

Genome-wide association study of sugar content and development of diagnostic KASP markers in peanut (*Arachis hypogaea* L.)

Jingjing Deng

Hebei Agricultural University

Mingyu Hou

Hebei Agricultural University

Shunli Cui

Hebei Agricultural University

Yingru Liu

Hebei Agricultural University

Hongtao Deng

Hebei Agricultural University

Xiukun Li

Hebei Agricultural University

Xiaoshu Chen

Peanut Institute, Jilin Academy of Agricultural Sciences

Lifeng Liu

lifengliu@cau.edu.cn

Hebei Agricultural University

Research Article

Keywords: peanut, sugar content, genome-wide association study, SNP, KASP

Posted Date: February 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8599999/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Sugar content is a critical factor influencing the quality of peanut kernels, determining the sweetness and taste of peanuts, which in turn affects consumer acceptance and market value. However, the genetic mechanisms underlying sugar content in peanut kernels remain poorly understood.

Results

In this study, GWAS analysis was conducted using 103 peanut germplasm resources with significant differences in total sugar content (TSC), soluble sugar content (SSC), and sucrose content (SC). The variation ranges for the three traits were 6.48%~26.10%, 5.22%~17.29%, and 2.01%~7.97%, respectively. Based on genome-wide association analysis, 98, 122, and 148 SNPs significantly associated with TSC, SSC, and SC were identified, respectively. Among them, 42 SNPs on chromosome A05 contributed 9.60%~20.29% phenotypic variance to SC. Further analysis detected 21 pleiotropic SNPs, including 2 SNPs significantly associated with both TSC and SC, and 18 SNPs significantly associated with both SSC and SC. Among the 21 multi-environment co-localized SNPs, 6 SNPs were consistently detected across three environments, with phenotypic explanation rates ranging from 10.03% to 20.29%. Based on the genome-wide association analysis results, within the decay region of two significant SNPs (AX-176820713 and AX-176816968), one SWEET gene (*Arahy.78PE2K*) and two STP genes (*Arahy.KFDF05* and *Arahy.IG4CN6*) were identified as participating in sugar metabolism. qRT-PCR results showed that the expression levels of *Arahy.78PE2K* and *Arahy.KFDF05* were significantly upregulated in high-sugar varieties, while the expression level of *Arahy.IG4CN6* was significantly downregulated. The KASP marker *qK-SCB03.1*(AX-176816968), developed based on the genome-wide association analysis and qRT-PCR results, can distinguish germplasm sugar content types, providing an important theoretical basis and technical support for peanut sugar content improvement and molecular marker-assisted breeding.

Conclusions

This study provides comprehensive insights into the identification and application of SNPs, candidate genes, and KASP molecular markers associated with peanut sugar content. It offers valuable theoretical support and genetic resources for molecular breeding projects aimed at improving peanut seed sugar content and developing high-sugar peanut varieties.

1 Introduction

Peanut (*Arachis hypogaea* L.), also known as groundnut, is a nutrient-rich oil and cash crop. In addition to being used for oil extraction, edible peanuts account for 40% of the total peanut production, with this proportion showing a year-on-year increasing trend^[1]. The level of cellulose is directly related to the

texture of peanuts. Soluble sugars are the primary factors influencing the quality and flavor of edible peanuts. Peanut soluble sugars include sucrose, glucose, fructose, and trehalose ^[2]. Glucose and fructose, as the predominant monosaccharides in peanut kernels, also play a crucial role in the flavor development of peanuts. During the peanut roasting process, the Maillard reaction occurring between these monosaccharides and free amino acids not only generates the characteristic aroma compounds of roasted peanuts but also imparts their unique nutty flavor profile. This series of complex chemical reactions and the formation of flavor compounds endow roasted peanuts with a distinct flavor quality that differentiates them from other processing methods ^[3–4]. Through correlation analysis, it was found that there exists a significant positive correlation between the sweetness and taste quality of peanuts, with a correlation coefficient reaching 0.88. This fully illustrates the importance of sucrose content in the quality evaluation of peanuts ^[5]. Moreover, when the sucrose content exceeds 6%, peanuts generally exhibit a superior taste ^[6].

The synthesis and metabolism of sugar content in peanut kernels are controlled by multiple enzymes, representing a quantitative trait governed by polygenes ^[5]. In previous studies, researchers have utilized recombinant inbred line (RIL) populations in conjunction with BSA-seq (Bulked Segregant Analysis Sequencing), genetic linkage mapping, and molecular marker techniques to localize quantitative trait locus (QTL) loci controlling key traits such as sucrose content and soluble sugar metabolism ^[7–9]. Natural populations offer several advantages in plant genetic mapping studies, including high genetic diversity, the elimination of the need to construct specific populations, broad environmental adaptability, and significant potential for mining superior alleles. Currently, genome-wide association study (GWAS) stands as one of the primary strategies for crop breeding improvement and has been applied in crops such as maize, wheat, and cotton ^[10–13]. In the context of peanut trait improvement and breeding, GWAS has been predominantly employed to identify loci associated with yield traits, plant architecture traits, and leaf spot disease resistance in peanut ^[14–16]. Regarding quality trait improvement, GWAS has been frequently utilized to identify variant loci linked to oil content and protein traits ^[17–18]. However, GWAS studies specifically targeting sugar content in peanuts have been rarely reported.

As a derivative method of fluorescence quantitative PCR, the kompetitive allele specific PCR (KASP) marker technology demonstrates characteristics of simple operation, rapid reaction, high throughput, and precise results ^[19]. KASP markers associated with sugar content have been developed in soybean, wheat, and potato ^[20–22]. In the field of crop breeding, the development of KASP markers has also been applied to distinguish drought tolerance in soybeans, disease resistance in wheat, and bacterial wilt resistance in cotton ^[23–25]. In peanut research, a series of KASP molecular markers associated with yield traits, seed weight, and quality have been developed ^[26–29]. Regarding the development of molecular markers for sugar content in peanuts, Li *et al.* and Zhang *et al.* have respectively developed Indel markers on chromosomes B08 and B06, which are associated with sucrose and soluble sugar content ^[30–31].

This study conducted a genome-wide association analysis on the sugar content of peanut kernels using 103 peanut germplasm resources. The objectives were to elucidate the genetic mechanisms underlying sugar content in peanut kernels, identify superior variant loci, develop relevant KASP markers, and mine candidate genes associated with sugar metabolism. These efforts hold significant importance for breeding peanuts with enhanced sugar content in their kernels.

2 Materials and Methods

2.1 Experimental Materials

The 103 peanut germplasm resources used in this study were sourced from the peanut genetic breeding innovation team at Hebei Agricultural University. These germplasms were planted in May 2021 and May 2022 at the Qingyuan Experimental Base of Hebei Agricultural University in Baoding City, Hebei Province (E115.48', N38.85'). In May 2023, the germplasms were planted at the Xushui Experimental Base in Baoding City, Hebei Province (E115°13', N39°06') (designated as 2021BD, 2022BD, and 2023BD). A completely randomized block design was adopted, with each material replicated twice. Ten hills were sown per row, with one seed per hill. The row length was 1.5 m, and the row spacing was 40 cm. Conventional field management practices were followed. Harvesting was conducted based on the growth periods of each germplasm in September 2021, September 2022, and September 2023, respectively. The harvested materials were naturally air-dried and subsequently used for the determination of total sugar content, soluble sugar content, and sucrose content in subsequent genome-wide association analyses.

In May 2024, the high-sugar variety "Jinongxian 1" and the low-sugar variety "Silihong" were planted at the Qingyuan Experimental Base in Baoding City, Hebei Province, under conventional field management. Sampling was performed at 70, 80, 90, and 100 days after flowering (DAF70, DAF80, DAF90, DAF100), with 10 seeds collected per time point and three replicates per sample. After rapid freezing in liquid nitrogen, the samples were stored at -80°C for subsequent validation experiments on the expression levels of candidate genes.

