



## OPEN ACCESS

## EDITED BY

Jianghua Chua,  
Key Laboratory of Tropical Plant  
Resource and Sustainable Use (CAS),  
China

## REVIEWED BY

Zhaoming Qi,  
Northeast Agricultural University,  
China  
Hengyou Zhang,  
Northeast Institute of Geography and  
Agroecology (CAS), China

## \*CORRESPONDENCE

Yanbo Cheng

✉ ybcheng@scau.edu.cn

Hai Nian

✉ hnian@scau.edu.cn

†These authors share first authorship

## SPECIALTY SECTION

This article was submitted to  
Functional and Applied Plant  
Genomics,  
a section of the journal  
Frontiers in Plant Science

RECEIVED 09 November 2022

ACCEPTED 15 December 2022

PUBLISHED 11 January 2023

## CITATION

Luo S, Jia J, Liu R, Wei R, Guo Z, Cai Z,  
Chen B, Liang F, Xia Q, Nian H and  
Cheng Y (2023) Identification of major  
QTLs for soybean seed size and seed  
weight traits using a RIL population in  
different environments.  
*Front. Plant Sci.* 13:1094112.  
doi: 10.3389/fpls.2022.1094112

## COPYRIGHT

© 2023 Luo, Jia, Liu, Wei, Guo, Cai,  
Chen, Liang, Xia, Nian and Cheng. This  
is an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction  
in other forums is permitted, provided  
the original author(s) and the  
copyright owner(s) are credited and  
that the original publication in this  
journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is  
permitted which does not comply with  
these terms.

# Identification of major QTLs for soybean seed size and seed weight traits using a RIL population in different environments

Shilin Luo<sup>1,2,3†</sup>, Jia Jia<sup>1,2,3†</sup>, Riqian Liu<sup>1,2,3†</sup>, Ruqian Wei<sup>1,2,3</sup>,  
Zhibin Guo<sup>1,2,3</sup>, Zhandong Cai<sup>1,2,3</sup>, Bo Chen<sup>1,2,3</sup>, Fuwei Liang<sup>1,2,3</sup>,  
Qiuju Xia<sup>4</sup>, Hai Nian<sup>1,2,3\*</sup> and Yanbo Cheng<sup>1,2,3\*</sup>

<sup>1</sup>The State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, South China Agricultural University, Guangzhou, Guangdong, China, <sup>2</sup>The Key Laboratory of Plant Molecular Breeding of Guangdong Province, College of Agriculture, South China Agricultural University, Guangzhou, Guangdong, China, <sup>3</sup>Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, Guangdong, China, <sup>4</sup>Rice Molecular Breeding Institute, Granlux Associated Grains, Shenzhen, Guangdong, China

**Introduction:** The seed weight of soybean [*Glycine max* (L.) Merr.] is one of the major traits that determine soybean yield and is closely related to seed size. However, the genetic basis of the synergistic regulation of traits related to soybean yield is unclear.

**Methods:** To understand the molecular genetic basis for the formation of soybean yield traits, the present study focused on QTLs mapping for seed size and weight traits in different environments and target genes mining.

**Results:** A total of 85 QTLs associated with seed size and weight traits were identified using a recombinant inbred line (RIL) population developed from Guizao1xB13 (GB13). We also detected 18 environmentally stable QTLs. Of these, *qSL-3-1* was a novel QTL with a stable main effect associated with seed length. It was detected in all environments, three of which explained more than 10% of phenotypic variance (PV), with a maximum of 15.91%. In addition, *qSW-20-3* was a novel QTL with a stable main effect associated with seed width, which was identified in four environments. And the amount of phenotypic variance explained (PVE) varied from 9.22 to 21.93%. Five QTL clusters associated with both seed size and seed weight were summarized by QTL cluster identification. Fifteen candidate genes that may be involved in regulating soybean seed size and weight were also screened based on gene function annotation and GO enrichment analysis.

**Discussion:** The results provide a biologically basic reference for understanding the formation of soybean seed size and weight traits.

## KEYWORDS

soybean, seed size, seed weight, stable QTLs, QTL clusters

## 1 Introduction

Soybean is one of the globally esteemed food crops with numerous nutritional substances including protein, carbohydrates, lipids, minerals, vitamins, and bioactive substances such as isoflavones, saponins, sterols and phospholipids. With its vital values in agriculture and economy, soybean plays an important role in agricultural and industrial activities such as processing food and feed and producing textiles. While the demand for soybean has been increasing globally, soybean yield enhancement is now receiving significant attention for its potential for evolving productivity.

Seed size and seed weight are quantitative traits controlled by multiple genes and constrained by environmental factors. The forming of both traits is a complex biological process (Li et al., 2020). Seed size usually measure s seed length, seed width, and seed thickness (Salas et al., 2006). Many studies have been done worldwide on QTL localization of seed size and weight in soybean. In contrast, studies on QTL localization for seed size were relatively less than that for seed weight. For example, on the soybean data website SoyBase (<http://www.soybase.org/>) as of March 2022, the number of QTL publications are 18 on seed length, 16 on seed width, and 23 on seed thickness versus the number of QTL publications on seed weight are 300.

Salas et al. (2006) detected a total of 19 QTLs associated with seed shape using three RIL populations in two environments. Xu et al. (2011) used MAJ Multi-QTL Joint Analysis (MAJ) and composite interval mapping (CIM) for QTLs mapping. A total of 121 main QTLs were detected in the  $F_{2:3}$ ,  $F_{2:4}$  and  $F_{2:5}$  populations from the direct and reciprocal crosses of Lishuizhongzihuang with Nannong493-1. Sun et al. (2021) used a RIL population to detect QTLs for seed size traits in four environments. Ten QTLs controlling related traits were identified, of which five QTLs distributed on chromosomes 02, 04, 06, 13 and 16 were detected in at least two environments with PVE ranging from 3.6 to 9.4%. The results of similar studies (Salas et al., 2006; Xu et al., 2011; Hu et al., 2013; Niu et al., 2013; Yang et al., 2017; Teng et al., 2018; Gao et al., 2019; Cui et al., 2020; Sun et al., 2021; Kumawat and Xu, 2021) vary from different localization methods and different genetic backgrounds.

The first study on QTL localization of seed weight in bean was reported by Sax (1923), in which he used seed colors as the markers. With the advancement of molecular marker technology, scholars had developed various molecular markers such as restriction fragment length polymorphism (RFLP), simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) with different mapping populations to identify QTLs of soybean seed weight. In 1996, Maughan et al. (1996) detected 6 QTLs associated with seed weight in 150  $F_2$  populations and their 150  $F_{2:3}$  lines using 91 polymorphic genetic markers including RFP, RAPD and SSR. Zhang et al. (2016) performed a genome-wide association study (GWAS)

using 309 soybean germplasm resources with 31045 SNP markers and associated QTLs related to seed weight to chromosome 04 and 19. More results from previous studies (Mian et al., 1996; Csanadi et al., 2001; Lee et al., 2001; Chapman et al., 2003; Fasoula et al., 2004; Hyten et al., 2004; D. Li et al., 2008; Han et al., 2012; Pathan et al., 2013; Kato et al., 2014; Yan et al., 2014; Kulkarni et al., 2017; Liu D. et al., 2018) are not identical, but these QTLs have laid a foundation for molecular marker-assisted breeding.

In summary, most researchers use different numbers of mapping populations to detect QTL in various environments. However, studies using more mapping populations had fewer experimental settings, in contrast to studies using fewer mapping populations tended to have more experimental environments. In this study, one mapping population is used to detect QTL in six environments suggesting a higher standard for the identification of stable QTL. In addition, the QTLs localization of soybean seed size and weight traits by RIL population in different environments and the mining of main effect genes can help us to understand the biological basis of soybean yield trait forming process, which has great theoretical and practical significance in guiding the molecular design breeding and in consistently enhancing soybean yield.

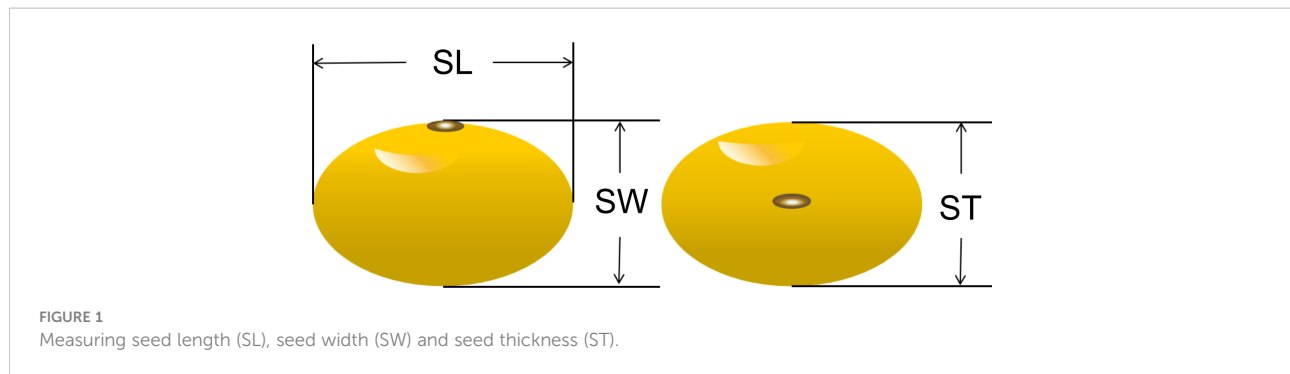
## 2 Materials and methods

### 2.1 Materials and planting

GB13 RIL population, comprising 248 lines, constructed with Guizao1 as the female parent and B13 as the male parent was planted on three replications in July 2018, July 2019 and July 2020 at the Teaching and Research Base of Zengcheng (23°14'N, 113°38'E), and in July 2019 and July 2020 at the Experimental Teaching Base of Guangzhou (23°10'N, 113°22'E) (designated as 18ZC, 19ZC, 20ZC, 19GZ and 20GZ, respectively). In addition, one  $F_1$  and 248  $F_{8:11}$  RILs of the GB13 RIL population were developed using the single seed descent method (Jiang et al., 2019). The experimental materials were planted in a completely randomized zonal design with one row per strain, a row length of 2.0 m, a row spacing of 0.5 m, and a plant spacing of 0.1 m. Field management was carried out according to conventional standards, and no pest or disease outbreaks occurred during growth.

### 2.2 Phenotypic statistics and analysis

Seed length (SL), seed width (SW) and seed thickness (ST) of soybean seeds were measured using an electronic vernier caliper (Figure 1). Twenty seeds from each strain were randomly selected for seed size measurement, after which they were



weighed on an electronic balance. After three repetitions, five times the average of the weights were used as phenotypic data for the seed weight trait (100-seed weight, HSW).

