



OPEN ACCESS

EDITED BY

Silvas Prince,
Upstream Biotechnology, United States

REVIEWED BY

Jun You,
Chinese Academy of Agricultural
Sciences, China
Chandra Mohan Singh,
Banda University of Agriculture and
Technology, India

*CORRESPONDENCE

Kwang-Soo Cho,
✉ kscholove@korea.kr

[†]These authors have contributed equally
to this work

RECEIVED 06 September 2023

ACCEPTED 27 November 2023

PUBLISHED 11 December 2023

CITATION

Kim S, Lee E, Lee J, An YJ, Oh E, Kim JI,
Kim SW, Kim MY, Lee MH and Cho K-S
(2023), Identification of QTLs and allelic
effect controlling lignan content in
sesame (*Sesamum indicum* L.) using
QTL-seq approach.
Front. Genet. 14:1289793.
doi: 10.3389/fgene.2023.1289793

COPYRIGHT

© 2023 Kim, Lee, Lee, An, Oh, Kim, Kim,
Kim, Lee and Cho. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Identification of QTLs and allelic effect controlling lignan content in sesame (*Sesamum indicum* L.) using QTL-seq approach

Sungup Kim^{1†}, Eunsoo Lee^{1†}, Jeongeun Lee¹, Yeon Ju An¹,
Eunyoung Oh¹, Jung In Kim¹, Sang Woo Kim¹, Min Young Kim¹,
Myoung Hee Lee¹ and Kwang-Soo Cho^{2*}

¹Upland Crop Breeding Research Division, Department of Southern Area Crop Science, National Institute of Crop Science, Rural Development Administration, Miryang, Republic of Korea, ²Central Crop Breeding Research Division, Department of Central Area Crop Science, National Institute of Crop Science, Rural Development Administration, Suwon, Republic of Korea

Sesame (*Sesamum indicum* L.), an oilseed crop, is gaining worldwide recognition for its healthy functional ingredients as consumption increases. The content of lignans, known for their antioxidant and anti-inflammatory effects, is a key agronomic trait that determines the industrialization of sesame. However, the study of the genetics and physiology of lignans in sesame is challenging, as they are influenced by multiple genes and environmental factors, therefore, the understanding of gene function and synthetic pathways related to lignan in sesame is still limited. To address these knowledge gaps, we conducted genetic analyses using F7 recombinant inbred line (RIL) populations derived from Goenbaek and Gomazou as low and high lignin content variants, respectively. Using the QTL-seq approach, we identified three loci, *qLignan1-1*, *qLignan6-1*, and *qLignan11-1*, that control lignan content, specifically sesamin and sesamol. The allelic effect between loci was evaluated using the RIL population. *qLignan6-1* had an additive effect that increased lignan content when combined with the other two loci, suggesting that it could be an important factor in gene pyramiding for the development of high-lignan varieties. This study not only highlights the value of sesame lignan, but also provides valuable insights for the development of high-lignan varieties through the use of DNA markers in breeding strategies. Overall, this research contributes to our understanding of the importance of sesame oil and facilitates progress in sesame breeding for improved lignan content.

KEYWORDS

DNA marker, lignan, oilseed, sesame, QTL-seq

1 Introduction

Sesame (*Sesamum indicum* L.) is a major oil crop and a special crop in East Asia, including Korea, China, and Japan, with a focus on its functional ingredients (Langyan et al., 2022). With an oil content of approximately 50% within its seeds (Uzun et al., 2008; Wang et al., 2014; Wei et al., 2015), sesame is an optimal candidate for industrial oil production and use. Extensive research has highlighted the nutritional advantages and health effects of sesame oil in humans, which include anti-inflammatory and anti-

cancer activity, antioxidant properties, and nootropic effects (Yokota et al., 2007; Monteiro et al., 2014; Wan et al., 2015; Xu et al., 2015; Kim et al., 2023). Lignans, predominantly sesame oil, are categorized into fat-soluble and water-soluble types (Dar and Arumugam, 2013). Among these, sesamin and sesamol are the major components of the fat-soluble aglycons (Dar and Arumugam, 2013; Michailidis et al., 2019), demonstrating the substantial health-promoting potential of sesame both clinically and pharmacologically (Majdalawieh et al., 2017). Given their significance, in-depth exploration of functional elements such as oil and lignans may significantly elevate the agricultural value and industrial potential of sesame.

In recent years, major crop breeding strategies have changed significantly, moving from traditional breeding using physical markers to the concept of marker-assisted selection (MAS) employing contemporary molecular markers (Collard and Mackill, 2008; Nadeem et al., 2018). In addition, the availability of high quality reference genomes and pan-genomes has made a variety of genomic studies possible, such as genetic mapping, structural variant studies of whole genomes, and syntenic analyses that compare regions between and within species (Bayer et al., 2020; Khan et al., 2020; Sun et al., 2022). The progress related to DNA markers based on next-generation sequencing (NGS) technology has made it possible to use genome-based markers such as single nucleotide polymorphisms (SNPs) and insertion-deletions (InDels), which are higher in throughput than previous RFLPs, RAPDs, and SSRs. This has led to rapid and accurate breeding as well as detailed research at the gene level, giving rise to a paradigm of digital breeding (Collard and Mackill, 2008; Ray and Satya, 2014; Nadeem et al., 2018; Varshney et al., 2021).