2.2 Phenotypic Data Measurement

The total sugar content, soluble sugar content, and sucrose content in mature peanut kernels were determined using near-infrared models for these traits, which had been previously established in our laboratory^[32]. After warming up the instrument for 30 minutes, mature, plump, and undamaged seeds from each material were placed into a sample cup, ensuring the surface of the cup was level. Each sample was scanned three times, and the loading process was repeated three times. The average values obtained were used for genome-wide association analysis. During peanut development, the soluble sugar content was measured following Chen's anthrone colorimetric method^[32]. For sucrose content determination, a sucrose content detection kit (anthrone colorimetric method) (Tiangen Biotech (Beijing) Co., Ltd., Catalog No.: G0506W) was employed for extraction.

2.3 Genome-Wide Association Analysis (GWAS)

Leveraging 17,223 SNPs obtained from resequencing conducted in our laboratory^[33], quality control was performed on these SNPs using TASSEL 5.0 software. Low-quality SNPs were filtered out based on the following criteria: call rate < 0.95 and minor allele frequency (MAF) < 0.05. Ultimately, 12,342 high-quality SNPs were retained for further analysis. Association analysis between the SNPs and the traits of total sugar content, soluble sugar content, and sucrose content was conducted using the general linear model (GLM) in TASSEL 5.0 software. SNP loci with a $-\log_{10}P$ value ≥ 3 were defined as significantly associated with the respective traits. Manhattan plots and Q-Q plots were generated using the CMplot package in R software^[34].

2.4 Candidate Gene Identification

Based on the results of the genome-wide association analysis and in accordance with the linkage disequilibrium (LD) decay distance, candidate genes were screened within a 160kb region upstream and downstream of the target SNPs from the "Tifrunner" genome sequence (<https://peanutbase.org/>). The identified genes were annotated using the "Tifrunner" genome annotation file. Genes involved in sugar metabolism were defined as candidate genes.

2.5 Real-Time Fluorescence Quantitative PCR Validation

Three genes were validated using real-time quantitative reverse transcription PCR (qRT-PCR). Specific primers for the DEGs were designed using Primer Premier 5.0 software (Primer, Canada). *Alcohol dehydrogenase (ADH)* was used as the internal reference gene, and the relative levels of expression of the genes were measured. The primer sequences utilized for qRT-PCR are provided in Supplementary Table 1. The qRT-PCR reactions were conducted using a Fast Real qPCR Premix (SYBR Green) kit (TIANGEN, Beijing, China). Each RT-qPCR reaction system was composed of a volume of 20 μL , including 0.6 μL of upstream and downstream primers, 4 μL of cDNA template, 10 μL of SYBR reagent, and 4.8 μL of ddH₂O. The PCR program included an initial pre-denaturation step at 95 °C for 10 min, followed by thermal cycling: denaturation at 95 °C for 5 s, annealing at 58 °C for 10 s, and an extension at 72 °C for 15 s, which was repeated for a total of 40 cycles. A LightCycler®96 SW 1.1 (Roche Diagnostics, Rotkreuz, Switzerland) was used to read the qPCR results, and the relative levels of expression of the genes were calculated using the $2^{-\Delta\Delta\text{CT}}$ method.

2.6 DNA Extraction

Thirty-one peanut germplasms with varying sugar contents were selected from the germplasm repository maintained by the Peanut Genetic Breeding Innovation Team at Hebei Agricultural University. These germplasms were cultivated in a greenhouse, and fresh peanut leaves were collected. Genomic DNA was extracted from the peanut leaves using the CTAB method. The concentration, integrity, and purity of the DNA were assessed using a NanoDrop spectrophotometer. The DNA was diluted to a working concentration of approximately 50 ng/ μL and stored at -20°C for future use.

2.7 Development and Validation of KASP Markers

Based on the results of the genome-wide association analysis, KASP molecular markers were developed for significantly associated SNPs. Sequences spanning 200 bp upstream and 200 bp downstream of the SNPs were downloaded from the PeanutBase website (<https://peanutbase.org/>) for KASP primer design. The primers were designed using Prime 5.0 software. The KASP assay system consisted of the following components: 0.84 µL of genomic DNA template at a concentration of 50 ng/µL, 5 µL of 2× AQP master mix, 0.14 µL of SNP-specific primers, and 4 µL of double-distilled water (ddH₂O). The PCR program was set as follows: initial denaturation at 95°C for 10 minutes; followed by 10 cycles of amplification, with each cycle consisting of denaturation at 95°C for 20 seconds and annealing/extension at 61°C for 40 seconds, with the annealing temperature decreasing by 0.6°C per cycle; and finally, 31 cycles of amplification, with each cycle consisting of denaturation at 95°C for 20 seconds and annealing/extension at 55°C for 40 seconds.

2.8 Statistical Analysis

All phenotypic data were subjected to descriptive statistical analysis, correlation analysis, normality tests, and analysis of variance (ANOVA) using IBM® SPSS® Statistics 26 software. Calculation of Broad-Sense Heritability Using Excel, with the Formula as Follows: $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{eg}^2/n + \sigma_e^2/nr)$, where σ_g^2 , σ_{eg}^2 , and σ_e^2 represent the genotypic variance, the interaction between 103 genotypes and environment variance, and the error (residual) variance component, respectively. r is the number of environment trials and n is the number of replications in each environment trial [35]. The bar charts of sucrose content and soluble sugar content at different developmental stages, as well as the box plot, were generated using GraphPad Prism 8.

3 Results

3.1 Phenotypic Variation Analysis of Sugar Content in Peanut Kernels

In this study, statistical and phenotypic variation analyses were conducted on TSC, SSC, and SC across 103 peanut germplasms. The distributions of TSC, SSC, and SC exhibited normality (Fig. 1). Among these traits, TSC demonstrated the widest variation range, spanning from 6.48% to 26.10%, while SC showed the narrowest range, from 2.01% to 7.97% (Table 1). Correlation analysis revealed significant positive correlations among TSC, SSC, and SC. Analysis of variance (ANOVA) indicated that TSC, SSC, and SC were influenced not only by genotype but also by environmental factors and genotype-environment interactions. The broad-sense heritability (H^2) values for TSC, SSC, and SC were estimated to be 81.37%, 72.89%, and 86.80%, respectively (Table 2). These findings suggest that TSC, SSC, and SC are primarily genetically controlled quantitative traits with heritable components.

Table 1 Phenotypic variation of sugar content in 103 peanut germplasms							
Trait Name	Environment	Range (%)	Average (%)	SD	CV(%)	Skewness	Kurtosis
TSC	2022BD	6.48–25.34	14.40	4.00	27.78	0.50	-0.28
	2023BD	7.92–26.10	16.11	4.46	27.68	-0.34	0.53
SSC	2021BD	7.41–14.37	11.32	1.69	14.93	-0.08	-0.78
	2022BD	5.22–17.29	9.98	2.34	23.45	0.36	-0.13
	2023BD	5.68–17.05	10.97	2.47	22.52	0.59	-0.06
SC	2021BD	2.31–7.71	5.51	0.97	17.60	-0.34	0.53
	2022BD	2.19–7.68	4.76	1.17	24.58	0.02	-0.36
	2023BD	2.01–7.97	5.00	1.34	26.80	0.25	0.50

Note: TSC: total sugar content; SSC: soluble sugar content; SC: sucrose content; SD: Standard Deviation; CV: Coefficient of Variation