The mean of the phenotypic data from the GB13 RIL population in five natural environments was calculated as the phenotypic data of the sixth environment (combined environment, designated as CE). Descriptive statistics, linear regression analysis and calculation of the skewness and the kurtosis were performed on phenotypic data from six environments using GraphPad Prism software (GraphPad Software Inc. v.7.0.4, United States), and the corresponding tables or figures were generated. Analysis of variance (ANOVA) and estimation of broad-sense heritability ( $h^2$ ) were performed on phenotypic data from the five natural environments using Genstat 12<sup>th</sup> Edition software. The formulae (Nyquist and Baker, 1991; Elattar et al., 2021) were as follows:

$$h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr)$$

where  $\sigma_g^2$  denotes genotype variance,  $\sigma_{ge}^2$  denotes genotype  $\times$  environment interaction variance, and  $\sigma_e^2$  denotes variance error. “n” denotes the number of environments, and “r” denotes the number of replicates per environment.

### 2.3 QTL localization for seed size and weight traits and analysis

The composite interval mapping method (Zeng, 1994) was used to map QTLs with a high-density bin linkage map previously constructed in the laboratory, as described, the map contained 56,561 SNPs and 3715 bin markers covering a total genome length of 3049.2 cM, with the length of the linkage cluster ranging from 120.22–211.37 cM, The average distance between markers was 0.80 cM, and the number of bin markers on each cluster varied from 147 to 259. (Jiang et al., 2019). The additive effect (ADD) and PVE of each QTL were analyzed by the WinQTLCart 2.5 software. The threshold of LOD for different traits was set to 2.5 at the 5% significance level, and LOD thresholds greater than 2.5 were considered as the basis for

determining the presence of QTLs. QTLs located on the same chromosome with adjacent marker intervals were combined into one QTL if their confidence intervals overlapped. QTLs were named according to “q + trait + chromosome + sequence number”, with the symbol “-” in between (McCouch et al., 1997).

### 2.4 Stable QTLs and QTL clusters identification

Generally, QTLs detected in at least two environments were considered stable QTLs (Qi et al., 2017), but there were six experimental environments in this study, so QTLs detected in at least three environments were considered as stable QTLs in this study to identify the stable QTLs.

All QTLs were sorted by chromosome as the primary condition and physical position as the secondary condition. QTLs with identical chromosomes in overlapping or adjacent physical position (less than 5cM) were grouped into a cluster and identified as QTL cluster if it was associated with at least two traits (Liu et al., 2017; Wang et al., 2022). To mine the main effect and stable candidate genes, the identified QTL clusters contain at least one stable QTL in this study.

### 2.5 Candidate gene prediction

First, GO (Gene Ontology) term enrichment analysis was performed for the genes within the interval of the identified QTL clusters. This is useful for selecting candidate genes based on the sequence variation of genes between the parents. The GO annotation numbers were downloaded from the soybean data website SoyBase (<https://www.soybase.org/>), and then the WEGO2.0 online gene enrichment analysis mapping web page (<http://wego.genomics.org.cn/>) was used to create the gene enrichment cellular composition, molecular function and biological processes visualization diagrams. Next, genes that were highly or specifically expressed in soybean seeds were to be further screened. Gene expression data were obtained from the soybean data website SoyBase. Finally, the selected genes

were functionally annotated by the Phytozome database (<https://phytozome-next.jgi.doe.gov/>), and then the candidate genes were identified based on the relevant literature and related gene functions. Meanwhile, the heat map of candidate genes expression was employed to analyze tissue-specific expression (Jia et al., 2022) by the online site (<https://www.omicshare.com/tools/Home/Soft/heatmap>), and the structures of candidate genes were mapped using the online site of the Gene Structure Display Server (<http://gsds.gao-lab.org/index.php>). The resequencing data of the parental lines were referred to compare the variation of candidate genes between parental lines. The whole genome of parental lines were sequenced using the Illumina HiSeq X Ten p, with an average sequencing depth of 8× (Xian et al., 2022). High-quality sequencing data were assessed to predict the gene structural variations.

### 3 Result

#### 3.1 Correlation analysis between seed size and seed weight

The regression linear analysis of seed size with seed weight shows that the SL, SW and ST were all highly significantly and positively correlated with HSW ( $p < 0.01$ ), with R-Squared ( $R^2$ ) of regression linear analysis ranging from 0.440 to 0.804 (Figure 2). In 18ZC, 19ZC, 19GZ, 20ZC and CE, the  $R^2$  between SW and HSW were the highest at 0.804, 0.721, 0.753, 0.552 and 0.742, respectively, while in 20GZ, the  $R^2$  between ST and HSW was the highest at 0.754. Thus, seed size is indeed closely related to seed weight. 18ZC, 19ZC and CE had the next largest  $R^2$  between SL and HSW at 0.685, 0.680 and 0.614, respectively and the smallest  $R^2$  between ST and HSW at 0.679, 0.626 and 0.588, respectively. 19GZ and 20ZC had the next largest  $R^2$  between ST and HSW at

0.713 and 0.467, and the smallest  $R^2$  between SL and HSW at 0.664 and 0.440. 20GZ had the next largest  $R^2$  between SL and HSW at 0.721, and the smallest  $R^2$  between SW and HSW at 0.717. Also, the above results suggest that the correlation between seed size and seed weight seems to be applied in field selection breeding as a reference for selecting high-yielding varieties.

#### 3.2 Descriptive statistics analysis of seed size and weight traits

The phenotypic data were analyzed by descriptive statistics, and the results showed that the SL, SW, ST (Supplementary Figure 1) and HSW of the female parent Guizao1 were higher than those of the male parent B13 with significant differences (Table 1), which laid the foundation for QTL mapping. The GB13 RIL population differed remarkably in size and weight traits, both of which showed transgressive segregation. The coefficient of variation (CV) for SL ranged from 4.33 to 5.25%, for SW from 3.96 to 4.78%, for ST from 4.71 to 5.18%, and for HSW from 10.68 to 12.98%. The CV for HSW was greater than that for SL, SW and ST, but they were relatively stable, which indicated that both parents contained genes that were acting in the phenotypic variation.

The absolute values of the skewness and the kurtosis of SL, SW, ST and HSW of the GB13 RIL population were less than 1 (Table 1). In addition, the frequency distribution graph (Figure 3) showed that the phenotypic data of seed size and weight traits displayed continuous variation. The above results indicated that seed size and weight traits of the GB13 RIL population obeyed normal distribution, which was consistent with the characteristics of the RIL population and belonged to quantitative genetic traits.

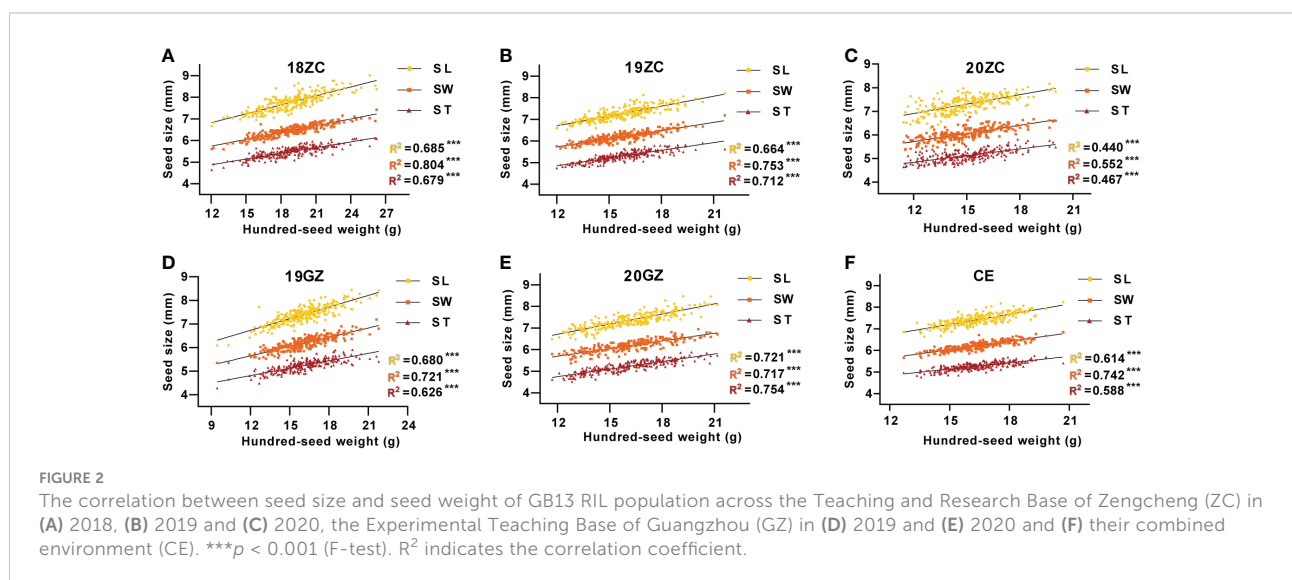


TABLE 1 Statistical analysis of seed size and weight traits for the parents and the lines at different environments.

