Article

Not peer-reviewed version

QTL and Candidate Genes that Control Seed Sugars Contents in the Soybean ‘Forrest’ By ‘Williams 82’ RIL Population

Dounya Knizia , Nacer Bellaloui , Jiazheng Yuan , Naoufal Lakhssassi , Erdem Anil , Tri Vuong , Mohamed G. Embaby , Henry T Nguyen , Alemu Mengistu , Khalid Meksem , Abdelmajid My Kassem *

Posted Date: 31 August 2023

doi: 10.20944/preprints202308.2073.v1

Keywords: soybean; RIL; Forrest; Williams 82; linkage map; RFOs; Sucrose; Raffinose; Stachyose; SNP

Article

QTL and Candidate Genes that Control Seed Sugars Contents in the Soybean 'Forrest' by 'Williams 82' RIL Population

Dounya Knizia [1], Nacer Bellaloui [3], Jiazheng Yuan [2], Naoufal Lakhssasi [1], Erdem Anil [1], Tri Vuong [4], Mohamed Embaby [1], Henry T. Nguyen [4], Alemu Mengistu [5], Khalid Meksem [1] and My Abdelmajid Kassem [2]*

[1] School of Agricultural Sciences, Southern Illinois University, Carbondale, IL 62901, USA
[2] Plant Genomics and Biotechnology Lab, Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC, USA.
[3] USDA, Agriculture Research Service, Crop Genetics Research Unit, 141 Experiment Station Road, Stoneville, Mississippi, 38776, USA.
[4] Division of Plant Science and Technology, University of Missouri, Columbia, MO, USA
[5] USDA, Agriculture Research Service, Crop Genetics Research Unit, 605 Airways Blvd Jackson, Tennessee, 38301, USA.
* Correspondence: author: mkassem@uncfsu.edu

Abstract: Soybean seed sugars are among the most abundant beneficial compounds for human and animal consumption in soybean seeds. Higher seed sugars such as sucrose are desirable as it contributes to taste and flavor in soy-based food. Therefore, the objectives of this study were to use 'Forrest' by 'Williams 82' (F×W82) recombinant inbred line (RIL) soybean population (n=309) to identify quantitative trait loci (QTL) and candidate genes that control seed sugar (sucrose, stachyose, and raffinose) contents in two environments (North Carolina and Illinois) over two years (2018 and 2020). A total of 26 QTL that control seed sugars contents were identified and mapped on 16 soybean chromosomes (chrs.). Interestingly, five QTL regions were identified in both locations, Illinois and North Carolina, in this study on chrs. 2, 5, 13, 17, and 20. Amongst 57 candidate genes identified in this study, 16 were located within 10 Megabase (MB) of the identified QTL. Amongst them a cluster of four genes involved in the sugars' pathway was collocated within 6 MB with two QTL that were detected in this study on chr. 17. Further functional validation of the identified genes could be beneficial in breeding programs to produce soybean lines with high beneficial sucrose and low raffinose family oligosaccharides.

Keywords: soybean; RIL; Forrest; Williams 82; linkage map; RFOs; sucrose; raffinose; stachyose; SNP

1. Introduction

Sugars including, sucrose, stachyose, glucose, raffinose, galactose, fructose, rhamnose, and starch, play a major role in seed and fruit development and seed desiccation tolerance (DT) [1–4]. Sucrose and raffinosaccharides (raffinose and stachyose), also called raffinose family oligosaccharides (RFOs), make up 5–7%, 1%, and 3–4% of total carbohydrates, respectively, of soybean seed dry weights [5]. RFOs are synthesized from sucrose by a series of additions of galactinol units and are involved in DT, freezing, stress tolerance, and seed longevity [6–9]. Galactinol synthase (GolS) is the key enzyme in the RFOs biosynthetic pathway converting galactinol and myo-inositol as the main precursors to form RFOs. Galactinol synthase (GolS) converts myo-inositol and UDP-galactose into galactinol, while sucrose and galactinol are converted into raffinose by the raffinose synthase [9,10]. In addition to being involved in stress tolerance, RFOs are reported to play a role in several signal transduction pathways [11], exports of specific mRNAs [12], and trafficking of certain vesical membranes [13].

 doi:10.20944/preprints202308.2073.v1

Like most seed composition traits, seed sugars [4] are influenced by many factors, including abiotic and biotic stresses, and environmental factors, such as temperature, soil moisture, freezing, seed maturity, and growth conditions. [1,14–19]. It was shown that stachyose contents increased drastically in drying seeds but not in seeds kept at high humidity levels, which reveals the critical role of stachyose in DT [1]. The effect of water deficit (WD) on enzymes involved in sugar biosynthetic pathways in soybean nodules was investigated. Sucrose synthase activity declined drastically with increased WD while sucrose content increased [14]. Other studies showed that WD impacts negatively sucrose biosynthesis and translocation from sources to sinks more than other sugar (raffinose and stachyose) biosynthesis [16,19]. Investigating 'Clark' and 'Harosoy' near-isogenic lines (NILs) revealed that Clark's sugars contents decreased with increased days of maturity for both cultivars while both positive and negative effects were observed concerning the effects of temperature in two different years (2004 and 2005) [15]. In 2004, seed sugars contents increased with temperature increase, while the contents decreased with increased temperatures in 2005 [15]. The effect of WD on several seed composition traits, including sugars on several Phomopsis susceptible and resistant soybean cultivars, was investigated. In fact, sugar (sucrose, raffinose, and stachyose) contents were higher in seeds of resistant maturity group III cultivars than their susceptible counterparts [16]. A recent study investigated the effect of soil moisture on seed sugars (sucrose, raffinose, stachyose) and starch contents among other compounds in two soybean cultivars in maturity group V (Asgrow, AG6332, and Progeny 5333RY) and showed that sucrose, stachyose, and raffinose contents in addition to the mineral nutrient (N, P, K, and Ca) contents decreased with increased soil moisture in both cultivars [17].

During the last decades, more than 53 QTL that control seed sucrose and RFOs, other sugar (glucose, galactose, fructose, fucose, rhamnose), and starch contents have been reported using different biparental and natural populations and mapping methods including single marker analysis, interval mapping (IM), composite interval mapping (CIM), and genome-wide association studies (GWAS) [18,20]. However, to our knowledge, only one of these studies identified candidate genes within these QTL regions [18,21]; The *Glyma.01g127600* that encodes for a protein phosphatase on chr. 1, *Glyma.03g019300* that encodes for a MADS-box protein, *Glyma.03g064700* that encodes for a phosphatidylinositol monophosphate-5-kinase on chr. 3, and *Glyma.06g194200* that encodes for a gibberellin-regulated protein on chr. 6 [18,21].

To improve seed quality, several attempts to manipulate seed sugars, phytic acid, and the content of other beneficial compounds have been conducted in recent years [22–24]. Monogastric animals (such as poultry and pigs) and humans do not produce α-galactosidase and cannot digest RFOs which reduces gastrointestinal performance, flatulence, and diarrhea. Therefore, reducing raffinose and stachyose and increasing sucrose in soybean seed content are desirable traits and the main goal in breeding programs [22–27]. The objective of this study was to genetically map QTL for seed sucrose, raffinose, and stachyose contents using the 'Forrest' by 'Williams 82' RIL population, in addition to identifying candidate genes involved in soybean seed sugars biosynthesis.

2. Materials and Methods

2.1. Plant Materials

The 'Forrest' × 'Williams 82' RIL population (F×W82, *n*=309) was previously studied and described in detail in our previous research [28,29]. The parents and RILs were evaluated in two locations: Spring Lake, NC (35.17° N, 78.97° W, 2018) and Carbondale, IL (37° N, 89° W, 2020). Details about growth conditions, crop management, and seed harvesting were carried out as described earlier [28,29].

2.2. Seed Sugars Quantification

RILs and parents (Forrest and Williams 82) and soybean germplasm seeds were harvested at maturity, and sugars (sucrose, raffinose, and stachyose) contents (%) were quantified using near-

 doi:10.20944/preprints202308.2073.v1

infrared reflectance (NIR) with an AD 7200 array feed analyzer (Perten, Springfield, IL) as described earlier [15,30].

2.3. DNA Isolation, SNP Genotyping, and Genetic Map Construction

Parents and RILs genomic DNA was extracted by cetyltrimethylammonium bromide (CTAB) method as previously described [31]. A NanoDrop spectrophotometer (NanoDrop Technologies Inc., Centreville, DE) was used to quantify DNA samples (50 ng/µl), and genotyping was done using the Illumina Infinium SoySNP6K BeadChips (Illumina, Inc. San Diego, CA) as described earlier [15] at the Soybean Genomics and Improvement Laboratory (USDA-ARS, Beltsville, MD 20705). The F×W82 genetic linkage map was constructed using JoinMap 4.0 [28,32] as previously described to detect QTL for seed isoflavones [28] and seed tocopherols contents [29].

2.4. Sugars QTL Detection

WinQTL Cartographer [33] Interval mapping (IM) and composite interval mapping (CIM) methods were used to identify QTL that control seed sugars (sucrose, stachyose, and raffinose) contents in this RIL population. The following parameters (Model 6, 1 cM step size, 10 cM window size, 5 control markers, and 1,000 permutations) have been used, and chromosomes were drawn using MapChart 2.2 [34].

2.5. Sugars Biosynthesis Candidate Genes Identification

The Glyma numbers of the sucrose and RFOs biosynthesis genes were obtained by reverse BLAST of the genes underlying the RFOs pathway in *Arabidopsis* using the available data at SoyBase. The sequences of the *Arabidopsis* genes were obtained from the Phytozome database (https://phytozome-next.jgi.doe.gov ; accessed on 08/15/2023). These sequences were used for Blast in SoyBase. The obtained genes that control the RFOs pathway were mapped to the identified sugars QTL.

2.6. Expression Analysis

The expression analysis of the identified candidate genes was performed using the publicly available data from SoyBase [20] to produce the expression profiles of these candidate genes in the soybean reference genome Williams 82 in the Glyma1.0 Gene Models version.

2.7. Comparison of the Williams 82 and Forrest Sequences

Sequences of Forrest and Williams 82 cv. were obtained from the variant calling and haplotyping analysis that was performed using the 108 soybean germplasm sequenced lines as described previously [35].