QTL-seq is an analytical technique that detects significant genetic regions associated with a trait by pooling individuals with contrasting phenotypes and comparing their genotypes to identify quantitative trait loci (QTLs) (Takagi et al., 2013). This approach combines bulked-segregant analysis (BSA), which previously used a small number of markers, with whole genome sequencing (WGS) using NGS technology to develop and use a large number of markers at low cost, thus improving the efficiency and accuracy of genetic studies. Unlike traditional QTL analysis in a bi-parental population, where all individuals must be genotyped to detect QTLs, QTL-seq mixes individuals with the same phenotype into groups, assuming that all genetic regions are equally likely to segregate due to random crossover, and only identifies genetic regions that are extremely segregated between groups as statistically significant (Abe et al., 2012; Takagi et al., 2013; Sugihara et al., 2022). Several studies in rice, maize, and tomato have demonstrated that QTL-seq provides comparable reliable results, especially for genotyping, thus overcoming the high cost of conventional QTL and genome wide association study (GWAS) (Chen et al., 2021; Topcu et al., 2021; Yang et al., 2021). Thus, a strategy to find useful trait-associated loci from a large number of materials based on different resources from different environments and ecotypes using low-cost and time-saving methods such as QTL-seq to accelerate breeding is essential. Secondary metabolites, are particularly regarded as typical quantitative features and are known to be influenced by a variety of hereditary variables (Alseekh et al., 2015).

In this study, we aimed to find genetic factors controlling sesame lignan content by using QTL-seq on an F7 recombinant inbred lines (RILs) constructed from individuals with different lignan content. Additionally, we looked at the impact of QTL stacking by comparing lignan content according to the haplotypes in the identified QTLs, which will be used as a basis for breeding the selection of new sesame varieties with improved economic value and functional content.

2 Materials and methods

2.1 Plant materials and phenotypic evaluation

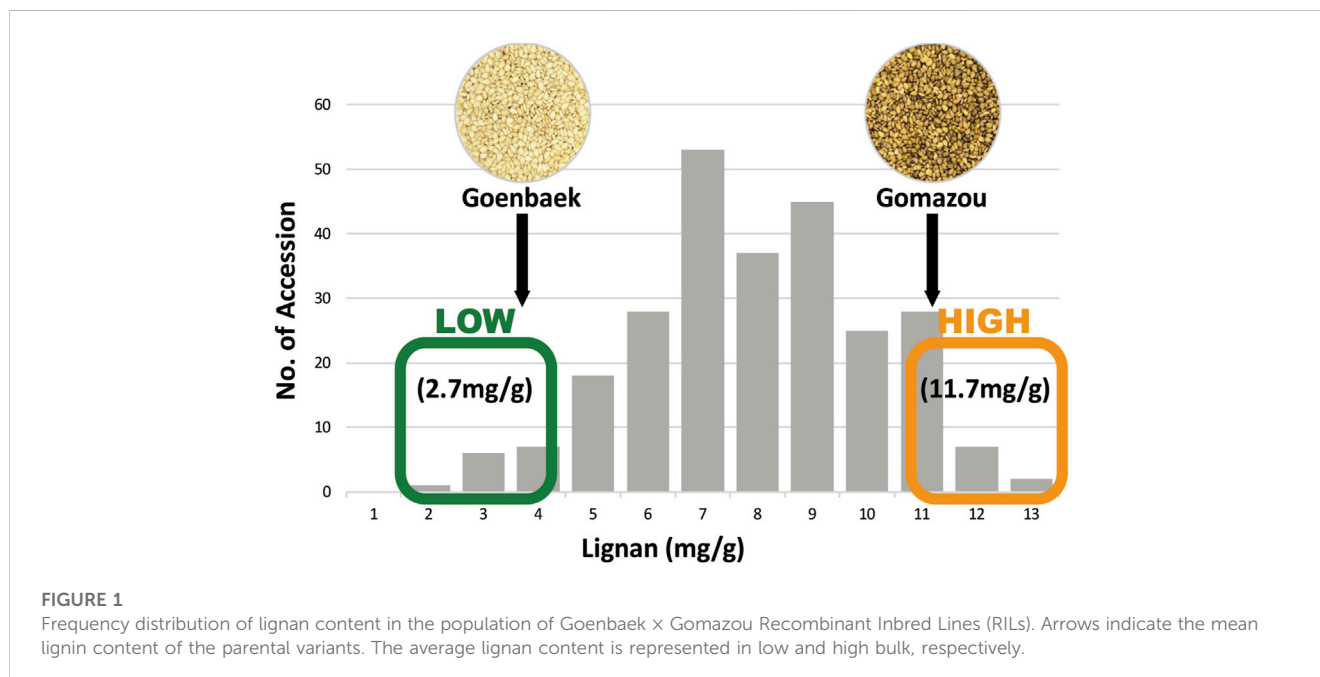
A total of 257 F7 populations were derived by crossing low-lignan Goenbaek with high-lignan Gomazou through generational advancement using the single-seed descent method. Goenbaek is an elite variety most widely grown in South Korea for blight resistance and agronomic stability (Kim et al., 2018), while Gomazou is a variety bred in Japan for functionality (Yasumoto and Katsuta, 2006). The parental lines and RILs were planted at the National Institute of Crop Science experimental greenhouse, Miryang, South Korea, in June 2018 at 55 cm × 15 cm spacing; ten individuals of each line were harvested in August, and these planting formats and periods are adapted to the weather conditions in Korea. After harvesting, they were pooled and examined for sesamin and sesamol content using HPLC to estimate the total lignan content. The sample was extracted in methanol and the analytical conditions were as follows: mobile phase A was 0.1% trifluoroacetic acid in water mobile phase B was 0.1% trifluoroacetic acid in methanol, and analysis under 60% (v/v) methanol conditions. The column was a YMC Triart C-18 (1.9 μm, 2.0 mm × 50 mm, 30°C; YMC Co., Kyoto, Japan) with a flow rate of 0.3 mL/min and was analyzed under UV at 290 nm using a Dionex 3000 HPLC (Thermo Fisher Scientific, Waltham, MA, United States of America). Sesamin and sesamol were measured using an analytical standard (Sigma-Aldrich, St. Louis, MO, United States of America).

