Indian J. Genet. Plant Breed., (2025); 85(4 Suppl.) : 745-753
https://doi.org/10.31742/ISGPB.85.suppl.7

ISSN: 0975-6906

www.isgpb.org

RESEARCH ARTICLE

Identification, mapping of the genomic regions and mining of the candidate genes for Hundred Seed Weight (HSW) in soybean [*Glycine max* (L.) Merr.]

Manisha Saini, Akshay Talukdar*, R. Ambika Rajendran and S.K. Lal

Abstract

Hundred seed weight (HSW) in soybean is an important trait determining yield and quality and is also desirable for specialty soy foods like tofu, natto, miso, and edamame. In order to identify the quantitative trait loci (QTLs) and mining of the candidate genes for HSW in soybean (*Glycine max*.), the current study was conducted on bold seeded soybean genotype EC1023 and VLS61, a medium seeded soybean genotypes derived F_2 and $F_{2:3}$ mapping populations. Genetic polymorphism studied with 517 SSR markers indicated the significant polymorphism between the parental genotypes to be 19.95% (103 markers); however, distribution of the polymorphism was not uniform across the chromosomes; Chr. 14 had 30.00 % polymorphism as against 7.14 % on Chrs.12. Through inclusive composite interval mapping approach, two major quantitative trait loci (QTLs) viz., *qHSW-5.1* (PVE=10.70%) and *qHSW-17.1* (PVE=19.86%) were mapped on Chr. 5 and Chr.17, respectively. The mapped QTLs were validated on interspecific RILs with varying level of HSW. QTL (*qHSW-17.1*) on Chr. 17 in the marker region Satt301-Sat_326 overlapped with the previously reported QTL for seed viability in soybean. Based on Protein Analysis Through Evolutionary Relationships (PANTHER), gene annotation information, andliterature search, 64 genes within two QTLs were predicted to be possible candidate genes that might regulate the HSW in soybean. The current study identified the two major QTLs and key candidate genes which govern the HSW in soybean, paving the way for developing soybean varieties with improved HSW through marker-assisted breeding.

Keywords: Soybean, hundred seed weight, QTLs, candidate genes, gene mining, marker assisted breeding

Introduction

Glycine max (L.) Merrill or soybean (2n = 40), is the most important oilseed crop in the world, contributing almost 57% of global oilseed production. Valued for its high oil (18–22%) and protein (40–45%) content, soybean is a nutrient-dense crop, offering an excellent source of plant-based protein, essential fatty acids, carbohydrates, antioxidants, and other vital nutrients for human health. Owing to its multifaceted use in food, feed, and industry, soybean has been aptly named the "miracle bean or golden bean" (Orf, 2010). A wide range of soy-based products such as full-fat soy flour, soymilk, tofu, natto, soy cheese, curd, sprouted snacks, and fortified baked goods are produced using soybeans (Kumar, 2005), with seed morphological traits particularly seed weight, size, and shape playing a critical role in determining their specific utility (Cui et al. 2004; Yan et al. 2017). For instance, in specialty soy food products like tofu, miso, and edamame (Cui et al. 2004) seed weight, size and shape are important factors and large seeds are preferred, whereas small seeds are best for making natto (Wilson 1995; Yan et al. 2017). Round seeds are preferred over other types for food-type soybeans. Thus, seed size and shape largely determine the seed weight of the seed and in soybean seed weight often expressed as Hundred Seed Weight (HSW) thus, HSW is considered a key agronomic parameter, significantly influencing both seed yield and product quality of the soybean (Burton 1987; Salas et al. 2006, Xu et al. 2011and Yan et al. 2017).

Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India

***Corresponding Author:** Akshay Talukdar, Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India, E-Mail:akshay.talukdar1@gmail.com

How to cite this article: Saini M., Talukdar A., Rajendran R.A. and Lal S.K. 2025. Identification, mapping of the genomic regions and mining of the candidate genes for Hundred Seed Weight (HSW) in soybean [*Glycine max* (L.) Merr.]. Indian J. Genet. Plant Breed., **85**(4 Suppl.): 745-753.

Source of support: Nil

Conflict of interest: None.

Received: July 2025 **Revised:** Oct. 2025 **Accepted:** Nov. 2025

HSW is a complex quantitative character regulated by several genes and significantly impacted by the growing environment (Brim and Cookham 1961). Tao et al. (2017) reported that seed size, which is expressed as HSW possess a high correlation with yield (Xu et al. 2011) and one of the fitness characteristic crucial for environmental adaptation and adaptability. Thus, improvement of HSW has become a major objective in soybean breeding programs. The phenotypic selection of the offspring from crosses between desired contrasting experimental lines has been the foundation of a conventional crop improvement (Lande and Thompson 1990; Dekkers and Hospital 2002). However, phenotypic selection of soybeans using conventional methods for HSW is challenging due to the quantitative nature of the trait (Main et al. 1996; Maughan et al. 1996; Li et al. 2008). Natural selection for larger seeds in soybeans has resulted in the accumulation of minor QTLs; these evolutionary changes are evident from Quantitative trait loci (QTL mapping) studies (Salas et al. 2006; Hina et al. 2020). QTL analysis is a useful technique for deciphering complex trait architecture. Thus, understanding of the genetic basis of HSW and finding the QTLs linked to HSW, exploring useful alleles linked to the QTLs and mining of the candidate genes have significant theoretical and practical implications for enhancing soybean yield and quality.

