Received: 2 February 2023 | Accepted: 3 August 2023 | Published online: 23 August 2023

DOI: 10.1002/csc2.21080

Crop Science

ORIGINAL ARTICLE

Crops for Nutrition & Health

Identification of quantitative trait loci and candidate genes for seed sucrose and soluble sugar concentrations in soybean

Changkai Liu[1] | **Heng Chen**[1] | **Qintan Yu**[1,2] | **Haidong Gu**[1,2] | **Yansheng Li**[1] | **Bingjie Tu**[3] | **Hengyou Zhang**[1] | **Qiuying Zhang**[1] | **Xiaobing Liu**[1]

[1]Key Laboratory of Soybean Molecular Design and Breeding, Northeast Institute of Geography and Agroecology, CAS, Harbin, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]Harbin University, Harbin, China

Correspondence
Qiuying Zhang, Key Laboratory of Soybean Molecular Design and Breeding, Northeast Institute of Geography and Agroecology, CAS, Haping Road 138, Harbin, Heilongjiang, 150081, China.
Email: zhangqiuying@iga.ac.cn

Assigned to Associate Editor Jason Gillman

Funding information
Strategic Priority Research Program of Chinese Academy of Sciences, Grant/Award Number: XDA28070402; National Key R&D Program of China, Grant/Award Number: 2021YFD1201103-03; Project for high-tech industrialization of science and technology between Jilin province and Chinese Academy of Sciences, Grant/Award Number: 2022SYHZ0039

Abstract

This study aimed to investigate the genetics of sucrose and soluble sugar accumulation in soybean [*Glycine max* (L.) Merrill.] and the impact on food quality. A recombinant inbred line population, constructed from ZD27 and HF25 with parents having significantly different levels of sucrose and soluble sugar content, was used. Using the QTL-Sequencing method, we identified 16 QTLs affecting sucrose and soluble sugar, including a major sucrose-associated QTL, *qSU1901*, with 10.6%–13.2% phenotypic variation for sucrose. To narrow down the list of candidate genes, we analyzed the polymorphisms within the major QTL interval. We further explored the temporal expression patterns of key genes involved in sucrose and soluble sugar transport and metabolism through dynamic transcriptome analysis. We generated a gene regulatory network that correlated with sucrose and soluble sugar content using weighted gene co-expression network analysis. By combining QTL-Seq and RNA-Sequencing data, we narrowed down 233 candidate genes. In conclusion, our findings provide a better understanding of sucrose and soluble sugar regulation in soybean.

1 | INTRODUCTION

Soy-based food products, such as tofu, soymilk, tempeh and natto, have been gaining immense popularity in recent times due to the growing demand for plant-based diets (Cai et al., 2021). The quality of these products is largely determined by the soluble carbohydrates that constitute around 10% of the dry weight of soybeans (Maughan et al., 2000). These soluble sugars, including glucose, fructose, sucrose, raffinose, and stachyose, are a heritable trait found in soybean seeds, with sucrose being the most abundant and

Abbreviations: ADD, additive effect; BLUP, best linear unbiased prediction; bZIP, basic leucine zipper; DEGs, differentially expressed genes; GO, gene ontology; GPI, glucose-6-phosphate isomerase; GRN, gene regulatory network; HK, hexokinase; ICIM, inclusive composite interval mapping; Indel, insertions and deletions; INV, invertase; KEGG, Kyoto Encyclopedia of Genes and Genomes; LOD, likelihood of odds; MYB, MYB DNA-binding domain-containing protein; PVE, phenotypic variation explained; QTL, quantitative trait locus; QTNs, quantitative trait nucleotides; RIL, recombinant inbred line; RNA-Seq, RNA sequencing; RS, raffinose synthase; ScrK, fructokinase; SNP, single nucleotide polymorphism; SPS, sucrose phosphate synthase; SSR, simple sequence repeat; SUS, sucrose synthase; SUT, sucrose transporter; SWEET, seed-expressed water-soluble sugar transport protein; TFs, transcription factors; UGP2, UTP-glucose-1-phosphate uridylyltransferase; WGCNA, weighted gene co-expression network analysis.

contributing to the sweetness of soybean products (Hu et al., 2023; Lee et al., 2016; Xu et al., 2022). Fermentation of raffinose and stachyose in soybeans by gut microbes can lead to flatulence and discomfort. Thus, it is desirable to minimize raffinose and stachyose levels while maximizing sucrose accumulation.

Given the importance of soluble sugar and sucrose in determining the quality of soy-based foods and feeds, soybean breeding programs have been focused on developing varieties with a high soluble sugar or sucrose content (Ficht et al., 2022; Lu et al., 2022; Pan et al., 2022; Sui et al., 2020; Xu et al., 2022) to meet the growing demand for high-grade soybeans. As such, soybean varieties with high soluble sugar content are increasingly valuable for soyfood production and soybean meal processing (Lee et al., 2016; Li et al., 2012).

To speed up the breeding process of high-sugar soybeans, molecular marker technology has been proposed as a viable approach (Maughan et al., 2000). Early studies have identified several sucrose-specific marker loci and quantitative trait loci (QTLs) associated with soluble sugar components using a range of markers including restriction fragment length polymorphism, simple sequence repeat (SSR), and random amplified polymorphic DNA. For example, Maughan et al. (2000) used an F_2 population with 178 polymorphic markers and identified 17 sucrose-specific marker loci that were mapped to seven different genomic regions. Zeng et al. (2014) identified eight sucrose QTLs on five chromosomes using SSR and single nucleotide polymorphism (SNP) markers. Wang et al. (2014) preliminarily identified 11 QTLs associated with five soluble sugar components using an F_2 population derived from soybean varieties V97-3000 and V99-5089. Although these early findings provided a foundation for subsequent gene localization, the localization interval and accuracy were limited due to the restrictions of earlier research methods.

Recent advancements in technology and methods, such as high-throughput sequencing technology, have significantly improved the accuracy and shortened the localization interval for gene localization. Lee et al. (2016) used Axiom 180 K SoyaSNP genotyping technology and a recombinant inbred line (RIL) population mapped five QTLs associated with sucrose and oligosaccharide on four chromosomes. Pan et al. (2022) identified three candidate genes associated with soybean soluble sugar content by using re-sequencing and meta-overview-collinearity integrated analysis methods on a population of chromosome segment substitution lines. Lu et al. (2022) used genome-wide association study and transcriptome sequencing methods and identified 13 quantitative trait nucleotides (QTNs) and several candidate genes for total soluble sugar content in soybean seeds. Similarly, Hu et al. (2023) identified 72 QTNs associated with fructose, glucose, sucrose raffinose, stachyose, and 14 QTNs with soluble sugar. By using a high-quality de novo genome assembly, Liu et al. (2022) reported four sucrose synthase (*SUS*) genes, one sucrose-phosphate synthase gene, and four sugar transport genes as candidate genes related to sucrose or soluble sugar accumulation in vegetable soybean.

Core Ideas

- Sixteen quantitative trait loci (QTLs) were found for sucrose and sugar levels in soybean, including a major sucrose-associated QTL qSU1901.
- A total of 233 candidate genes narrowed down in QTL interval, enlightening sugar transport and metabolism.
- Gene regulatory network generated to understand gene expression in sucrose and sugar regulation.

Sucrose is a major transport molecule for photosynthetic products in higher plant (Li et al., 2021). Several candidate genes involved in sucrose metabolism and accumulation, including genes related to *SUS*, sucrose phosphate synthase (*SPS*), sucrose phosphate phosphatase, invertase (*INV*), seed-expressed water-soluble sugar transport protein (*SWEET*) family, and sucrose transporter (*SUT*) family genes, have been reported in soybean (Li et al., 2021; Wang et al., 2019) and other plants (Chen et al., 2015; Zhen et al., 2018). The raffinose synthase (RS), which belongs to the hydrolase family, has also been shown to affect sucrose levels in soybean (Dierking & Bilyeu, 2008). For instance, soybean RS2, encoded by *Glyma06g18890*, catalyzes the reaction: sucrose + galactinol → raffinose + myo-inositol (Kumar et al., 2010). Additionally, several transcription factors (TFs), including WRKY DNA-binding protein (Li et al., 2020), MYB DNA-binding domain-containing protein (MYB) (Wei et al., 2018), and basic leucine zipper (bZIP) (Chen et al., 2020), have been implicated in regulating sucrose accumulation. Overall, multiple processes contribute to the accumulation of sucrose in seeds. However, compared to other soybean nutritional traits such as protein and oil, research on the genetic regulation of soluble sugar and sucrose content is still in its early stages (Pan et al., 2022).

RNA sequencing (RNA-Seq) technology is an effective tool in dissecting the regulatory network of sucrose metabolism during seed development. By combining RNA-Seq analysis with QTL mapping, candidate genes and their expression patterns can be swiftly identified and located, leading to a significant acceleration of the target gene mapping process (Guo et al., 2021; Lu et al., 2022; Perlo et al., 2022). Therefore, utilizing precise localization techniques in conjunction with multi-omics analyses to construct suitable mapping populations is a desirable approach for identifying candidate genes responsible for soluble sugar and sucrose

content (Guo et al., 2021; Lu et al., 2022; Zhang et al., 2021). In this study, whole genome resequencing technology was employed to locate QTLs in a more stable RIL mapping population established by Chen et al. (2021), which involved crossing the high sucrose and soluble sugar soybean variety ZD27 with the low sucrose and soluble sugar variety HF25. Additionally, transcriptome expression differences between ZD27 and HF25 during seed filling were also examined. The QTLs and candidate genes identified in this study will facilitate the design of molecular modules and breeding lines/varieties with high sugar levels in soybean.

2 | MATERIALS AND METHODS

2.1 | Plant materials and sample collection

The plant materials used in this study comprised 158 F_7 RILs that were generated from a cross between "Hefeng 25" (HF25) and "Zhongdou 27" (ZD27). Compared to HF25, ZD27 has a ripening period that is 7 days later and higher total soluble sugar and sucrose content. In 2020, the RILs were grown in the fields of three different locations, Harbin (45°N 127°E, active accumulated temperature >2700°C), Mudanjiang (44°N 129°E, active accumulated temperature 2500°C–2700°C), and Hailun (47°N 126°E, active accumulated temperature 2300°C–2500°C). Each field experiment was carried out using a completely randomized design with three replicates. While both parents and the RIL population reached full maturity at the Harbin and Mudanjiang sites, only 128 mature lines were selected for sucrose and soluble sugar detection at the Hailun experimental site because 30 lines did not reach full maturity.

Furthermore, in 2021 HF25 and ZD27 were planted in Harbin using the same experimental design for transcriptome detection.

Seedling leaves from the RIL population were collected to extract DNA for resequencing, while mature seeds were harvested to measure sucrose and soluble sugar content. Seed samples for RNA-Seq testing were collected at 2, 3, 4, 5, 6, 7, and 8 weeks after beginning seed stage (R5) using the classification by Fehr and Caviness (1977).

