 4.24 g, respectively (Table S2). Cumulatively, 154 SNPs were detected across the three environments (E1, E2, and E3). In comparing the results across three environments, 75 of 154 SNPs were environmentally sensitive, 52 were detected in two environments, and 27 were detected in all environments (E1, E2, and E3). Therefore, the 27 SNPs could be considered stable SNPs (Table 2, Table 3, Fig. 2 and Fig. 3). Among the 27 environmentally stable SNPs, 1, 4, 1, 4, 3,

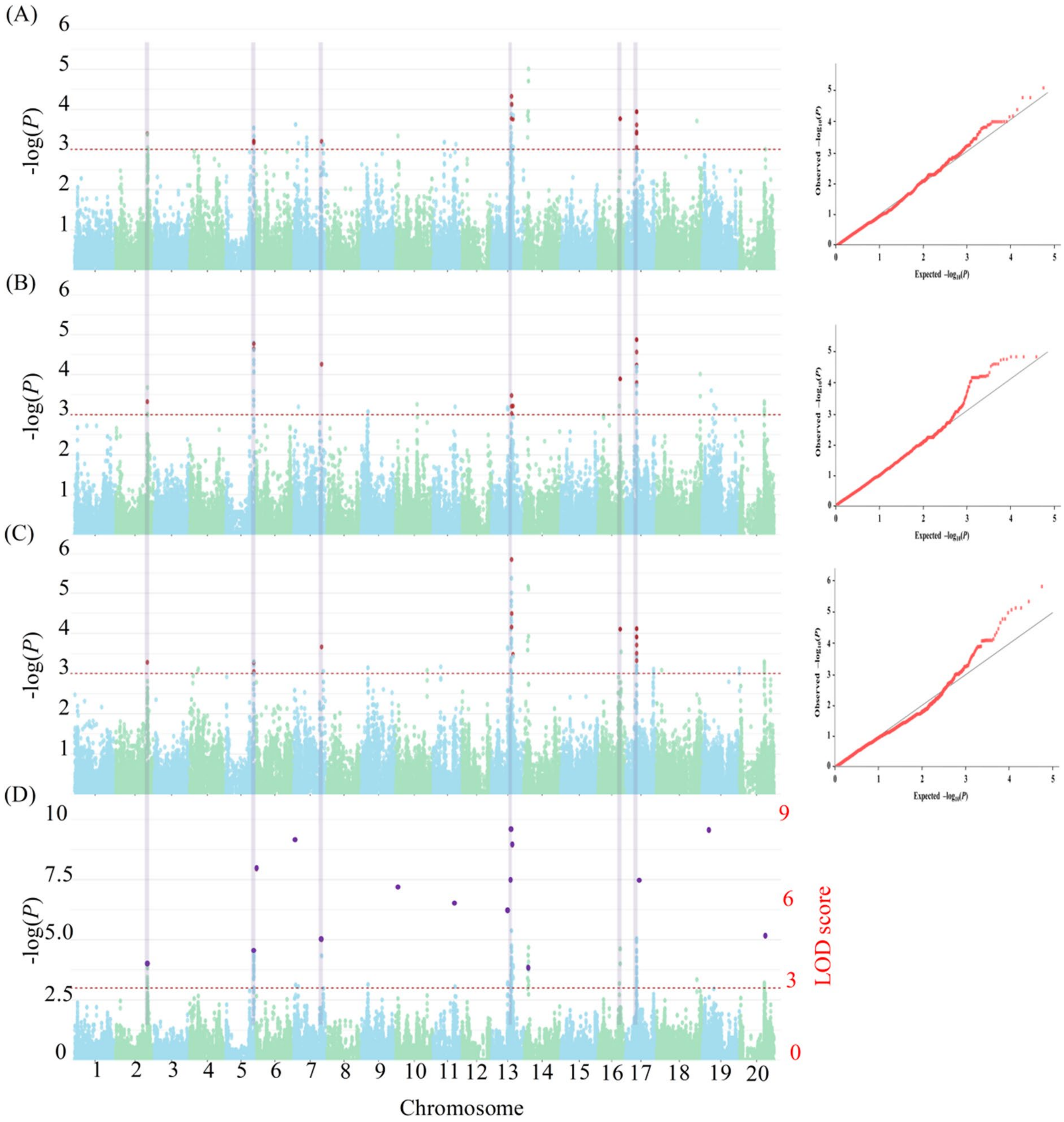

Fig. 2 Genome-wide association study (GWAS) results for HSW by SL-GWAS and ML-GWAS models. **A**, **B**, and **C** represent the GWAS and Q-Q plot results of the HSW in E1, E2, and E3 environments using the SL-GWAS model, respectively. The dashed horizontal line depicts a significant threshold level [$-\log_{10}(P) \geq 3$]. Red dots represent the SNPs that could be detected in all environments. **D** GWAS results of the HSW using three ML-GWAS models based on the average phenotypic values of the three environments. Purple dots represent detected QTNs

and 14 were distributed on Chr.02, Chr.05, Chr.07, Chr.13, Chr.16, and Chr.17 respectively. The LD block analyses were conducted for stable QTNs, and these 27 SNPs were located within 7 LD block regions with the distance of each block ranging from 40 to 610 Kb (Table 3, Fig. 4).

The average HSW phenotype of each soybean accession in three environments was used to conduct SNP-trait

Table 2 Summary of single-locus GWAS model (MLM) results in different environments

Chromosome	E1[a]	E2[a]	E3[a]	Single environment[b]	Two environments[b]	Three environments[b]
2	3	4	1	3	1	1
4	0	0	2	2	0	0
5	7	15	6	6	5	4
7	12	2	2	11	1	1
9	0	2	2	0	2	0
10	1	1	1	3	0	0
11	6	1	1	8	0	0
13	16	7	26	15	11	4
14	10	0	8	2	8	0
16	3	4	4	2	0	3
17	15	39	34	12	17	14
18	1	2	1	4	0	0
19	0	3	2	5	0	0
20	1	7	8	2	7	0
Total	75	87	98	75	52	27

[a] E1, E2, and E3 represent the test environment in 2018, 2019, and 2020 respectively

[b] Single environment, two environments, and three environments represent the number of SNPs detected in a single environment, two environments, and three environments respectively

association mapping by three ML-GWAS approaches (mrMLM, FASTmrMLM, and FASTmrEMMA). A total of 11, 9, and 4 significant QTNs with the LOD, QTN effect, and PVE range from 3.18 to 8.63, −2.25 to 1.98 g, and 2.22 to 11.95% were detected by mrMLM, FASTmrMLM, and FASTmrEMMA, respectively

