



OPEN ACCESS

EDITED BY

Ainong Shi,
University of Arkansas, United States

REVIEWED BY

Jiban Shrestha,
Nepal Agricultural Research Council,
Nepal
Atta Ullah,
Balochistan Agriculture College, Pakistan

*CORRESPONDENCE

Lijuan Qiu

✉ qiu lijuan@caas.cn

Jun Wang

✉ wjsoybean_2008@163.com

[†]These authors have contributed equally to this work

RECEIVED 20 February 2026

REVISED 15 April 2026

ACCEPTED 16 April 2026

PUBLISHED 12 May 2026

CITATION

Liu K, Yu H, Jadoon S, Shen M, Xiong Y, Chen Y, Liu Z, Liu J, Gyebe-Nimako B, Zhang Y, Zhang F, Qiu L and Wang J (2026) Phenotypic and genetic dissection of seed shape related traits in soybean (*Glycine max* L.). *Front. Plant Sci.* 17:1814388. doi: 10.3389/fpls.2026.1814388

COPYRIGHT

© 2026 Liu, Yu, Jadoon, Shen, Xiong, Chen, Liu, Liu, Gyebe-Nimako, Zhang, Zhang, Qiu and Wang. 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.

Phenotypic and genetic dissection of seed shape related traits in soybean (*Glycine max* L.)

Kanglin Liu^{1†}, Huan Yu^{1†}, Sawaira Jadoon^{1†}, Maoting Shen¹, Yajun Xiong¹, Yijie Chen¹, Zhiyu Liu¹, Junjun Liu¹, Bernard Gyebe-Nimako¹, Yu Zhang¹, Fan Zhang¹, Lijuan Qiu^{2*} and Jun Wang^{1,2*}

¹The Shennong Laboratory/Institute of Crop Molecular Breeding, Henan Academy of Agricultural Sciences, Zhengzhou, China, ²State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

Introduction: Seed shape and hundred-seed weight (HSW) are yield-related traits, but the interrelationships between them and a systematic genetic analysis are lacking.

Methods: In this study, phenotypic evaluation of seed shape-related traits and HSW was performed in a recombinant inbred line (RIL) population derived from the cross between “ZD41” and “ZYD02878.” All seed shape-related traits (SW, seed width; SL, seed length; ST, seed thickness; SP, seed perimeter; SA, seed area; SLW, seed length-to-width ratio; SLT, seed length-to-thickness ratio; SWT, seed width to thickness ratio) were evaluated. Broad-sense heritability was calculated. Correlation analysis, path analysis, QTL mapping (using ICIM, EMMAX, and TASSEL), and genomic selection (using rrBLUP) were performed.

Results: All seed shape-related traits exhibited approximately normal distributions in the RIL population. The broad-sense heritability of seed shape-related traits ranged from 0.67 to 0.90, indicating that additive genes played a major role in expressing all studied traits. Correlation analysis revealed that SW showed the highest positive correlation with HSW ($r = 0.86-0.87$, $P < 0.001$) among the three measured seed shape-related traits (SL, SW, and ST), which agrees with the finding that SW was the only trait showing a positive direct effect on HSW in path analysis. A total of 98 QTL corresponding to 59 unique regions were identified by ICIM, EMMAX, and TASSEL. Three genes, *Glyma.02G255900*, *Glyma.02G270100*, and *Glyma.02G275200*, were identified as candidates for seed shape regulation. Genomic selection for seed shape-related traits using rrBLUP revealed that the prediction accuracy of SL, SW, ST, SP, and SA reached 0.65, 0.68, 0.65, 0.67, and 0.68, respectively, based on genome-wide SNP markers.

Discussion: This study provided a systematic dissection of seed shape-related traits in terms of both phenotypic and genetic aspects, which laid a solid foundation for gene cloning and could be used in breeding programs in the future.

KEYWORDS

genetic dissection, genomic selection, QTL mapping, seed shape related traits, soybean, trait network

1 Introduction

Soybean is a globally significant source of plant-based protein and vegetable oil (Qin et al., 2022; Zhang et al., 2022; Fang et al., 2017; Hu et al., 2023b). In 2024, China imported more than 100 million tons of soybeans in order to meet the domestic need of the country (<https://www.stats.gov.cn/sj/>). Increasing domestic soybean yield represents a critical strategy for reducing this reliance on imports (Liu et al., 2025). Hundred-seed weight is a key agronomic trait determining crop yield (Vogel et al., 2021; Hu et al., 2023b). Seed shape-related traits, including seed length (SL), seed width (SW), and seed thickness (ST), are major factors influencing both hundred-seed weight (HSW) and appearance quality in soybean (Liang et al., 2005; Kumawat and Xu, 2021).

To date, over 300 QTL associated with seed shape have been reported by various researchers (www.soybase.org). Recombinant inbred lines (RILs) and chromosome segment substitution lines (CSSLs) are commonly used as mapping populations for quantitative trait loci (QTL) mapping of soybean seed size (Cheng et al., 2016; Hina et al., 2020; Kumar et al., 2023; Liang et al., 2008; Luo et al., 2023; Teng et al., 2018, 2019; Kumawat and Xu, 2021; Moongkanna et al., 2011; Salas et al., 2006; Yang et al., 2017; Zhang et al., 2025). Using those populations, two pleiotropic QTL associated with SL, S21–S22 and O23–O19/O19–O21, have been fine-mapped (Niu et al., 2013a). Analysis of a natural population comprising 257 accessions from six geographical ecoregions in China identified 59 QTL exhibiting genotype-by-environment interactions (Niu et al., 2013b). Linkage analysis in an RIL population of 184 lines detected 12 QTL, of which five were related to seed shape, while association mapping in 219 cultivated soybean accessions identified 41 significant SNP loci, of which seven were related to seed shape (Hu et al., 2013). Using RIL and sub-F₂ populations across three environments, a major seed shape QTL (*qSS3*) was mapped, and the effects were validated in NIL populations (Cui et al., 2020). Integrating findings from multiple studies, a meta-analysis localized 23 consensus QTL onto 17 linkage groups, and 37 candidate genes were designated (Gao et al., 2019).

To date, at least 17 genes have been identified that directly regulate soybean seed traits, including SL, SW, ST, and SP (Duan et al., 2023). Knocking out *GmGA3ox1* suppressed gibberellin (GA) biosynthesis in soybean, leading to significantly reduced seed length without affecting seed thickness and width (Hu et al., 2022). *GmPDAT*, involved in acyl lipid metabolism, was reported to regulate oil deposition as well as seed length and thickness (Liu et al., 2020). Genes that positively regulate SL and SW include *GmST05*, *GmUSPL1*, and *GmJAZ3*, while *Dt2* acts as a negative regulator of both SL and SW (Duan et al., 2022; Liang et al., 2022; Hu et al., 2023a; Yuan et al., 2024). Overexpression of *GmMFT* and *GmSW17* resulted in significant increases in seed length, width, and thickness (Cai et al., 2023; Liang et al., 2024). Conversely, plants overexpressing *SW16.1* exhibited a significant decrease in SL, SW, and ST (Chen et al., 2023a). Furthermore, *Gmdtm1-1* and *Gmdtm1-2* had smaller seed area and perimeter than Williams 82 due to *GmNAPI* mutations (Tang et al., 2020).

Genomic selection (GS), initially proposed for animal breeding, has since been extensively studied in crop science (Meuwissen et al.,

2001). The application of GS to complex traits may enable a faster rate of genetic improvement over time (Ravelombola et al., 2021). The factors influencing GS prediction accuracy include trait heritability, marker number and density, genetic relationship between the training population and the breeding population, training population size, and modeling methods (Alemu et al., 2024). GS has been widely applied in soybean for traits such as yield, plant height, seed weight, protein and oil content, as well as amino acid composition (Ma et al., 2018; Qin et al., 2019; Bandillo et al., 2023; Van der Laan et al., 2025). However, scarce studies have been conducted on GS for seed shape-related traits, which need to be investigated.

To address these questions, in this study, QTL identification related to seed shape-related traits was performed using both linkage and association analyses by phenotyping across two distinct environments over a RIL population derived from the cross between large-seeded cultivated soybean cultivar “ZD41” and the small-seeded wild soybean “ZDY02878,” and then candidate genes were screened based on SNP variation, haplotype analysis, and expression analysis with the purpose of delving into the relationship between seed shape-related traits and hundred-seed weight. This study provided a solid genetic basis for the molecular cloning of genes regulating seed shape.

2 Materials and methods

2.1 Plant materials

The RIL population, consisting of 364 lines, was developed previously and described (Chen et al., 2023b). Basically, the maternal and paternal parents of the RIL are ZD41 and ZDY02878, with significant difference in seed size. The RIL-F₈ and F₉ were planted at Yangtze University (30.37°N, 112.06°E) and A’jian Farm (34.41°N, 115.98°E) in the summers of 2020 and 2021, respectively, and hereafter were designated as 20JZ and 21SQ. The spaced planting design with uniform 1-meter row spacing was used (Wang et al., 2024). The NAM population, constructed from crosses between ZD41 and 35 paternal parents, was planted in 20JZ for candidate gene haplotype analysis (Song et al., 2024).

