



OPEN ACCESS

EDITED BY

Yi Tian,
Agricultural University of Hebei, China

REVIEWED BY

Jianbo He,
Nanjing Agricultural University, China
Yang Li,
Henan University, China

*CORRESPONDENCE

Jiajia Li
✉ ljia6862@163.com

[†]These authors have contributed
equally to this work

RECEIVED 02 February 2026
REVISED 06 March 2026
ACCEPTED 11 March 2026
PUBLISHED 02 April 2026

CITATION

Wang G, Zhao D, Hu X, Zheng H,
Wang W, Bai P, Miao L, Gao H, Wang L,
Sun H and Li J (2026) Natural
variation of the *GmDt1* gene affects
the 100-seed weight of soybean.
Front. Plant Sci. 17:1801920.
doi: 10.3389/fpls.2026.1801920

COPYRIGHT

© 2026 Wang, Zhao, Hu, Zheng, Wang,
Bai, Miao, Gao, Wang, Sun and Li. 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.

Natural variation of the *GmDt1* gene affects the 100-seed weight of soybean

Guoji Wang[†], Duo Zhao[†], Xiaoyu Hu[†], Haowei Zheng,
Wei Wang, Pengyu Bai, Long Miao, Huihui Gao,
Longlong Wang, Hongye Sun and Jiajia Li*

School of Agronomy, Anhui Agricultural University, Hefei, China

100-seed weight (100-SW) is a critical determinant of soybean yield. The identification and functional characterization of its underlying genes are therefore essential for the genetic improvement of seed and yield-related traits. A residual heterozygous line (RHL) segregating for 100-SW was derived from a recombinant inbred line (RIL) population generated by crossing small-seed (19.75 ± 1.93 g) and large-seed (26.20 ± 0.82 g) soybean parents. Phenotypic segregation of 100-SW was analyzed, and Chi-square test was used to verify the segregation ratio. Bulk segregant analysis combined with whole-genome sequencing (BSA-seq) was performed using both Euclidean distance and index algorithms to map the target gene. Functional annotation, molecular marker validation, and germplasm resequencing were conducted to identify the key candidate gene and its haplotypes. Phenotypic analysis showed significant segregation and normal distribution of 100-SW in the RHL, with a Chi-square-verified 1:2:1 segregation ratio, indicating control by a single nuclear gene. BSA-seq mapped the gene to a 5.46 Mb region on chromosome 19, where 74 non-synonymous SNPs in coding sequences were identified (including one causing initiation codon loss), distributed across 36 genes. *GmDt1* (*Glycine max Determinant stem 1*) was confirmed as the key candidate gene, with a G-to-T non-synonymous mutation in its first exon as the functional locus (validated in the original RIL population). Resequencing of diverse germplasm classified *GmDt1* into five haplotypes; the large-seed haplotype *GmDt1*-H2 was absent in wild soybeans, present in 9.07% of landraces, and 15.83% of cultivated soybeans. The gradual increase in the frequency of *GmDt1*-H2 from wild to cultivated soybeans suggests that this haplotype has been positively selected during soybean breeding. Identification of *GmDt1* and its functional mutation provides a valuable molecular target for the genetic improvement of soybean seed traits and yield.

KEYWORDS

100-seed weight, BSA-seq, haplotype, soybean, variation

1 Introduction

As one of China's major grain, oil, and cash crops, soybean provides a substantial source of plant protein and edible oil for human consumption, and also serves as a primary protein source for animal feed (Li et al., 2014; Miao et al., 2020). According to data from the National Bureau of Statistics of China, the average soybean yield in 2025 was 2039 kg per hectare, which

is not only lower than the global average (approximately 2930 kg per hectare) but also far below that of major soybean-producing countries such as the United States and Brazil (approximately 3750 kg per hectare). Moreover, China's large population exacerbates the imbalance between domestic supply and demand, resulting in heavy reliance on soybean imports. Consequently, improving soybean yield in China has thus become a pressing issue in Chinese agriculture. Soybean yield per unit area is determined by both planting density and individual plant productivity, with 100-seed weight (100-SW) being a key factor influencing the latter (Paran and Zamir, 2003). Therefore, research on soybean 100-SW holds important theoretical and practical significance.

During the domestication of wild soybeans into cultivated varieties, seed size has been progressively selected, with humans preferentially favoring seeds with higher 100-seed weight (Lee et al., 2011; Han et al., 2016). Seed size is a complex trait influenced by multiple genetic and environmental factors, with genetic factors playing a predominant regulatory role. Studies have reported that the broad-sense heritability of seed size can reach as high as 0.97 (Zhang et al., 2016; Yan et al., 2017). In soybean, several genes and loci have been identified as key regulators of seed size. For instance, the *POWRI* gene undergoes truncation in its CCT domain due to transposon insertion, resulting in increased 100-seed weight (Goettel et al., 2022). *GmSWEET10a/b* modulates seed size by transporting carbohydrates from the seed coat to the inner seed, and *GmSWEET10a* has been shown to interact with *GmDtl* to coordinately regulate soybean seed size (Wang et al., 2020; Li X. et al., 2024). *GmSSS1* encoding a SPINDLY homolog that affects cell proliferation and division, thereby influencing 100-SW (Zhu et al., 2022). A T/C polymorphism at the *SW9-1* locus on chromosome 9 accounts for 10.05%–10.93% of the phenotypic variation in seed size (Li J. et al., 2019). *GmSMS6* interacts both physically and genetically with the transcription factor *GmbZIP151* and the RING-type E3 ligase *GmUBQ1*; as a regulatory hub, it coordinates the transcriptional activation by *GmbZIP151* and the *GmUBQ1*-mediated protein stability, and primarily repressing cell expansion in soybean seeds (Li et al., 2025). Natural variation in the promoter region of *GmST05* affects its transcriptional level, thereby regulating seed size (Duan et al., 2022). Several transcription factors including *GmMYB73*, *GmMYB181*, *GmBS1*, *GmBS2*, *GmWRKY15a*, and *GmSW14*, have been implicated in seed size regulation (Liu et al., 2014; Ge et al., 2016; Gu et al., 2017; Yang et al., 2018; Zhang et al., 2025) while substrate enzymes such as *GmFAD3*, *GmCYP78A10*, *GmCYP78A72*, *GmCYP78A5*, *GA20OX*, *GmCIF1*, *GmUBC1*, and *GmSW17* also contribute to the regulation of seed size (Singh et al., 2011; Wang J. et al., 2015; Lu et al., 2016; Zhao et al., 2016; Du et al., 2017; Tang et al., 2017; Mao et al., 2020; Liang et al., 2024).

Bulked segregant analysis (BSA) is a target gene mapping approach that involves pooling individuals with extreme phenotypes from segregating progeny within a population, followed by comparison of polymorphisms between the contrasting bulks and their associated phenotypes (Michelmore et al., 1991; Abe et al., 2012). Compared with traditional QTL mapping methods, BSA analysis of only a small subset of extreme individuals, thereby simplifying the experimental process and reducing costs (Yang

et al., 2013). BSA-seq can be applied to various population types, including segregating genetic populations, natural populations, and multi-family populations. In agricultural research, segregating populations are most commonly used and include F_1 , F_2 or $F_{2:3}$ -derived populations, backcross populations, RIL populations, and doubled haploid (DH) populations. Secondary segregating populations, such as chromosome segment substitution lines (CSSLs) and residual heterozygous lines (RHLs), are also suitable for BSA-based bulk analysis (Wang et al., 2023).

In this study, two major cultivated soybean cultivars ZP661 and ZH13 exhibiting significant differences in 100-SW were used to construct a mapping population. The candidate gene *GmDtl* was identified through BSA-based bulked pool sequencing, and further analysis revealed that a non-synonymous mutation in the first exon of *GmDtl* serves as the functional locus regulating 100-SW. This study thus provides a valuable molecular target for the improvement of soybean yield.

2 Materials and methods

2.1 Experimental materials

Two soybean cultivars ZP661 (100-SW: 19.75 g) and ZH13 (100-SW: 26.20 g) were used in this study. These parental lines were crossed in 2019 to produce F_1 progeny, which were subsequently self-pollinated to generate F_2 individuals in 2020. A RIL population was ultimately established after ten successive generations of selfing. From this population, a RHL exhibiting significant phenotypic segregation for 100-SW was identified (Supplementary Figure 1). To minimize environmental variation, the experimental materials were planted at two locations, Suzhou and Mingguang, in Anhui Province, China. All materials were arranged in a randomized complete block design (RCBD) and managed according to standard field practices. The 100-SW of each line was measured at soybean maturity.