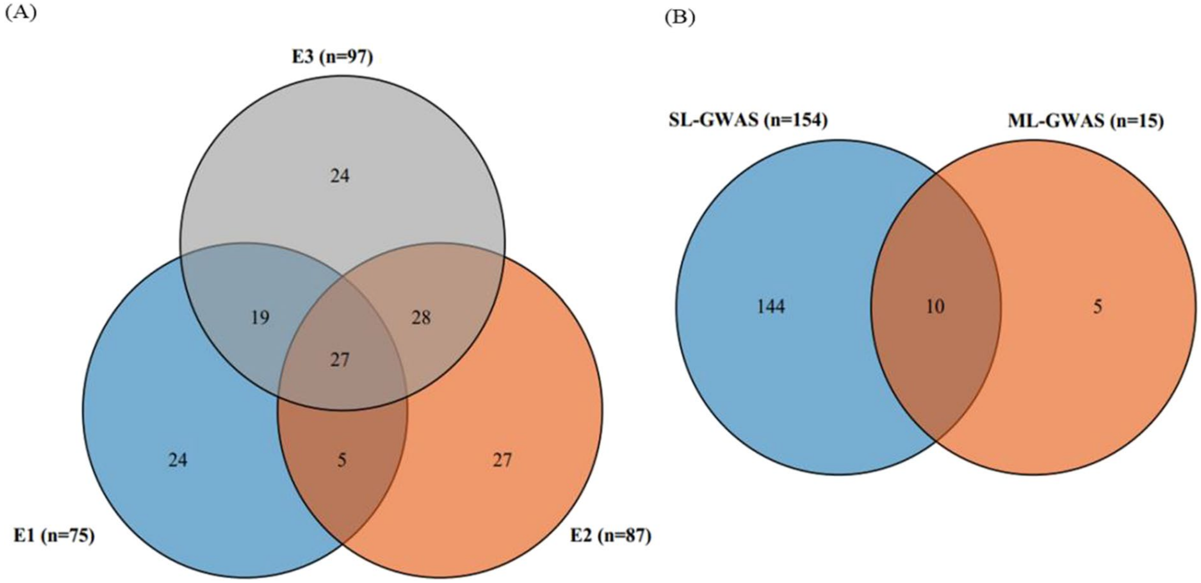

Fig. 3 Venn diagram of SNPs detected in this study. **A** SNPs by single-locus-GWAS (SL-GWAS) across the three environments (2018 [E1], 2019 [E2], and 2020 [E3]). **B** SNPs detected by both SL-GWAS and multi-locus-GWAS models

Table 3 Stable loci for soybean 100-seed weight detected in all experimental environments

Locus name	Chromosome	Number of stable SNPs	Genomic interval on Wm82.a1 (Wm82.a2) (Mb)[a]	– Log(P)	R^2 (%)[b]	References [c]
Locus 1	2	1	42.53–42.78 (39.47–39.72)	3.28–3.39	4.46–4.62	
Locus 2	5	4	38.13–38.57 (41.76–42.22)	3.05–4.77	4.08–6.89	Ikram et al. (2020)
Locus 3	7	1	37.71–37.99 (37.61–37.89)	3.20–4.26	4.31–6.04	Wang et al. (2015a, 2015b)
Locus 4	13	3	27.01–27.62 (28.21–28.82)	3.00–5.86	3.99–8.81	Hyten et al. (2004); Funatsuki et al. (2005); Wang et al. (2015a, 2015b); Kato et al. (2014)
Locus 5	13	1	29.26–29.30 (30.46–30.50)	3.22–3.76	4.33–5.21	Wang et al. (2015a, 2015b)
Locus 6	16	3	29.88–30.15 (30.22–30.52)	3.77–4.11	5.23–5.82	Karikari et al. (2020)
Locus 7	17	14	15.09–15.45 (14.86–15.22)	3.05–4.87	4.06–7.07	Teng et al. (2009)

[a]Represents the intervals of LD blocks containing stable SNPs. The data inside and outside the brackets represent the position of LD blocks on soybean Wm82.a1 and Wm82.a2 reference genome, respectively

[b]Represents the phenotypic variance (%) explained by the stable SNPs

[c]Represents that this locus overlaps or is close to the previously reported QTL/QTN position

(Table 4, Fig. 2D). Among these QTNs, 2 SNPs (Gm07_37852330 and Gm13_21923558) were detected by all ML-GWAS models, 5 SNPs (Gm06_155985, Gm07_2584000, Gm13_28334337, Gm19_9403950, and Gm20_35141779) were detected by two ML-GWAS models, and other 8 SNPs were only detected by one ML-GWAS model (Table 4).

We further compared the results of different GWAS models; among the 15 significant QTNs identified by the ML-GWAS models, 10 SNPs could be detected by the SL-GWAS model (Fig. 3B). Four (locus 1–4) of the seven stable loci (LD block regions) significantly associated with HSW detected by the SL model contained QTNs detected by the ML-GWAS models. Within the LD block 5 to 7 regions, no QTN was detected by the ML-GWAS models; however, in these three regions, the $-\log_{10}(P)$ values of some SNPs were greater than 4 (Fig. 2D), hence can be regarded as potential QTNs. The above results indicate that the results of the two models are largely consistent. However, the results of MLM (SL-GWAS) were not been subjected to the Bonferroni correction ($P < 0.05/58{,}112$); if this correction was performed, no significant SNP could be identified, suggesting that the ML-GWAS models seem to have higher accuracy and detection power.

Validation and allelic effect analysis of stable loci for soybean HSW

Depending on the results of GWAS, seven LD block regions were significantly associated with HSW in all experimental environments. Therefore, we focused on these seven intervals and considered them as important and environmentally stable loci controlling HSW in the studied panel and location. In order to validate these stable loci, the distribution of HSW was further examined in individuals with different alleles of the most significant SNPs in these loci (Fig. 5). Seven SNP markers: Gm02_42561640 (T/C), Gm05_38473956 (T/G), Gm07_37852330 (T/A), Gm13_27285558 (T/A), Gm13_29266613 (G/A), Gm16_30060864 (T/C), and Gm17_15123072 (C/A) were used to analyze the distribution of HSW with different alleles. The results showed that HSW differs significantly ($P < 0.05$) among the different alleles of the seven markers. For example, Gm02_42561640 (T/C) SNP with T allele accessions had an HSW of 20.60 g which was significantly higher ($P = 1.1 \times 10^{-6}$) than those with C allele (18.01 g). Similarly, Gm05_38473956 (T/G), Gm07_37852330 (T/A), Gm13_27285558 (T/A), Gm13_29266613 (G/A), Gm16_30060864 (T/C), and Gm17_15123072 (C/A), alleles T, A, A, A, C, and C, could increase HSW by 1.36, 3.04, 1.96, 1.44, 3.84, and 3.64 g, respectively, relative to their alternative alleles. In addition, for all the above seven SNPs, the HSW of allele TTAAACC was also significantly greater than that of allele CGAAACA. Furthermore, we also analyzed the distributions of HSW in different haplotypes at each LD block. Significant phenotypic differences between some haplotypes can be found in each locus ($P < 0.05$). For example, in locus 1, we extracted