Table 2 Analysis of variance on sugar content of 103 peanut germplasm resources							
Trait	Source	<i>df</i>	SS	MS	<i>F</i> -Value	<i>P</i> -Value	<i>H</i> ² (%)
TSC	Environment	1	134.36	134.36	55.33	< 0.0001	81.37
	Genotype	99	4479.64	45.25	18.63	< 0.0001	
	Environment×Genotype	73	1423.07	19.49	8.03	< 0.0001	
	Error	174	422.56	2.436			
SSC	Environment	2	135.26	67.63	59.26	< 0.0001	72.89
	Genotype	102	964.36	9.46	8.28	< 0.0001	
	Environment×Genotype	163	1031.39	6.33	5.54	< 0.0001	
	Error	176	200.85	1.41			
SC	Environment	2	29.58	14.75	47.74	< 0.0001	86.80
	Genotype	101	404.11	4.00	12.95	< 0.0001	
	Environment×Genotype	173	168.33	0.97	3.15	< 0.0001	
	Error	174	54.39	0.31			
Notes: <i>df</i> : degree of freedom; SS: sum of squares; MS: mean square							

3.2 Genome-Wide Association Analysis

Association analysis was performed on 103 peanut accessions using the general linear model (GLM + PCA) in TASSEL 5.0 software. A total of 320 significantly associated SNPs were identified, with phenotypic contribution rates ranging from 9.60% to 20.55%, distributed across all 20 chromosomes. Under two environmental conditions, 98 SNPs significantly associated with total sugar content (TSC) were detected, spanning 14 chromosomes (A02, A03, A04, A05, A06, A07, A10, B03, B04, B05, B06, B07, B08, and B10), with phenotypic contribution rates ranging from 10.84% to 20.55%. Under three environmental conditions, 122 SNPs significantly associated with soluble sugar content (SSC) were identified, primarily located on chromosomes A09, B03, and B08, with phenotypic contribution rates ranging from 9.54% to 16.94%. Among these, 11 SNPs were located on chromosome A09, with phenotypic contribution rates ranging from 10.22% to 12.97%. Additionally, 148 SNPs significantly associated with sucrose content (SC) were detected under three environmental conditions, distributed across 19 chromosomes (excluding A08), with phenotypic contribution rates ranging from 9.60% to 20.29% (Fig. 2, Table S1).

In this study, 21 pleiotropic SNPs were identified, all detected under the same environmental conditions. Two SNPs, located on chromosomes A02 and B08, were significantly associated with both TSC and SC, with phenotypic contribution rates ranging from 10.84% to 14.24%. Nineteen SNPs were significantly associated with both SSC and SC, distributed on chromosomes A01, A03, A04, A09, B03, and B06, with phenotypic contribution rates ranging from 10.18% to 17.49%. Among these, 11 SNPs significantly associated with both SSC and SC were located on chromosome B03, with phenotypic contribution rates ranging from 12.50% to 16.17%. Two closely located SNPs, AX-176805505 and AX-176805573, were found within the 108820416 bp–108848601 bp interval on chromosome B03, with phenotypic contribution rates ranging from 12.54% to 15.09% (Table S2).

Furthermore, 21 multi-environment co-localized SNPs were identified, all significantly associated with SC, distributed on chromosomes A01, A05, A07, A10, B05, B08, B09, and B10, with phenotypic explanation rates ranging from 10.03% to 20.29%. Among these, six SNPs were consistently detected across all three environments, with phenotypic contribution rates ranging from 10.03% to 20.29%. On chromosomes A05 and B08, five and eight multi-environment co-localized significant SNPs were detected, respectively, with phenotypic contribution rates ranging from 10.35% to 20.29% and 10.19% to 18.56%. Notably, the SNP AX-177642263 on chromosome B08 was both a pleiotropic and multi-environment co-localized significant SNP, with phenotypic contribution rates ranging from 10.84% to 12.25% (Table S3).

3.3 Candidate Gene Identification

Based on the results of the genome-wide association study (GWAS) and the linkage disequilibrium (LD) decay region, candidate genes were screened within a 160 kb region flanking 21 SNPs significantly associated with pleiotropic effects and 21 SNPs significantly co-localized across multiple environments. All genes within these intervals were searched using the PeanutBase database. Among them, 12 genes were closely related to sugar metabolism processes (Table 3).

Table 3
Candidate gene information table

Gene ID	Chromosome	Position(bp)	Function Annotation
<i>Arahy.1BTX43</i>	A05	90381350–90392960	Beta glucosidase
<i>Arahy.JGCP84</i>	A05	90361118–90361498	Beta glucosidase
<i>Arahy.US8HBC</i>	A05	89909000–89917250	α -galactosidase
<i>Arahy.B8GWAG</i>	A07	6623190–6625824	Galactosyltransferase protein
<i>Arahy.VSC23Z</i>	B03	112236549–112240021	Glucose dehydrogenase
<i>Arahy.IG4CN6</i>	B03	108190615–108192754	Sugar transporter
<i>Arahy.KFDF05</i>	B03	108361004–108363118	Sugar transporter
<i>Arahy.78PE2K</i>	B05	139799001–139802016	SWEET sugar transporter
<i>Arahy.OA1ABD</i>	B08	117873277–117883963	Glucose dehydrogenase
<i>Arahy.OAY0ZE</i>	B08	28509472–28509678	Galactosyltransferase protein
<i>Arahy.CNB6AU</i>	B08	26964142–26964932	Starch branching enzyme
<i>Arahy.44QEU2</i>	B08	27791257–27797548	Sodium/calcium exchanger protein

3.4 Expression Analysis of Candidate Genes

Expression analysis of candidate genes was conducted using four seed developmental stages from the high-sugar cultivar "Jinongxian1" and the low-sugar cultivar "Silihong" to investigate their functional roles in sugar synthesis. Soluble sugar content (SSC) and sucrose content (SC) were measured during the DAF70-DAF100 (Days After Flower) period in both cultivars. Significant differences in SSC and SC were detected between the two cultivars, though their temporal variation patterns differed. Specifically, in "Jinongxian 1", SSC reached its maximum at DAF70 and minimum at DAF100, while in "Silihong", SSC was highest at DAF100 and lowest at DAF90. For SC, "Jinongxian 1" exhibited the highest levels at DAF70 and the lowest at DAF90, whereas "Silihong" showed the highest SC at DAF100 and the lowest at DAF70 (Fig. 3).

Significant differential expression between cultivars was observed for the SWEET-encoding gene *Arahy.78PE2K* and two STP-encoding genes (*Arahy.KFDF05* and *Arahy.1G4CN6*). *Arahy.78PE2K* exhibited higher expression in the high-sugar cultivar "Jinongxian 1" compared to "Silihong" across most development stages (Fig. 6). This trend was consistent except at DAF90, where expression was lower in the high-sugar cultivar. *Arahy.KFDF05* exhibited elevated expression in the low-sugar cultivar "Silihong" relative to "Jinongxian 1" at all stages except DAF100. Conversely, *Arahy.1G4CN6* showed reduced expression in "Jinongxian 1" compared to "Silihong", with statistically significant differences observed at all stages except DAF80 (Fig. 4).

3.5 KASP Molecular Marker Development

Based on genome-wide association analysis and qRT-PCR results, two SNPs (AX-176820713 and AX-176816968) were ultimately identified. Specifically, AX-176820713 represents a stably inherited SNP locus significantly associated with seed sucrose content (SC), while AX-176816968 is a co-localized SNP locus demonstrating significant association with both seed sugar content (SSC) and seed sucrose content (SC) under the 2023BD environmental condition. To facilitate functional validation, these two SNP loci were used to design primers targeting for experimental verification.