2.2 Phenotypic and genetic analysis

Phenotypic evaluation of 100-SW in RHLs and RILs was conducted as follows: 100 seeds were randomly selected from each line, and their total weight was measured using an electronic balance.

Individual plants in the segregating RIL population were tagged and sampled at the R6 growth stage. Following harvest, the 100-SW data of mature plants were used to perform a chi-square test to determine the genetic inheritance of this trait. The analysis was carried out using the following formula:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

O_i represents the observed number of plants; E_i represents the expected number of plants, k represents the total number of categories.

2.3 BSA-Seq based candidate interval mapping

Based on the phenotypic traits of the two parental lines and the RHLs, four DNA pools were constructed: two parental pools, a small-seed bulk (Bulk A), and a large-seed bulk (Bulk B). Each parental pool consisted of 10 individual plants from ZP661 or ZH13, respectively. For the progeny bulks, 30 ZP661-like plants and 30 ZH13-like plants were selected from the segregating RHL population, and leaf samples from these plants were pooled to construct the two progeny bulks. Sequencing was performed at a depth of 30× for the progeny bulks and 10× for the parental pools. The experimental procedure was as follows: young top leaves were collected from each selected plant, and genomic DNA was extracted from individual leaf samples using the CTAB method. After assessing DNA quality and quantifying DNA concentration, equal amounts of DNA from each plant within a pool were combined to generate the four bulks. Subsequent sequencing and bioinformatic analyses were performed by Biomarker Technologies Corporation (Beijing, China).

To ensure the quality of subsequent analyses, raw sequencing reads generated by the Illumina platform were stringently filtered to obtain high-quality clean reads. All data processing followed a standardized workflow, with the software and specific steps described below. The processed data were subsequently used for alignment variant detection, and other downstream analyses.

FastQC (v0.11.9) was used for preliminary quality assessment of the raw sequencing data. This included evaluating sequence quality distribution, base composition, adapter contamination, and other metrics. Trimmomatic (v0.39) was employed to remove low-quality reads, adapter sequences, and abnormal sequences; Samtools (v1.9) was used for format conversion, sorting of alignment results, and related statistical analyses. MultiQC (v1.14) was applied to visualize FastQC results and integrate data quality control reports.

Raw reads were filtered using Trimmomatic software with the following core parameter settings: LEADING: 3, TRAILING: 3, SLIDINGWINDOW: 4: 15, MINLEN: 100. The specific filtering criteria are described below, fully consistent with the subsequent results section. Reads containing adapter sequences were removed using the built-in Illumina adapter library. After adapter removal, paired-end reads shorter than 100 bp were filtered out. Paired-end reads longer than 100 bp were retained for subsequent analysis.

Filtering of reads with high N content: The proportion of unrecognized bases (N) in each read was calculated. Reads with an N content exceeding 10%, along with their paired-end sequences were filtered out. Removal of low-quality reads: A base quality threshold of Q10 corresponding to a sequencing error rate of 10%, was applied. If more than 50% of the bases in a read had a quality value below 10, the read and its paired-end sequence were discarded.

Euclidean Distance (ED) algorithm for association analysis: This method uses sequencing data to identify markers showing significant differences between bulks, thereby pinpointing genomic regions associated with a trait. In theory, loci not associated with the target trait should show little to no difference between the two bulks,

while loci linked to the target trait will display differences. Consequently, ED values for non-target loci are expected to approach 0. Larger ED values indicate greater differences in marker allele frequencies between the two bulks. The ED is calculated using the following formula, where “mut” denotes the allele frequency in the mutant bulk, and “Wt” denotes the allele frequency in the wild-type bulk.

$$ED = \sqrt{\frac{(A_{mut} - A_{Wt})^2 + (C_{mut} - C_{Wt})^2}{(G_{mut} - G_{Wt})^2 + (T_{mut} - T_{Wt})^2}}$$

Index association analysis is a marker-trait association method based on differences in genotype frequencies between bulks. It identifies significant variations in genotype frequencies and quantifies these differences using the Δ_{index} metric. The calculation formula is as follows, where “aa” represents the wild-type progeny bulk, “ab” represents the mutant progeny bulk, and “M” and “P” denote the read counts of alleles derived from the wild-type parent and the mutant parent in their respective bulks. The Δ_{index} allows the assessment of differences at each locus between the mutant and wild-type bulks. Specifically, the stronger the association between a marker and the target trait, the closer the Δ_{index} value is to 1.

$$\text{Index (aa)} = M_{aa} / (M_{aa} + P_{aa})$$

$$\text{Index (ab)} = M_{ab} / (M_{ab} + P_{ab})$$

$$\Delta_{index} = \text{Index (aa)} - \text{Index (ab)}$$

2.4 Alignment of sequencing reads, annotation of variants and haplotype analysis

Sequencing reads were aligned, and detected variants (e.g., the variant located 183 bp downstream of the ATG initiation codon of *GmDt1*) were annotated using the soybean reference genome (*Glycine max* Wm82.a4.v1) and the corresponding gene annotation files. These resources were downloaded from the Phytozome database (<https://phytozome-next.jgi.doe.gov/>), a comprehensive plant genome repository hosted by the US Department of Energy Joint Genome Institute (JGI). Haplotype analysis of variant sites was conducted using the Soybean Genomic Variation Database (SoyGVD, <https://yanglab.hzau.edu.cn/SoyGVD/#/variation>).

2.5 RNA extraction and reverse transcription from plant tissues

Total RNA was extracted from soybean tissues using the TaKaRa MiniBEST Plant RNA Extraction Kit (TaKaRa, Dalian, China). RNA integrity was verified by electrophoresis, and only samples showing clear 28S and 18S ribosomal RNA bands were used for subsequent experiments. RNA concentration was then measured using a micro-spectrophotometer, and the samples were stored at -80 °C until use. First-strand cDNA was synthesized using the PrimeScriptTM RT.

Reagent Kit with gDNA Eraser (Perfect Real Time) (TaKaRa, Dalian, China) following the manufacturer's instructions.

2.6 Primer design

The soybean reference genome from the Phytozome database (<https://phytozome-next.jgi.doe.gov/>) was used to retrieve the target amplicon sequences. Specific primers were designed based on these sequences using Primer 5.0 software. The primers used in this study are listed in Table 1.

2.7 Development of dCAPS molecular markers

The dCAPS markers were developed using SnapGene software. PCR amplification was performed on the selected polymorphic loci using Taq-Master DNA polymerase. The specific PCR reaction system (15 μ L total volume) and conditions are detailed as follows: PCR reaction system (15 μ L): ddH₂O 5.3 μ L, Taq Master Mix 7.5 μ L, forward/reverse primers (F/R) 0.6 μ L each, and genomic DNA 1 μ L. PCR reaction conditions: pre-denaturation at 95 °C for 5 min; followed by 36 cycles of denaturation at 98 °C for 30 s, annealing at 56 °C for 30 s, and extension at 72 °C for 45 s; followed by a final extension at 72 °C for 5 min. PCR products were analyzed by 1.5% agarose gel electrophoresis to verify the presence of a single target band. Qualified PCR products were then digested with the appropriate restriction endonuclease for 8 hours. The fragment lengths of the digested products were determined by 2.5% agarose gel electrophoresis.

2.8 Expression analysis of *GmDtl*

GmTubulin was used as the reference gene for quantitative real-time PCR (qRT-PCR) normalization. The qRT-PCR assay was performed using AceQ Universal SYBR qPCR Master Mix, which contains the fluorescent SYBR Green I DNA-binding dye (Vazyme, Nanjing, China). The assay was conducted away from direct strong light to prevent dye degradation. The total reaction volume was 20 μ L, consisting of 10 μ L of 2 $\times$ AceQ Universal SYBR qPCR Master Mix, 7.2 μ L of ddH₂O, 0.4 μ L each of forward and reverse primers, and 2 μ L of cDNA template. Each reaction was performed in three biological replicates to ensure reproducibility.

2.9 Statistical analysis

All experimental data were obtained from three or more independent replicate experiments and analyzed using Student's t-test. Data collation and statistical analyses were performed using Microsoft Excel. Graphs were generated using GraphPad Prism 8 software. A statistically significant difference between two data groups was defined at the 95% confidence level.