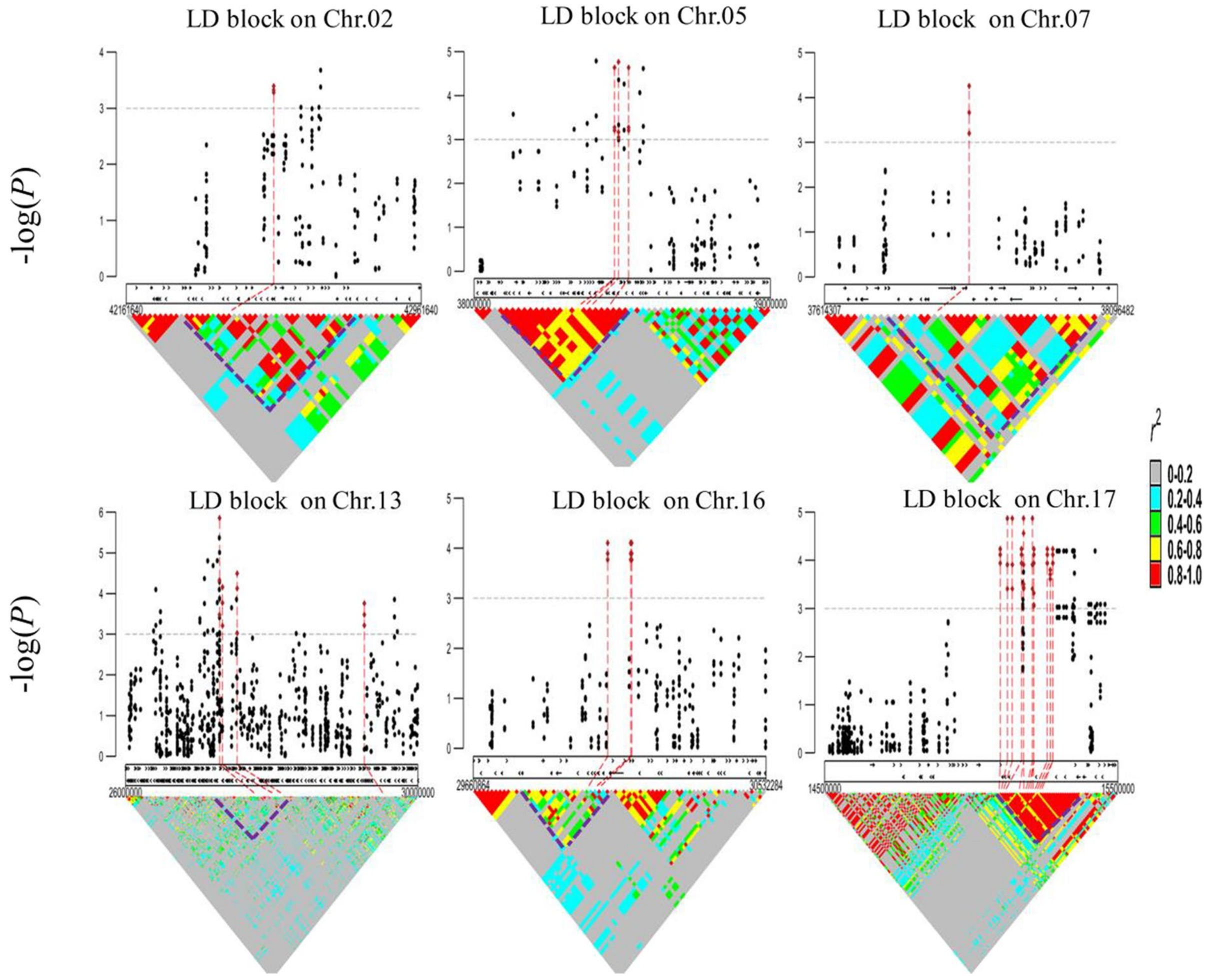

Fig. 4 Results of LD block analysis of the region around the position of stable QTN. Twenty-seven stable SNPs were located within 7 LD block regions with the distance of each block ranging from 40 to 610 Kb

the first four haplotypes and found that the average phenotypic value of soybeans with haplotype 1 was 21.18 g, which was significantly higher than that of accessions with haplotype 4 (18.08 g). Similar results were found at other LD blocks (Figure S1). Combining all 7 LD blocks, the average HSW phenotype of accessions with increasing haplotypes (25.13 g) was larger than that of accessions with decreasing haplotypes (18.87 g) (Figure S1). These results indicate that these stable loci could explain the genetic variation of HSW among the accessions, and the beneficial alleles from these LD block regions would be useful for MAS in soybean with high and stable HSW.

Prediction candidate genes of the stable loci for HSW in soybean

Potential candidate genes were mined from the 7 stable loci. A search for putative candidate genes resulted in a total of 217 genes within the 7 stable loci, of which 197 genes have homologous genes in *Arabidopsis*. According to the functional annotations, gene ontology (GO) enrichment analysis, and the expression data as well as available literature, the possible candidate genes were predicted. In all, 11 genes are involved in several biological processes such as gibberellin biosynthetic process, seed development, cell growth, embryo

Table 4 GWAS results of three multi-locus models based on the average HSW phenotype

QTN	Chr.[a]	Position on Wm82.a1 (bp)	Effect (g)[b]	LOD[c]	$-\text{Log}_{10}(P)$	R^2 (%)[d]	Model	References[e]
Gm02_42561640	2	42,561,640	1.89	4.26	5.02	4.72	FASTmrEMMA	
Gm05_38411981	5	38,411,981	0.71	4.83	5.62	4.68	mrMLM	Ikram et al. (2020)
Gm06_155985	6	155,985	0.84	6.63–7.07	7.48–7.93	6.03	mrMLM, FASTmrMLM	
Gm07_2584000	7	2,584,000	0.67–0.91	4.88–8.25	5.67–9.15	4.48–8.15	mrMLM, FASTmrMLM	Teng et al. (2009)
Gm07_37852330	7	37,852,330	−2.25–0.70	3.18–4.95	3.89–5.74	2.22–5.71	mrMLM, FASTmrMLM, FASTmrEMMA	Wang et al. (2015a, 2015b)
Gm10_4316320	10	4,316,320	1.17	6.26	7.10	4.03	mrMLM	
Gm11_29670751	11	29,670,751	0.97	5.59	6.41	2.78	FASTmrMLM	Specht et al. (2001)
Gm13_21923558	13	21,923,558	0.73–1.33	3.28–5.40	4.00–6.21	4.37–5.22	mrMLM, FASTmrMLM, FASTmrEMMA	Kato et al. (2014)
Gm13_27285558	13	27,285,558	0.71	6.63	7.48	4.56	mrMLM	Hyten et al. (2004); Funatsuki et al. (2005); Wang et al. (2015a, 2015b); Kato et al. (2014)
Gm13_27286305	13	27,286,305	1.98	8.60	9.51	9.68	FASTmrEMMA	Hyten et al. (2004); Funatsuki et al. (2005); Wang et al. (2015a, 2015b); Kato et al. (2014)
Gm13_28334337	13	28,334,337	−1.02–0.91	7.52–8.03	8.40–8.92	7.80–9.87	mrMLM, FASTmrMLM	
Gm14_6369300	14	6,369,300	−0.77	3.59	4.32	2.26	mrMLM	Liu et al. (2011)
Gm17_15976558	17	15,976,558	1.09	6.35	7.19	7.78	FASTmrMLM	Teng et al. (2009)
Gm19_9403950	19	9,403,950	−1.36–1.27	6.45–8.63	7.30–9.54	10.45–11.95	mrMLM, FASTmrMLM	Hoeck et al. (2003)
Gm20_35141779	20	35,141,779	−1.23–1.06	3.31–4.38	4.02–5.15	2.39–3.19	mrMLM, FASTmrMLM	Sun et al. (2012); Kato et al. (2014)