Trait <sup>#</sup>	Env. <sup>a</sup>	Parents <sup>b</sup>		RILs <sup>e</sup>					
		Guizao1 <sup>c</sup>	B13 <sup>d</sup>	Range <sup>f</sup>	Mean <sup>g</sup>	SD <sup>h</sup>	CV (%) <sup>i</sup>	Skew. <sup>j</sup>	Kurt. <sup>k</sup>
SL	18ZC	8.34 ± 0.02	7.40 ± 0.01	6.66-9.01	7.77	0.41	5.25	0.12	-0.20
	19GZ	7.40 ± 0.02	6.90 ± 0.02	6.08-8.44	7.40	0.38	5.15	-0.19	0.66
	19ZC	7.40 ± 0.02	6.68 ± 0.02	6.52-8.21	7.27	0.32	4.33	0.32	-0.20
	20GZ	7.61 ± 0.01	6.86 ± 0.03	6.48-8.46	7.35	0.37	5.04	0.20	0.03
	20ZC	7.87 ± 0.03	6.85 ± 0.01	6.29-8.17	7.30	0.35	4.82	-0.35	-0.20
	CE	7.72 ± 0.07	6.94 ± 0.05	6.64-8.23	7.42	0.28	3.80	0.07	-0.03
SW	18ZC	6.65 ± 0.01	6.24 ± 0.01	5.59-7.43	6.46	0.29	4.44	-0.17	0.40
	19GZ	5.99 ± 0.02	5.84 ± 0.05	5.35-7.18	6.18	0.30	4.78	-0.01	0.22
	19ZC	6.02 ± 0.02	5.77 ± 0.02	5.65-7.18	6.20	0.25	3.96	0.31	0.46
	20GZ	5.97 ± 0.01	6.08 ± 0.02	5.37-6.95	6.18	0.28	4.55	-0.12	-0.23
	20ZC	6.10 ± 0.02	5.68 ± 0.03	5.37-6.75	6.05	0.27	4.50	-0.01	-0.27
	CE	6.34 ± 0.11	6.19 ± 0.08	5.66-6.85	6.21	0.21	3.39	0.14	-0.21
ST	18ZC	5.53 ± 0.01	5.22 ± 0.01	4.65-6.11	5.49	0.26	4.79	-0.22	0.07
	19GZ	5.17 ± 0.02	4.84 ± 0.04	4.28-5.90	5.24	0.25	4.83	-0.23	0.60
	19ZC	5.12 ± 0.01	4.90 ± 0.01	4.75-5.90	5.30	0.24	4.49	0.01	-0.47
	20GZ	5.11 ± 0.01	5.06 ± 0.01	4.52-5.88	5.23	0.27	5.18	-0.14	-0.39
	20ZC	5.22 ± 0.01	4.88 ± 0.02	4.61-5.81	5.12	0.24	4.71	0.25	-0.20
	CE	5.16 ± 0.01	5.05 ± 0.05	4.76-5.75	5.27	0.18	3.39	-0.10	-0.06
HSW	18ZC	18.85 ± 0.1	16.32 ± 0.06	12.05-26.22	18.88	2.45	12.98	0.14	0.28
	19GZ	15.35 ± 0.19	12.90 ± 0.24	9.45-21.80	16.09	1.91	11.89	0.02	0.61
	19ZC	14.97 ± 0.06	12.32 ± 0.05	12.00-21.63	15.77	1.68	10.68	0.29	0.02
	20GZ	15.40 ± 0.05	13.27 ± 0.13	11.70-21.17	16.08	2.01	12.52	0.14	-0.65
	20ZC	17.30 ± 0.12	13.02 ± 0.09	11.43-20.03	14.96	1.75	11.71	0.31	-0.17
	CE	16.37 ± 0.30	14.14 ± 0.51	12.70-20.70	16.35	1.41	8.62	0.15	-0.34

<sup>#</sup>Seed length (SL), seed width (SW), seed thickness (ST) and 100-seed weight (HSW).

<sup>a</sup>Environment.

<sup>b</sup>Parents of GB13 RIL population.

<sup>c</sup>Female parent of GB13 RIL population.

<sup>d</sup>Male parent of GB13 RIL population.

<sup>e</sup>Recombinant inbred lines.

<sup>f</sup>Range of seed size and 100-seed weight for GB13 RIL population.

<sup>g</sup>Mean of seed size and 100-seed weight for GB13 RIL population.

<sup>h</sup>Standard deviation.

<sup>i</sup>Coefficient of variation.

<sup>j</sup>Skewness.

<sup>k</sup>Kurtosis.

### 3.3 Analysis of variance and estimates of broad-sense heritability

The results of ANOVA for SL, SW, ST and HSW of the GB13 RIL population in five natural environments (Table 2)

show that the genotype, environment and interaction between genotype and environment had significant effects on seed size and weight traits of the GB13 RIL population ( $p < 0.0001$ ). The  $h^2$  for seed size and weight traits of the GB13 RIL population were relatively high ranging from 0.74 to

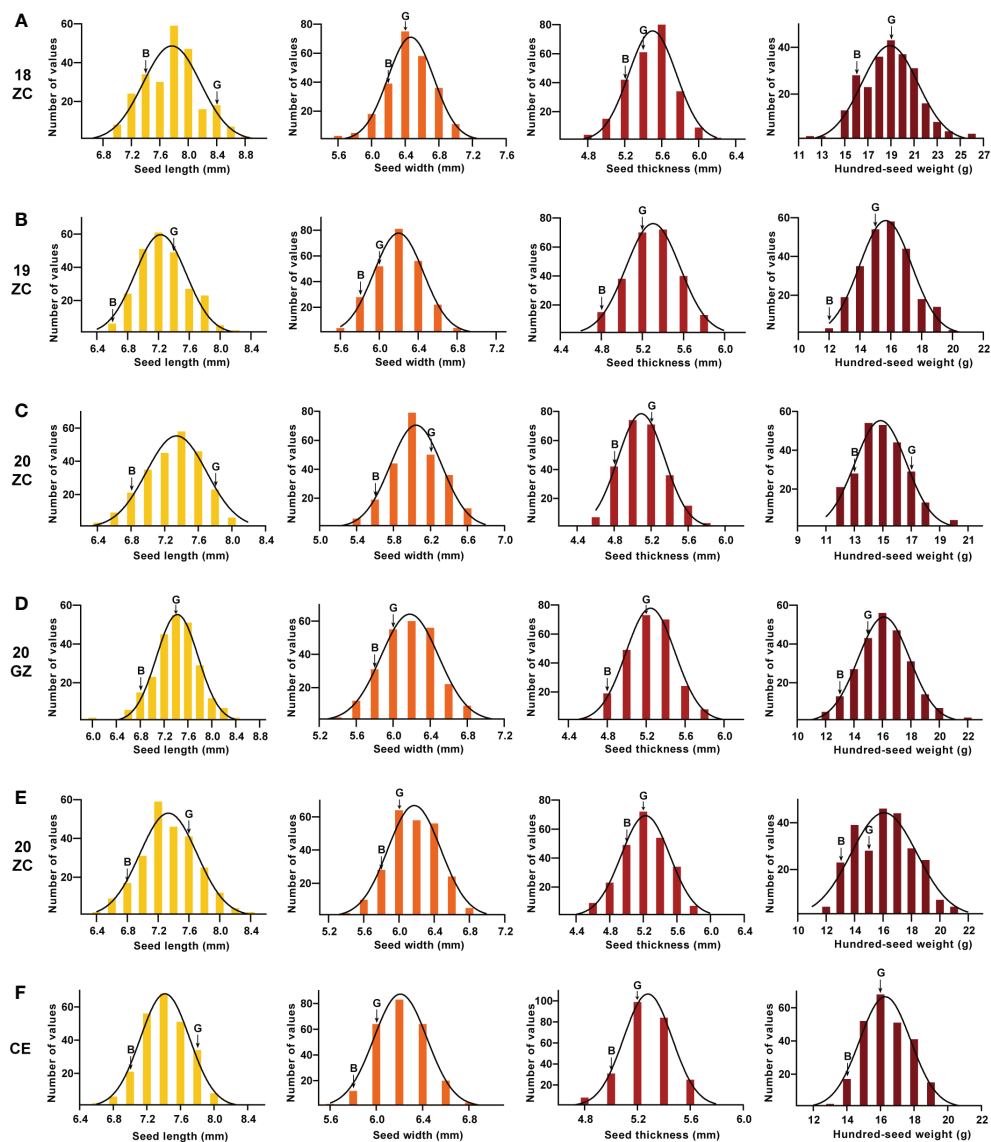


FIGURE 3

Frequency distribution and its fitted curve of seed size and weight traits for GB13 RIL population across the Teaching and Research Base of Zengcheng (ZC) in (A) 2018, (B) 2019 and (C) 2020, the Experimental Teaching Base of Guangzhou (GZ) in (D) 2019 and (E) 2020 and (F) their combined environment (CE). The letter "G" indicates the female parent Guizao1, and the letter "B" indicates the male parent B13. The horizontal coordinates of the different colored figures from left to right were seed length, seed width, seed thickness and 100-seed weight.

0.83 (Table 2). The traits are suitable for further QTL localization analysis.

In terms of the proportion of total variation accounted for variation of different sources (Table 2), the largest source of variation in SL and SW was genotype, while the largest source of variation in ST and HSW was genotype  $\times$  environment interactions. The most influenced by the environment was HSW, accounted for 30.71% of the total variation. It was followed by SL, SW and ST, which accounting for 19.51%, 18.19% and 17.53% of the total variation, respectively. The above results indicated that the seed weight trait was more

influenced by the environment compared to the seed size trait.

### 3.4 Identification of stable QTL for seed size and weight traits

A total of 85 QTL were detected for seed size and weight in six environments with PVE in the range of 3.14% - 21.93%, of which 19 QTLs were identified for SL, 23 for SW, 18 for ST, and 25 for HSW (Supplementary Tables 1–4). By collation and

TABLE 2 Analysis of variance and broad-sense heritability for GB13 RIL population at five natural environments.

Trait <sup>#</sup>	Sources <sup>a</sup>	Df <sup>b</sup>	SS <sup>c</sup>	MS <sup>d</sup>	P value <sup>e</sup>	VC <sup>f</sup>	PV (%) <sup>g</sup>	h <sup>2</sup> (%) <sup>h</sup>
SL	Repeat	2	0.74	0.37	<0.0001			0.83
	Genotype	247	294.11	1.19	<0.0001	0.07	46.49	
	Environment	4	123.42	30.86	<0.0001	0.04	19.51	
	Interaction	988	200.95	0.20	<0.0001	0.07	31.77	
	Error	2478	13.38	0.01				
	Total variation	3719	632.59					
SW	Repeat	2	1.09	0.54	<0.0001			0.82
	Genotype	247	164.87	0.67	<0.0001	0.04	45.42	
	Environment	4	66.05	16.51	<0.0001	0.02	18.19	
	Interaction	988	119.22	0.12	<0.0001	0.04	32.84	
	Error	2478	11.80	0.01				
	Total variation	3719	363.04					
ST	Repeat	2	1.09	0.54	<0.0001			0.74
	Genotype	247	118.51	0.48	<0.0001	0.02	38.23	
	Environment	4	54.34	13.58	<0.0001	0.02	17.53	
	Interaction	988	124.30	0.13	<0.0001	0.04	40.1	
	Error	2478	11.72	0.01				
	Total variation	3719	309.96					
HSW	Repeat	2	45.00	22.50	<0.0001			0.75
	Genotype	247	7210.17	29.19	<0.0001	1.46	33.32	
	Environment	4	6645.42	1661.36	<0.0001	2.22	30.71	
	Interaction	988	7256.28	7.34	<0.0001	2.38	33.54	
	Error	2478	479.53	0.19				
	Total variation	3719	21636.40					

<sup>#</sup> Seed length (SL), seed width (SW), seed thickness (ST) and 100-seed weight (HSW).  
<sup>a</sup>Sources of variation.  
<sup>b</sup>Degree of freedom.  
<sup>c</sup>Sum of deviation squares.  
<sup>d</sup>mean square.  
<sup>e</sup>The P value of the F-test (joint hypotheses test).  
<sup>f</sup>Variance components for different sources of variation.  
<sup>g</sup>Proportion of variation.  
<sup>h</sup>Broad-sense heritability.

analysis, a total of 18 stable QTLs were identified (Table 3), including 3 for SL, 6 for SW, 4 for ST, and 5 for HSW (Figure 4). One QTL, *qSL-3-1*, associated with SL was detected in all environments, and it explained more than 10% of PV in three environments in the range of 11.93% - 15.91%. Three QTLs (*qHSW-11-3*, *qHSW-20-3* and *qSW-20-3*) were detected in four environments. Among them, *qSW-20-3* explained 9.22%, 10.79%, 20.89% and 21.93% of PV in four environments,

respectively. Another 14 QTLs distributed on chromosomes 03, 04, 05, 07, 11, 12, 13, 18 and 20 were detected in three environments. In summary, integrating the PVE for QTL, the number of detections in various environments, and the relevant references, *qSL-3-1* and *qSW-20-3* can be considered as novel stable QTL with the main effects. Meanwhile, these stable QTLs above were important references for the excavation of genes controlling seed size and weight traits.