3 Results

3.1 Phenotypic identification and genetic analysis of 100-SW in residual heterozygous lines

In the previously constructed F_{2:10} RIL population derived from a cross between ZP661 (100-SW: 19.75 $\pm$ 1.93 g) and ZH13 (100-SW: 26.20 $\pm$ 0.82 g), a segregating RHL was identified. This RHL line was planted in Suzhou and Mingguang, Anhui Province, where significant segregation of the 100-SW trait was observed among the segregating individuals. The 100-SW values of these plants exhibited a normal distribution at both locations (Figure 1). The mean 100-SW was 23.12 $\pm$ 3.47 g with a coefficient of variation (CV) of 14.98% in Suzhou, whereas the mean 100-SW in Mingguang was 21.42 $\pm$ 3.94 g with a CV of 18.39%. Due to the larger sample size and more stable genetic variation at the Suzhou site, phenotypic characterization was performed on 464 individual plants of this RHL line grown in Suzhou for subsequent genetic studies. Chi-square test results revealed that the phenotypic segregation ratio fitted 1:2:1 ($P > 0.05$), indicating that 100-SW in this population is likely controlled by a single major gene (Figure 1A, Table 2).

3.2 Candidate gene mapping based on BSA bulked sequencing

3.2.1 Data quality assessment

To further refine the mapping of candidate genes, two extreme bulks (Bulk A and Bulk B) were constructed. The average 100-SWs of Bulk A and Bulk B were 26.11 $\pm$ 1.12 g and 19.92 $\pm$ 1.38 g,

TABLE 1 Primers used in this study.

Primer names	Primer sequence 5'-3'
dCAPS-19g181900-F	ATGCTATCAGATGTAGGGATGAACCAGAGTTACTTCAAAATGTAAACC
dCAPS-19g181900-R	GAATCTTGGCTTCTTTTGTGTGTGTGAG
<i>Dtl</i> -frag1-F	GGGGCAAAACACACTCGAT
<i>Dtl</i> -frag1-R	AGTTCTGTAATGTTTGTGTTGAGACT
<i>Dtl</i> -frag2-F	GTTTCTCTTAATAACTTAACCTCTT
<i>Dtl</i> -frag2-R	TACTACAGAACGTACACAACATCT
<i>qDtl</i> -F	TGTTCTCTCTTACAATGGCA
<i>qDtl</i> -R	GCTTCTGTGTAACTCACAGTC

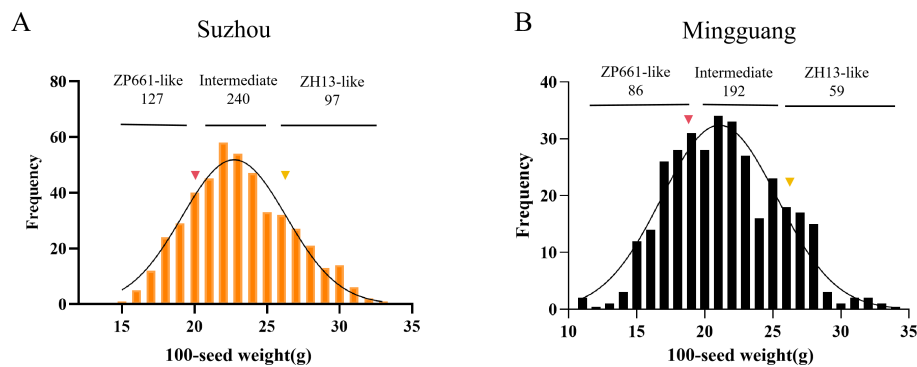


FIGURE 1

Characteristics of 100-SW distribution in residual heterozygous lines. (A) Distribution of 100-seed weight in recombinant heterozygous lines grown in Suzhou, Anhui Province, ranging from 15.21 to 32.72 g ($n = 464$). (B) Distribution of 100-seed weight in recombinant heterozygous lines grown in Mingguang City, Anhui Province, ranging from 11.56 to 33.08 g ($n = 337$). Red triangles indicate the 100-SW of ZP661, and yellow triangles indicate the 100-SW of ZH13. Each individual plant's 100-seed weight was measured in three technical replicates.

respectively, showing an extremely significant difference (Figure 2). Candidate gene mapping was conducted using the BSA-based bulked sequencing approach. Whole-genome resequencing was performed on the four samples, followed by data filtering. Reads containing adapters were removed. Reads with an N content exceeding 10% were discarded. Reads in which more than 50% of bases had a quality score below 10 were also excluded. The final results indicated that the Q30 value exceeded 80%, demonstrating that the sequencing data were of sufficient quality for subsequent analyses (Table 3).

3.2.2 Association analysis

Association analysis of these high-quality SNPs was conducted using two algorithms: the ED algorithm and the Index algorithm. Results showed that the ED algorithm-based SNP association analysis identified mapping intervals on chromosomes 1, 3, 4, 6, 13, and 19, with respective sizes of 3.66 Mb, 0.58 Mb, 0.18 Mb, 1.62 Mb, 3.98 Mb, and 6.06 Mb (Figure 3A, Table 3). In comparison, the Index algorithm-based SNP association analysis identified a single 5.46 Mb mapping interval on chromosome 19 (Figure 3B, Table 4). The intersection of results from the two association analysis methods defined a final 5.46 Mb candidate region on chromosome 19 (Table 4).

3.2.3 Mining of candidate genes

Analysis of variant sites within the 5.46 Mb candidate interval identified 74 single-nucleotide polymorphisms (SNPs) that induced non-synonymous mutations in genes and one SNP that resulted in the loss of the start codon. These functional variants were distributed across the coding regions of 36 genes (Supplementary Table 1). We analyzed the expression patterns and predicted biological functions of these 36 genes, and excluded those specifically expressed in tissues such as roots, flowers, and leaves. Finally, two auxin response factors, *Glyma.19G181900* and *Glyma.19G206100*, as well as one pod setting habit regulatory

gene *Glyma.19G194300* (*GmDt1*, an ortholog of *Arabidopsis TFL1*), were identified as candidate genes underlying the target phenotype (Table 5). Further analysis of the BSA sequencing results for these three putative candidate genes showed that *Glyma.19G194300* and *Glyma.19G181900* exhibited relatively high ED and Δ index values. Therefore, these two genes were selected for subsequent investigation. First, validation analysis was conducted on *Glyma.19G181900*. A dCAPS molecular marker was developed based on the variant site to examine whether the phenotype of segregated progeny was tightly linked to this variation (Supplementary Figure 2A). The results showed that among 19 ZP661-like and intermediate individuals, one plant exhibited no linkage between the variation and phenotype. Among 15 ZH13-like individuals, three plants showed unlinked variation and phenotype (Supplementary Figure 2B). These results excluded *Glyma.19G181900* as a regulator of 100-SW. Subsequently, CAPS markers and sequencing were used to perform a correlation analysis between genetic variations in the *GmDt1* gene and phenotypic traits. The results revealed complete co-segregation between the variation and phenotype. Individuals with the T/T genotype had a 100-SW ranging from 24.68 to 28.20 g; those with the G/G genotype ranged from 16.66 to 21.54 g; and those with the G/T genotype ranged from 22.83 to 24.85 g (Figure 4; Supplementary Table 2). These results indicate that *GmDt1* is a candidate gene regulating 100-SW in soybean, with the functional variant located at position 183 downstream of the ATG initiation codon.

3.3 Analysis of relative expression level of *GmDt1* in parents

To verify the differential expression of *GmDt1* in seeds of the parental lines, total RNA was extracted from seeds of ZH13 and ZP661 at the R5.2 and R5.3 stages, cDNA was then synthesized for quantitative real-time PCR analysis. The results showed that the expression level of *GmDt1* in ZH13 was significantly higher than that in ZP661 at both developmental stages (Figure 5).

TABLE 2 Genetic analysis of 100-SW in soybean.

Plant materials	Sum	Expected value	Observed value	Expected ratio	χ^2	P-value
ZP661-like	464	116	127	1:2:1	4.43	0.11
Intermediate		232	240			
ZH13-like		116	97			

3.4 Analysis of 100-SW among different allelic types of *GmDt1*

To further verify the effect of *GmDt1* alleles on 100-SW, sequencing was used to identify the *GmDt1-G* and *GmDt1-T* alleles in the RIL population derived from ZH13 and ZP661. Forty lines carrying each allele were randomly selected to compare their 100-SW values. The results showed that the average 100-SW of RIL^{*Dt1-G*} was 18.68 ± 1.21 g, whereas that of RIL^{*Dt1-T*} was 26.42 ± 1.46 g. This represents a significant increase of 41.45% in RILDt1-T compared with RILDt1-G (Figure 6).

3.5 *GmDt1* haplotype analysis

To investigate whether *GmDt1* alleles have been exploited during soybean domestication and improvement, genotyping of *GmDt1* was performed using resequencing data from public databases. A total of four non-synonymous mutations were

identified in the coding sequence (CDS) region, which were classified into five haplotypes (Figure 7A). The parental lines ZP661 and ZH13 belonged to haplotypes H1 and H2, respectively. Analysis of the evolutionary characteristics of these five haplotypes revealed that wild soybean (*G. soja*) predominantly harbored the H1 haplotype (99.61%), with a small proportion of the H4 haplotype (0.39%). No other haplotypes were detected in wild soybeans. In contrast, three additional haplotypes (H2, H3, and H5) emerged in landraces and improved cultivars.