[a] Chr. represents chromosome. [b] Represents value of allelic effect. [c] Represents the log of odds (LOD) value of the QTN. [d] Represents the phenotypic variance (%) explained by each QTN. [e] Represents that this QTN overlaps or is close to the previously reported QTL/QTN position

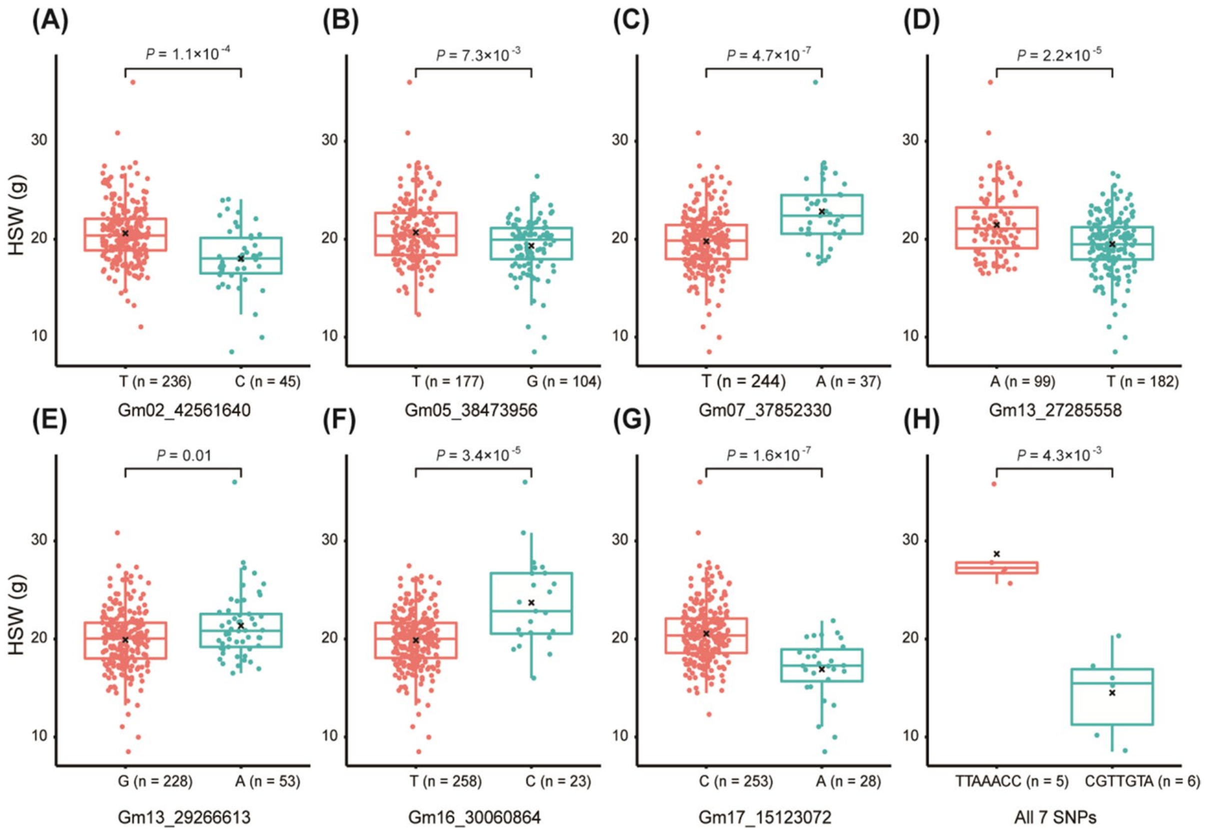

Fig. 5 The distributions of the HSW in individuals with different alleles of the most significant SNPs in stable loci. **A** to **G** are the results of HSW distribution at Gm02_42561640, Gm05_38473956, Gm07_37852330, Gm13_27285558, Gm13_29266613, Gm16_30060864, and Gm17_15123072, respectively. **H** is the result of the combination of all 7 SNPs. Where "*n*" represents the number of soybean accessions

development, and lipid metabolic process which may play key roles in regulating seed size/weight. Except *Glyma.02G209900*, the RNA-Seq data of 10 other prediction candidate genes are in Fig. 6 based on RNA-seq data obtained from Severin et al. (2010). Expression analysis for different soybean tissues showed that these 10 candidate genes have a high expression at one or various developmental stages in seed tissue (Fig. 6). Hence, these were considered to be the candidate genes for regulating HSW within the 7 stable loci (Table S3 and Fig. 6).

Discussion

Soybean seed weight is a typical quantitative trait, controlled by numerous genes with small effects, and significantly affected by environmental conditions (Panthee et al. 2005; Liang et al. 2016; Wu et al. 2018). It is difficult to improve soybean seed weight by traditional conventional breeding methods. MAS can effectively identify and track target genes by using markers closely linked to traits by reducing the number of lines to be tested in the selection and saves time in breeding programs (Hoeck et al. 2003; Mian et al. 1996; Kim et al. 2020). However, most of the reported QTL/QTNs were specific to an environment or located in large genomic intervals (Han et al. 2012; Niu et al. 2013; Zhang et al. 2016; Li et al. 2020; Zhao et al. 2019; Karikari et al. 2020). In contrast, breeders prefer to use stable QTL/QTNs with small genomic intervals in molecular-assisted breeding programs. Hence, it is imperative to identify the environmentally stable QTL/QTNs within a small region for seed weight for marker-assisted breeding.