2.2 Construction of bulks, DNA extraction, and sequencing

Based on the distribution of lignan content in the RIL population, ten individuals belonging to each of the two extremes were selected to form high and low bulk. Genomic DNA was extracted from fresh leaves collected at the early stage of growth using the GeneAll® Exgene™ Plant SV kit (GeneAll Biotechnology, Seoul, Korea). The DNA concentration of each individual was uniformly diluted to 10 ng/mol using the Nanodrop 3,000 spectrometer (Thermo Scientific, Wilmington, DE, United States of America) and pooled into high and low bulk by mixing an equal amount of DNAs, respectively. After constructing the pair-end sequencing libraries, Goenbaek (read length: 101 bp), Gomazou (read length: 151 bp), high bulk (read length: 151 bp), and low bulk

TABLE 1 Variation of sesamin, sesamol, and lignan content in the parental lines and Recombinant Inbred Lines (RILs) population.

Lignan content (mg/g)	Parents		RIL population			
	Goenbaek	Gomazou	Average	SD	Min	Max
Sesamin	2.7	8.2	5.2	1.7	1.0	9.4
Sesamol	1.0	3.1	2.2	0.6	0.7	3.7
Lignan	3.7	11.3	7.5	2.1	1.7	12.8



(read length: 151 bp) were sequenced using the Illumina HiSeq platform. Sequenced data were deposited into the National Center for Biotechnology Information (NCBI) under BioProject (<https://www.ncbi.nlm.nih.gov/bioproject>) as PRJNA1028140.

2.3 Genotyping and QTL-seq analysis

The raw reads were trimmed to remove residual adapter sequences using trimmomatic v0.39 (Bolger et al., 2014). The trimmed reads were aligned to the sesame reference genome (cv. Zhongzhi No.13. v 2.0) (Wang et al., 2016) using BWA v0.7.17 (Li and Durbin, 2009), and BAM files were generated with a Q-score of 30 or higher using SAMtools v.1.15.1 (Li et al., 2009). To select variants for analysis, only positions with a mapping quality ≥ 30 or more, a mapping depth ≥ 8 but < 250 , and where both parental lines were confirmed to be homozygous alleles were filtered from the bulk samples using VCFtools v0.1.16 (Danecek et al., 2011). For the QTL-seq study, the SNP-index was calculated according to well-established principles from previous studies (Abe et al., 2012; Takagi et al., 2013), and the analysis pipeline was applied using VCF files (Sugihara et al., 2022). To reduce the possibility of missing positions and to find major genetic regions, the variant index of both

Goenbaek and Gomazou was calculated and compared as consensus genomes. The average SNP-index was calculated through sliding window analysis with a 1 Mb window size and a step size of 50 Kb. Under the null hypothesis of no QTL, the statistical test of $\Delta(\text{SNP-index})$ was performed for all SNP positions based on read depth. The $\Delta(\text{SNP-index})$ was plotted along the physical chromosome position with 99% and 95% confidence intervals.

