Received: 22 March 2020 | Accepted: 2 September 2020 | Published online: 14 October 2020

DOI: 10.1002/tpg2.20059

The Plant Genome

ORIGINAL RESEARCH

Genome-wide association analysis of sucrose concentration in soybean (*Glycine max* L.) seed based on high-throughput sequencing

Meinan Sui | **Yue Wang** | **Yuyue Bao** | **Xi Wang** | **Ruiqiong Li** | **Yan Lv** | **Ming Yan** | **Chao Quan** | **Chunxia Li** | **Weili Teng** | **Wenbin Li** | **Xue Zhao** | **Yingpeng Han**

Key Laboratory of Soybean Biology and Breeding/Genetics, Chinese Ministry of Education, Northeast Agricultural University, Harbin, 150030, China

Correspondence
Xue Zhao and Yingpeng Han, Key Laboratory of Soybean Biology and Breeding/Genetics, Chinese Ministry of Education, Northeast Agricultural University, Harbin, 150030, China.
Email: hyp234286@aliyun.com; xuezhao@neau.edu.cn.

Meinan Sui, Yue Wang, Yuyue Bao, Xue Zhao, and Yingpeng Han contributed equally to this work.

Funding information
Heilongjiang Provincial Project, Grant/Award Numbers: GX17B002, JC2018007, SC2019ZX16B0039, GJ2018GJ0098, C2018016; Chinese National Natural Science Foundation, Grant/Award Numbers: 31701446, 31871650, 31971967, 31671717, 31471517; National Key R & D Project, Grant/Award Numbers: 2017YFD0101306, 2016YFD0100304, 2017YFD0101302; Ministry of Science and Technology in China, Grant/Award Number: 2015RA228; Heilongjiang Province, Grant/Award Numbers: LBH-Z15017, LBH-Q17015

Abstract

The sucrose concentration in soybean seed significantly affects the flavor of soybean-derived products. In this study, an association panel of 178 elite accessions and 33,149 single-nucleotide polymorphisms (SNPs) was utilized to identify quantitative trait nucleotides (QTNs) of sucrose concentration in soybean seeds by genome-wide association study (GWAS). Five QTNs (rs2688589, rs29026218, rs5926884, rs6886889, and rs10299216) distributed across five genomic regions in five chromosomes were identified in two or more locations by GWAS. A total of 60 candidate genes near the 200-kb flanking region of these five identified loci were identified. Three of these genes (*Glyma.04G032600*, *Glyma.04G034600*, and *Glyma.11G092100*) have been reported to be involved in the process of sugar biosynthesis. Based on gene-based association and haplotype analyses, a total of 35 SNPs from 10 genes associated with sucrose concentration were identified. Of them, *Glyma.04G032600* was the only gene that has been reported to be related to sucrose content; the other nine genes were novel and may be associated with sucrose content. These beneficial alleles and candidate genes may be of great value in improving sucrose content in soybean seeds.

1 | INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] seeds, on a dry matter basis, usually contain about 40% protein, 20% oil, 35%

Abbreviations: GWAS, genome-wide association study; QTN, quantitative trait nucleotides; SNP, single-nucleotide polymorphism.

carbohydrates, and 5% ash (Krober, & Cartter, 1962). They are often directly utilized as raw products in the creation of various oriental soy foods items (Wang, Chen, & Zhang, 2014). About 60% and 40% of carbohydrates are insoluble and soluble sugars, respectively. Soluble sugar in soybean dry matter is comprised of 5% sucrose, 1.5% raffinose, and 3% stachyose (Wilson, 2004). The soluble sugar concentration in soybean seeds, particularly sucrose concentration, significantly affect the taste and flavor of soy foods such as soymilk and tofu (Kim, Kang, Cho, Choung, & Suh, 2005). Increased sucrose content in soybean seeds could effectively improve the sweetness of soy foods (Hou et al., 2009). In addition, seed sucrose content is negatively associated with yield of soymilk and tofu, simultaneously affects the hardness and firmness of tofu (Poysa, & Woodrow, 2002; Taira, Tanaka, Saito, & Saito, 1990). Sucrose plays an important role in natto fermentation and directly affects natto fermentation process (Taira et al., 1990). However, only a small percentage of soybean production is directly utilized in the manufacture of soy foods, and the demand for "food-grade" soybean seed continues to grow at an annual rate of 3–5% (Maughan, Maroof, & Buss, 2000). In addition to the significant nutritional value in seeds, sucrose is the major photosynthesis product of higher plants, which is not only the carbon base of physiological metabolism in plants, but also a signaling molecule that coordinates the relationship between plant sources and sinks. Additionally, sucrose participates in various metabolic and biosynthetic processes and plays a vital role in plant growth and seed development (Ruan, Jin, Yang, Li, & Boyer, 2010; Ruan, 2012). Furthermore, sucrose is involved in the responses to various abiotic stresses (Du et al., 2020). The role of sucrose in plants cannot be underestimated.

Sucrose content significantly varies among soybean varieties, and the sucrose content in cultivated and wild soybean seeds is 4–6% and 3–4%, respectively. Hymowitz, Dudley, Collins, and Brown (1974) indicated that sucrose content is negatively and positively correlated protein and oil content, respectively. Additionally, the environment and agronomic management (planting date and irrigated methods) significantly affect sucrose concentrations (Bellaloui et al., 2011). Genetic studies on the inheritance of sucrose concentration in soybean seeds are limited, thereby hampering breeding of new varieties with higher sucrose content. The application of the latest developments in molecular markers can be helpful in improving plant breeding efficiency (Paterson et al., 1991). By early 2020, only a few quantitative trait loci (QTLs) related to sucrose content were reported by linkage analysis according to the SoyBase database (www.soybase.org). Skoneczka, Maroof, Shang, and Buss (2009) identified one main effect QTL in the Sat_213-Satt643 interval of the chromosome 6 through two F_2 populations, which could explain 76% of the observed genetic variations in sucrose content. Kim et al. (2005) found four sucrose content-related QTLs in chromosomes 2, 11, and 19 using the RIL populations derived from a cross between 'Keunolkong' × 'Shinpaldalkong'. Kim, Kang, and Oh (2006) reported the two additional QTLs on chromosomes 12 and 16 using the same populations. Maughan et al. (2000) identified 17 sucrose content-related QTLs in chromosomes 5, 7, 8, 13, 15, 19, and 20 based on the RIL populations derived from a cross between V71-370 and PI 407162. Wang et al. (2014) reported one sucrose content-related QTL on chromosome 11 through a F_2 population from a cross between V97-3000 × V99-5089. Patil et al. (2018) identified four QTLs in chromosomes 6, 8, 16 and 20, and the QTL in chromosomes 8 (qSuc_08) was a major QTL for sucrose based on the RIL populations derived from a cross between *G. max* (Williams 82) and *G. soja* (PI 483460B).

As an alternative to linkage analysis, genome wide association study (GWAS) has been used to elucidate the molecular basis of complex traits in soybean due to that it significantly increases the range of phenotypic variations based on the accumulation of historic recombination events. Advances in genome sequencing have accelerated the resolution and accuracy of GWAS in soybean (Li et al., 2015). Presently, the molecular basis of few quality traits in soybean have been analyzed using GWAS, including protein, oil, and fatty acid content (Cao et al., 2017; Hwang et al., 2014). However, only a few studies that identified QTLs underlying sucrose content in soybean seeds have been conducted through GWAS (Patil et al., 2018).

In this study, we performed a GWAS based on 33,149 SNPs and 178 soybean germplasms to screen SNPs that related to the sucrose concentration. Then, analyzed genes located near the candidate regions around significant SNPs to verify the potential role of candidate genes related to sucrose concentration.

Core Ideas

- 178 elite accessions were genotyped and phenotyped for sucrose concentration.
- Five QTNs were identified for sucrose concentration.
- Haplotypes of candidate genes significantly associated with sucrose concentration were detected.
- Ten candidate genes related to sucrose concentration in soybean seeds were predicted.