In the last few decades, numerous QTLs for characteristics pertaining to soybean seed weight and size have been discovered. The USDA Soybean Genome Database (SoyBase; http://www.soybase.org) currently contains more than 400 QTLs for seed size and shape; however, the majority of these QTLs are unconfirmed, meaning they have not been mapped across multiple environments or validated on mapping populations with different genetic backgrounds (Hina et al. 2020). Mian et al. (1996) discovered 16 QTLs for seed size and shape on 12 distinct soybean chromosomes. In a cross between two early-maturing soybean cultivars, Csanádi et al. (2001) discovered QTLs affecting seed weight. Hoeck et al. (2003) discovered 27 QTLs associated with seed size on 16 soybean chromosomes, whereas Li et al. (2008) discovered three QTLs for seed length (SL) on Chr. 7, Chr. 13, and Chr. 16. Three epistatic-effect QTLs and 19 main-effect QTLs for SL were discovered on eight chromosomes by Lü et al. (2011). Using two recombinant inbred line (RIL) populations, Kato et al. (2014) discovered a significant and stable QTL for soybean seed weight that accounts for 9.4–20.9% of phenotypic variation. They also showed that the markers linked to these QTLs were consistent in all environments and had effects significant enough to be appropriate for a marker-assisted breeding (MAB) program. Liu et al. (2013) identified 18 seed-weight QTLs using a RIL population evaluated in six different environments; five of these were found to be stable across all environments and not involved in epistatic interactions. Xie et al. (2014) finely mapped QTLs for soybean seed size traits on Chr.6 in the RIL population derived from cross of Li-shui-zhong-zi-huang and Nannong493-1. Later, Xin et al. (2016) identified 12 main-effect QTLs underlying 100-seed weight using 194 chromosome-segment substitution lines that covered 82.5% of the wild soybean's (*Glycine soja*) genome. Hina et al. (2020) identified 88 main and epistatic-effect QTLs for six soybean seed size and shape traits. Similarly, Li et al. (2020) mapped 42 QTLs with additive effects for seed traits. In a diverse genetic background of vegetable type and seed type soybean, Kumar et al. (2023) mapped 5 stable QTLs viz; *qSL-10-1*, *qSW-4-1*, *qSV-4-1*, *qSLW-10-1* and *qSLH-10-1* in $F_{2:3}$ mapping population.

Although a number of QTL mapping studies have revealed a multiple QTLs information concerning soybean HSW, molecular genetic basis of HSW is still unknown. Finding and validating QTLs under a range of circumstances and backgrounds is therefore crucial for the efficient use of QTLs in MAB. In order to identify the most important genomic regions and putative genes for soybean HSW, the current study was conducted using an F_2 and $F_{2:3}$ mapping population derived from a cross between EC1023, a bold seeded soybean genotype and VLS61, a medium size seed type of soybean. The results are expected to help develop soybean varieties with higher yield and quality through MAB for HSW.

Materials and methods

Plant materials

The experiment was conducted on 125 F_2 derived F_3 ($F_{2:3}$) intra-specific (*Glycine max*.) population derived from the hybridization of a bold yellow seeded high seed viable soybean genotype EC1023 (having 9.70 g of HSW) with medium size yellow seeded poor viable genotype VLS61 (5.90 g of HSW). The F_1 and F_2 seeds were grown under controlled conditions of the National Phytotron Facility (NPF), IARI, New Delhi and harvested separately and used in the inheritance and mapping study.

For studying the inheritance of HSW, seeds of soybean parental genotypes were obtained from the Soybean Laboratory, Division of Genetics, ICAR-IARI, New Delhi. Fresh seeds of parental soybean genotype were selected for subsequent experimentations. All post-harvest operations including threshing and cleaning were done manually to maintain purity of the seeds and avoiding mechanical injuries. Uniform moisture content was ensured in all the seed samples prior to the experiment.

Molecular analysis was done in 119 $F_{2:3}$ plants, whereas, phenotypic data could be collected from 49 plants as several plants died during the unavoidable experimental crises.

For validation of the findings of the present study, an interspecific RILs population was used, developed by the crossing small seeded *Glycine soja* accession DC2008-1(0.96g HSW) with a bold seeded cultivated variety DS9712 (9.12g HSW) (Yashpal et al. 2015). Out of 300 RILs, a set of 40 RILs

(F_8 generation) were used in the present study and seeds of the RILs were obtained from the Soybean Breeding Lab, Genetics Division, IARI, New Delhi.