2.4 Variant mining, KEGG pathway mapping, and qRT-PCR analysis

The parent sequence variation and two bulks were analyzed using the SnpEff software to assign their changes and impacts on the function of the gene (Cingolani et al., 2012). For functional annotation of the genes, homologous genes were searched through comparative analysis using the BLASTp v2.10.0+ software with default parameters based on *Arabidopsis thaliana* gene-protein information. The functional descriptions of sesame genes were inferred by referring to *A. thaliana* homologs, and protein sequences and gene information were downloaded from TAIR version 10 (<https://www.arabidopsis.org>). The assignment of candidate genes to metabolic pathways was mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG, <https://www.genome>.

TABLE 2 Number of significant loci and SNP/InDel from QTL-seq analysis.

Consensus genome	Chromosome	Confidence interval (bp)	SNP/InDel index		Δ (SNP/InDel-index)
			Lignan-L	Lignan-H	
Goenbaek	chr1	3,950,000	0.9305	0.284	-0.6465
		5,650,000–10,100,000	0.9044	0.2155	-0.6889
		11,400,000	0.8871	0.2597	-0.6273
		11,600,000–11,750,000	0.9123	0.3052	-0.6071
		12,850,000–12,900,000	0.8936	0.264	-0.6285
		15,100,000–16,750,000	0.9774	0.2078	-0.7696
	chr2	14,850,000–15,250,000	0.8667	0.2131	-0.6536
	chr3	23,150,000–23,600,000	0.9971	0.3881	-0.609
	chr6	15,100,000–16,050,000	0.3565	0.9978	0.6414
	chr9	6,000,000–6,350,000	0.9549	0.3348	-0.6201
		6,550,000–6,700,000	0.9558	0.3378	-0.6181
	chr11	500,000–2,100,000	0.8743	0.0206	-0.8537
Gomazou	chr1	3,950,000	0.081	0.7174	0.6364
		5,650,000–10,100,000	0.0992	0.7779	0.6787
		11,400,000	0.1228	0.7369	0.6142
		11,600,000	0.088	0.6968	0.6087
		12,850,000–12,900,000	0.1111	0.7391	0.628
		15,100,000–16,750,000	0.0243	0.7923	0.768
	chr2	14,950,000–15,250,000	0.1358	0.7853	0.6495
	chr3	23,150,000–23,600,000	0.003	0.6196	0.6166
	chr6	14,600,000–16,500,000	0.8129	0.021	-0.792
	chr11	500,000–2,100,000	0.1277	0.99	0.8623
		2,450,000–3,150,000	0.8427	0.1523	-0.6904
	chr12	14,350,000–15,750,000	0.7222	0.0054	-0.7167
	chr13	2,550,000–2,600,000	0.2645	0.869	0.6045

jp/kegg/pathway.html) pathway. To compare gene expression between the Goenbaek and Gomazou varieties, total RNA was extracted from immature seeds at 15 and 30 days after flowering using the GeneAll Ribospin™ Plant kit (GeneAll, Korea) and converted into cDNA using the RNA to cDNA EcoDry™ Premix (Oligo dT) kits (TaKaRa, Japan). Primers for qRT-PCR were designed using Primer3 (<https://primer3plus.com>) as shown in Supplementary Table S1. qRT-PCR was performed using SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad Laboratories, CA, United States of America) on QuantStudio™ 5 (Applied Biosystems, United States of America). The amplification conditions were as follows: 50°C for 2 min and denaturation at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min, 72°C for 15 s. The sesame housekeeping gene 18s rRNA (qFw: 5'-CGTCCCTGCCCTTTGTACAC-3' and qRv: 5'-CGAACACTT CACCGGACCAT-3') was used as a reference gene (Noguchi et al., 2008). The relative quantification of samples was calculated

according to the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) in three biological replicates, each with two technical replicates. Student's t-test was used to determine statistical significance between groups.

2.5 Development of KASP, InDel, and CAPS markers and evaluation in RILs

Based on variant annotation results between the parents, InDel, Kompetitive Allele Specific PCR (KASP), and Cleaved Amplified Polymorphic Sequences (CAPS) markers were designed using bulk-biased SNPs and InDels located in the intron of *qLignan1-1*, and in the CDS of *qLignan6-1*, and in the intergenic region of *qLignan11-1*. Primers for InDel marker were designed using Primer3 (<https://primer3plus.com>), while primers for CAPS marker were designed using dCAPS Finder 2.0 (<http://helix.wustl.edu/dcaps>). PCR