2 | MATERIALS AND METHODS

2.1 | Planting and phenotyping

The 178 soybean germplasms collected (landraces and improved cultivars) (Supplemental Table S1) were grown under field conditions at Harbin, Gongzhuling and Shenyang in 2016. All accessions planted in the three locations were used for phenotypic evaluation and genotyping. Field experimental trials were performed with three replications and a randomized complete block design. The single row plots were 3-m in length with a row spacing of 0.65 m in all environments. Ten fully mature plants were randomly selected in each row to measure sucrose content.

2.2 | Measurement of sucrose content

Sucrose content in soybean seeds was extracted and quantified under with using the method of Teixeira, Ribeiro, Rezende, Barros, and Moreira (2012). Approximately 5 g of mature soybean seeds was milled to a fine powder and then dried at 105 °C for 5 h. Then, 20 mg of sample and 1 ml of 80% (v/v) ethanol were mixed in a 2-ml tube, vortexed for 1 min, and then placed in a 70 °C water bath for 90 min. The supernatant was transferred to a new tube after centrifugation for 10 min at 16,100*g* and topped up to 1.0 ml with 80% ethanol. The sucrose content in the sample was measured in a 96-well ELISA plate. Approximately 5 μl of alcohol extract from each sample, 85 μl of distilled water, and 10 μl of invertase (10 mg/ml) were placed in each well. The sealed plate was incubated at 55 °C for 10 min, followed by the addition of 200 μl of GOD (Bioclin kit) reagent into each well. The plate was sealed and then placed in a water bath at 37 °C for 15 min. Five minutes later and at room temperature, the absorbance at a wavelength of 490 nm was read using a Titertek Multiskan Plus spectrophotometer equipped for reading ELISA plates. The concentration of sucrose in each sample was determined by the standard curve. Each sample analysis was performed in triplicate.

2.3 | DNA extraction and sequencing

The cetyltrimethylammonium bromide (CTAB) method was used to extract DNA from fresh leaves of each sample. All soybean varieties were sequenced on basis of the extracted DNA using the specific locus amplified fragment sequencing (SLAF-seq) methodology (Sun et al., 2013). Digestive enzymes *MseI* (EC 3.1.21.4) and *HaeIII* (EC 3.1.21.4) (Thermo Fisher Scientific Inc., Waltham, MA, USA) were used to generate >50,000 sequencing tags (about 300–500 bp in length) of all tested accessions based on preliminary analysis of the soybean Williams 82 reference genome and were distributed in unique genomic regions of 20 soybean chromosomes. The sequencing libraries of each accession were defined based on the sequencing tags. The similarities and differences between the 45-bp sequence reads at both ends of the sequencing tag of each registered library and the soybean reference genome were assessed by SOAP2 software. The SLAF groups were defined using raw reads from the same genomic location through more than 58,000 SLAF tags in each test sample. The SNPs with minor allele frequencies (MAFs) ≥0.05 were defined. Genotypes where the depth of minor allele accounts for more than a third of the total depth of the sample were considered heterozygous genotypes.

2.4 | Population genetics analysis

The principle component analysis (PCA) approach was conducted to assess the population structure of the association panel through GAPIT software (Lipka et al., 2012). LD between pairs of SNPs (MAF ≥0.05 and missing data ≤10%) and r^2 (squared allele frequency correlations) was calculated via TASSEL 3.0 (Bradbury et al., 2007). In contrast to the GWAS, missing SNP genotypes were not imputed with the major allele before LD analysis. Parameter settings in the software program: MAF ≥0.05 and the integrity of each SNP ≥ 80%.

2.5 | Association analysis

The signals associated with the sucrose content in soybean seeds were identified using 33,149 SNPs from 178 tested accessions. The significance threshold for SNP trait associations was determined with compressed mixed linear model (CMLM) in GAPIT (Lipka et al., 2012) by $P < .001$ (Hwang et al., 2014; Yan et al., 2017), and this was used as the threshold to determine whether there was a meaningful signal that was associated with the sucrose content (Holm, 1979).

2.6 | Prediction and haplotype analysis of candidate genes for sucrose concentration

We selected genes within the 200-kb genomic region of significant SNPs as candidate genes, and the function of these genes was annotated according to the soybean reference genome (Wm82. a2. v1, http://www.soybase.org) (Cheng et al., 2017). Based on genomic re-sequencing data, genomic regions variations (exon regions, splice sites, untranslated regions [UTRs], intron regions, upstream and

TABLE 1 Statistical analysis and variation of sucrose concentration of association panel based on SAS 9.0 software

Location[a]	Min[b]	Max[c]	Mean	CV (%)[d]	Skewness	Kurtosis	F
Harbin	41.57	99.38	71.61	16.38	0.16	−0.25	86.69**
Gongzhuling	33.34	87.13	60.02	15.94	0.35	0.11	
Shenyang	37.08	84.57	58.00	17.51	0.47	−0.01	

[a]Harbin: at Harbin in 2016, Gongzhuling: at Gongzhuling in 2016, Shenyang: at Shenyang in 2016.
[b]Minimum.
[c]Maximum.
[d]Coefficient of variation.
**$P < .01$.

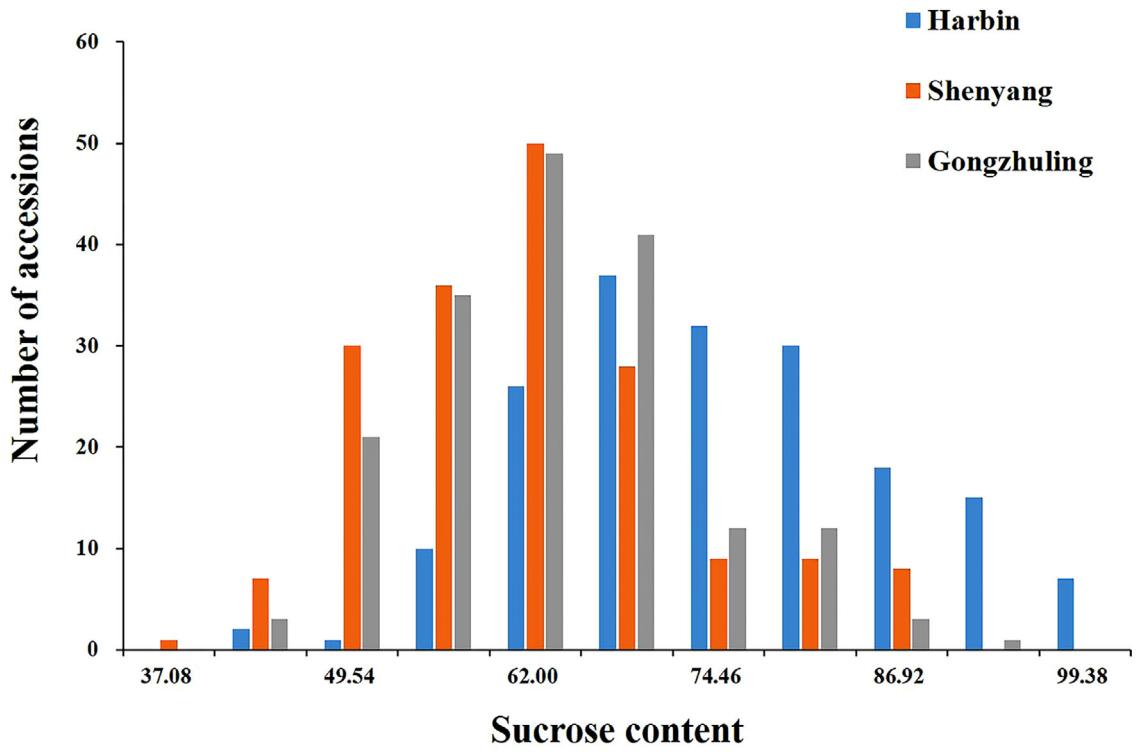

FIGURE 1 Variation of sucrose concentration among 178 accessions in three tested environments ('Harbin', 'Gongzhuling', 'Shenyang')

downstream regions) of candidate genes were detected in 16 lines and analyzed using the general linear model (GLM) method as implemented in TASSEL version 3.0 for identifying sucrose-related haplotypes (Bradbury et al., 2007). Significant SNPs affecting the target trait were asserted when the test statistics reached $P < .01$.

