QTLs for soybean seed isoflavones are linked to laccases, BANYULS and the MBW complex

Xueying Li[1], Jun Zhang[1], Liangyu Chen[1], Zhenhui Wang[1], Cheng Yu[1], Xiao Han[1], Yang Song[1], Dan Yao[1], Sujie Fan[1], Jeremy Murray[2], Songnan Yang[1], Jian Ma[1], and Jian Zhang[1]

[1]Jilin Agricultural University Library
[2]Chinese Academy of Sciences Shanghai Branch

May 9, 2022

Abstract

Isoflavones are secondary metabolites present in seeds of soybean [*Glycine max* (L.) Merr.] which have been recognized their benefit to human health. In this study, QTL mapping for soybean isoflavone gylcones including daidzin, glycitin and genistin and total isoflavones content was performed in population of 178 F2:6 recombinant inbred lines (RILs) which was generated from cross between varieties Jinong17 and Jinong18. A genetic linkage map covering 1248 cM was constructed using the simple sequence repeat (SSR) molecular markers. The results revealed 22 isoflavone- related QTLs, 5 for daidzin, 7 for genistin, 6 for glycitin, and 4 for total isoflavone content. Seven of these represent new QTLs. Twenty candidate genes were identified, including eight laccases with presumed role in lignin biosynthesis, and the transcriptional regulator BANYULS and all three components of the MYB-bHLH-WD40 (MBW) complex that regulate its expression. These findings suggest that alterations in lignin and proanthocyanidin metabolism influence isoflavone accumulation in seeds. These leads might be helpful in the efforts to breed new soybean varieties with improved isoflavone composition and content.

Article

QTLs for soybean seed isoflavones are linked to laccases, BANYULS and the MBW complex

Xueying Li [1, #], Jun Zhang[1, 2, #], Liangyu Chen [1], Zhenhui Wang[1]. Cheng Yu[1], Xiao Han [1], Yang Song[1], Dan Yao [3], Sujie Fan[1], Jeremy Murray[4], Songnan Yang[1, *], Jian Ma[1, *] and Jian Zhang[1,5*]

[1] College of Agronomy, Jilin Agricultural University, 130118 Changchun, Jilin, China; xueyingl@jlau.edu.cn (X.L.); zhangjun@jlau.edu.cn (Jun.Z.); 20200045@mails.jlau.edu.cn (L.C.); wzhjlau@163.com(Z.H.W.); yucheng199811@163.com (C.Y.); hanxiaoyy@jlau.edu.cn (X.H.); songyangjlnd@163.com (Y.S.); fansujie@jlau.edu.cn (S.F.)

[2] National Crop Variety Approval and Characteristic Identification Station, Jilin Agricultural University, 130118 Changchun, Jilin, China

[3] College of Life Science, Jilin Agricultural University, 130118 Changchun, Jilin, China; dyao@jlau.edu.cn (D.Y.)

[4] CAS Center for Excellence in Molecular Plant Science(CEMPS), Chinese Academy of Science, Shanghai, China; jmurray@sippe.ac.cn (JM)

[5] Department Biology, University of British Columbia, Okanagan, Kelowna, BC, Canada

* Correspondence: soy@jlau.edu.cn (S.Y.); majian197916@jlau.edu.cn (J.M.); zhangjian@jlau.edu.cn (Jian.Z.)

# Xueying Li and Jun Zhang contributed equally to this work

Abstract: Isoflavones are secondary metabolites present in seeds of soybean [*Glycine max* (L.) Merr.] which have been recognized their benefit to human health. In this study, QTL mapping for soybean isoflavone gylcones including daidzin, glycitin and genistin and total isoflavones content was performed in population of 178 F2:6 recombinant inbred lines (RILs) which was generated from cross between varieties Jinong17 and Jinong18. A genetic linkage map covering 1248 cM was constructed using the simple sequence repeat (SSR) molecular markers. The results revealed 22 isoflavone- related QTLs, 5 for daidzin, 7 for genistin, 6 for glycitin, and 4 for total isoflavone content. Seven of these represent new QTLs. Twenty candidate genes were identified, including eight laccases with presumed role in lignin biosynthesis, and the transcriptional regulator BANYULS and all three components of the MYB-bHLH-WD40 (MBW) complex that regulate its expression. These findings suggest that alterations in lignin and proanthocyanidin metabolism influence isoflavone accumulation in seeds. These leads might be helpful in the efforts to breed new soybean varieties with improved isoflavone composition and content.

Keywords: soybean isoflavones; genetic mapping; quantitative trait loci; simple sequence repeat.

1. Introduction

Soybean is an important cash crop in many parts of the world. Soybean seeds are a good source of isoflavones, metabolites that are widely considered to be beneficial to human health as potential anticancer agents, and as therapy to reduce menopausal symptoms [1,2]. Soybean isoflavones are found in different parts of the plant, particularly in the hypocotyl, cotyledon, and seed coat [3]. The chemical structure of isoflavones is similar to that of female estrogens, and consequently they are often called phytoestrogens. Daidzin, glycitin and genistin are the main isoflavone compounds in soybean seeds[4]. Daidzin is a specific inhibitor of aldehyde dehydrogenase which may help to suppress ethanol consumption[5], and has also been shown to help prevent bone loss[6], provide neuroprotection and neuronutrition [7], and to have anti-oxidant and anti-inflammatory activities [8,9]. Glycitin may be useful for suppressing cartilage destruction in osteoarthritis [10] and protecting lung tissues from lipopolysaccharide-induced inflammation [11]. Genistin was also shown to be a potent anti-adipogenic and anti-lipogenic agent [12]. From the plant perspective, isoflavones function in plant disease resistance, plant-environment interactions [13,14], and act as important signals in interactions with beneficial N-fixing bacteria [15]. Soybean is also an important anti-bacterial plant protection element [16,17]. It plays an important role in the defensive response that protects plants from UV-induced damage [18-20]. In addition, soybean isoflavones are signaling substances between plants and microorganisms [21].It was found that the levels of genistein, daidzein and daidzein increased significantly when *Sclerotinia sclerotiorum*infected soybean [22].Other studies showed that the content of isoflavone metabolites in soybean leaf tissue changed significantly before and after aphid attack, indicating that the substances had a relationship with soybean aphid resistance [23,24].These research results indicated that increasing of soybean isoflavone content would be accompanied by the introduction of other good traits.

Since soybean products are rich source of isoflavones, increasing the total isoflavone content in seeds has become an important objective in soybean breeding.