2.2 | Phenotypic data measurement

The anthrone sulfuric acid method is a quick and efficient method for the analyzing sucrose and soluble sugar concentrations in soybean and other plants and has been widely used (Liu et al., 2017; Morse, 1947; Tang et al., 2021). Therefore, this method was employed to determine the concentrations of soluble sugar and sucrose in this study (Tu et al., 2017). Samples weighing over 0.5 g were placed in a 10-mL centrifuge tube, and then extracted with 6 mL 80% ethanol for 40 min at 80°C in a water bath. The homogenate was vortexed every 10 min. The mixture was then centrifuged at 4500× *g* for 3 min, and the supernatant was removed. The residue was extracted consecutively for three times. The supernatant was removed to a 50-mL volumetric flask with 80% ethanol to dilute to the 50 mL volume. For total soluble sugar determination, 4 mL of anthrone (1000 mL 80% H_2SO_4 + 2.5 g anthrone) was added to 1 mL supernatant, bathed in 90°C water for 10 min, then measured at 620 nm with a spectrophotometer (Xinshiji T6) after cooling. The soluble sugar concentration was calculated with glucose as standard. For sucrose determination, 100 μL 3 mol L^{-1} of NaOH was added to 1 mL supernatant and water bathed at 100°C for 10 min in order to remove the monosaccharides. After cooling, 4 mL anthrone was added and bathed in water at 40°C for 10 min, measured at 620 nm. Sucrose was used as the calculation standard.

2.3 | High-density genetic map construction

The high-density genetic map used in this study was previously constructed by Chen et al. (2021) using the whole genome resequencing method. For SNP marker analysis, DNA was extracted from five young leaves from each line. Among the parents, a total of 1,283,813 SNPs were identified, out of which 1,093,273 SNPs were found in the RILs. Using 5338 bin markers based on the SNP data, a high-density genetic map extending over a total length of 2487.17 cM was constructed with an average distance of 0.47 cM. The Wm82.a2.v1 reference genome (Schmutz et al., 2010) comprising 995.708 Mbp of annotated sequences was used in this study (Table S1).

2.4 | Quantitative trait loci identification

QTL mapping was constructed using inclusive composite interval mapping (ICIM) in IciMapping V4.2 (http://www.isbreeding.net/software/) (Luo et al., 2021). The pre-adjusted mapping parameters were as follows: step = 1.0 cM, Probability of Inclusion (PIN) = 0.001 and likelihood of odds (LOD) ≥ 2.5 with 1000-permutation testing. The QTLs were named as "*q*" plus trait name along with the chromosome information at the end.

The best linear unbiased prediction (BLUP) method applied in this study reduced the impact of environmental differences (Bernardo, 1996).

The reference QTL data were provided by the website www.soybase.org.

2.5 | Statistical analyses of phenotypic data

Variance analysis of phenotypic data was performed using SPSS 12.0 with environment, replication within environment, genotype, and interaction between genotype and environment as random effects. The broad-sense heritability (h^2) of every trait was estimated as $h^2 = \sigma_g{}^2/(\sigma_g{}^2 + \sigma_{ge}{}^2/n + \sigma^2/nr)$, where $\sigma_g{}^2$ is the genotypic variance, $\sigma_{ge}{}^2$ is the interaction between genotype and environmental variance, σ^2 is the error variance, n is the number of environments, and r is the number of replicates (Nyquist & Baker, 1991).

The distribution features and correlations of sucrose and soluble sugar were analyzed using R package "ggplot2" (Wickham, 2016).

2.6 | Prediction of candidate genes and expression analysis

Linkage mapping that identified QTLs whose physical positions led to the prediction of candidate genes was based on the annotation of Glyma.Wm82.a2.v1 on Phytozome (https://phytozome.jgi.doe.gov/). To predict the functions of the candidate genes, Basic Local Alignment Search Tool for Proteins (BLASTP) was used to search for proteins in *Arabidopsis thaliana* with the amino acid sequences of the candidate genes.

2.7 | cDNA library construction and RNA-Seq

A total of 28 RNA samples were used to construct the cDNA library in Shanghai Majorbio Biopharm Technology Co., Ltd. Illumina HiSeq 4000 was used to sequence each cDNA library. The data were analyzed on the online platform of Majorbio Could platform (www.majorbio.com), including differentially expressed genes (DEGs) selection, gene function annotation, transcription factor prediction, and weighted gene co-expression network analysis (WGCNA) analysis, and so forth (Ren et al., 2022).

2.8 | Gene expression verification by quantitative real-time PCR (qRT-PCR)

From the RNA-Seq data, 11 DEGs were selected for qRT-PCR verification. Total RNA (0.3 μg) was extracted with a Plant RNA Extraction Kit (TIANGEN), and the first-strand cDNA templates were synthesized with a cDNA synthesis kit (R223-01; Vazyme). qRT-PCR (ChamQ SYBR qPCR Master Mix (Q311-02/03; Vazyme) was conducted with the Light Cycler 480 instrument (Roche) through denaturation at 95°C for 10 min, followed by 40 cycles at 95°C for 10 s and at 60°C for 15 s, 72°C 20 s (Liu et al., 2020). The *GmActin4* gene (*Glyma.12G020500*) was selected as an internal control (Pan et al., 2022), and the relative expression level was calculated as the $2^{-\Delta\Delta Ct}$ value (Pfaffl, 2001). The primer sequences are listed in Table S2.

3 | RESULTS

3.1 | Phenotypic variation and correlations among soluble sugar and sucrose in the ZH population

To evaluate the phenotypic variation of the ZD27 × HF25 (ZH) population, soluble sugar, and sucrose concentrations were investigated in three environments (Harbin, Mudanjiang, and Hailun) as well as in a combined BLUP model. Table 1 shows that the high-sugar parent ZD27 had concentrations of sucrose and soluble sugar ranging from 81.0 to 100.0 mg kg^{-1} and 165.2 to 179.7 mg kg^{-1}, respectively, while the low-sugar parent HF25 had concentrations ranging from 68.2 to 83.4 mg kg^{-1} and 133.0 to 168.2 mg kg^{-1}, respectively. Among RILs, the minimum concentration of sucrose and soluble sugar ranged from 52.5 to 75 mg kg^{-1} and 95.3 to 141.1 mg kg^{-1}, respectively, while, the corresponding maximum value ranged from 107.3 to 127.5 mg kg^{-1} and 223.1 to 249.5 mg kg^{-1}, respectively. The coefficient of variation (CV) for sucrose ranged from 4.7% to 19.4%, while for soluble sugar, it ranged from 4.7% to 19.8%. The analysis of variance in sucrose and soluble sugar concentrations demonstrated significant genotypic differences among the RILs as well as among different environments. However, there was no significant interaction between genotype and environment for both sucrose and soluble sugar concentrations (Table 1). The heritability (*H*2) of sucrose and soluble sugar was determined to be 58.5% and 54.6%, respectively.

Sucrose and soluble sugar concentrations of the ZH population followed a normal distribution (Table 1) and were significantly correlated (0.731; Table 2). Furthermore, the mean values of soluble sugar and sucrose concentration showed a decreasing trend with an increase in accumulated active accumulated temperature in the following order: HL (2300°C–2500°C) > M (2500°C–2700°C) > H (>2700°C) (Table 1).

3.2 | Identification of QTLs for soluble sugar and sucrose in the RIL population

Using the ICIM mapping method, a total of 16 QTLs were identified as significantly related to sucrose and soluble sugar concentrations in soybean seed (Table 3). These QTLs were distributed on 12 different chromosomes.

TABLE 1 Means ranges and parental values for sucrose and soluble sugar concentrations in soybeans grown at three locations.

		Parents		RIL population					
Traits	**Loc**	**ZD27**	**HF25**	**Minimum**	**Maximum**	**Mean**	**CV (%)**	**Skewness**	**Kurtosis**
Sucrose	H	81.0 ± 4.7[a]	71.9 ± 3.6	52.5	112.0	79.8	19.4	−0.035	−0.172
	HL	100.0 ± 8.1	83.4 ± 7.5	75.0	127.5	99.6	14.2	0.226	0.298
	M	80.8 ± 9.1	68.2 ± 9.0	58.6	107.3	84.5	16.8	0.121	0.036
	BLUP	86.8	74.9	76.7	101.1	87.9	4.7	0.139	0.380
Soluble sugar	H	165.2 ± 12.2	144.6 ± 5.4	95.3	223.1	145.1	19.8	0.246	0.436
	HL	179.7 ± 3.2	168.2 ± 2.5	141.1	249.5	202.3	15.3	−0.294	−0.131
	M	158.8 ± 9.9	133.0 ± 4.9	121.6	245.5	189.8	18.4	0.109	0.026
	BLUP	167.4	149.0	157.5	198.7	179.0	4.7	−0.103	−0.291

Abbreviations: BLUP, best linear unbiased prediction; CV, coefficient of variation; H, Harbin; HL, Hailum; Loc, locations; M, Mudanjiang; RIL, recombinant inbred line.
[a]Mean ± standard deviation.

TABLE 2 Analysis of variance for sucrose and soluble sugar concentrations, heritability coefficient (H^2), and correlation between sucrose and soluble sugar.

Trait	**Source**	***df***	**MS**	***p* value**	**H^2 (%)**	**Correlation**
Sucrose	Genotype	157	452.4	***	58.5	0.731***
	Environment	2	43264.8	***		
	G × E	275	207.6	ns		
	Error	870	229.5			
Soluble sugar	Genotype	157	2113.4	***	54.6	
	Environment	2	371523.8	***		
	G × E	275	1151.9	ns		
	Error	870	1067.1			

Abbreviations: E, environment; G, genotype; MS, mean square; ns, not significant.
*** and * represent signifcant at the 0.001 and 0.05 probability levels, respectively.

A total of eight QTLs were detected for sucrose, which were mapped to chromosomes 6, 7, 9, 13, 17, 19, and 20. Among them, *qSU1901* (Chr19: 45,311,975–45,464,136) was detected in Harbin, Mudanjiang, and the BLUP conditions, explaining 10.63%–13.18% of the phenotypic variation with an LOD value of 3.62–6.29 and additive effect (ADD) 1.53–3.94. *qSU2002* (Chr20: 40,523,599–41,882,459) was detected in Mudanjiang and BLUP conditions, explaining 10.26%–14.44% of the phenotypic variation with an LOD value of 3.50–7.90 and an ADD of 1.71–2.85. The remaining six QTLs were exclusively detected under one condition. Notably, one QTL, *qSU0701*, had been previously identified QTL by Maughan et al. (2000). The QTL with the highest phenotypic variation explained (PVE) was *qSU1701* (4,517,109–7,005,804) at 16.81%, with an LOD of 7.61 and an ADD of −4.60. This QTL was only detected in Hailun.