Phenotyping of the HSW

One hundred (100) seeds from each genotype were collected randomly and weighted in gram. The process was repeated 3 times and the average was worked out. On the basis of average weight (g), the seeds were categorized as large (> 10 g), medium (5.0-10.0 g) and small (< 5 g) as per Anonymous (2009).

Development of segregating progenies

For developing segregating progenies, both the parental genotype EC1023 and VLS61, were sown on staggered dates for synchronized flowering in the National Phytotron Facility (NPF), IARI, New Delhi. Upon flowering (30-45 days after sowing), for effective hybridization, a novel technique i.e. pollination without emasculation as given by Talukdar and Shivakumar (2012) was utilized. The F_1 seeds so produced were grown under controlled conditions and hybridity was tested with polymorphic SSR markers satt523 (Fig. 1). The F_2 seeds of the selected F_1 plants were grown in the NPF and the $F_{2:3}$ seeds were harvested separately. The 125 $F_{2:3}$ seeds from the cross EC1023 ×VLS61were obtained and used for further analysis.

DNA extraction and molecular genotypingof the parents with SSR marker

Genomic DNA was isolated from tender soybean leaves using modified CTAB procedure (Lodhi et al., 1994). Quality and quantity of the DNA extracted from the mapping population was ascertained through spectrophotometer analysis. The DNA samples were diluted to a concentration of 20ng/µl. Based on the consensus soybean genetic linkage map published by Cregan et al. (1999) and Song et al. (2004), SSR markers scattered throughout the 20 genetic linkage groups were chosen. A set of 517 SSR markers were used for the molecular genotyping of the parental genome, out of which 103 found to be polymorphic and were used for the molecular genotyping of the F_2 population. Genomic DNA of 119 F_2 plants were amplified by PCR and size separated in 3% metaphor gel through gel electrophoresis.

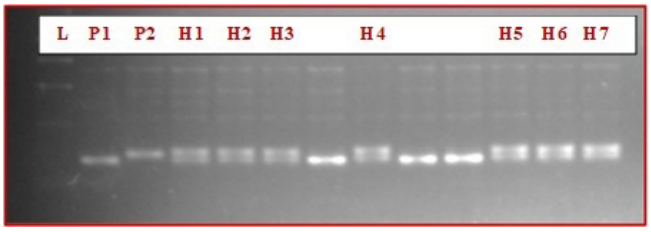

Fig. 1. Hybridity testing of the F_1 plants by using polymorphic SSR marker Satt523 (L: ladder, P1: Parent 1, P2: Parent 2, H1, H2, H3…H7: Hybrid 1, 2, 3….7.)

Linkage map construction and QTL mapping

For linkage map construction and to map the QTL for seed viability, software QTL IciMapping V4.2 was used. A genetic distance of 50 cM and a minimum LOD score of 3 was used to construct the linkage map connecting the markers. Kosambi's mapping function (Kosambi, 1944) was used to calculate map distances. Method for QTL analysis was Inclusive Composite Interval Mapping of ADDitive (and dominant) QTL (ICIM-ADD) (Zeng, 1994). The phenotypic data i.e. HSW of $F_{2:3}$ progenies of the 49 $F_{2:3}$ plants and the molecular genotypic data point of 97 SSR markers (out of 103 polymorphic markers on F_2 population, 6 showed segregation distortion and hence discarded) were used to map QTL for HSW. A LOD score of 3 was maintained to confirm the presence of a QTL in a particular genomic region. The threshold levels for each trait for ICIM-ADD mapping was computed by conducting a permutation test with 1000 permutations at 0.05 type -I error.

Allele mining and identification of candidate gene for seed viability

The QTLs identified and validated in this study were considered as a stable QTL. Model genes were downloaded from SoyBase (http://www.soybase.org) and EnsemblPlants (https://plants.ensembl.org) at the genomic location of the stable QTLs on the soybean genome (Glyma2.0). Gene Ontology (GO) enrichment analysis was performed using Phytozome 13 (http://phytozome-next.jgi.doe.gov) for all the genes in each QTL region. The predicted candidate genes were then subjected to Protein Analysis Through Evolutionary Relationships (PANTHER) Classification System in order to permit high-throughput analysis according to family and sub-family, molecular function, biological activity, and pathway.