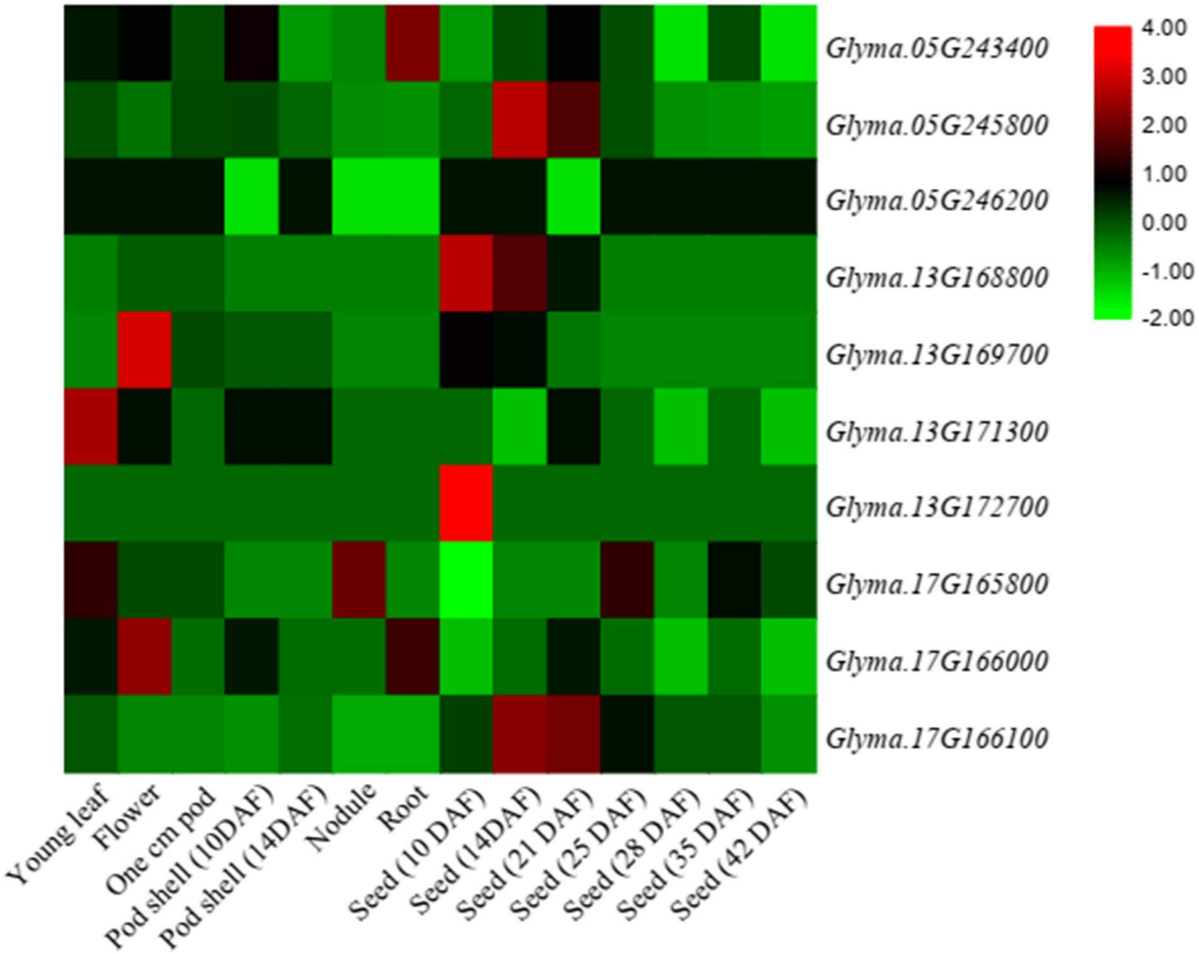

Fig. 6 Heat map exhibiting the expression profiles of 10 candidate genes among the different soybean tissues and development stages from the stable loci. DAF is days after flowering. RNA-seq data obtained from Severin et al. (2010) is available on the SoyBase website (https://soybase.org/, accessed on 12.12.2021)

Association mapping is an effective method to analyze the genetic basis of complex traits. Two main approaches of the GWAS are presented, one such approach is the SL-GWAS model (MLM) (Zhang et al. 2005; Yu et al. 2006), which has widely been used in the genetic dissection of complex traits in crops. But, this model usually requires Bonferroni's correction for multiple tests to reduce the false-positive rates between the markers and the phenotype, which may lead to the exclusion of important loci (Zhang et al. 2019). To overcome this issue, several ML-GWAS methods that do not need stringent Bonferroni's correction for multiple comparisons have been developed, such as mrMLM, FASTmrMLM, and FASTmrEMMA (Wang et al. 2016; Wen et al. 2018; Tamba and Zhang 2018). However, previous studies have demonstrated that none of the methods could identify all QTNs (Khan et al. 2019; Karikari et al. 2020; Ikram et al. 2020). Therefore, in order to obtain more possible related SNPs, it is necessary to combine the results from various GWAS models (Zhang et al. 2019). This strategy has been used in several studies (Khan et al. 2019; He et al. 2018; Xu et al. 2017; Karikari et al. 2020).

In this study, we used 281 soybean accessions with 58,112 SNPs to dissect the genetic basis of HSW by 2 contrasting GWAS approaches. For the SL-GWAS model, the main purpose was to obtain more possible associated SNPs and to examine the stability of significantly associated SNPs in different environments. According to the results of the Q-Q plot, $-\log_{10}(P) \geq 3$ was adopted as a threshold to declare a significant association of SNPs with HSW in the SL-GWAS model (MLM) (Fig. 2). As a result, 27 SNPs were distributed in 7 loci that could be detected in all experimental environments (Fig. 2, Fig. 3). For ML-GWAS models, the main purpose was to verify the important locus obtained from the SL-GWAS model. Finally, the 7 LD block regions associated with HSW detected by the single-locus model could be verified directly or indirectly by the results of the ML-GWAS models (Table 4, Fig. 3B). The results show that the results of the two models are largely consistent, indicating these seven loci are major genomic regions for controlling HSW in soybean. In addition, the allelic effect analysis was performed on two alleles of peak SNP from each of the seven LD blocks. The results showed that the HSW of soybean accessions with n excellent allele was significantly higher than that of accessions with another allele for all the 7 SNP peaks, and in particular the allelic effects of 7 SNPs could also be accumulated (Fig. 5). Phenotypic analysis based on the haplotypes of LD blocks had similar results (Fig. S1). These results provide useful information for soybean HSW molecular-assisted breeding and polymerization breeding in this region.

We further compared our results to already known loci/QTNs, 6 loci and 12 QTNs located near or overlapped with the genomic regions of known QTL/QTN for soybean HSW (Tables 2 and 3), giving credence to our results. For example, the position of QTN Gm05_38411981 is similar to the one identified by Ikram et al. (2020) at 38,490,635 bp on Chr.05. Locus 3 and locus 4 overlapped with the genomic regions of *Seed weight 42–5* and *Seed weight 15–3* reported by Wang et al. (2015a) and Hyten et al. (2004), respectively. However, except for locus 4, all loci in this study were detected in a small physical interval (< 500 Kb). For instance, *Seed weight 42–5* was located in a large genomic interval (greater than 2.3 Mb) between marker *Sat_121* and marker *Satt210* in SoyBase, but in the current study, locus 3 is located in an approximately 280 Kb genomic interval (Table 3). *Seed weight 15–3* was detected in approximately 4.7 Mb genomic interval between marker *Sat_133* and marker *Satt334* and overlapped with locus 4 (610 Kb); the resolution of QTL interval is greatly improved in our study. Hence, these QTL/QTNs could be considered important targets for soybean HSW to clone candidate genes in future studies. In addition, one locus (locus 1) and three QTNs (Gm02_42561640, Gm06_155985, and Gm10_4316320) were newly identified in this study, which can add to the growing knowledge of the genetic control of HSW.