2.2 Phenotyping of seed shape related traits

Matured seeds from one randomly selected plant per line from both the RIL and NAM populations were threshed. Seed shape data—including measured traits, namely seed area (SA, mm²), seed perimeter (SP, mm), seed length (SL, mm), seed width (SW, mm), seed thickness (ST, mm), as well as derived traits, namely seed length-to-width (SLW), seed length-to-thickness (SLT), and seed width-to-thickness (SWT), were collected using a seed shape scanner (SC-G, Hangzhou Wanshen Detection Technology Co., Ltd., China) (Ren et al., 2021). Generally, the seeds on the plate were scanned to capture a two-dimensional image, in which the thinnest dimension (seed thickness) of the seeds was placed vertically; then,

the length, width, and area were calculated using models, and the thickness of the seeds was obtained by determining the width of seeds fixed by two large combs, in which the seeds were placed linearly between comb teeth and the thinnest dimension was placed horizontally. Hundred-seed weight (HSW, g) was measured using a 0.01 g precision balance.

2.3 Statistical analysis

First, the 3-sigma principle was applied to remove outliers using R code (Khan et al., 2021). Descriptive statistical analysis of seed shape-related traits was then performed using R packages such as Hmisc (Harrell, 2003). Transgressive segregation was defined as the phenotype value of RIL population individuals being higher or lower than that of the parents. Correlation analysis was conducted using functions such as Pearson within the Hmisc (5.2-3) R package (Khan et al., 2021). Principal Component Analysis (PCA) was carried out using factoMineR (2.11) and factoextra (1.0.7) (Irnawati et al., 2021). Path analysis was implemented using the sem (3.1-16), lavaan (0.6-19), and agricolae (1.3-7) R packages, employing the average values of HSW and other seed shape-related traits across two environments, with HSW as the dependent variable (Y) and the other seed shape-related traits as independent variables. Total effect, direct effect, and indirect effect were used to characterize path coefficients, and the top three total effects among all traits were plotted in a network using visNetwork (Fox, 2006; Rosseel, 2012; De Mendiburu, 2019). Based on Pearson correlation analysis, Plant Trait Networks (PTN) were constructed among different traits with default parameters of $r = 0.2$ and $P = 0.05$ using the MultiTraits R package, and the network is characterized by the metrics of degree (K), closeness (C), betweenness (B), and clustering coefficient (CC) (He et al., 2020b; Li et al., 2022).

The variance among the RIL population and across different environments was estimated by two-way ANOVA (Nandigam et al., 2025). The broad-sense heritability (h^2) of each seed shape-related trait was calculated using the following formula:

$$h^2 = \frac{V_g}{V_g + \frac{V_{gl}}{L} + \frac{V_e}{PL}}$$

where V_g represents the genotypic variance, V_{gl} denotes the genotype-by-environment interaction variance, V_e is the error variance, L indicates the number of environments, and P signifies the number of replications. Broad-sense heritability calculations were implemented using a completely randomized model via the lme4 (1.1-35.5) package (Bates et al., 2015).

2.4 Genotyping and QTL mapping

This RIL population was planted at the College of Agriculture, Yangtze University, in 2022. Young leaves were collected at the V5 stage for genomic DNA extraction. DNA was extracted using the CTAB method (Doyle and Doyle, 1990). Genotyping was conducted using a 200K SNP chip by Beijing Compass Biotechnology Co., Ltd (Song et al., 2024). The raw genotype data were then sorted, filtered, and imputed using PLINK (Purcell et al., 2007), TASSEL, and Beagle software, respectively, resulting in 85,313 high-quality

SNPs (minor allele frequency <0.05, site completeness >0.8). Based on the high-density genetic map, ICIM analysis was performed using the BIP module in QTL IciMapping software to execute ICIM-ADD, with the LOD threshold set to 2.5, a scanning step of 1 cM, and other parameters set to default (Meng et al., 2015). Based on the Mixed Linear Model (MLM), the kinship matrix was calculated using emmax-kin-inter64 in EMMAX as a covariate; the top five principal components (PCs) were computed using GCTA and incorporated as covariates along with the kinship matrix into the association study using TASSEL. The significance threshold for both was set at $-\log_{10}(1/m)$ ($m = 85,313$). Significant loci were selected, with flanking 100 kb regions defined as candidate intervals, and Manhattan and QQ plots were generated using the R package CMplot (Chen et al., 2023b).

2.5 Candidate gene screening

Using the Wm82.a2.v1 genome as a reference, LD analysis of association regions was performed with the R package IntAssoPlot (He et al., 2020a). The target interval was narrowed based on r^2 (>0.8), and genes containing four mutation types—nonsynonymous, splicing, stopgain, and stoploss—were further screened through single nucleotide polymorphism (SNP) variation analysis. For potential candidate genes, expression profile datasets from the public soybean transcriptome database PPRD were used to analyze gene expression levels in different tissues (root, stem, leaf, seed). Expression data were normalized and visualized as heatmaps using the R package pheatmap. Haplotype analysis was conducted using the HaploAssistant (Yu et al., 2024). Haplotypes with fewer than 10 genotypes were excluded. The significance of phenotypic differences between haplotypes was tested using the Wilcoxon rank-sum test. Combining candidate gene expression profiles and haplotype analysis results, candidate genes were functionally annotated based on the SoyBase database to identify genes regulating seed development and shape formation.

2.6 Genomic selection

To evaluate the impact of different marker sets on prediction accuracy in genomic selection (GS) for seed shape-related traits, we constructed three marker sets: marker set 1 (M1, 85,313 genome-wide SNPs), marker set 2 (M2, 5,101 SNPs linked to seed shape-related QTL), and marker set 3 (M3, 80,212 genome-wide SNPs excluding trait-associated SNPs). Models were evaluated by 5-fold cross-validation for 20 iterations. All marker sets were treated as random effects in rrBLUP (Endelman, 2011).

3 Results

3.1 Phenotypic analysis of seed shape-related traits

The 364 RILs used in this study were planted in two environments (20JZ and 21SQ) and phenotyped for seed shape-related traits. After removing outliers, the number of valid measurements

for seed shape-related traits ranged from 337 to 364. The results showed that “ZD41” had significantly larger mean values for SL (7.32–7.64 vs 3.34–3.97), SW (6.23–6.47 vs 2.73–3.02), ST (5.98–6.16 vs 2.08–2.35), SA (38.30–38.60 vs 7.12–9.13), and SP (22.73–23.65 vs 10.15–11.67) than “ZYD02878.” In contrast, “ZD41” exhibited lower values for key shape descriptors—specifically, the SLW (1.17–1.18 vs 1.22–1.31), SLT (1.22–1.24 vs 1.61–1.69), and SWT (1.04–1.05 vs 1.29–1.31) (Table 1). These findings collectively demonstrated that “ZD41” produces seeds that are both larger and more spherical than those of the wild soybean “ZYD02878.” All seed shape-related traits of the RIL population in 20JZ and 21SQ approximately followed a normal distribution (Figure 1). Except for ST in 21SQ, all traits showed skewness >0.00 and kurtosis <0.00, exhibiting platykurtic and right-skewed distributions (Table 1; Figures 1A–H). For the parental lines (ZYD02878 and ZD41), the range of all measured traits (excluding SWT) were broader in the 20JZ than in 21SQ. Likewise, in the RIL population, 20JZ displayed broader ranges and higher mean values for every trait except SWT compared with 21SQ. Transgressive segregation was observed in the derived traits, e.g., SLT, SLW, and SWT, but not in measured traits (SL, SW, ST, SP, and SA) (Table 1). One-way ANOVA revealed that >70% of the total variance observed in each trait (SA (84%), SP (83%), SL (73%), SW (90%), ST (86%), SLW (72%), SLT (77%), SWT (72%)) was attributable to variance among genotypes, suggesting that the variation observed was mostly genetic rather than environmental (Appendix A). Similarly, SW exhibited the highest heritability (0.90), followed by SA (0.89), SP (0.88), ST (0.87), SL (0.83), SLW (0.81), and SLT (0.75), with SWT showing the lowest (0.67) (Table 1). This suggests that the expression of these traits is stable across different environments and is predominantly influenced by genetic variation.

3.2 Interrelationships between HSW and seed shape related traits

Pearson correlation coefficients among different seed-related traits revealed that HSW, SA, SP, SL, SW, and ST exhibited significant positive correlations ($r = 0.65$ – 0.98 , $P < 0.001$) across two environments. Among these, SA and SP consistently showed the strongest correlation ($r = 0.98$, $P < 0.001$). HSW exhibited strong positive correlations ($r > 0.79$, $P < 0.001$) with the directly measured traits, specifically SA, SP, SW, ST, and SL, in descending order (Figures 2A, B). In contrast, the traits weakly associated with HSW were the derived traits SLW, SLT, and SWT.

Based on the correlation analysis, a plant trait network was constructed. The results indicated that ST, SL, and SW possessed the three highest values for degree (K), closeness (C), and betweenness centrality (B), identifying them as the central hubs within the network (Figure 2C; Appendix B). Within this network, HSW was directly connected to all five measured traits (SL, SW, ST, SP, and SA) but not to the derived traits (SLT, SWT, and SLW). This network architecture is consistent with the results of the correlation analysis.

Path analysis revealed that all five measured seed shape traits had positive total effects on HSW, ranging from 0.81 to 0.93. Among these, only SW exhibited a positive direct effect, while the

other traits influenced HSW indirectly (Figure 2D; Appendix C). In contrast, all derived traits showed negative total effects on HSW.