Results

Phenotypic characterization of parents and $F_{2:3}$ population for HSW in soybean

HSW of bold seeded soybean parental genotype EC1023and medium seeded parental soybean genotype VLS61 were 9.70g and 5.90g, respectively, showing the significant phenotypic variation for HSW among both the parental genotype. The HSW of the seeds of $F_{2:3}$ progenies were varied from 1.57 to 13.07g with a mean of 6.49g (Table 1.). Based on HSW, the $F_{2:3}$ progenies were classified as small seeded (<5g), medium size (5-10g) and large seeded (> 10g). Out of the 49 progenies, 17 had small seeds, 23 had intermediate and 9 had large seeds respectively. The range of HSW in the $F_{2:3}$ progenies surpassed the range of HSW of the parental genotypes i.e. 5.90 g to 9.70 g, and it showed the continuous distribution from low to high (Fig. 2). The frequency distribution of HSW indicates the involvement of more than one gene or QTL in controlling the HSW trait in

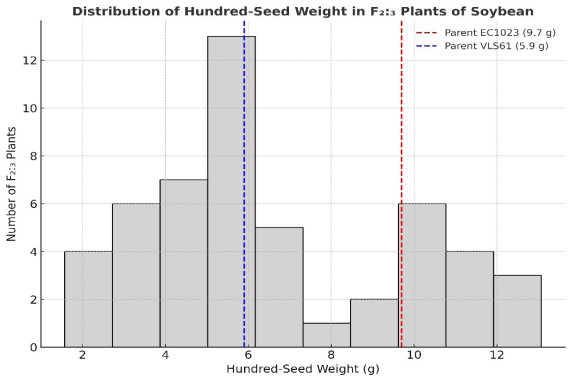

Fig. 2. Frequency distribution of HSW in the seedlings of F2:3 population

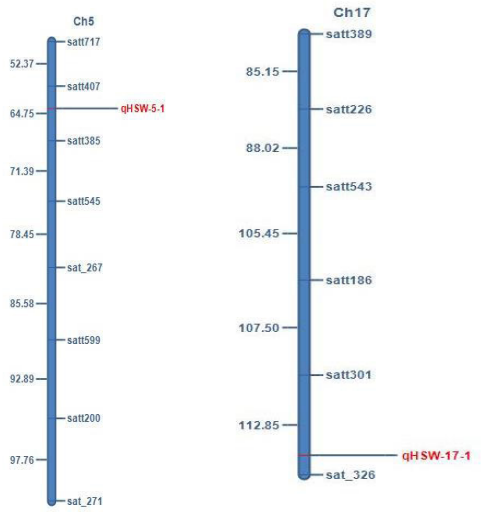

Fig. 3. Linkage map of chromosome 5 and 17 showing QTLs for hundred seed weight with their map position

Table 1. List of hundred seed weight (HSW) recorded from the progenies of F_2 plants

S. No.	F_2 Plant No.	Hundred seed weight of $F_{2:3}$ (g)
1	C6 P-3	4.61
2	C6 P-5	4.00
3	C6 P-6	5.60
4	C6 P-7	5.26
5	C6 P-8	4.14
6	C6 P-9	3.69
7	C6 P-19	12.02
8	C6 P-21	2.00
9	C13 P-1	4.25
10	C13 P-2	5.30
11	C13 P-3	6.40
12	C13 P-4	6.76
13	C13 P-7	10.77
14	C13 P-8	5.41
15	C13 P-10	3.55
16	C13 P-11	4.82
17	C13 P-12	9.57
18	C13 P-13	11.25
19	C13 P-41	11.05
20	C13 P-42	12.00
21	C13 P-43	9.16
22	C10 P-1	5.17
23	C10 P-2	6.12
24	C10 P-3	5.27
25	C10 P-4	3.16
26	C10 P-5	2.66
27	C10 P-6	1.57
28	C10 P-7	5.08
29	C10 P-8	5.36
30	C10 P-9	3.07
31	C10 P-10	6.16
32	C10 P-11	9.79
33	C10 P-12	11.54
34	C10 P-13	6.00
35	C10 P-14	3.44
36	C10 P-15	9.63
37	C10 P-16	2.00
38	C10 P-20	8.16
39	C10 P-21	6.55
40	C10 P-22	6.35
41	C10 P-24	7.16
42	C10 P-25	10.66
43	C10 P-26	9.65
44	C10 P-27	4.75
45	C10 P-28	10.30
46	C10 P-35	13.07
47	C10 P-36	4.20
48	C10 P-37	3.65
49	C10 P-42	5.83
50	EC1023	9.70
51	VL5-61	5.90

soybean. The presence of a greater number of phenotypic classes and appearance of the transgressive segregants in the $F_{2:3}$ population also confirms the involvement of a large number of genes and their recombination in the expression of the phenotype.

Parental polymorphism survey

Genomic diversity analysis of the parental genotypes was carried out using 517 SSR markers (approximately 25 markers per chromosome), which were selected in such a way so that to ensure uniform coverage across the entire soybean genome. Among these, 103 markers exhibited polymorphism, corresponding to a polymorphism rate of 19.92%. This relatively low level of polymorphism is expected, as both parents belong to the same species, *Glycine max*. The distribution of polymorphic SSR loci across the genome was found to be non-uniform, with certain chromosomes showing higher polymorphism than others. The highest proportion of polymorphic loci (30.00%) was observed on chromosome 14, whereas the lowest (7.14%) was recorded on chromosomes 12 (Table 2).