In flowering plants, a mature seed consists of the embryo, the endosperm, and the seed coat. These three major anatomical components of seed have different genetic compositions and are known to be controlled by many genes (Lafon-Placette and Kōhler 2014). However, compared with the molecular mechanisms underlying seed size/weight in *Arabidopsis* and rice (Li and Li 2014; Zhu et al. 2015; Ren et al. 2019), only a few genes controlling HSW have been cloned and predicted in soybean (Lu et al. 2017; Nguyen et al. 2021; Lu et al. 2016; Du et al. 2017; Gu et al. 2017; Di et al. 2015). Based on gene function annotation, GO, pathway analysis, and gene expression data as well as related literature, 11 candidate genes regulating soybean seed weight were predicted in this study. *Glyma.02G209900* within locus 1 encoding a soluble cytoplasmic pyrophosphatase is homologous to *AT1G01050*, which can regulate the oil biosynthesis in developing seeds and thus affects seed weight, indicating that it is a candidate gene for soybean HSW (Meyer et al. 2012). *Glyma.05G243400* underlying the stable locus 2 is homologous to *AT1G18070*, which annotates to be involved in the translation elongation factor EF1A/initiation factor IF2gamma family protein in *Arabidopsis*, and it has been reported that this gene may be correlated with the regulation of seed size/weight in soybean (Li et al. 2019). *Glyma.13G169700* is a sucrose efflux transporter gene homologous to the *AtSWEET5* gene (*AT5G62850*). *AtSWEET5* belongs to SWEET (sugars will eventually be exported transporter) family. A couple of genes from the SWEET family have been demonstrated to regulate seed size/weight and other seed quality traits in soybean and other crops (Wang et al. 2015b; Chen et al. 2015; Wang et al. 2020; Miao et al. 2020). *Glyma.13G171300* is homologous to *AT5G45920*, which co-expressed with ABSCISIC ACID INSENSITIVE 4 and involved in seed storage metabolism (Yan and Chen 2017). *Glyma.13G172700* is homologous to *WUSCHEL RELATED HOMEOBOX 9* (*WOX9*) in *Arabidopsis*. *WOX9 is* required for maintaining cell division and essential for *Arabidopsis* embryonic development (Wu et al. 2007). *Glyma.17G165800* is a pentatricopeptide repeat (PPR) protein gene. PPR proteins have been shown to affect seed development and size in rice and maize (Huang et al. 2020; Li et al. 2014). *Glyma.17G166000* encodes geranyl diphosphate synthase 1 (GPS1) and is homologous to *AT2G34630*. *GPS* gene was found to be involved in gibberellin biosynthesis (Van-Schie et al. 2007). It is well known that gibberellin plays an important role in regulating seed development (Daviere and Achard 2013). Recently, a key enzyme gene in the gibberellin synthesis pathway (*GmGA3ox1*) was also cloned, which has an important impact on soybean yield (Hu et al. 2022). *Glyma.17G166100* annotates to encode GDSL-like lipase, which plays an important role in oil synthesis and seed development of oil and other plants (Ding et al. 2019; Clauss et al. 2011).

In addition to the homologous genes of the above eight genes identified to be related to seed weight in *Arabidopsis* or other crops, three other candidate genes also deserve attention. *Glyma.05G245800* is associated with the biological process of cell proliferation and has high expression specifically in seed 14 DAF (days after flowering) and 21 DAF stages in seed tissues. *Glyma.13G168800* is highly expressed in the early stage of seed development (seed 10 DAF and seed 14 DAF), and the gene annotation information and GO enrichment analysis show that this gene is involved in anther development, pollination, vegetative to reproductive phase transition of the meristem, and lipid storage biological processes. *Glyma.05G246200* is associated with the biological process of seed development and is highly expressed in seed tissues at various seed developmental stages. This information indicates that these genes may play an important role in seed set and development and affect seed weight. Therefore, the above genes could be targeted for further screening and possible functional validation to deepen our understanding of regulating seed size/weight in soybean.

Conclusions

In this study, we performed a GWAS using 281 accessions and 58,112 SNPs for dissecting the genetic architecture of HSW in soybean. As a result, a total of 154 SNPs and 15 QTNs significantly associated with HSW were dissected by the SL- and ML-GWAS models, respectively. Combined with the results of different GWAS models and the LD block and allelic effect analyses, 7 LD block regions were considered major genomic regions for controlling soybean HSW. And 11 candidate genes underlying the major genomic regions that may regulate seed weight in soybean were predicted. The significantly associated SNPs and the stable loci as well as predicted candidate genes might be of great usefulness for marker-assisted breeding and gene discovery for HSW in soybean.

Author contribution T. Z. and Y. C. conceived and designed the project. Y. C. and S. J. performed the experiments. Y. C., B. K., S. J., L. C., and S. Z. analyzed the data. Y. C. and B. K. drafted the manuscript. T. Z. and B. K. revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding This research was funded by the National Natural Science Foundation of China (Grant Nos. 32160469), the Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2020JQ-795), the Young Talent fund of the University Association for Science and Technology in Shaanxi, China (2020205), the Special Scientific Research Project of Yan'an University (YDY2020-30), the Doctoral Research Startup Program of Yan'an University (YDBK2018-02), and the College Students' innovation Program of Yan'an University (D2021125).

Declarations

Ethics approval and consent to participate All of the authors have read and have abided by the statement of ethical standards for the manuscript submitted to *Molecular Breeding*.

Consent for publication All of the authors approved the manuscript published in *Molecular Breeding*.

Conflict of interest The authors declare no competing interests.

References

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2):263–265

Beche E, Gillman JD, Song Q, Nelson R, Beissinger T, Decker J, Shannon G, Scaboo A (2020) Nested association mapping of important agronomic traits in three interspecific soybean populations. Theor Appl Genet 133(3):1039–1054

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635

Cao Y, Zhang X, Jia S, Karikari B, Zhang M, Xia Z, Zhao T, Liang F (2021) Genome-wide association among soybean accessions for the genetic basis of salinity-alkalinity tolerance during germination. Crop Pasture Sci 72(4):255–267

Carter TE, Nelson RL, Sneller CH, Cui Z (2004) Genetic diversity in soybean. In: Boerma HR, Specht JE, eds Soybeans: improvement, production, and uses MadisonWI: American Society of Agronomy:303–416

Chen L, Lin I, Qu X, Sosso D, McFarlane H, Londoño A, Samuels A, Frommer W (2015) A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the Arabidopsis embryo. Plant Cell 27(3):607–619

Chung G, Singh RJ (2008) Broadening the genetic base of soybean: a multidisciplinary approach. Crit Rev Plant Sci 27(5):295–341

Clarke EJ, Wiseman J (2000) Developments in plant breeding for improved nutritional quality of soya beans I. Protein and amino acid content. The Journal of Agricultural Science 134(2):111–124

Clauss K, von Roepenack-Lahaye E, Böttcher C, Roth M, Welti R, Erban A, Kopka J, Scheel D, Milkowski C, Strack D (2011) Overexpression of sinapine esterase *BnSCE3* in oilseed rape seeds triggers global changes in seed metabolism. Plant Physiol 155(3):1127–1145

Daviere JM, Achard P (2013) Gibberellin Signaling in Plants. Development 140(6):1147–1151

Di S, Yan F, Rodas FR, Rodriguez TO, Murai Y, Iwashina T, Sugawara S, Mori T, Nakabayashi R, Yonekura-Sakakibara K, Saito K, Takahashi R (2015) Linkage mapping, molecular cloning and functional analysis of soybean gene *Fg3* encoding flavonol 3-O-glucoside/galactoside (1 → 2) glucosyltransferase. BMC Plant Biol 15(1):1–13

Ding L, Li M, Wang W, Cao J, Wang Z, Zhu K, Yang Y, Li Y, Tan X (2019) Advances in plant GDSL lipases: from sequences to functional mechanisms. Acta Physiol Plant 41(9):1–11

Du J, Wang S, He C, Zhou B, Ruan YL, Shou H (2017) Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. J Exp Bot 68(8):1955–1972

Food FAO. Agriculture Organization of the United Nations. Faostat. 2019. http://faostat.fao.org. Accessed 10 Dec 2020

Friedman M, Brandon DL (2001) Nutritional and health benefits of soy proteins. J Agric Food Chem 49(3):1069–1086