The content and composition of soybean isoflavones are quantitative traits controlled by multiple genes and influenced by the environment [25]. Isoflavonoids are produced from the p-coumaryl CoA by the action of Chalcone Synthase (CHS) and/or Chalcone Reductase (CHR), followed by the sequential action of Chalcone Isomerase (CHI) and Isoflavone Synthase (IFS) [26].The availability of a soybean genetic linkage map has greatly boosted soybean genetic research on isoflavonoids. Liang (2010) identified six QTLs for isoflavone content localized in linkage groups J, N, D2, and G by using an RIL population [27].WANG et al. [28]detected 34 QTLs for isoflavone content in another RIL population, of which 23 were newly discovered loci, with one marker, qGTD2_1, explaining 3.4-11.0% of the phenotypic variation. Zhang et al. [29] detected 21 QTLs for isoflavones content, distributed on 8 chromosomes, which explained 4.48-8.83% of the phenotypic variation; of them, only one QTLs showed negative effects on isoflavones content. To date, in addition to known genes in the isoflavone synthesis pathway, researchers have identified 61 daidzein-related QTLs, 68 genistein-related

QTLs, 71 glycitin -related QTLs and 62 total isoflavone content-related QTLs, which are recorded in the Soybase Genome Database (http://www.soybase.org). There have been some reports on transcription factors that regulate key enzymes of isoflavone synthesis. Chu et al. [30] reported one candidate gene, *GmMYB29* , that is significantly correlated with isoflavone content and can activate *IFS2* and *CHS8* promoters. Vadivel et al. [31] studied transcription factor *GmMYB176* , and showed it regulates isoflavones by activating the expression of *CHS8* . However, the available markers density and population sizes used in previous studies were not sufficient to indentify the underlying genes [32]. Identification of QTLs affecting the isoflavone content will provide a theoretical basis for soybean quality improvement through molecular marker-assisted breeding and can provide leads on the key components of the pathway, and how they are regulated. Newly identified QTLs from our study would advance the understanding of the epistatic interactions associated with isoflavones content. SSR markers have multiple alleles in a single locus and are highly polymorphic, which is very beneficial to genetic research. In the field of soybean research, polymorphism of SSR molecular markers has been fully confirmed [33]. RIL population consists of F2 individuals who continuously self-cross, sib mate, or randomly cross within the population until the genotypes of the individuals in the family are completely homozygous.

In this study, a genetic linkage map was constructed using 58 SSR molecular markers and population of 178 F2:6 RILs, and QTLs were mapped for daidzin, glycitin, genistin and total isoflavone content. We present new QTLs and identify candidate genes and discuss their potential for metabolic engineering of soybean seed isoflavones.

2. Results

2.1. Variation of seed isoflavone content

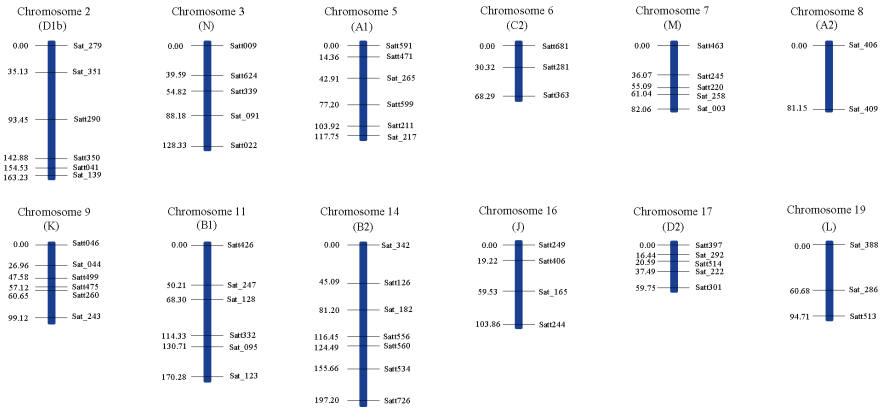
The mean isoflavone content in dry seeds of the 178 RILs and the parental lines is shown in Table 1. Jinong17 had significantly higher values for all isoflavone contents compared to Jinong18, indicated that the two parents differed in the genes controlling individual and total isoflavone contents. The average total isoflavone content was 1235.6 μg/g, or 1.2% of the dry weight. The three main isoflavones, daidzin, glycitin and genistin, accounted for more than 90% of the total isoflavone content. The average content of the three components in the population was daidzin > genistin > glycitin.

The population exhibited positively skewed distribution for individual and total isoflavones with values ranging from 0.068 to 2.18, indicating their seed isoflavone composition and content was more like Jinong17 which has a higher isoflavones content. The peak shape of daidzin and glycitin was platykurtic because their kurtosis values were negative, while genistein and the total content of isoflavones had a sharp distribution.

Table 1. Isoflavone content of the RIL population and their parents Jinong17 and Jinong18.

Trait	**Parents**	**Parents**	**RIL population**	**RIL population**	**RIL population**	**RIL po**
	Jinong17	**Jinong18**	**Mean**	**Range**	**Variance**	**Kurtosi**
Daidzin(μg/mg)	1.67	0.41	0.95	0.30-1.76	0.17	-1.26
Glycitin(μg/mg)	0.13	0.01	0.05	0.00-0.14	0.00	-1.00
Genistin(μg/mg)	2.21	0.18	0.24	0.10-2.32	0.11	8.03
Total Isoflavones(μg/mg)	4.01	0.60	1.24	0.49-4.09	0.36	7.45

2.2. SSR-based genetic map construction



A total of 278 polymorphic SSR markers were used to genotype the 178 RILs and 58 SSR markers showed polymorphism in the population. These SSR markers were assigned to 12 linkage groups (Figure 1) with a total length of 1365.73 cM and an average distance between adjacent markers of 23.41 cM. The number of markers on each linkage group ranged from 2 to 9. The order of markers in the constructed linkage map and the soybean common linkage map (GmComposite 2003, http://www.soybase.org) were consistent, except that satt591 and satt471 on linkage group A1 and satt560 and satt534 on the B2 linkage group were inversed (Figure 2).

Figure 1. The genetic map of soybean based on SSR markers. The genetic distances (cM) were shown on the left side and the markers on the right.