Marker segregation analysis, linkage map construction and QTL mapping of the HSW in soybean

As 103 polymorphic SSR markers were used for the molecular genotyping of the F_2 population, the segregation data of each marker was subjected to chi-square (χ^2) test for goodness of fit to 1:2:1 ratio. Out of 103 polymorphic markers used, 97 markers showed goodness of fit to the expected 1:2:1 ratio, while 6 markers showed segregation distortion at significance level of $P< 0.05$ and hence these were excluded from further analysis. The phenotypic data i.e. HSW of $F_{2:3}$ progenies of the 49 F_2 plants and the molecular genotypic data point of 97 SSR markers on F_2 plants were used to map QTL for HSW.

The marker trait analysis mapped two QTL for 100-seed weight viz., *qHSW-5.1* and *qHSW-17.1* were mapped on Chr.5 and Chr.17, respectively. Both the QTL were major QTL and seems as a novel QTLs for HSW (PVE ranged from 10.70 to 19.86 %). Linkage map of the QTL showing map position has been depicted in Fig 3. QTL *qHSW-17.1* showed negative additive effects (-1.60) indicating that these alleles from were contributed by the reduced seed weight, while the other

Table 2. Chromosome-wise distribution of SSR markers used and their level of polymorphism in cross combination EC1023 xVLS61

Chr. No.	Linkage Group	Total SSR used (No.)	Polymorphic SSR (No.)	Monomorphic SSR (No.)	Polymorphism Level (%)
1	D1a	25	6	19	24.00
2	D1b	34	4	30	11.76
3	N	22	3	19	13.63
4	C1	22	3	19	13.63
5	A1	33	8	25	24.24
6	C2	27	6	21	22.22
7	M	28	7	21	25.00
8	A2	26	3	23	11.53
9	K	24	6	18	25.00
10	O	27	5	22	18..51
11	B1	24	5	19	20.83
12	H	28	2	26	07.14
13	F2	33	9	24	27.27
14	B2	20	6	14	30.00
15	E2	17	4	13	23.52
16	J	28	5	23	17.85
17	D2	30	6	24	20.00
18	G	25	6	19	24.00
19	L	21	5	16	23.80
20	I	23	4	19	17.39
Total		517	103	414	19.92

Table 3. List of QTLs detected for hundred seed weight in soybean

S. No.	Chromosome Number	QTL	Map Position (cM)	Marker Interval (Lt marker-Rt marker)	LOD	PVE%	Additive Effect
1	17	qHSW-17.1	556.20	Satt301-Sat_326	4.23	19.86	-1.60
2	5	qHSW-5.1	130.95	Satt407-Satt385	2.04	10.70	1.10

LOD: LOD score, PVE (%): Phenotypic variation explained by QTL, Add: Estimated additive effect of the QTL.

QTL *qHSW-5.1* showing the positive additive effect (1.10) had contributed toward enhanced the seed weight. Map position of the identified QTL, markers bracketing them, phenotypic variance explained by the QTL, LOD and additive effect of the QTL are presented in Table 3.

Validation of the novel reported markers in RILs population

Out of the 4 SSR markers flanking to the QTLs mapped in present study for HSW, only two markers viz., Satt301 and Sat_326, both mapped on chromosome 17 overlapped with the same genomic region bracketed by the markers governing the seed viability QTLs (*qSv-17.1*).

Gene ontology and candidate gene prediction within the identified QTLs

Based on themapping results, identified QTLs viz; qHSW-5.1 and qHSW-17.1 studied for gene ontology (GO) andcandidate gene prediction analysis. Within the physical genomic interval of qHSW-5.1 and qHSW-17.1, 64 model genes were present, downloaded from SoyBase((http://www.soybase.org) and Ensembl Plants (https://plants.ensembl.org), respectively. Phytozome 13 was used to annotate all of the genes found in each QTL region. Each ofthe QTLs had a higher proportion of genes associatedwith the cell part, cell organelle, catalytic activity, binding, metabolic process, and cellular process which demonstrates the importance ofthese processes in the development of soybean seeds. Further, PANTHER analysis, gene annotation data, and literature search revealed the potential candidate genes underlying aforementioned QTLs accountable for HSW in soybean.

Discussion

Hundred seed weight (HSW) in soybean is the economically important traits that significantly influence soybean production, market value, and end-use quality (Gandhi 2009). It not only determines the consumer preference and processing suitability but also directly affects yield potential and seed composition. Therefore, the development of the soybean varieties with desirable seed size is one of the primary objectives of modern soybean breeding programs. In this context, the present study aimed to identify, map the QTLs and mine candidate genes governing HSW in soybean.