Funatsuki H, Kawaguchi K, Matsuba S, Sato Y, Ishimoto M (2005) Mapping of QTL associated with chilling tolerance during reproductive growth in soybean. Theor Appl Genet 111(5):851–861

Gu Y, Li W, Jiang H, Wang Y, Gao H, Liu M, Chen Q, Lai Y, He C (2017) Differential expression of a WRKY gene between wild and cultivated soybeans correlates to seed size. J Exp Bot 68(11):2717–2729

Gupta K, Mishra SK, Gupta S, Pandey S, Panigrahi J, Wani SH (2021) Functional role of miRNAs: key players in soybean improvement. Phyton 90(5):1339

Han Y, Li D, Zhu D, Li H, Li X, Teng W, Li W (2012) QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. Theor Appl Genet 125(4):671–683

He L, Xiao J, Rashid KY, Yao Z, Li P, Jia G, Wang X, Cloutier S, You FM (2018) Genome-wide association studies for pasmo resistance in flax (*Linum usitatissimum* L.). Front Plant Sci 9:1982

He F, Ding S, Wang H, Qin F (2020) *IntAssoPlot*: an R package for integrated visualization of genome-wide association study results with gene structure and linkage disequilibrium matrix. Front Genet 11:260

Hina A, Cao Y, Song S, Li S, Sharmin R, Elattar M, Bhat J, Zhao T (2020) High-resolution mapping in two ril populations refines major “QTL hotspot” regions for seed size and shape in soybean (*Glycine max* L.). Int J Mol Sci 21(3):1040

Hoeck JA, Fehr WR, Shoemaker RC, Welke GA, Johnson SL, Cianzio SR (2003) Molecular marker analysis of soybean seed size. Crop Sci 43:68–74

Hopper NW, Overholt JR, Martin JR (1979) Effect of cultivar, temperature and seed size on the germination and emergence of soya beans (*Glycine max* (L.) Merr.). Ann Bot 44(3):301–308

Hu D, Li X, Yang Z, Liu S, Hao D, Chao M, Zhang J, Yang H, Su X, Jiang M, Lu S, Zhang D, Wang L, Kan G, Wang H, Cheng H, Wang J, Huang F, Tian Z, Yu D (2022) Downregulation of a gibberellin 3b-hydroxylase enhances photosynthesis and increases seed yield in soybean. New Phytol 2022:1. https://doi.org/10.1111/nph.18153

Huang J, Lu G, Liu L, Raihan M, Xu J, Jian L, Zhao L, Tran T, Zhang Q, Liu J, Li W, Wei C, Braun D, Li Q, Fernie A, Jackson D, Yan J (2020) The kernel size-related quantitative trait locus *qKW9* encodes a pentatricopeptide repeat protein that affects photosynthesis and grain filling. Plant Physiol 183:1696–1709

Huang W, Hou J, Hu Q, An J, Zhang Y, Han Q, Li X, Wu Y, Zhang D, Wang J, Xu R, Li L, Sun L (2021) Pedigree-based genetic dissection of quantitative loci for seed quality and yield characters in improved soybean. Mol Breeding 41:14

Hymowitz T (1970) On the domestication of the soybean. Econ Bot 24(4):408–421

Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME (2004) Seed quality QTL in a prominent soybean population. Theor Appl Genet 109(3):552–561

Ibrahim AK, Zhang L, Niyitanga S, Afzal MZ, Xu Y, Zhang L, Zhang L, Qi J (2020) Principles and approaches of association mapping in plant breeding. Tropical Plant Biology 13(3):212–224

Ikram M, Han X, Zuo JF, Song J, Han CY, Zhang YW, Zhang YM (2020) Identification of QTNs and their candidate genes for 100-seed weight in soybean (*Glycine max* L.) using multi-locus genome-wide association studies. Genes 11(7):714

Karikari B, Chen S, Xiao Y, Chang F, Zhou Y, Kong J, Bhat J, Zhao T (2019) Utilization of interspecific high-density genetic map of RIL population for the QTL detection and candidate gene mining for 100-seed weight in soybean. Front Plant Sci 10:1001

Karikari B, Wang Z, Zhou Y, Yan W, Feng J, Zhao T (2020) Identification of quantitative trait nucleotides and candidate genes for soybean seed weight by multiple models of genome-wide association study. BMC Plant Biol 20(1):1–14

Kato S, Sayama T, Fujii K, Yumoto S, Kono Y, Hwang TY, Kikuchi A, Takada Y, Tanaka Y, Shiraiwa T, Ishimoto M (2014) A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. Theor Appl Genet 127(6):1365–1374

Khan SU, Yangmiao J, Liu S, Zhang K, Khan MHU, Zhai Y, Olalekan A, Fan C, Zhou Y (2019) Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in *Brassica napus* L. Ind Crops Prod 142:111877

Kim JM, Kim KH, Jung J, Kang BK, Lee J, Ha BK, Kang S (2020) Validation of marker-assisted selection in soybean breeding program for pod shattering resistance. Euphytica 216(11):1–9

Kumawat G, Xu D (2021) A major and stable quantitative trait locus *qSS2* for seed size and shape traits in a soybean RIL Population[J]. Front Genet 12:646102

Lafon-Placette C, Köhler C (2014) Embryo and endosperm, partners in seed development. Curr Opin Plant Biol 17:64–69

Lee GA, Crawford GW, Li L, Yuka S, Xuexiang C (2011) Archaeological soybean (*Glycine max*) in East Asia: does size matter? PLoS ONE 6(11):e26720

Li N, Li Y (2014) Ubiquitin-mediated control of seed size in plants. Front Plant Sci 5:332

Li X, Zhang Y, Hou M, Sun F, Shen Y, Xiu Z, Wang X, Chen Z, Sun S, Small I, Tan B (2014) *Small kernel 1* encodes a pentatricopeptide repeat protein required for mitochondrial *nad7* transcript editing and seed development in maize (*Zea mays*) and rice (*Oryza sativa*). Plant J 79(5):797–809

Li X, Zhang X, Zhu L, Bu Y, Wang X, Zhang X, Zhou Y, Wang X, Guo N, Qiu L, Zhao J, Xing H (2019) Genome-wide association study of four yield-related traits at the R6 stage in soybean. BMC Genet 20(1):1–15

Li M, Chen L, Zeng J, Razzaq MK, Xu X, Xu Y, Wang W, He J, Xing G, Gai J (2020) Identification of additive–epistatic QTLs conferring seed traits in soybean using recombinant inbred lines. Front Plant Sci 11:566056

Liang H, Xu L, Yu Y, Yang H, Dong W, Zhang H (2016) Identification of QTLs with main, epistatic and QTL by environment interaction effects for seed shape and hundred-seed weight in soybean across multiple years. J Genet 95(2):475–477

Liu W, Kim MY, Van K, Lee YH, Li H, Liu X, Lee SH (2011) QTL identification of yield-related traits and their association with flowering and maturity in soybean. J Crop Sci Biotechnol 14(1):65–70

Liu N, Niu Y, Zhang G, Feng Z, Bo Y, Lian J, Wang B, Gong Y (2022). Genome sequencing and population resequencing provide insights into the genetic basis of domestication and diversity of vegetable soybean.Horticulture Research, 9. https://doi.org/10.1093/hr/uhab052