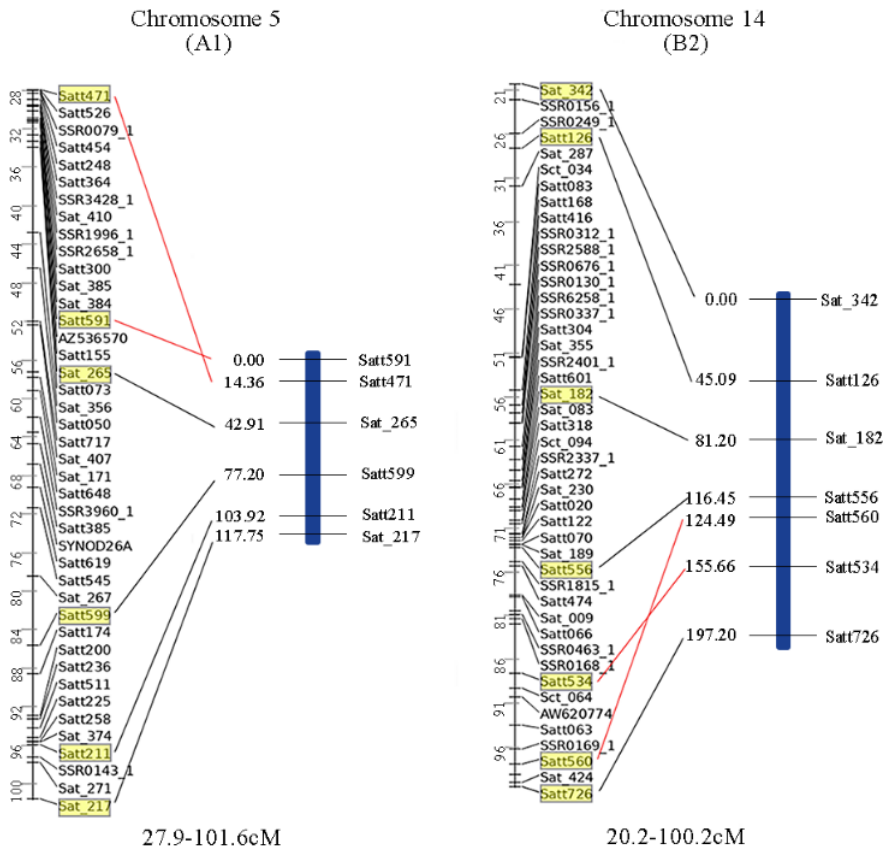


Figure 2. Comparison of the soybean A1 and B2 linkage groups (right) constructed by the Jinong17 × Jinong18 RIL population (right) with the corresponding linkage group of GmComposite 2003 genetic map (left). The genetic distances (cM) are shown on the left side and the markers on the right. The common markers between the two maps are highlighted in yellow boxes and connected by black lines, except for the ones of whose positions are rearranged which are connected by red lines.

2.3. QTL mapping for isoflavone content in soybean seeds

In total, we identified 22 main effect QTLs associated with the main soybean isoflavones including daidzin, glycitin and genistin and total isoflavone content, explained 0.35% to 2.06% of the phenotypic variation. The detailed information for all QTLs is shown in Table 2. All QTLs located on 11 linkage groups. The range of LOD scores and additive effects for the QTLs ranged from 2.98 to 37.30 and -0.34 to 1.27, respectively.

Table 2. The QTLs of soy isoflavones detected by ICIM method.

Traits	Linkage Group	Locus	LOD Value	Additive Effect	Phenotypic Variation Explained (%)
Daidzin	2 (D1b)	*qDaidzin-2-1*	3.76	0.25	2.06
	6(C2)	*qDaidzin-6-1*	4.67	0.21	1.91
	8(A2)	*qDaidzin-8-1*	3.19	-0.34	1.65
	16(J)	*qDaidzin-16-1*	3.71	0.18	1.62
	16(J)	*qDaidzin-16-2*	3.76	0.16	1.75

Traits	Linkage Group	Locus	LOD Value	Additive Effect	Phenotypic Variation Explained (%)
Glycitin	2(D1b)	*qGlycitin-2-1*	8.13	0.05	1.76
	3(N)	*qGlycitin-3-1*	6.92	0.18	1.86
	7(M)	*qGlycitin-7-1*	4.65	-0.0021	1.74
	11(B1)	*qGlycitin-11-1*	7.82	-0.01	1.76
	16(J)	*qGlycitin-16-1*	4.76	-0.04	1.71
	17(D2)	*qGlycitin-17-1*	2.98	-0.02	0.35
Genistin	8(A2)	*qGenistin-8-1*	23.88	0.18	0.79
	9(K)	*qGenistin-9-1*	27.23	-0.0033	0.79
	11(B1)	*qGenistin-11-1*	37.30	0.76	0.91
	14(B2)	*qGenistin-14-1*	24.43	0.0023	0.79
	14(B2)	*qGenistin-14-2*	31.97	0.04	0.80
	19(L)	*qGenistin-19-1*	28.70	0.02	0.79
	19(L)	*qGenistin-19-2*	29.84	0.78	0.79
Isoflavone	3(N)	*qIsoflavone-3-1*	11.41	-0.14	0.90
	9(K)	*qIsoflavone-9-1*	11.08	0.07	1.00
	11(B1)	*qIsoflavone-11-1*	11.17	1.20	1.65
	19(L)	*qIsoflavone-19-1*	10.51	1.27	0.96

For daidzin, 5 QTLs were distributed on 4 linkage groups, including chromosome 2 (D1b), chromosome 6 (C2), chromosome 8 (A2) and chromosome 16 (J). Six QTLs related to glycitin content were located on chromosome 2 (D1b), chromosome 3 (N), chromosome 7 (M), chromosome 11 (B1), chromosome 16 (J) and chromosome 17 (D2). Seven QTLs related to genistin content were detected on chromosome 8 (A2), chromosome 9 (K), chromosome 11 (B1), chromosome 14 (B2) and chromosome 19 (L). Four QTLs related to the total amount of isoflavones were found on chromosome 3 (N), chromosome 9 (K), chromosome 11(B1) and chromosome 19 (L). The location of all QTLs for soybean isoflavones detected is shown in Figure 3.

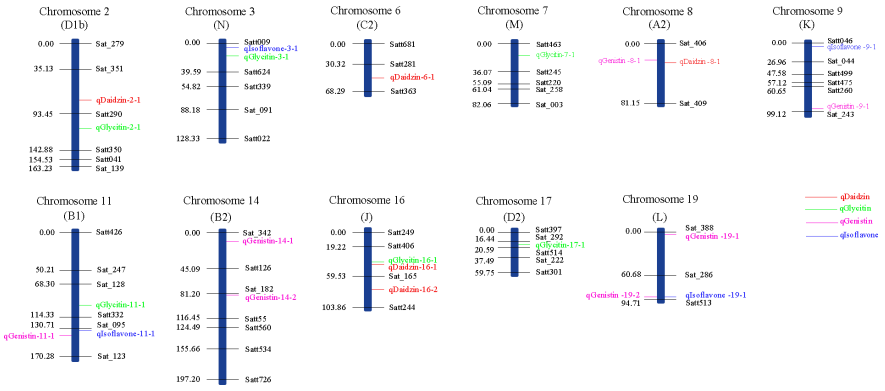


Figure 3. An SSR based linkage map of 22 QTLs for seed isoflavone content in soybean. Note: marker names are labeled at the right side of each linkage group, and distance between makers labeled at the left side, the isoflavones-related QTLs are indicated by colorful labels..

Five loci associated with daidzin content were found in 4 linkage groups. The QTL qDaidzin-8-1, was positioned between Sat_406 and Sat_409 on the chromosome 8 (A2) linkage group, with a LOD value of 3.19 and accounted for 1.65% of the phenotypic variation explained (PVE). The additive effect value for this QTL was -0.34, with the negative value indicating that the allele associated with increased daidzin content derived from the low-isoflavone male parent Jinong 18. Four other QTLs with positive additive effects were located on chromosome 2 (D1b), chromosome 6 (C2) and chromosome 16 (J) were identified for daidzin explaining 1.65-2.06% of PVE with LOD scores ranging from 3.71 to 4.67 (Table 2).

For glycitin content, six loci were dispersed on 6 linkage groups. QTLs qGlycitin-7-1, qGlycitin-11-1, qGlycitin-16-1 and qGlycitin-17-1 had PVEs ranging from 0.35% to 1.76%, and LOD values ranging from 2.98 to 7.82 and had additive effects with negative values, indicating that the increased glycitin content originated from Jinong 18. Loci qGlycitin-2-1 and qGlycitin-3-1 located on chromosome 2 (D1b) and chromosome 3 (N), with PVE of 1.76% and 1.86% and additive effects of 0.05 and 0.18 (Table 2).