TABLE 3 Stable QTLs associated with seed size and weight traits for GB13 RIL population across different environments.

QTL #	Chr <sup>a</sup>	Interval	Position (cM)	PVE (%) <sup>h</sup>						LOD <sup>i</sup>						ADD <sup>j</sup>					
				E1 <sup>b</sup>	E2 <sup>c</sup>	E3 <sup>d</sup>	E4 <sup>e</sup>	E5 <sup>f</sup>	E6 <sup>g</sup>	E1	E2	E3	E4	E5	E6	E1	E2	E3	E4	E5	E6
<i>qSL-3-1</i>	3	bin32-bin43	43.1	13.57	6.03	11.93	4.38	6.61	15.91	9.37	4.03	7.49	3.15	4.61	11.45	0.15	0.09	0.11	0.08	0.09	0.11
<i>qSL-3-2</i>	3	bin51-bin55	49.4			12.15		9.07	14.56			7.63		6.42	10.38			0.11		0.11	0.11
<i>qHSW-3-2</i>	3	bin53-bin55	50.1			5.12		4.41	6.81			3.68		3.20	5.43			0.38		0.37	0.37
<i>qHSW-4-2</i>	4	bin75-bin78	57.5	5.34		4.68			5.28	3.75		3.25			4.18	0.59		0.37			0.33
<i>qST-4-3</i>	4	bin77-bin83	65.2	4.86	8.61				6.10	3.18	5.44				4.70	0.06	0.08				0.04
<i>qST-5-2</i>	5	bin155-bin158	135.1	4.67			4.52		3.69	3.06			3.13		2.89	-0.06			-0.06		-0.04
<i>qST-7-1</i>	7	bin159-bin161	139.9				3.85	5.75	6.99			2.69	4.06	5.37					0.05	0.06	0.05
<i>qSW-11-1</i>	11	bin95-bin100	101	3.89			7.48		6.34	2.90			5.23		5.14	-0.06			-0.08		-0.05
<i>qHSW-11-3</i>	11	bin131-bin134	133.1		4.90		8.26	3.33	4.41		2.99		5.76	2.56	3.41		0.42		0.59	0.32	0.31
<i>qSW-12-2</i>	12	bin103-bin109	111.1			5.79		3.69	5.39			4.37		2.83	4.36			0.06		0.05	0.05
<i>qHSW-13-1</i>	13	bin115-bin116	90.0		4.34			5.19	7.21		2.91			3.79	5.69		0.39			0.40	0.38
<i>qSL-13-1</i>	13	bin115-bin117	90	3.47				4.12	4.47	2.55				3.03	3.54	0.08				0.07	0.06
<i>qST-13-3</i>	13	bin143-bin149	104.6		5.09	5.52			4.56		3.27	3.74			3.50		0.06	0.06			0.04
<i>qSW-18-1</i>	18	bin237-bin239	152.4	5.62		4.91			3.91	4.03		3.72			3.24	0.07		0.05			0.04
<i>qSW-20-2</i>	20	bin46-bin49	49.1	7.46	11.80	8.64				5.30	8.29	6.24				-0.08	-0.10	-0.07			
<i>qSW-20-3</i>	20	bin64-bin68	55.3	9.22		10.79		21.93	20.89	6.63		7.89		15.04	15.24	-0.09	-0.10	-0.08		-0.14	-0.10
<i>qHSW-20-3</i>	20	bin68-bin73	55.9		6.12	3.31		4.68	7.21		4.06	2.52		3.42	5.70		-0.47	-0.31		-0.38	-0.38
<i>qSW-20-4</i>	20	bin84-bin86	64.2			3.80		7.30	5.81			2.65		4.56	3.86			-0.05		-0.09	-0.05

<sup>#</sup>The name of each QTL. <sup>a</sup>Chromosome. <sup>b</sup>at Zengcheng in 2018. <sup>c</sup>at Guangzhou in 2019. <sup>d</sup>at Zengcheng in 2019. <sup>e</sup>at Guangzhou in 2020. <sup>f</sup>at Zengcheng in 2020. <sup>g</sup>Combined environment. <sup>h</sup>Phenotypic variation explained. <sup>i</sup>Log of odd value. <sup>j</sup>Additive effect.



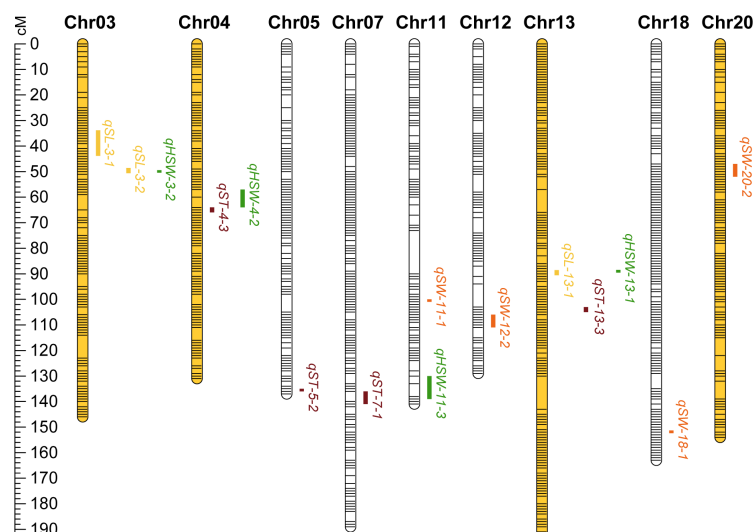


FIGURE 4

The linkage map of stable QTLs for soybean seed size and weight traits. The ruler on the left side indicates the interval distance between markers using cM (centimorgan) as the unit. Bars with yellow emphasis indicate the chromosome on which the QTL clusters are located. Between the bars in the graph, the yellow indicates QTL for SL; the orange indicates QTL for SW; the reddish brown indicates QTL for ST; the green indicates QTL for HSW.

### 3.5 Identification of QTL clusters

According to the classification criteria of QTL clusters, we obtained five QTL clusters (Table 4) from the QTL mapping results (Supplementary Tables 1–4). These five QTL clusters were distributed on chromosomes 03, 04, 13 and 20. All of them were associated with HSW, which verified that the seed size was closely related to HSW from another aspect. The four QTL

clusters related to SL were *qLH3-1*, *qLH3-2*, *qLWTH4* and *qLH13*; the two QTL clusters related to SW were *qLWTP4* and *qWTH20*; the two QTL clusters related to ST were *qLWTP4* and *qWTP20*. In terms of the number of traits controlled, one QTL cluster for four traits was *qLWTP4*, one QTL cluster for three traits was *qWTH20*, and three QTL clusters for two traits were *qLH3-1*, *qLH3-2*, and *qLH13*. Among them, *qLP3-1* and *qLP3-2* containing major and stable QTL(s) were located on the forearm

TABLE 4 QTL clusters associated with at least two traits.

Clusters	Chr <sup>a</sup>	Interval	Position(bp)	QTLs <sup>b</sup>	ADD <sup>c</sup>	PVE(%) <sup>d</sup>	Env. <sup>e</sup>
<i>qLH3-1</i>	3	bin32-bin43	4387438-6456915	<i>qSL-3-1</i>	0.15	13.57	18ZC
					0.09	6.03	19GZ
					0.11	11.93	19ZC
					0.08	4.38	20GZ
					0.09	6.61	20ZC
					0.11	15.91	CE
				<i>qHSW-3-1</i>	0.33	3.77	19ZC
	0.35	6.07	CE				
<i>qLH3-2</i>	3	bin51-bin55	8784766-17199876	<i>qSL-3-2</i>	0.11	12.15	19ZC
					0.11	9.07	20ZC
					0.11	14.56	CE
				<i>qHSW-3-2</i>	0.38	5.12	19ZC

(Continued)

TABLE 4 Continued

Clusters	Chr <sup>a</sup>	Interval	Position(bp)	QTLs <sup>b</sup>	ADD <sup>c</sup>	PVE(%) <sup>d</sup>	Env. <sup>e</sup>
					0.37	4.41	20ZC
					0.37	6.81	CE
<i>qLWTH4</i>	4	bin74-bin83	9040807-12316836	<i>qSL-4-2</i>	0.06	4.36	CE
				<i>qSW-4-2</i>	0.06	3.78	18ZC
					0.04	3.70	CE
				<i>qST-4-3</i>	0.06	4.86	18ZC
					0.08	8.61	19GZ
					0.04	6.10	CE
				<i>qHSW-4-2</i>	0.59	5.34	18ZC
					0.37	4.68	19ZC
				0.33	5.28	CE	
<i>qLH13</i>	13	bin115-bin117	27473391-27671175	<i>qSL-13-1</i>	0.08	3.47	18ZC
					0.07	4.12	20ZC
					0.06	4.47	CE
				<i>qHSW-13-1</i>	0.39	4.34	19GZ
					0.40	5.19	20ZC
					0.38	7.21	CE
<i>qWTH20</i>	20	bin64-bin73	27890104-33222868	<i>qSW-20-3</i>	-0.09	9.22	18ZC
					-0.08	10.79	19ZC
					-0.14	21.93	20ZC
					-0.10	20.89	CE
				<i>qST-20-2</i>	-0.05	7.55	CE
				<i>qHSW-20-3</i>	-0.47	6.12	19GZ
					-0.31	3.31	19ZC
					-0.38	4.68	20ZC
					-0.38	7.21	CE

<sup>a</sup>Chromosome. <sup>b</sup>Individual QTLs of cluster. <sup>c</sup>Additive effect. <sup>d</sup>Phenotypic variation explained. <sup>e</sup>Environment.

of chromosome 03 at the physical positions between 4387438 and 6456915 bp and between 8784766 and 17199876 bp, respectively. Similarly, *qWTH20* also contained major and stable QTLs and was located at the physical position of chromosome 20 between 12056958 and 33222868 bp.

In summary, *qLP3-1*, *qLP3-2* and *qWTH20* may be QTL clusters containing major effect genes that regulate soybean seed size and weight traits; *qLWTP4* and *qLP13* may be QTL clusters containing micro-effector genes that harmonize the control of soybean seed size and weight traits. Therefore, these five QTL clusters can provide a reference for mining the target genes controlling seed size and weight

traits, and were used as the target intervals for gene GO enrichment analysis.