3.3 QTL identification for seed shape related traits

In this study, we employed ICIM for linkage analysis and EMMAX and TASSEL for genome-wide association studies to identify quantitative trait loci (QTL) regulating seed shape. A total of 98 QTL corresponding to 68 unique QTL loci were identified by all three methods (Table 2; Appendixes D, E, F, G). Linkage mapping with ICIM detected a total of 44 QTL associated with seed-related traits (Figure 3; Appendixes D). The number of QTL and their collectively explained phenotypic variation for each trait were as follows: eight for SL (29.67%), five for SW (26.59%), nine for ST (43.76%), seven for SP (29.66%), seven for SA (36.83%), four for SLT (15.87%), two for SLW (8.44%), and two for SWT (7.21%). Through EMMAX analysis, a total of 10 seed shape-associated QTL were identified on chromosomes 2, 5, 6, and 8, including four QTL located on each of chromosomes 2 and 8 (Figure 4; Appendix E). TASSEL-based mapping detected 44 QTL, which were predominantly distributed on chromosomes 2, 8, 9, 13, and 18, with 5, 11, 4, 5, and 4 QTL identified on these chromosomes, respectively (Figure 5; Appendixes F, I).

Among these identified QTL, 21 genomic regions overlapped with 39 QTL detected by multiple mapping methods or for different traits. These overlapping QTL are considered robust, stable, and highly confident, as their repeated identification underscores their reliability (Appendix G). Notably, the locus *qSLW2-1* (Chr02: 43,973,626–46,364,051) was co-localized with several other QTL—*qSLT2-1*, *qST2-2*, *qSLW2-2*, *qSLW2-3*, and *qSLW2-4*—using both mapping methods and across two environments (20JZ and 21SQ). Among these, *qSLW2-4* displayed the highest phenotypic variance explained (PVE) value of 9.83%. Furthermore, *qSWT5-1* (Chr05: 40,434,511–42,218,304) was consistently detected by both methods in the 21SQ environment, explaining 7.52% of the phenotypic variation in SWT; this locus was designated *qSWT5-2*. Similarly, *qSA8-1* (Chr08: 35,457,204–36,135,707) overlapped with *qSP8-1*, *qSW8-2*, and *qST8-4* across analytical methods and environments. Of these, *qSW8-2* exhibited the highest PVE value, at 8.01% (Appendixes D, E, F, G).

3.4 Candidate gene mining

qSWT5-1 and *qSA8-1* were consistently detected by both approaches and exhibited high PVE. Notably, within the *qSWT5-1* interval (Chr05:40,434,511–42,218,304), the previously cloned gene *GmST05* (Chr05:41,852,863–41,854,961) (Duan et al., 2022) was physically located. Linkage disequilibrium (LD) analysis of *qSLW2-1* revealed an LD block spanning Chr02:43,973,626–46,363,056 encompassing 285 genes. For *qSA8-1*, one LD block (Chr08:35,652,893–35,848,556) was identified to contain 11 genes (Figure 6; Appendixes E).

SNP variation analysis revealed that, among the aforementioned 296 genes, 291 contained SNPs classified as nonsynonymous, stop-gain, stop-loss (loss of stop codon), or alternative splicing mutations (Appendix H). Gene expression analysis showed that 12 of those 291

TABLE 1 Descriptive statistics and broad-sense heritability of seed shape-related traits.

Trait	Env.	P1	P2	Max	Min	Mean	Sd	Skenes	Kurtosis	P Value	h ²
SL (mm)	20JZ	7.64	3.97	6.86	4.14	5.49	0.51	0.29	−0.23	0.00	0.83
	21SQ	7.32	3.34	6.30	3.99	5.06	0.45	0.32	−0.28	0.00	
SW (mm)	20JZ	6.47	3.02	5.14	3.34	4.19	0.34	0.03	−0.28	0.32	0.90
	21SQ	6.23	2.73	5.12	3.12	4.11	0.37	0.19	−0.17	0.14	
ST (mm)	20JZ	6.16	2.35	4.74	2.83	3.73	0.37	0.18	−0.33	0.14	0.87
	21SQ	5.98	2.08	4.46	2.66	3.56	0.35	−0.02	−0.31	0.23	
SP (mm)	20JZ	23.65	11.67	19.80	12.47	16.23	1.33	0.28	−0.22	0.00	0.88
	21SQ	22.70	10.15	18.75	12.39	15.38	1.22	0.29	−0.33	0.02	
SA (mm2)	20JZ	38.60	9.43	25.62	10.63	18.11	2.84	0.33	−0.26	0.00	0.89
	21SQ	38.30	7.12	23.46	10.62	16.35	2.61	0.39	−0.26	0.00	
SLW	20JZ	1.18	1.31	1.56	1.10	1.31	0.09	0.25	−0.45	0.00	0.81
	21SQ	1.17	1.22	1.43	1.07	1.22	0.08	0.59	−0.31	0.00	
SLT	20JZ	1.24	1.69	1.80	1.18	1.47	0.12	0.22	−0.27	0.02	0.75
	21SQ	1.22	1.61	1.67	1.12	1.40	0.11	0.17	−0.27	0.02	
SWT	20JZ	1.05	1.29	1.29	0.98	1.13	0.06	0.31	−0.36	0.00	0.67
	21SQ	1.04	1.31	1.30	1.00	1.14	0.06	0.21	−0.13	0.04	

SA, seed area; SP, seed perimeter; SL, seed length; SW, seed width; SLT, seed length to thickness; SWT, seed width to thickness; SLW, seed length to width.

TABLE 2 Statistic of QTL identified by different methods.

Traits	20JZ	21SQ	Total QTL	Chromosome	Total PVE (%)	Method
SL	2	6	8	2; 3; 4; 5; 8; 9; 11; 17	29.67	ICIM
SW	2	3	5	7; 10; 11; 17	26.59	ICIM
ST	5	4	9	3; 11; 14; 17; 20	43.76	ICIM
SP	3	4	7	4; 5; 9; 10; 11; 19; 20	29.66	ICIM
SA	3	4	7	9; 10; 11; 20	36.83	ICIM
SLT	1	3	4	5; 6; 11	15.87	ICIM
SLW	1	1	2	6	8.44	ICIM
SWT	0	2	2	5; 6	7.21	ICIM
SL	0	0	0	NA	NA	EMMAX
SW	0	0	0	NA	NA	EMMAX
ST	1	0	1	6	NA	EMMAX
SP	0	1	1	8	NA	EMMAX
SA	0	1	1	8	NA	EMMAX
SLT	1	1	2	2; 8	NA	EMMAX
SLW	2	1	3	2; 8	NA	EMMAX
SWT	0	2	2	2; 5	NA	EMMAX
SL	0	0	0	NA	NA	TASSEL
SW	5	3	8	2; 8; 13; 18; 19; 20	62.46	TASSEL
ST	4	9	13	1; 2; 3; 5; 7; 8; 18; 20	92.44	TASSEL
SP	12	1	13	5; 7; 8; 9; 10; 12; 13; 15; 17; 18	97.69	TASSEL
SA	6	1	7	8; 9; 13	50.21	TASSEL
SLT	0	0	0	NA	NA	TASSEL
SLW	1	1	2	2	17.3	TASSEL
SWT	0	1	1	5	7.52	TASSEL

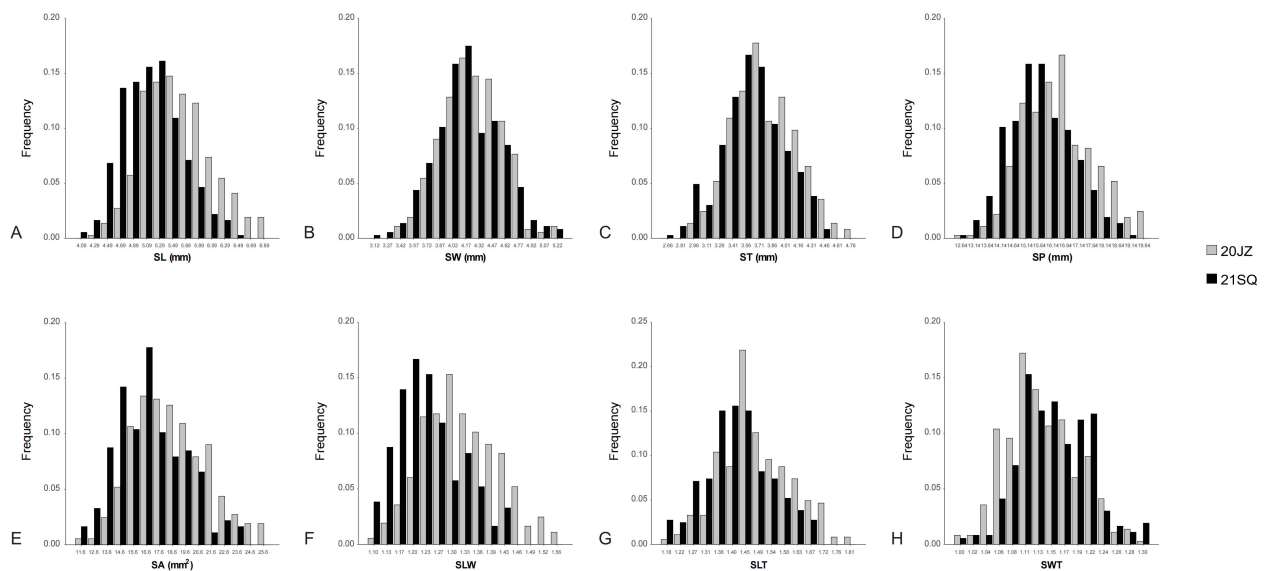


FIGURE 1

Frequency distribution of seed shape-related traits. (A–H) represent the frequency distribution of SA, SP, SL, SW, ST, SLW, SLT, and SWT, respectively, under two environments (20JZ and 21SQ). SA, seed area; SP, seed perimeter; SL, seed length; SW, seed width; SLT, seed length to seed thickness; SWT, seed width to seed thickness.