Among these, AX-176816968 exhibited robust genotyping performance and was formally designated as *qK-SCB03.1*. As illustrated in Fig. 5A, fluorescence-based genotyping revealed AA genotype in blue-labeled materials and GG genotype in red-labeled materials. Boxplot analysis further confirmed that AA genotype materials exhibit significantly higher sucrose content compared to GG genotype materials (Fig. 5B). Quantitative analysis showed that kernel sucrose content in AA genotype materials ranged from 5.45% to 7.82% (mean: 6.37%), whereas GG genotype materials displayed a range of 2.03% to 3.67% (mean: 3.19%). These findings provide critical molecular markers for precision breeding strategies aimed at optimizing sugar content in peanut cultivars (Table S4).

4 Discussion

4.1 Abundant Variation in Sugar Content of Peanut Germplasm in Natural Populations

Peanut kernel sweetness is a heritable quantitative trait ^[5], primarily derived from soluble sugars (sucrose, glucose, fructose) that exhibit substantial variation across germplasm. As a core resource for dissecting complex trait genetics, GWAS populations derive their value from extensive genetic diversity accumulated through natural selection and artificial domestication. Such populations typically encompass wild relatives, landraces, and commercial inbred lines, providing unparalleled advantages for association mapping. Significant phenotypic variation in sugar content (> 3-fold) has been documented in natural populations of crops like soybean, sugarcane, and vegetable soybean ^[36–38]. Similarly, sucrose content in peanut natural populations shows considerable variation ^[39]. In this study, a natural population of 103 peanut accessions was employed to quantify total sugars, soluble sugars, and sucrose content across three environments. The phenotypic ranges were 6.48%~26.10% for TSC, 5.22%~17.29% for SSC, and 2.01%~7.97% for SC, indicating substantial variation. This broad phenotypic spectrum not only enhances statistical power for GWAS but also underscores the critical role of genetic diversity in elucidating the genetic architecture of complex traits.

4.2 SNP variations related to sugar content lead to differences in the sugar content of peanut kernels.

SNPs, as the most prevalent form of genetic variation in the genome, exert significant impacts on gene function and phenotypic traits, occupying a pivotal position in crop trait improvement and gene

functional research ^[40]. When SNPs induce base alterations in coding regions, they may lead to amino acid substitutions (e.g., synonymous mutations) or translation termination (e.g., frameshift mutations). The ΔG deletion in the third exon of the wheat *FT-D1* ^[41] induces protein dysfunction, resulting in increased spikelet number and delayed heading. In tomato, a single-base mutation in the W-box motif of the *WRKY33* promoter disrupts the low-temperature response element, leading to reduced cold tolerance in cultivated varieties ^[42]. The mechanism underlying the first high oleic acid peanut mutant F435 involves the regulation of $\Delta 12$ fatty acid desaturase (FAD2) ^[43]. Specifically, a G:C/A:T mutation at position 448 bp in the *ahFAD2A* gene causes a D150N missense mutation, while an A:T insertion at position 442 bp in the *ahFAD2B* gene triggers a frameshift mutation and premature stop codon^[44–45]. These findings underscore the central role of identifying key SNPs in elucidating the regulatory mechanisms of complex traits, with systematic functional SNP identification serving as a fundamental basis for advancing crop genetic improvement. In this study, the SNP AX-177638808, which is significantly associated with soluble sugar content and exhibits a G/A mutation, co-localizes with the major sucrose-related QTL *qSUCA08.1* reported by Li ^[30]. Similarly, the SNP AX-176802474, associated with total sugar content and also showing a G/A mutation, aligns with the sucrose-related *qSCA06.2* identified by Wang ^[8]. Additionally, novel SNP such as AX-177642263, AX-176820713, and AX-176816968 were newly characterized in this study, and their functional roles in peanut sugar metabolism warrant further investigation. Our results demonstrate that the GWAS panel used in this study harbors abundant genetic variants related to sugar content, confirming its suitability for sugar-related GWAS analyses.

4.3 Expression of *AhSTP* and *AhSWEET* Influences Sugar Content in Peanut Kernels

The expression of transport and translocation-related genes significantly impacts sugar content in peanut kernels. In plants, sucrose serves as the primary photosynthetic assimilate, providing substrates for the accumulation of storage substances like proteins and lipids ^[46]. After phloem transport to sink tissues, sucrose enters parenchyma cells via sucrose transporters (SUT) on the cell membrane. In the cytoplasm, it is hydrolyzed to hexoses (glucose and fructose) by sucrose synthase (SUS) and neutral invertase (NINV). Alternatively, sucrose can be cleaved extracellularly by cell wall invertase (CWINV), with the resulting hexoses transported into cells via hexose transporters (STP) ^[47–50]. SWEET proteins, a novel class of sugar transporters facilitating passive diffusion along concentration gradients, mediate transmembrane transport of various sugars including sucrose, fructose, and glucose ^[51]. In *Arabidopsis*, *AtSWEET11* and *AtSWEET12* cooperate with *AtSUT1* in phloem loading ^[52], while *AtSWEET16* and *AtSWEET17* regulate vacuolar fructose storage ^[53]. STPs, specialized hexose transporters with high substrate affinity, are subject to glucose-mediated repression of *AtSTP1* expression in *Arabidopsis* ^[54]. In tomato, silencing three *SIHT* genes identified via QTL mapping reduced fruit hexose concentrations by 55% ^[55–56]. In this study, two *STP* genes (*Arahy.KFDF05* and *Arahy.1G4CN6*) and one *SWEET* gene (*Arahy.78PE2K*) were identified within candidate loci. Previous studies confirmed differential expression

of *Arahy.78PE2K* between high- and low-sugar peanut cultivars during development^[57]. Cis-element analysis revealed that these genes contain not only light and stress-responsive elements but also hormone-responsive motifs (e.g., TCA element for salicylic acid, ABRE for abscisic acid) and MYB transcription factor binding sites. This suggests their involvement in regulating cytoplasmic sugar levels through light, stress, and hormonal signaling, thereby influencing vacuolar sugar storage. RT-qPCR analysis demonstrated cultivar-specific expression patterns: *Arahy.KFDF05* showed significantly higher expression in the high-sugar cultivar "Jinongxian 1" than in "Silihong" across four developmental stages, while *Arahy.1G4CN6* exhibited the opposite pattern with significant differences at all stages except DAF80. *Arahy.78PE2K* displayed consistently higher expression in the high-sugar cultivar. These findings indicate that *Arahy.1G4CN6* negatively regulates sugar accumulation in kernels, whereas *Arahy.78PE2K* and *Arahy.KFDF05* act as positive regulators. The three genes are located within the attenuation regions of significant SNPs AX-176820713 and AX-176816968, exhibiting germline variation among germplasms. These loci serve as pivotal targets for investigating sugar content variation across germplasms. Further in-depth studies are warranted to elucidate their haplotypic patterns and the regulatory mechanisms governing sugar content.

4.4 *qK-SCB03.1* Enables Genotyping of Sucrose Content in Peanut

KASP technology, optimized for detecting Indels or SNPs, enables efficient multiplex SNP genotyping across diverse samples. Currently, only one indel marker associated with sucrose content has been developed in peanut^[30]. In this study, GWAS was performed on a peanut population to identify SNPs significantly associated with kernel sugar content. A KASP marker, *qK-SCB03.1*, was developed based on these associations. Genotyping results revealed that peanut accessions with the AA genotype generally exhibited higher sucrose content than those with the GG genotype. This molecular marker is anticipated to facilitate marker-assisted selection for breeding peanut varieties with optimized sucrose content. Furthermore, this locus not only shows significant association with sucrose but also with soluble sugar content, indicating its pleiotropic effect. Thus, *qK-SCB03.1* holds potential for selecting peanut germplasm with varying soluble sugar contents, providing a valuable tool for targeted breeding strategies.