Eight QTLs located on seven chromosomes were associated with soluble sugar concentration, explaining 6.26%–12.31% of the trait variation. Among them, *qSS0201* was a previously identified QTL that was detected by Kim et al. (2005). *qSS0301* (Chr3: 6,004,297–6,444,222) was detected in both Mudanjiang and BLUP conditions, explaining 9.43%–11.88% PVE with an LOD value of 3.51–5.20. The ADD of *qSS0301* ranged from −6.92 to −2.82, while *qSS1701* (Chr17: 7,470,395–10,014,816) and *qSS1702* (Chr17: 7,969,537–10,599,548) were overlapping QTLs on chromosome 17. They explained 8.74%–12.31% PVE with an LOD value of 3.95–5.07. The ADD of *qSS1701* and *qSS1702* were negative. The physical location of *qSS0601* (Chr6: 43709851–44392137) on chromosome 6 was completely consistent with that of *qSU0601*, so this location might affect both sucrose and soluble sugar concentrations simultaneously. The PVE of *qSS0601* was 11.57% with an LOD value of 4.93 and ADD of −8.03.

3.3 | Candidate genes among the QTLs

Based on the gene annotations of Glyma.Wm82. a2.v1 on the Phytozome website (https://phytozome.jgi.doe.gov/), 896

TABLE 3 QTL mapping for sucrose and soluble sugar concentrations in the recombinant inbred line (RIL) population.

Trait	QTL	Chr.	Genetic (cM)	Physical	LOD	PVE%	ADD	Number of genes	Location
Sucrose	*qSU0601*	6	82.11–83.08	43709851–44392137	4.54	9.14	−3.36	21	HL
	qSU0701[a]	7	9.67–10.30	414946–581698	3.39	7.58	2.94	18	H
	qSU0901	9	82.59–87.22	22562116–23609852	2.97	6.69	5.16	5	H
	qSU1301	13	116.78–117.10	38932430–39299827	3.23	6.34	−2.94	38	HL
	qSU1701	17	0.318–6.57	4517109–7005804	7.61	16.81	−4.60	278	HL
	qSU1901	19	67.89–68.53	45311975–45464136	5.22	13.18	3.94	13	H
	qSU1901	19	67.89–68.53	45311975–45464136	3.62	10.63	2.84	13	M
	qSU1901	19	67.89–68.53	45311975–45464136	6.29	11.94	1.53	13	BLUP
	qSU2001	20	4.76–6.73	701169–807404	3.38	5.68	3.50	14	BLUP
	qSU2002	20	80.71–81.35	40523599–41882459	3.50	10.26	2.85	125	M
	qSU2002	20	80.71–81.35	40523599–41882459	7.90	14.44	1.71	125	BLUP
Soluble sugar	*qSS0201*[b]	2	99.62–105.28	41077675–41427601	3.14	8.40	9.70	34	HL
	qSS0301	3	17.55–20.28	6004297–6444222	3.51	9.43	−6.92	11	M
	qSS0301	3	17.55–20.28	6004297–6444222	5.20	11.88	−2.82	11	BLUP
	qSS0601	6	82.11–83.08	43709851–44392137	4.93	11.57	−8.03	21	HL
	qSS0901	9	111.16–115.41	27628060–29278485	2.76	7.22	12.66	28	M
	qSS1301	13	58.74–59.38	25129596–25259816	2.81	6.26	−2.05	10	BLUP
	qSS1701	17	12.18–13.48	7470395–10014816	3.95	8.74	−2.47	276	BLUP
	qSS1702	17	13.48–18.35	7969537–10599548	5.07	12.31	−8.50	282	HL
	qSS1801	18	26.17–27.47	48132212–48604215	3.28	9.25	6.54	19	H

Abbreviations: ADD, additive effect; BLUP, best linear unbiased prediction; Chr., chromosome; H, Harbin; HL, Hailum; LOD, likelihood of odds; M, Mudanjiang; PVE, phenotypic variation explained; QTL, quantitative trait locus.

[a]This QTL was detected by Maunghan et al. (2000) with 3-LOD value interval.

[b]This QTL was detected by Kim et al. (2005) with 2-LOD value interval.

genes, including 59 transcript factors, were identified across 16 QTLs (Table S3). Among them, 17 genes derived from five QTLs were found to be related to sucrose or soluble sugar metabolism or transport (Table 4). Additional polymorphism analysis was carried out on these 17 genes (Table S4). Three genes from the *qSU1701* interval, namely xylose isomerase (*Glyma.17G066200*), probable alpha, alpha-trehalose-phosphate synthase [UDP-forming] 11 (*Glyma.17G067800*), and UDP-glucuronate 4-epimerase 1 (*Glyma.17G069600*), exhibited SNP and insertions and deletions (Indel) variations in the exon region. Additionally, a UDP-glucose 6-dehydrogenase 1 gene (*Glyma.17G077000*) also displayed Indel variation in the exon region (Table 4). A probable UDP-N-acetylglucosamine gene (*Glyma.19G196000*) was detected in *qSU1901*, which exhibited SNP and Indel variations in the exon region. In addition, a probable alkaline/neutral invertase D gene (*Glyma.20G177200*) and a sucrose transport protein SUC5 gene (*Glyma.20G174100*) from the *qSU2002* interval showed SNP and Indel variations, with a phosphoglucomutase (alpha-D-glucose-1,6-bisphosphate-dependent) gene (*Glyma.20G179900*) exhibiting SNP variation in the exon region. Furthermore, a GDP-fucose transporter 1 gene (*Glyma.17G119300*) from the *qSS1701*/*qSS1702* intervals displayed SNP variation in upstream, while a triose-phosphate transporter family protein gene (*Glyma.17G115200*) displayed SNP and Indel variations in the exon region (Table S4).

As *qSU1901* was detected in multiple conditions, we further investigated the genes located within this QTL interval. There are a total of 14 genes involved in this interval (Table S5), which comprise a UDP-N-acetylglucosamine gene (*Glyma.19G196000*), a mRNA-decapping enzyme-like protein gene (*Glyma.19G196300*), a B3 domain-containing protein gene (*Glyma.19G196600*), a lysosomal Pro-X carboxypeptidase gene (*Glyma.19G196700*), a TLD domain-containing protein 1 (*Glyma.19G197300*), three secoisolariciresinol dehydrogenase genes (*Glyma.19G197000*, *Glyma.19G197100*, and *Glyma.19G197200*), two momilactone A synthase genes (*Glyma.19G196800* and *Glyma.19G196900*), and four genes with unknown functions (*Glyma.19G196100*, *Glyma.19G196200*, *Glyma.19G196400*, and *Glyma.19G196500*). After conducting polymorphism analysis, it was found that multiple SNP and Indel were found within the genes (Table S5). Notably, *Glyma.19G196000* gene contained 93 SNPs, of which four were non-synonymous coding mutations located in the exon region. Similarly, a

TABLE 4 Gene annotations related to sucrose or soluble sugar metabolism or transport in QTL intervals detected.

QTL	Chr.	Env.	Candidate genes	Polymorphic changes	Annotations
qSU1701	17	HL	*Glyma.17G061300*	–	putative glucose-6-phosphate 1-epimerase-like [Glycine max]
			Glyma.17G066200	SNP/Indel	xylose isomerase [Glycine max]
			Glyma.17G077000	Indel	UDP-glucose 6-dehydrogenase 1 [Glycine max]
			Glyma.17G081400	–	probable galacturonosyltransferase 11 [Glycine max]
			Glyma.17G067800	SNP/Indel	probable alpha,alpha-trehalose-phosphate synthase [UDP-forming] 11 [Glycine max]
			Glyma.17G069600	SNP/Indel	UDP-glucuronate 4-epimerase 1 [Glycine max]
	19	H/M/BLUP	*Glyma.19G196000*	SNP/Indel	probable UDP-N-acetylglucosamine [Glycine max]
qSU2002	20	M/BLUP	*Glyma.20G177200*	SNP/Indel	probable alkaline/neutral invertase D [Glycine max]
			Glyma.20G174200	–	Sucrose transport protein SUC5 [Glycine soja]
			Glyma.20G174100	SNP/Indel	Sucrose transport protein SUC5 [Glycine soja]
			Glyma.20G179900	SNP	Phosphoglucomutase (alpha-D-glucose-1,6-bisphosphate-dependent)
qSS0201	2	HL	*Glyma.02G225000*	–	uncharacterized protein LOC100799660 isoform X1 [Glycine max];Glycosyl hydrolase family 3 N terminal domain
qSS1701/qSS1702	17	HL/BLUP	*Glyma.17G109700*	–	probable sucrose-phosphate synthase [Glycine max]
			Glyma.17G098400	–	aldose 1-epimerase [Glycine max]
			Glyma.17G119300	SNP	GDP-fucose transporter 1 [Glycine max]
			Glyma.17G111400	–	probable galactinol–sucrose galactosyltransferase 6 isoform X1 [Glycine max]; Raffinose synthase or seed imbibition protein Sip1
			Glyma.17G115200	SNP/Indel	triose-phosphate transporter family protein [Glycine max]

Abbreviations: BLUP, best linear unbiased prediction; Chr., chromosome; Env., environment; H, Harbin; HL, Hailum; Indel, insertions and deletions; M, Mudanjiang; PVE, phenotypic.
variation explained; QTL, quantitative trait locus; SNP, single nucleotide polymorphism.

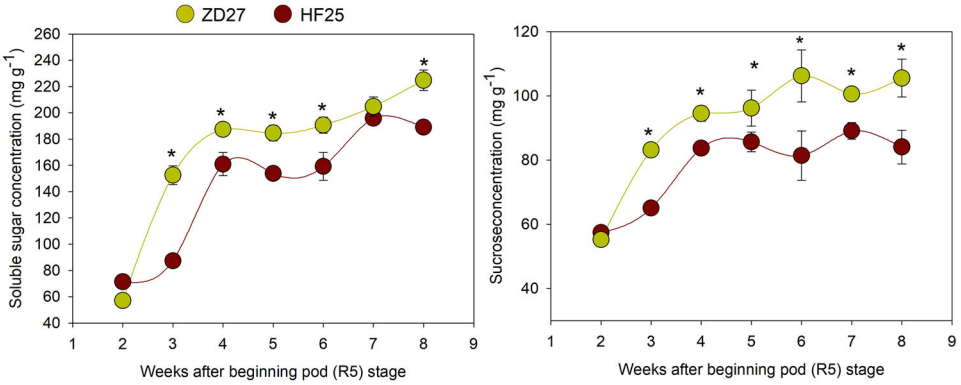

FIGURE 1 Sucrose and soluble sugar accumulation in ZD27 and HF25 during seed development. *significant at $p < 0.05$ level; bars mean SD ($n = 3$).

non-synonymous coding mutation and a codon deletion were detected in the exon region of the *Glyma.19G196600* gene (Table S5).

3.4 | Sucrose and soluble sugar concentrations during seed development in "Zhongdou27" and "Hefeng25"

As depicted in Figure 1, ZD27 consistently exhibited higher sucrose and soluble sugar concentrations from 3 to 8 weeks (S3 to S8) after beginning seed stage (R5). On average, the sucrose and soluble sugar concentrations of ZD27 were 17.4% and 18.1% higher than those of HF25. Both ZD27 and HF 25 showed a trend of oscillation in sucrose and soluble sugar concentrations, rising from S2 to S8. The period from S2 to S4 corresponded to a phase of rapid increase in sucrose and soluble sugar concentrations.