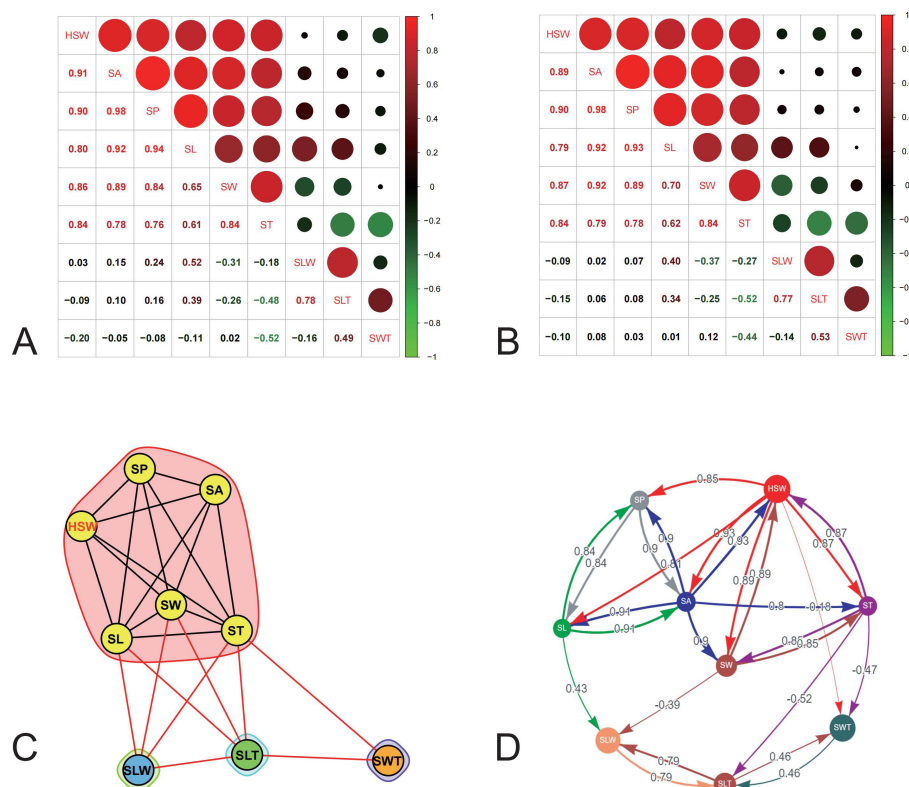


FIGURE 2

The correlation between HSW and seed shape-related traits. (A, B) Correlation heatmap of seed shape-related traits and HSW in 20JZ and 21SQ. (C) Plant trait network of seed shape-related traits and HSW, which was generated using the MultiTraits R package. (D) Path analysis of seed shape-related traits and HSW. Top three total effects of each target trait were displayed in network using visNetwork. SA, seed area; SP, seed perimeter; SL, seed length; SW, seed width; SLT, seed length to thickness; SWT, seed width to thickness; SLW, seed length to width.

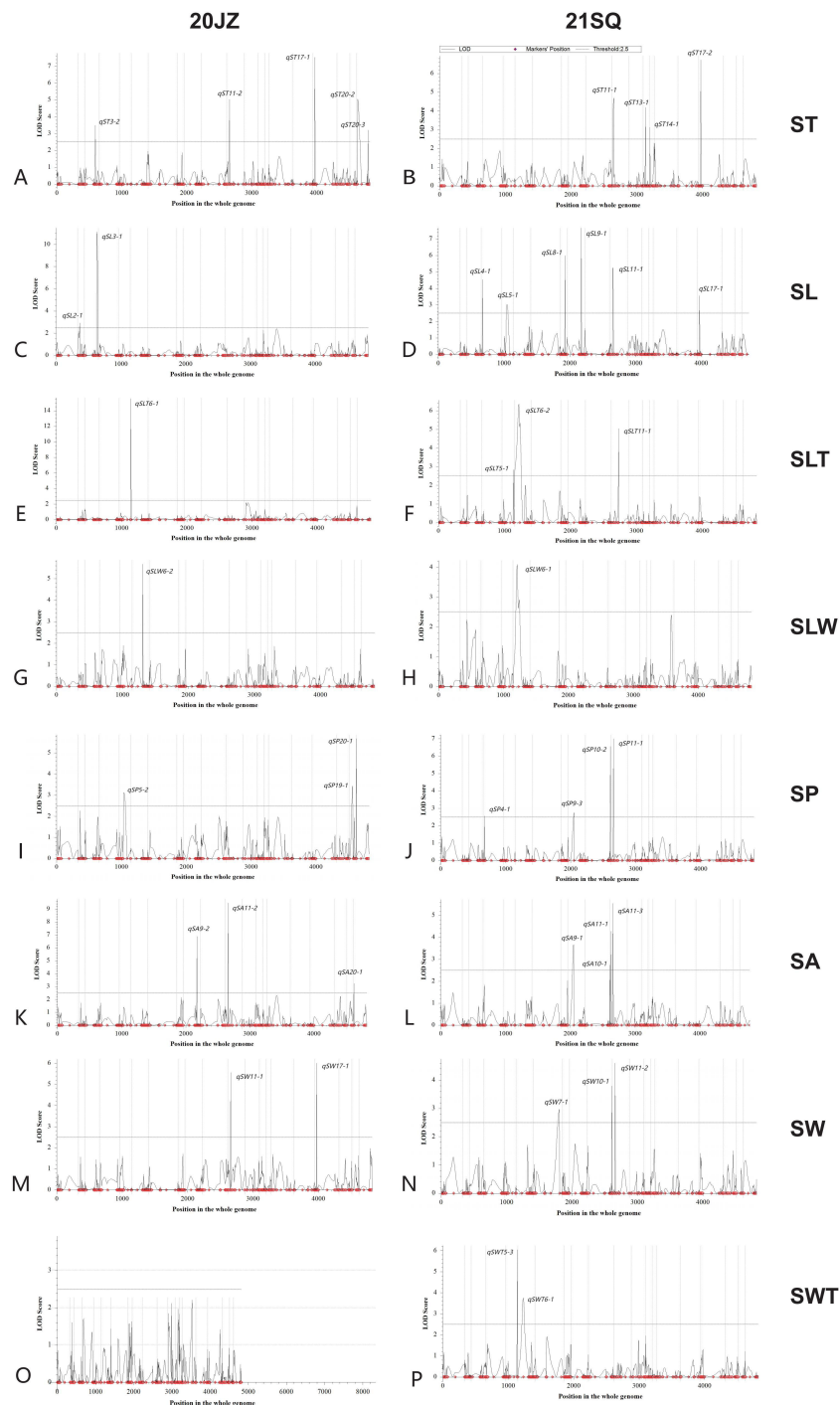


FIGURE 3

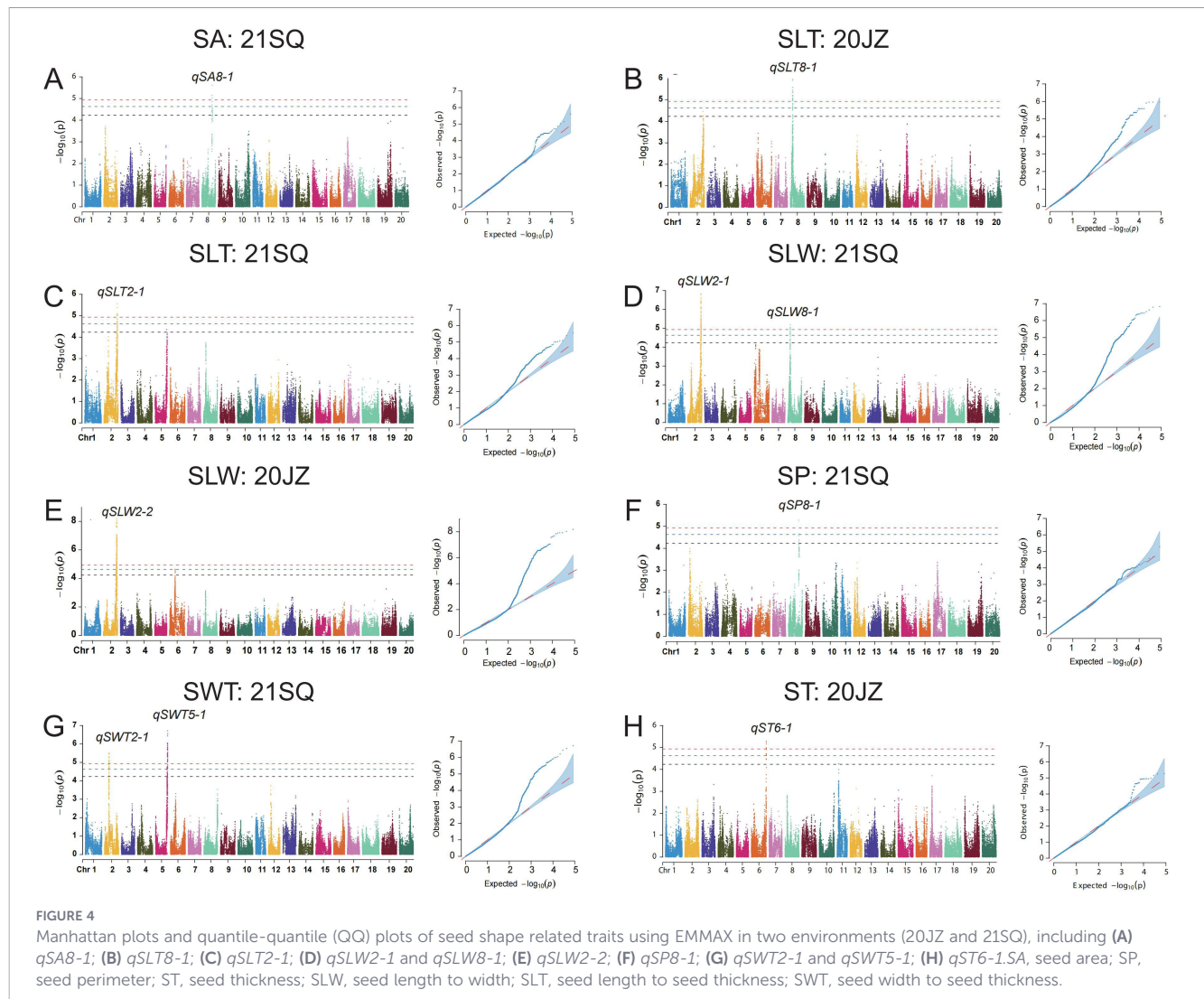
Seed traits related QTL mapped for (A, B) ST, seed thickness; (C, D) SL, seed length; (E, F) SLT, seed length to seed thickness; (G, H) SLW, seed length to width; (I, J) SP, seed perimeter; (K, L) SA, seed area; (M, N) SW, seed width; (O, P) SWT, seed width to seed thickness by ICIM in two environments (20JZ and 21SQ).