A total of 523 soybean accessions with clear provenance were randomly selected to analyze the differences in 100-seed weight among various haplotypes of *GmDt1*. The results showed that Hap2 exhibited the highest mean 100-SW (20.46 ± 3.85 g), which was significantly higher than that of Hap1 (17.83 ± 4.72 g) (Figure 8). Moreover, the proportion of H2, the haplotype associated with high 100-SW identified in this study, continuously increased from landraces to improved cultivars, rising from 9.07% to 15.83% (Figure 7B). These results indicate that the *GmDt1-H2* haplotype related to high 100-SW may have been gradually selected by breeders during soybean improvement.

To further investigate whether the gradual selection of the *GmDt1-H2* haplotype was driven by 100-SW rather than stem growth habit, we selected 165 released soybean cultivars carrying this haplotype. These accessions included lines from the Northern ecological zone, Huang-Huai-Hai ecological zone, and Southern ecological zone in China. Their stem growth habits were examined. The results showed that 130 accessions were determinate, 31 were semi-determinate, and 4 were indeterminate (Supplementary Table 3). These findings further confirm that the *GmDt1-H2* haplotype carrying *GmDt1-T* was selected during breeding primarily for high 100-SW, rather than for the determinate stem growth habit.

4 Discussion

4.1 Phenotypic variation in 100-SW of soybean

As a major oilseed and cash crop, soybean provides humans and livestock with abundant nutrients, including proteins and fats. In recent years, although global soybean yield has continued to increase, projections indicate that soybean production will need to double by 2050 to meet the growing demand (Jeong et al., 2012; Bodirsky et al., 2015). 100-SW is a key determinant of soybean yield. Developing soybean cultivars with an optimal 100-SW has thus become one of the primary objectives of plant breeders (Chen et al., 2023; Kumar et al., 2023; Xu et al., 2023). Notably, 100-SW in soybean is a

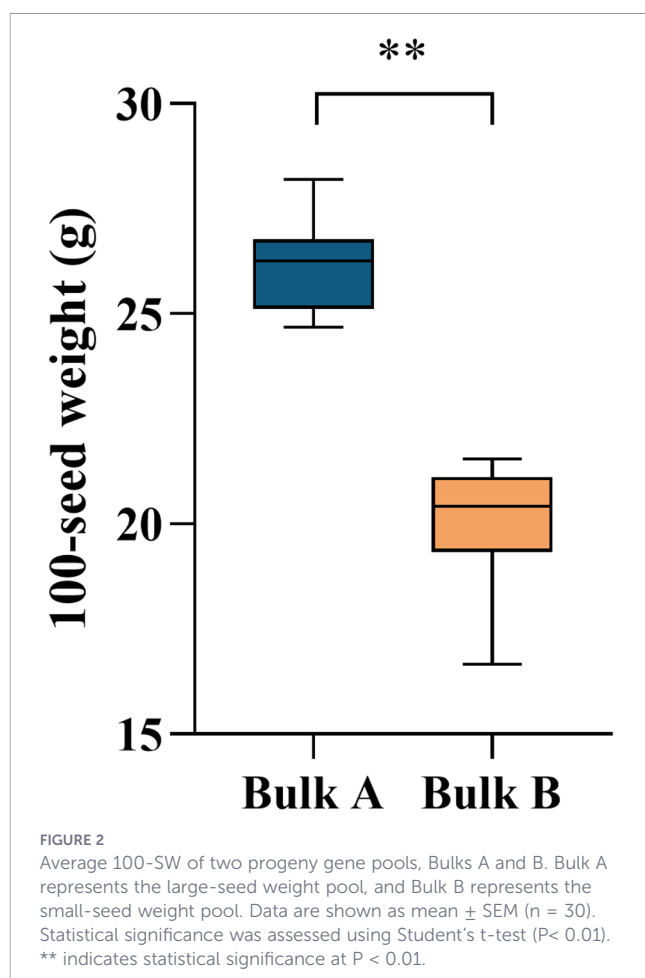


TABLE 3 The QC data of the whole genome sequencing.

Library ID	Clean_reads	Clean_base	Q20 (%)	Q30 (%)	GC (%)
Bulks A	201945448	30157281558	99.58	96.97	35.79
Bulks B	218916416	32690298970	99.56	96.76	36.01
ZP661	91995608	13733918996	99.57	96.92	35.39
ZH13	97118664	14499465106	99.59	97.02	35.56

quantitative trait controlled by multiple genes (Yan et al., 2017; Li X. et al., 2019). Therefore, identifying key genes regulating 100-SW will not only facilitate soybean yield improvement but also provide molecular targets for high-yield breeding programs.

In this study, RHLs derived from the cross between the small-seeded cultivar ZP661 and the large-seeded cultivar ZH13 were used to examine phenotypic variation and perform genetic analysis of soybean 100-SW. The 100-SW of these materials varied widely, ranging from approximately 15 g to 33 g, and followed a normal distribution. Previous studies have shown that the 100-SW of cultivated soybean varieties is roughly six times that of wild soybeans, reflecting substantial genetic variation (Zhang et al., 2016; Yu et al., 2017; Zhao et al., 2019). Consistent with these reports, the RHL population in this study

also exhibited significant genetic variation, making it suitable for further genetic analysis.

Genetic analysis indicated that 100-SW in this RHL population is regulated by *GmDt1*. This finding differs from most previous mapping studies on 100-SW, which generally reported regulation by multiple quantitative trait loci (QTLs) (Li Y. et al., 2024). This discrepancy may be due to the multiple generations of selfing in RHLs, which lead to homozygosity of non-target traits and minor-effect loci, while retaining only a few major-effect loci. Consequently, *GmDt1* acts as a major QTL in this specific genetic background. Such single-gene regulation is relatively common when mapping of 100-SW genes using soybean mutants. For example, progeny from the cross between the *sss1* mutant and its wild-type parent ZP661 showed a single-gene regulatory pattern, and *GmSSS1* was identified as the