3. Results

3.1. Sugars Frequency Distribution

The frequency distributions among sucrose, raffinose, and stachyose contents were quite different in the F×W82 RIL population based on Shapiro–Wilk's method for the normality test. Raffinose (2018), stachyose (2018), and sucrose (2020) were normally distributed. Only positive or negative skewness were identified in the RIL population, and all kurtosis values of these variables were positive (Table 1; Figure 1). In terms of coefficient of variation (CV), the value of sucrose 2018 showed the highest percentage of variation (62.86%) compared to other assessed traits, and the rest of the CVs appeared to be less varied within these two years. The histogram of sucrose 2018 was extremely skewed, and the other traits evaluated were normally distributed.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

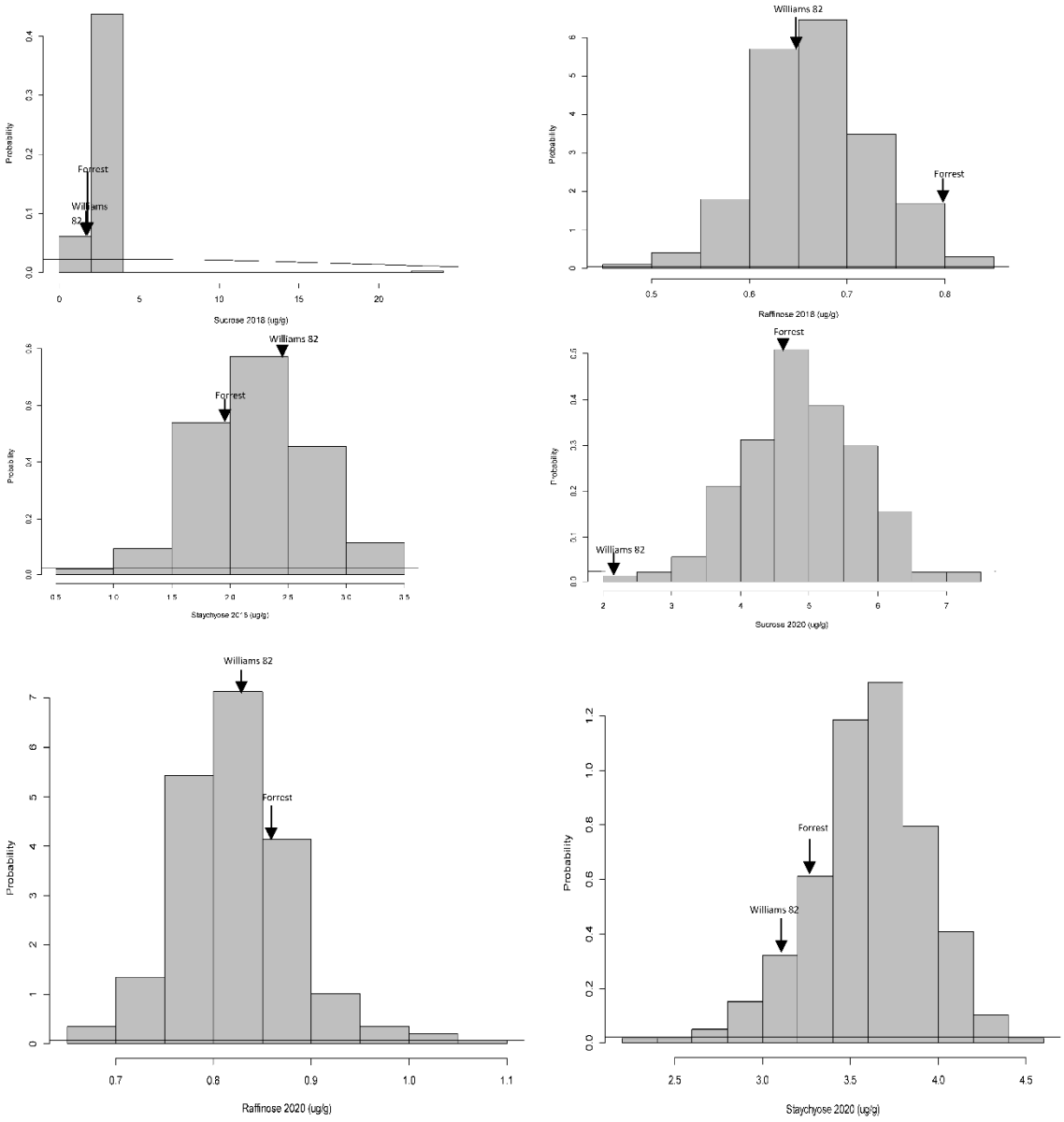

Figure 1. Frequency distribution of sugars (sucrose, raffinose, and stachyose) in the FXW82 RIL population grown in two environments over two years (Spring Lake, NC in 2018 and Carbondale, IL in 2020.

The broad-sense heritability (h^2_b) of seed sugar weight for sucrose, raffinose, and stachyose contents across two different environments appeared quite different. Stachyose had the highest heritability (92%), and the h^2_b for sucrose was 36.8% (Table 2). However, no negative h^2_b values for sugar contents were observed. The RILs-Year interactions still played a significant role in the molecular formation among three sugar contents in soybean seeds, The Sum Sq and Mean Sq to determine σ_G^2 and σ_{GE}^2 for each trait (Table 2) using type I sum of squares (ANOVA (model)) function in R program were implemented.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

Table 1. Seed sugars contents means, ranges, CVs, skewness, and kurtosis in the FxW82 RIL population evaluated in Spring Lake, NC (2018) and Carbondale, IL (2020). Mean and range values are expressed in μg/g of seed weight.

Year	Trait	Mean	Range	CV (%)	SE	Skewness	Kurtosis	W value (P<0.05)
2018	Sucrose	2.58	22.7	62.86	0.12	12.2	161.38	0.22***
	Raffinose	0.67	0.26	9.16	0.01	0.18	3.26	0.99
	Stachyose	2.23	2.55	21.74	0.03	-0.07	2.85	0.99
2020	Sucrose	4.92	4.98	17.2	0.05	-0.13	3.15	0.99
	Raffinose	0.83	0.41	7.28	0.01	0.65	4.83	0.97***
	Stachyose	3.61	2.15	9.06	0.02	-0.48	3.8	0.98**

Table 2. Two-way ANOVA of seed sugars (sucrose, stachyose, and raffinose) contents in the FxW82 RIL population evaluated in Spring Lake, NC (2018) and Carbondale, IL (2020).

Response: Sucrose				
	Df	**Sum Sq**	**Mean Seq**	**H^2**
Line	369	1134.22	3.0738	0.378
Year	1	5.6	5.5975	
Line × Year	2	3.82	1.9108	
Residuals	0	0	NA	
Response: Raffinose				
	Df	Sum Sq	Mean Seq	H^2
Line	369	3.4552	0.0093891	0.739
Year	1	0.0253	0.0253139	
Line × Year	2	0.0048	0.0023972	
Residuals	0	0	NA	
Response: Stachyose				
	Df	Sum Sq	Mean Seq	H^2
Line	369	246.73	0.66865	0.92
Year	1	1.611	1.61115	
Line × Year	2	0.106	0.05307	
Residuals	0	0	NA	

3.2. Sugars Contents QTL

IM and CIM have been used to identify QTL for seed sugar contents in this FxW82 RIL population; however, only QTL identified by CIM are presented here, although the QTL identified by the IM method were still reported in Tables S1 and S2. A total of 26 QTL that control seed sugar contents have been identified in both NC-2018 (19 QTL) and IL-2020 (7 QTL) by CIM (Tables 3 and 4; Figure S1).

In Spring Lake, NC in 2018 (NC-2018), 12 QTL that control seed sucrose content (qSUC-1–qSUC-12) have been identified and mapped on Chrs. 1, 2, 3, 4, 5, 6, 9, 10, 13, 17, 18, and 19; 4 QTL that control seed stachyose content (qSTA-1–qSTA-4) have been identified and mapped on Chrs. 13 and 19; and 3 QTL that control seed raffinose content (qRAF-1–qRAF-3) have been identified and mapped on Chr. 9 and 12 (Tables 3 and 5; Figure S1). Likewise, in Carbondale, IL in 2020 (IL-2020), 3 QTL that control seed sucrose content (qSUC-1–qSUC-3) have been identified and mapped on Chrs. 2, 5, and 8; and 4 QTL that control seed stachyose content (qSTA-1–qSTA-4) have been identified and mapped on Chrs. 13, 16, 17, and 20 (Tables 4 and 6; Figure S1). No QTL that controls seed raffinose content have been identified in this location.

No QTL for seed sugar contents have been identified by other studies within the QTL regions on chr. 4 (qSUC-4-NC-2018, 6.5–16.5 cM), chr. 10 (qSUC-8-NC-2018, 214.1–216.1 cM), and chr. 18 (qSUC-11-NC-2018, 20.1–17.5 cM), which indicates they are novel QTL regions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

Table 3. Quantitative trait loci (QTL) that control sugars (sucrose, stachyose, and raffinose) contents in FxW82 RIL population in Spring Lake, NC in 2018. These QTL have been identified by CIM method. * Indicate novel QTL.

Trait	QTL	Chr.	Marker/Interval	Position (cM)	LOD	R^2	Add. Eff.
Sucrose	*qSUC-1*	1	Gm01_3504836-Gm01_34668250	01-12.1	39.19	20.46	-3.05
	qSUC-2	2	Gm02_5155733-Gm02_9925870	128.5-142.2	42.77	47.90	4.42
	qSUC-3	3	Gm03_4595422-Gm03_411354	639.2-39.8	32.62	20.50	3.05
	*qSUC-4**	4	Gm04_7672403	6.5-16.5	54.35	37.50	4.62
	qSUC-5	5	Gm05_3867435-Gm05_327341	831.5-37.01	20.65	17.51	2.60
	qSUC-6	6	Gm06_1737718-Gm06_501439	48.5-52.4	5.36	10.50	-1.37
	qSUC-7	9	Gm09_1888876	173.9-178.1	32.62	20.50	3.05
	*qSUC-8**	10	Gm10_621706	214.01-216.01	34.25	19.10	-4.48
	qSUC-9	13	Gm13_3891723-Gm13_35248	280.2-58.2	19.12	17.51	2.60
	qSUC-10	17	Gm17_4967175-Gm17_52944	750.4-1.0	33.22	20.50	3.05
	*qSUC-11**	18	Gm18_1620585-Gm18_202082	394.7-96.5	20.10	17.51	2.60
	qSUC-12	20	Gm19_2552468	172.11	6.98	9.10	1.41
Stachyose	*qSTA-1*	13	Gm13_3524828	96.2-98.2	2.52	14.8	0.19
	qSTA-2	13	Gm13_3884070-Gm13_3803273	121.8-123.2	2.60	5.2	0.11
	qSTA-3	19	Gm19_3789399-Gm19_436261	698.01-124.1	4.21	8.5	-0.16
	qSTA-4	19	Gm19_4946208-Gm19_5032228	184.1-186.1	2.53	5.3	0.11
Raffinose	*qRAF-1*	9	Gm09_4024436-Gm09_4082234	108.01-110.9	2.26	4.6	-0.01
	qRAF-2	9	Gm09_1888876	173.9-178.1	2.47	7.6	0.08
	qRAF-3	12	Gm12_6023395-Gm12_2379195	114.6-118.6	2.15	4.7	-0.01

Table 4. Quantitative trait loci (QTL) that control sugars (sucrose, stachyose, and raffinose) contents in FxW82 RIL population in Carbondale, IL in 2020. These QTL have been identified by CIM method. * Indicate novel QTL.