3 | RESULTS

3.1 | Statistical and variation analysis of sucrose concentration

The phenotypic value of 178 accessions in three locations were assessed (Supplemental Table S1), and the mean, standard deviation, coefficient of variation, skewness, and kurtosis of sucrose content were calculated, which showed a larger variation (Table 1; Figure 1). Phenotypic variation ranged from high concentration (99.38 mg/g) to low concentration (33.34 mg/g), and the average concentration among these three environments was 63.21 mg/g. The coefficients of variation were observed from 15.94% to 17.51%. Correspondingly, the kurtosis and skewness of the association panel observed in the three locations distributed normal without saliency, which indicated that the association panels were suitable for GWAS analysis. In addition, the sucrose content varied greatly among the three locations and showed extremely significant differences ($P < .01$) (Table 1).

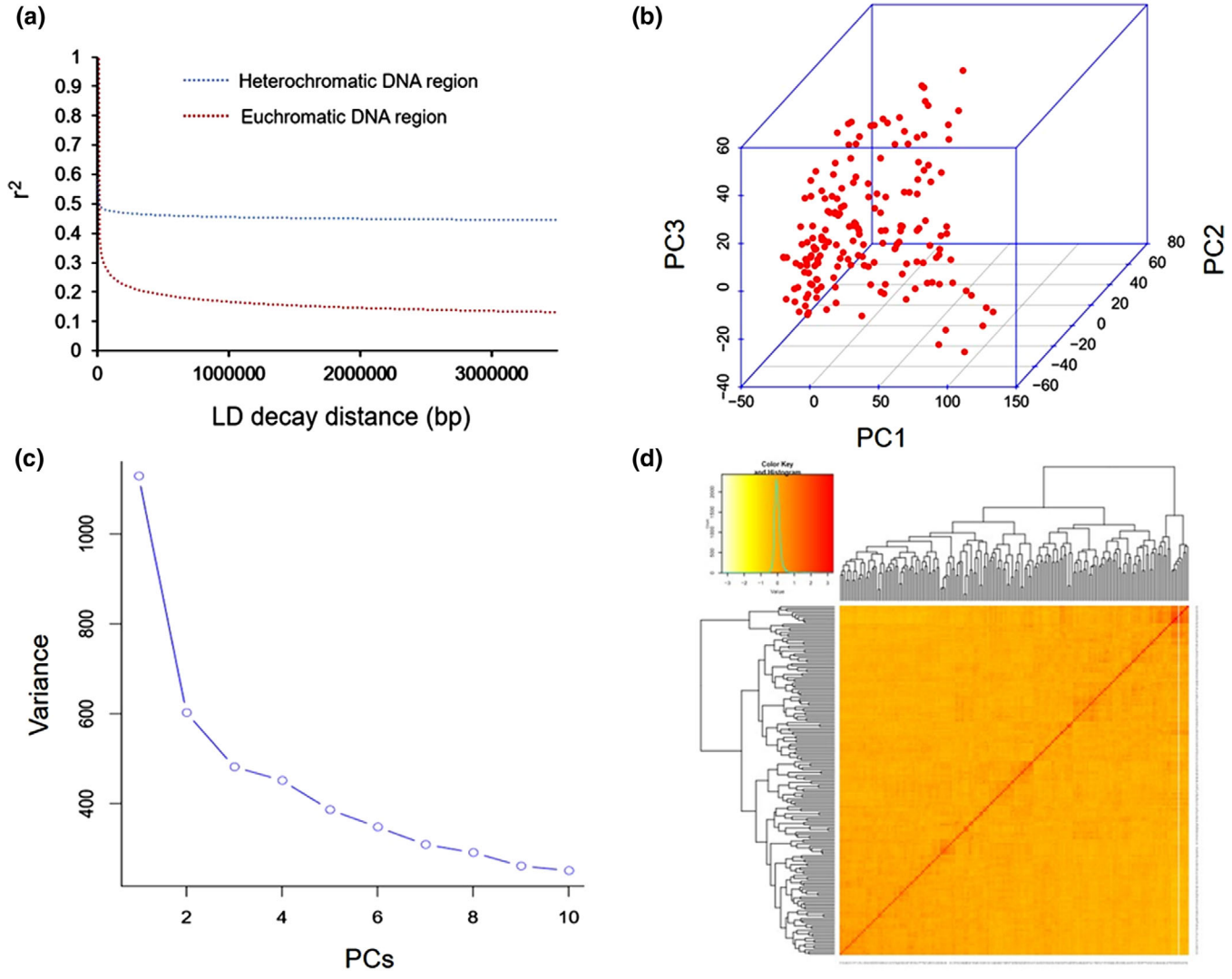

FIGURE 2 Mapping genetic data of populations. (a) The linkage disequilibrium (LD) decay of genome-wide association study (GWAS) population in heterochromatic regions and euchromatic regions. (b) The first three PCs of 33,149 SNPs used in the GWAS. (c) Population structure of soybean germplasms. (d) A heatmap of the kinship matrix of the 178 soybean accessions

3.2 | Sequencing and genotyping

The genomic DNA of the association panel was partially sequenced based on the SLAF-seq approach. A total of 33,149 high-quality markers (SNPs with MAF ≥ 0.05, missing data ≤ 10%) genotyped from the 178 accessions were exploited (Supplemental Table S2). These SNPs distributed on the 20 chromosomes of soybean genome that spanned approximately 951.4 Mbp, covering roughly 86.49% of the soybean genome (Supplemental Figure S1). Marker density was approximately 1 SNP per 28.7 kb, and the mean number of SNPs per chromosome was 1,657, which varied across different chromosomes, from 638 SNPs on chromosome 11 to 3,556 SNPs on chromosome 6 (Supplemental Table S3).

3.3 | Analysis of linkage disequilibrium (LD) and population structure

The average distance of LD decay in euchromatic DNA and the heterochromatic regions was analyzed to depict the mapping resolution for genome scans and GWA mapping. In euchromatic regions, the mean level of LD measured by r^2 dropped sharply to 0.2 at 387 kb. In heterochromatic regions, the LD pattern was greatly different from that in euchromatic regions (Figure 2a). The group stratifications of the association panels were scanned by the principal component (PC) and kinship analysis in view of 33,149 SNP markers. The results indicated that the first three PCs accounted for 14.01% of the genetic variation. The overall genetic variation declined drastically at PC2,

TABLE 2 Peak SNP and beneficial alleles associated with the sucrose concentration identified by GWAS

SNP	Chromosome	Position	Location	-log10(P)	MAF	Allele 1	Allele 2	Average sucrose content of accessions with allele 1	Average sucrose content of accessions with allele 2	Average sucrose content of population
rs2688589	4	2688589	Harbin	3.01	0.23	A	C	76.20	70.41	71.61
			Shenyang	3.50	0.23			63.12	56.57	58.00
			Gongzhuling	3.97	0.23			64.86	58.69	60.02
rs29026218	6	29026218	Harbin	3.14	0.18	T	G	73.05	64.80	71.61
			Shenyang	3.28	0.18			59.37	51.97	58.00
rs5926884	7	5926884	Gongzhuling	3.41	0.15	C	T	67.03	58.77	60.02
			Shenyang	3.01	0.15			66.04	56.53	58.00
rs6886889	11	6886889	Harbin	3.89	0.14	A	G	72.56	65.97	71.61
			Gongzhuling	3.41	0.14			60.76	55.49	60.02
rs10299216	12	10299216	Harbin	4.08	0.24	T	C	78.46	69.29	71.61
			Shenyang	3.96	0.24			64.10	56.03	58.00

and the inflection point arose at PC3 (Figure 2b and 2c). The results showed that the mapping population was predominated by the first three PCs. In addition, the results of pairwise relative kinship coefficients indicated lower levels of genetic relatedness among 178 soybean germplasms (Figure 2d).