Conclusions

In this study, we conducted a genome-wide association study (GWAS) using 103 peanut germplasm accessions, identifying 98 SNPs significantly associated with total sugar content, 122 SNPs significantly associated with soluble sugar content, and 148 SNPs significantly associated with sucrose content. A total of 21 pleiotropic SNPs and 21 stable genetic SNPs significantly associated with sucrose were detected. Three candidate genes related to sucrose were screened. Furthermore, based on GWAS results and qRT-PCR validation, a KASP molecular marker (*qK-SCB03.1*) associated with sucrose content in peanut kernels was developed. This work provides theoretical support for peanut sugar content research

and holds significant importance for improving peanut kernel sugar content and breeding high-sugar peanut varieties.

Declarations

Ethics, Consent to Participate, and Consent to Publish declarations

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data generated and analyzed in this study are included in the manuscript and supplementary files. Further inquiries can be directed to the corresponding authors.

Conflict of interest

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

Funding

This work was financially sponsored by the National Key Research and Development Program of China(2023YFD1202800); Hebei Agriculture Research System (HBCT2024040205); the China Agriculture Research System (CARS-13); the National Key Research and Development Program of seed industry technology innovation team of Hebei province(21326316D-2); Natural Science Foundation of Hebei Province(C2022204145)

Author contributions

J. D.: Investigation, Writing – original draft. M. H.: Methodology, Supervision, Writing – review & editing. S. C.: Investigation, Writing – review & editing. H. D.: Investigation, Writing – review & editing. Y. L.: Investigation, Writing – review & editing. X. L.: Investigation, Writing – review & editing. X. C.: Investigation, Writing – review & editing. L. L.: Formal analysis, Funding acquisition, Supervision, Writing – review & editing, Writing – original draft.

References

1. Zhang X Q, Tang Z. Study on the Consumption Trend of Peanut in China[J]. Journal of Peanut Science, 2008, (01): 32-36.

2. Pattee H E, Isleib T G, Giesbrecht F G. Variation in intensity of sweet and bitter sensory attributes across peanut genotypes1[J]. Peanut Science, 1998, 25(2): 63-9.
3. Mason M E, Newell J A, Johnson B R, *et al.* Nonvolatile flavor components of peanuts[J]. Journal of Agricultural and Food Chemistry, 1969, 17(4): 728-732.
4. Savage G P, Keenan J I. The composition and nutritive value of groundnut kernels[M]. The groundnut crop: A scientific basis for improvement. Dordrecht: Springer Netherlands, 1994: 173-213.
5. Pattee H E, Isleib T G, Giesbrecht F G, *et al.* Relationships of sweet, bitter, and roasted peanut sensory attributes with carbohydrate components in peanuts[J]. Journal of Agricultural and Food Chemistry, 2000, 48(3): 757-63.
6. Oupadissakoon C, Young C T, Mozingo R W. Evaluation of free amino acid and free sugar contents in five lines of Virginia-type peanuts at four locations[J]. Peanut Science, 1980, 7(1):55-60.
7. Huai D, Zhi C, Wu J, *et al.* Unveiling the molecular regulatory mechanisms underlying sucrose accumulation and oil reduction in peanut kernels through genetic mapping and transcriptome analysis[J]. Plant Physiology and Biochemistry, 2024, 208: 108448.
8. Wang F, Miao H, Zhang S, *et al.* Identification of a major QTL underlying sugar content in peanut kernels based on the RIL mapping population[J]. Frontiers in Plant Science, 2024, 15: 1423586.
9. Wang Z, Zhang Y, Huai D, *et al.* Detection of two homologous major QTLs and development of diagnostic molecular markers for sucrose content in peanut[J]. Theoretical and Applied Genetics, 2024, 137(3): 61.
10. Li H, Peng Z, Yang X, *et al.* Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels[J]. Nature Genetics, 2013, 45(1): 43-50.
11. Manickavelu A, Joukhadar R, Jighly A, *et al.* Genome wide association mapping of stripe rust resistance in Afghan wheat landraces[J]. Plant Science, 2016, 252: 222-229.
12. Su J, Wang C, Ma Q, *et al.* An RTM-GWAS procedure reveals the QTL alleles and candidate genes for three yield-related traits in upland cotton[J]. BMC Plant Biology, 2020,20: 416.
13. Liu R, Gong J, Xiao X, *et al.* GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high-density SNP markers[J]. Frontiers in Plant Science, 2018, 9: 1067.
14. Wang J, Yan C, Li Y, *et al.* GWAS discovery of candidate genes for yield-related traits in peanut and support from earlier QTL mapping studies[J]. Genes, 2019, 10(10): 803.
15. Zhang H, Chu Y, Dang P, *et al.* Identification of QTLs for resistance to leaf spots in cultivated peanut (*Arachis hypogaea* L.) through GWAS analysis[J]. Theoretical and Applied Genetics, 2020, 133: 2051-2061.
16. Zhang X, Zhang J, He X, *et al.* Genome-wide association study of major agronomic traits related to domestication in peanut[J]. Frontiers in Plant Science, 2017, 8: 1611.
17. Zhang H, Wang M L, Dang P, *et al.* Identification of potential QTLs and genes associated with seed composition traits in peanut (*Arachis hypogaea* L.) using GWAS and RNA-Seq analysis[J]. Gene,

2021, 769: 145215.

18. Umer M J, Lu Q, Huang L, *et al.* Genome-wide association study reveals the genetic basis of amino acids contents variations in Peanut (*Arachis hypogaea* L.) [J]. *Physiologia Plantarum*, 2024, 176(5): e14542.
19. Semagn K, Babu R, Hearne S, *et al.* Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement[J]. *Molecular breeding*, 2014, 33: 1-14.
20. Pan W J, Xue H A N, Huang S Y, *et al.* Identification of candidate genes related to soluble sugar contents in soybean seeds using multiple genetic analyses[J]. *Journal of Integrative Agriculture*, 2022, 21(7): 1886-1902.
21. Fu L, Wu J, Yang S, *et al.* Genome-wide association analysis of stem water-soluble carbohydrate content in bread wheat[J]. *Theoretical and Applied Genetics*, 2020, 133: 2897-2914.
22. Gong L, Guo X, Nie F, *et al.* Identification of single nucleotide polymorphism in StCWIN1 and development of Kompetitive Allele-Specific PCR (KASP) marker associated with tuber traits in potato[J]. *Plant Growth Regulation*, 2023, 101(2): 453-462.
23. Jia Q, Zhou M, Xiong Y, *et al.* Development of KASP markers assisted with soybean drought tolerance in the germination stage based on GWAS[J]. *Frontiers in Plant Science*, 2024, 15: 1352379.
24. Kumar S, Pradhan A K, Kumar U, *et al.* Validation of Novel spot blotch disease resistance alleles identified in unexplored wheat (*Triticum aestivum* L.) germplasm lines through KASP markers[J]. *BMC Plant Biology*, 2022, 22(1): 618.
25. Zhao Y, Chen W, Cui Y, *et al.* Detection of candidate genes and development of KASP markers for Verticillium wilt resistance by combining genome-wide association study, QTL-seq and transcriptome sequencing in cotton[J]. *Theoretical and Applied Genetics*, 2021, 134(4): 1063-1081.
26. Wang Z, Yan L, Chen Y, *et al.* Detection of a major QTL and development of KASP markers for seed weight by combining QTL-seq, QTL-mapping and RNA-seq in peanut[J]. *Theoretical and Applied Genetics*, 2022, 135(5): 1779-1795.
27. Huang B, Hua L I U, Fang Y, *et al.* Identification of oil content QTL on Arah12 and Arah16 and development of KASP markers in cultivated peanut (*Arachis hypogaea* L.) [J]. *Journal of Integrative Agriculture*, 2023, 12(11): 10-11.
28. Sun Z, Qi F, Liu H, *etal.* QTL mapping of quality traits in peanut using whole-genome resequencing[J]. *The crop journal*, 2022, 10(1): 177-184.
29. Li X, Hao J, Deng H, *et al.* Identification of a pleiotropic QTL and development KASP markers for HPW, HSW, and SP in peanut[J]. *Journal of Integrative Agriculture*, 2024.
30. Li W, Huang L, Liu N, *et al.* Identification of a stable major sucrose-related QTL and diagnostic marker for flavor improvement in peanut[J]. *Theoretical and Applied Genetics*, 2023, 136(4): 78.
31. Zhang Xiaojun, Zhao Ruihua, Yu Xiaona, *et al.* Molecular Marker Methods for Loci Associated with Soluble Sugar Content in Peanut Kernels and Their Applications, CN111334606A[P]. 2020.