Genistin content was associated with seven loci in 5 linkage groups. The QTL qGenistin-9-1, found on chromosome 9 (K) linkage group between Satt260 and Sat_243, with a genetic distance from Sat_243 of 5.12 cM, had a PVE of 0.79% and an additive effect of -0.0033. However, the remaining QTL associated with genistin, which were located in chromosomes 8 (A2), 11 (B1), 14 (B2) and 19 (L), with PVEs ranging from 0.79-0.91% and LOD scores ranging from 24.43- 37.30 (Table 2), all had positive additive effect values, indicating the associated increases in genistein were derived from the female Jinong 17.

Four loci associated total isoflavone content were located in 4 linkage groups. The QTL qIsoflavone-3-1 was detected on chromo-some 3 (N) between Satt009 and Satt624, 3.0 cM from Satt009, with a LOD values of 11.41, and it had additive effect of -0.14. However, the remining three QTLs had positive additive effects values. These QTL, qIsoflavone-9-1, qIsoflavone-11-1, qIsoflavone-19-1, were located on chromosomes 9 (K), 11 (B1) and 19 (L), and had PVEs from 0.96% to 1.65% with LOD scores ranging from 10.51-11.17 (Table 2).

2.4. Identification of candidate genes in flavone or isoflavone-related QTL regions

According to the annotation from SoyBase database (https://www.soybase.org), fifty-eight candidate genes were identified (Table S1). Among them, many genes were annotated as transcript factors and enzymes. Interestingly, twenty of the candidate genes are involved in phenylpropanoid metabolism, which is a metabolic pathway associated with flavonoid metabolism and isoflavonoid metabolism [34](Table 3). Of the twenty genes, seven genes relates to lactase Laccase 2,3,5,6,7 and Laccase 15/TT10 were identified here whne compared with *A. thaliana* , and several studies have shown that laccase can regulate the synthesis of phenylpropanoid and lignin [35-37], as phenylpropanoid, lignin also associated with synthesis of flavonoid and isoflavonid metabolism [38]. Meanwhile, several transcript factors of *A. thaliana* homolog, e.g. *bHLH42* , *MYB20* and *TTG1* , to directly regulate flavones or isoflavones synthesis in *A. thaliana* [39,44,52]. These results imply the accuracy of the presemted QTL mapping here. Finally, the RNA-Seq atlas of soybean different tissues the was obtained from Phytozome [40,41] to use select candidate genes. The expression dynamic variation of all these identified genes expressed in the seed were differed than other tissues (Figure 4). We found that nine of the fifty-six genes with expression data were highly expressed in seeds, *Glyma.02G076300* ,*Glyma.02G147800* , *Glyma.06G136900* , *Glyma.08G062000* ,*Glyma.08G062100* , *Glyma.09G020300* , *Glyma.16G158400* ,*Glyma.17G156000* and *Glyma.19G102000* , respectively. Some genes were also expressed in seeds, although their expression was not higher than in other tissues, such as *Glyma.02G125100* ,*Glyma.06G16500* and *Glyma.11G189100* . As we expected, some of phenylpronoid-related genes were also identified as seed-high or seed-available expression genes. The parental lines of the mapping population are not the same as the soybean used in the public RNA-seq altas, and the differences in expression of a gene among tissues are releavent candidate genes with expression in the seeds.

Table 3. Phenylpronoid-related candidate genes associated with QTL for soybean seed isoflavone content.

Locus	**Gene Name**	*A. thaliana* **homolog**	*A. thaliana* **gene names**

Locus	**Gene Name**	*A. thaliana* **homolog**	*A. thaliana* **gene names**
qDaidzin-2-1	Glyma.02G130400	AT5G13930	Chalcone Synthase/TT4
qDaidzin-2-1	Glyma.02G147800	AT4G09820	bHLH42/TT8
qDaidzin-6-1	Glyma.06G118500	AT5G13930	Chalcone Synthase/TT4
qDaidzin-6-1	Glyma.06G136900	AT5G24520	TTG1
qDaidzin-16-2	Glyma.16G158400	AT5G48100	Laccase 15/TT10
qDaidzin-8-1/qGenistin-8-1	Glyma.08G062000	AT1G61720	BANYULS
qDaidzin-8-1/qGenistin-8-1	Glyma.08G062100	AT1G61720	BANYULS
qGenistin-9-1	Glyma.09G205700	AT2G23910	CCR6
qGenistin-14-1	Glyma.14G056100	AT2G40370	Laccase 5
qGenistin-19-1	Glyma.19G105100	AT5G13930	Chalcone Synthase/TT4
qGenistin-19-1	Glyma.19G155300	AT4G09820	bHLH42/TT8
qGlycitin-2-1	Glyma.02G171700	AT3G28430	TT9
qGlycitin-11-1	Glyma.11G137500	AT2G29130	Laccase 2
qIsoflavone-11-1/qGenistin-11-1	Glyma.11G164000	AT2G30210	Laccase 3
qIsoflavone-11-1/qGenistin-11-1	Glyma.11G215800	AT1G66230	MYB20
qIsoflavone-3-1/qGlycitin-3-1	Glyma.03G077900	AT2G38080	Laccase
qIsoflavone-3-1/qGlycitin-3-1	Glyma.U027300	AT3G09220	Laccase 7
qIsoflavone-3-1/qGlycitin-3-1	Glyma.U027400	AT2G29130	Laccase 2
qIsoflavone-9-1	Glyma.09G123500	AT1G23230	Mediator 23