3.4 | Quantitative trait nucleotides (QTNs) associated with sucrose concentration by GWAS

Five QTNs associated with sucrose content, distributed across five chromosomes, were identified on basis of the CMLM models in more than one tested environment (Table 2; Figure 3). Among these, one QTN (rs2688589 on chromosome 4) was identified in three environments simultaneously, and the other four QTNs (rs29026218, rs5926884, rs6886889, and rs10299216, located on chromosomes 6, 7, 11, and 12, respectively) were detected in two environments. The allelic effects of five significant SNP loci were analyzed to verify that whether these alleles are related to sucrose content. By comparison, it was found that the average sucrose concentration of all germplasms with allele 1 were significantly higher than all germplasms with allele 2 and the entire population (Table 2). The allelic effects of the five SNPs controlling sucrose content were analyzed. Among them, the direction of the contribution to the phenotype value of three SNPs, including rs2688589, rs5926884 and rs10299216, were opposite to the value of each allelic effect in all the three locations. Of the other two SNPs (rs29026218 and rs6886889), the allelic effect values were positive for sucrose content in the three locations (Figure 4). The results demonstrated that the differences of alleles in all QTNs had significant effects on sucrose concentration, which could be useful for assisted selection for soybeans with higher sucrose content germplasms.

3.5 | Prediction of candidate genes controlling sucrose concentration

A total of 60 genes, located within the 200-kb flanking regions of the five identified QTNs were considered as candidate genes. All of these genes were categorized into various functional groups on the basis of the Gene Ontology database (http://geneontology.org/) to further expound the potential functions of these genes. (Table 3; Supplemental Figure S2). Of these, two genes were classified as the protein of unknown function. The other 58 candidate genes were related to the cell wall, development, DNA, glycolysis, hormone metabolism, lipid metabolism,

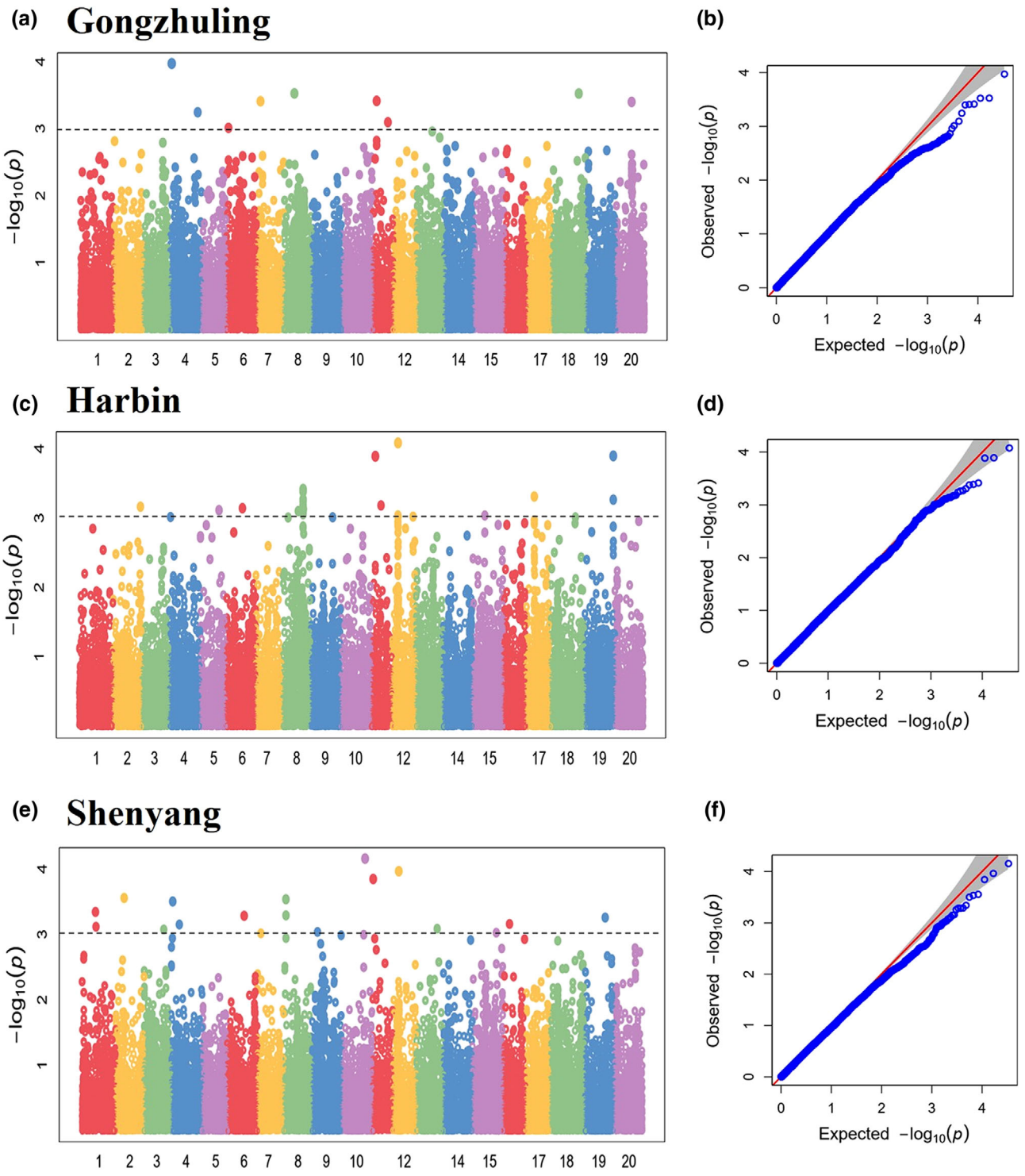

FIGURE 3 GWAS results of the sucrose concentration from 178 soybean accessions. a, c, e represent the Manhattan plot at 'Gongzhuling', 'Harbin', 'Shenyang', respectively. b, d, f represent the Q-Q plot at 'Gongzhuling', 'Harbin', 'Shenyang', respectively

metal handling, mitochondrial electron transport/ATP synthesis, N-metabolism, nucleotide metabolism protein, RNA, secondary metabolism, stress, and transport (Supplemental Figure S2). Among these, three genes directly involved in the process of sugar biosynthesis. *Glyma.04G032600* (located 99.132 kb near chromosome 4:2688589) (EC: 5.3.1.9), belonged to the sugar isomerase (SIS) family protein, as a 6-phosphate glucose isomerase, catalyzed the interconversion of D-glucopyranose 6-phosphate and β-D-fructofuranose 6-phosphate in the process of sucrose biosynthesis and sucrose degradation (Yu, Lue, Wang, & Chen, 2000). *Glyma.04G034600*

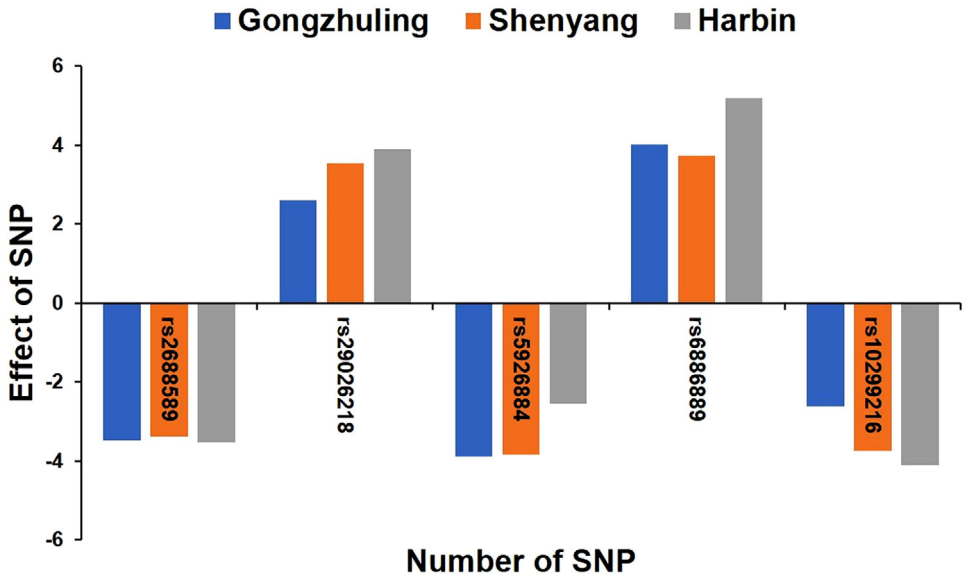

FIGURE 4 The allelic effects of the significant SNPs

(located 74.012 kb near chromosome 4:2688589) (EC: 2.4.1.43), a type of galacturonosyltransferase, participated in homogalacturonan biosynthesis, which catalyzed the conversion of a glycosyl group from UDP-a-D-galacturonate to 1,4-a-D-galacturonosyl (Sterling, Quigley, Orellana, & Mohnen, 2001). *Glyma.11G092100* (located 92.44 kb near chromosome 11:6886889), as the major facilitator superfamily protein, played a catalytic role in sugar transport (Pao, Paulsen, & Saier, 1998).