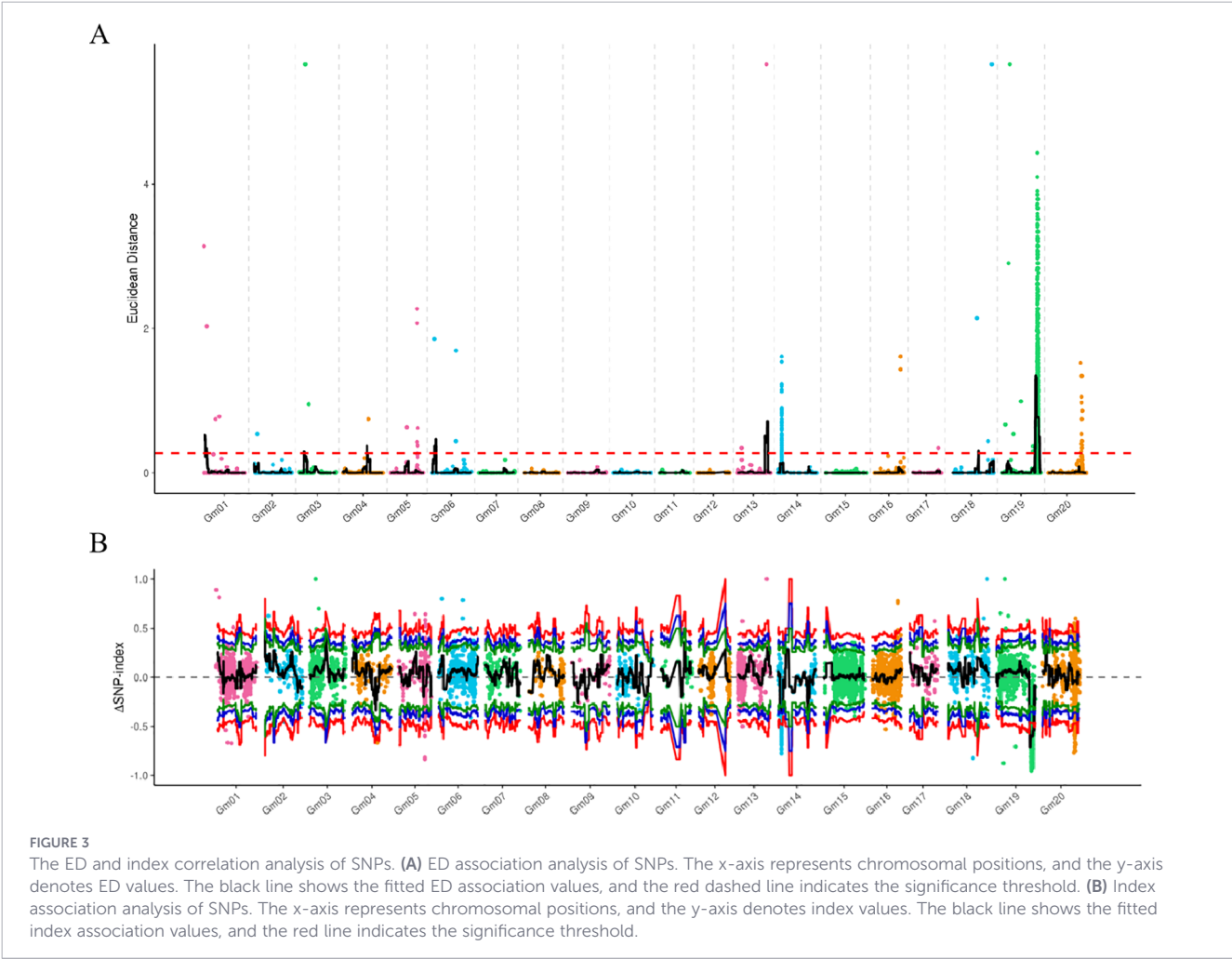


TABLE 4 Associated region on chromosome 19 based on SNP analysis.

Association analysis type	Chromosome ID	Start position	End position	Region size (Mb)
SNP correlates results (ED)	Gm01	0	2000000	2
	Gm01	2100000	3720000	1.62
	Gm01	3780000	3820000	0.04
	Gm03	6400000	6980000	0.58
	Gm04	30040000	30220000	0.18
	Gm06	4840000	6460,000	1.62
	Gm13	34900000	38880000	3.98
	Gm19	42660000	48720000	6.06
SNP correlates results (SNP-index)	Gm19	43240000	48700000	5.46
The intersection of association results	Gm19	43240000	48700000	5.46

TABLE 5 Analysis of candidate gene variation loci.

Gene	Site	Variant base and amino acid	ED	Delta-index
<i>Glyma.19g194300 (GmDt1)</i>	45625529 (1st exon)	AGG->AGT, R->S	1.018	0.720
<i>Glyma.19G181900</i>	44489593 (3rd exon)	AGG->AAG, R->K	0.997	0.705
<i>Glyma.19G206100</i>	46699857 (3rd exon)	ATG->TTG, M->L	0.893	0.631

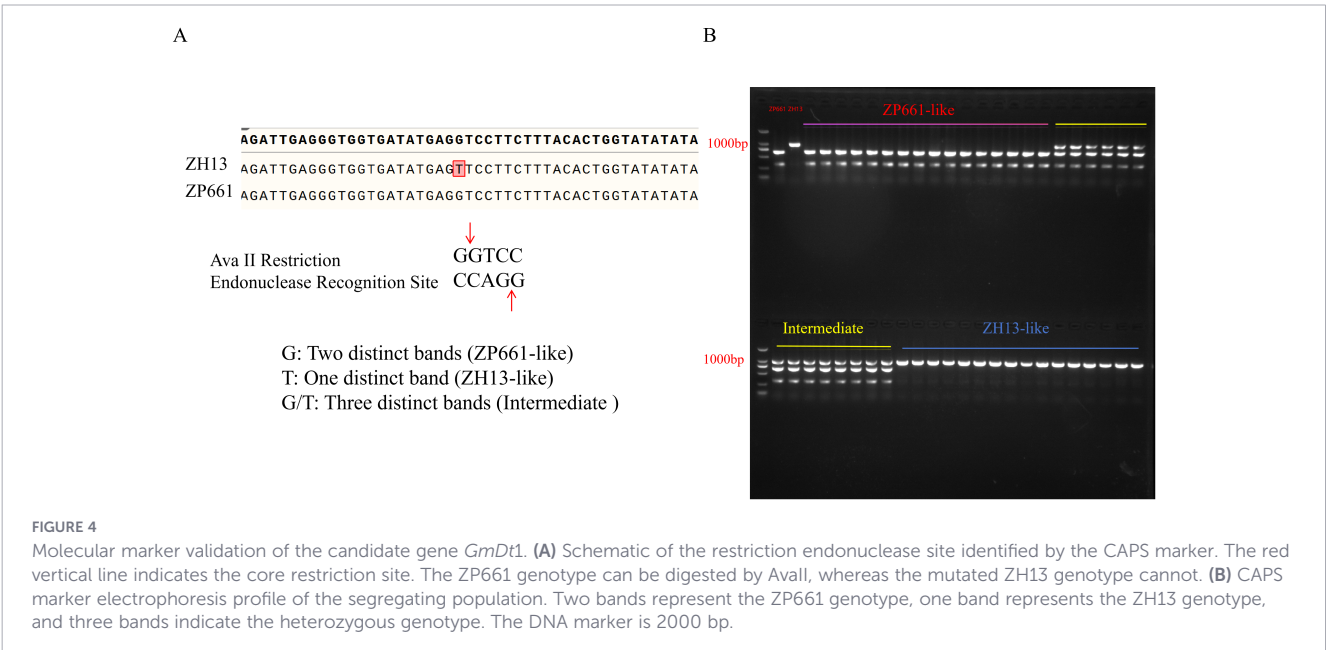
candidate gene through fine mapping (Zhu et al., 2022). Therefore, the 100-SW regulatory gene identified in this study can be inferred as a major-effect candidate gene.

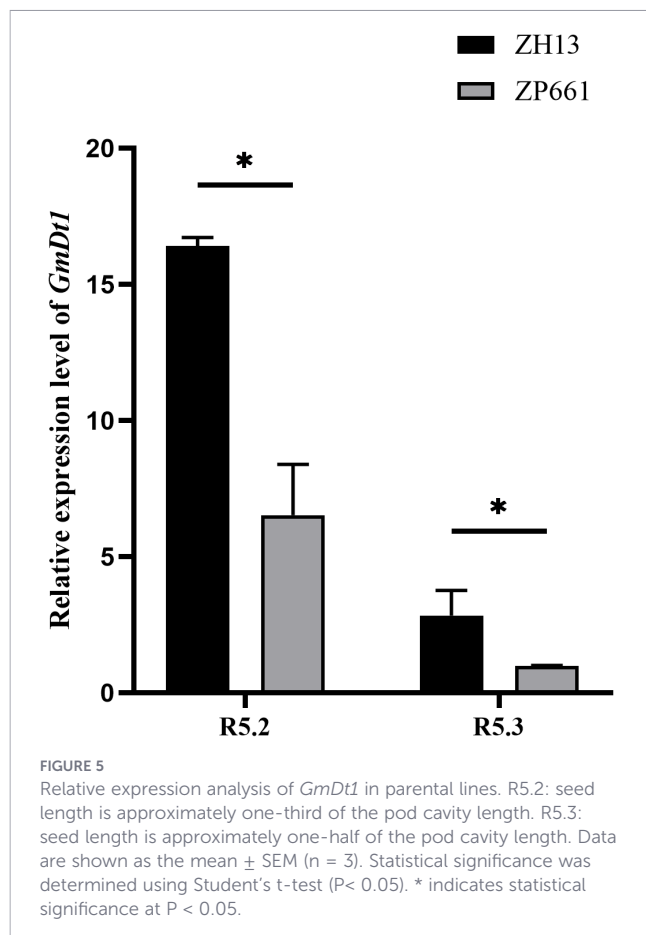
4.2 Analysis of candidate genes for soybean 100-SW

BSA-seq is a cost-effective strategy for primary QTL detection, but it usually results in relatively large genomic intervals. Further fine mapping or GWAS is typically required to narrow down the target region. In the present study, BSA-seq was employed to map the gene regulating soybean 100-SW, which was initially localized to a 5.46 Mb interval on

chromosome 19. This result is supported by numerous previous studies focusing on genetic loci associated with the 100-SW trait. Multiple QTLs (seed weight 3-5; seed weight 7-7; seed weight 12-3; seed weight 13-9; seed weight 17-1; seed weight 42-4) on chromosome 19 have been consistently reported to be closely linked to soybean 100-SW (Mian et al., 1996; Orf et al., 1999; Csanádi et al., 2001; Hoeck et al., 2003; Stombaugh et al., 2004; Wang X. et al., 2015). Collectively, these pieces of evidence confirm that this region of chromosome 19 is a conserved key segment involved in regulating soybean 100-SW. This finding validates the reliability of our initial mapping result.