Trait	QTL	Chr.	Marker	Position (cM)	LOD	R^2	Add. Eff.
Sucrose	*qSUC-1*	2	Gm02_1199805-Gm02_1373746	196.4-205.6	2.63	3.60	-0.16
	qSUC-2	5	Gm05_3803682-Gm05_3748078	18.01-22.1	2.10	0.03	-0.14
	qSUC-3	8	Gm08_5960619-Gm08_8268861	47.1-55.9	2.37	0.04	0.16
Stachyose	*qSTA-1*	13	Gm13_2748576	0.5-4.5	2.03	0.09	0.21
	qSTA-2	16	Gm16_3183754-Gm16_3010888	81.6-94.7	2.85	3.92	0.10
	qSTA-3	17	Gm17_8449684-Gm17_8352493	136.5-136.7	2.37	3.00	-0.08
	qSTA-4	20	Gm20_294157-Gm20_1133712	145.4-148.5	3.59	4.50	-0.12

Table 5. QTL and candidate genes that control sugars (sucrose, stachyose, and raffinose) contents in FxW82 RIL population in Spring Lake, NC in 2018. These QTL have been identified by CIM method. Genes with (***) are apart from the identified QTL with less than 10 MB; Genes with (**) are apart from the identified QTL with less than 20 MB; Genes with (*) are apart from the identified QTL with more than 20 MB.

Trait	QTL	Marker/Interval	LOD	R2	Wm82.a2.v1	Start	End	Wm82.a1.v1.1	Start	End	Dis. (MB)
Sucrose	*qSUC-1*	Gm01_3504836-Gm01_3466825	39.19	20.46	*Glyma.01G225800**	55452580	55456886	*Glyma01g43540*	54536305	54540597	51.03
	qSUC-2	Gm02_5155733-Gm02_9925870	42.77	47.9	*Glyma.02G016700****	1490049	1491170	*Glyma02g02030*	1475851	1476528	3.6
	qSUC-3	Gm03_4595422-Gm03_4113546	32.62	20.5	*Glyma.03G222000**	43660855	43663317	*Glyma03g38080*	44498027	44500613	39.9
					*Glyma.03G229800**	43172456	43175687	*Glyma03g38910*	45176126	45179418	40.5
					*Glyma.03G137900**	35393011	35398758	*Glyma03g29440*	37419739	37425659	32.8
					*Glyma.03G216300**	42037913	42044153	*Glyma03g37441*	44041487	44047783	39.4
	qSUC-4	Gm04_7672403	54.35	37.5	*Glyma.04G145800***	27037731	27039621	*Glyma18g23060*	26644665	26645606	18.97
					*Glyma.04G190000**	46076888	46080907	*Glyma04g36410*	42932203	42936043	35.2
	qSUC-5	Gm05_3867435-Gm05_3273418	20.65	17.51	*Glyma.05G040300****	3593378	3598821	*Glyma05g02510*	1870330	1875692	1.3
					*Glyma.05G003900****	307460	312091	*Glyma05g08950*	8806144	8810647	4.9
					*Glyma.05G217100**	39735138	39739763	*Glyma05g36850*	40599128	40603658	36.7
					*Glyma.05G185500**	37243691	37249494	*Glyma05g31920*	36953899	36959702	33.08
					*Glyma.05G236600**	41293446	41294570	*Glyma05g34830*	39054363	39055344	35.18
					*Glyma.05G204700**	38804305	38807296	*Glyma05g38120*	41530564	41533554	37.6
	qSUC-6	Gm06_1737718-Gm06_5014399	5.36	10.5	*Glyma.06G175500****	14845358	14849994	*Glyma06g18480*	14802178	14807061	9.7
					*Glyma.06G179200***	15217419	15223877	*Glyma06g18890*	15175181	15181763	10.16
	qSUC-7	Gm09_1888876	32.62	20.5	*Glyma.09G073600****	7809852	7816248	*Glyma09g08550*	7845409	7851685	5.9
					*Glyma.09G016600****	1285132	1290884	*Glyma09g01940*	1270010	1276140	0.6
					*Glyma.09G167000**	39103764	39109664	*Glyma09g29710*	36530532	36536435	34.6
	qSUC-8	Gm10_621706	34.25	19.1	*Glyma.10G017300****	1523661	1524691	*Glyma10g02170*	1519053	1519546	0.8
					*Glyma.10G214700**	44674211	44679550	*Glyma10g35890*	44094080	44098889	43.4
					*Glyma.10G145600**	38035440	38039395	*Glyma10g28640*	37509189	37513105	36.88
					*Glyma.10G145300**	38014452	38016396	*Glyma10g28610*	37488202	37490030	36.8

doi:10.20944/preprints202308.2073.v1

	qSUC-9	Gm13_3891723-Gm13_3524828	19.12	17.51	*Glyma.13G160100**	27576191	27579282	*Glyma13g22890*	26380083	26383137	22.48
					*Glyma.13G114000***	22767704	22773231	*Glyma13g17420*	21211880	21217237	17.3
	qSUC-10	Gm17_4967175-Gm17_5294475	33.22	20.5	*Glyma.17G037400****	2732048	2737399	*Glyma17g04160*	2739794	2745132	2.2
					*Glyma.17G045800****	3404918	3410491	*Glyma17g05067*	3412682	3418160	1.5
					*Glyma.17G035800****	2629011	2639005	*Glyma17g03990*	2637080	2646732	2.3
					*Glyma.17G111400****	8744555	8747526	*Glyma17g11970*	9015075	9018145	3.7
	qSUC-11	Gm18_1620585-Gm18_2020823	20.1	17.51	*Glyma.18G145700**	24414069	24415225	*Glyma18g21870*	24645144	24646447	22.6
	qSUC-12	Gm19_2552468	6.98	9.1	*Glyma.19G140700**	40199041	40201038	*Glyma19g32250*	40004601	40006724	37.4
					*Glyma.19G004400****	359933	363588	*Glyma19g00441*	238429	242106	2.3
					*Glyma.19G217700**	47033812	47037286	*Glyma19g40550*	46915407	46918937	44.3
					*Glyma.19G212800**	46633685	46639818	*Glyma19g40041*	46515393	46521627	43.9
					*Glyma.19G219100**	47148224	47150373	*Glyma19g40680*	47029812	47032065	44.4
					*Glyma.19G227800**	47911129	47914214	*Glyma19g41550*	47789168	47792321	45.2
Stachyose	*qSTA-1*	Gm13_3524828	2.52	14.8	*Glyma.13G160100**	27576191	27579282	*Glyma13g22890*	26380083	26383137	22.8
					*Glyma.13G114000***	22767704	22773231	*Glyma13g17420*	21211880	21217237	17.6
	qSTA-2	Gm13_3884070-Gm13_3803273	2.6	5.2	*Glyma.13G160100**	27576191	27579282	*Glyma13g22890*	26380083	26383137	22.4
					*Glyma.13G114000***	22767704	22773231	*Glyma13g17420*	21211880	21217237	17.3
	qSTA-3	Gm19_3789399-Gm19_4362616	4.21	8.5	*Glyma.19G004400****	359933	363588	*Glyma19g00440*	241366	241903	3.5
					*Glyma.19G140700**	40199041	40201038	*Glyma19g32250*	40004601	40006724	35.6
					*Glyma.19G217700**	47033812	47037286	*Glyma19g40550*	46915407	46918937	42.5
					*Glyma.19G212800**	46633685	46639818	*Glyma19g40041*	46515393	46521627	42.1
					*Glyma.19G219100**	47148224	47150373	*Glyma19g40680*	47029812	47032065	42.6
					*Glyma.19G227800**	47911129	47914214	*Glyma19g41550*	47789168	47792321	43.4
	qSTA-4	Gm19_4946208-Gm19_5032228	2.53	5.3	*Glyma.19G004400****	359933	363588	*Glyma19g00440*	241366	241903	4.7
					*Glyma.19G140700**	40199041	40201038	*Glyma19g32250*	40004601	40006724	34.9
					*Glyma.19G217700**	47033812	47037286	*Glyma19g40550*	46915407	46918937	41.8
					*Glyma.19G212800**	46633685	46639818	*Glyma19g40041*	46515393	46521627	41.4
					*Glyma.19G219100**	47148224	47150373	*Glyma19g40680*	47029812	47032065	41.9

					*Glyma.19G227800**	47911129	47914214	*Glyma19g41550*	47789168	47792321	42.7
Raffinose	*qRAF-1*	Gm09_4024436-Gm09_4082234	2.26	4.6	*Glyma.09G073600****	7809852	7816248	*Glyma09g08550*	7845409	7851685	3.7
					*Glyma.09G016600****	1285132	1290884	*Glyma09g01940*	1270010	1276140	2.7
					*Glyma.09G167000****	39103764	39109664	*Glyma09g29710*	36530532	36536435	
	qRAF-2	Gm09_1888876	2.47	7.6	*Glyma.09G073600****	7809852	7816248	*Glyma09g08550*	7845409	7851685	5.9
					*Glyma.09G016600****	1285132	1290884	*Glyma09g01940*	1270010	1276140	0.6
					*Glyma.09G167000**	39103764	39109664	*Glyma09g29710*	36530532	36536435	32.4
	qRAF-3	Gm12_6023395-Gm12_2379195	2.15	4.7	*Glyma.12G162600**	30862398	30862873	*Glyma12g26693*	30087270	30088386	24.06

Table 6. QTL and candidate genes that control sugars (sucrose, stachyose, and raffinose) contents in FxW82 RIL population in Carbondale, IL in 2020. These QTL have been identified by CIM method. Genes with (***) are apart from the identified QTL with less than 10 MB; Genes with (**) are apart from the identified QTL with less than 20 MB; Genes with (*) are apart from the identified QTL with more than 20 MB.