To further verify the potential effects of candidate genes for sucrose, gene-based association analysis was performed by the GLM method. A total of 215 SNPs in 60 candidate genes were identified among 16 lines (eight higher/lower sucrose lines) based on MAF $\geq$ 0.05 via genome resequencing. Finally, 35 SNPs from 10 genes (*Glyma.04G032600*, *Glyma.04G033000*, *Glyma.04G034700*, *Glyma.07G065400*, *Glyma.07G065500*, *Glyma.07G065600*, *Glyma.07G066100*, *Glyma.07G066400*, *Glyma.07G066600*, and *Glyma.07G066800*) were significantly associated with the sucrose concentration in more than two environments (Table 4). Among them, three nonsynonymous SNPs were detected from three genes (*Glyma.07G065400*, *Glyma.07G065500* and *Glyma.07G066100*), and the change of these three SNPs caused the changes of amino acid. Seven SNPs located in the upstream regions were identified from 5 genes, including 2 SNPs from *Glyma.04G032600*, 1 SNP from *Glyma.04G033000*, 1 SNP from *Glyma.04G034700*, 2 SNPs from *Glyma.07G065600* and 1 SNP from *Glyma.07G066600*. In total, 8 SNPs located in the downstream regions were identified, which were from four genes (1 SNP from *Glyma.04G033000*, 2 SNPs from *Glyma.07G066100*, 3 SNPs from *Glyma.07G066400*, 2 SNPs from *Glyma.07G066600*). Three SNPs were synonymous, which were detected from three genes (*Glyma.07G065400*, *Glyma.07G065500* and *Glyma.07G066800*). Two SNPs located in the UTR regions were detected from *Glyma.07G066800*. Other twelve SNPs were located in intronic the regions, which were detected from *Glyma.04G034700*, *Glyma.07G065400*, *Glyma.07G065500*, *Glyma.07G065600*, *Glyma.07G066100* and *Glyma.07G066800* (Table 4). Of these, *Glyma.04G032600* has been reported to be directly involved in sucrose biosynthesis in plants. Other genes could be considered as new genes in regulating the sucrose concentration of soybean. In addition, the effects of different alleles from peak SNPs of candidate genes were analyzed. The haplotype analysis showed that the sucrose content in soybean accessions with two different haplotypes among six genes (*Glyma.04G032600*, *Glyma.04G033000, Glyma.04G034700*, *Glyma.07G065500*, *Glyma.07G066100* and *Glyma.07G066400*) were significantly different ($P < .01$). Sucrose content of soybean accessions under different haplotypes among *Glyma.07G065400* showed significant difference at two of the three environments ($P < .05$). Sucrose content of soybean accessions with two different haplotypes among *Glyma.07G065600* were significantly different ($P < .01$ or $P < .05$). Sucrose content of soybean accessions under three different haplotypes among two genes (*Glyma.07G066600* and *Glyma.07G066800*) had significant difference (Figure 5). The results showed that there were significantly different in the sucrose concentration of soybean germplasms under different alleles of candidate genes. These beneficial alleles from candidate genes would be of interest and may contribute to MAS in soybeans with higher sucrose content.

TABLE 3 Gene models in the flanking regions of peak SNP

Peak SNP	Chr.	Physical position (bp)	Gene model	Distance to SNP(Kbp)	Functional annotation
rs2688589	4	2688589	Glyma.04G032600	99.13	Sugar isomerase (SIS) family protein
			Glyma.04G032700	88.13	GYF domain-containing protein
			Glyma.04G032800	78.59	Protein of unknown function (DUF1195)
			Glyma.04G033000	64.15	ovate family protein 7
			Glyma.04G033300	38.96	C2H2-like zinc finger protein
			Glyma.04G033500	25.03	J-domain protein required for chloroplast accumulation response 1
			Glyma.04G033600	13.95	Phospholipid/glycerol acyltransferase family protein
			Glyma.04G033700	8.16	DNA glycosylase superfamily protein
			Glyma.04G033800	0.89	Brassinosteroid signalling positive regulator (BZR1) family protein
			Glyma.04G034200	46.18	metalloendopeptidases;zinc ion binding
			Glyma.04G034300	55.72	thaumatin-like protein 3
			Glyma.04G034400	62.56	GNS1/SUR4 membrane protein family
			Glyma.04G034500	64.77	RING/U-box superfamily protein
			Glyma.04G034600	74.01	galacturonosyltransferase-like 2
			Glyma.04G034700	82.11	exocyst complex component sec3A
			Glyma.04G034800	96.65	Biotin/lipoate A/B protein ligase family
rs29026218	6	29026218	Glyma.06G222400	40.956	Transmembrane amino acid transporter family protein
			Glyma.06G222500	56.998	Zinc-binding dehydrogenase family protein
			Glyma.06G222600	88.031	Tetratricopeptide repeat (TPR)-like superfamily protein
			Glyma.06G222700	90.521	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
rs5926884	7	5926884	Glyma.07G065200	94.75	ATP binding;GTP binding;nucleotide binding;nucleoside-triphosphatases
			Glyma.07G065300	91.62	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
			Glyma.07G065400	80.93	Protein of unknown function (DUF674)
			Glyma.07G065500	76.11	NB-ARC domain-containing disease resistance protein
			Glyma.07G065600	48.18	NB-ARC domain-containing disease resistance protein
			Glyma.07G065700	3.27	ACC synthase 1
			Glyma.07G065800	10.42	Heavy metal transport/detoxification superfamily protein
			Glyma.07G066000	27.07	Erythronate-4-phosphate dehydrogenase family protein
			Glyma.07G066100	33.86	Homeodomain-like superfamily protein
			Glyma.07G066300	51.78	Erythronate-4-phosphate dehydrogenase family protein
			Glyma.07G066400	59.00	coenzyme Q biosynthesis Coq4 family protein / ubiquinone biosynthesis Coq4 family protein
			Glyma.07G066600	73.74	Protein kinase superfamily protein
			Glyma.07G066700	84.46	coenzyme Q biosynthesis Coq4 family protein / ubiquinone biosynthesis Coq4 family protein
			Glyma.07G066800	91.14	MAP kinase 4
			Glyma.07G066900	97.01	Pectin lyase-like superfamily protein
rs6886889	11	6886889	Glyma.11G089700	97.42	GDA1/CD39 nucleoside phosphatase family protein
			Glyma.11G089800	92.57	nodulin MtN21 /EamA-like transporter family protein

(Continues)

TABLE 3 (Continued)

Peak SNP	Chr.	Physical position (bp)	Gene model	Distance to SNP(Kbp)	Functional annotation
			Glyma.11G089900	86.25	nodulin MtN21 /EamA-like transporter family protein
			Glyma.11G090000	80.31	nodulin MtN21 /EamA-like transporter family protein
			Glyma.11G090200	45.58	ATP-binding cassette 14
			Glyma.11G090300	37.88	zinc knuckle (CCHC-type) family protein
			Glyma.11G090400	30.46	SWIB/MDM2 domain superfamily protein
			Glyma.11G090500	26.75	SWIB/MDM2 domain superfamily protein
			Glyma.11G090700	23.99	F-box family protein
			Glyma.11G090800	17.74	glycine/proline-rich protein
			Glyma.11G090900	12.47	RNAse l inhibitor protein 2
			Glyma.11G091000	1.98	xyloglucanase 113
			Glyma.11G091200	6.61	phosphatidylinositol synthase 1
			Glyma.11G091300	15.66	SBP (S-ribonuclease binding protein) family protein
			Glyma.11G091400	31.75	farnesylated protein 6
			Glyma.11G091500	35.32	Transducin family protein / WD-40 repeat family protein
			Glyma.11G091600	42.40	4-coumarate:CoA ligase 3
			Glyma.11G091700	51.12	P-loop containing nucleoside triphosphate hydrolases superfamily protein
			Glyma.11G091800	56.04	NADH-ubiquinone oxidoreductase-related
			Glyma.11G091900	66.34	Transcription factor DP
			Glyma.11G092000	86.34	GRAS family transcription factor
			Glyma.11G092100	92.44	Major facilitator superfamily protein
rs10299216	12	10299216	Glyma.12G109000	97.78	zinc ion binding;nucleic acid binding
			Glyma.12G109200	25.65	glutamine synthetase 2
			Glyma.12G109500	8.81	Protein phosphatase 2C family protein