HSW is the critical economic trait that has a direct impact on soybean yield and market value. Consequently, breeders have consistently focused on developing cultivars with improved HSW with desirable seed shapes to enhance productivity and quality. However, similar to yield, HSW is quantitative in nature and polygenic in nature regulated by multiple genes with small additive effects and are strongly influenced by environmental conditions and genotype × environment interactions making their genetic improvement through conventional breeding approaches challenging (Hina et al. 2020). Achieving this objective requires a comprehensive understanding of the genetic architecture, inheritance pattern, and molecular mechanisms governing HSW in soybean. For this purposein the present study, F_2 and $F_{2:3}$ mapping populations were generated from a cross between EC1023, characterized by bold seeded soybean genotype, with VLS61, a medium-sized seeds soybean genotype. Both the parents differed genetically by 19.90% (103 markers) in HSW, ensuring adequate genetic variability for QTL mapping. Whenever parents were more distinctly related to each other high is the genetic diversity for a character in a population. Similarly, 21.07% of variability was observed among the vegetable and seed type soybean by Kumar et al. (2023). Kulkarni et al. (2017) also reported significant variation for HSW among the Williams 82 and PI 366121. The wide phenotypic variation and normal distribution of HSW observed in the segregating generations support its polygenic inheritance and quantitative nature, as also reported earlier (Liang et al. 2016: Khosla et al. 2020: Kumar et al. (2023). The appearance of transgressive segregants exhibiting seed weight beyond both parents further suggests recombination of favourable alleles from both genotypes, a phenomenon frequently observed in soybean (Li et al. 2008; Zhang et al. 2010).

Numerous QTLs associated with HSW and related parameters viz., size, and shape have been reported over the past few decades, whereas most of have not been effectively utilized in breeding programs due to their instability or lack of validation across diverse genetic backgrounds. For this a comprehensive phenotypic evaluation coupled with molecular genotyping was performed to detect significant QTLs and to identify potential candidate genes controlling HSW. The significant difference in HSW between the parents contributed to the identification of new genomic regions associated with HSW. Two major QTLs, *qHSW-5.1* and *qHSW-17.1*, were found in the present study to play significant roles in regulating seed weight. Similarly, Mian et al. (1996), Hoeck et al. (2003), Hina et al. (2020) and Kumar et al. (2023) reported 16, 27, 88 and 5 QTLs, respectively, for HSW in

soybean. Kato et al. (2014) discovered a significant and stable QTL for soybean seed weight possessing 9.4–20.9% of phenotypic variation. The positive and negative values of the QTLs (additive value) were used as the criteria for selecting the superior alleles. If QTL additive effect value is positive, contributes toward the enhancement of the HSW whereas if QTL value is negative reduced the superior alleles and reduced the HSW. Qi et al. (2020) identified several superior alleles present in the larger seeded lines and believe that superior alleles may have an important effect on the HSW of soybeans. Over the past few decades, several QTLs associated with HSW viz; seed size, seed length, seed width and seed thicknesshave been reported in soybean using various mapping populations and molecular markers (Csanádi et al. 2001; Zhang et al. 2004; Niu et al. 2013; Kato et al. 2014; Xie et al. 2014 and Wu et al. 2018). However, most of these QTLs have shown inconsistent expression across different genetic backgrounds and environments due to factors such as small population size, use of low-density genetic linkage maps, and lack of validation through fine mapping or candidate gene analysis. Consequently, the findings of the present study are expected to contribute to a better understanding of the genetic basis of the HSW and facilitate the development of soybean cultivars with improved seed quality through molecular breeding approaches.

Interestingly, qHSW-17.1 was co-localized with previously reported QTLs for seed viability (*qSv-17.1*), oil, and protein content (Csanadi et al. 2001; Junyi et al. 2007; Qi et al. 2011; Priolli et al. 2015; Saini et al. 2025), suggesting possible linkage or pleiotropic effects between loci governing seed morphology and composition. Such associations indicate that certain genomic regions may simultaneously regulate seed size, shape, oil, and protein content (Hina et al. 2020), providing valuable opportunities for simultaneous improvement of yield and quality traits.

To explore further, the functional basis of these QTLs, genes located within their physical intervals were analyzed. Out of several model genes extracted from the genomic regions underlying *qHSW-5.1* and *qHSW-17.1*, a total of 64 potential candidate genes were identified through PANTHER classification, gene annotation, and literature mining (Karikari et al. 2019). These genes are mainly associated with key biological processes such as catalytic activity, transport, metabolism, and signal transduction, all of which are critical for seed growth and development (Fan et al. 2006; Li and Li 2014; Kumar et al. 2023). The results of this investigation provide deeper insights into the genetic control of HSW in soybean. The identified QTLs and associated candidate genes serve as valuable genetic resources for developing molecular markers and conducting fine-mapping or functional validation studies. The QTLs identified herein, especially *qHSW-5.1* and *qHSW-17.1*, warrant further characterization to develop tightly linked markers for use in marker-assisted and genomic selection programs.Thus, this study provides the major and stable QTLs and candidate genes regulating HSW in soybean, and these findings will be of great use for MAB of soybean varieties with improved HSW.

Authors' contribution

Conceptualization of research (AT, MS); Designing of the experiments (AT, MS); Contribution of experimental materials (MS, AT, PAR, SKL); Execution of field/lab experiments and data collection ( ); Analysis of data and interpretation (MS, AT); Preparation of the manuscript (MS, AT).