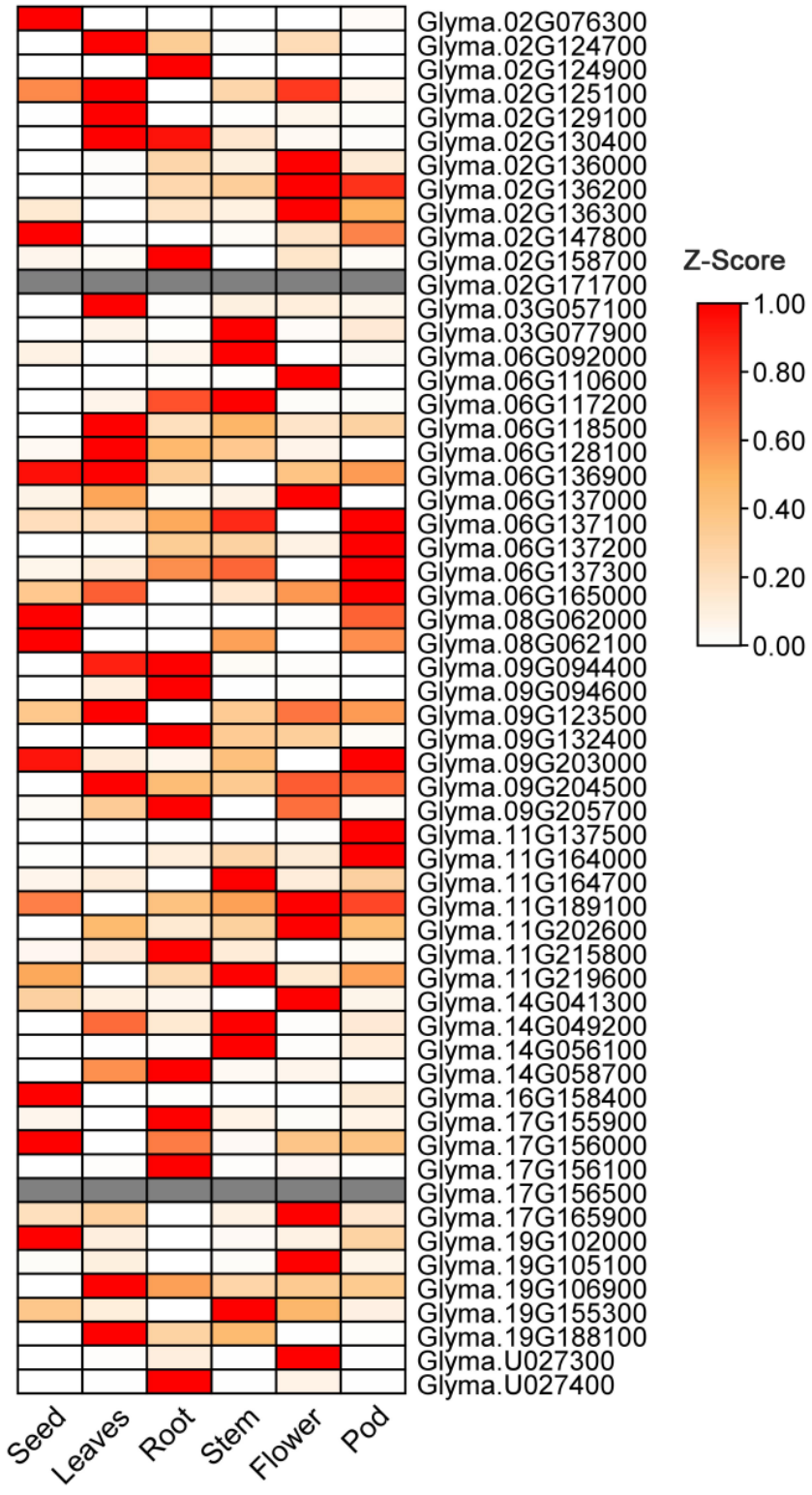


Figure 4. Identification of isoflavone-related candidate genes through transcriptome profile analysis. Normalized FPKM values are depicted on the Z-Score scale, gray indicates no expression data of the target gene

in the database.

3. Discussion

The isoflavone content of soybean seeds is a quantitative trait that is determined by multiple QTLs and environmental factors [42]. Ideally, breeders would like to introduce multiple genes into elite soybean cultivars to produce a superior variety with high yield and high isoflavone content in seeds. Achieving this goal through traditional methods has proven to be a challenge due to long breeding cycles. Identification of QTLs would provide the much-needed tools for efficient selection of high isoflavone content [43]. The RILs described here, as a permanent isolated population, provided good material for genetic map construction and QTL mapping. Using an RIL population of 178 F2:6 and 52 SSR molecular markers, we assembled a genetic linkage map covering 11 chromosomes. This yielded a total of 22 QTLs related to isoflavone content. The sequence of SSR markers on the linkage groups was mostly consistent with previously reported maps. Among the detected sites, 5, 7 and 6 were associated with daidzin, genistin and glycitin content, respectively. Four QTLs were found to be associated with total isoflavones, two which corresponded to genistin QTLs, and one to a glycitin QTL. Referring to the 2003 soybean public genetic map (GmComposite 2003), 15 QTLs were classified corresponded to those in the SoyBase database (https://www.soybase.org), and 7 new ones were identified including qDaidzin-2-1, qGlycitin-11-1, qGlycitin-16-1, Genistin-8-1, qGenistin-9-1, qGenistin-14-1 and qGenistin-19-2.

Of the candidate genes identified (Table S1), many were involved in phenypropanoid metabolism (Table 3). This included many genes involved in lignin biosynthesis, including eight laccases associated with five different QTL. Given their use of common precursors, it can be expected that downregulation of lignin biosynthesis could enhance the production of flavonoids. Indeed, previous research revealed a significant increase of flavonoids in Laccase 1 RNAi lines in cotton (Gossypium hirsutum) [44]. Similarly, decreasing lignin levels through downregulation of other lignin biosynthetic genes was shown to result in flavonoid accumulation in both Arabidopsis thaliana and Medicago sativa [45-47]. On the other hand, it was shown, using a population of inbred Brassica napus lines, that lines with high seed lignin showed increased expression of both lignin and flavonoid biosynthesis genes, suggesting that concordant regulation of these pathways can also occur [48]. Homologs of two arabidopsis phenypropanoid-related transcriptional regulators were also identified as candidates. One of these was a homolog of MYB20, a transcription factor that promotes expression of lignin biosynthesis genes in arabidopsis. In addition to activating lignin biosynthetic genes, MYB20 was shown to directly activate transcriptional repressors of flavonoid biosynthesis genes, which was reflected in increased flavonoid levels in the myb20 mutant [49]. Another candidate was homologous to MEDIATOR 23, a transcriptional regulator that was shown to regulate the production of phenylpropanoids [50]. Another candidate is homologous to Cinnamoyl CoA reductase 6 (CCR6), which potentially encodes an enzyme responsible for the final step in monolignol biosynthesis and that was previously linked to a QTL for lignin content in arabidopsis [51]. More research is needed to determine whether the QTLs described here can be explained through perturbation of lignin metabolism.

Several other candidates are predicted to be flavonoid-related, including a homolog of BANYULS, and its regulators. BANYULS a negative regulator of flavonoid synthesis in *A. thaliana* , that inhibits the production of proanthocyanidins in the seed coat [52] BANYULS expression is positively regulated by the so-called MBW transcriptional complex, which is comprised of the R2R3-MYB TRANSPARENT TESTA 1 (TT2) [53,54], the bHLH protein TT8 [55], and the WD40-repeat protein TRANSPARENT TESTA GLABROUS 1 (TTG1) [32,56]. Our study found that homologs of BANYULS and all three MBW complex members are associated with isoflavonoid QTLs. As discussed above for lignins, this could be explained by changes in proanthocyanidin synthesis affecting levels of common precursors required for production isoflavonoids. Two QTL were associated with CHS, the key branchpoint enzyme directing carbon flux from general phenylpropanoid metabolism to the flavonoid pathway.

In total, five of the 22 isoflavonoid QTL identified were found to be linked to homologs of 'transparent testa' genes. The TT genes were identified as having reduced seed coat pigments due to lower levels of proanthocyanidins [57] and hence are all expressed in the seed coat. For example, one candidate is homologous

to TT10/Laccase 15, a gene expressed in developing testa, where it colocalizes with the flavonoid end products proanthocyanidins and flavonols [35]. Our data suggests that competition with lignin and other flavonoids for precursors could be a major factors affecting isoflavonoid production in seeds. This work highlights two research avenues that deserve further investigation: the identification or creation of lines with a) reduced lignin production to divert precursors to the flavonoid branch or b) increased proanthocyanidin biosynthesis, which may be associated with a general increase in flavonoid content.