4 | DISCUSSION

The sucrose in soybean seeds is one of the desirable traits for taste and flavor of soybean-derived food, and affects the feeding efficiency of soymeal. Sucrose has the highest soluble sugars in soybean seeds, accounting for 5% of total carbohydrates (Smith & Circle, 1972; Wilson, 2004; Mozzoni, Shi, & Chen, 2013). Among the three main components of soluble sugar (sucrose, raffinose and stachyose), sucrose is the easiest to digest and absorb by monogastric animals (Peterbauer & Richter, 2001; Smith & Circle, 1972; Wilson, 2004; Mozzoni et al., 2013; Lunn, 2008). Therefore, breeding soybean lines with high sucrose concentration was of considerable significance for breeding in soybean. Major seed composition characteristics such as protein, oil, and sucrose content are significantly affected by the environment (Bandillo et al., 2015; Chaudhary et al., 2015; Thomas, Boote, Allen, Gallo-Meagher, & Davis, 2003). In current study, 178 cultivated soybeans from three environments (Supplemental Table S1) were collected mainly in China and assessed in terms of sucrose content to determine its correlation. The sucrose content at three locations showed extremely significant variation ($P < .01$) (Table 1), likely due to the diversity of temperatures among three locations, including temperatures during plant growth or seed development and the accumulated temperature (Zeng et al., 2014). Other studies showed that increased temperature reduced the sucrose level (Hou et al., 2009; Wolf, Cavins, Kleiman, & Black, 1982). Evidently, the sugar content was influenced by environmental conditions, justifying the significance of multiple locations for detecting sugar concentration.

To date, only 37 QTLs related with sucrose concentration have been reported (http://www.soybase.org) on the basis of linkage analysis. Identification of candidate loci associated with sucrose concentration in soybean *via* GWAS is an effective method. Patil et al. (2018) identified a major QTL on chromosome 8 (qSuc_08) through a population derived from the cross of *G. max* (Williams 82) and *G. soja* (PI 483460B) with more than 91,000 SNPs, which harbored putative genes involved in sugar transport. Additionally, GWAS analysis was performed to verify the major QTLs.

TABLE 4 Haplotype analysis of candidate genes

Gene ID	Chromosome	Physical position (bp)	Location	Region	alleles	-log10 (P)	Functional annotation
Glyma.04G032600	4	2588648	Gongzhuling	upstream	A/C	2.10	Sugar isomerase (SIS) family protein
			Shenyang			2.47	
		2588796	Gongzhuling	upstream	A/C	2.25	
			Shenyang			2.08	
Glyma.04G033000	4	2625734	Gongzhuling	downstream	A/T	2.11	ovate family protein 7
			Shenyang			2.22	
		2649062	Gongzhuling	upstream	A/G	2.48	
			Shenyang			2.43	
Glyma.04G034700	4	2770408	Gongzhuling	upstream	G/A	2.61	exocyst complex component sec3A
			Shenyang			2.99	
		2772734	Gongzhuling	intronic	C/T	2.29	
			Shenyang			2.37	
Glyma.07G065400	7	5848103	Gongzhuling	nonsynonymous	G/T	1.90	Protein of unknown function (DUF674)
			Harbin			2.04	
			Shenyang			2.15	
		5848125	Gongzhuling	synonymous	A/C	2.17	
			Harbin			2.16	
			Shenyang			2.34	
		5848397	Gongzhuling	intronic	A/C	2.13	
			Harbin			1.64	
			Shenyang			2.24	
Glyma.07G065500	7	5857497	Gongzhuling	intronic	G/C	2.83	NB-ARC domain-containing disease resistance protein
			Harbin			2.79	
			Shenyang			2.90	
		5858061	Gongzhuling	intronic	A/G	2.17	

(Continues)

TABLE 4 (Continued)

Gene ID	Chromosome	Physical position (bp)	Location	Region	alleles	-log10 (P)	Functional annotation
			Harbin			2.16	
			Shenyang			2.34	
		5858126	Gongzhuling	intronic	G/A	2.80	
			Harbin			2.28	
			Shenyang			2.88	
		5867023	Gongzhuling	synonymous	T/A	2.13	
			Harbin			2.59	
			Shenyang			2.31	
		5867040	Gongzhuling	nonsynonymous	G/A	2.26	
			Harbin			2.43	
			Shenyang			2.03	
Glyma.07G065600	7	5877918	Gongzhuling	upstream	G/C	2.12	NB-ARC domain-containing disease resistance protein
			Harbin			2.08	
			Shenyang			1.97	
		5877969	Gongzhuling	upstream	C/T	2.36	
			Harbin			1.99	
			Shenyang			2.11	
		5881608	Gongzhuling	intronic	T/G	2.64	
			Harbin			1.92	
			Shenyang			2.01	
Glyma.07G066100	7	5960972	Gongzhuling	nonsynonymous	C/T	2.21	Homeodomain-like superfamily protein
			Harbin			2.20	
			Shenyang			2.69	
		5961998	Gongzhuling	intronic	G/A	2.41	
			Harbin			2.07	
			Shenyang			2.83	

(Continues)

TABLE 4 (Continued)

Gene ID	Chromosome	Physical position (bp)	Location	Region	alleles	-log10 (P)	Functional annotation
		5962677	Gongzhuling	intronic	C/T	2.03	
			Harbin			2.03	
			Shenyang			2.51	
		5967336	Gongzhuling	downstream	G/A	2.24	
			Harbin			2.20	
			Shenyang			3.09	
		5967355	Gongzhuling	downstream	G/A	2.48	
			Harbin			2.38	
			Shenyang			3.15	
Glyma.07G066400	7	5985000	Gongzhuling	downstream	T/C	3.08	coenzyme Q biosynthesis Coq4 family protein / ubiquinone biosynthesis Coq4 family protein
			Harbin			2.17	
			Shenyang			3.00	
		5985221	Gongzhuling	downstream	C/G	3.63	
			Harbin			2.13	
			Shenyang			2.89	
		5986532	Gongzhuling	downstream	A/C	2.48	
			Harbin			1.91	
			Shenyang			2.39	
Glyma.07G066600	7	5999812	Gongzhuling	upstream	C/T	2.04	Protein kinase superfamily protein
			Shenyang			2.29	
		6004431	Gongzhuling	downstream	T/C	2.59	
			Shenyang			2.64	
		6004796	Gongzhuling	downstream	G/A	1.90	
			Shenyang			2.06	
Glyma.07G066800	7	6018200	Gongzhuling	UTR5	T/G	2.21	MAP kinase 4
			Harbin			2.20	
			Shenyang			2.69	

(Continues)

TABLE 4 (Continued)

Gene ID	Chromosome	Physical position (bp)	Location	Region	alleles	-log10 (P)	Functional annotation
		6021161	Gongzhuling	intronic	C/G	2.27	
			Harbin			2.44	
			Shenyang			2.55	
		6021778	Gongzhuling	intronic	T/A	2.45	
			Harbin			2.68	
			Shenyang			3.14	
		6022066	Gongzhuling	intronic	A/G	3.20	
			Harbin			3.13	
			Shenyang			3.44	
		6022276	Gongzhuling	intronic	C/T	2.66	
			Harbin			2.94	
			Shenyang			3.05	
		6023539	Gongzhuling	UTR3	T/A	2.00	
			Harbin			2.01	
			Shenyang			2.42	
		6028854	Gongzhuling	synonymous	C/T	2.11	
			Harbin			2.09	
			Shenyang			2.40	