genes were seed-specific expressed (Appendix I). Based on the phenotypic and genotypic data of the NAM population, haplotypes of 291 genes were analyzed. A total of 682 haplotypes were inferred for 217 genes, of which 154 genes showed significant differences between haplotypes (Appendix J). Taken together, three genes exclusively expressed in seeds, namely *Glyma.02G255900*, *Glyma.02G270100*, and *Glyma.02G275200* showed significant phenotypic differences between haplotypes as well (Figure 7; Appendixes I, K), which were functionally characterized as nucleolar histone methyltransferase-

related protein, FAM91A1, and early nodulin-like protein 20, respectively, and were therefore selected as candidate genes for seed shape regulation.

3.5 Genomic selection of seed shape related traits

To construct the GS model for seed-related traits, three different marker sets, namely genome-wide markers (M1), trait-associated



markers (M2), and genome-wide markers excluding trait-associated markers (M3), were used. Results showed that the trait-associated marker set (M2) showed relatively lower prediction accuracy compared to genome-wide markers (M1 and M3). For the measured seed shape traits (SL, SW, ST, SP, and SA), the prediction accuracy by genome-wide markers reached 0.60–0.68. This represents a 40.2% to 60.8% increase over the accuracy achieved using the trait-associated marker set M2 (0.35–0.49). Minimal differences in accuracy were observed between the genome-wide markers and sets M1 and M3 (Figure 8; Appendix L). The derived traits (SLT, SLW, and SWT) exhibited lower prediction accuracy, ranging from 0.41 to 0.52, in contrast to the higher accuracy achieved for the directly measured traits. The prediction accuracy of genomic selection showed no discernible pattern across the different environments.

4 Discussion

In this study, seed shape was characterized by eight related traits, namely SL, SW, ST, SP, SA, SLT, SLW, and SWT, of which

SLT, SLW, and SWT are derived traits obtained by division among SL, SW, and ST. Of the five measured traits, all exhibited high correlations with each other (0.61–0.98) (Figures 2A, B), clustered together and distinctly separated from derived traits in the plant trait network (Figure 2C). SA and SP are also derived from seed length and width through calculation. In this context, SL, SW, and ST are the key components determining seed size and shape. Correlation analysis revealed that SL exhibited the weakest correlation with HSW among the three seed shape-related traits in different environments (20JZ and 21SQ) (Figures 2A, B). Meanwhile, path analysis indicated that all measured traits (SL, SW, ST, SP, and SA) had positive total effects on hundred-seed weight, while SW was the only trait exhibiting a positive direct effect (Figure 2D; Appendix C), suggesting its decisive role in determining soybean HSW in this study.

However, this pattern is not universally consistent. Multiple studies have reported significant correlations between seed thickness (ST) and HSW under varying environmental conditions. For instance, in a recombinant inbred line (RIL) population developed from a cross between Dongnong46 and L-100, HSW demonstrated

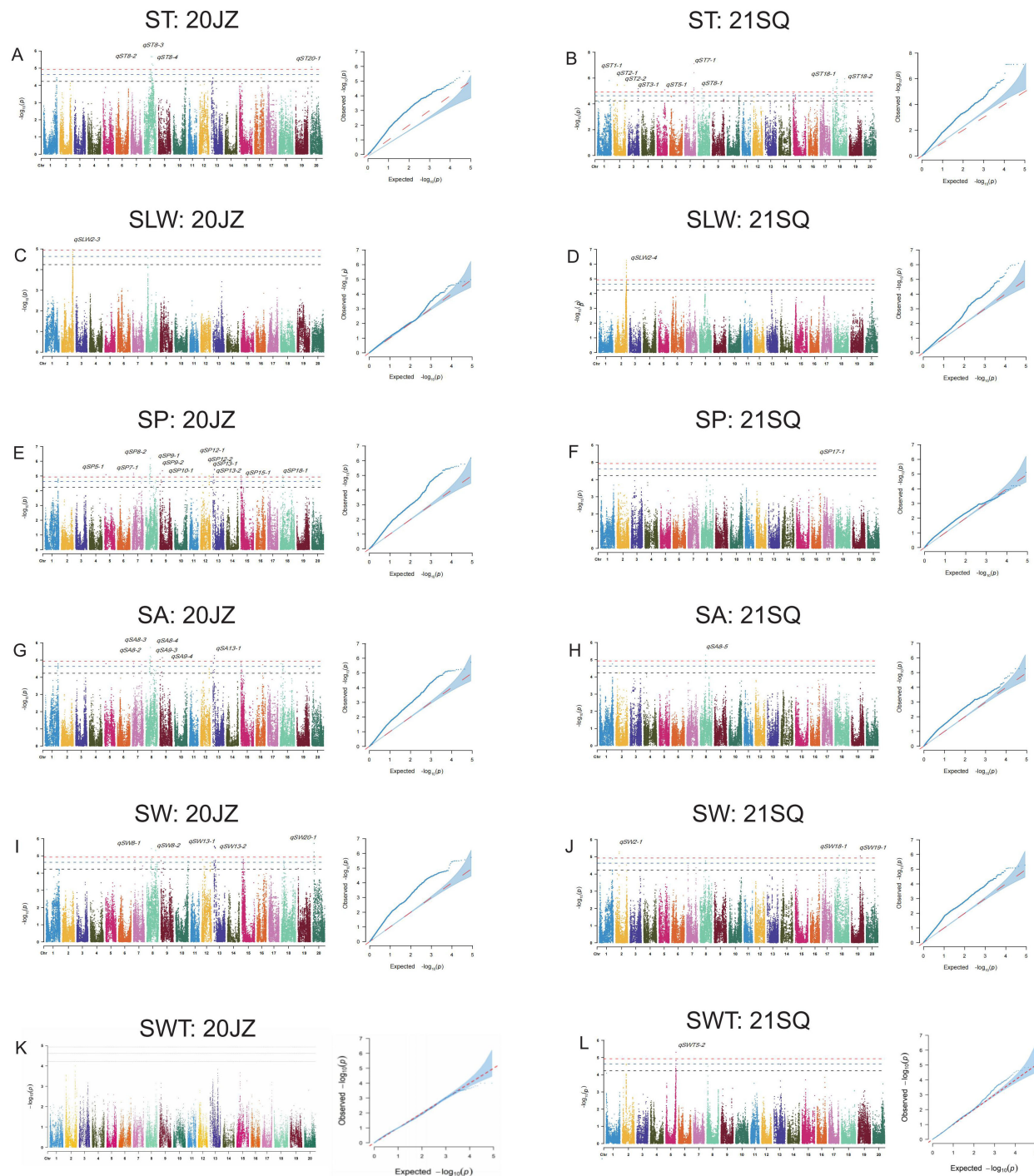
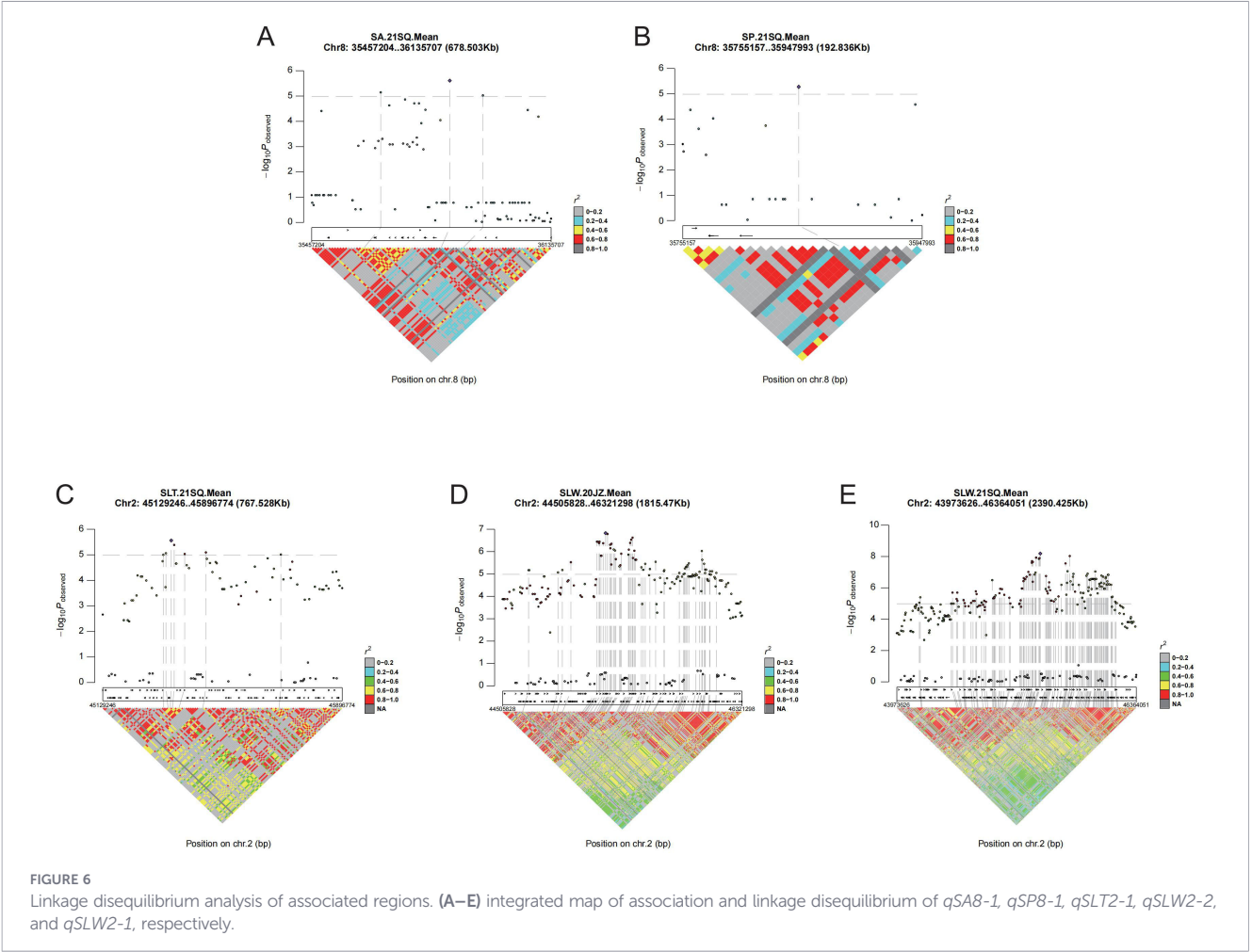