4. Materials and Methods

4.1. Plant materials and field experiments

An F2:6 recombinant inbred line (RIL) population consisting of 178 lines was derived from the cross of the parental genotypes Jinong 17 and Jinong 18 that differed significantly in the composition and total content of isoflavones in soybean seeds. Jinong 17 was used as the female parent and contains higher daidzin, glycitin, genistin and total isoflavones than Jinong 18, the male parent.

The field experiment was conducted in the experimental field of Jilin Agricultural University in Changchun, Jilin Province (43°13'N, 125deg19'E), China. The experiment was laid out in a randomized complete block design (RCBD) with three replications. Parents and RIL populations were planted with a row length of 4.5 meters, a distance between rows of 0.65 meters, 2 rows for each RIL material, and the planting density was 180,000-200,000 plants per hectare. Seeds were sampled after full maturity for further phenotypic characterization.

4.2. Soybean seed isoflavone extraction and determination

Soybean seed isoflavones were extracted and characterized as follows: Soybean seeds were ground into powder and were sifted through a 40-mesh screen. The soybean powder obtained was degreased with petroleum ether at 65 for 2 hours, and then was dried at 37 until reaching a constant weight. The skimmed powder (250 mg) was dissolved in 10 mL 80% methanol at room temperature for 2 hours and then distilled at 80degC for 12 hours. The supernatant was collected by centrifugation at 12000 rpm for 15 minutes. 1 mL of supernatant solution was filtered into a separate HPLC vial using a glass syringe equipped with a 0.45 μm nylon filter (Amicon, USA) for further isoflavones determination. The standards (genistein, genistin, daidzein, glycitein, glycitin and daidzin with purity more than 98%, Sigma-Aldrich, USA) were dissolved in methanol to a concentration of 100mg/L. Then the isoflavones standard solutions were serially diluted to 2 mg/L, 5 mg/L and 10mg/L for the calibration curves. An internal standard (2-methoxyflavone, Sigma-Aldrich, USA) solution (5 mg/mL) was also prepared in the same solvent.

A high-performance liquid chromatography (HPLC, Shimadzu, Japan) system was employed with a Shim-pack VP-DOS column (150mm×4.6mm, Shimadzu, Japan) and a mobile phase of methanol-water (a cubage ratio of 30:70). The column temperature was set at 40, the wave length was 254nm, the flow rate was 1mL/min, and the filling amount was 10μL. Analytes were quantified on the basis of the internal standard method. All samples were analyzed in triplicates.

4.3. Genotype analysis and construction of Genetic Maps

Fresh leaves of soybean plants at the three-leaf stage were used for total genome DNA extraction with the cetyltrimethylammonium bromide (CTAB) method [58]. DNA was dissolved in ddH2O to a concentration 100ng/ml with RNAase A (Thermo Fisher Scientific, USA)added. All primer sequences were obtained from Soybase database (*https://www.soybase.org*). A total of 274 primers from 14 linkage groups were used to detect polymorphism of SSR loci in the parents (in the parents), and 58 of them were available and synthesized.

PCR amplification was performed using the 2×Taq PCR StarMix kit (Genstar Kangrun Bio, Beijing, China). The PCR reactions used 1.0 μg template DNA, 1.0 μL of 10 μM forward and reverse primers, 10 μL of 2×Taq PCR Star Mix, and 7 μL of ddH2O. PCR was carried out as follows: 94 pre-denaturation for 5 min, 94 denaturation for 30 s, 53 renaturations for 30 s, 72 extension for 30 s, repeated 35 cycles, and 5min final

extension at 72 after the last cycle. The ICIMapping v4.0 software was used to identify the QTLs related to isoflavone contents and composition. A LOD value of 2.5 was used as a minimum to declare the detection of a QTL in a particular region.

5. Conclusions

The present study used an RIL population derived two parents with different isoflavones components, Jinong 17 and Jinong 18, to detect QTLs as well as mine possible candidate genes controlling soybean isoflavones content in multiple environments. A total of 22 QTLs were found, and 7 of these were novel. 61 candidate genes were identified based on the annotations. 10 genes were identified through integration of the QTL analysis, gene annotation and gene expression profile analysis. These three genes were not only in the isoflavones-related QTL region, but also highly expressed when isoflavones accumulated rapidly in soybean seeds. These results advance our understanding of the genetic basis of isoflavones synthesis and accumulation in soybeans.

Author Contributions: X.L., Jun.Z., L.C., C.Y., X.H., S.Y. were involved in the manuscript writing and methodology; Y.S., D.Y., S.F. J.M., Jian Z. were involved in manuscript refinement, data analysis, visualization and important intellectual content discussion; Jun Z., Jian Z. helpe to obtaine funding; J.M, Jian.Z helped frame the paper

Funding: This research was supported by Science and Technology Project of Jilin Provincial Department of Education (JJKH20210348KJ), Jilin Provincial Science and Technology Department of Science and Technology Development Project (20210202113NC), Science and Technology Project of Jilin Provincial Department of Education (JJKH20180658KJ) and Jilin Agricultural University high level researcher grant (JAUHLRG20102006).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tai, T.Y.; Tsai, K.S.; Tu, S.T.; Wu, J.S.; Chang, C.I.; Chen, C.L.; Shaw, N.S.; Peng, H.Y.; Wang, S.Y.; Wu, C.H. The effect of soy isoflavone on bone mineral density in postmenopausal Taiwanese women with bone loss: A 2-year randomized double-blind placebo-controlled study. *Osteoporosis Int.* **2012** , *23* , 1571-1580.

2. Gutierrez-Gonzalez, J.J.; Vuong, T.D.; Zhong, R.; Yu, O.; Lee, J.; Shannon, G.; Ellersieck, M.; Nguyen, H.T.; Sleper, D.A. Major locus and other novel additive and epistatic loci involved in modulation of isoflavone concentration in soybean seeds. *Theor. Appl. Genet.***2011** , *123* , 1375-1385.

3. Dhaubhadel, S.; Mcgarvey, B.D.; Williams, R.; Gijzen, M. Isoflavonoid biosynthesis and accumulation in developing soybean seeds. *Plant Mol. Biol.* **2003** , *53* , 733-743.

4. Li, X.; Kamala, S.; Tian, R.; Du, H.; Li, W.; Kong, Y.; Zhang, C. Identification and validation of quantitative trait loci controlling seed isoflavone content across multiple environments and backgrounds in soybean. *Mol. Breeding* **2018** , *38* , 1-14.

5. Lowe, E.D.; Gao, G.; Johnson, L.N.; Keung, W.M. Structure of daidzin, a naturally occurring anti-alcohol-addiction agent, in complex with human mitochondrial aldehyde dehydrogenase. *J. Med. Chem.***2008** , *51* , 4482-4487.