References

Anonymous (2009). Guidelines for the conduct of test for distinctiveness, uniformity and stability (DUS) on Soybean [*Glycine max* (L.) Merrill]. Plant Var J. of India, **3**(10): 289-98.

Brim C.A. and Cockerham C.C. 1961. Inheritance of quantitative characters in soybean. Crop Sci., **1**: 187–190. doi:10.2135/cropsci1961.0011183X000100030009x.

Burton J.W. 1987. Quantitative genetics: results relevant to soybean breeding. In 'Soybeans: improvement, production and uses'. 2nd edn. Agronomy Monograph 16. (Ed. JR Wilcox) pp. 211–247. (ASA, CSSA, SSSA:Madison, WI, USA).

Cregan P. B., Mudge J., Fickus E. W., Danesh D., Denny R. and Young N. D. 1999. Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor. Appl. Genet., **99**: 811–818.

Csanádi G. Y., 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): 912–919.

Cui Z., James A. T., Miyazaki S., Wilson R. F. and Carter T. 2004. Breeding specialty soybeans for traditional and new soyfoods (Champaign, IL, USA: AOSC Press), 274–332.

Dekkers J.C. and Hospital F. 2002. The use of molecular genetics in the improvement of agricultural populations. Nature Reviews Genet., **3**: 22–32. doi: 10.3390/ijms21031040

Fan C., Xing Y., Mao H., Lu T., Han B., Xu C., et al. 2006. GS3, a major QTLfor 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

Gandhi A. P. 2009. Quality of soybean and its food products. Int. Food Res. J., **16**(1): 11–19.

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 forseed size and shape in soybean (Glycine max L.). Int. J. Mol. Sci., **21**(3): 1040.

Hoeck J. A., Fehr W. R., Shoemaker R. C., Welke G. A., Johnson S. L. and Cianzio S. R. 2003. Molecular marker analysis of seed size in soybean. Crop Sci., **43**(1): 68–74. doi: 10.2135/cropsci2003.0068.

Junyi G., Wang Y., Wu X. and Chen S. 2007. A comparative study onsegregationanalysis and QTL mapping of quantitative traits in plants—with a casein soybean. Front. Agri. China, **1**(1): 1–7.

Karikari B., Chen S., Xiao Y., Chang F., Zhou Y., Kong J., et al. 2019. Utilization of interspecific high-density genetic map of RIL

population for the QTLdetection and candidate gene mining for 100-seed weight in soybean. Front. Plant Sci., **10**: 1001. doi:10.3389/fpls.2019.01001

Kato S., Sayama T., Fujii K., Yumoto S., Kono Y., Hwang T. Y., 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.

Khosla G., Gill B. S., Sirari A., Sharma P. and Dhaliwal I. 2020. IdentifyingQTL for seed weight in a cross between vegetable and grain type soybeans. Indian J. Genet. Plant Breed., **80**(04): 475–479. doi: 10.31742/IJGPB.80.4.15.

Kosambi D. D. 1944. The estimation of map distance from recombination values. Ann. Eug., **12**: 172–175.

Kulkarni K.P., Asekova S., Lee D.H., Bilyeu K., Song J.T. and Lee J.D. 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): pp.148-155.

Kumar A. 2005. Studies on storability and vigour parameters in soybean. Division of Seed Science and Technology, M. Sc. Thesis. IARI, New Delhi.

Kumar R., Saini M., Taku M., Debbarma P., Mahto R.K., Ramlal A., Sharma D., Rajendran A., Pandey R., Gaikwad K. and Lal S.K. 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**: p.1074245.

Lande R. and Thompson R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, **124:** 743–756.

Li M., Chen L., Zeng J., Razzaq M. K., Xu X., Xu Y., et al. 2020. Identification of additive–epistatic QTLs conferring seed traits in soybean using recombinant inbred lines. Front. Plant Sci., **11**: 566056. doi: 10.3389/ fpls.2020.566056.

Li W., Zheng D. H., Van K. J. and Lee S. H. 2008. QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. merr.). J. Crop Sci. Biotechnol., **11**(3): 171 176.

Liang H., Xu L., Yu Y., Yang H., Dong W. and Zhang H. 2016. Identification of QTLs with main, epistatic and QTL by environment interactioneffects for seed shape and hundred-seed weight in soybean across multiple years. J. Genet., **95**(2): 475–477. doi:10.1007/s12041-016-0648-8.

Liu Y.L., Li Y.H., Reif J.C., Mette M.F., Liu Z.X., Liu B., Zhang S.S., Yan L., Ru-zhen C. and Qiu L.J. 2013. Identification of quantitative trait loci underlying plant height and seed weight in soybean. The Plant Genome, **6**: doi:10.3835/plantgenome2013.03.0006.

Lodhi M. A., Ye G. N., Weeden N. F. and Reisch B. I. 1994. A simple and efficient method for DNA extraction from grapevine cultivars andVitis species. Plant Mol. Biol. Reporter, **12**(1): 6–13.