FIGURE 5

Manhattan plots and quantile-quantile (QQ) plots for (A, B) ST, seed thickness; (C, D) SLW, seed length to width; (E, F) SP, seed perimeter; (G, H) SA, seed area; (I, J) SW, seed width; (K, L) SWT, seed width to seed thickness using TASSEL under two environments (20JZ and 21SQ).

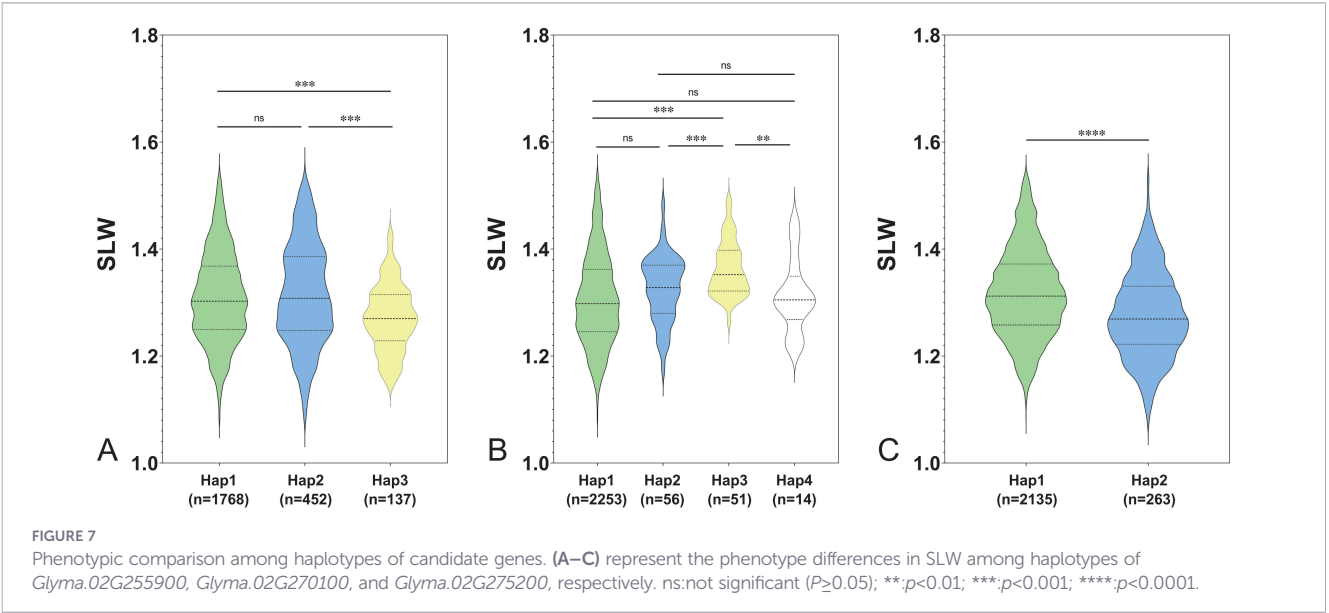
the strongest association with ST across three environments (18XY, 18AC, and 18HL), whereas it correlated most strongly with SW in the 17XY environment (Teng et al., 2019). Similarly, in a RIL population derived from Guizao1 and B13, the highest regression coefficient (R^2) was observed between HSW and SW in five environments, while the strongest relationship with ST was detected

in one environment (Luo et al., 2023). Another study using a RIL population from the cross “Zheng92116” × “Qihuang30” found that HSW showed the highest correlation with SW in two environments, with ST in one environment, and exhibited equal correlation strength with both SW and ST in an additional environment (Sun et al., 2021). It has been reported that soybean seed shape and size



have undergone domestication selection, and the importance of ST for hundred-seed weight progressively increases from wild soybeans to landraces and improved cultivars (Niu et al., 2012). Collectively, SW and ST may be the key phenotypic determinants influencing HSW in soybean.

Fifty-seven loci overlapped with previous studies, accounting for 58.16% of the total QTL, while 41 loci were novel (Appendix M). As the seed shape-related traits varied in different environments with broad-sense heritability over 0.60, genetic effect on the seed shape related traits is probably higher than environmental effects. A





number of regulatory genes controlling soybean seed shape have already been cloned, with major efforts focused on transcriptional regulation, phytohormone signaling and homeostasis, and metabolic pathways (Zhang et al., 2024). In this study, three seed-specific genes with significant phenotypic differences between haplotypes were selected as candidate genes regulating seed shape. *Glyma.02G255900* encodes a nucleolar histone methyltransferase-related protein and might influence the global methylation status of genes in seeds. *Glyma.02G270100* was annotated as a FAM91A1 protein. The homologue of which (*LTD1*) in rice located on the plasma membrane, endoplasmic reticulum, and multivesicular bodies, could interact with TFB2, and CycH1;1 and is involved in cell cycle regulation. Knockdown of *LTD1* resulted in reduced tiller number, grain number per plant, and as well as 1,000-grain weight (Yang et al., 2025b). *Glyma.02G270100* might be involved in HSW and seed shape regulation via a similar mechanism in soybean; however, further investigation is needed. *Glyma.02G275200* encodes a nodulin-like protein 20. Although nodulin-like proteins were presumed to function in the nodulation process, recent studies also suggest that early nodulin-like (ENODL) genes may play a vital role in the formation and development of underground storage organs in plants (Zhao et al., 2023). Seed shape is formed during seed development, which could be possibly regulated by early nodulin-like genes as well.

This study identified 59 unique QTLs related to seed shape-related traits, further confirming that seed morphology is a complex quantitative trait. Because the broad-sense heritability of seed shape-related traits was relatively high (0.67–0.90), genomic selection analysis could be performed. Using marker set M2 (trait-associated markers), the prediction accuracy for seed shape-related traits ranged from 0.43 to 0.48. Notably, the M1

marker set, which contained an additional 80,212 SNPs not associated with seed shape, increased prediction accuracy by 36.2%–85.7%. These findings are inconsistent with earlier research on soybean branching and stem termination traits (Wang et al., 2024; Yang et al., 2025a). However, the results of genomic selection (GS) and marker-assisted selection (MAS) for predicting soybean hundred-seed weight showed that the average prediction accuracy of GS was superior to that of MAS, which is consistent with the results of this study (Zhang et al., 2016). This discrepancy may be due to incomplete identification of seed shape-associated markers in the current study, where undiscovered markers and missing heritability might play an important role in shaping prediction outcomes.