32. Chen M, Hou M Y, Cui S L, *et al.* Construction of Near-Infrared Model of Peanut Sugar Content in Different Seed Coat Colors[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(09): 2896-2902.
33. Li, L., Cui, S., Dang, P. *et al.* GWAS and bulked segregant analysis reveal the Loci controlling growth habit-related traits in cultivated Peanut (*Arachis hypogaea* L.). *BMC Genomics***23**, 403 (2022).
34. Yin L, Zhang H, Tang Z, *et al.* rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study[J]. *Genomics, proteomics & bioinformatics*, 2021, 19(4): 619-628.
35. Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. Estimating and interpreting heritability for plant breeding: an update. *Plant Breed. Rev.* 2003: 22, 9–112.
36. Hou A, Chen P, Alloatti J, *et al.* Genetic variability of seed sugar content in worldwide soybean germplasm collections[J]. *Crop Science*, 2009, 49(3): 903-912.
37. Tai P Y P, Miller J D. Germplasm diversity among four sugarcane species for sugar composition[J]. *Crop science*, 2002, 42(3): 958-964.
38. Xu W, Liu H, Li S, *et al.* GWAS and identification of candidate genes associated with seed soluble sugar content in vegetable soybean[J]. *Agronomy*, 2022, 12(6): 1470.
39. Bishi S K, Kumar L, Dagla M C, *et al.* Characterization of Spanish peanut germplasm (*Arachis hypogaea* L.) for sugar profiling and oil quality[J]. *Industrial Crops and Products*, 2013, 51: 46-50.
40. Shastry B S. SNPs: impact on gene function and phenotype[J]. *Single nucleotide polymorphisms: methods and protocols*, 2009: 3-22.
41. Chen Z, Ke W, He F, *et al.* A single nucleotide deletion in the third exon of *FT-D1* increases the spikelet number and delays heading date in wheat (*Triticumaestivum* L.)[J]. *Plant biotechnology journal*, 2022, 20(5): 920-933.
42. Guo M, Yang F, Liu C, *et al.* A single-nucleotide polymorphism in WRKY33 promoter is associated with the cold sensitivity in cultivated tomato[J]. *New Phytologist*, 2022, 236(3): 989-1005.
43. Norden A J, Gorbet D W, *et al.*, 1987, Variability in oil quality among peanut genotypes in the Florida breeding program, *Peanut Sci.*, 14(1): 7-11.
44. Jung S, Powell G, Moore K, *etal.* The high oleate trait in the cultivated peanut [*Arachishypogaea* L.]. II. Molecular basis and genetics of the trait[J]. *Molecular and General Genetics MGG*, 2000a, 263: 806-811.
45. Lopez Y, Nadaf H L, Smith O D, *etal.* Isolation and characterization of the $\Delta 12$ -fatty acid desaturase in peanut (*Arachishypogaea* L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines[J]. *Theoretical and Applied Genetics*, 2000, 101: 1131-1138.
46. Hajduch M, Matusova R, Houston N L, *et al.* Comparative proteomics of seed maturation in oilseeds reveals differences in intermediary metabolism[J]. *Proteomics*, 2011, 11(9): 1619-1629.
47. Zhang L, Peng Y, Pelleschi-Travier S, *et al.* Evidence for apoplasmic phloem unloading in developing apple fruit[J]. *Plant Physiology*, 2004, 135: 574–586.

48. Fan R, Peng C, Xu Y, *et al.* Apple sucrose transporter SUT1 and sorbitol transporter SOT6 interact with cytochrome b5 to regulate their affinity for substrate sugars[J]. *Plant Physiology*, 2009, 150: 1880–1901.
49. Li M, Feng F, Cheng L. Expression patterns of genes involved in sugar metabolism and accumulation during apple fruit development[J]. *PLoS ONE*, 2012, 7: e33055.
50. Wang D, Zhao J T, Qin Y Q, *et al.* Molecular cloning, characterization and expression profile of the sucrose synthase gene family in *Litchi chinensis*[J]. *Horticultural Plant Journal*, 2021, 7 (6): 520–528
51. Jeena G S, Kumar S, Shukla R K. Structure, evolution and diverse physiological roles of SWEET sugar transporters in plants[J]. *Plant Molecular Biology*, 2019, 100: 351-365.
52. Le Hir R, Spinner L, Klemens P A W, *et al.* Disruption of the sugar transporters *AtSWEET11* and *AtSWEET12* affects vascular development and freezing tolerance in *Arabidopsis*. *Molecular Plant*, 2015, 8(11): 1687–1690.
53. Chardon F, Bedu M, Calenge F, *et al.* Leaf fructose content is controlled by the vacuolar transporter *SWEET17* in *Arabidopsis*[J]. *Current Biology*, 2013, 23(8): 697-702.
54. Cordoba E, Aceves-Zamudio D L, Hernández-Bernal A F, *et al.* Sugar regulation of Sugar Transporter Protein 1 (STP1) expression in *Arabidopsis thaliana*[J]. *Journal of Experimental Botany*, 2015, 66(1): 147-159.
55. McCurdy D W, Dibley S, Cahyanegara R, *et al.* Functional characterization and RNAi-mediated suppression reveals roles for hexose transporters in sugar accumulation by tomato fruit[J]. *Molecular Plant*, 2010, 3: 1049–1063.
56. Prudent M, Lecomte A, Bouchet J P, *et al.* Combining ecophysiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration[J]. *Journal of Experimental Botany*, 2011, 62:907–919.
57. Li W, Huang L, Liu N, *et al.* Key regulators of sucrose metabolism identified through comprehensive comparative transcriptome analysis in peanuts[J]. *International Journal of Molecular Sciences*, 2021, 22(14): 7266.