6. Yang, M.H.; Jung, S.H.; Chinnathambi, A.; Alahmadi, T.A.; Alharbi, S.A.; Sethi, G.; Ahn, K.S. Attenuation of STAT3 signaling cascade by daidzin can enhance the apoptotic potential of bortezomib against multiple myeloma. *Biomolecules* **2019** , *10* , 23.

7. Wei, G.; Liang, T.; Wei, C.; Nong, X.; Lu, Q.; Zhao, J. Daidzin inhibits RANKL-induced osteoclastogenesis in vitro and prevents LPS-induced bone loss in vivo. *J. Cell. Biochem.* **2019** ,*120* , 5304-5314.

8. Zhao, L.; Chen, Q.; Brinton, R.D. Neuroprotective and neurotrophic efficacy of phytoestrogens in cultured hippocampal neurons. *Exp. Biol. Med.* **2002** , *227* , 509-519.

9. Xiao, F.; Cui, H.; Zhong, X. Beneficial effect of daidzin in dry eye rat model through the suppression of inflammation and oxidative stress in the cornea. *Saudi J. Biol. Sci.* **2018** , *25* , 832-837.

10. Wang, W.; Yang, R.; Zhang, M.; Li, J.; Peng, J.; Xu, M.; Zhao, Y.; Li, H.; Pan, X. Glycitin suppresses cartilage destruction of osteoarthritis in mice. *Inflammation* **2020** , *43* , 1312-1322.

11. Chen, Y.; Guo, S.; Jiang, K.; Wang, Y.; Yang, M.; Guo, M. Glycitin alleviates lipopolysaccharide-induced acute lung injury via inhibiting NF-ϰB and MAPKs pathway activation in mice. *Int. Immunopharmacol.***2019** , *75* , 105749.

12. Choi, Y.R.; Shim, J.; Kim, M.J. Genistin: A novel potent anti-adipogenic and anti-lipogenic agent. *Molecules***2020** , *25* , 2042.

13. Graham, T.L. Flavonoid and isoflavonoid distribution in developing soybean seedling tissues and in seed and root exudates. *Plant Physiol.* **1991** , *95* , 594-603.

14. Lozovaya, V.V.; Lygin, A.V.; Zernova, O.V.; Li, S.; Hartman, G.L.; Widholm, J.M. Isoflavonoid accumulation in soybean hairy roots upon treatment with Fusarium solani. *Plant Physiol. Bioch.***2004** , *42* , 671-679.

15. Dong, W.; Song, Y. The significance of flavonoids in the process of biological nitrogen fixation. *Int. J. Mol. Sci.* **2020** ,*21* , 5926.

16. Dodds, P.N.; Rathjen, J.P. Plant immunity: Towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.***2010** , *11* , 539-548.

17. Algar, E.; Gutierrez-Mañero, F.J.; Garcia-Villaraco, A.; García-Seco, D.; Lucas, J.A.; Ramos-Solano, B. The role of isoflavone metabolism in plant protection depends on the rhizobacterial MAMP that triggers systemic resistance against Xanthomonas axonopodis pv. Glycines in Glycine max (L.) Merr. Cv. Osumi. *Plant Physiol. Bioch.***2014** , *82* , 9-16.

18. Kosslak, R.M.; Bohlool, B.B. Suppression of nodule development of one side of a split-root system of soybeans caused by prior inoculation of the other side. *Plant Physiol.* **1984** , *75* , 125-130.

19. Kim, H.; Jang, Y.; Baek, I.; Lee, J.; Park, M.J.; Chung, Y.; Chung, J.; Kim, J. Polymorphism and expression of isoflavone synthase genes from soybean cultivars. *Molecules & Cells (Springer Science & Business Media BV)* **2005** , *19* .

20. Mortier, V.; Fenta, B.A.; Martens, C.; Rombauts, S.; Holsters, M.; Kunert, K.; Goormachtig, S. Search for nodulation-related CLE genes in the genome of Glycine max. *J. Exp. Bot.* **2011** , *62* , 2571-2583.

21. Naoumkina, M.; Farag, M.A.; Sumner, L.W.; Tang, Y.; Liu, C.; Dixon, R.A. Different mechanisms for phytoalexin induction by pathogen and wound signals in Medicago truncatula. *Proceedings of the National Academy of Sciences* **2007** , *104* , 17909-17915.

22. Wegulo, S.N.; Yang, X.; Martinson, C.A.; Murphy, P.A. Effects of wounding and inoculation with Sclerotinia sclerotiorum on isoflavone concentrations in soybean. *Can. J. Plant Sci.* **2005** ,*85* , 749-760.

23. Yao, L.; Zhong, Y.; Wang, B.; Yan, J.; Wu, T. BABA application improves soybean resistance to aphid through activation of phenylpropanoid metabolism and callose deposition. *Pest Manag. Sci.* **2020** , *76* , 384-394.

24. Joshi, K.; Baumgardner, J.L.; Macphail, M.; Acharya, S.R.; Blotevogel, E.; Dayan, F.E.; Nachappa, P.; Nalam, V.J. The source of Rag5-mediated resistance to soybean aphids is located in the stem.*Front. Plant Sci.* **2021** , 1425.

25. Eldridge, A.C.; Kwolek, W.F. Soybean isoflavones: Effect of environment and variety on composition. *J. Agr. Food Chem.***1983** , *31* , 394-396.

26. Jung, W.; Yu, O.; Lau, S.C.; O'Keefe, D.P.; Odell, J.; Fader, G.; Mcgonigle, B. Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. *Nat. Biotechnol.* **2000** , *18* , 208-212.

27. Liang, H.; Yu, Y.; Wang, S.; Yun, L.; Wang, T.; Wei, Y.; Gong, P.; Liu, X.; Fang, X.; Zhang, M. QTL mapping of isoflavone, oil and protein contents in soybean (Glycine max L. Merr.). *Agricultural Sciences in China* **2010** , *9* , 1108-1116.

28. Wang, Y.; Han, Y.; Zhao, X.; Li, Y.; Teng, W.; Li, D.; Zhan, Y.; Li, W. Mapping isoflavone QTL with main, epistatic and QTL× environment effects in recombinant inbred lines of soybean. *PLoS One***2015** , *10* , e118447.

29. Zhang, H.J.; Li, J.W.; Liu, Y.J.; Jiang, W.Z.; Du, X.L.; Li, L.; Li, X.; Su, L.T.; Wang, Q.Y.; Wang, Y. Quantitative trait loci analysis of individual and total isoflavone contents in soybean seeds. *J. Genet.* **2014** , *93* , 331-338.

30. Chu, S.; Wang, J.; Zhu, Y.; Liu, S.; Zhou, X.; Zhang, H.; Wang, C.; Yang, W.; Tian, Z.; Cheng, H. An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. *PLoS Genet.* **2017** , *13* , e1006770.

31. Anguraj Vadivel, A.K.; Renaud, J.; Kagale, S.; Dhaubhadel, S. GmMYB176 regulates multiple steps in isoflavonoid biosynthesis in soybean. *Front. Plant Sci.* **2019** , *10* , 562.