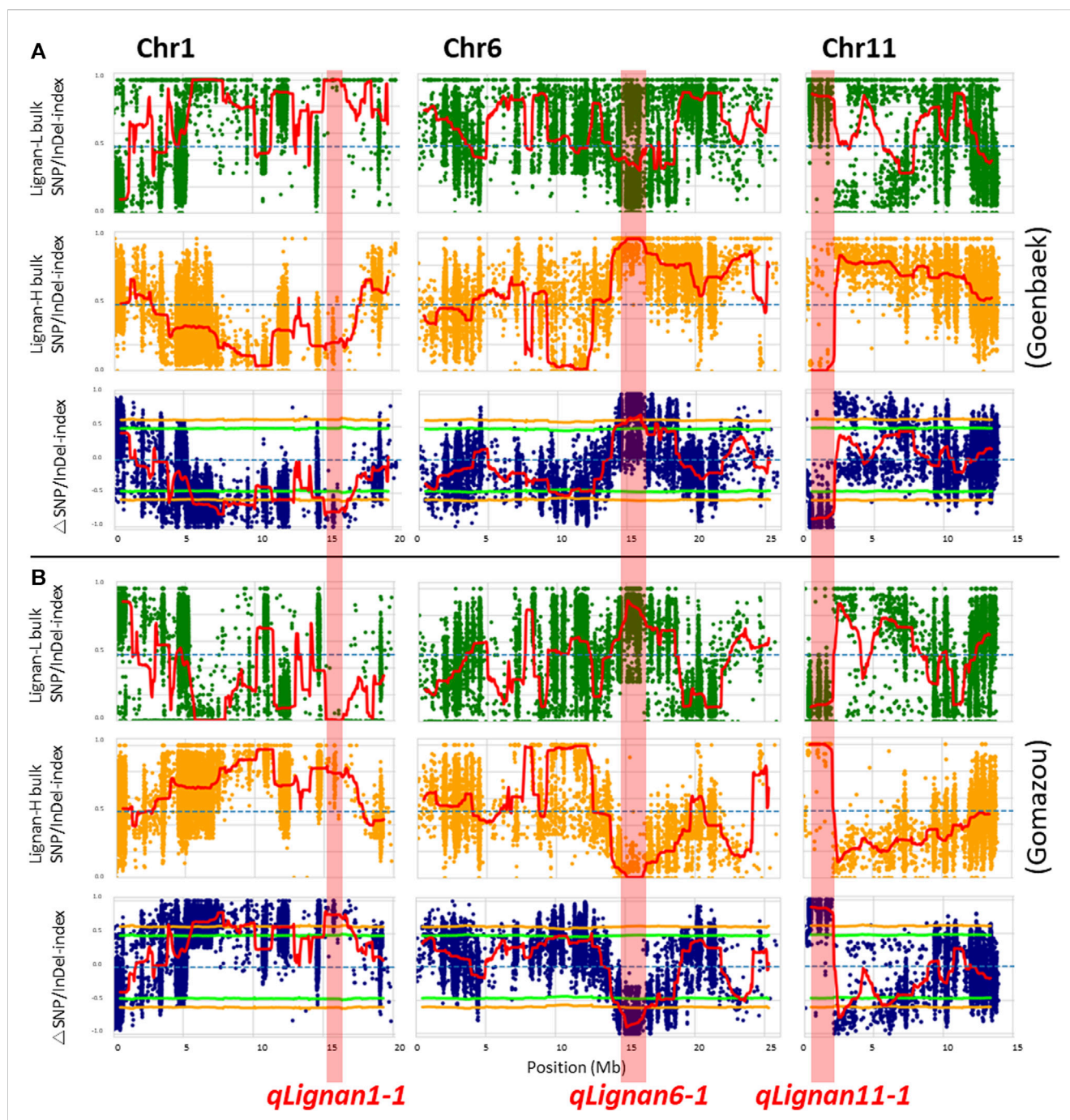


FIGURE 2

QTL-seq analysis for the identification of major effect QTLs controlling lignan content in sesame. (A) SNP/InDel-index plot for high- (green) and low-lignan (yellow) bulks, and Δ SNP/InDel-index (blue) based on 'Goenbaek' as a reference. (B) SNP/InDel-index plots for high- (green) and low-lignan (yellow) bulks, and Δ SNP/InDel-index (blue) based on 'Gomazou' as a reference. The average index value was plotted using red lines using a sliding window approach of 1 Mb intervals with 50 kb increments. The Δ SNP/InDel-index obtained by subtracting the SNP/InDel-index of low-lignan bulk from the SNP/InDel-index of high-lignan bulk was calculated with the statistical confidence interval under the null hypothesis of no QTL (orange, $p < 0.01$; and light green, $p < 0.05$). The three major effect QTLs with the $|\Delta$ SNP/InDel-index ≥ 0.7 were shadowed in red on chromosome 1, 6, and 11, and designated as *qLignan1-1*, *qLignan6-1*, and *qLignan11-1*, respectively.

amplification conditions were: 95°C for 5 min, followed by 30 cycles of 95°C for 10 s, 60°C for 15 s, and 72°C for 15 s. For CAPS marker assay, amplified products were digested with *Bsp*HI restriction enzymes. Amplified or digested products were separated and visualized using QIAxcel (an automated capillary electrophoresis

system by Qiagen, Hilden, Germany) for genotyping. The KASP primer design was performed by LGC genomics (London, UK) for two allele-specific forward primers and a common reverse primer based on 100 bp of both flanking sequences. The KASP-PCR amplification was conducted following the KASP technology

TABLE 3 Summary of the number of counter-SNPs/InDels near/or within genes in QTL regions.