Figures

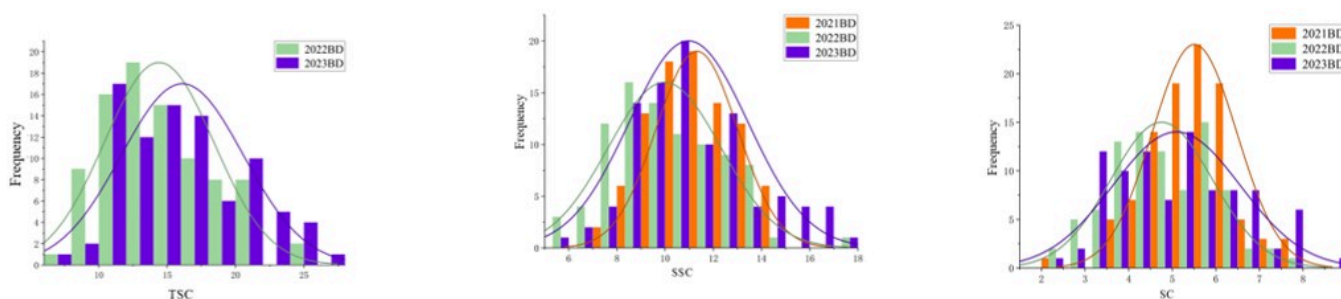


Figure 1

Histogram of frequency distribution for three traits

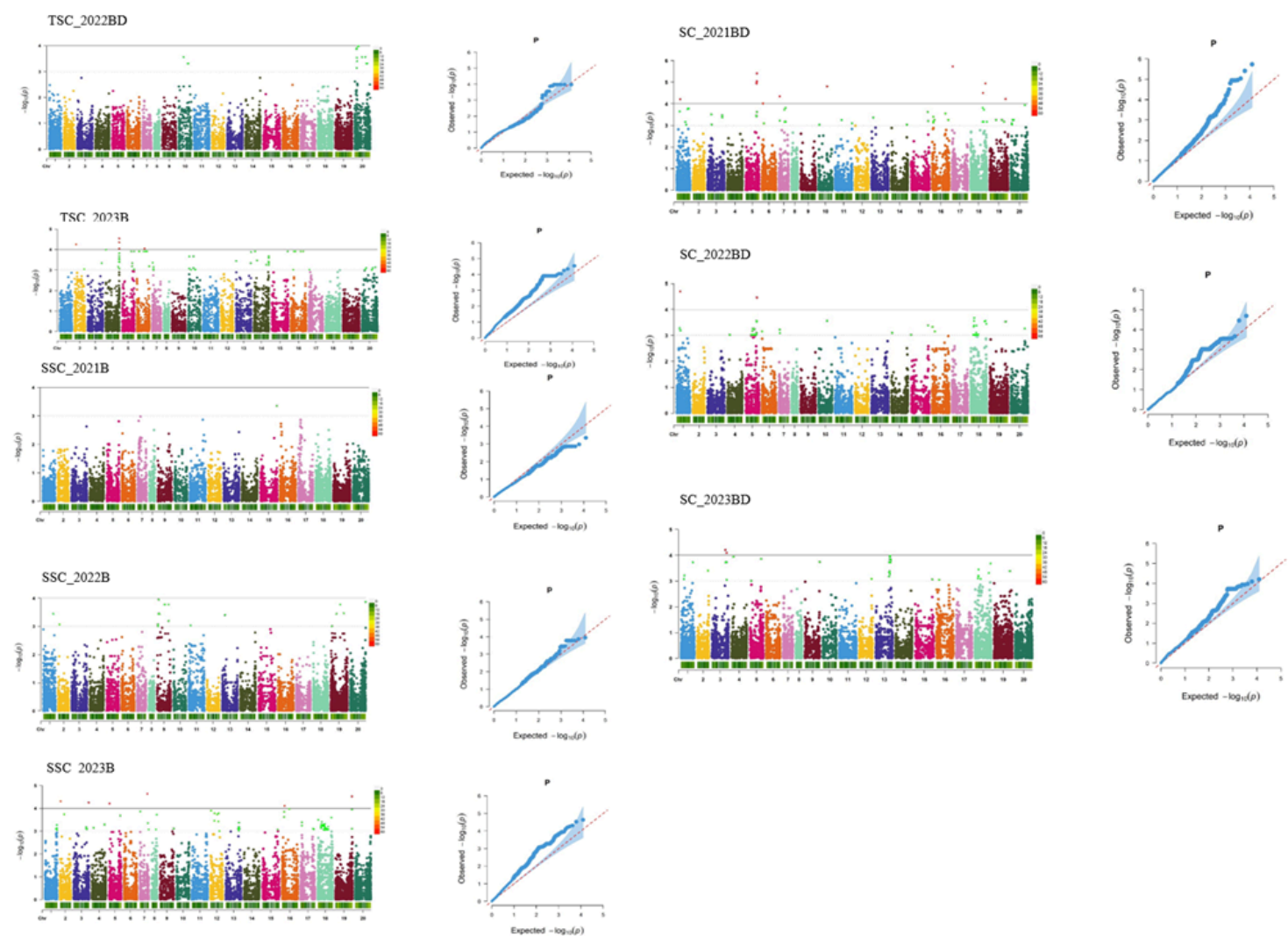


Figure 2

Manhattan Plot and Q-Q Plot of Sugar Content in Peanut Kernels

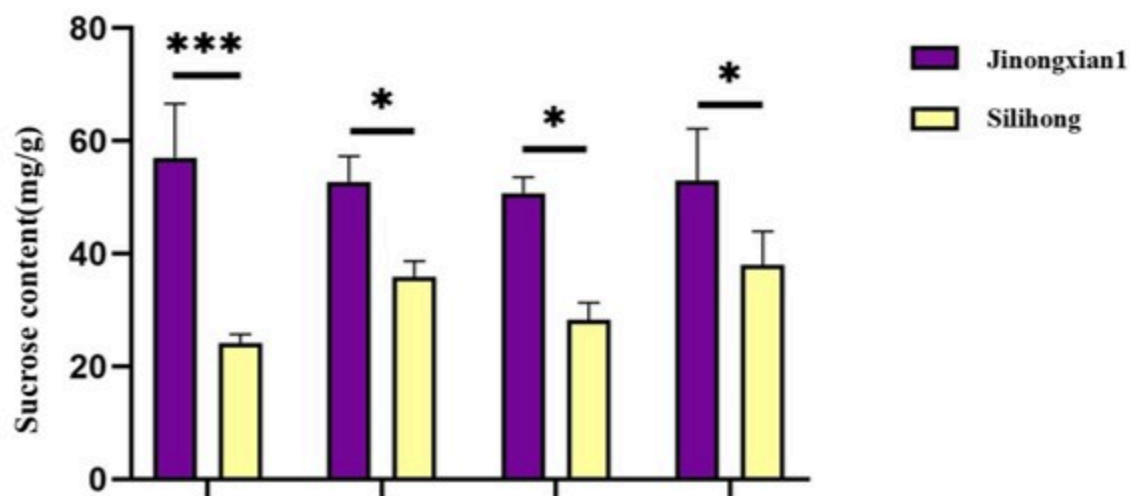
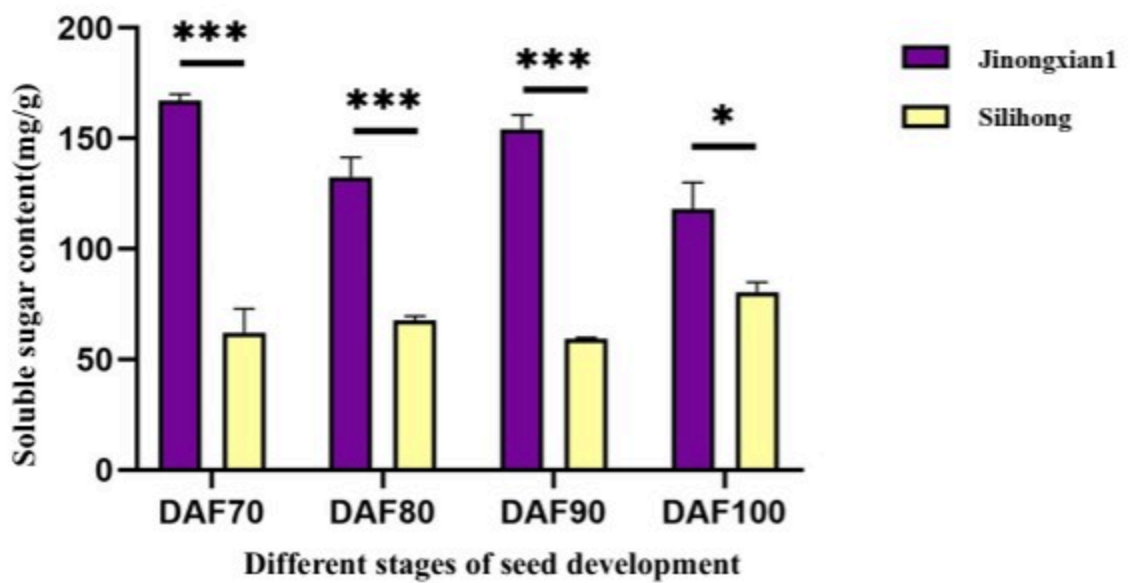