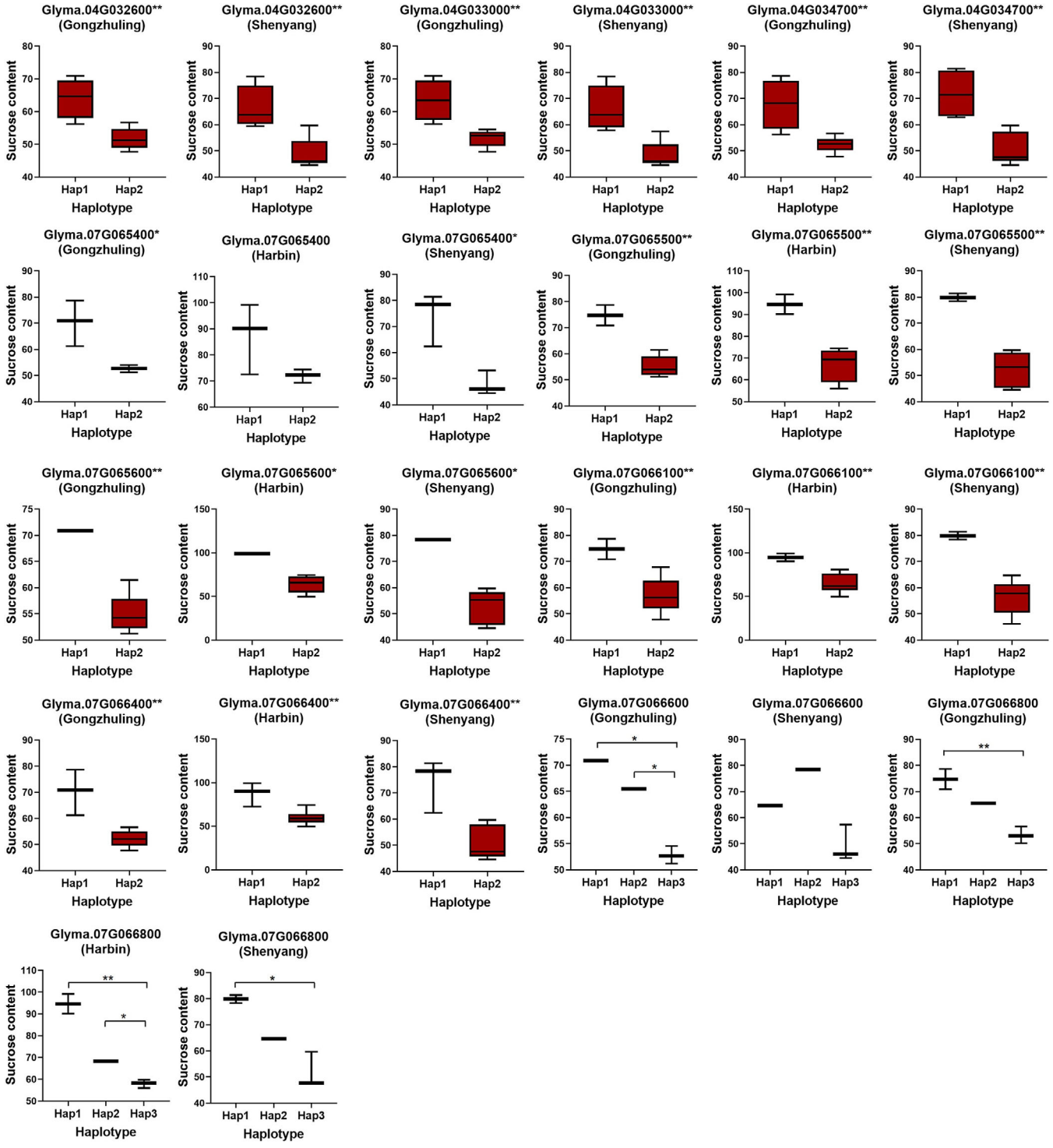

FIGURE 5 Haplotypes analysis of genes related to sucrose concentration. Significance of ANOVA: $*P < .05$, $**P < .01$

In this study, five QTNs related to sucrose concentration were identified across five chromosomes (chromosomes 4, 6, 7, 11, and 12) in more than one environment. Among these, the SNP loci chromosome 4:2688589 located could be detected in all three environments and the other four loci can be detected in two environments. All of the five QTNs had no overlap region with QTLs that were previously detected and thus could be considered as novel loci associated with sucrose concentration. In addition, based on the detected QTNs, the sucrose content of germplasms with exceptional alleles was higher than those of other tested germplasms with different alleles. Therefore, these

identified QTNs are more likely to control sucrose content in soybean.

Currently, about 362 candidate genes were identified to participate in the regulation of sucrose content in soybean (http://www.soybase.org). GWAS has become the main choice for mining and identifying important genes due to the relatively short LD fragments (Li et al., 2015). In this study, a total of 60 candidate genes were screened in 200-kb genomic regions of the five peak SNPs. Three genes, including two genes located near chromosome 4:2688589 (*Glyma.04G032600*, *Glyma.04G034600*) and 1 gene located near chromosome 11:6886889 (*Glyma.11G092100*) were demonstrated to directly influence anabolic metabolism of sucrose in soybean (Pao et al., 1998; Yu et al., 2000). Gene-based association and haplotype analyses were conducted to verify the accuracy of candidate genes. As a result, 10 candidate genes (*Glyma.04G032600*, *Glyma.04G033000*, *Glyma.04G034700*, *Glyma.07G065400*, *Glyma.07G065500*, *Glyma.07G065600*, *Glyma.07G066100*, *Glyma.07G066400*, *Glyma.07G066600*, and *Glyma.07G066800*) were detected and nine beneficial haplotypes were identified in different environments. The results confirmed once again that *Glyma.04G032600* has definite effects in regulating sucrose content. Hence, the value of this gene for increasing the concentration of sucrose in soybean was worth exploring in the future. The other nine genes can also be labeled as potential candidate genes for regulating the sucrose content in soybean and should be discussed and verified in further studies.

5 | CONCLUSION

A total of 60 genes located near five peak SNPs were detected to be associated with sucrose concentration based on 178 diverse soybean accessions in three environments and 33,149 SNPs by GWAS. Through gene-based association analysis and haplotype analysis, one gene with known definite effects in regulating sucrose content and nine novel genes were detected. The beneficial alleles and candidate genes identified may be helpful for improving the concentration of sucrose in soybean seed.

ACKNOWLEDGMENTS

This study was financially supported by the Heilongjiang Provincial Project (GX17B002, JC2018007, SC2019ZX16B0039, GJ2018GJ0098, C2018016), the Chinese National Natural Science Foundation (31701446, 31871650, 31971967, 31671717, 31471517), National Key R & D Project (2017YFD0101306, 2016YFD0100304, 2017YFD0101302), the National Project (2014BAD22B01, 2016ZX08004001-007), the Youth Leading Talent Project of the Ministry of Science and Technology in China (2015RA228), The National Ten-thousand Talents Program (W03020275), Postdoctoral Fund in Heilongjiang Province (LBH-Z15017, LBH-Q17015), The National Project (CARS-04-PS04), The 'Youth Innovation Talent' Project of the general undergraduate universities in Heilongjiang province (UNPYSCT-2016145), the 'Academic Backbone' Project of Northeast Agricultural University (17XG22).

AUTHOR CONTRIBUTIONS

MNS, YW, and YYB conceived the study and contributed to population development. WLT and WBL conducted genotyping. XW, RQL, YL, MY, CQ, and CXL contributed to phenotypic evaluation. XZ and YPH contributed to the experimental design and writing paper. All authors contributed to and approved the final manuscript.

CONFLICT OF INTEREST DISCLOSURE

The authors declare that there is no conflict of interest.

ORCID

Yingpeng Han https://orcid.org/0000-0002-9829-6588

REFERENCES

Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J., & Lorenz, A. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *The Plant Genome*, *8*, 1–13. https://doi.org/10.3835/plantgenome2015.04.0024.

Bellaloui, N., Ebelhar, M. W., Gillen, A. M., Fisher, D. K., Abbas, H. M., Mengisu, A., … Paris, R. L. (2011). Soybean seed protein, oil, and fatty acids are altered by S and S + N fertilizers under irrigated or non-irrigated environments. *Agricultural Sciences*, *2*, 465–476. https://doi.org/10.4236/as.2011.24060.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308.

Cao, Y., Li, S., Wang, Z., Chang, F., Kong, J., & Gai, J., & Zhao, T. (2017). Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. *Frontiers in Plant Science*, *8*, 1222. https://doi.org/10.3389/fpls.2017.01222.