Trait	QTL	Marker	LOD	R2	Wm82.a2.v1	Start	End	Wm82.a1.v1.1	Start	End	Dis. (MB)
Sucrose	*qSUC-1*	Gm02_1199805-Gm02_1373746	2.63	3.6	*Glyma.02G016700****	1490049	1491170	*Glyma02g02030*	1475851	1476528	0.2
	qSUC-2	Gm05_3803682-Gm05_3748078	2.1	0.03	*Glyma.05G040300****	3593378	3598821	*Glyma05g02510*	1870330	1875692	1.8
					*Glyma.05G003900****	307460	312091	*Glyma05g08950*	8806144	8810647	5.002
					*Glyma.05G217100**	39735138	39739763	*Glyma05g36850*	40599128	40603658	36.7
					*Glyma.05G185500**	37243691	37249494	*Glyma05g31920*	36953899	36959702	33.1
					*Glyma.05G236600**	41293446	41294570	*Glyma05g34830*	39054363	39055344	35.2
					*Glyma.05G204700**	38804305	38807296	*Glyma05g38120*	41530564	41533554	37.7
	qSUC-3	Gm08_5960619-Gm08_8268861	2.37	0.04	*Glyma.08G043800****	3450235	3451725	*Glyma08g04860*	3446035	3447462	2.5
					*Glyma.08G143500****	10949673	10956219	*Glyma08g15220*	11038816	11045375	2.7
					*Glyma.08G011800****	942037	944988	*Glyma08g01480*	939512	942346	5.01
					*Glyma.08G023100****	1852651	1856671	*Glyma08g02690*	1848105	1853380	4.1
Stachyose	*qSTA-1*	Gm13_2748576	2.03	0.09	*Glyma.13G160100**	27576191	27579282	*Glyma13g22890*	26380083	26383137	23.6
					*Glyma.13G114000***	22767704	22773231	*Glyma13g17420*	21211880	21217237	18.4
	qSTA-2	Gm16_3183754-Gm16_3010888	2.85	3.92	*Glyma.16G217200**	37414228	37419838	*Glyma16g34290*	36921346	36926746	33.7

doi:10.20944/preprints202308.2073.v1

qSTA-3	Gm17_8449684-Gm17_8352493	2.37	3	*Glyma.17G037400****	2732048	2737399	*Glyma17g04160*	2739794	2745132	5.6
				*Glyma.17G045800****	3404918	3410491	*Glyma17g05067*	3412682	3418160	4.9
				*Glyma.17G035800****	2629011	2639005	*Glyma17g03990*	2637080	2646732	*5.8*
				*Glyma.17G111400****	8744555	8747526	*Glyma17g11970*	9015075	9018145	0.5
qSTA-4	Gm20_294157-Gm20_1133712	3.59	4.5	*Glyma.20G177200**	41446962	41451980	*Glyma20g31730*	40330117	40334860	40.03
				*Glyma.20G095200**	33827363	33831352	*Glyma20g22780*	32686241	32690264	32.3
				*Glyma.20G094500**	33759416	33761555	Glyma20g22700	32618509	32620443	32.3

doi:10.20944/preprints202308.2073.v1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

3.3. In silico Sucrose, Raffinose and Stachyose Biosynthetic Pathway Genes in Soybean

The sugars (sucrose, raffinose, and stachyose) biosynthetic pathway was studied in many plants, including the plant model *Arabidopsis thaliana* [36,37] and the leguminous model *Medicago sativa* L. [38]. A reverse BLAST of the genes identified in *Arabidopsis thaliana* was conducted using the SoyBase [20] to reconstruct the sugars (sucrose, raffinose, and stachyose) biosynthetic pathway in soybean (Figure 2).

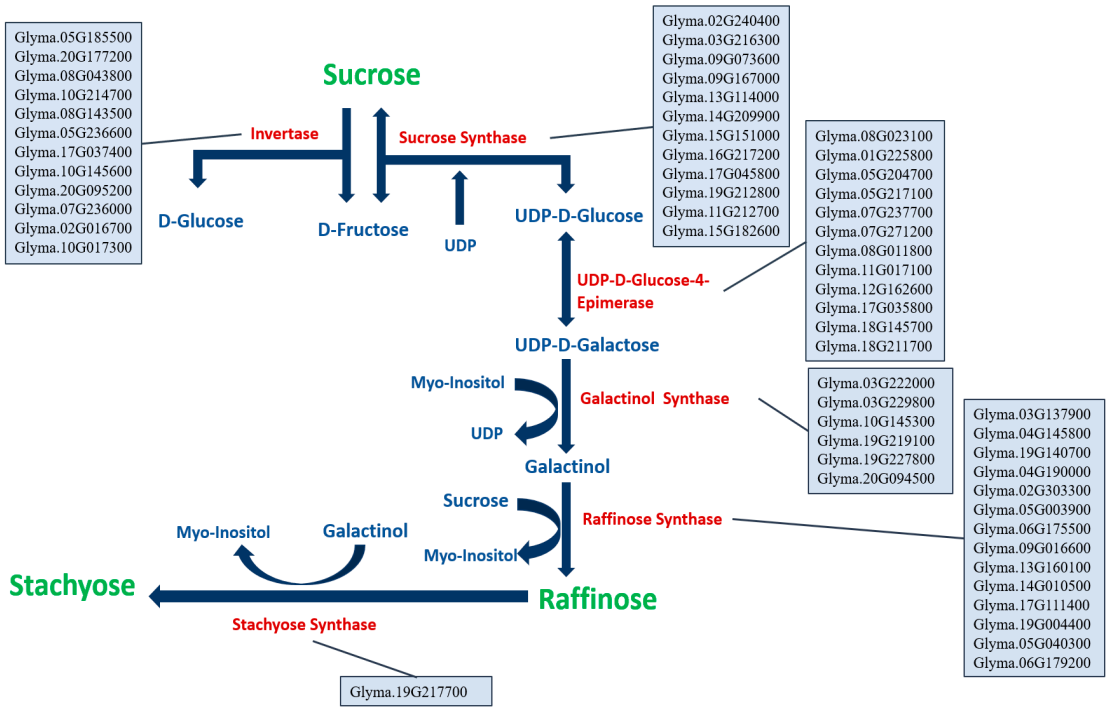

Figure 2. The sugars (sucrose, raffinose, and stachyose) biosynthetic pathway with the identified candidate genes in soybean. The genes are in Wm82.a2.v1 annotation.

A total of fifty-seven candidate genes were identified to underly the sugar (sucrose, raffinose, and stachyose) biosynthetic pathway (Figure 2). In this pathway, twelve candidate genes were identified for invertase including *Glyma.05G185500, Glyma.20G177200, Glyma.08G043800, Glyma.10G214700, Glyma.08G143500, Glyma.05G236600, Glyma.17G037400, Glyma.10G145600, Glyma.20G095200, Glyma.07G236000, Glyma.02G016700,* and *Glyma.10G017300*. Twelve candidate genes were identified for sucrose synthase including *Glyma.02G240400, Glyma.03G216300, Glyma.09G073600, Glyma.09G167000, Glyma.13G114000, Glyma.14G209900, Glyma.15G151000, Glyma.16G217200, Glyma.17G045800, Glyma.19G212800, Glyma.11G212700,* and *Glyma.15G182600*. Twelve candidate genes were identified for UDP-D-Glucose-4-Epimerase, *Glyma.08G023100, Glyma.01G225800, Glyma.05G204700, Glyma.05G217100, Glyma.07G237700, Glyma.07G271200, Glyma.08G011800, Glyma.11G017100, Glyma.12G162600, Glyma.17G035800, Glyma.18G145700,* and *Glyma.18G211700*. For the galactinol synthase, six candidate genes were identified, including *Glyma.03G222000, Glyma.03G229800, Glyma.10G145300, Glyma.19G219100, Glyma.19G227800,* and *Glyma.20G094500*. Fourteen candidate genes were identified for raffinose synthase, *Glyma.03G137900, Glyma.04G145800, Glyma.19G140700, Glyma.04G190000, Glyma.02G303300, Glyma.05G003900, Glyma.06G175500, Glyma.09G016600, Glyma.13G160100, Glyma.14G010500, Glyma.17G111400, Glyma.19G004400, Glyma.05G040300,* and *Glyma.06G179200*. For the stachyose synthase, only one candidate gene was identified *Glyma.19G217700* (Figure 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

3.4. Association between the Identified sugar (sucrose, raffinose, and stachyose) Biosynthetic Pathway Candidate Genes and Reported QTL

The identified genes for the sugar (sucrose, raffinose, and stachyose) biosynthesis in soybean have been mapped to the identified QTL. Amongst fifty-seven candidate genes, sixteen have been located less than 10 MB to the identified QTL on chrs. 2, 5, 6, 8, 9, 10, 17, and 19 (Tables 3, 4, 5, 6).

The sucrose synthase candidate gene *Glyma.09G073600* and the raffinose synthase candidate gene *Glyma.09G016600* are positioned close to the *qSUC-7-IL-2018*, *qRAF-1-IL-2018*, and *qRAF-2-IL-2018* on Chr.9 (Tables 3, 4, 5, 6). The invertase candidate gene *Glyma.02G016700* is located 3.6 and 0.2 MB apart from the *qSUC-1-IL-2018* and *qSUC-1-NC-2020,* respectively, on Chr. 2 (Tables 3, 4, 5, 6). The raffinose synthase candidate genes *Glyma.05G003900* and *Glyma.05G040300* are located close to the *qSUC-5-IL-2018* and *qSUC-2-NC-2020* on Chr. 5 (Tables 3, 4, 5, 6). On chr. 6, the raffinose synthase candidate gene *Glyma.06G175500* is located close to the *qSUC-6-IL-2018* (Tables 3, 4, 5, 6). The invertase candidate genes *Glyma.08G043800,* and *Glyma.08G143500;* and the UDP-D-Glucose-4-Epimerase candidate genes *Glyma.08G011800* and *Glyma.08G023100* on chr. 8 are located close to the *qSUC-3-NC-2020* (Tables 3, 4, 5, 6, S3, S4). On chr. 10, the invertase candidate gene *Glyma.10G017300* is located close to the *qSUC-8-IL-2018* (Tables 3, 4, 5, 6). On Chr. 17, a cluster of four genes involved in the sugars' pathway is collocated within 6 MB with two QTL (qSUC-10-NC-2018 and qSTA-3-IL-2020) that were identified in this study. These genes are the *Glyma.17G037400* encoding for an invertase, *Glyma.17G045800* encoding for sucrose synthase, *Glyma.17G111400* encoding for raffinose synthase, and *Glyma.17G035800* encoding for UDP-D-glucose-4-epimerase (Tables 3, 4, 5, 6, Figure S3.). The raffinose synthase candidate gene *Glyma.19G004400* is positioned close to the *qSTA-3-IL-2018* and the *qSTA-4-IL-2018* (Tables 3, 4, 5, 6), as well as the *qRAF-8-IL-2018* and *qRAF-9-IL-2018* identified using the IM method (Tables 3, 4).

3.5. Association between the Identified Candidate Genes and the Previously Reported QTL

Several studies have identified and mapped QTL underlying seed sugar content using different populations and mapping methods [39–42], as summarized in [18].