To identify the causal gene within this 5.46 Mb interval, further bioinformatic analysis and functional annotation suggested that

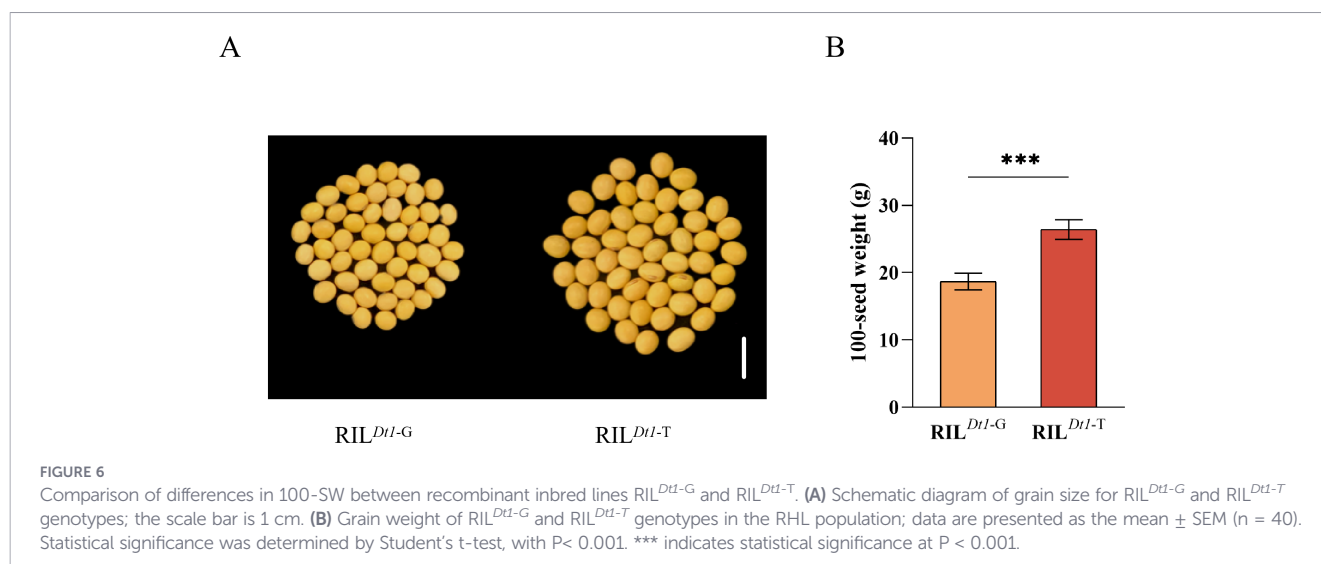


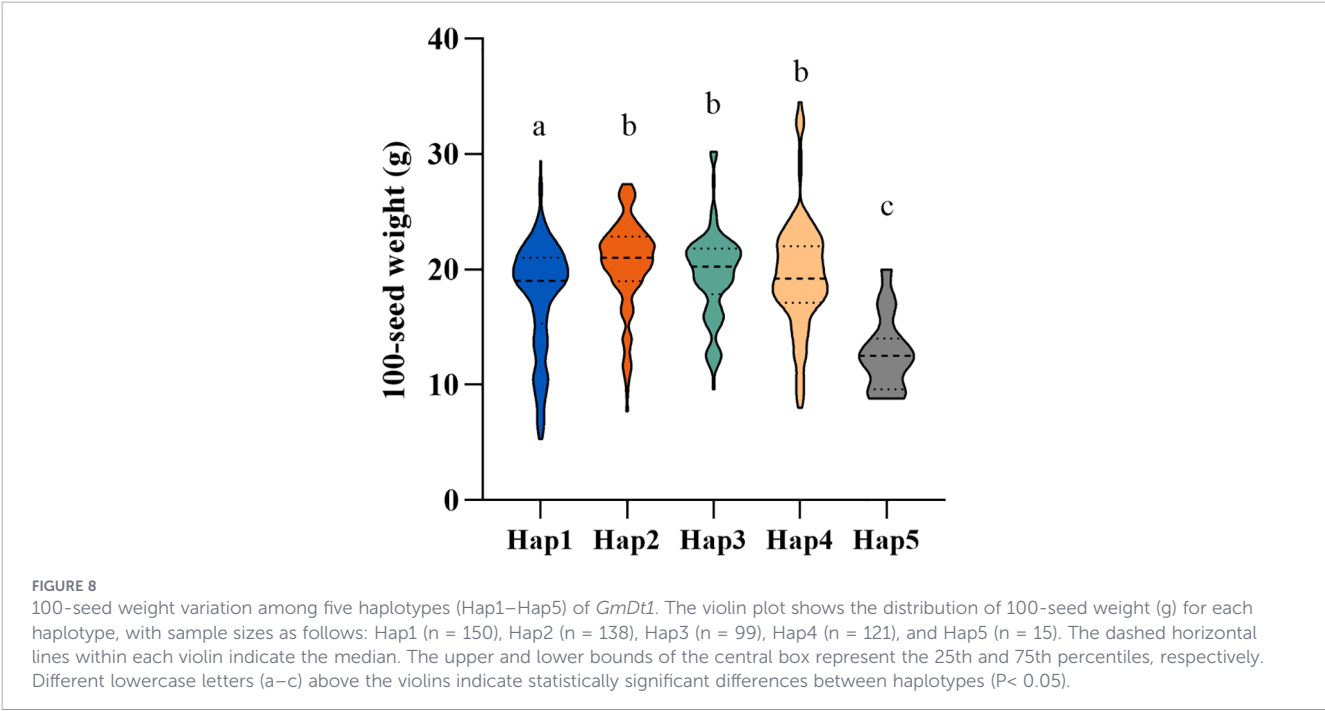
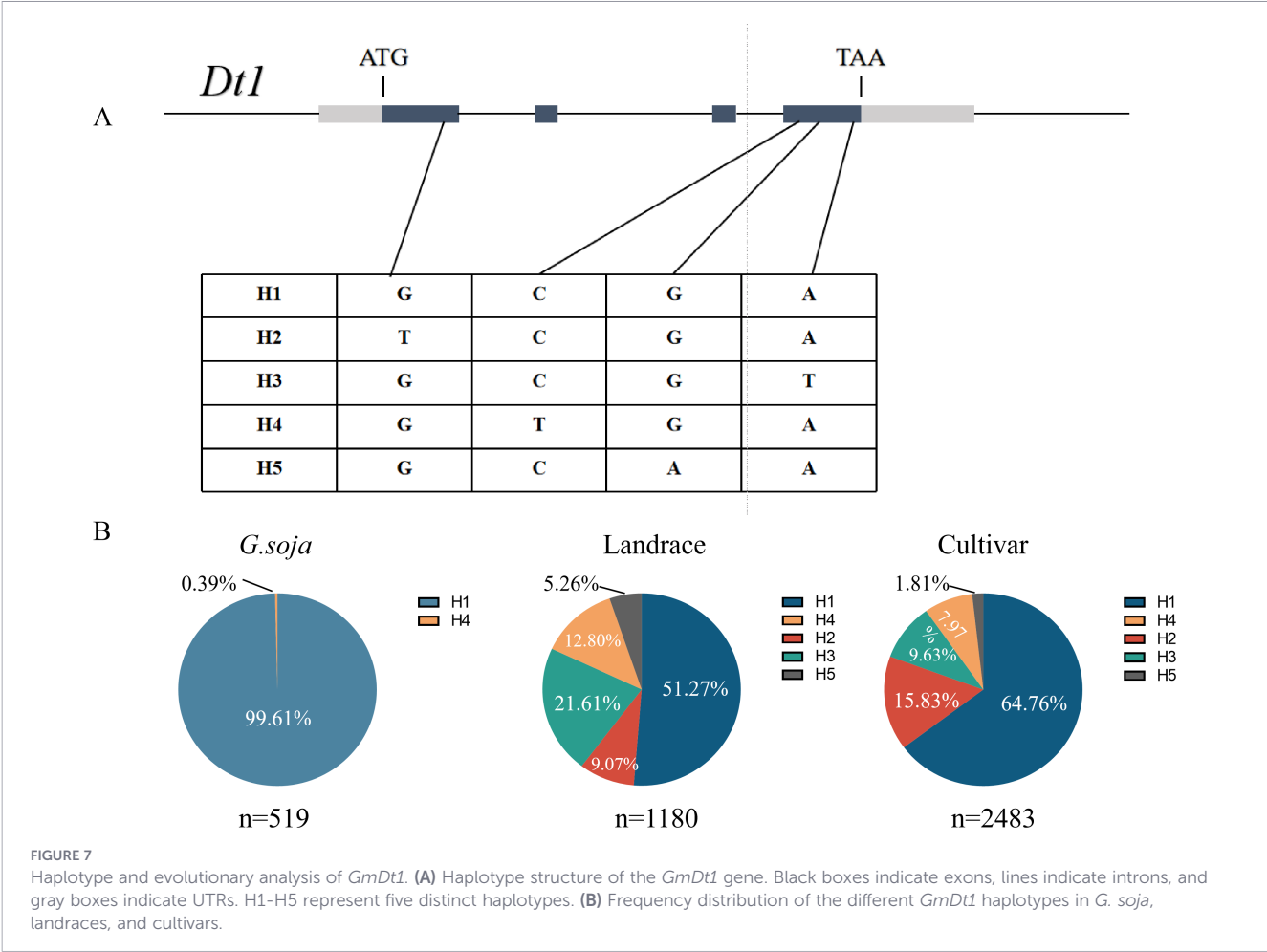