3.5 | RNA-Seq analysis and differentially expressed genes during seed development between ZD27 and HF25

Comparisons of gene expression based on correlations and principal component analysis showed significant differences between ZD27 and HF25 seeds at different developmental stages (Figures S1 and S2). Furthermore, the expression patterns of the two cultivars differed significantly (Figure S2). Eleven DEGs were selected for RT qPCR verification, and the experimental results were found to be significantly positively correlated with RNA-Seq, with a correlation coefficient of 0.85 (Figure S3), which validated the reliability of transcriptome sequencing results.

As depicted in Figure 2A, there were 12,393 DEGs between ZD27 and HF25, of which 4922 were individually up-regulated and 7097 were individually down-regulated. The expression patterns of 304 DEGs varied from S2 to S8. DEGs were mostly down-regulated in all 7 sampling periods. The highest number of DEGs was observed in S7 (8231), while the lowest in S6 (333) (Figure 2B). Across the seven sampling periods, 25 genes were co-upregulated and 27 genes were co-downregulated (Figure 2C,D).

All DEGs were annotated using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases and categorized into "cellular component," "molecular function," and "biological process," which encompassed 14, 16, and 21 second categories of GO terms (Figure S4A), respectively. Among the "biological process" terms, the most DEGs were associated with "cellular process," "metabolic process," and "biological regulation" had the most DEGs. In the "molecular function" group, the most common terms were "binding," "catalytic activity," and "transporter activity." The most common terms in the "cellular component" group were "cell part," "membrane part," and "organelle."

Using KEGG annotation, all DEGs were classified into five types of the first category: "metabolism (10)," "genetic information processing (4)," "environmental information processing (2)," "cellular processes (1)," and "organismal systems (1)," and into 18 types of the second category (Figure S4B). Among the "metabolism" type, "carbohydrate metabolism" had the highest number of DEGs (465). These DEGs were involved in 15 pathways (Figure S4C), with the "starch and sucrose metabolism" pathway contained the most DEGs (97) (Table S6).

3.6 | The detailed information of DEGs involved in sucrose metabolism and transportation

There were 25 genes involved in sucrose metabolism and 20 genes involved in sucrose transportation among the DEGs (Figure 3). Note that 5 hexokinase (*HK*), 3 fructokinase

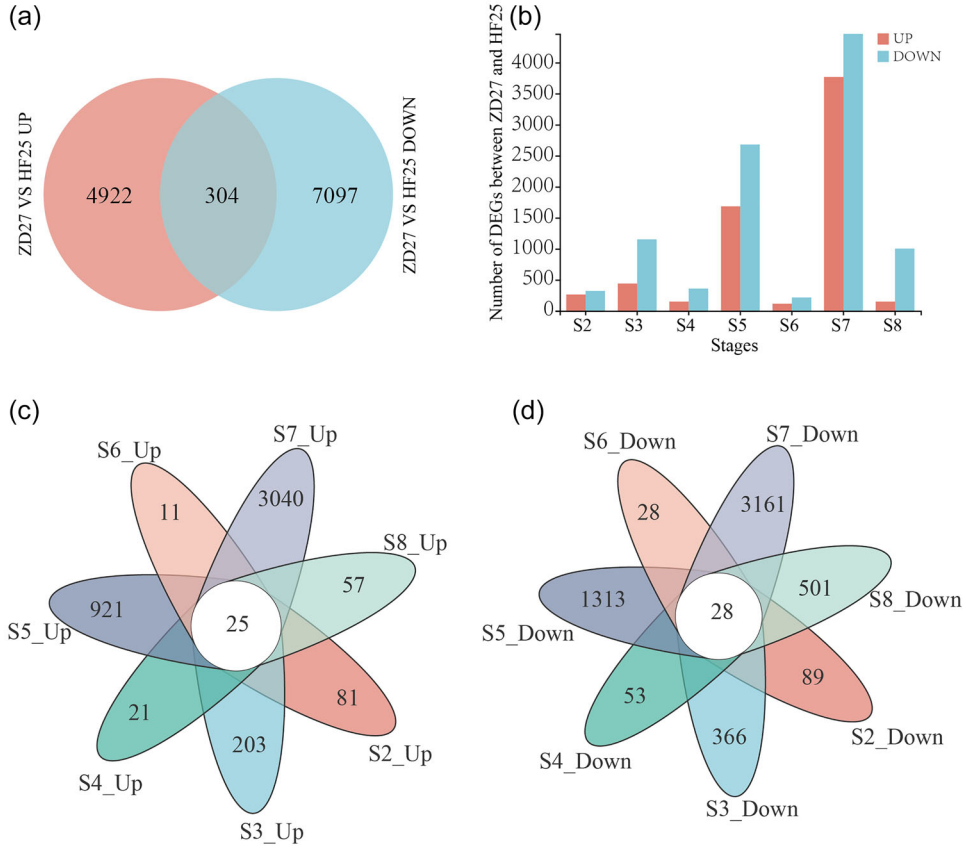

FIGURE 2 Comparison of differentially expressed genes (DEGs) between ZD27 and HF25. (A) Venn diagram of up-regulated and down-regulated DEGs of all sampling periods; (B) number of up-regulated and down-regulated DEGs from S2 to S8 stages; (C) Venn diagram of up-regulated DEGs at each stage from S2 to S8; and (D) Venn diagram of down-regulated DEGs at each stage from S2 to S8.

(*scrK*), 6 invertase (*INV*), 5 sucrose synthase (*SUS*), 4 sucrose-phosphate synthase (*SPS*), 1 glucose-6-phosphate isomerase (*GPI*), and 1 UTP-glucose-1-phosphate uridylyltransferase (*UGP2*) genes were involved in sucrose metabolism pathway (Figure 3A). Notably, down-regulation of INV DEGs, which regulate sucrose degradation, was observed during S2–S8 stage, while the up-regulation of *SPS* DEGs, which regulate sucrose synthesis, was observed with seed development, particularly the expression of *Glyma.17G109700* (*SPS*). Seventeen *SWEET* genes and three *SUT* genes were involved in sucrose transport (Figure 3B). Phylogenetic analysis revealed that these *SWEET* genes were homologous to *SWEET* genes in *Arabidopsis* (Figure 3C), and were divided into four subfamilies (Clade I–IV). Among them, eight genes were classified into Clade IV subfamily (*Glyma.08G077200*, *Glyma.15G049200*, *Glyma.08G183500*, *Glyma.05G202700*, *Glyma.08G010000*, *Glyma.04G198400*, *Glyma.06G167000*, *Glyma.06G166800*), five genes were classified into Clade I subfamily (*Glyma.13G264400*, *Glyma.12G234500*, *Glyma.14G120300*, *Glyma.13G041300*, and *Glyma.14G160100*), three genes were classified into Clade III subfamily (*Glyma.06G200800*, *Glyma.03G065500*, and *Glyma.19G066350*), and one gene was classified into Clade II subfamily (*Glyma.20G066500*). In Figure 3B, *SWEET* and *SUT* genes exhibited inconsistent expression levels from S2 to S8, with some genes showing high expression levels at S7 stage.

3.7 | Weighted correlation network analysis

The study conducted WGCNA analysis to identify co-expressed genes related to sucrose/soluble sugar concentration. A total of 9794 genes were used to build a co-expression network that resulted in 17 modules, ranging from 22 to 2087 genes each (Figure 4A, Table S7). Among these modules, five were significantly correlated with both sucrose and soluble sugar concentrations. ME yellowgreen, containing 149 DEGs, showed a positive correlation with both sucrose and soluble sugar concentrations. On the other hand, ME blue, ME red, ME magenta, and ME yellow exhibited negative correlations with both sucrose and soluble sugar concentrations (Figure 4B). Additionally, ME midnightblue was significantly positively correlated with sucrose, while ME brown was significantly positively correlated with soluble sugar. Lastly, significant negative correlations were

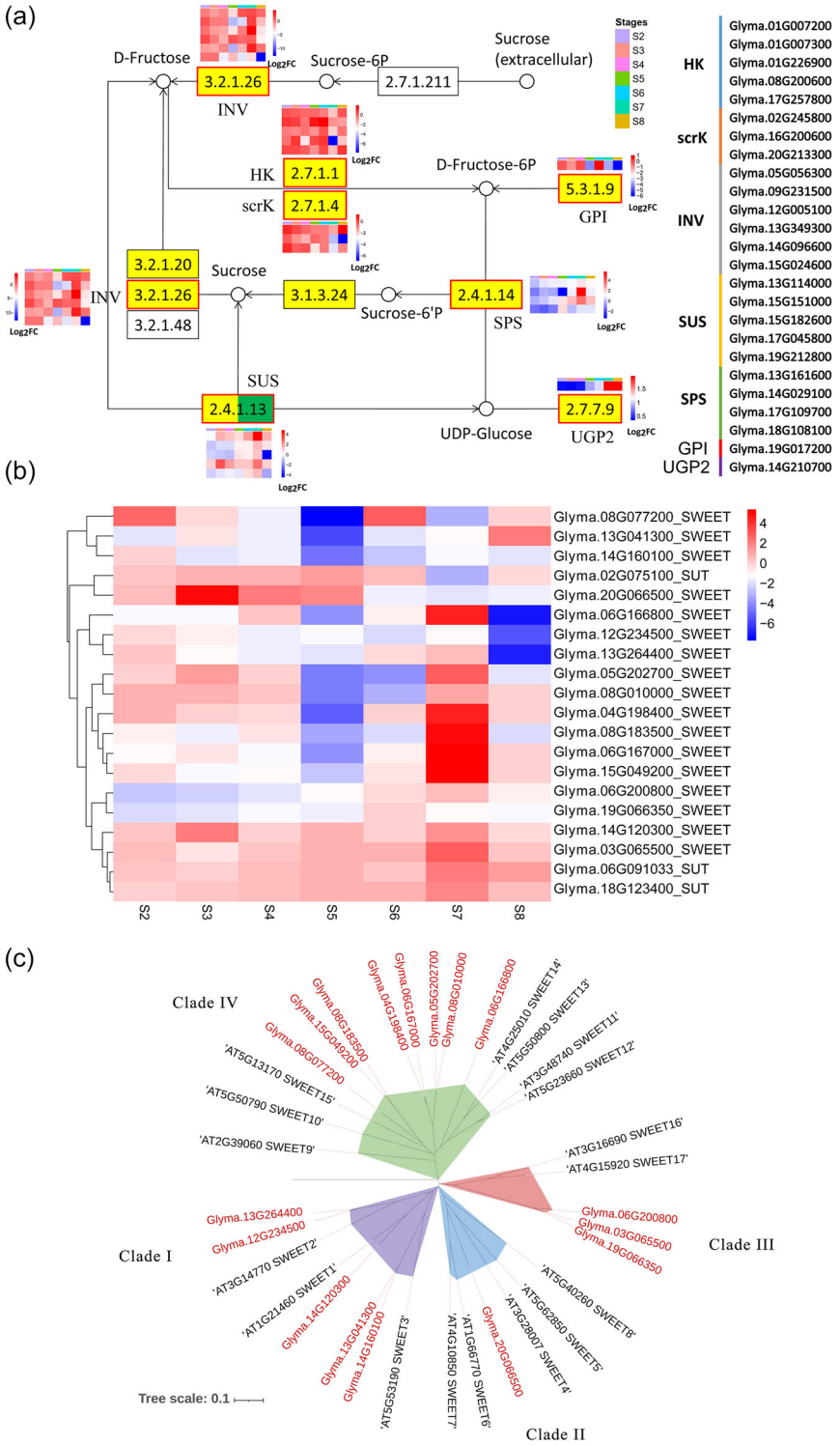

FIGURE 3 Differentially expressed genes (DEGs) of two soybean varieties involved in sucrose metabolism and transportation. (A) DEGs involved in sucrose metabolism; (B) DEGs involved in sucrose transportation; and (C) phylogenetic tree representing the relationships between seed-expressed water-soluble sugar transport protein (SWEET) genes in *Arabidopsis* and DEGs. GPI, glucose-6-phosphate isomerase; HK, hexokinase; INV, invertase; scrK, fructokinase; SPS, sucrose phosphate synthase; SUS, sucrose synthase; UGP2, UTP-glucose-1-phosphate uridylyltransferase.

found for ME salomen with sucrose, and for ME turquoise and ME purple with soluble sugar.