The identified genes have been mapped to the previously reported QTL regions associated with the seed sugar content using the data from SoyBase [18,20,43] six candidate genes have been located within the identified seed sugars QTLs and 18 have been located < 9 MB apart from these regions (Table 7). Among them is the invertase candidate gene *Glyma.08G143500* that is located within the seed sucrose 1-2 QTL on Chr. 8 [20,39]. Also, the galactinol-sucrose galactosyl-transferase 6-related candidate gene *Glyma.13G160100* is situated within the seed sucrose 1-5 QTL [20,39](Table 7). Likewise, the raffinose synthase candidate gene *Glyma.19G140700* is collocated within the seed sucrose 1-8 QTL [20,39], less than < 0.5 MB apart from seed sucrose 2-11 and seed sucrose 2-10 [20,41], and 1.9 MB from seed oligosaccharide 2-7 [20,40].

Table 7. Candidate genes controlling sugars (sucrose, stachyose, and raffinose) contents associated with previously reported QTL.

Gene ID	Start	End	QTL	QTL Start	QTL End	Reference
Glyma.02G240400	42892680	42898279	Seed sucrose 2-2	39547350	41441274	[41]
			Seed oligosaccharide 1-1	39547350	41441274	[41]
Glyma.05G236600	41293446	41294570	Seed sucrose 1-1	3924139	4279362	[39]
Glyma.08G043800	3450235	3451725	Seed sucrose 1-3	7892162	8937354	[39]
Glyma.08G143500	10949673	10956219	Seed sucrose 1-2	10865328	13126779	[39]
Glyma.09G073600	7809852	7816248	Seed sucrose 4-2	2973041	5901485	[44]
Glyma.13G114000	22767704	22773231	Seed sucrose 1-5	26196486	28912864	[39]
Glyma.14G209900	47515899	47521687	Seed sucrose 3-1	38859467	40060720	[40]
			Seed oligosaccharide 2-1	38859467	40060720	[40]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

Glyma.15G151000	12497113	12508050	Seed sucrose 3-3	13755345	17021739	[40]
			Seed oligosaccharide 2-3	13755345	17021739	[40]
Glyma.19G140700	40199041	40201038	Seed sucrose 1-8	40205349	40265091	[39]
			Seed oligosaccharide 2-7	42119600	43329204	[40]
Glyma.19G212800	46633685	46639818	Seed oligosaccharide 2-7	42119600	43329204	[40]
			qSU1901	45311975	45464136	[43]
Glyma.19G217700	47033812	47037286	Seed oligosaccharide 2-7	42119600	43329204	[40]
			qSU1901	45311975	45464136	[43]
Glyma.20G095200	33827363	33831352	Seed sucrose 1-4	2716974	25498552	[39]
Glyma.08G011800	942037	944988	Seed sucrose 1-3	7892162	8937354	[39]
			Seed sucrose 1-13	8283676	9192408	[39]
Glyma.08G023100	1852651	1856671	Seed sucrose 1-3	7892162	8937354	[39]
			Seed sucrose 1-13	8283676	9192408	[39]
Glyma.19G219100	47148224	47150373	Seed sucrose 1-8	40205349	40265091	[39]
			Seed sucrose 2-10	40637071	41616190	[41]
			Seed sucrose 2-11	40637071	41616190	[41]
			Seed oligosaccharide 2-7	42119600	43329204	[40]
Glyma.19G227800	47911129	47914214	Seed sucrose 1-8	40205349	40265091	[39]
			Seed sucrose 2-10	40637071	41616190	[41]
			Seed sucrose 2-11	40637071	41616190	[41]
			Seed oligosaccharide 2-7	42119600	43329204	[40]
Glyma.20G094500	33759416	33761555	Seed sucrose 1-4	2716974	25498552	[39]
Glyma.20G177200	41446962	41451980	qSU2002	40523599	41882459	[43]
Glyma.15G182600	17910130	17916426	Seed sucrose 3-3	13755345	17021739	[40]
			Seed oligosaccharide 2-3	13755345	17021739	[40]
Glyma.05G003900	307460	312091	Seed sucrose 1-1	3924139	4279362	[39]
Glyma.09G016600	1285132	1290884	Seed sucrose 4-2	2973041	5901485	[44]
Glyma.17G111400	8744555	8747526	qSS1701	7470395	10014816	[43]
			qSS1702	7969537	10599548	[43]
Glyma.13G160100	27576191	27579282	Seed sucrose 1-5	26196486	28912864	[39]
Glyma.19G004400	359933	363588	Seed sucrose 2-3	4244065	12744826	[41]
			Seed oligosaccharide 1-2	4244065	12744826	[41]
			Seed sucrose 2-6	9284015	34059981	[41]
			Seed oligosaccharide 1-5	9284015	34059981	[41]

The sucrose synthase candidate gene *Glyma.02G240400* was located close (< 1.5 MB) to two QTL controlling seed sugar contents, the seed sucrose 2-2 and seed oligosaccharide 1-1 [20,41]. Moreover, the raffinose synthase candidate gene *Glyma.05G003900* is located less than < 4 MB apart from the seed sucrose 1-1 [20,39]. The raffinose synthase candidate gene *Glyma.19G004400* is located less than 9 MB apart from four QTL controlling the sugar contents, namely seed sucrose 2-3, seed oligosaccharide 1-2, seed sucrose 2-6, and seed oligosaccharide 1-5 [20,41] (Table 7). On chr. 8, the seed sucrose 1-3 and seed sucrose 1-13 are located close to the invertase candidate genes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

Glyma.08G043800, and *Glyma.08G143500*; as well as UDP-D-glucose-4-epimerase candidate genes *Glyma.08G011800* and *Glyma.08G023100* [20,39] (Table 7). The sucrose synthase candidate gene *Glyma.09G073600* and the raffinose candidate gene *Glyma.09G016600* are positioned less than < 2 MB apart from the seed sucrose 4-2 [20,44] (Table 7). Interestingly, the sucrose synthase candidate genes *Glyma.15G182600* and *Glyma.15G151000* are located less than < 1.25 MB from the seed sucrose 3-3 and seed oligosaccharide 2-3 [20,40].

3.6. Organ-specific Expression of the Identified Candidate Genes

The expression pattern of the identified candidate genes was investigated in Williams 82 cv. using the RNA-seq data available at SoyBase [20]. The dataset includes several plant tissues, including leaves, nodules, roots, pods, and seeds (Figure 3A, 3B, and S2). Four of the fifty-seven identified candidate genes have no available RNA-seq data, including the sucrose synthase candidate genes *Glyma.03G216300*, *Glyma.17G045800*, and *Glyma.19G212800*, as well as the UDP-D-glucose-4-epimerase candidate genes *Glyma.18G211700* (Figure S2). The raffinose synthase candidate gene *Glyma.04G145800* is not expressed in any of the analyzed tissues, whilst the rest of the identified genes showed different expression patterns.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

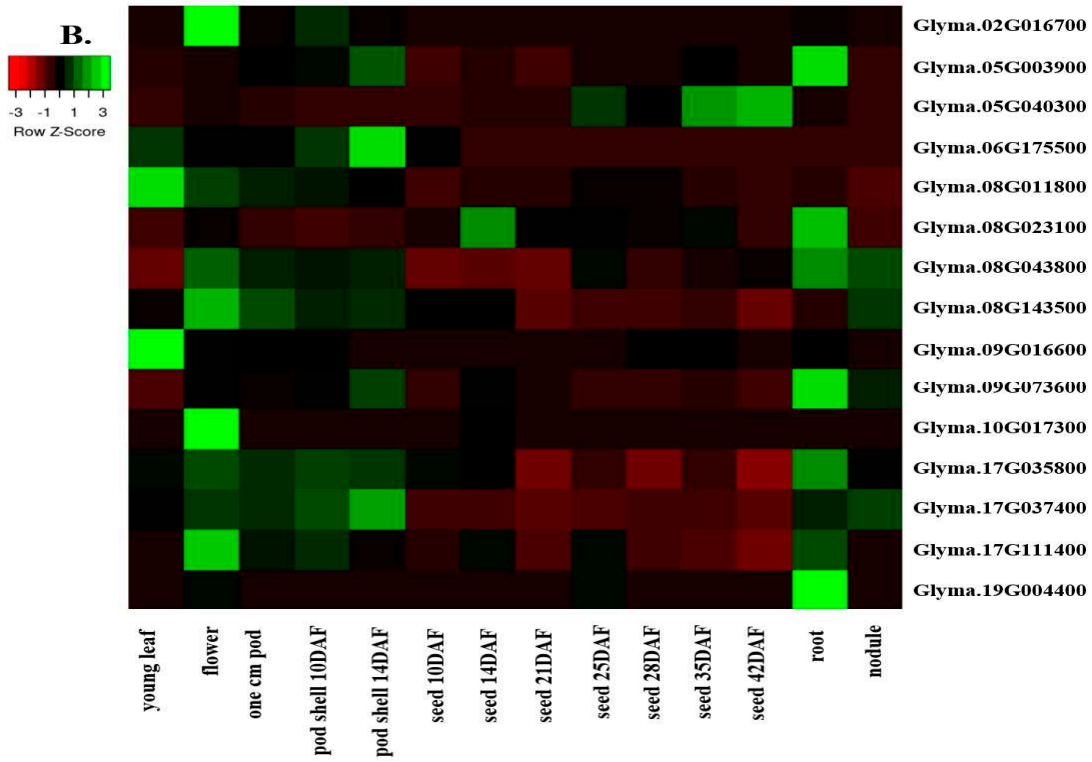

Figure 3. A. Tissue specific expression of the identified sugars candidate genes. **B.** Expression HeatMap of the identified candidate genes located within 10 MB to the identified sugars QTL regions in Williams 82 (RPKM) were retrieved from publicly available RNA-seq data from Soybase database [20]. RNA-seq data is not available at Soybase for the *Glyma.17G045800* candidate gene.

The sucrose synthase candidate genes *Glyma.09G073600* and *Glyma.13G114000* present a high expression profile in all the analyzed tissues except for the young leaves, while the raffinose synthase candidate gene *Glyma.17G111400* is abundantly expressed in all the analyzed tissues except for the seeds and nodules. Interestingly, the sucrose synthase candidate gene *Glyma.15G182600* is highly expressed in all the tissues excluding the young leaves and the nodules. The raffinose synthase candidate gene *Glyma.03G137900* is abundantly expressed in flowers, nodules, and roots. The raffinose synthase candidate gene *Glyma.14G010500* and the invertase candidate gene *Glyma.05G236600* are highly expressed in the flowers and pods. Also, The UDP-D-glucose-4-epimerase candidate gene *Glyma.05G204700* is abundantly expressed in the flowers and seeds. While the invertase candidate gene *Glyma.20G177200* is highly expressed in nodules and roots, the raffinose synthase candidate gene *Glyma.06G179200* was found to be highly expressed in seed (Figure 3A, Figure S2).