5 Conclusion

This study conducted both phenotypic and genetic analyses of seed shape-related traits in a RIL population. The broad-sense heritability of different traits ranged from 0.67 to 0.90, with SW ranking the highest. The interrelationship analysis between HSW and seed shape-related traits revealed that ST and SW might be key traits contributing to HSW. A total of 98 QTL were mapped with 60 QTL co-localized and corresponding to 59 unique QTL regions. Three genes, *Glyma.02G255900*, *Glyma.02G270100*, and *Glyma.02G275200*, were identified as candidates for seed shape regulation, which are under investigation. Genomic prediction accuracy of seed shape-related traits ranged from 0.62 to 0.68, making the application of genomic selection feasible in the breeding program.

Data availability statement

The data presented in the study are deposited in the Soybean Functional Genomics and Breeding repository, direct access to data through link <https://sfgb.rmbreeding.cn/about/data>.

Author contributions

KL: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. HY: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. SJ: Formal analysis, Investigation, Software, Visualization, Writing – review & editing. MS: Data curation, Writing – review & editing. YX: Writing – review & editing, Investigation. YC: Data curation, Writing – review & editing. ZL: Data curation, Writing – review & editing. JL: Investigation, Writing – review & editing. BG-N: Data curation, Writing – review & editing. YZ: Methodology, Writing – review & editing. FZ: Methodology, Writing – review & editing. LQ: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. JW: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was funded by the Key Research Project of the Shennong Laboratory (SN01-2024-01) and the Major Agricultural Science and Technology Project of the Ministry of Agriculture and Rural Affairs (Grant No. NK202308010302).

References

- Alemu, A., Åstrand, J., Montesinos-López, O. A., Isidro y Sánchez, J., Fernández-González, J., Tadesse, W., et al. (2024). Genomic selection in plant breeding: Key factors shaping two decades of progress. *Mol. Plant Breed.* 17, 552–578. doi: 10.1016/j.molp.2024.03.007
- Bandillo, N. B., Jarquin, D., Posadas, L. G., Lorenz, A. J., and Graef, G. L. (2023). Genomic selection performs as effectively as phenotypic selection for increasing seed yield in soybean. *Plant Genome* 16, e20285. doi: 10.1002/tpg2.20285
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67, 1–48. doi: 10.18637/jss.v067.i01
- Cai, Z. D., Xian, P. Q., Cheng, Y. B., Zhong, Y. W., Yang, Y. Y., Zhou, Q. H., et al. (2023). MOTHER-OF-FT-AND-TFL1 regulates the seed oil and protein content in soybean. *New Phytol.* 239, 905–919. doi: 10.1111/nph.18792
- Chen, X. L., Liu, C., Guo, P. F., Hao, X. S., Pan, Y. P., Zhang, K., et al. (2023a). Differential SW16.1 allelic effects and genetic backgrounds contributed to increased seed weight after soybean domestication. *JIPB* 65, 1734–1752. doi: 10.1111/jipb.13480
- Chen, Y. J., Xiong, Y. J., Hong, H. L., Li, G., Gao, J., Guo, Q. Y., et al. (2023b). Genetic dissection of and genomic selection for seed weight, pod length, and pod width in soybean. *Crop J.* 11, 832–841. doi: 10.1016/j.cj.2022.11.006
- Cheng, Q., Rui, L., Deng, Y. Y., Xiao, E. L., Liu, B. Q., Yang, C. Y., et al. (2016). QTL mapping of soybean grain size and shape traits. *Acta Agron. Sin.* 42, 1309–1318. doi: 10.3724/SP.J.1006.2016.01309
- Cui, B. F., Chen, L., Yang, Y. Q., and Liao, H. (2020). Genetic analysis and map-based delimitation of a major locus qSS3 for seed size in soybean. *Plant Breed.* 139, 1145–1157. doi: 10.1111/pbr.12853
- De Mendiburu, F. (2019). Package ‘agricolae’. *R. Package* 1, 1143–1149. doi: 10.32614/CRAN.package.agricolae
- Doyle, J. J. T., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus San Franc Calif.* 12, 13–15.
- Duan, Z. B., Li, Q., Wang, H., He, X. M., and Zhang, M. (2023). Genetic regulatory networks of soybean seed size, oil and protein contents. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1160418
- Duan, Z. B., Zhang, M., Zhang, Z. F., Liang, S., Fan, L., Yang, X., et al. (2022). Natural allelic variation of GmST05 controlling seed size and quality in soybean. *Plant Biotechnol. J.* 20, 1807–1818. doi: 10.1111/pbi.13865
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Fang, C., Ma, Y. M., Wu, S. W., Liu, Z., Wang, Z., Yang, R., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 18, 161. doi: 10.1186/s13059-017-1289-9
- Fox, J. (2006). Teacher’s corner: Structural equation modeling with the sem package in R. *Struct. Equation Modeling: A. Multidiscip. J.* 13, 465–486. doi: 10.1207/s15328007sem1303_7

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.2026.1814388/full#supplementary-material>