KEGG enrichment analysis was conducted for each module that was found to be significantly correlated with sucrose or soluble sugar (Table S8). We focused mostly on positively correlated modules (Figure 5). In the ME yellowgreen module, there were 149 DEGs enriched in 38 pathways which mostly belonged to the carbohydrate metabolism category. Notably, the pyruvate metabolism (map00620), pentose and glucuronate interconversions (map00040), and starch and sucrose metabolism (map00500) pathways were significantly enriched ($p < 0.05$) (Figure 5A). In the ME brown module, there were 1185 DEGs enriched in 102 pathways with amino acid metabolism (14%) and carbohydrate metabolism (13%) categories having the highest incidence of KEGG pathways. Galactose metabolism (map00052),

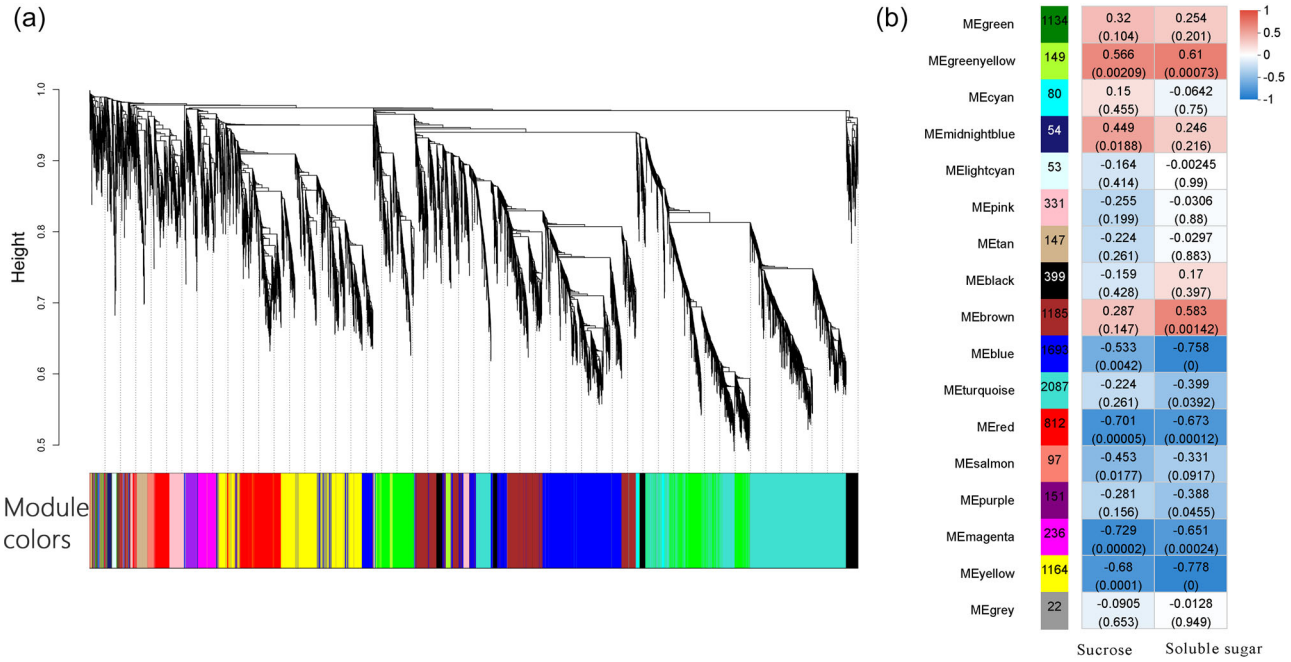

FIGURE 4 Co-expression network (A) and correlations (B) between module eigengene (ME) and sucrose and soluble sugar concentration in soybean varieties.

ascorbate and aldarate metabolism (map00053), and citrate cycle (TCA cycle) (map00020) pathways, which belong to the carbohydrate metabolism category, showed significant enrichment. Additionally, the glycerolipid metabolism (map00561) pathway was also found to be significantly enriched in the ME brown module (Figure 5B, Table S8). Finally, carbohydrate metabolism was noted as one of the most enriched categories in the MEmidnightblue module, with the amino sugar and nucleotide sugar metabolism (map00520) pathway being significantly enriched (Figure 5C).

It is worth noting that two essential genes involved in sucrose metabolism, namely the *SUS* (*Glyma.15G151000*) and *SPS* (*Glyma.18G108100*) genes, as well as four transcript factor (TF) genes (*Glyma.04G028100*, *Glyma.05G095900*, *Glyma.11G239200*, and *Glyma.12G076200*), were identified in the MEgreenyellow module, and can be integrated with the other 40 related DEGs to construct a gene regulatory network (GRN) (Figure S5).

3.8 | Candidate gene analysis of DEGs within the QTLs

The Venn analysis identified 233 genes that overlap between DEGs and QTL intervals (Figure 6A), including an *SPS* gene (*Glyma.17G109700*) and 11 TFs containing 29 transcripts (Table S9). Among these genes, six were most annotated for carbohydrate metabolism in the KEGG annotations, specifically *Glyma.17G101700*, *Glyma.17G100600*, *Glyma.17G116400*, *Glyma.17G089000*, *Glyma.17G076100*, and *Glyma.17G109700* (Figure 6C, Table S9).

Since *qSU1901* and *qSU2002* were identified in at least two conditions, DEGs in these two QTL intervals were investigated (Figure 6B). Additionally, *qSU1901* contained three DEGs (*Glyma.19G197100*, *Glyma.19G197200*, and *Glyma.19G197300*). Notably, *Glyma.19G197200* gene, annotated as secoisolariciresinol dehydrogenase, was only expressed in ZD27, during seed development (Table S10). Non-synonymous coding changes were found in *Glyma.19G197100* (secoisolariciresinol dehydrogenase gene) and *Glyma.19G197300* (TLD domain-containing protein 1 gene), in the exon region (Table S5). In *qSU2002*, a total of 35 DEGs were identified (Table S11), including three TFs from LBD, MYB, and GATA families (*Glyma.20G177600*, *Glyma.20G178500*, and *Glyma.20G180100*). Many polymorphism variations existed among those three genes. Notably, four non- synonymous coding changes were observed in *Glyma.20G178500* gene (Table S12).

4 | DISCUSSION

4.1 | Major QTLs and candidate genes controlling sucrose/soluble sugar accumulation

The current study determined the heritability (H^2) of sucrose and soluble sugar to be 58.5% and 54.6%, respectively, with CVs ranging from 14.2% to 19.5% and 15.3% to 19.8%, respectively. The concentration of sucrose and soluble sugars

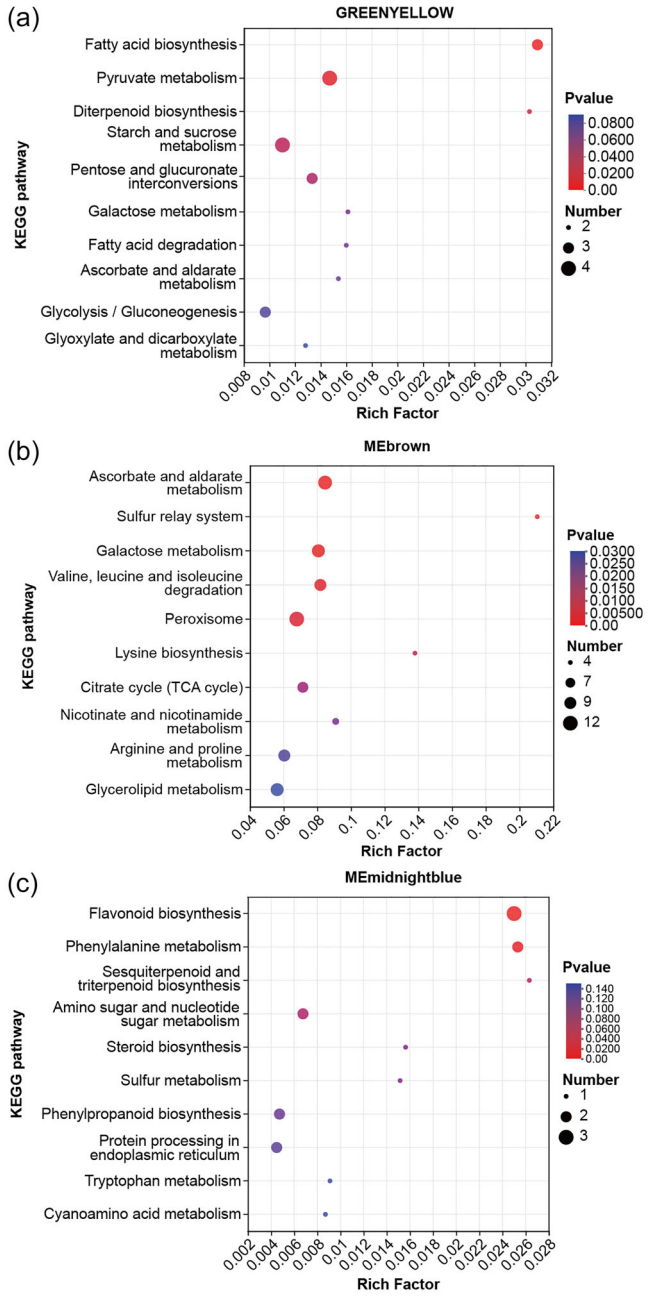

FIGURE 5 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment of the MEgreenyellow (A); MEbrown (B); and MEmidnightblue (C) modules. ME, module eigengene.

was influenced by both genotype and environmental factors. The absence of a significant genotype × environment interaction suggests that the data can be pooled and analyzed as a single dataset (Maughan et al., 2000). The average seed sucrose and soluble sugar concentrations for the RIL population decreased with the increase of active accumulated temperature during the growing season. Additionally, the highest sucrose and soluble sugar concentrations were found in Hailun location, which is consistent with the previous observation (Zeng et al., 2014). Given that sucrose is the main component of soluble sugar in soybean (Li et al., 2012; Liu et al., 2017), and the two sugars are positively correlated, the RIL population is a suitable choice for QTL localization analysis. However, when considering the sugar profile as a quantitate trait, it is important to note that environmental conditions can significantly influence the results. Multiple locations or years of observation and phenotyping may be required (Pan et al., 2022; Wang et al., 2014).