Seventeen of the identified candidate genes were situated less than 10 MB apart from the identified QTL regions. *Glyma.09G073600* is highly expressed in seeds in Williams 82 cv., followed by *Glyma.17G111400*, *Glyma.17G035800*, and *Glyma.08G043800* with moderated expression profile. The remaining genes have lower expression patterns, excluding the *Glyma.02G016700, Glyma.06G175500, Glyma.09G016600, Glyma.10G017300,* and *Glyma.19G004400* genes that are not expressed in seeds in Williams 82 cv.

3.7. Comparison of the Williams 82 and Forrest Sequences

The sequences of the candidate genes that are located less than 10 MB from the identified QTL were compared. The results have shown that six of them have SNPs and InDels between Forrest and Williams 82 sequences, *Glyma.09G073600, Glyma.08G143500, Glyma.05G003900, Glyma.17G035800, Glyma.17G111400*, and *Glyma.09G016600* (Tables. S4, Figure 4).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2023 doi:10.20944/preprints202308.2073.v1

The sucrose synthase *Glyma.09G073600* has in total 28 SNPs and 7 InDels; three of these SNPs are located upstream the 5'UTR, two are downstream the 3'UTR, and seven are located in the exons (Table S4, Figure 4). For the invertase candidate gene *Glyma.08G143500*, there are 20 SNPs and 5 InDels. One of these SNPs is located in exon 7, causing a missense mutation, and two SNPs are located upstream the 5'UTR (Table S4, Figure 4). The raffinose synthase candidate gene *Glyma.05G003900* has 9 SNPs and one InDel, four of those SNPs are in the exons, from which two SNPs resulted in missense mutations (Table S4, Figure 4). Likewise, the raffinose synthase candidate gene *Glyma.09G016600* possesses 12 SNPs and 2 InDels. Amongst these SNPs, there are two located in exons that resulted in missense mutations in addition to the 2 InDels located in the exons (Table S4, Figure 4). For the raffinose candidate gene *Glyma.17G111400*, 8 SNPs were found from which one is located upstream the 5' UTR, another one is downstream the 3'UTR, and the last six are in exons causing silent mutations (Table S4, Figure 4). Finally, the UDP-D-Glucose-4-Epimerase candidate gene *Glyma.17G035800* has two SNPs that are positioned in introns (Table S4).

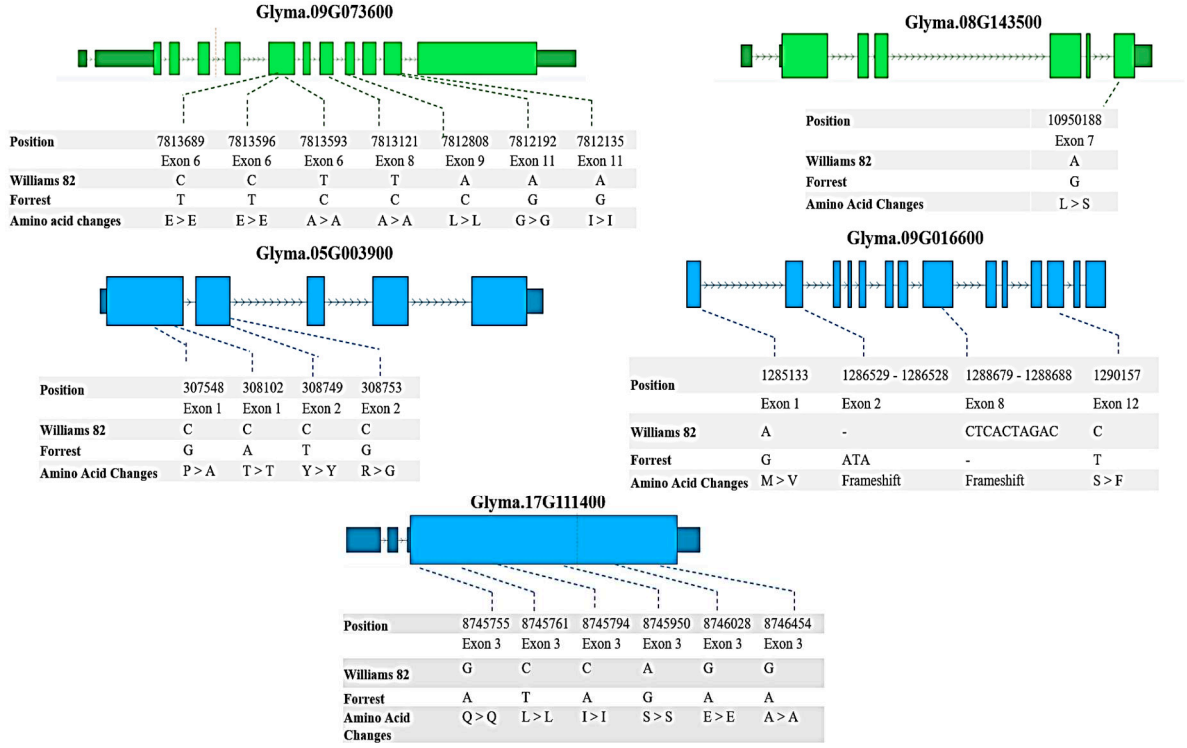

Figure 4. Positions of SNPs between Forrest and Williams 82 cultivars in *Glyma.09G073600, Glyma.08G143500, Glyma.05G003900, Glyma.17G111400,* and *Glyma.09G016600* coding sequences. In the gene model diagram, the light blue/light green boxes represent exons, blue/green bars represent introns, dark blue/dark green boxes represent 3'UTR or 5'UTR. SNPs were positioned relative to the genomic position in the genome version W82.a2.

4. Discussion

Soybean seed sugars play a major role in seed and fruit development. Recently, soy products became very popular as a result of a growing demand for vegan diets [45]. The quality and taste of these products are determined by soybean seed sugar content [39]. These sugars include sucrose, raffinose, and stachyose that make up to 5–7%, 1%, and 3–4% of total carbohydrates, respectively [5]. However, the raffinose and stachyose fermentation by humans and monogastric animal intestines microbes leads to a reduced gastrointestinal performance, flatulence, and diarrhea. Thus, reducing raffinose and stachyose and increasing sucrose in soybean seed content are desirable[22,27].

Knowing the importance of soybean seed sucrose content in the quality of the soybean based products for food and feed, breeding programs are focused on developing soybean seeds with high

sucrose content and low RFOs content [43,46]. Thus, soybean varieties with high sucrose content are valuable for soybean food and feed products [47].

The identification of QTL associated with sugars components using different types of molecular markers is one of the breeding process approaches that researches use to breed for a high sucrose soybean.

In the current study, all seed sugar (sucrose, raffinose, and stachyose) phenotypic data exhibited normal distributions in all environments studied (years and locations), showing that these traits are polygenic and complex as shown earlier [21,39–41,44,47–53].

The SNP-based genetic linkage map facilitated the identification of several QTL including QTL for seed isoflavone contents [28], seed tocopherol contents [29], and seed sugar (sucrose, stachyose, and raffinose) contents as reported in the current study.

A total of 26 QTL that control seed sugar contents have been identified in both IL-2018 and NC-2020 by CIM. Among these, three are novel QTL regions, including qSUC-4, qSUC-8, and qSUC-11 mapped on chrs. 4, 10, and 18, respectively. All the other sugar QTL reported in this study have been located within or very close to other sugar QTL previously reported [30,39–41,44] as summarized in [18]. Five other genomic regions on chrs. 2, 6, 12, 16, and 19 harboring sugar QTL either from this study or from other studies are of particular interest. On chr. 2, qSUC-2-NC-2018 may correspond to *suc 1-1* identified previously [39]. This QTL region contains the *Glyma.02G016700* candidate gene that encodes for invertase.

Interestingly, several QTL have been identified previously including a QTL that controls seed raffinose content within the qSUC-1-NC-2018 region (chr. 1) [30], two QTL (suc 2-2 and suc 3-2) that control seed sucrose content within the qSUC-2-NC-2018 region (chr. 2) [20,40,41], a QTL that controls seed sucrose content (suc-001) within the qSUC-3-NC-2018 region (chr. 3), [30]; 2 QTL that control seed sucrose (suc 1-1 and suc 4-1) content within the qSUC-5-NC-2018 region (chr. 5) [39,44]; a QTL that controls seed raffinose content (raf003 and raf004) within the qSUC-6-NC-2018 and qSUC-7-NC-2018 regions (chrs. 6 and 9), [30]; a QTL that controls seed sucrose (suc 1-5) content within the qSUC-9-NC-2018 region (chr. 13), [39]; and a QTL that controls seed sucrose (suc 1-4) content within the qSUC-12-NC-2018 region (chr. 20) [39].

Likewise, several other QTL have been identified previously a QTL that controls seed sucrose (suc 2-2, 3-2) content within the qSUC-1-IL-2020 region (chr. 2)[40,41]; QTL that control seed sucrose (suc 1-1, 4-1) content within the qSUC-2-IL-2020 region (chr. 5) [39,44]; and within qSUC-3-IL-2020 region on chr. 8, QTL that control seed sucrose (suc 1-2, 1-3, 1-13) content within the qSUC-3-IL-2020 region (chr. 8)[39]. Within the QTL regions that were found to control seed stachyose contents (qSTA-1-IL-2020, qSTA-2-IL-2020, and qSTA-4-IL-2020) reported in the current study on chrs. 13, 16, and 19, several QTL that control seed sucrose (suc 1-4, 1-5, 3-5, 3-6) and seed raffinose (raff007) contents have been identified previously [39–41].

On chr. 6, qSUC-6-NC-2018 most likely corresponds to *suc 2-2* [41] and raffinose (*raf003*) QTL regions identified previously [30,39]. The QTL region contains *Glyma.06G175500* candidate gene encoding for raffinose synthase. Interestingly, the genomic region on chr. 19 comprising a cluster of sucrose QTL (suc 1-6 to 1-8, 2-3 to 2-11) [39,41] also contains two stachyose QTL identified in this study (qSTA-3-NC-2018 and qSTA-4-NC-2018). The candidate gene *Glyma.19G004400*, that also encodes for raffinose synthase was identified within this QTL region.

No candidate genes have been identified on chrs. 12 (qRAF-3-NC-2018), 16 (qSTA-2-NC-2018), and 20 (qSTA-4-NC-2018).

Remarkably, within the novel QTL regions reported here on chrs. 4, 10, and 18, seven candidate genes have been identified; including the *Glyma.18G145700* encoding for UDP-D-glucose-4-epimerase on chr. 18 (Tables 5, 6, and Figure 2).

Interestingly, five QTL regions were detected in both locations, IL and NC; The first QTL region contains qSUC-5-NC-2018 and qSUC-2-IL-2020 that were detected in the same location on chr. 5. Additionally, the qSUC-9-NC-2018, qSTA-1-NC-2018, and qSTA-2-NC-2018 were located only 1 MB apart from qSTA-1-IL-2020 on chr.13. Moreover, qSUC-12-NC-2018 was 1.3 MB away from qSTA-4-IL-2020 on chr. 20. Furthermore, qSUC-10-NC-2018 and qSTA-3-IL-2020 were positioned 3.1 MB apart

from each other on chr. 17. Additionally, qSUC-2-NC-2018 and qSUC-1-IL-2020 were located ~4 MB apart on chr. 2. The QTL regions that were not detected in both locations may be affected by environmental conditions.