QTLs (genes)	Number of counter-SNPs/InDels (genes)											
	1 kb upstream		5'UTR	CDS	Intron	3'UTR	1 kb downstream	Total				
<i>qLignan1-1</i>	0	(0)	0	(0)	11	(1)	0	(0)	0	(0)	11	(1)
<i>qLignan6-1</i>	125	(35)	0	(0)	141	(25)	62	(25)	140	(44)	468	(60)
<i>qLignan11-1</i>	74	(14)	0	(0)	76	(15)	31	(9)	51	(11)	232	(18)

manual (LGC Genomics, Beverly, MA, United States of America): 94°C for 15 min of activation, 10 cycles of 94°C for 20 s and a gradual decrease from 61°C (decrease at 0.6°C per cycle), and 26 cycles of 94°C for 20 s and 55°C for 1 min.

3 Results

3.1 Sesamin, sesamol, and lignan content in the parent and RIL, and the construction of two extreme bulks

The sesamin and sesamol content in the Goenbaek, Gomazou, and 257 F7 RILs seeds were investigated and the two compounds were combined to give the total lignan content. Goenbaek and Gomazou seeds had mean sesamin, sesamol, and lignan contents of 2.7 and 8.2 mg/g, 1.0 and 3.1 mg/g, and 3.7 and 11.3 mg/g, respectively; Gomazou seeds had higher amounts of all components (Table 1). The 257 RIL accessions had mean sesamin, sesamol, and lignan contents of 5.2, 2.2, and 7.5 mg/g, respectively, and the observed minimum and maximum within-population values suggested that some accessions resulted from transgressive segregation beyond the parental values (Table 1). Statistically significant correlations were observed between the three compounds (Supplementary Figure S1). In particular, sesamin ($r = 0.98$, $p < 0.001$) and sesamol ($r = 0.85$, $p < 0.001$) showed strong positive correlations with lignan. Based on the normal distribution of lignan content in the population, 10 individuals were selected from each end of the extremes to comprise a low bulk (Lignan-L) ranging from 1.7 to 3.5 mg/g and a high bulk (Lignan-H) ranging from 10.8 to 12.8 mg/g, with a mean of 2.7 and 11.7 mg/g, respectively (Figure 1, Supplementary Table S2).

3.2 Whole genome resequencing of parents and lignan-high and lignan-low bulks

A total of four samples were sequenced, including parents, Lignan-H, and Lignan-L, generating a total of 80.4 Gb of raw data (Supplementary Table S3). After filtering by quality score and adapter trimming, 170,831,013 reads (101 bp in length) were mapped for Goenbaek and 216,367,131 reads (151 bp in length) for Gomazou at average depths of 51x and 103x, respectively. Similarly, Lignan-L and Lignan-H mapped 48,240,157 and 48,403,785 reads, respectively, with an average depth of approximately 21x. All samples covered more than 95% of the sesame reference genome in terms of bases for which at least one read was mapped. Sequenced reads from the parental lines were deposited in the National Center for Biotechnology Information (NCBI) database under the BioProject accession number PRJNA1028140.

3.3 Calculation of SNP/InDel-index and QTL-seq analysis

To perform the QTL-seq analysis, Goenbaek and Gomazou were used as consensus genomes with replacement variants for the reference genome. The variants for analysis were selected by ensuring that only alleles that are homozygous and polymorphic

TABLE 4 Lists of 79 genes functionally annotated in three QTLs and functional description based on *Arabidopsis thaliana* homologues with KEGG orthology.