A total of eight sucrose-associated QTLs were detected in this study on seven chromosomes. Interestingly, a previously reported QTL, *qSU0701*, on Chr7 was also detected (Maughan et al., 2000). Furthermore, two novel QTLs, *qSU1901* on Chr19 (45,311,975–45,464,136) and *qSU2002* on Chr20 (40,523,599–41,882,459) were identified, which explained 10.63%–13.18% and 10.26%–14.44% of the phenotypic variation for sucrose, respectively, across multiple conditions. A previous study by Kim et al. (2005) also identified a major sucrose QTL on Chr19 using SSR markers, which explained 21.4% of the total phenotypic variation. additionally, major sucrose-associated QTLs have been reported on both Chr19 and Chr20, with the most collinear QTL relationships observed on Chr19 (Maughan et al., 2000; Pan et al., 2022). Therefore, Chr19 and Chr20 may contain major QTLs regulating sucrose accumulation in soybeans. Notably, *qSU1901* was detected in all conditions except for the Hailun location, possibly due to the small number of population lines at this experimental site, which can affect the outcome of QTL mapping (Chen et al., 2021). Therefore, we performed a BLUP model to fix this. Besides, the physical position of *qSU1901* was found to be close to that of a previously reported sucrose QTL (Pan et al., 2022). Our results confirm that *qSU1901* is a major sucrose-associated QTL with a relatively stable genetic effect.

In terms of soluble sugar, a total of eight soluble sugar-linked QTLs were found on seven chromosomes. *qSS0301* (Chr3: 6,004,297–6,444,222) was detected in both Mudanjiang and BLUP conditions, whereas *qSS1701* (Chr17: 7,470,395–10,014,816) and *qSS1702* (Chr17: 7,969,537–10,599,548) overlapped on Chr17 and were detected in Hailun and BLUP conditions. Interestingly, *qSS0601* (Chr6: 43,709,851–44,392,137) and sucrose-associated *qSU0601* (Chr6: 43,709,851–44,392,137) were located precisely on Chr6, indicating a potential simultaneous impact on the contents of sucrose and soluble sugar. Of note, we identified a *qSS0201*, in Hailun location, had also been detected by Kim et al. (2005). Since these QTLs were not consistently detected in multiple environments, they may be greatly affected by environmental conditions, and further experiments are required to confirm their effects (Park et al., 2019).

After annotating the genes within all detected QTL intervals, we identified 17 genes related to sucrose or soluble sugar transport and metabolism. These genes

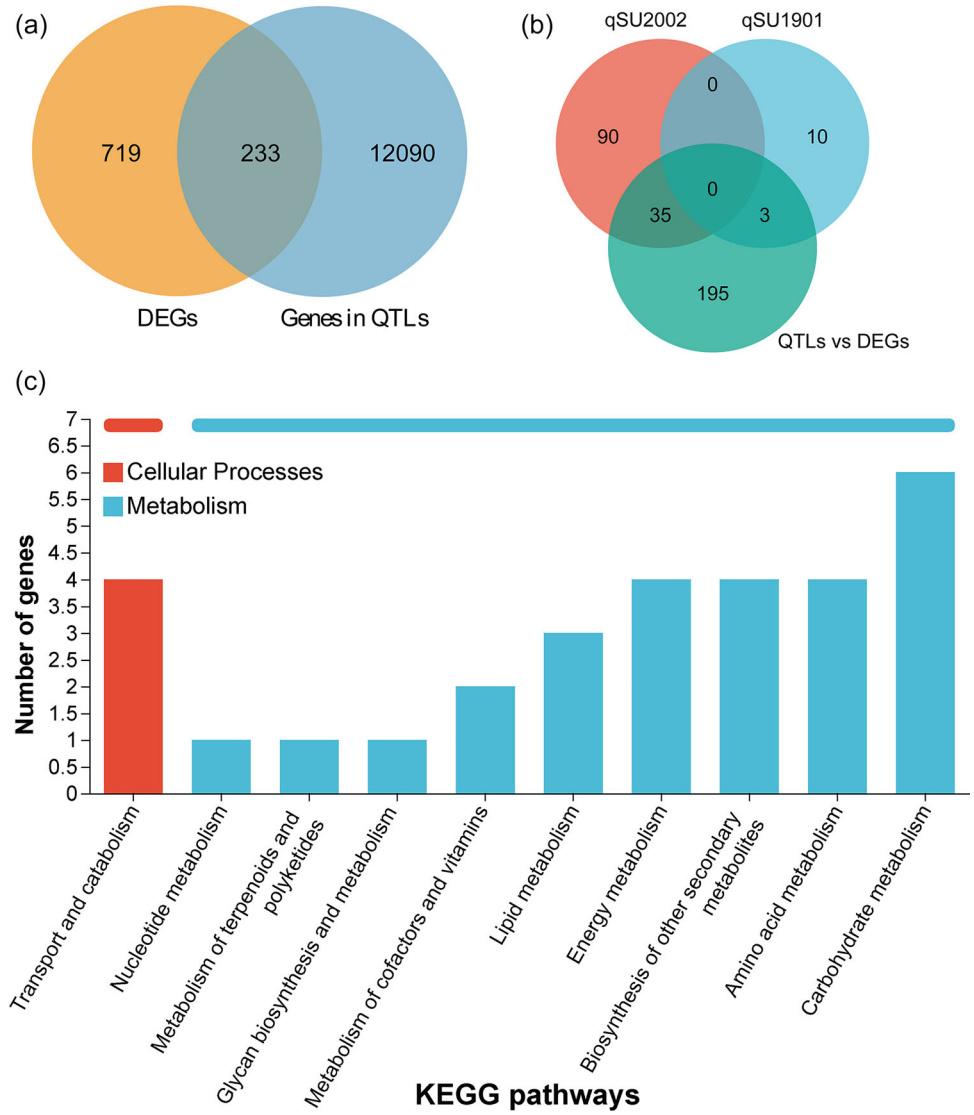

FIGURE 6 Venn analysis of genes among all detected differentially expressed genes (DEGs) and quantitative trait loc (QTLs) (A); Venn analysis of genes among *qSU19001*, *qSU2002* and all detected DEGs (B); and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation of all DEGs within the major QTLs (C).

were derived from six QTLs: *qSU1701* (6), *qSU1901* (1), and *qSU2002* (4) were sucrose-associated QTLs, while *qSS0201* (1) and *qSS1701/qSS1702* (5) were soluble sugar-associated QTLs. Due to their important roles in sucrose transport and degradation, SUCs and INV are crucial factors to consider when studying sucrose transport and metabolism. The identification of an *INV* gene (*Glyma.20G177200*) and two sucrose transport protein SUC5 genes (*Glyma.20G174100*, *Glyma.20G174200*) in *qSU2002* makes them prime candidates for further investigation. Moreover, *Glyma.20G177200* and *Glyma.20G174100* were found to have polymorphic variations in the exon region, further emphasizing their importance. Sucrose-phosphate synthase (*SPS1*, *Glyma.17G109700*), a main enzyme facilitating sucrose synthesis and playing a major role in seed sucrose accumulation, was identified in *qSS1701/qSS1702* intervals. This gene was found highly expressed in ZD27. However, no significant polymorphism differences were detected between the two parents of this gene, which might be regulated by other factors, such as transcription factors (TFs) (Payyavula et al., 2013; Perlo et al., 2022). Furthermore, 59 TFs were identified within all of the detected QTLs. Of particular interest, a predicted B3 domain-containing protein transcript factor (*Glyma.19G196600*) was found in the major QTL *qSU1901*. Previous research has shown that B3 domain-containing protein transcript factors can regulate *SUS* genes by controlling downstream O2 transcription factors (bZIP family) (Yang et al., 2021). Polymorphism analysis performed on *Glyma19G196600* gene, revealed the presence of both SNP and Indel variations were detected in the exon region. Therefore, it is highly likely that *Glyma19G196600* acts as a regulator of sucrose accumulation in soybean.

Nevertheless, additional genetic and functional studies are necessary to confirm its exact function.

4.2 | Transcriptome analysis of sucrose and soluble sugar accumulation

During seed development, HF25 and ZD27 varieties grew rapidly from 2 (S2) to 4 (S4) weeks after beginning seed stage (R5). Afterward, seed sucrose and soluble sugar concentrations fluctuated, consistent with findings from previous studies (Liu et al., 2017; Tu et al., 2017). Typically, ZD27 has a higher sugar content than HF25. Likewise, gene expression in seed followed a certain time sequence and demonstrated different patterns between the two varieties. Comparable outcomes were observed in the high-sugar peanut variety ICG 12525 and low-sugar peanut variety Zhonghua 10 (Li et al., 2021). Thus, the accumulation of sucrose and soluble sugar in seed is a dynamic process that is regulated by the temporal expression of specific genes (Du et al., 2020). During this process, the precise regulation of key genes leads to differences in sugar content among varieties (Li et al., 2021).

This study identified numerous DEGs, with specific focus on those involved in sugar transport and metabolism. Given that carbohydrates are transported and metabolized through a series of enzymes (Ruan, 2014), identifying the DEGs within these pathways is crucial. During seed development, specific differences were observed in carbohydrate metabolism between ZD27 and HF25, particularly within the sucrose metabolism pathway. A total of five HK genes, three *scrK* genes, six *INV* genes, five *SUS* genes, four *SPS* genes, one *GPI*, and one *UGP2* gene were found to be differentially expressed between the two varieties. Notably, ZD27 exhibited generally lower levels of differentially expressed *INV* genes, but showed higher levels of *SPS* genes during seed development. SPS is the primary enzyme responsible for stimulating sucrose synthesis, whereas INV catalyzes sucrose decomposition (Winter & Huber, 2000). Soybean seeds accumulate sucrose more effectively when enzyme activity increases toward sucrose synthesis and decreases toward sucrose degradation (Liu et al., 2017). Thus, the expression patterns of key enzymes involved in sucrose metabolism might contribute to the differences in sugar content among varieties. Of these enzyme genes, *GmSPS1* (*Glyma.17G109700*), *GmSPS2* (*Glyma.14G029100*), *GmSuSy1* (*Glyma.13G114000*), and *GmA-INV* (*Glyma.05G056300*) have been previously reported in soybeans and warrant further investigation (Du et al., 2020).