In a previous study [54], 31,245 SNPs and 323 soybean germplasm accessions grown in three different environments were used to identify 72 QTL associated with individual sugars and 14 associated with total sugar [54]. In addition, ten candidate genes that are within the 100 Kb flanking regions of the lead SNPs in six chromosomes were significantly associated with sugar content in soybean; eight of them are involved in the sugar metabolism in soybean [54]. Amongst these candidate genes, the raffinose synthase gene *Glyma.05G003900* is also reported in this study.

A recent study used a RIL population from a cross of ZD27 by HF25 to identify 16 QTL controlling seed sucrose and soluble sugars content in soybean [43]. Amongst these QTL, qSU1701[43] with a LOD = 7.61 and phenotypic variation explained (PVE)= 16.8 % was identified on chr. 17 to be associated with the seed sucrose content. This QTL region is collocated with the qSUC-10-NC-2018 identified in this study for the same trait with a LOD = 33.2 and an R^2= 20.5. On the same chr., qSS1701 [43] and qSS1702 identified to be associated with the seed soluble sugar content are collocated with the qSTA-3-IL-2020. These QTL are positioned within less than 8 MB with a cluster of four genes involved in the sugars' pathway, including the *Glyma.17G037400* encoding for invertase, *Glyma.17G045800* encoding for sucrose synthase, *Glyma.17G111400* encoding for raffinose synthase (showing 7 SNPs variations in exons) (Figure 4), and *Glyma.17G035800* encoding for UDP-D-glucose-4-epimerase. Our results confirm that this region on chr. 17 is a major QTL associated with seed sugars content in soybean. In the same study [43], qSU2001 identified on chr. 20 with LOD=3.38 and PVE=5.6 % is collocated with the qSUC-12-NC-2018, and 0.3 MB apart from the qSTA-4-IL-2020. The invertase candidate gene Glyma.20G177200 is positioned within the qSU2002 [43] identified on chr. 20 with LOD=7.9 and PVE=14.4 %. These results confirm that this region on chr 20 is involved in soybean seed sugar contents. On chr. 3, qSS0301 was previously identified [43] to be associated with soluble sugar content in soybean with a LOD= 5.2 and PVE= 11.8. This QTL is located 1.4 MB apart from qSUC-3-NC-2018.

Although the major QTL qSU1901 reported in a previous study [43] on chr. 19 is ~40MB away from the qSTA-3-NC-2018 and qSTA-4-NC-2018, it could be that the gene(s) underlying this QTL are different or not due to chromosomal rearrangement that happened in ZD27 by HF25 population versus Forrest by Williams 82 population. Those QTL regions on chr. 19 were reported in several studies and could be subject to further high-density genetic mapping to isolate genes that underly sugar content in soybean seeds.

The sucrose synthase gene *Glyma.09G073600* was highly expressed in seeds, followed by *Glyma.17G111400*, *Glyma.17G035800*, and *Glyma.08G043800* with moderated expression patterns in seeds. *Glyma.09G073600* and *Glyma.09G016600* are located close to the qSUC-7-IL-2018, qRAF-1-IL-2018, and qRAF-2-IL-2018 on Chrs.9. *Glyma.08G143500* is located close to the qSUC-3-NC-2020, and *Glyma.05G003900* is positioned close to the qSUC-5-IL-2018 and qSUC-2-NC-2020 on Chr. 5. These genes could be useful in gene editing technology or breeding programs to develop soybean cultivars with reduced amounts of RFOs, and high amounts of sucrose which is beneficial for human consumption and animal feed.

Further studies are needed to characterize these genes, identify their enzymes and protein products, understand their roles in the sugar's biosynthetic pathway in soybean.

5. Conclusions

In summary, we have identified 26 QTL associated with the seed sugars contents and 57 candidate genes involved in sucrose, raffinose, and stachyose biosynthetic pathway. Amongst these candidate genes, 16 were located less than 10 MB apart from the QTL regions identified in this study.

On chr. 17, a cluster of four genes controlling the sugar pathway is collocated within 6 MB with two QTL (*qSUC-10-NC-2018* and *qSTA-3-IL-2020)* that were identified in this study. Moreover, the raffinose synthase candidate gene *Glyma.06G175500* is 9.7MB apart from the qSUC-6-NC-2018 QTL on chr. 6. The invertase candidate gene *Glyma.02G016700* is located 3.6 and 0.2 MB apart from qSUC-

1-NC-2018 (R^2=47.9) *and qSUC-1-IL-2020* (R^2=3.6) respectively, on chr. 2. Moreover, the sucrose synthase candidate gene *Glyma.09G073600* and the raffinose synthase candidate gene *Glyma.09G016600* were found close to the qSUC-7-IL-2018, qRAF-1-IL-2018, qRAF-2-IL-2018, and qRAF-1-IL-2018 on chr. 9.

Five QTL regions were commonly identified in the two environments, NC and IL, on chrs. 2, 5, 13, 17 and 20, ((qSUC-5-NC-2018 and qSUC-2-IL-2020), (qSUC-9-NC-2018 and qSTA-1-NC-2018, qSTA-1-IL-2020), (qSUC-12-NC-2018, qSTA-4-IL-2020), (qSUC-10-NC-2018 and qSTA-3-IL-2020), and (qSUC-2-NC-2018 and qSUC-1-IL-2020)).

Five genes (*Glyma.09G073600, Glyma.08G143500, Glyma.17G111400, Glyma.05G003900,* and *Glyma.09G016600*) have SNPs and InDels between Forrest and Williams 82 sequences. These SNPs could potentially explain the difference in sugar content between Forrest and Williams 82 cultivars.

Further studies are required to functionally characterize these genes understand and validate their roles in the sugar's biosynthetic pathway in soybean, before being used in breeding programs to produce soybean lines with high beneficial sucrose and low RFOs.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, K.M. and M.A.K. and; methodology, D.K., J.Y., T.V., N.L., A.M., E.A., N.B. and M.E.; validation, M.A.K., K.M. and H.T.N.; formal analysis, D.K., J.Y. and N.B.; investigation, K.M., and M.A.K.; resources and data curation, K.M., M.A.K. and H.T.N.; writing—original draft preparation, D.K., M.A.K. and K.M.; review and editing, D.K., J.Y., N.B., N.L., T.V., M.A.K., K.M. and H.T.N.; supervision, M.A.K., K.M.; project administration, M.A.K., K.M., and H.T.N. All authors have read and agreed to the published version of the manuscript.

Acknowledgements: Technical support provided by Sandra Mosley is appreciated. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture (USDA). USDA is an equal opportunity provider and employer. This research was partially funded by the U.S. Department of Agriculture, Agricultural Research Service Project 6066-21220-014-000D. This project was partially funded by the United Soybean Board, project # 2220-152-0104 and Southern Illinois University at Carbondale.

References

1. Blackman, S.A.; Obendorf, R.L.; Leopold, A.C. Maturation Proteins and Sugars in Desiccation Tolerance of Developing Soybean Seeds 1. *Plant Physiology* **1992**, *100*, 225-230, doi:10.1104/pp.100.1.225.
2. Hitz, W.D.; Carlson, T.J.; Kerr, P.S.; Sebastian, S.A. Biochemical and Molecular Characterization of a Mutation That Confers a Decreased Raffinosaccharide and Phytic Acid Phenotype on Soybean Seeds. *Plant Physiology* **2002**, *128*, 650-660, doi:10.1104/pp.010585.
3. Koch, K. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Current Opinion in Plant Biology* **2004**, *7*, 235-246, doi:https://doi.org/10.1016/j.pbi.2004.03.014.
4. Redekar, N.R.; Glover, N.M.; Biyashev, R.M.; Ha, B.-K.; Raboy, V.; Maroof, M.A.S. Genetic interactions regulating seed phytate and oligosaccharides in soybean (Glycine max L.). *PLOS ONE* **2020**, *15*, e0235120, doi:10.1371/journal.pone.0235120.
5. Skoneczka, J.A.; Maroof, M.A.S.; Shang, C.; Buss, G.R. Identification of Candidate Gene Mutation Associated With Low Stachyose Phenotype in Soybean Line PI200508. *Crop Science* **2009**, *49*, 247-255, doi:10.2135/cropsci2008.07.0403.
6. Horbowicz, M.; Obendorf, R.L. Seed desiccation tolerance and storability: dependence on flatulence-producing oligosaccharides and cyclitols—review and survey. *Seed Science Research* **1994**, *4*, 385-405.
7. Sprenger, R.; Schlagenhaufer, R.; Kerb, R.; Bruhn, C.; Brockmöller, J.; Roots, I.; Brinkmann, U. Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype–phenotype correlation. *Pharmacogenetics and Genomics* **2000**, *10*, 557-565.