- 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 Breed.* 138, 733–740. doi: 10.1111/pbr.12745
- Harrell, F. E. (2003). Hmisc: harrell miscellaneous. *Hmisc* 5, 1–3. doi: 10.32614/CRAN.package.Hmisc
- He, F. Y., Ding, S. C., Wang, H. W., and Qin, F. (2020a). IntAssoPlot: An R package for integrated visualization of genome-wide association study results with gene structure and linkage disequilibrium matrix. *Front. Genet.* 11. doi: 10.3389/fgene.2020.00260
- He, N. P., Li, Y., Liu, C. C., Xu, L., Li, M. X., Zhang, J. H., et al. (2020b). Plant trait networks: Improved resolution of the dimensionality of adaptation. *Trends Ecol. Evol.* 35, 908–918. doi: 10.1016/j.tree.2020.06.003
- Hina, A., Cao, Y. G., Song, S. Y., Li, S. G., 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.). *Trends Ecol. Evol.* 21, 1040. doi: 10.3390/ijms21031040
- Hu, D. Z., Li, X., Yang, Z. Y., Liu, S. L., Hao, D. R., Chao, M. N., et al. (2022). Downregulation of a gibberellin 3 β -hydroxylase enhances photosynthesis and increases seed yield in soybean. *New Phytol.* 235, 502–517. doi: 10.1111/nph.18153
- Hu, Y., Liu, Y., Tao, J. J., Lu, L., Jiang, Z. H., Wei, J. J., et al. (2023a). GmJAZ3 interacts with GmRR18a and GmMYC2a to regulate seed traits in soybean. *JIPB* 65, 1983–2000. doi: 10.1111/jipb.13494
- Hu, Y., Liu, Y., Wei, J. J., Zhang, W. K., Chen, S. Y., and Zhang, J. S. (2023b). Regulation of seed traits in soybean. *ABIOTECH* 4, 372–385. doi: 10.1007/s42994-023-00122-8
- Hu, Z. B., Zhang, H. R., Kan, G. Z., Ma, D. Y., Zhang, D., Shi, G. X., 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, 247–254. doi: 10.1007/s10709-013-9723-8
- Irnawati, I., Riswanto, F. D. O., Riyanto, S., Martono, S., and Rohman, A. (2021). The use of software packages of R factextra and FactoMineR and their application in principal component analysis for authentication of oils. Indonesian. *J. Chemom. Pharm. Anal.* 1, 1–10. doi: 10.22146/jcpa.482
- Khan, I. U., Javaid, N., Taylor, C. J., Gamage, K. A. A., and Ma, X. (2021). Big data analytics for electricity theft detection in smart grids. *IEEE Madrid PowerTech*, 1–6. doi: 10.1109/PowerTech46648.2021.9495000
- Kumar, R., Saini, M., Taku, M., Debbarma, P., Mahto, R. K., Ramlal, A., et al. (2023). Identification of quantitative trait loci (QTLs) and candidate genes for seed shape and 100-seed weight in soybean [*Glycine max* (L.) Merr.]. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1074245
- 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, 646102. doi: 10.3389/fgene.2021.646102
- Li, Y., Liu, C., Sack, L., Xu, L., Li, M., Zhang, J., et al. (2022). Leaf trait network architecture shifts with species-richness and climate across forests at continental scale. *Ecol. Lett.* 25, 1442–1457. doi: 10.1111/ele.14009
- Liang, H. Z., Li, W. D., Wang, H., and Fang, X. J. (2005). Genetic effects on seed traits in soybean. *Yi Chuan Xue Bao* 32, 1199–1204.
- Liang, H. Z., Wang, S. F., Yu, Y. L., Wang, T. F., and Gong, P. T. (2008). QTL mapping of six soybean grain shape traits. *J. Henan Agric. Sci.* 9, 45–51. doi: 10.15933/j.cnki.1004-3268.2008.09.018
- Liang, Q., Chen, L., Yang, X., Yang, H., Liu, S., Kou, K., et al. (2022). Natural variation of Dt2 determines branching in soybean. *Nat. Commun.* 13, 6429. doi: 10.1038/s41467-022-34153-4
- Liang, S., Duan, Z. B., He, X. M., Yang, X., Yuan, Y. Q., Liang, Q. J., et al. (2024). Natural variation in GmSW17 controls seed size in soybean. *Nat. Commun.* 15, 7417. doi: 10.1038/s41467-024-51798-5
- Liu, J. Y., Zhang, Y. W., Han, X., Zuo, J. F., Zhang, Z., Shang, H., et al. (2020). An evolutionary population structure model reveals pleiotropic effects of GmPDAT for traits related to seed size and oil content in soybean. *J. Exp. Bot.* 71, 6988–7002. doi: 10.1093/jxb/eraa426
- Liu, K., Wang, H., and Mu, Y. Y. (2025). Risk dispersion and optimization of soybean imports for China: Based on substitution and dependence risks. *China Oils Fats* 2, 1–7. doi: 10.19902/j.cnki.zgyz.1003-7969.230456
- Luo, S. L., Jia, J., Liu, R. Q., Wei, R. Q., Guo, Z. B., Cai, Z. D., et al. (2023). Identification of major QTLs for soybean seed size and seed weight traits using a RIL population in different environments. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1094112
- Ma, Y. S., Liu, Z. X., Wen, Z. X., Wei, S. H., Yang, C. M., Wang, H. C., et al. (2018). Effect of population structure on prediction accuracy of soybean 100-seed weight by genomic selection [J]. *Acta Agron. Sin.* 44, 43–52. doi: 10.3724/SP.J.1006.2018.00043
- Meng, L., Li, H. H., Zhang, L. Y., and Wang, J. K. (2015). QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819
- Moongkanna, J., Nakasathien, S., Novitzky, W. P., Kwanyuen, P., Sinchaisri, P., and Srivines, P. (2011). SSR markers linking to seed traits and total oil content in soybean. *Thai J. Agric. Sci.* 44, 233–241.
- Nandigam, S., Mahendrakar, M. D., Srungarapu, R., Chand, U., Gopalakrishnan, S., Thati, S., et al. (2025). Rapid generation advancement of RIL population and assessing the impact of Rhizobium nodulation on crop yields in Chickpea. *Sci. Rep.* 15, 13945. doi: 10.1038/s41598-025-98965-2
- Niu, Y., Xie, F. T., Bu, S. H., Xie, S. Q., and Han, S. F. (2013a). Fine mapping of QTL for soybean grain shape traits. *Acta Agron. Sin.* 39, 609–616. doi: 10.3724/SP.J.1006.2013.00609
- Niu, Y., Xu, Y., Li, G. J., Wang, Y. Q., Liu, X. F., Li, H. N., et al. (2012). Domestication of seed size and shape traits in soybean. *Soybean Sci.* 13, 522–528. doi: 10.14260/jemds/516
- Niu, Y., Xu, Y., Li, G. J., Wang, Y. Q., Liu, X. F., Li, H. N., et al. (2013b). Association mapping for seed size and shape traits in soybean cultivars. *Mol. Breed.* 31, 785–794. doi: 10.1007/s11032-012-9833-5
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Qin, J., Shi, A. N., Song, Q. J., Li, S., Wang, F. M., Cao, Y. H., et al. (2019). Genome wide association study and genomic selection of amino acid concentrations in soybean seeds. *Front. Plant Sci.* 10, 1145. doi: 10.3389/fpls.2019.01445
- Qin, P. X., Wang, T. R., and Luo, Y. C. (2022). A review on plant-based proteins from soybean: Health benefits and soy product development. *J. Agric. Food Res.* 7, 100265. doi: 10.1016/j.jafr.2021.100265
- Ravelombola, W., Qin, J., Shi, A. N., Song, Q. J., Yuan, J., Wang, F. M., et al. (2021). Genome-wide association study and genomic selection for yield and related traits in soybean. *PLoS One* 16, e0255761. doi: 10.1371/journal.pone.0255761
- Ren, X. L., Zhi, L. Y., Liu, L., Meng, D. Y., Su, Q. N., Batool, A., et al. (2021). Alternative splicing of TaGS3 differentially regulates grain weight and size in bread wheat. *Int. J. Mol. Sci.* 22, 11692. doi: 10.3390/ijms222111692
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Stat. Software* 48, 1–36. doi: 10.18637/jss.v048.i02
- 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, 1459–1466. doi: 10.1007/s00122-006-0392-1
- Song, J., Xiong, Y. J., Chen, Y. J., Xu, R. X., Liu, K. L., Guo, Q. Y., et al. (2024). Genetic analysis of seed coat and flower color based on a soybean nested association mapping population. *Acta Agronom Sin.* 50, 556–575. doi: 10.3724/SP.J.1006.2024.34094
- 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 Breed.* 140, 1058–1069. doi: 10.1111/pbr.12968
- Tang, K. Q., Yang, S. X., Feng, X. X., Wu, T., Leng, J. T., Zhou, H. K., et al. (2020). GmNAP1 is essential for trichome and leaf epidermal cell development in soybean. *Plant Mol. Biol.* 103, 609–621. doi: 10.1007/s11103-020-01013-y
- 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. Sci.* 156, 3–12. doi: 10.1017/S002185961700082X
- Teng, W. L., Wang, Q. J., Feng, W. J., Li, Y., Zhao, X., Han, Y. P., et al. (2019). QTL mapping of grain shape traits and 100-grain weight in soybean. *J. Northeast. Agric. Univ.* 50, 21–32. doi: 10.19720/j.cnki.issn.1005-9369.2019.12.003
- Van der Laan, L., Parmley, K., Saadati, M., Pacin, H. T., Panthulugiri, S., Sarkar, S., et al. (2025). Genomic and phenomic prediction for soybean seed yield, protein, and oil. *Plant Genome* 18, e70002. doi: 10.1002/tpg2.70002
- Vogel, J. T., Liu, W., Olhott, P., Crafts-Brandner, S. J., Pennycooke, J. C., and Christiansen, N. (2021). Soybean yield formation physiology – a foundation for precision breeding based improvement. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.719706
- Wang, J., Yang, Q. C., Chen, Y. J., Liu, K. L., Zhang, Z., Xiong, Y. J., et al. (2024). QTL mapping and genomic selection of stem and branch diameter in soybean (*Glycine max* L.). *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1388365
- Yang, X. R., Hu, C., Zhang, X. Y., Wang, X. L., Chen, L. F., Zhang, H. S., et al. (2025b). LTD1 plays a key role in rice tillering regulation through cooperation with CycH1;1 and TFB2 subunits of the TFIIB complex. *Plant J.* 122, e70119. doi: 10.1111/tpj.70119
- 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. Biol. Plants* 23, 877–889. doi: 10.1007/s12298-017-0468-1
- Yang, Q. C., Wang, J., Xiong, Y. J., Mao, A. L., Zhang, Z. Q., Chen, Y. J., et al. (2025a). Identification of QTL for branch traits in soybean (*Glycine max* L.) and its application in genomic selection. *Front. Genet.* 16. doi: 10.3389/fgene.2025.1484146
- Yu, H., Chen, Y. J., Liu, K. L., and Zhang, F. (2024). Development and application of a haplotype analysis R package (HaploAssistant). *Mol. Plant Breed.* 98, 1–11

- Yuan, X. B., Jiang, X. Y., Zhang, M. Z., Wang, L. F., Jiao, W., Chen, H. T., et al. (2024). Integrative omics analysis elucidates the genetic basis underlying seed weight and oil content in soybean. *Plant Cell* 36, 2160–2175. doi: 10.1093/plcell/koae062
- Zhang, Y., Bhat, J. A., Zhang, Y., and Yang, S. (2024). Understanding the molecular regulatory networks of seed size in soybean. *Int. J. Mol. Sci.* 25, 1441. doi: 10.3390/ijms25031441
- Zhang, M., Liu, S. L., Wang, Z., Yuan, Y. Q., Zhang, Z. F., Liang, Q. J., et al. (2022). Progress in soybean functional genomics over the past decade. *Plant Biotechnol. J.* 20, 256–282. doi: 10.1111/pbi.13682
- Zhang, J. P., Song, Q. J., Cregan, P. B., and Jiang, G. L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* 129, 117–130. doi: 10.1007/s00122-015-2614-x
- Zhang, Y., Yang, X. J., Bhat, J. A., Zhang, Y. H., Bu, M. R., Zhao, B. F., et al. (2025). Identification of superior haplotypes and candidate gene for seed size-related traits in soybean (*Glycine max* L.). *Mol. Breed.* 45, 3. doi: 10.1007/s11032-024-01525-1
- Zhao, G., Wu, S. M., Zhang, Y., Huang, S. Q., Cai, H., and Wen, J. (2023). Molecular cloning and expression analysis of ThENODL1 and ThENODL2 of *Tetrastigma hemsleyanum* Diels et Gilg. *Life. Sci. Res.* 27, 23–30. doi: 10.16605/j.cnki.1007-7847.2022.07.0167

Appendix A

The ANOVA analysis of seed shape related traits and the proportion of variance explained by genotype.

Appendix B

PTN analysis of seed shape related traits and HSW.

Appendix C

Path analysis of seed shape related traits and HSW.

Appendix D

Seed traits related QTL mapped by ICIM in two different environments.

Appendix E

Seed traits related QTL mapped by EMMAX in two different environments.

Appendix F

Seed shape related QTL identified by TASSEL.

Appendix G

The overlapping QTL detected by either multiple mapping methods or for different traits.

Appendix H

SNP variation related to genes in *qSLW2-1* and *qSA8-1*.

Appendix I

Expression profile of genes within candidate region among different tissues.

Appendix J

Haplotype sequence of genes within *qSLW2-1* and *qSA8-1*.

Appendix K

Statistic significance between haplotypes of genes within *qSLW2-1* and *qSA8-1* by Wilcoxon.

Appendix L

GS prediction accuracy of soybean seed shape related traits based on different marker sets using RRBLUP.

Appendix M

Seed trait related QTL identified overlapped with references.