Gene ID	1 kb upstream	CDS				Intron	1 kb downstream	A.thaliana gene ID	Gene description ^a	KO ^b
		Synonymous	Missense	Nonsense	Frameshift					
SIN_1023526	0	0	0	0	0	11	0	AT2G41830	Uncharacterized protein	
SIN_1015666	0	0	1	0	0	0	0	AT2G14095	unknown protein	
SIN_1015667	0	0	0	0	0	1	0	AT5G55230	microtubule-associated proteins 65-1	
SIN_1015668	0	3	1	0	0	1	0	AT1G12420	ACT domain repeat 8 (ACR8)	
SIN_1015669	0	0	0	0	0	0	6	AT4G12080	AT-hook motif nuclear-localized protein 1 (AHL1)	
SIN_1015671	1	0	0	0	0	0	3	AT1G12410	CLP protease proteolytic subunit 2	
SIN_1015672	0	0	1	0	0	0	2	AT3G05100	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	
SIN_1015673	0	0	0	0	0	0	1	AT4G22790	MATE efflux family protein	
SIN_1015674	0	1	0	0	0	0	0	AT1G04110	Subtilase family protein	
SIN_1015676	2	0	0	0	0	0	1	AT2G27830	unknown protein	
SIN_1015679	1	0	0	0	0	0	0	AT1G12370	photolyase 1 (PHR1)	
SIN_1015680	0	0	0	0	0	1	1	AT1G12360	Sec1/munc18-like (SM) proteins superfamily	
SIN_1015683	0	0	0	0	0	2	0	AT4G22740	glycine-rich protein	
SIN_1015687	0	2	0	0	0	10	0	AT4G12230	alpha/beta-Hydrolases superfamily protein	K19367
SIN_1015688	0	0	0	0	0	0	1	AT1G02000	UDP-D-glucuronate 4-epimerase 2 (GAE2)	K08679
SIN_1015689	1	0	0	0	0	16	0	AT4G12290	Copper amine oxidase family protein	K00276
SIN_1015690	1	1	0	0	0	4	0	AT4G12290	Copper amine oxidase family protein	K00276
SIN_1015691	2	0	2	0	0	0	0	AT4G12300	polypeptide 4 (CYP706A4)	
SIN_1015692	0	0	0	0	0	7	0	AT4G12320	polypeptide 6 (CYP706A6)	
SIN_1015693	0	0	3	0	0	8	3	AT2G30370	CHALLAH (CHAL)	
SIN_1018396	6	0	0	0	0	0	2	AT4G12300	polypeptide 4 (CYP706A4)	
SIN_1018406	0	0	0	0	0	0	2	AT2G26900	Sodium Bile acid symporter family	
SIN_1018407	9	1	0	0	0	11	1	AT4G12350	myb domain protein 42 (MYB42)	
SIN_1018408	3	2	1	1	0	0	2	AT3G24255	RNA-directed DNA polymerase (reverse transcriptase)-related family protein	
SIN_1018409	2	0	0	0	0	0	7	AT5G64880	unknown protein	

(Continued on following page)

TABLE 4 (Continued) Lists of 79 genes functionally annotated in three QTLs and functional description based on *Arabidopsis thaliana* homologues with KEGG orthology.

Gene ID	1 kb upstream	CDS				Intron	1 kb downstream	A.thaliana gene ID	Gene description ^a	KO ^b
SIN_1018410	3	0	0	0	0	4	2	AT1G62780	unknown protein	
SIN_1018411	9	0	0	0	0	0	1	AT4G12390	pectin methylesterase inhibitor 1 (PME1)	
SIN_1018412	2	0	0	0	0	0	0	AT5G62360	Plant invertase/pectin methylesterase inhibitor superfamily protein	
SIN_1018413	10	2	0	0	0	1	2	AT1G62750	Translation elongation factor EFG/EF2 protein	
SIN_1018414	3	0	0	0	0	2	1	AT3G10195	Putative membrane lipoprotein	
SIN_1018415	2	0	0	0	0	1	1	AT4G12400	stress-inducible protein, putative	
SIN_1018416	0	0	0	0	0	0	10	AT4G22620	SAUR-like auxin-responsive protein family	K14488
SIN_1018417	0	0	0	0	0	12	7	AT1G12250	Pentapeptide repeat-containing protein	
SIN_1018420	0	5	1	0	0	4	0	AT1G12240	Glycosyl hydrolases family 32 protein	K01193
SIN_1018421	1	0	0	0	0	5	5	AT1G12230	Aldolase superfamily protein	K00616
SIN_1018422	1	1	0	0	0	1	5	AT4G22590	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein	K01087
SIN_1018423	2	0	0	0	0	0	2	AT5G55550	RNA-binding (RRM/RBD/RNP motifs) family protein	K14411
SIN_1018424	0	0	0	0	0	0	4	AT4G22600	unknown protein	
SIN_1018425	2	0	0	0	0	0	6	AT4G22580	Exostosin family protein	
SIN_1018426	6	0	3	0	0	0	2	AT1G63370	Flavin-binding monooxygenase family protein	
SIN_1018427	0	0	1	0	0	5	0	AT1G63370	Flavin-binding monooxygenase family protein	
SIN_1018428	1	0	0	0	0	0	4	AT1G62530	Plant protein of unknown function (DUF863)	
SIN_1018429	6	0	1	0	0	0	2	AT4G12440	adenine phosphoribosyl transferase 4 (APT4)	K00759
SIN_1018430	5	0	0	0	0	0	2	AT1G62520	unknown protein	
SIN_1018431	7	0	1	0	0	0	2	AT5G49500	Signal recognition particle, SRP54 subunit protein	K03106
SIN_1018432	2	3	0	0	0	8	2	AT4G22540	OSBP(oxysterol binding protein)-related protein 2A (ORP2A)	
SIN_1018433	8	2	1	0	0	3	2	AT1G12110	nitrate transporter 1	
SIN_1018434	0	0	0	0	0	6	0	AT3G12810	SNF2 domain-containing protein/helicase domain-containing protein	K11661
SIN_1018435	7	0	0	0	0	0	11	AT1G12100	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	

(Continued on following page)

TABLE 4 (Continued) Lists of 79 genes functionally annotated in three QTLs and functional description based on *Arabidopsis thaliana* homologues with KEGG orthology.