8. Pennycooke, J.C.; Jones, M.L.; Stushnoff, C. Down-regulating α-galactosidase enhances freezing tolerance in transgenic petunia. *Plant Physiology* **2003**, *133*, 901-909.
9. ElSayed, A.I.; Rafudeen, M.S.; Golldack, D. Physiological aspects of raffinose family oligosaccharides in plants: protection against abiotic stress. *Plant Biology* **2014**, *16*, 1-8.
10. Keller, F.; Pharr, D.M. Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides. In *Photoassimilate distribution in plants and crops*; Routledge: 1996; pp. 157-183.
11. Xue, H.; Chen, X.; Li, G. Involvement of phospholipid signaling in plant growth and hormone effects. *Current opinion in plant biology* **2007**, *10*, 483-489.
12. Okada, M.; Ye, K. Nuclear phosphoinositide signaling regulates messenger RNA export. *RNA biology* **2009**, *6*, 12-16.
13. Thole, J.M.; Nielsen, E. Phosphoinositides in plants: novel functions in membrane trafficking. *Current opinion in plant biology* **2008**, *11*, 620-631.
14. González, E.M.; Gordon, A.J.; James, C.L.; Arrese-lgor, C. The role of sucrose synthase in the response of soybean nodules to drought. *Journal of Experimental Botany* **1995**, *46*, 1515-1523, doi:10.1093/jxb/46.10.1515.
15. Bellaloui, N.; Smith, J.R.; Gillen, A.M.; Ray, J.D. Effect of Maturity on Seed Sugars as Measured on Near-Isogenic Soybean (Glycine max) Lines. *Crop Science* **2010**, *50*, 1978-1987, doi:10.2135/cropsci2009.10.0596.
16. Bellaloui, N.; Mengistu, A.; Fisher, D.K.; Abel, C.A. Soybean Seed Composition Constituents as Affected by Drought and Phomopsis in Phomopsis Susceptible and Resistant Genotypes. *Journal of Crop Improvement* **2012**, *26*, 428-453, doi:10.1080/15427528.2011.651774.
17. Wijewardana, C.; Reddy, K.R.; Bellaloui, N. Soybean seed physiology, quality, and chemical composition under soil moisture stress. *Food Chemistry* **2019**, *278*, 92-100, doi:https://doi.org/10.1016/j.foodchem.2018.11.035.
18. Kassem, M.A. *Soybean Seed Composition: Protein, Oil, Fatty Acids, Amino Acids, Sugars, Mineral Nutrients, Tocopherols, and Isoflavones*; Springer Nature: 2021.
19. Taiz, L. *Mineral nutrition*, Plant Physiology ed.; Sinauer Associates Inc.: P.O. Box 407, Sunderland, MA 01375, USA., 1998.
20. Brown, A.V.; Conners, S.I.; Huang, W.; Wilkey, A.P.; Grant, D.; Weeks, N.T.; Cannon, S.B.; Graham, M.A.; Nelson, R.T. A new decade and new data at SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Research* **2021**, *49*, D1496-D1501, doi:10.1093/nar/gkaa1107.
21. Salari, M.W.; Ongom, P.O.; Thapa, R.; Nguyen, H.T.; Vuong, T.D.; Rainey, K.M. Mapping QTL controlling soybean seed sucrose and oligosaccharides in a single family of soybean nested association mapping (SoyNAM) population. *Plant Breeding* **2021**, *140*, 110-122, doi:10.1111/pbr.12883.
22. Wang, T.L.; Domoney, C.; Hedley, C.L.; Casey, R.; Grusak, M.A. Can we improve the nutritional quality of legume seeds? *Plant Physiology* **2003**, *131*, 886-891.
23. Arendt, E.K.; Zannini, E. *Cereal grains for the food and beverage industries*; Elsevier: 2013.
24. Avilés-Gaxiola, S.; Chuck-Hernández, C.; Serna Saldívar, S.O. Inactivation methods of trypsin inhibitor in legumes: a review. *Journal of Food Science* **2018**, *83*, 17-29.
25. Kerr, P.S.; Pearlstein, R.W.; Schweiger, B.J.; Becker-Manley, M.F.; Pierce, J.W. Nucleotide sequences of galactinol synthase from zucchini and soybean. **1998**.
26. Frías, J.; Bakhsh, A.; Jones, D.; Arthur, A.; Vidal-Valverde, C.; Rhodes, M.; Hedley, C.L. Genetic analysis of the raffinose oligosaccharide pathway in lentil seeds. *Journal of Experimental Botany* **1999**, *50*, 469-476.
27. Hedley, C.L. *Carbohydrates in grain legume seeds: improving nutritional quality and agronomic characteristics*; CABI: 2001.
28. Knizia, D.; Yuan, J.; Bellaloui, N.; Vuong, T.; Usovsky, M.; Song, Q.; Betts, F.; Register, T.; Williams, E.; Lakhssassi, N. The Soybean High Density 'Forrest'by 'Williams 82'SNP-Based Genetic Linkage Map Identifies QTL and Candidate Genes for Seed Isoflavone Content. *Plants* **2021**, *10*, 2029.
29. Knizia, D.; Yuan, J.; Lakhssassi, N.; El Baze, A.; Cullen, M.; Vuong, T.; Mazouz, H.; T. Nguyen, H.; Kassem, M.A.; Meksem, K. QTL and Candidate Genes for Seed Tocopherol Content in 'Forrest'by 'Williams 82'Recombinant Inbred Line (RIL) Population of Soybean. *Plants* **2022**, *11*, 1258.
30. Akond, M.; Liu, S.; Kantartzi, S.K.; Meksem, K.; Bellaloui, N.; Lightfoot, D.A.; Kassem, M.A. Quantitative trait loci underlying seed sugars content in "MD96-5722" by "Spencer" recombinant inbred line population of soybean. *Food and Nutrition Sciences* **2015**, *6*, 964.
31. Allen, G.C.; Flores-Vergara, M.; Krasynanski, S.; Kumar, S.; Thompson, W. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature protocols* **2006**, *1*, 2320-2325.
32. Wu, X.; Vuong, T.D.; Leroy, J.A.; Grover Shannon, J.; Sleper, D.A.; Nguyen, H.T. Selection of a core set of RILs from Forrest× Williams 82 to develop a framework map in soybean. *Theoretical and applied genetics* **2011**, *122*, 1179-1187.
33. Wang, S.; Basten, C.; Zeng, Z. Windows QTL Cartographer 2.5 _011. *North Carolina State University, Raleigh* **2012**.

34. Voorrips, R. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of heredity* **2002**, *93*, 77-78.
35. Patil, G.B.; Lakhssassi, N.; Wan, J.; Song, L.; Zhou, Z.; Klepadlo, M.; Vuong, T.D.; Stec, A.O.; Kahil, S.S.; Colantonio, V.; et al. Whole-genome re-sequencing reveals the impact of the interaction of copy number variants of the rhg1 and Rhg4 genes on broad-based resistance to soybean cyst nematode. *Plant Biotechnology Journal* **2019**, *17*, 1595-1611, doi:10.1111/pbi.13086.
36. Iftime, D.; Hannah, M.A.; Peterbauer, T.; Heyer, A.G. Stachyose in the cytosol does not influence freezing tolerance of transgenic Arabidopsis expressing stachyose synthase from adzuki bean. *Plant Sci* **2011**, *180*, 24-30, doi:10.1016/j.plantsci.2010.07.012.
37. González-Morales, S.I.; Chávez-Montes, R.A.; Hayano-Kanashiro, C.; Alejo-Jacuinde, G.; Rico-Cambron, T.Y.; de Folter, S.; Herrera-Estrella, L. Regulatory network analysis reveals novel regulators of seed desiccation tolerance in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **2016**, *113*, E5232-E5241, doi:doi:10.1073/pnas.1610985113.
38. Blöchl, A.; March, G.G.-d.; Sourdioux, M.; Peterbauer, T.; Richter, A. Induction of raffinose oligosaccharide biosynthesis by abscisic acid in somatic embryos of alfalfa (Medicago sativa L.). *Plant Science* **2005**, *168*, 1075-1082, doi:https://doi.org/10.1016/j.plantsci.2004.12.004.
39. Maughan, P.; Maroof, M.; Buss, G. Identification of quantitative trait loci controlling sucrose content in soybean (Glycine max). *Molecular Breeding* **2000**, *6*, 105-111.
40. Kim, H.K.; Kang, S.T.; Oh, K.W. Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of Glycine max seeds. *Journal of Plant Research* **2006**, *119*, 533-538.
41. Kim, H.-K.; Kang, S.-T.; Cho, J.-H.; Choung, M.-G.; Suh, D.-Y. Quantitative trait loci associated with oligosaccharide and sucrose contents in soybean (Glycine max L.). *Journal of Plant Biology* **2005**, *48*, 106-112.
42. Mainali, H.R.; Vadivel, A.K.A.; Li, X.; Gijzen, M.; Dhaubhadel, S. Soybean cyclophilin GmCYP1 interacts with an isoflavonoid regulator GmMYB176. *Scientific reports* **2017**, *7*, 1-12.
43. Liu, C.; Chen, H.; Yu, Q.; Gu, H.; Li, Y.; Tu, B.; Zhang, H.; Zhang, Q.; Liu, X. Identification of quantitative trait loci (QTLs) and candidate genes for seed sucrose and soluble sugar concentrations in soybean. *Crop Science* **2023**.
44. Zeng, A.; Chen, P.; Shi, A.; Wang, D.; Zhang, B.; Orazaly, M.; Florez-Palacios, L.; Brye, K.; Song, Q.; Cregan, P. Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Science* **2014**, *54*, 554-564.
45. Cai, J.-S.; Feng, J.-Y.; Ni, Z.-J.; Ma, R.-H.; Thakur, K.; Wang, S.; Hu, F.; Zhang, J.-G.; Wei, Z.-J. An update on the nutritional, functional, sensory characteristics of soy products, and applications of new processing strategies. *Trends in Food Science & Technology* **2021**, *112*, 676-689.
46. Sui, M.; Wang, Y.; Bao, Y.; Wang, X.; Li, R.; Lv, Y.; Yan, M.; Quan, C.; Li, C.; Teng, W. Genome-wide association analysis of sucrose concentration in soybean (Glycine max L.) seed based on high-throughput sequencing. *The Plant Genome* **2020**, *13*, e20059.
47. Lee, J.S.; Kim, S.-M.; Kang, S. Fine mapping of quantitative trait loci for sucrose and oligosaccharide contents in soybean [Glycine max (L.) Merr.] using 180 K Axiom® SoyaSNP genotyping platform. *Euphytica* **2016**, *208*, 195-203.
48. Stombaugh, S.; Orf, J.H.; Jung, H.; Chase, K.; Lark, K.; Somers, D. Quantitative trait loci associated with cell wall polysaccharides in soybean seed. *Crop science* **2004**, *44*, 2101-2106.
49. Feng, C.; Morsy, M.; Giannoccaro, E.; Zhang, B.; Chen, P. Soybean seed sugar content and quantitative trait loci mapping. In Proceedings of the Plant nutrition for food security, human health and environmental protection; Fifteenth International Plant Nutrition Colloquium, 2005.
50. Jaureguy, L.M. *Identification of molecular markers associated with seed size, protein and sugar content in soybean*; University of Arkansas: 2009.
51. Wang, X.; Jiang, G.-L.; Green, M.; Scott, R.A.; Song, Q.; Hyten, D.L.; Cregan, P.B. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. *Molecular genetics and genomics* **2014**, *289*, 935-949.
52. Dhungana, S.K.; Kulkarni, K.P.; Park, C.W.; Jo, H.; Song, J.T.; Shin, D.H.; Lee, J.D. Mapping quantitative trait loci controlling soybean seed starch content in an interspecific cross of 'Williams 82'(Glycine max) and 'PI 366121'(Glycine soja). *Plant Breeding* **2017**, *136*, 379-385.
53. Patil, G.; Vuong, T.D.; Kale, S.; Valliyodan, B.; Deshmukh, R.; Zhu, C.; Wu, X.; Bai, Y.; Yungbluth, D.; Lu, F. 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* **2018**, *16*, 1939-1953.

54. Hu, L.; Wang, X.; Zhang, J.; Florez-Palacios, L.; Song, Q.; Jiang, G.-L. Genome-Wide Detection of Quantitative Trait Loci and Prediction of Candidate Genes for Seed Sugar Composition in Early Mature Soybean. *International Journal of Molecular Sciences* **2023**, *24*, 3167.