Recent studies have highlighted the role of SWEET transporters in sugar transport and regulation of sugar translocation efficiency across membranes, thereby influencing seed sucrose and soluble sugar accumulation (Jeena et al., 2019). These SWEET transporters are located within different compartments throughout the cell (Chen et al., 2012; Li et al., 2021). In *Arabidopsis*, *SWEET* gene family of sugar transporters can be divided into four subclades (Clade I–IV), depending on their preferential transportation of distinct sugar moieties (Chen et al., 2012; Jeena et al., 2019). As reported, Clade I, II, and IV transported primarily hexoses, while SWEETs from clade III transported primarily sucrose, although it is also possible for them to transport hexoses (Feng & Frommer, 2015). In this study, a total of 17 SWEET homologous genes from across the four Clades were found differentially expressed with distinct temporal expression patterns. Specifically, in ZD27, SWEET family genes were highly expressed at S7 stage, including *GmSWEET12* (*Glyma.05G202700*), *GmSWEET15* (*Glyma.06G166800*), and *GmSWEET21* (*Glyma.08G100000*), which have been reported closely linked to sucrose accumulation (Du et al., 2020; Wang et al., 2019). In addition, three *SUT*/*SUC* genes were identified as differentially expressed, and they are linked to sugar accumulation in sink organs (Peng et al., 2020). Taken together, these *SWEET* and *SUT*/*SUC* family genes may also play a crucial role in soybean sugar accumulation. Our identification results can serve as a foundation for future gene function analysis studies.

To better understand the complex biological processes involved in soybean sucrose/soluble sugar metabolism and identify the synergistic relationship between DEGs, we performed WGCNA, breaking down the co-expressed genes into different pathways (Li et al., 2021). Three significantly positively correlated modules with sucrose and/or soluble sugar concentration modules were constructed. KEGG enrichment analysis indicated that carbohydrate metabolism was a principal process involved in these modules. Among the three modules, the MEgreenyellow module was found to be positively correlated with both sucrose and soluble sugar. This module contains a *SUS* (*Glyma.15G151000*) gene and an *SPS* (*Glyma.18G108100*) gene, and four correlated transcript factor (TF) genes (*Glyma.04G028100*, *Glyma.05G095900*, *Glyma.11G239200*, and *Glyma.12G076200*), which might lead to a potential GRN for soybean sucrose/soluble sugar accumulation. As TFs interact with the cis-elements of downstream gene promoters, they are crucial for sucrose storage and transportation (Chen et al., 2020; Li et al., 2020; Xiao et al., 2016). The four TFs involved in the GRN, belong to the *AP2*/*ERF* (*Glyma.04G028100* and *Glyma.11G239200*), *HSF* (*Glyma.05G095900*), and *ARF* (*Glyma.12G076200*) families respectively, and these families have been associated with sucrose accumulation in previous studies (Li et al., 2021). Overall, the results of WGCNA are useful for designing sucrose and soluble sugar molecular breeding modules, as they provide a better understanding of the complex biological processes involved in soybean sugar metabolism and the synergistic relationships between different DEGs.

By combining QTL mapping and RNA-seq data, we were able to pinpoint candidate genes more precisely (Park et al., 2019). Based on this, we narrowed the list down to 233 candidate genes. Most of the candidate genes identified were annotated in the carbohydrate metabolism pathway, highlighting the crucial role of this pathway in soybean sugar metabolism. Interestingly, the SPS gene, *GmSPS1* (*Glyma.17G109700* belongs to *qSS1701/qSS1702*), was also detected in this part. Therefore, understanding the upstream genetic regulation of *GmSPS1* and identifying potential regulators of this gene are important future research directions.

Additionally, we detected 11 TF genes from 11 TF families that might be significant upstream regulators of sugar accumulation in soybean (Chen et al., 2020; Li et al., 2021; Xiao et al., 2016). Thus, investigating the regulatory mechanisms of these TF genes could provide valuable insights into the genetic regulation of soybean sugar metabolism, and might lead to the development of novel molecular breeding strategies for enhanced sugar accumulation in soybean.

In conclusion, our study identified 16 QTLs associated with sucrose or soluble sugar concentrations in soybean, and we detected several candidate genes within these QTL intervals through polymorphism analysis. The major sucrose-associated QTL *qSU1901* was detected across multiple conditions and explained 10.6%–13.2% of the phenotypic variation in sucrose concentration. Our dynamic transcriptome analysis revealed temporal expression patterns of key enzyme genes involved in sucrose and soluble sugar transport and metabolism. Additionally, a GRN was constructed containing a *SUS* gene, an *SPS* gene, and four correlated TFs, providing valuable insights into the genetic regulation of sucrose metabolism in soybean. Through the integration of QTL mapping and RNA-Seq, we identified several candidate genes that could be targeted in future breeding strategies aimed at producing soybeans with improved sugar content and quality. Overall, our study offers valuable implications for the genetic improvement of soybean sugar metabolism.

AUTHOR CONTRIBUTIONS

Changkai Liu: Conceptualization; data curation; funding acquisition; investigation, writing—original draft. **Heng Chen**: Data curation. **Qintan Yu**: Data curation. **Haidong Gu**: Methodology; software. **Yansheng Li**: Investigation; software. **Tu Bingjie**: Formal analysis; software. **Hengyou Zhang**: Writing—review and editing. **Qiuying Zhang**: Conceptualization; funding acquisition. **Xiaobing Liu**: Supervision; writing—review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI repository, accession numbers: PRJNA778303, PRJNA778408, and PRJNA895728.

ORCID

Changkai Liu https://orcid.org/0000-0001-8455-2897

REFERENCES

Barth, I., & Sauer, M. N. (2003). PmSUC3: Characterization of a SUT2/SUC3-type sucrose transporter from *Plantago major*. *Plant Cell*, *15*, 1375–1385. https://doi.org/10.1105/tpc.010967

Bernardo, R. (1996). Best linear unbiased prediction of maize single-cross performance. *Crop Science*, *36*, 50–56. https://doi.org/10.2135/cropsci1996.0011183X003600010009x

Cai, J. S., Feng, J. Y., Ni, Z. J., Ma, R. H., Thakur, K., Wang, S., & Wei, Z. J. (2021). An update on the nutritional, functional, sensory characteristics of soy products, and applications of new processing strategies. *Trends in Food Science & Technology*, *112*, 676–689. https://doi.org/10.1016/j.tifs.2021.04.039

Chen, H., Pan, X., Wang, F., Liu, C., Wang, X., Li, Y., & Zhang, Q. (2021). Novel QTL and meta-QTL mapping for major quality traits in soybean. *Frontiers in Plant Science*, *12*, 774270. https://doi.org/10.3389/fpls.2021.774270

Chen, L. Q., Lin, I. W., Qu, X. Q., Sosso, D., McFarlane, H., Londoño, A., Samuels, A. L., & Frommer, W. B. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the Arabidopsis embryo. *Plant Cell*, *27*, 607–619. https://doi.org/10.1105/tpc.114.134585

Chen, L. Q., Qu, X. Q., Hou, B. H., Sosso, D., Osorio, S., Fernie, A. R., & Frommer, W. B. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science*, *335*, 207–211. https://doi.org/10.1126/science.1213351

Chen, Q., Tang, Y. M., Wang, Y., Sun, B., Chen, T., Lei, D. Y., Zhang, F., Luo, Y., Zhang, Y., Wang, X. R., & Tang, H. R. (2020). Enhance sucrose accumulation in strawberry fruits by eliminating the translational repression of *FabZIPs1.1*. *Scientia Horticulturae*, *259*, 108850. https://doi.org/10.1016/j.scienta.2019.108850

Dierking, E. C., & Bilyeu, K. D. (2008). Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome*, *1*(2), 135–145.

Du, Y., Zhao, Q., Chen, L., Yao, X., Zhang, H., Wu, J., & Xie, F. (2020). Effect of drought stress during soybean R2-R6 growth stages on sucrose metabolism in leaf and seed. *International Journal of Molecular Sciences*, *21*, 618. https://doi.org/10.3390/ijms21020618

Fehr, W. R., & Caviness, C. E. (1977). *Stages of soybean development*. Iowa State University of Science and Technology. Special Report.87. http://lib.dr.iastate.edu/specialreports/87

Feng, L., & Frommer, W. B. (2015). Structure and function of SemiSWEET and SWEET sugar transporters. *Trends in Biochemical Sciences*, *40*, 480–486. https://doi.org/10.1016/j.tibs.2015.05.005

Ficht, A., Bruce, R., Torkamaneh, D., Grainger, C. M., Eskandari, M., & Rajcan, I. (2022). Genetic analysis of sucrose concentration in soybean seeds using a historical soybean genomic panel. *Theoretical and Applied Genetics*, *135*, 1375–1383. https://doi.org/10.1007/s00122-022-04040-z

Guo, J., Gu, X., Lu, W., & Lu, D. (2021). Multiomics analysis of kernel development in response to short-term heat stress at the grain formation stage in waxy maize. *Journal of Experimental Botany*, *72*, 6291–6304. https://doi.org/10.1093/jxb/erab286

Hu, L., Wang, X., Zhang, J., Florez-Palacios, L., Song, Q., & Jiang, G. L. (2023). Genome-wide detection of quantitative trait loci and prediction of candidate genes for seed sugar composition in early mature soybean. *International Journal of Molecular Sciences*, *24*, 3167. https://doi.org/10.3390/ijms24043167

Hubbard, N. L., Huber, S. C., & Pharr, D. M. (1989). Sucrose phosphate synthase and acid invertase as determinants of sucrose concentration in developing muskmelon (*Cucumis melo* L.) fruits. *Plant Physiology*, *91*, 1527–1534. https://doi.org/10.1104/pp.91.4.1527

Jeena, G. S., Kumar, S., & Shukla, R. K. (2019). Structure, evolution and diverse physiological roles of SWEET sugar transporters in plants. *Plant Molecular Biology*, *100*, 351–365. https://doi.org/10.1007/s11103-019-00872-4

Kim, H. K., Kang, S. T., Cho, J. H., Choung, M. G., & Suh, D. Y. (2005). Quantitative trait loci associated with oligosaccharide and sucrose contents in soybean (*Glycine max* L.). *Journal of Plant Biology*, *48*, 106–112. https://doi.org/10.1007/BF03030569

Kumar, V., Rani, A., Goyal, L., Dixit, A. K., Manjaya, J. G., Dev, J., & Swamy, M. (2010). Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *Journal of Agricultural and Food Chemistry*, *58*, 5081–5085. https://doi.org/10.1021/jf903141s

Lee, J. S., Kim, S. M., & Kang, S. (2016). Fine mapping of quantitative trait loci for sucrose and oligosaccharide contents in soybean [*Glycine max* (L.) Merr.] using 180 K Axiom SoyaSNP genotyping platform. *Euphytica*, *208*, 195–203. https://doi.org/10.1007/s10681-015-1622-x

Li, W., Huang, L., Liu, N., Pandey, M. K., Chen, Y., Cheng, L., Guo, J., Yu, B., Luo, H., Zhou, X., Huai, D., Chen, W., Yan, L., Wang, X., Lei, Y., RK Varshney, K., & Liao, B. (2021). Key regulators of sucrose metabolism identified through comprehensive comparative transcriptome analysis in peanuts. *International Journal of Molecular Sciences*, *22*, 7266. https://doi.org/10.3390/ijms22147266

Li, X., Guo, W., Li, J., Yue, P., Bu, H., Jiang, J., Liu, W., Xu, Y., Yuan, H., Li, T., & Wang, A. (2020). Histone acetylation at the promoter for the transcription factor puwrky31 affects sucrose accumulation in pear fruit. *Plant Physiology*, *182*, 2035–2046. https://doi.org/10.1104/pp.20.00002

Li, Y. S., Du, M., Zhang, Q. Y., Wang, G. H., & Liu, X. B. (2012). Greater differences exist in seed protein, oil, total soluble sugar and sucrose content of vegetable soybean genotypes [*Glycine max* (L.) Merrill] in Northeast China. *Australian Journal of Crop Science*, *6*, 1681–1686.