Lu X, Li QT, Xiong Q, Li W, Bi YD, Lai Y, Man W, Zhang W, Ma B, Chen S, Zhang J (2016) The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. Plant J 86(6):530–544

Lu X, Xiong Q, Cheng T, Li QT, Liu XL, Bi Y, Li W, Zhang W, Ma B, Lai Y, Du W, Man W, Chen S, Zhang JS (2017) A *PP2C-1* allele underlying a quantitative trait locus enhances soybean 100-seed weight. Mol Plant 10(5):670–684

Meyer K, Stecca KL, Ewell-Hicks K, Allen S, Everard J (2012) Oil and protein accumulation in developing seeds is influenced by the expression of a cytosolic pyrophosphatase in *Arabidopsis*. Plant Physiol 159(3):1221–1234

Mian MAR, Bailey MA, Tamulonis JP, Shipe ER, Carter TE, Parrott WA, Ashley D, Hussey R, Boerma HR (1996) Molecular markers associated with seed weight in two soybean populations. Theor Appl Genet 93(7):1011–1016

Miao L, Yang S, Zhang K, He J, Wu C, Ren Y, Gai J, Li Y (2020) Natural variation and selection in *GmSWEET39* affect soybean seed oil content. New Phytol 225(4):1651–1666

Nguyen CX, Paddock KJ, Zhang Z, Stacey MG (2021) *GmKIX8y* regulates organ size in soybean and is the causative gene for the major seed weight QTL *qSw17la*. New Phytol 229(2):920–934

Niu Y, Xu Y, Liu X, Yang S, Wei S, Xie F, Zhang Y (2013) Association mapping for seed size and shape traits in soybean cultivars. Mol Breeding 31(4):785–794

Nyquist WE, Baker R (1991) Estimation of heritability and prediction of selection response in plant populations. Crit Rev Plant Sci 10(3):235–322

Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci 45(5):2015–2022

Qi Z, Song J, Zhang K, Liu S, Tian X, Wang Y, Fang Y, Li X, Wang J, Yang C, Jiang S, Sun X, Tian Z, Li W, Ning H (2020) Identification of QTNs controlling 100-seed weight in soybean using multilocus genome-wide association studies. Front Genet 11:689

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13(2):174–180

Ren D, Wang X, Yang M, Yang L, He G, Deng X (2019) A new regulator of seed size control in *Arabidopsis* identified by a genome-wide association study. New Phytol 222(2):895–906

Severin A, Woody J, Bolon Y, Joseph B, Diers B, Farmer A, Muehlbauer G, Nelson R, Grant D, Specht J, Graham M, Cannon S, May G, Vance C, Shoemaker R (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol 10(1):1–16

Specht J, Chase K, Macrander M, Graef G, Chung J, Markwell J, Germann M, Orf J, Lark K (2001) Soybean response to water: a QTL analysis of drought tolerance. Crop Sci 41(2):493–509

Sun Y, Pan J, Du Shi, X, Wu Q, Qi Z, Jiang H, Xin D, Liu C, Hu G, Chen Q (2012) Multi-environment mapping and meta-analysis of 100-seed weight in soybean. Mol Biol Rep 39(10):9435–9443

Sun J, Shu K, Zhuang W, Wang H, Xu X (2017) Analysis of China's soybean supply and demand based on international trade. Journal of Northeast Agricultural Sciences 42(6):64–68

Tamba C, Zhang Y (2018) A fast mrMLM algorithm for multi-locus genome-wide association studies. bioRxiv. https://doi.org/10.1101/341784

Teng W, Han Y, Du Y, Sun D, Zhang Z, Qiu L, Sun G, Li W (2009) QTL analyses of seed weight during the development of soybean (*Glycine max* L. Merr.). Heredity 102(4):372–380

Van-Schie C, Ament K, Schmidt A, Lange T, Haring M, Schuurink R (2007) Geranyl diphosphate synthase is required for biosynthesis of gibberellins. The Plant Journal 52(4):752–762

Wang J, Chen P, Wang D, Shannon G, Shi A, Zeng A, Orazaly M (2015a) Identification of quantitative trait loci for oil content in soybean seed. Crop Sci 55(1):23–34

Wang S, Yokosho K, Guo R, Whelan J, Ruan Y, Ma J, Shou H (2015b) The soybean sugar transporter *GmSWEET15* mediates sucrose export from endosperm to early embryo. Plant Physiol 180(4):2133–2141

Wang S, Feng J, Ren W, Huang B, Zhou L, Wen Y, Zhang J, Dunwell J, Xu S, Zhang Y (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep 6(1):1–10

Wang S, Liu S, Wang J, Yokosho K, Zhou B, Yu Y, Liu Z, Frommer W, Ma J, Chen L, Guan Y, Shou H, Tian Z (2020) Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. Natl Sci Rev 7(11):1776–1786

Wen Y, Zhang H, Ni Y, Huang B, Zhang J, Feng J, Wang S, Dunwell J, Zhang Y, Wu R (2018) Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Brief Bioinform 19(4):700–712

Wu X, Chory J, Weigel D (2007) Combinations of *WOX* activities regulate tissue proliferation during *Arabidopsis* embryonic development. Dev Biol 309(2):306–316

Wu D, Zhan Y, Sun Q, Xu L, Lian M, Zhao X, Han Y, Li W (2018) Identification of quantitative trait loci underlying soybean (*Glycine max* [L.] Merr.) seed weight including main, epistatic and QTL× environment effects in different regions of Northeast China. Plant Breed 137(2):194–202

Xu Y, Xu C, Xu S (2017) Prediction and association mapping of agronomic traits in maize using multiple omic data. Heredity 119(3):174–184

Yan A, Chen Z (2017) The pivotal role of abscisic acid signaling during transition from seed maturation to germination. Plant Cell Rep 36(5):689–703

Yu J, Pressoir G, Briggs W, Bi I, Yamasaki M, Doebley J, McMullen M, Gaut B, Nielsen D, Holland J, Kresovich S, Buckler E (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38(2):203–208

Zhang Y, Mao Y, Xie C, Smith H, Luo L, Xu S (2005) Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). Genetics 169(4):2267–2275

Zhang J, Song Q, Cregan P, Jiang G (2016) Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycinemax*). Theor Appl Genet 129(1):117–130

Zhang Y, Jia Z, Dunwell J (2019) The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. Front Plant Sci 10:100

Zhang W, Xu W, Zhang H, Liu X, Cui X, Li S, Song L, Zhu Y, Chen X, Chen X (2021) Comparative selective signature analysis and high-resolution GWAS reveal a new candidate gene controlling seed weight in soybean. Theor Appl Genet 134:1329–1341

Zhao X, Dong H, Chang H, Zhao J, Teng W, Qiu L, Li W, Han Y (2019) Genome wide association mapping and candidate gene analysis for hundred seed weight in soybean [*Glycine max* (L.) Merrill]. BMC Genomics 20(1):1–11

Zhu X, Liang W, Cui X, Chen M, Yin C, Luo Z, Zhu J, Lucas W, Wang Z, Zhang D (2015) Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein. Plant J 82(4):570–581