### 3.6 Gene GO enrichment analysis

By Gene GO enrichment analysis, we found that most of the genes within these five QTL cluster intervals were involved in cellular processes and metabolic processes (Figure 5). Most of the genes in the *qLP3-1* were involved in intracellular metabolic processes, bioregulation and reaction to stimuli, and were mostly involved in binding and catalytic activities in terms of molecular

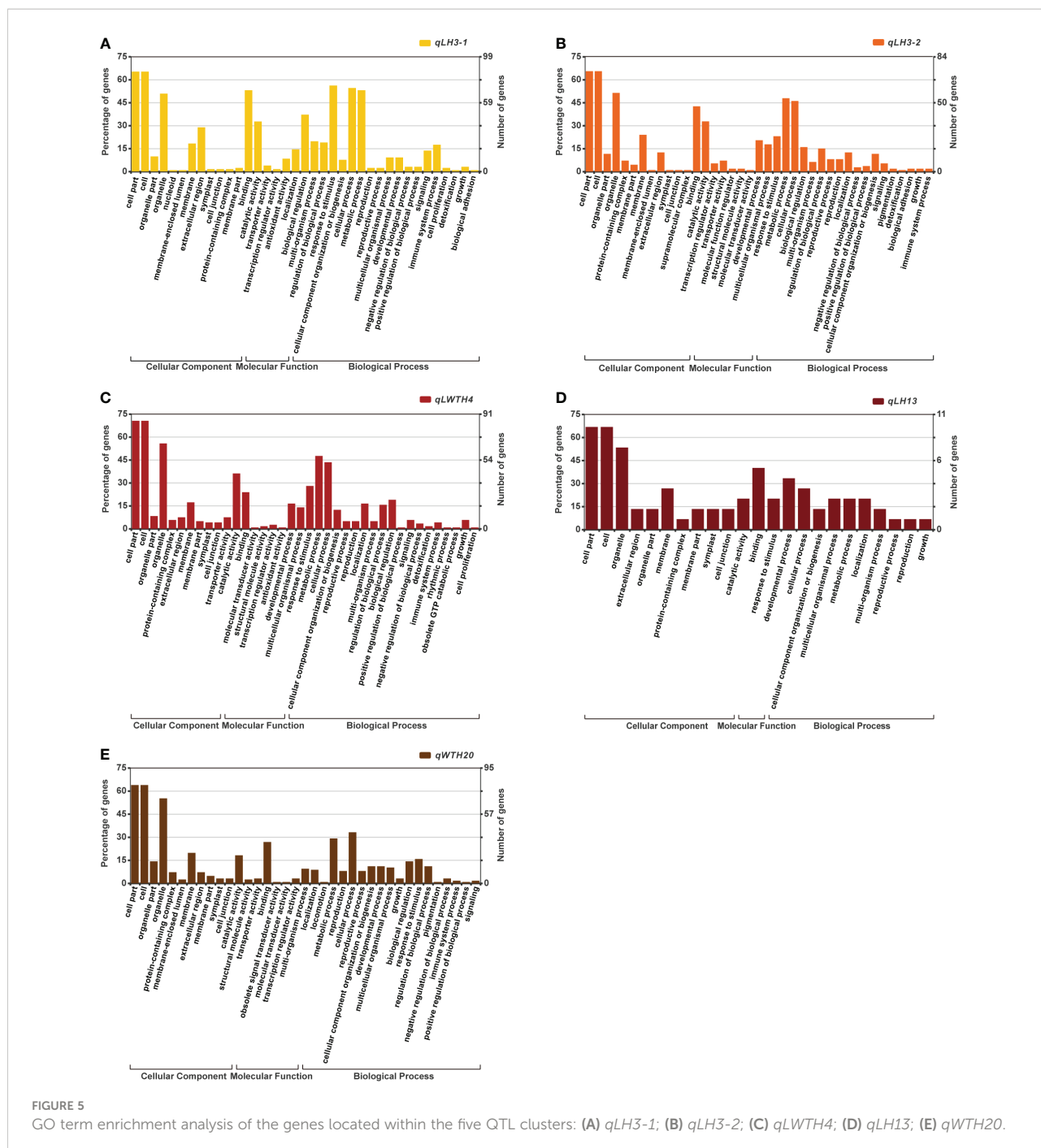


FIGURE 5 GO term enrichment analysis of the genes located within the five QTL clusters: (A) *qLH3-1*; (B) *qLH3-2*; (C) *qLWTH4*; (D) *qLH13*; (E) *qWTH20*.

functions. While most of the genes in the *qLP3-2* were similar in function to those in the *qLP3-1*. In comparison, most of the genes in the *qLWTP4* were also involved in intracellular metabolic processes, but a small number of genes were involved in regulating growth and cell proliferation. The *qLP13* contained only 11 annotated genes, most of which were involved in growth and cellular processes. Similarly, most of the genes in the *qWTP20-2* were involved in intracellular metabolic processes.

### 3.7 Candidate genes in QTL clusters

Through the above gene GO enrichment analysis, followed by gene expression screening, gene function annotation, and sequence variation analysis, we obtained a total of 15 candidate genes (Table 5) that might be participating in the regulation of seed size and weight traits. Among them, six genes were located on chromosome 03, three on chromosome 04, two on chromosome 13, and four on chromosome 20. Among the 15

TABLE 5 Candidate genes identified within the five QTL clusters that were highly expressed or specific expression in soybean seed.

Gene	Position (bp)	Gene functional annotation
<i>Glyma.03g036600</i>	4432910-4435002	Negative regulation of transcription; organ development; lipid transport calmodulin binding
<i>Glyma.03g040400</i>	5050463-5052307	Protease inhibitors (LTP family); Seed storage
<i>Glyma.03g044200</i>	5586338-5589780	Transporter protein activity; integral to membrane
<i>Glyma.03g045400</i>	5760642-5763461	Metabolic process; methyltransferase activity
<i>Glyma.03g065700</i>	11896538-11909845	Regulation of transcription; sequence-specific DNA binding transcription factor activity; DNA-dependent
<i>Glyma.03g065900</i>	12064850-12067709	Metabolic process; hydrolase activity; catalytic activity
<i>Glyma.04g100100</i>	9177207-9180239	Negative regulation of gene expression; response to abscisic acid stimulus; ubiquitin-protein ligase activity
<i>Glyma.04g100400</i>	9219564-9221412	Sterol biosynthetic process; response to oleuropein lactone stimulus
<i>Glyma.04g107100</i>	11239281-11250948	Regulation of meristem growth; Ubiquitin ligases are involved in synaptic protein degradation
<i>Glyma.13g159500</i>	27476780-27479031	Ubiquitin system components suggestive of protein; cell growth; cell morphogenesis
<i>Glyma.13g160400</i>	27606020-27608359	Lipid transport; lipid binding; Hydrophobic protein
<i>Glyma.20g081600</i>	30778586-30779164	Regulation of gene expression; response to auxin stimulus; protein binding
<i>Glyma.20g084000</i>	31536334-31538924	RNA methylation; RNA processing
<i>Glyma.20g084500</i>	31666190-31674959	Ubiquitin-protein ligase activity; protein ubiquitination; CUL4-RING ubiquitin ligase complex
<i>Glyma.20g087000</i>	32527435-32530216	Negative regulation of ethylene mediated signaling pathway; regulation of transcription; ethylene binding

genes screened, *Glyma.03g040400* was the gene with the same function as the gene (*LOC\_Os07g19000*) associated with seed size in rice; *Glyma.03g045400* and *Glyma.20g084000* were related to DNA methylation; *Glyma.04g100400*, *Glyma.20g081600* and *Glyma.20g087000* were hormone-related and might be involved in the regulation of seed size and weight traits by hormone signaling pathway; *Glyma.04g100100*, *Glyma.04g107100*, *Glyma.13g159500* and *Glyma.20g084500* might negatively regulate seed size and weight through the ubiquitin-proteasome pathway. *Glyma.03g036600* and *Glyma.03g065700* were associated with transcription factor regulation and might be involved in the regulation of seed size and weight traits through the transcription factor regulation pathway; *Glyma.03g044200*, *Glyma.03g065900* and *Glyma.13g160400* might be related to the metabolic synthesis of protein and oil, and indirectly regulated the seed size and weight traits.

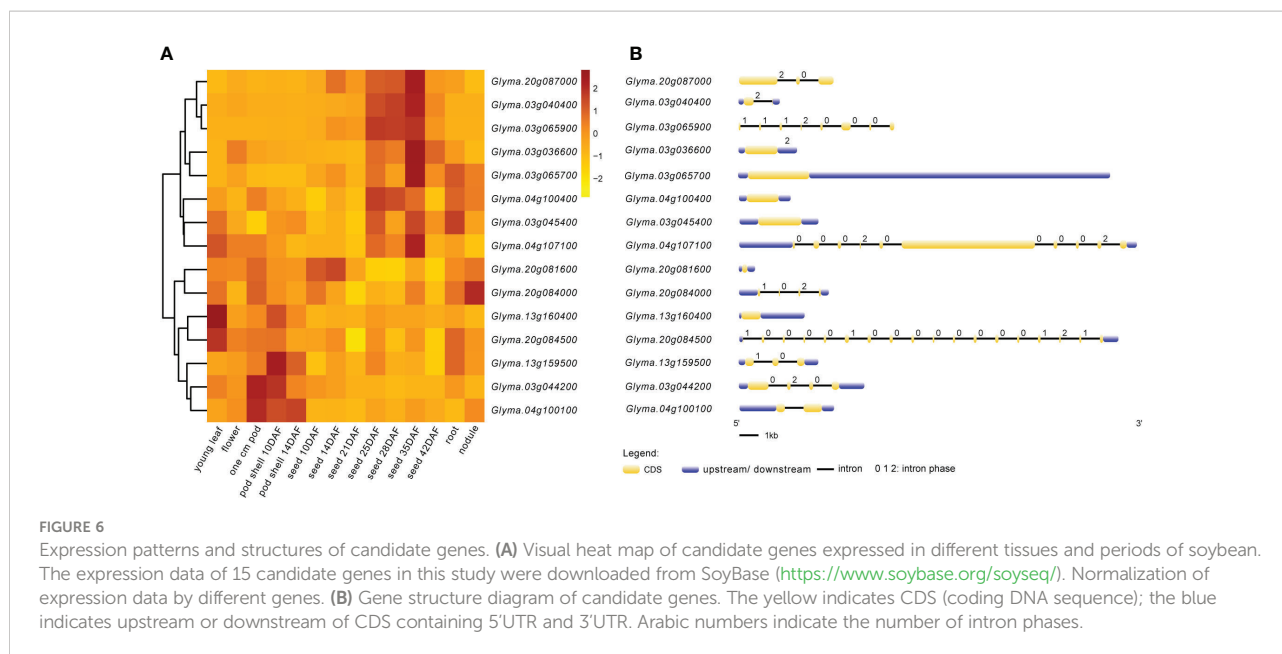
The expression heat map (Figure 6A) of these 15 candidate genes showed that two genes each were highly expressed in young leaf and nodule of soybean; three genes were specifically expressed at the pod shell development stage in soybean; eight genes including *Glyma.03g036600*, *Glyma.03g040400*, *Glyma.03g045400*, *Glyma.03g065700*, *Glyma.03g065900*, *Glyma.04g100400*, *Glyma.04g107100* and *Glyma.20g087000* were specifically and progressively expressed at the seed maturity stage in soybean. Meanwhile, the structural maps for most of the candidate genes (Figure 6B) were matched with their sequence in the parental lines, but ten candidate genes varied in the sequence of the parental lines, with the variant region on introns, 3'UTR (untranslated region) and 5'UTR (Table 6). They

are *Glyma.03g036600*, *Glyma.03g040400*, *Glyma.03g044200*, *Glyma.03g045400*, *Glyma.03g065700*, *Glyma.04g100100*, *Glyma.04g107100*, *Glyma.13g160400*, *Glyma.20g084000* and *Glyma.20g084500*, and deserve to be investigated in depth.