GmDtl is a promising candidate gene for 100-SW regulation. As a well-characterized hub gene in soybean, *GmDtl* plays a crucial role in modulating podding habit and plant height (Liu et al., 2010; Tian et al., 2010). Notably, a recent study by Li Y. et al. (2024) demonstrated that *GmDtl* interacts with *GmSWEET10a* to coordinately modulate soybean seed size. This finding directly corroborates our results and expands our understanding of *GmDtl*'s functional versatility beyond plant architecture control.

To further decipher the molecular mechanism underlying *GmDtl*-mediated regulation of 100-SW, we analyzed its sequence variation and genetic pattern. Four non-synonymous mutations have been previously reported in *GmDtl* (Tian et al., 2010). Our study identified a novel natural variation located in the first exon of *GmDtl*. This variation is distinct from that reported by Li Y. et al. (2024), which was situated in the fourth exon, indicating a new functional locus contributing to soybean 100-SW regulation. In the present study, non-synonymous SNPs in coding regions were prioritized during candidate gene screening, as they are more likely to affect protein function. Variations in non-coding regions warrant further investigation in future studies. Additionally, genetic analysis revealed that *GmDtl* exerts differential effects on 100-SW in dominant homozygous and heterozygous genotypes, displaying an incomplete dominance inheritance pattern. This genetic characteristic is consistent with that of *GmSSS1*, another well-characterized 100-SW regulatory gene in soybean (Zhu et al., 2022). These findings suggest a conserved genetic regulatory mechanism for soybean 100-SW among these genes.

The RHL population used in this study is characterized by large seeds with determinate podding habit and small seeds with a semi-determinate podding habit. Investigation of agronomic traits in several lines from the ZP661 $\times$ ZH13 RIL population revealed that, although the population generally displayed the tendency of large seeds associated with dwarf plants and fewer nodes, and small seeds associated with tall plants and more nodes, some exceptional lines exhibited tall stature, more nodes, and large seeds simultaneously (Supplementary Table 4). Furthermore, although most cultivated soybean varieties carrying the *GmDtl*-H2 haplotype show a determinate podding habit, those with a semi-erect stem growth habit still maintain relatively high 100-SW, such as ZH55 (26 g) and ZH74 (25.7 g) (Supplementary Table 3). Such incomplete linkage among 100-SW, podding habit, and plant height is highly valuable for breeding elite soybean varieties adapted to diverse ecological regions. Collectively, these results verify the key role of *GmDtl* in regulating soybean 100-SW and identify a novel functional variant. This finding lays a solid foundation for further functional characterization and molecular breeding applications.





4.3 The role of *GmDt1* gene in domestication and improvement of soybean 100-SW

Cultivated soybean (*Glycine max*) was domesticated from wild soybean (*Glycine soja*) (Asaf et al., 2017). During domestication, humans preferred larger seeds due to their higher yield potential (Zhu et al., 2022). Almost all wild soybeans harbor the small 100-SW haplotype *GmDt1-H1*. As humans selected varieties, other haplotypes emerged in landraces and improved cultivars. Notably, the high 100-SW haplotype *GmDt1-H2* identified in this study absent in wild soybeans was detected at a frequency of 9.07% in landraces and increased to 15.83% in improved cultivars. Although *GmDt1-H2* has not yet become the dominant haplotype, its gradual enrichment indicates that this elite haplotype will continue to be favored during crop improvement. Similar selection patterns have been observed for elite haplotypes of other seed-related genes, including *GmSW16.1*, *POWR1*, *GmST05*, and *GmCYP82C4* (Duan et al., 2022; Goettel et al., 2022; Chen et al., 2023; Li et al., 2024).

5 Conclusion

This study mapped a core candidate gene *GmDt* controlling 100-SW to chromosome 19 using BSA-seq technology. A G-to-T non-synonymous mutation in the first exon was confirmed as the functional locus driving 100-SW phenotypic variation. This represents a novel functional variant distinct from previously reported mutations in *GmDt1*. Whole-genome resequencing classified *GmDt1* into five distinct haplotypes. Notably, *GmDt1-H2* has undergone gradual artificial selection during the domestication transition from wild to cultivated soybean. Together, these results validate the critical role of *GmDt1* in 100-SW regulation and provide a valuable molecular target for improving soybean seed traits and enhancing yield potential in molecular breeding programs.

Data availability statement

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

Author contributions

GW: Data curation, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. DZ: Data curation, Investigation, Methodology, Visualization, Writing – original draft. XH: Data curation, Investigation, Methodology, Visualization, Writing – review & editing. HZ: Investigation, Validation, Visualization, Writing – review & editing. WW: Investigation, Validation, Visualization, Writing – review & editing. PB: Investigation, Validation, Visualization, Writing – review & editing. LM: Supervision, Validation, Visualization, Writing – review & editing. HG: Supervision, Validation,

Visualization, Writing – review & editing. LW: Supervision, Validation, Visualization, Writing – review & editing. HS: Supervision, Validation, Visualization, Writing – review & editing. JL: Conceptualization, Funding acquisition, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by Key Science and Technology Project of Anhui Province (202423i10050012), the Agriculture Science and Technology Major Project, the Special Fund for Anhui Agriculture Research System.

Conflict of interest

The authors 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.2026.1801920/full#supplementary-material>

SUPPLEMENTARY TABLE 1
BLAST analysis of candidate gene protein sequences.

SUPPLEMENTARY TABLE 2
GmDt1 gene variation loci and phenotypic analysis.