Lü H. Y., Liu X. F., Wei S. P. and Zhang Y. M. 2011. Epistatic association mapping in homozygous crop cultivars. PloS One, **6**(3): e17773. doi: 10.1371/ journal.pone.0017773.

Maughan P. J., Saghai 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**: 574-579.

Mian M. A. R., Bailey M. A., Tamulonis J. P., Shipe E. R., Carter T. E., 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.

Niu Y., Xu Y., Liu X. F., Yang S. X., Wei S. P., Xie F. T., et al. 2013. Association mapping for seed size and shape traits in soybean cultivars. Mol. Breed., **31**(4): 785–794. doi: 10.1007/ s11032-012-9833-5.

Orf, James. 2010. Introduction. In: Genetics, genomics and breeding of soybean. Bilyeu, K., Ratnaparkhe, M.B. and Kole, C. (Eds.). Science Publishers, NH, USA. pp: 1-18.

Priolli R. H.G., Campos J. B., Stabellini N. S., Pinheiro J. B. and Vello N. A. 2015. Association mapping of oil content and fatty acid components in soybean. Euphytica, **203**(1): 83–96.

Qi Z. M., Wu Q., Han X., Sun Y. N., Du X. Y., Liu C. Y., et al. 2011. Soybeanoil content QTL mapping and integrating with meta-analysis method for mininggenes. Euphytica, **179**(3): 499–514.

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

Saini M., Yadav R.R., Kumar R., Chandra S., Rathod N.K.K., Taku M., Yadav M., Basu S., Rajendran A., Lal S.K. and Talukdar A., 2025. Mapping of quantitative trait loci and mining of candidate genes for seed viability in soybean [*Glycine max* (L.) Merr.]. Front. Plant Sci., **15**: p.1372037.

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.

Song Q. J., Marek L. F., Shoemaker R. C., Lark K. G., Concibido V. C., Delannay X., et al. 2004. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet., **109**(1): 122–128. doi: 10.1007/s00122-004-1602-3.

Talukdar A. and Shivakumar M. 2012. Pollination without emasculation: an efficient method of hybridization in soybean (Glycine max (L.) Merrill). Curr. Sci., **103**(6): 628-630.

Tao Y., Mace E. S., Tai S., Cruickshank A., Campbell B. C., Zhao X., et al. 2017. Whole-genome analysis of candidate genes associated with seed size and weight in sorghum bicolor reveals signatures of artificial selection and insights into parallel domestication in cereal crops. Front. Plant Sci., **8**: 1237. doi: 10.3389/ fpls.2017.01237.

Wu D., Zhan Y., Sun Q., Xu L., Lian M., Zhao X., et al. 2018. Identificationof quantitative trait loci underlying soybean (*Glycine max* [L.] merr.) seed weightincludingmain, epistatic and QTL_ environment effects in different regions ofnortheast China. Plant Breed., **137**: 194–202. doi: 10.1111/ pbr.12574.

Xie F. T., Niu Y., Zhang J., Bu S. H., Zhang H. Z., Geng Q. C., et al. 2014. Fine mapping of quantitative trait loci for seed size traits in soybean. Mol. Breed., **34**(4): 2165–2178. doi: 10.1007/ s11032-014-0171-7.

Xin D., Qi Z., Jiang H., Hu Z., Zhu R., Hu J., Han H., Hu G., Liu C. and Chen Q. 2016. QTL location and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* Sieb. &Zucc.) chromosome segment substitution lines. PLoS One, **11**: e0149380. doi:10.1371/journal.

Xu Y., Li H. N., Li G. J., Wang X., Cheng L. G. and Zhang Y. M. 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.

Yan L., Hofmann N., Li S., Ferreira M. E., Song B., Jiang G., et al.

2017. Identification of QTL with large effect on seed weight in a selective population of soybean with genome-wide association and fixation index analyses. BMC Genomics, **18**(1): 1–11.

Yashpal Rathod D. R., Jyoti D., Anil K., Keya M., Deepika C., Subhash C. and Akshay T. 2015. Genomic variation studies in Glycine max and Glycine soja using SSR markers. Curr. Sci., **109**(11): 1929-1931.

Yu M., Liu Z. and Wilson J. D. O. 1995. "Storage of orthodox seeds," in Seed quality: basic mechanisms agricultural implications'. (Ed. AS Basra) pp. 173–208.

Zhang D., Cheng H., Wang H., Zhang H., Liu C. and Yu D. 2010. Identification of genomic regions determining flower and pod numbers development in soybean (*Glycine max* L.). J. Genet. Genomics, **37**(8): 545–556. doi: 10.1016/S1673-8527(09)60074-6.

Zhang W. K., Wang Y. J., Luo G. Z., Zhang J. S., He C. Y., Wu X. L., et al. 2004. QTL mapping of ten agronomic traits on the soybean (*Glycine max* L.Merr.) genetic map and their association with EST markers. Theor. Appl. Genet., **108**(6): 1131–1139.