## 4 Discussion

Based on the results of the correlation analysis between seed size and seed weight in this study, the R<sup>2</sup> between seed width and seed weight was the largest in most environments, so we suggest that varieties with wide seeds can be selected for varietal improvement for soybean seed weight trait. However, from the results of the QTLs localized for soybean seed size and weight traits, most of the QTLs for HSW overlapped or were close to the QTL interval for SL. It suggests that most of the genes with multi-effects may be genes that coordinate the control on both SL and HSW. Nevertheless, the specific regulatory mechanism is still unclear (He et al., 2021) and needs to be further investigated.

In this study, a total of 60 QTLs related to seed size trait and 25 QTLs related to seed weight trait were identified in the GB13 RIL population in five natural environments and their combined environment. Nineteen QTLs for SL included 4 stable QTLs and 15 sensitive QTLs (Supplementary Table 1); 23 QTLs for seed width included 6 stable QTLs and 17 sensitive QTLs (Supplementary Table 2); 18 QTLs for ST included 5 stable QTLs and 13 sensitive QTLs (Supplementary Table 3). The 25 QTLs associated with HSW included 5 stable QTLs and 20 sensitive QTLs (Supplementary Table 4). By comparing the



results with those of previous studies, ten sensitive QTLs for SL (*qSL-2-1*, *qSL-3-3*, *qSL-4-1*, *qSL-4-2*, *qSL-4-3*, *qSL-6-2*, *qSL-11-1*, *qSL-11-2*, *qSL-18-2* and *qSL-20-1*) overlapped with the intervals of previous localization results or were within their subsets (Salas et al., 2006; Xu et al., 2011; Niu et al., 2013; Yang et al., 2017; Cui et al., 2020; Hina et al., 2020; Kumawat and Xu, 2021). Eight sensitive QTLs for SW (*qSW-2-1*, *qSW-2-2*, *qSW-4-1*, *qSW-9-1*, *qSW-10-1*, *qSW-15-1*, *qSW-20-1*, and *qSW-20-5*) overlap with or were within a subset of the previous localization result intervals (Xu et al., 2011; Niu et al., 2013; Yang et al., 2017; Hina et al., 2020). Five sensitive QTLs for ST (*qST-4-4*, *qST-5-2*, *qST-6-1*, *qST-13-1* and *qST-20-3*) overlapped with or included the intervals of previous localization results (Xu et al., 2011; Niu et al., 2013; Hina et al., 2020). Five sensitive QTLs for HSW (*qHSW-4-3*, *qHSW-13-2*, *qHSW-17-2*, *qHSW-20-1*, and *qHSW-20-5*) were found to have regions of overlap with the localization results of previous studies (Mansur et al., 1993; Teng et al., 2009; Han et al., 2012; Kato et al., 2014). Overall, the results of the present study correlated well with those of previous studies.

Besides, comparing the localization results of seed size traits in this study with the results of other related trait localization studies, eight QTLs (*qSL-4-1*, *qSL-17-1*, *qSL-20-1*, *qSW-12-2*, *qSW-15-2*, *qSW-17-1*, *qST-13-3*, *qST-20-1*) were found to overlap or interval overlap with QTLs related to yield traits from previous studies (Mansur et al., 1993; Kabelka et al., 2004; Du et al., 2009; Palomeque et al., 2009; Liu et al., 2011; Wang X. et al., 2014). Fifteen QTLs (*qSL-6-2*, *qSL-13-1*, *qSL-17-1*, *qSL-18-1*, *qSL-20-1*, *qSW-11-1*, *qSW-12-1*, *qSW-12-2*, *qSW-20-2*, *qSW-20-3*, *qST-5-2*, *qST-13-1*, *qST-13-3*, *qST20-1*, *qST-20-2*) overlapped QTLs associated with traits related to protein, oil and fatty acids (Sebolt et al., 2000; Tajuddin et al., 2003; Bachlava et al., 2009; Qi et al., 2011; Eskandari et al., 2013; Mao et al., 2013; Wang J.

et al., 2014; Fan et al., 2015; Warrington et al., 2015; Han et al., 2015; Leite et al., 2016; Fliege et al., 2022). Five QTLs (*qSW-15-1*, *qSW-15-3*, *qSW-18-2*, *qST-4-1* and *qSW-5-2*) had overlapping intervals with QTLs related to seed-set of pod number and soybean maturity (Yao et al., 2015; Ning et al., 2018). Two QTLs (*qSL-5-1* and *qST-5-2*) had interval overlapping with isoflavonoid-related QTLs (Yang et al., 2011; Cai et al., 2018). The *qSW-4-1* overlapped with the QTLs associated with the long juvenility (Yue et al., 2017). Researchers sometimes studied the seed weight of soybean for QTLs mapping in collaboration with protein and oil content (Fasoula et al., 2004; Pathan et al., 2013; Yesudas et al., 2013; Wang X. et al., 2014; Whiting et al., 2020; He et al., 2021). In this study, *qHSW-13-1*, *qHSW-15-1*, *qHSW-17-2*, *qHSW-20-2* and the stable QTL *qHSW-20-3* had overlap intervals with QTLs related to protein or oil content studied previously (Panthee et al., 2005; Mao et al., 2013; Wang X. et al., 2014; Han et al., 2015; Fliege et al., 2022). Besides, *qHSW-15-1* and *qHSW-17-1* also had overlap intervals with yield-related QTL studied previously (Palomeque et al., 2009; Liu et al., 2011). In summary, many QTLs with stable main effects related to seed size and weight traits in this study were consistent with previous studies results and correlated with QTLs related to other traits including protein, oil content, pod number, yield, maturity, and long juvenile stage. Thus, we suggest that genes regulating soybean seed size and weight traits may be correlated with genes regulating other traits, especially protein and oil content, so genes regulating the synthesis or metabolism of protein and oil content were also used as the basis for candidate gene speculation in this study.

In addition, compared with soybean, rice (*Oryza sativa L.*) has been studied much more intensively than soybean in terms of its seed size and weight. Currently, genes associated with seed size and weight, such as *GW2* (Song et al., 2007), *OsGRF4* (Li et al.,

TABLE 6 Sequence variations between the parental lines of candidate genes.

Gene	Chr <sup>a</sup>	Loci(bp) <sup>b</sup>	Guizao1 <sup>c</sup>	B13 <sup>d</sup>	Region <sup>e</sup>
<i>Glyma.03g036600</i>	3	4433324	T	TA	UTR3
<i>Glyma.03g040400</i>	3	5051148	A	ACG	intronic
	3	5051173	A	ATCACCC	intronic
	3	5051386	T	TC	intronic
	3	5052301	G	GAT	UTR3
<i>Glyma.03g044200</i>	3	5588615	TCA	T	intronic
	3	5588681	AAAC	A	intronic
	3	5588699	AAC	A	intronic
<i>Glyma.03g045400</i>	3	5760781	T	TTC	UTR3
	3	5760989	TG	T	UTR3
	3	5761009	A	AAAAACAAGAG	UTR3
	3	5761114	AACTCTC	A	UTR3
	3	5763119	T	TATG	UTR5
<i>Glyma.03g065700</i>	3	11907032	G	GA	UTR3
<i>Glyma.03g065900</i>	3	NA	NA	NA	NA
<i>Glyma.04g100100</i>	4	9179539	G	GAAAACAAAAC	UTR5
<i>Glyma.04g100400</i>	4	NA	NA	NA	NA
<i>Glyma.04g107100</i>	4	11246102	GA	G	intronic
<i>Glyma.13g159500</i>	13	NA	NA	NA	NA
<i>Glyma.13g160400</i>	13	27607021	TA	T	UTR3
<i>Glyma.20g081600</i>	20	NA	NA	NA	NA
<i>Glyma.20g084000</i>	20	31536882	A	AAAAACCCTACATCTTCAT	intronic
	20	31537182	ATT	A	intronic
	20	31537333	TTA	T	intronic
	20	31537718	A	AGAATCAATGTCATT	intronic
	20	31538735	TTA	T	UTR3
<i>Glyma.20g084500</i>	20	31666328	GTCA	G	UTR3
	20	31669438	A	AAC	intronic
	20	31673766	TCG	T	intronic
	20	31674346	T	TAA	intronic
	20	31674567	T	TA	intronic
<i>Glyma.20g087000</i>	20	NA	NA	NA	NA

<sup>a</sup>Chromosome.  
<sup>b</sup>Positions with variation in genes between parental lines.  
<sup>c</sup>Female parent of GB13 RIL population.  
<sup>d</sup>Male parent of GB13 RIL population.  
<sup>e</sup>Region of variation in genes.

2016), *GS3* (Fan et al., 2006), *qSW5* (Shomura et al., 2008), *GW5* (Weng et al., 2008), *qGL3* (Zhang et al., 2012), *LGY3* (Liu Q. et al., 2018), *GS5* (Li and Li, 2016), *GS6* (Sun et al., 2013), *GLW7* (Wang et al., 2015), etc., have been cloned. Furthermore, the regulatory pathways of the genes mainly include the ubiquitin-proteasome pathway, G protein signaling pathway, Mitogen-activated protein kinases (MAPK) signaling pathway, hormone signaling pathway and transcription factor regulatory pathway, etc. (Li and Li, 2016). Among them, in the ubiquitin-proteasome pathway, E3 ubiquitin ligase plays a major role and negatively regulates seed width and seed weight (Song et al., 2007). In the hormone signaling pathway, Brassinolide (BR) and Auxin (IAA) can promote cell growth and expansion (Li et al., 2019), it mainly regulates glume development to positively regulate seed size in rice (Li et al., 2011). Therefore, in this study, candidate genes were selected and predicted by screening genes within five clusters based on their specific expression in soybean, combined with homologous or identical functions of genes related to grain size and weight traits in rice.