Chaudhary, J., Patil, G. B., Sonah, H., Deshmukh, R. K., Vuong, T. D., Valliyodan, B., & Nguyen, H. T. (2015). Expanding omics resources for improvement of soybean seed composition traits. *Frontiers in Plant Science*, *6*, 1021. https://doi.org/10.3389/fpls.2015.01021.

Cheng, P., Gedling, C. R., Patil, G., Vuong, T. D., Shannon, J. G., Dorrance, A. E., & Nguyen, H. T. (2017). Genetic mapping and

haplotype analysis of a locus for quantitative resistance to *Fusarium graminearumin* soybean accession PI 567516C. *Theoretical and Applied Genetics*, *130*, 999–1010. https://doi.org/10.1007/s00122-017-2866-8.

Du, Y., Zhao, Q., Chen, L., Yao, X., Zhang, W., Zhang, B., & Xie, F. (2020). Effect of drought stress on sugar metabolism in leaves and roots of soybean seedlings. *Plant Physiology and Biochemistry*, *146*, 1–12. https://doi.org/10.1016/j.plaphy.2019.11.003.

Hwang, E. Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., & Cregan, P. B. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics*, *15*, 1. https://doi.org/10.1186/1471-2164-15-1.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70. https://doi.org/10.2307/4615733.

Hou, A., Chen, P., Alloatti, J., Li, D., Mozzoni, L., Zhang, B., & Shi, A. (2009). Genetic variability of seed sugar content in worldwide soybean germplasm collections. *Crop Science*, *49*, 903. https://doi.org/10.2135/cropsci2008.05.0256.

Hymowitz, T. J., Dudley, J. W., Collins, F. I., & Brown, C. M. (1974). Estimations of protein and oil concentration in corn, soybean, and oat seed by near-infrared light reflectance. *Crop Science*, *14*, 713. https://doi.org/10.2135/cropsci1974.0011183X0014000 50031x.

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.

Kim, H. K., Kang, S. T., & Oh, K. W. (2006). Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of *Glycine max* seeds. *Journal of Plant Research*, *119*, 533–538. https://doi.org/10.1007/s10265-006-0004-9.

Krober, O. A., & Cartter, J. L. (1962). Quantitative interrelations of protein and nonprotein constituents of soybeans. *Crop Science*, *2*, 171–172. https://doi.org/10.2135/cropsci1962.0011183X000200020028x.

Li, Y., Reif, J. C., Ma, Y., Hong, H., Liu, Z., Chang, R., & Qiu, L. (2015). Targeted association mapping demonstrating the complex molecular genetics of fatty acid formation in soybean. *BMC Genomics*, *16*, 841. https://doi.org/10.1186/s12864-015-2049-4.

Lipka, A. E., Feng, T., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., … Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, *28*, 2397–2399. https://doi.org/10.1093/bioinformatics/bts444.

Lunn, J. E. (2008). Sucrose metabolism. In *Encyclopedia of life sciences*. Chichester: Widely Ltd. https://doi.org/10.1002/9780470015902.a0021259.

Maughan, P. J., Maroof, M. A. S., & Buss, G. R. (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.

Mozzoni, L., Shi, A., & Chen, P. (2013). Genetic analysis of high sucrose, low raffinose, and low stachyose content in V99-5089 soybean seeds. *Journal of Crop Improvement*, *27*, 606–616. https://doi.org/10.1080/15427528.2013.812998.

Pao, S. S., Paulsen, I. T., & Saier, M. H. (1998). Major facilitator superfamily. *Microbiology and Molecular Biology Reviews*, *62*, 1–34. https://doi.org/10.1128/mmbr.62.1.1-34.1998.

Paterson, A. H., Tanksley, S. D., & Sorrells, M. E. (1991). DNA markers in plant improvement. *Advances in Agronomy*, *46*, 39–90. https://doi.org/10.1016/S0065-2113(08)60578-7.

Patil, G., Vuong, T. D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., … Nguyen, H. T. (2018). Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnology Journal*, *16*, 1939–1953. https://doi.org/10.1111/pbi.12929.

Poysa, V., & Woodrow, L. (2002). Stability of soybean seed composition and its effect on soymilk and tofu yield and quality. *Food Research International*, *35*, 337–345. https://doi.org/10.1016/S0963-9969(01)00125-9.

Peterbauer, T., & Richter, A. (2001). Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Science Research*, *11*, 185–197. https://doi.org/10.1079/SSR200175.

Ruan, Y. L. (2012). Signaling role of sucrose metabolism in development. *Molecular Plant*, *5*, 763–765. https://doi.org/10.1093/mp/sss046.

Ruan, Y. L., Jin, Y., Yang, Y. J., Li, G. J., & Boyer, J. S. (2010). Sugar input, metabolism, and signaling mediated by invertase: Roles in development, yield potential, and response to drought and heat. *Molecular Plant*, *3*, 942–955. https://doi.org/10.1093/mp/ssq044.

Skoneczka, J. A., Maroof, M. A. S., Shang, C., & Buss, G. R. (2009). Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. *Crop Science*, *49*, 247. https://doi.org/10.2135/cropsci2008.07.0403.

Smith, A. K., & Circle, S. J. (1972). Chemical composition of the seed. In Soybeans: Chemistry and Technology. Vol. 1, Proteins (pp. 61–92). Westport: Avi Publishing.

Sterling, J. D., Quigley, H. F., Orellana, A., & Mohnen, D. (2001). The catalytic site of the pectin biosynthetic enzyme alpha-1,4-galacturonosyltransferase is located in the lumen of the golgi. *Plant Physiology*, *127*, 360–371. https://doi.org/10.1104/pp.127.1.360.

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., … Zeng, H. (2013). SLAF-seq: An efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLOS One*, *8*, e58700. https://doi.org/10.1371/journal.pone.0058700.

Taira, H., Tanaka, H., Saito, M., & Saito, M. (1990). Effect of cultivar, seed size, and crop year on total and free sugar contents of domestic soybeans. *Nippon Shokuhin Kogyo Gakkaishi*, *37*, 203–213. https://doi.org/10.3136/nskkk1962.37.3_203.

Teixeira, A. I., Ribeiro, L. F., Rezende, S. T., Barros, E. G., & Moreira, M. A. (2012). Development of a method to quantify sucrose in soybean grains. *Food Chemistry*, *130*, 1134–1136. https://doi.org/10.1016/j.foodchem.2011.07.128.

Thomas, J. M. G., Boote, K. J., Allen, L. H., Gallo-Meagher, M., & Davis, J. M. (2003). Elevated temperature and carbon dioxide effects on soybean seed composition and transcript abundance. *Crop Science*, *43*, 1548. https://doi.org/10.2135/cropsci2003.1548.

Wang, Y., Chen, P., & 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.

Wilson, R. F. (2004). Seed composition. *Soybeans Improvement*, *16*. https://doi.org/10.2134/agronmonogr16.3ed.c13.

Wolf, R. B., Cavins, J. F., Kleiman, R., & Black, L. T. (1982). Effect of temperature on soybean seed constituents: Oil, protein, moisture, fatty acids, amino acids and sugars. *Journal of the American Oil Chemists' Society*, *59*, 230–232. https://doi.org/10.1007/BF02582182.

Yan, L., Hofmann, N., Li, S., Ferreira, M. E., Song, B., Jiang, G., … Song, Q. (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. https://doi.org/10.1186/s12864-017-3922-0.

Yu, T. S., Lue, W. L., Wang, S. M., & Chen, J. (2000). Mutation of Arabidopsis plastid phosphoglucose isomerase affects leaf starch synthesis and floral initiation. *Plant Physiology*, *123*, 319–326. https://doi.org/10.1104/pp.123.1.319.

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., … Cregan, P. (2014). Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Science*, *54*, 554. https://doi.org/10.2135/cropsci2013.01.0036.

SUPPORTING INFORMATION

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

How to cite this article: Sui M, Wang Y, Bao Y, et al. Genome-wide association analysis of sucrose concentration in soybean (*Glycine max* L.) seed based on high-throughput sequencing. *Plant Genome*. 2020;13:e220059. https://doi.org/10.1002/tpg2.20059