References

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt.2095
- Asaf, S., Khan, A. L., Aaqil Khan, M., Muhammad Imran, Q., Kang, S. M., Al-Hosni, K., et al. (2017). Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. *PLoS One* 12, e0182281. doi: 10.1371/journal.pone.0182281
- Bodirsky, B. L., Rolinski, S., Biewald, A., Weindl, I., Popp, A., and Lotze-Campen, H. (2015). Global food demand scenarios for the 21st century. *PLoS One* 10, e0139201. doi: 10.1371/journal.pone.0139201
- Chen, X., Liu, C., Guo, P., Hao, X., Pan, Y., Zhang, K., et al. (2023). Differential SW16.1 allelic effects and genetic backgrounds contributed to increased seed weight after soybean domestication. *J. Integr. Plant Biol.* 65, 1734–1752. doi: 10.1111/jipb.13480
- 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, 912–919. doi: 10.1007/s001220100621
- Du, J., Wang, S., He, C., Zhou, B., Ruan, Y. L., and Shou, H. (2017). Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J. Exp. Bot.* 68, 1955–1972. doi: 10.1093/jxb/erw460
- Duan, Z., Zhang, M., Zhang, Z., 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
- Ge, L., Yu, J., Wang, H., Luth, D., Bai, G., Wang, K., et al. (2016). Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *PNAS* 113, 12414–12419. doi: 10.1073/pnas.1611763113
- Goettel, W., Zhang, H., Li, Y., Qiao, Z., Jiang, H., Hou, D., et al. (2022). POWR1 is a domestication gene pleiotropically regulating seed quality and yield in soybean. *Nat. Commun.* 13, 3051. doi: 10.1038/s41467-022-30314-7
- Gu, Y., Li, W., Jiang, H., Wang, Y., Gao, H., Liu, M., et al. (2017). Differential expression of a WRKY gene between wild and cultivated soybeans correlates to seed size. *J. Exp. Bot.* 68, 2717–2729. doi: 10.1093/jxb/erw460
- Han, Y., Zhao, X., Liu, D., Li, Y., Lightfoot, D. A., Yang, Z., et al. (2016). Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytol.* 209, 871–884. doi: 10.1111/nph.13626
- 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, 68–74. doi: 10.2135/cropsci2003.6800
- Jeong, N., Suh, S. J., Kim, M. H., Lee, S., Moon, J. K., Kim, H. S., et al. (2012). Ln is a key regulator of leaflet shape and number of seeds per pod in soybean. *Plant Cell* 24, 4807–4818. doi: 10.1105/tpc.112.104968
- 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
- Lee, G. A., Crawford, G. W., Liu, L., Sasaki, Y., and Chen, X. (2011). Archaeological soybean (*Glycine max*) in East Asia: does size matter? *PLoS One* 6, e26720. doi: 10.1371/journal.pone.0026720
- Li, B., Yang, C., Yong, B., Wang, Y., Zhu, W., Gu, Y., et al. (2025). A 14-3-3 modulator of seed weight and quality for unlocking the yield potential of soybean. *Nat. Commun.* 16, 10547. doi: 10.1038/s41467-025-65598-y
- Li, J., Zhao, J., Li, Y., Gao, Y., Hua, S., Nadeem, M., et al. (2019). Identification of a novel seed size associated locus sw9-1 in soybean. *Crop J.* 7, 548–559. doi: 10.1016/j.cj.2018.12.010
- Li, X., Chen, Z., Li, H., Yue, L., Tan, C., Liu, H., et al. (2024). Dtl inhibits SWEET-mediated sucrose transport to regulate photoperiod-dependent seed weight in soybean. *Mol. Plant* 17, 496–508. doi: 10.1016/j.molp.2024.02.007
- Li, X., Zhang, X., Zhu, L., Bu, Y., Wang, X., Zhang, X., et al. (2019). Genome-wide association study of four yield-related traits at the R6 stage in soybean. *BMC Genet.* 20, 39. doi: 10.1186/s12863-019-0737-9
- Li, Y., Zhao, W., Tang, J., Yue, X., Gu, J., Zhao, B., et al. (2024). Identification of the domestication gene GmCYP82C4 underlying the major quantitative trait locus for the seed weight in soybean. *Theor. Appl. Genet.* 137, 62. doi: 10.1007/s00122-024-04571-7
- Li, Y. H., Zhou, G., Ma, J., Jiang, W., Jin, L. G., Zhang, Z., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979
- Liang, S., Duan, Z., He, X., Yang, X., Yuan, Y., Liang, Q., et al. (2024). Natural variation in GmSW17 controls seed size in soybean. *Nat. Commun.* 15, 7417. doi: 10.1038/s41467-024-51798-5
- Liu, Y. F., Li, Q. T., Lu, X., Song, Q. X., Lam, S. M., Zhang, W. K., et al. (2014). Soybean GmMYB73 promotes lipid accumulation in transgenic plants. *BMC Plant Biol.* 14, 73. doi: 10.1186/1471-2229-14-73
- Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., et al. (2010). The soybean stem growth habit gene Dtl is an ortholog of Arabidopsis TERMINAL FLOWER1. *Plant Physiol.* 153, 198–210. doi: 10.1104/pp.109.150607
- Lu, X., Li, Q. T., Xiong, Q., Li, W., Bi, Y. D., Lai, Y. C., et al. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *Plant J.* 86, 530–544. doi: 10.1111/tpj.13181
- Mao, Z. Z., Gong, Y., Shi, G. X., Li, Y. L., Yu, Y., and Huang, F. (2020). Cloning of the soybean E2 ubiquitin-conjugating enzyme GmUBC1 and its expression in Arabidopsis thaliana. *Hereditas* 42, 788–798. doi: 10.16288/j.ycz.20-141
- Mian, M. A., 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, 1011–1016. doi: 10.1007/BF00230118
- Miao, L., Yang, S., Zhang, K., He, J., Wu, C., Ren, Y., et al. (2020). Natural variation and accumulation in GmSWEET39 affect soybean seed oil content. *New Phytol.* 225, 1651–1666. doi: 10.1111/nph.16250
- Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *PNAS* 88, 9828–9832. doi: 10.1073/pnas.88.21.9828
- Orf, J. H., Chase, K., Jarvik, T., Mansur, L. M., Cregan, P. B., Adler, F. R., et al. (1999). Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Sci.* 39, 1642–1651. doi: 10.2135/cropsci1999.3961642x
- Paran, I., and Zamir, D. (2003). Quantitative traits in plants: beyond the QTL. *Trends Genet.* 19, 303–306. doi: 10.1016/S0168-9525(03)00117-3
- Singh, A. K., Fu, D. Q., El-Habbak, M., Navarre, D., Ghabrial, S., and Kachroo, A. (2011). Silencing genes encoding omega-3 fatty acid desaturase alters seed size and accumulation of Bean pod mottle virus in soybean. *Mol. Plant-Microbe Interact.* 24, 506–515. doi: 10.1094/MPMI-09-10-0201
- Stombaugh, S. K., Orf, J. H., Jung, H. G., Chase, K., Lark, K. G., and Somers, D. A. (2004). Quantitative trait loci associated with cell wall polysaccharides in soybean seed. *Crop Sci.* 44, 2101–2106. doi: 10.2135/cropsci2004.2101
- Tang, X., Su, T., Han, M., Wei, L., Wang, W., Yu, Z., et al. (2017). Suppression of extracellular invertase inhibitor gene expression improves seed weight in soybean (*Glycine max*). *J. Exp. Bot.* 68, 469–482. doi: 10.1093/jxb/erw425
- Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J. E., Nelson, R. L., et al. (2010). Artificial selection for determinate growth habit in soybean. *PNAS* 107, 8563–8568. doi: 10.1073/pnas.1000088107
- Wang, J., Chen, P., Wang, D., Shannon, G., Shi, A., Zeng, A., et al. (2015). Identification of quantitative trait loci for oil content in soybean seed. *Crop Sci.* 55, 23–34. doi: 10.2135/cropsci2014.04.0280
- Wang, X., Han, L., Li, J., Shang, X., Liu, Q., Li, L., et al. (2023). Next-generation bulked segregant analysis for Breeding 4.0. *Cell Rep.* 42, 113039. doi: 10.1016/j.celrep.2023.113039
- Wang, X., Li, Y., Zhang, H., Sun, G., Zhang, W., and Qiu, L. (2015). Evolution and association analysis of GmCYP78A10 gene with seed size/weight and pod number in soybean. *Mol. Biol. Rep.* 42, 489–496. doi: 10.1007/s11033-014-3792-3
- Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y. C., et al. (2020). Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. *Natl. Sci. Rev.* 7, 1776–1786. doi: 10.1093/nsr/nwaa110
- Xu, M., Kong, K., Miao, L., He, J., Liu, T., Zhang, K., et al. (2023). Identification of major quantitative trait loci and candidate genes for seed weight in soybean. *Theor. Appl. Genet.* 136, 22. doi: 10.1007/s00122-023-04299-w
- 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, 529. doi: 10.1186/s12864-017-3922-0
- Yang, Z., Huang, D., Tang, W., Zheng, Y., Liang, K., Cutler, A. J., et al. (2013). Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. *PLoS One* 8, e68433. doi: 10.1371/journal.pone.0068433
- Yang, H., Xue, Q., Zhang, Z., Du, J., Yu, D., and Huang, F. (2018). GmMYB181, a Soybean R2R3-MYB protein, increases branch number in transgenic Arabidopsis. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01027
- Yu, C., Qu, Z., Zhang, Y., Zhang, X., Lan, T., Adelson, D. L., et al. (2017). Seed weight differences between wild and domesticated soybeans are associated with specific changes in gene expression. *Plant Cell Rep.* 36, 1417–1426. doi: 10.1007/s00299-017-2165-5
- Zhang, C., Li, W., Tan, C., Huang, M., Wu, H., Liu, S., et al. (2025). Natural allelic variation in SW14 determines seed weight and quality in soybean. *Nat. Commun.* 16, 8070. doi: 10.1038/s41467-025-63582-0

Zhang, J., Song, Q., 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

Zhao, B., Dai, A., Wei, H., Yang, S., Wang, B., Jiang, N., et al. (2016). Arabidopsis KLU homologue GmCYP78A72 regulates seed size in soybean. *Plant Mol. Biol.* 90, 33–47. doi: 10.1007/s11103-015-0392-0

Zhao, X., Dong, H., Chang, H., Zhao, J., Teng, W., Qiu, L., et al. (2019). Genome wide association mapping and candidate gene analysis for hundred seed weight in soybean [*Glycine max* (L.) Merr. *BMC Genomics* 20, 648. doi: 10.1186/s12864-019-6009-2

Zhu, W., Yang, C., Yong, B., Wang, Y., Li, B., Gu, Y., et al. (2022). An enhancing effect attributed to a nonsynonymous mutation in SOYBEAN SEED SIZE 1, a SPINDLY-like gene, is exploited in soybean domestication and improvement. *New Phytol.* 236, 1375–1392. doi: 10.1111/nph.18461