The major and stable QTL, *qSW-20-3*, which was detected in four environments with PVE varied from 9.22 to 21.93%, was contained in the *qWTP20-2* cluster. By the analysis of sequence variation, we found that among the candidate genes within this QTL interval, only *Glyma.20g084000* and *Glyma.20g084500* were different in the sequence of the parental lines. They were not specifically expressed at the seed maturation stage in soybean, moreover, they have a negative additive effect, but they were associated with RNA methylation and protein ubiquitination, respectively. Therefore, we speculate that the genes within this cluster may indirectly regulate seed size and weight by regulating other traits. While another major and stable QTL, *qSL-3-1*, which was detected in six environments with PVE varied from 4.38 to 15.91%, was contained in the *qLP3-1* cluster. Comparing the sequence variation between parental lines for candidate genes within this cluster, we found that five genes underwent natural variation between parental lines, and all of them were specifically and highly expressed at the seed maturation stage in soybean. Moreover, they have a positive additive effect. Therefore, we speculate that the five candidate genes within the *qLP3-1* cluster regulate seed size and weight positively through certain pathways based on their gene annotation. However, the exact signaling pathway remains to be further investigated.

It is important to study the molecular regulatory network of yield-related traits in soybean. And the loci and candidate genes in this study provide an important theoretical basis and genetic resources for solving the bottleneck problem of soybean yield, which deserves to be further investigated at the molecular level.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

## Author contributions

YC and SL designed the project. SL performed the experiments and drafted the manuscript. RL, ZG, BC, and FL helped the experiment to obtain part of the data. JJ and RW revised the manuscript. SL and QX analyzed the data. HN and ZC validated the manuscripts.

## Funding

This work was supported by the Key-Areas Research and Development Program of Guangdong Province (2020B020220008), the China Agricultural Research System (CARS-04-PS09), the Guangdong Agricultural Research System (2022KJ136), and the Research Project of the State Key Laboratory of Agricultural and Biological Resources Protection and Utilization in Subtropics.

## Acknowledgments

We thank Prof. Liangfa Ge (South China Agricultural University, Guangzhou, China) for critical reading and comments on the manuscript and Long Xiao (Perkins&Will, Vancouver, BC) for constructive suggestions on the language of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1094112/full#supplementary-material>

## References

- Bachlava, E., Dewey, R. E., Burton, J. W., and Cardinal, A. J. (2009). Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. *Crop Sci.* 49 (2), 433–442. doi: 10.2135/cropsci2008.06.0324
- Cai, Z., Cheng, Y., Ma, Z., Liu, X., Ma, Q., Xia, Q., et al. (2018). Fine-mapping of QTLs for individual and total isoflavone content in soybean (*Glycine max* L.) using a high-density genetic map. *Theor. Appl. Genet.* 131 (3), 555–568. doi: 10.1007/s00122-017-3018-x
- Chapman, A., Pantalone, V. R., Ustun, A., Allen, F. L., Landau-Ellis, D., Trigliano, R. N., et al. (2003). Quantitative trait loci for agronomic and seed quality traits in an F<sub>2</sub> and F<sub>4,6</sub> soybean population. *Euphytica.* 129 (3), 387–393. doi: 10.1023/A:1022282726117
- Csanadi, G., Vollmann, J., Stift, G., and Lelley, T. (2001). Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor. Appl. Genet.* 103 (6–7), 912–919. doi: 10.1007/s001220100621
- Cui, B., Chen, L., Yang, Y., and Liao, H. (2020). Genetic analysis and map-based delimitation of a major locus qSS3 for seed size in soybean. *Plant Breeding.* 139 (6), 1145–1157. doi: 10.1111/pbr.12853
- Du, W., Wang, M., Fu, S., and Yu, D. (2009). Mapping QTLs for seed yield and drought susceptibility index in soybean (*Glycine max* L.) across different environments. *J. Genet. Genomics* 36 (12), 721–731. doi: 10.1016/S1673-8527(08)60165-4
- Elattar, M. A., Karikari, B., Li, S., Song, S., Cao, Y., Aslam, M., et al. (2021). Identification and validation of major QTLs, epistatic interactions, and candidate genes for soybean seed shape and weight using two related RIL populations. *Front. Genet.* 12. doi: 10.3389/fgene.2021.666440
- Eskandari, M., Cober, E. R., and Rajcan, I. (2013). Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. *Theor. Appl. Genet.* 126 (2), 483–495. doi: 10.1007/s00122-012-1995-3
- Fan, S., Li, B., Yu, F., Han, F., Yan, S., Wang, L., et al. (2015). Analysis of additive and epistatic quantitative trait loci underlying fatty acid concentrations in soybean seeds across multiple environments. *Euphytica.* 206 (3), 689–700. doi: 10.1007/s10681-015-1491-3
- Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* 112 (6), 1164–1171. doi: 10.1007/s00122-006-0218-1
- Fasoula, V. A., Harris, D. K., and Boerma, H. R. (2004). Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. *Crop Sci.* 44 (4), 1218–1225. doi: 10.2135/cropsci2004.1218
- Fliege, C. E., Ward, R. A., Vogel, P., Nguyen, H., Quach, T., Guo, M., et al. (2022). Fine mapping and cloning of the major seed protein quantitative trait loci on soybean chromosome 20. *Plant J.* 110 (1), 114–128. doi: 10.1111/tpj.15658
- Gao, H. H., Gu, Y., Jiang, H. W., Li, Y. Y., Xin, D. W., Liu, C. Y., et al. (2019). Mapping and analysis of QTLs related to seed length and seed width in *Glycine max*. *Plant Breeding.* 138 (6), 733–740. doi: 10.1111/pbr.12745
- Han, Y., Teng, W., Wang, Y., Zhao, X., Wu, L., Li, D., et al. (2015). Unconditional and conditional QTL underlying the genetic interrelationships between soybean seed isoflavone, and protein or oil contents. *Plant Breeding.* 134 (3), 300–309. doi: 10.1111/pbr.12259
- Han, Y., Xie, D., Teng, W., Sun, J., and Li, W. (2012). QTL underlying developmental behaviour of 100-seed weight of soybean. *Plant Breeding.* 131 (5), 600–606. doi: 10.1111/j.1439-0523.2012.01987.x
- He, Q., Xiang, S., Yang, H., Wang, W., Shu, Y., Li, Z., et al. (2021). A genome-wide association study of seed size, protein content, and oil content using a natural population of sichuan and chongqing soybean. *Euphytica.* 217 (11), 198. doi: 10.1007/s10681-021-02931-8
- Hina, A., Cao, Y., Song, S., Li, S., Sharmin, R. A., Elattar, M. A., et al. (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. doi: 10.3390/ijms21031040
- Hu, Z., Zhang, H., Kan, G., Ma, D., Zhang, D., Shi, G., et al. (2013). Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. merr.). *Genetica.* 141 (4–6), 247–254. doi: 10.1007/s10709-013-9723-8
- Hyten, D. L., Pantalone, V. R., Sams, C. E., Saxton, A. M., Landau-Ellis, D., Stefaniak, T. R., et al. (2004). Seed quality QTL in a prominent soybean population. *Theor. Appl. Genet.* 109 (3), 552–561. doi: 10.1007/s00122-004-1661-5
- Jiang, B. Z., Li, M., Cheng, Y. B., Cai, Z. D., Ma, Q. B., Jiang, Z., et al. (2019). Genetic mapping of powdery mildew resistance genes in soybean by high-throughput genome-wide sequencing. *Theor. Appl. Genet.* 132 (6), 1833–1845. doi: 10.1007/s00122-019-03319-y
- Jia, J., Wang, H., Cai, Z., Wei, R., Huang, J., Xia, Q., et al. (2022). Identification and validation of stable and novel quantitative trait loci for pod shattering in soybean [*Glycine max* (L.) merr.]. *J. Integr. Agr.* 21 (11), 3169–3184. doi: 10.1016/j.jia.2022.08.082
- Kabelka, E. A., Diers, B. W., Fehr, W. R., LeRoy, A. R., Baianu, I. C., You, T., et al. (2004). Putative alleles for increased yield from soybean plant introductions. *Crop Sci.* 44 (3), 784. doi: 10.2135/cropsci2004.0784
- Kato, S., Sayama, T., Fujii, K., Yumoto, S., Kono, Y., Hwang, T., et al. (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. doi: 10.1007/s00122-014-2304-0
- Kulkarni, K. P., Asekova, S., Lee, D., Bilyeu, K., Song, J. T., and Lee, J. (2017). Mapping QTLs for 100-seed weight in an interspecific soybean cross of williams 82 (*Glycine max*) and PI 366121 (*Glycine soja*). *Crop Pasture Sci.* 68 (2), 148–155. doi: 10.1071/CP16246
- Kumawat, G., and Xu, D. (2021). A major and stable quantitative trait locus qSS2 for seed size and shape traits in a soybean RIL population. *Front. Genet.* 12. doi: 10.3389/fgene.2021.646102
- Lee, S. H., Park, K. Y., Lee, H. S., Park, E. H., and Boerma, H. R. (2001). Genetic mapping of QTLs conditioning soybean sprout yield and quality. *Theor. Appl. Genet.* 103 (5), 702–709. doi: 10.1007/s001220100595
- Leite, D. C., Pinheiro, J. B., Campos, J. B., Di Mauro, A. O., and Unêda-Trevisoli, S. H. (2016). QTL mapping of soybean oil content for marker-assisted selection in plant breeding program. *Genet. Mol. Res.* 15 (1), grm7685. doi: 10.4238/gmr.15017685
- Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43 (12), 1266–1269. doi: 10.1038/ng.977
- Li, S., Gao, F., Xie, K., Zeng, X., Cao, Y., Zeng, J., et al. (2016). The OsmiR396c-OsGRF4-OsGIF1 regulatory module determines grain size and yield in rice. *Plant Biotechnol. J.* 14 (11), 2134–2146. doi: 10.1111/pbi.12569
- Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi: 10.1016/j.pbi.2016.05.008
- Li, M., Liu, Y., Wang, C., Yang, X., Li, D., Zhang, X., et al. (2020). Identification of traits contributing to high and stable yields in different soybean varieties across three chinese latitudes. *Front. Plant Sci.* 10, 1642. doi: 10.3389/fpls.2019.01642
- Liu, Q., Han, R., Wu, K., Zhang, J., Ye, Y., Wang, S., et al. (2018). G-Protein  $\beta$  subunits determine grain size through interaction with MADS-domain transcription factors in rice. *Nat. Commun.* 9 (1), 852. doi: 10.1038/s41467-018-03047-9
- Liu, W., Kim, M. Y., Van, K., Lee, Y., Li, H., Liu, X., et al. (2011). QTL identification of yield-related traits and their association with flowering and maturity in soybean. *J. Crop Sci. Biotechnol.* 14 (1), 65–70. doi: 10.1007/s12892-010-0115-7
- Liu, N., Li, M., Hu, X., Ma, Q., Mu, Y., Tan, Z., et al. (2017). Construction of high-density genetic map and QTL mapping of yield-related and two quality traits in soybean RILs population by RAD-sequencing. *BMC Genomics* 18 (1), 466. doi: 10.1186/s12864-017-3854-8
- Liu, D., Yan, Y., Fujita, Y., and Xu, D. (2018). Identification and validation of QTLs for 100-seed weight using chromosome segment substitution lines in soybean. *Breed. Sci.* 68 (4), 442–448. doi: 10.1270/jsbbs.17127
- Li, N., Xu, R., and Li, Y. (2019). Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* 70 (1), 435–463. doi: 10.1146/annurev-arplant-050718-095851
- Li, W., Zheng, D., Van, K., and Lee, S. (2008). QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. merr.). *J. Crop Sci. Biotechnol.* 11 (3), 171–176.
- Mansour, L. M., Lark, K. G., Kross, H., and Oliveira, A. (1993). Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). *Theor. Appl. Genet.* 86 (8), 907–913. doi: 10.1007/BF00211040
- Mao, T., Jiang, Z., Han, Y., Teng, W., Zhao, X., and Li, W. (2013). Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. *Plant Breeding.* 132 (6), 630–641. doi: 10.1111/pbr.12091
- Maughan, P. J., Maroof, M. A., and Buss, G. R. (1996). Molecular-marker analysis of seed-weight: Genomic locations, gene action, and evidence for