32. Gutierrez-Gonzalez, J.J.; Wu, X.; Zhang, J.; Lee, J.; Ellersieck, M.; Shannon, J.G.; Yu, O.; Nguyen, H.T.; Sleper, D.A. Genetic control of soybean seed isoflavone content: Importance of statistical model and epistasis in complex traits. *Theor. Appl. Genet.* **2009** ,*119* , 1069-1083.

33. Akkaya, M.S.; Bhagwat, A.A.; Cregan, P.B. Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* **1992** ,*132* , 1131-1139.

34. Vogt, T. Phenylpropanoid biosynthesis. *Mol. Plant* **2010** ,*3* , 2-20.

35. Hu, Q.; Min, L.; Yang, X.; Jin, S.; Zhang, L.; Li, Y.; Ma, Y.; Qi, X.; Li, D.; Liu, H. Laccase GhLac1 modulates broad-spectrum biotic stress tolerance via manipulating phenylpropanoid pathway and jasmonic acid synthesis. *Plant Physiol.* **2018** , *176* , 1808-1823.

36. Xue, C.; Yao, J.; Xue, Y.; Su, G.; Wang, L.; Lin, L.; Allan, A.C.; Zhang, S.; Wu, J. PbrMYB169 positively regulates lignification of stone cells in pear fruit. *J. Exp. Bot.* **2019** , *70* , 1801-1814.

37. Wang, X.; Zhuo, C.; Xiao, X.; Wang, X.; Docampo-Palacios, M.; Chen, F.; Dixon, R.A. Substrate specificity of LACCASE8 facilitates polymerization of caffeyl alcohol for C-lignin biosynthesis in the seed coat of Cleome hassleriana. *The Plant Cell* **2020** , *32* , 3825-3845.

38. Dong, N.; Lin, H. Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *J. Integr. Plant Biol.* **2021** , *63* , 180-209.

39. Gonzalez, A.; Zhao, M.; Leavitt, J.M.; Lloyd, A.M. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *The Plant Journal* **2008** ,*53* , 814-827.

40. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; Xu, D.; Hellsten, U.; May, G.D.; Yu, Y.; Sakurai, T.; Umezawa, T.; Bhattacharyya, M.K.; Sandhu, D.; Valliyodan, B.; Lindquist, E.; Peto, M.; Grant, D.; Shu, S.; Goodstein, D.; Barry, K.; Futrell-Griggs, M.; Abernathy, B.; Du, J.; Tian, Z.; Zhu, L.; Gill, N.; Joshi, T.; Libault, M.; Sethuraman, A.; Zhang, X.; Shinozaki, K.; Nguyen, H.T.; Wing, R.A.; Cregan, P.; Specht, J.; Grimwood, J.; Rokhsar, D.; Stacey,

G.; Shoemaker, R.C.; Jackson, S.A. Genome sequence of the palaeopolyploid soybean. *Nature* **2010** , *463* , 178-183.

41. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.***2012** , *40* , D1178-D1186.

42. Han, Y.; Li, D.; Zhao, G.; Zhao, X.; Jiang, Z.; Hu, H.; Wu, L.; Wang, Y.; Gao, Y.; Li, Y. Dynamic quantitative trait loci underlies isoflavone accumulation in soybean seed. *Plant Breeding***2016** , *135* , 335-341.

43. Sh, U.; Fmd, S.; Mauro, A. *Marker-Assisted selection and genomic selection* ; Soybean Breeding, 2017.

44. Li, S. Transcriptional control of flavonoid biosynthesis: Fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signaling & Behavior* **2014** , *9* , e27522.

45. Besseau, S.; Hoffmann, L.; Geoffroy, P.; Lapierre, C.; Pollet, B.; Legrand, M. Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. *The Plant Cell* **2007** , *19* , 148-162.

46. Li, X.; Bonawitz, N.D.; Weng, J.; Chapple, C. The growth reduction associated with repressed lignin biosynthesis in Arabidopsis thaliana is independent of flavonoids. *The plant cell* **2010** ,*22* , 1620-1632.

47. Gill, U.S.; Uppalapati, S.R.; Gallego Giraldo, L.; Ishiga, Y.; Dixon, R.A.; Mysore, K.S. Metabolic flux towards the (iso) flavonoid pathway in lignin modified alfalfa lines induces resistance against Fusarium oxysporum f. Sp. Medicaginis. *Plant, Cell & Environment***2018** , *41* , 1997-2007.

48. Ding, Y.; Yu, S.; Wang, J.; Li, M.; Qu, C.; Li, J.; Liu, L. Comparative transcriptomic analysis of seed coats with high and low lignin contents reveals lignin and flavonoid biosynthesis in Brassica napus. *BMC Plant Biol.* **2021** , *21* , 1-16.

49. Geng, P.; Zhang, S.; Liu, J.; Zhao, C.; Wu, J.; Cao, Y.; Fu, C.; Han, X.; He, H.; Zhao, Q. MYB20, MYB42, MYB43, and MYB85 regulate phenylalanine and lignin biosynthesis during secondary cell wall formation. *Plant Physiol.* **2020** , *182* , 1272-1283.

50. Dolan, W.L.; Dilkes, B.P.; Stout, J.M.; Bonawitz, N.D.; Chapple, C. Mediator complex subunits MED2, MED5, MED16, and MED23 genetically interact in the regulation of phenylpropanoid biosynthesis. *The Plant Cell* **2017** , *29* , 3269-3285.

51. Capron, A.; Chang, X.F.; Hall, H.; Ellis, B.; Beatson, R.P.; Berleth, T. Identification of quantitative trait loci controlling fibre length and lignin content in Arabidopsis thaliana stems. *J. Exp. Bot.* **2013** , *64* , 185-197.

52. Baudry, A.; Heim, M.A.; Dubreucq, B.; Caboche, M.; Weisshaar, B.; Lepiniec, L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana.*The Plant Journal* **2004** , *39* , 366-380.

53. Nesi, N.; Jond, C.; Debeaujon, I.; Caboche, M.; Lepiniec, L. The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed.*The Plant Cell* **2001** , *13* , 2099-2114.

54. Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. MYB transcription factors in Arabidopsis. *Trends Plant Sci.* **2010** , *15* , 573-581.

55. Nesi, N.; Debeaujon, I.; Jond, C.; Pelletier, G.; Caboche, M.; Lepiniec, L. The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. *The Plant Cell* **2000** , *12* , 1863-1878.

56. Walker, A.R.; Davison, P.A.; Bolognesi-Winfield, A.C.; James, C.M.; Srinivasan, N.; Blundell, T.L.; Esch, J.J.; Marks, M.D.; Gray, J.C. The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *The Plant Cell* **1999** , *11* , 1337-1349.

57. Appelhagen, I.; Thiedig, K.; Nordholt, N.; Schmidt, N.; Huep, G.; Sagasser, M.; Weisshaar, B. Update on transparent testa mutants from Arabidopsis thaliana: Characterisation of new alleles from an isogenic collection. *Planta* **2014** , *240* , 955-970.

58. Doyle, J.J.; Doyle, J.L.; Brown, A. Analysis of a polyploid complex in Glycine with chloroplast and nuclear DNA. *Aust. Syst. Bot.***1990** , *3* , 125-136.