Figure 3

Differences in soluble sugars and sucrose content during the four stages of seed development between “Jinongxian 1” and “Silihong” peanut varieties

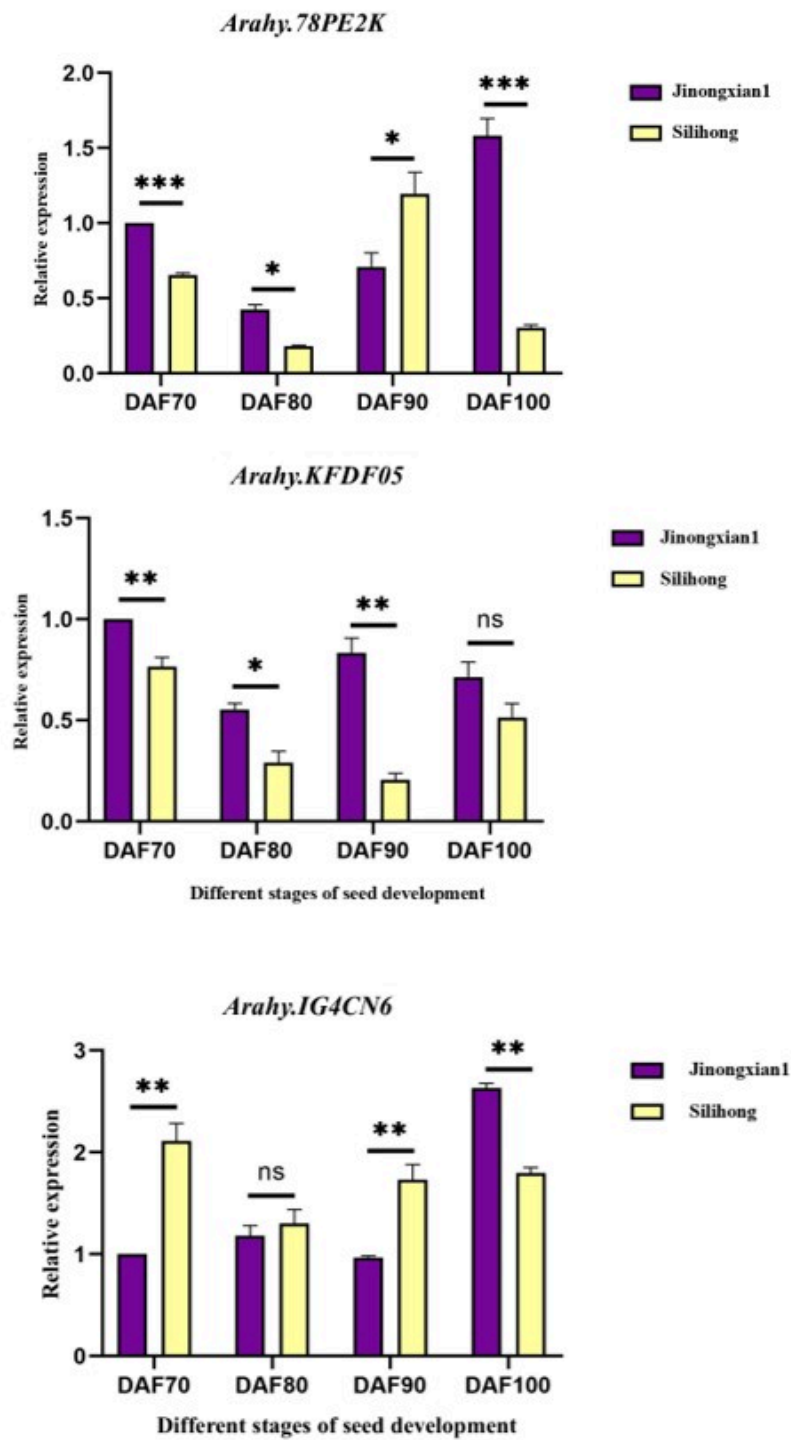
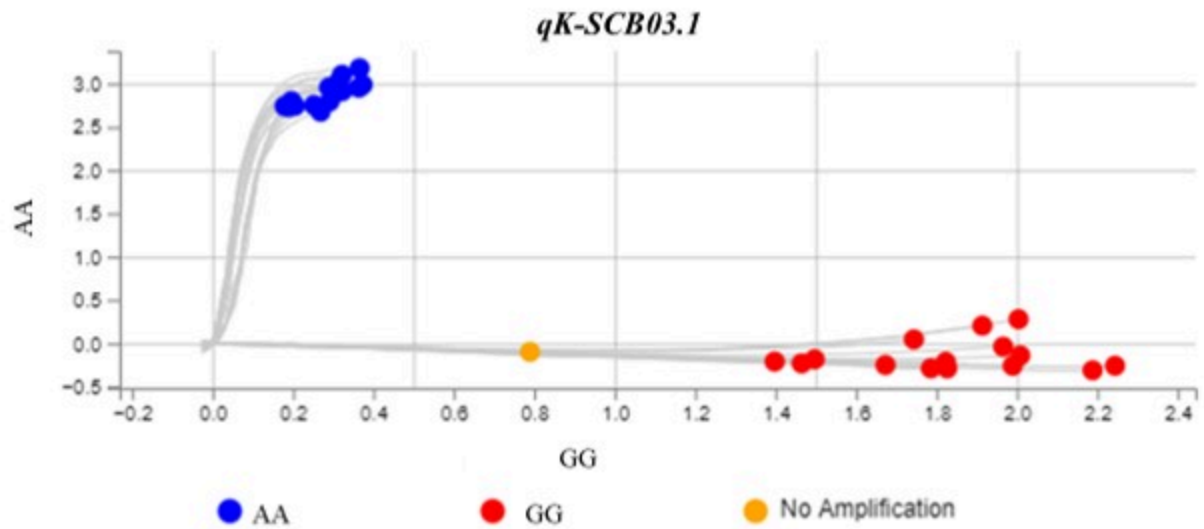


Figure 4

Expression differences of three genes during the four stages of seed development in “Jinongxian 1” and “Silihong”peanut varieties

A



B

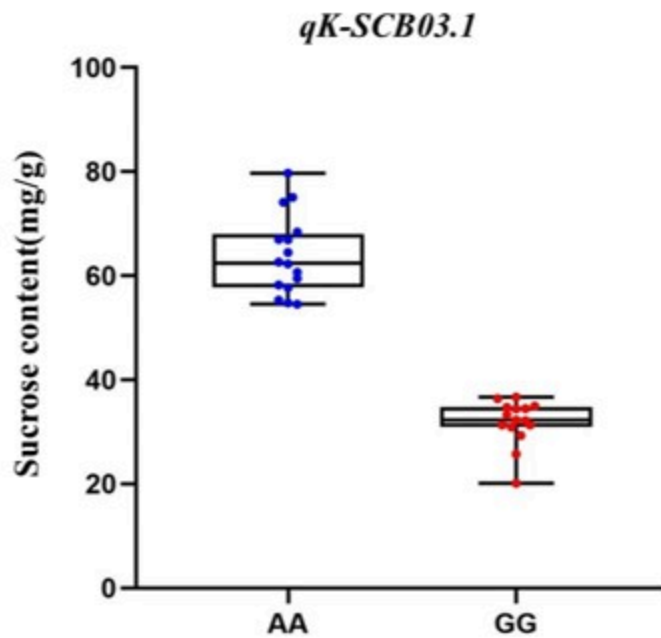


Figure 5

Typing results (A) and sucrose differences(B) of 31 germplasm resources in *qK-SCB03.1*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable.xlsx](#)