- orthologous evolution among three legume species. *Theor. Appl. Genet.* 93 (4), 574–579. doi: 10.1007/BF00417950
- Mccouch, S., Cho, Y., Yano, M., Paul, E., Blinstrub, M., Morishima, H., et al. (1997). Report on QTL nomenclature. *Rice Genet. Newsl.* 14, 11–13.
- Mian, M. A., Bailey, M. A., Tamulonis, J. P., Shipe, E. R., Carter, T. E. J., Parrott, W. A., et al. (1996). Molecular markers associated with seed weight in two soybean populations. *Theor. Appl. Genet.* 93 (7), 1011–1016. doi: 10.1007/BF00230118
- Ning, H., Yuan, J., Dong, Q., Li, W., Xue, H., Wang, Y., et al. (2018). Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean [*Glycine max* (L.) merril]. *PLoS One* 13 (4), e195830. doi: 10.1371/journal.pone.0195830
- Niu, Y., Xu, Y., Liu, X., Yang, S., Wei, S., Xie, F., et al. (2013). Association mapping for seed size and shape traits in soybean cultivars. *Mol. Breeding*. 31 (4), 785–794. doi: 10.1007/s11032-012-9833-5
- Nyquist, W. E., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10 (3), 235–322. doi: 10.1080/07352689109382313
- Palomeque, L., Li-Jun, L., Li, W., Hedges, B., Cober, E. R., and Rajcan, I. (2009). QTL in mega-environments: II. agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. *Theor. Appl. Genet.* 119 (3), 429–436. doi: 10.1007/s00122-009-1048-8
- Panthee, D. R., Pantalone, V. R., West, D. R., Saxton, A. M., and Sams, C. E. (2005). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci.* 45 (5), 2015–2022. doi: 10.2135/cropsci2004.0720
- Pathan, S. M., Vuong, T., Clark, K., Lee, J., Shannon, J. G., Roberts, C. A., et al. (2013). Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Sci.* 53 (3), 765–774. doi: 10.2135/cropsci2012.03.0153
- Qi, Z., Wu, Q., Han, X., Sun, Y., Du, X., Liu, C., et al. (2011). Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica*. 179 (3), 499–514. doi: 10.1007/s10681-011-0386-1
- Qi, Z. M., Zhang, X. Y., Qi, H. D., Xin, D. W., Han, X., Jiang, H. W., et al. (2017). Identification and validation of major QTLs and epistatic interactions for seed oil content in soybeans under multiple environments based on a high-density map. *Euphytica*. 213 (8), 162. doi: 10.1007/s10681-017-1952-y
- Salas, P., Oyarzo-Llaipen, J. C., Wang, D., Chase, K., and Mansur, L. (2006). Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (*Glycine max* L. merr.). *Theor. Appl. Genet.* 113 (8), 1459–1466. doi: 10.1007/s00122-006-0392-1
- Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in PHASEOLUS VULGARIS. *Genetics*. 8 (6), 552–560. doi: 10.1093/genetics/8.6.552
- Sebolt, A. M., Shoemaker, R. C., and Diers, B. W. (2000). Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci.* 40 (5), 1438–1444. doi: 10.2135/cropsci2000.4051438x
- Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., et al. (2008). Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* 40 (8), 1023–1028. doi: 10.1038/ng.169
- Song, X., Huang, W., Shi, M., Zhu, M., and Lin, H. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39 (5), 623–630. doi: 10.1038/ng2014
- Sun, L., Li, X., Fu, Y., Zhu, Z., Tan, L., Liu, F., et al. (2013). GS6, a member of the GRAS gene family, negatively regulates grain size in rice. *J. Integr. Plant Biol.* 55 (10), 938–949. doi: 10.1111/jipb.12062
- Sun, Y. Q., Tian, R., Shao, Z. Q., Chen, S. L., Zhang, H., Jin, Y., et al. (2021). Mining of quantitative trait loci and candidate genes for seed size and shape across multiple environments in soybean (*Glycine max*). *Plant Breeding*. 140 (6), 1058–1069. doi: 10.1111/pbr.12968
- Tajuddin, T. C. F. T., Watanabe, S., Yamanaka, N., and Harada, K. (2003). Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. *Breed. Sci.* 53 (2), 133–140. doi: 10.1270/jsbbs.53.133
- Teng, W., Han, Y., Du, Y., Sun, D., Zhang, Z., Qiu, L., et al. (2009). QTL analyses of seed weight during the development of soybean (*Glycine max* L. merr.). *Heredity*. 102 (4), 372–380. doi: 10.1038/hdy.2008.108
- Teng, W. L., Sui, M. N., Li, W., Wu, D. P., Zhao, X., Li, H. Y., et al. (2018). Identification of quantitative trait loci underlying seed shape in soybean across multiple environments. *J. Agric. Science*. 156 (1), 3–12. doi: 10.1017/S002185961700082X
- Wang, H., Jia, J., Cai, Z., Duan, M., Jiang, Z., Xia, Q., et al. (2022). Identification of quantitative trait loci (QTLs) and candidate genes of seed iron and zinc content in soybean [*Glycine max* (L.) merr.]. *BMC Genomics* 23 (1), 146. doi: 10.1186/s12864-022-08313-1
- Wang, X., Jiang, G., Green, M., Scott, R. A., Song, Q., Hyten, D. L., et al. (2014). Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. *Mol. Genet. Genomics* 289 (5), 935–949. doi: 10.1007/s00438-014-0865-x
- Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., et al. (2015). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* 47 (8), 949–954. doi: 10.1038/ng.3352
- Wang, J., Liu, L., Guo, Y., Wang, Y., Zhang, L., Jin, L., et al. (2014). A dominant locus, qBSC-1, controls  $\beta$  subunit content of seed storage protein in soybean (*Glycine max* (L.) merr.). *J. Integr. Agr.* 13 (9), 1854–1864. doi: 10.1016/S2095-3119(13)60579-1
- Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., et al. (2015). QTL for seed protein and amino acids in the benning × danbaekkong soybean population. *Theor. Appl. Genet.* 128 (5), 839–850. doi: 10.1007/s00122-015-2474-4
- Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., et al. (2008). Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* 18 (12), 1199–1209. doi: 10.1038/cr.2008.307
- Whiting, R. M., Torabi, S., Lukens, L., and Eskandari, M. (2020). Genomic regions associated with important seed quality traits in food-grade soybeans. *BMC Plant Biol.* 20 (1), 485. doi: 10.1186/s12870-020-02681-0
- Xian, P., Cai, Z., Jiang, B., Xia, Q., Cheng, Y., Yang, Y., et al. (2022). GmRmd1 encodes a TIR-NBS-BSP protein and confers resistance to powdery mildew in soybean. *Plant Commun.* 3 (6), 100418. doi: 10.1016/j.xplc.2022.100418
- Xu, Y., Li, H., Li, G., Wang, X., Cheng, L., and Zhang, Y. (2011). Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. merr.). *Theor. Appl. Genet.* 122 (3), 581–594. doi: 10.1007/s00122-010-1471-x
- Yang, K., Moon, J., Jeong, N., Chun, H., Kang, S., Back, K., et al. (2011). Novel major quantitative trait loci regulating the content of isoflavone in soybean seeds. *Genes Genom.* 33 (6), 685–692. doi: 10.1007/s13258-011-0043-z
- Yang, H., Wang, W., He, Q., Xiang, S., Tian, D., Zhao, T., et al. (2017). Chromosome segment detection for seed size and shape traits using an improved population of wild soybean chromosome segment substitution lines. *Physiol. Mol. Plant Biol.* 23 (4), 877–889. doi: 10.1007/s12298-017-0468-1
- Yan, L., Li, Y., Yang, C., Ren, S., Chang, R., Zhang, M., et al. (2014). Identification and validation of an over-dominant QTL controlling soybean seed weight using populations derived from *Glycine max* × *Glycine soja*. *Plant Breeding*. 133 (5), 632–637. doi: 10.1111/pbr.12197
- Yao, D., Liu, Z. Z., Zhang, J., Liu, S. Y., Qu, J., Guan, S. Y., et al. (2015). Analysis of quantitative trait loci for main plant traits in soybean. *Genet. Mol. Res.* 14 (2), 6101–6109. doi: 10.4238/2015.June.8.8
- Yesudas, C. R., Bashir, R., Geisler, M. B., and Lightfoot, D. A. (2013). Identification of germplasm with stacked QTL underlying seed traits in an inbred soybean population from cultivars Essex and Forrest. *Mol. Breeding*. 31 (3), 693–703. doi: 10.1007/s11032-012-9827-3
- Yue, Y., Liu, N., Jiang, B., Li, M., Wang, H., Jiang, Z., et al. (2017). A single nucleotide deletion in j encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. *Mol. Plant* 10 (4), 656–658. doi: 10.1016/j.molp.2016.12.004
- Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genet. (Austin)*. 136 (4), 1457–1468. doi: 10.1093/genetics/136.4.1457
- Zhang, J., Song, Q., Cregan, P. B., and Jiang, G. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* 129 (1), 117–130. doi: 10.1007/s00122-015-2614-x
- Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., et al. (2012). Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci.* 109 (52), 21534–21539. doi: 10.1073/pnas.1219776110