Liu, C., Tu, B., Li, Y., Tian, B., Zhang, Q., Liu, X., & Herbert, S. J. (2017). Potassium application affects key enzyme activities of sucrose metabolism during seed filling in vegetable soybean. *Crop Science*, *57*, 2707–2717. https://doi.org/10.2135/cropsci2016.08.0648

Liu, C., Tu, B., Wang, X., Li, Y., Zhang, Q., & Liu, X. (2020). Transcript profile in vegetable soybean roots reveals potential gene patterns regulating K uptake efficiency. *Agronomy*, *10*, 1796. https://doi.org/10.3390/agronomy10111796

Lu, W., Sui, M., Zhao, X., Jia, H., Han, D., Yan, X., & Han, Y. (2022). Genome-wide identification of candidate genes underlying soluble sugar content in vegetable soybean (*Glycine max* L.) via association and expression analysis. *Frontiers in Plant Science*, *13*, 930639. https://doi.org/10.3389/fpls.2022.930639

Luo, Q., Zheng, Q., Hu, P., Liu, L., Yang, G., Li, H., Li, B., & Li, Z. (2021). Mapping QTL for agronomic traits under two levels of salt stress in a new constructed RIL wheat population. *Theoretical and Applied Genetics*, *134*, 171–189. https://doi.org/10.1007/s00122-020-03689-8

Maughan, P., Maroof, M. S., & Buss, G. (2000). Identification of quantitative trait loci controlling sucrose content in soybean (*Glycine max*). *Molecular Breeding*, *6*, 105–111. https://doi.org/10.1023/A:1009628614988

Morse, E. E. (1947). Anthrone in estimating low concentrations of sucrose. *Analytical Chemistry*, *19*(12), 1012–1013. https://doi.org/10.1021/ac60012a021

Nyquist, W. E., & Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *Critical Reviews in Plant Sciences*, *10*(3), 235–322. https://doi.org/10.1080/07352689109382313

Pan, W. J., Han, X., Huang, S. Y., Yu, J., Zhao, Y., Qu, K. X., Zhang, Z. X., Yin, Z. G., Qi, H. D., Yu, G. L., Zhang, Y., Xin, D. W., Zhu, R. S., Liu, C. Y., Wu, X. X., Jiang, H. W., Hu, Z. B., Zuo, Y. H., Chen, Q. S., & Qi, Z. M. (2022). Identification of candidate genes related to soluble sugar contents in soybean seeds using multiple genetic analyses. *Journal of Integrative Agriculture*, *21*, 1886–1902. https://doi.org/10.1016/S2095-3119(21)63653-5

Park, M., Lee, J. H., Han, K., Jang, S., Han, J., Lim, J. H., Jung, J. W., & Kang, B. C. (2019). A major QTL and candidate genes for capsaicinoid biosynthesis in the pericarp of *Capsicum chinense* revealed using QTL-seq and RNA-seq. *Theoretical and Applied Genetics*, *132*, 515–529. https://doi.org/10.1007/s00122-018-3238-8

Patil, G., Valliyodan, B., Deshmukh, R., Prince, S., Nicander, B., Zhao, M., Song, L., Sonah, H., Lin, L., Chaudhary, J., Liu, Y., Joshi, T., Xu, D., & Nguyen, H. T. (2015). Soybean (*Glycine max*) SWEET gene family: Insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genomics*, *16*, 520. https://doi.org/10.1186/s12864-015-1730-y

Payyavula, R. S., Singh, R. K., & Navarre, D. A. (2013). Transcription factors, sucrose, and sucrose metabolic genes interact to regulate potato phenylpropanoid metabolism. *Journal of Experimental Botany*, *64*, 5115–5131. https://doi.org/10.1093/jxb/ert303

Peng, Q., Cai, Y., Lai, E., Nakamura, M., Liao, L., Zheng, B., Ogutu, C., & Han, Y. (2020). The sucrose transporter MdSUT4.1 participates in the regulation of fruit sugar accumulation in apple. *BMC Plant Biology*, *20*, 191. https://doi.org/10.1186/s12870-020-02406-3

Perlo, V., Margarido, G. R., Botha, F. C., Furtado, A., Hodgson-Kratky, K., Correr, F. H., & Henry, R. J. (2022). Transcriptome changes in the developing sugarcane culm associated with high yield and early-season high sugar content. *Theoretical and Applied Genetics*, *135*, 1619–1636. https://doi.org/10.1007/s00122-022-04058-3

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, *29*, e45. https://doi.org/10.1093/nar/29.9.e45

Ren, Y., Yu, G., Shi, C., Liu, L., Guo, Q., Han, C., Zhang, D., Zhang, L., Liu, B., Gao, H., Zeng, J., Zhou, Y., Qiu, Y., Wei, J., Luo, Y., Zhu, F., Li, X., Wu, Q., Li, B., … Huang, H. (2022). Majorbio Cloud: A one-stop, comprehensive bioinformatic platform for multiomics analyses. *iMeta*, *1*, e12. https://doi.org/10.1002/imt2.12

Ruan, Y. L. (2014). Sucrose metabolism: Gateway to diverse carbon use and sugar signaling. *Annual Review of Plant Biology*, *65*, 33–67. https://doi.org/10.1146/annurev-arplant-050213-040251

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., … Jackson, S. A (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*, 178–183. https://doi.org/10.1038/nature08670

Sui, M., Wang, Y., Bao, Y., Wang, X., Li, R., Lv, Y., Yan, M., Quan, C., Li, C., Teng, W., Li, W., Zhao, X., & Han, Y. (2020). Genome-wide association analysis of sucrose concentration in soybean (*Glycine max* L.) seed based on high-throughput sequencing. *Plant Genome*, *13*, e20059.

Susana, C., Cindy, L., Sara, C., Javier, C., Roland, L., Carolina, E., & Julia, H. (2013). Altered sucrose synthase and invertase expression affects the local and systemic sugar metabolism of nematode-infected *Arabidopsis thaliana* plants. *Journal of Experimental Botany*, *65*, 201–212.

Tang, Y., Ren, J., Liu, C., Jiang, J., Yang, H., & Li, J. (2021). Genetic characteristics and QTL analysis of the soluble sugar content in ripe tomato fruits. *Scientia Horticulturae*, *276*, 109785. https://doi.org/10.1016/j.scienta.2020.109785

Tu, B., Liu, C., Tian, B., Zhang, Q., Liu, X., & Herbert, S. J. (2017). Reduced abscisic acid content is responsible for enhanced sucrose accumulation by potassium nutrition in vegetable soybean seeds. *Journal of Plant Research*, *130*, 551–558. https://doi.org/10.1007/s10265-017-0912-x

Wang, S., Yokosho, K., Guo, R., Whelan, J., Ruan, Y. L., Ma, J. F., & Shou, H. (2019). The soybean sugar transporter *GmSWEET15* mediates sucrose export from endosperm to early embryo. *Plant Physiology*, *180*, 2133–2141. https://doi.org/10.1104/pp.19.00641

Wang, Y. Q., Chen, P. Y., & Zhang, B. (2014). Quantitative trait loci analysis of soluble sugar contents in soybean. *Plant Breeding*, *133*, 493–498. https://doi.org/10.1111/pbr.12178

Wei, L., Mao, W., Jia, M., Xing, S., Ali, U., Zhao, Y., Chen, Y., Cao, M., Dai, Z., Zhang, K., Dou, Z., Jia, W., & Li, B. (2018). FaMYB44.2, a transcriptional repressor, negatively regulates sucrose accumulation in strawberry receptacles through interplay with FaMYB10. *Journal of Experimental Botany*, *69*, 4805–4820. https://doi.org/10.1093/jxb/ery249

Wickham, H. (2016). Data analysis. *ggplot2: Elegant graphics for data analysis* (pp. 189–201). Springer.

Winter, H., & Huber, S. C. (2000). Regulation of sucrose metabolism in higher plants: Localization and regulation of activity of key enzymes. *Critical Reviews in Plant Sciences*, *19*, 31–67. https://doi.org/10.1080/07352680091139178

Xiao, G., Qin, H., Zhou, J., Quan, R., Lu, X., Huang, R., & Zhang, H. (2016). *OsERF2* controls rice root growth and hormone responses through tuning expression of key genes involved in hormone signaling and sucrose metabolism. *Plant Molecular Biology*, *90*(3), 293–302. https://doi.org/10.1007/s11103-015-0416-9

Xu, W., Liu, H., Li, S., Zhang, W., Wang, Q., Zhang, H., Liu, X., Cui, X., Chen, X., Tang, W., Li, Y., Zhu, Y., & Chen, H. (2022). GWAS and identification of candidate genes associated with seed soluble sugar content in vegetable soybean. *Agronomy*, *12*, 1470. https://doi.org/10.3390/agronomy12061470

Yang, T., Guo, L., Ji, C., Wang, H., Wang, J., Zheng, X., Xiao, Q., & Wu, Y. (2021). The B3 domain-containing transcription factor ZmABI19 coordinates expression of key factors required for maize seed development and grain filling. *Plant Cell*, *33*, 104–128. https://doi.org/10.1093/plcell/koaa008

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., Florez-Palacios, L., Brye, K., & Song Q Cregan, P. (2014). Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Science*, *54*, 554–564. https://doi.org/10.2135/cropsci2013.01.0036

Zhang, S., Du, H., Ma, Y., Li, H., Kan, G., & Yu, D. (2021). Linkage and association study discovered loci and candidate genes for glycinin and β-conglycinin in soybean (*Glycine max* L. Merr.). *Theoretical and Applied Genetics*, *134*, 1201–1215. https://doi.org/10.1007/s00122-021-03766-6

Zhen, Q., Fang, T., Peng, Q., Liao, L., Zhao, L., Owiti, A., & Han, Y. (2018). Developing gene-tagged molecular markers for evaluation of genetic association of apple SWEET genes with fruit sugar accumulation. *Horticulture Research*, *5*, 14. https://doi.org/10.1038/s41438-018-0024-3

Zheng, M., Peng, C., Liu, H., Tang, M., Yang, H., Li, X., Liu, J., Sun, X., Wang, X., Xu, J., Hua, W., & Wang, H. (2017). Genome-wide association study reveals candidate genes for control of plant height, branch initiation height and branch number in rapeseed (*Brassica napus* L.). *Frontiers in Plant Science*, *8*, 1246. https://doi.org/10.3389/fpls.2017.01246

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Liu, C., Chen, H., Yu, Q., Gu, H., Li, Y., Tu, B., Zhang, H., Zhang, Q., & Liu, X. (2023). Identification of quantitative trait loci and candidate genes for seed sucrose and soluble sugar concentrations in soybean. *Crop Science*, *63,* 2976–2992. https://doi.org/10.1002/csc2.21080