Gene ID	1 kb upstream	CDS				Intron	1 kb downstream	A.thaliana gene ID	Gene description ^a	KO ^b
SIN_1018437	2	0	0	0	0	0	6	AT1G12100	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	
SIN_1018438	2	0	0	0	0	5	3	AT2G13620	cation/hydrogen exchanger 15 (CHX15)	
SIN_1018440	1	5	4	0	0	0	3	AT4G15550	indole-3-acetate beta-D-glucosyltransferase (IAGLU)	
SIN_1018441	0	1	0	0	0	0	6	AT4G12490	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	
SIN_1018442	2	0	0	0	0	0	3	AT2G45180	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	
SIN_1018443	0	0	0	0	0	0	1	AT2G45180	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	
SIN_1018444	5	0	5	1	0	0	3	AT3G26120	terminal EAR1-like 1	
SIN_1018446	1	1	0	0	0	1	7	AT4G17260	Lactate/malate dehydrogenase family protein	K00016
SIN_1018447	0	2	1	0	0	22	1	AT4G16510	YbaK/aminoacyl-tRNA synthetase-associated domain	
SIN_1018448	7	0	0	0	0	0	1	AT4G16515	root meristem growth factor 6 (RGF6)	
SIN_1018492	0	0	0	0	0	0	1	AT4G16600	Nucleotide-diphospho-sugar transferases superfamily protein	
SIN_1018493	0	0	0	0	1	0	0	AT2G30575	los glycosyltransferase 5 (LGT5)	K13648
SIN_1005674	0	0	0	0	0	1	0	AT1G16760	Protein kinase protein with adenine nucleotide alpha hydrolases-like domain	
SIN_1005676	1	0	0	0	0	0	1	AT4G01250	WRKY family transcription factor22	K13425
SIN_1005696	1	0	0	0	0	0	0	AT4G18020	CheY-like two-component responsive regulator family protein	
SIN_1009908	2	0	0	0	0	14	0	AT2G45650	AGAMOUS-like 6 (AGL6)	
SIN_1009909	0	3	4	0	0	4	21	AT4G00230	xylem serine peptidase 1 (XSP1)	
SIN_1009910	0	3	0	0	1	4	0	AT4G00230	xylem serine peptidase 1 (XSP1)	
SIN_1009911	21	0	0	0	0	6	0	AT1G01910	P-loop containing nucleoside triphosphate hydrolases superfamily protein	
SIN_1009951	1	0	0	0	0	1	1	AT3G53740	Ribosomal protein L36e family protein	K02920
SIN_1009952	3	2	0	0	0	12	0	AT2G45880	beta-amylase 7	K01177
SIN_1009953	2	4	7	0	0	5	12	AT4G00440	Protein of unknown function (DUF3741)	
SIN_1009954	21	0	0	0	0	2	1	AT1G01650	SIGNAL PEPTIDE PEPTIDASE-LIKE 4 (SPPL4)	
SIN_1009955	0	1	0	0	0	6	1	AT3G61415	SKP1-like 21 (SK21)	

(Continued on following page)

TABLE 4 (Continued) Lists of 79 genes functionally annotated in three QTLs and functional description based on *Arabidopsis thaliana* homologues with KEGG orthology.

Gene ID	1 kb upstream	CDS	1 kb downstream	Intron	1 kb downstream	A.thaliana gene ID	Gene description ^a	KO ^b
SIN_1009956	6	0	0	0	0	ATI.G01610	glycerol-3-phosphate acyltransferase 4 (GPAT4)	K13508
SIN_1009957	4	0	0	3	3	AT4G00370	Major facilitator superfamily protein	
SIN_1009958	2	0	0	1	1	AT2G46020	transcription regulatory protein SNF2, putative	
SIN_1009959	1	1	0	8	1	ATI.G63800	ubiquitin-conjugating enzyme 5 (UBC5)	K10576
SIN_1009960	2	1	0	2	6	AT4G00335	RING-H2 finger B1A (RHB1A)	
SIN_1009961	7	1	0	7	3	ATI.G01580	ferric reduction oxidase 2	

^aThe description follows *Arabidopsis thaliana* genes of TAIR, version 10.

^bKO: Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology.

to parental align to one another, resulting in a total of 481,298 SNP/InDel variants (416,102 SNPs and 65,195 InDels). For these alleles mapped to consensus genomes, the SNP/InDel-index between Lignan-L and Lignan-H bulks was calculated using a sliding window approach with a window size of 1 Mb and a step size of 50 Kb. For example, if both bulks have an identical nucleotide base at an identified position, the SNP/InDel-index value at this position will be 0. Conversely, if the reference genome and the identified nucleotide base are completely inconsistent, the SNP/InDel-index value will be 1. Following the formula: the $\Delta(\text{SNP/InDel-index})$ value as the subtraction of the index between Lignan-L and Lignan-H was plotted as a mirror-like reflected image across all 13 chromosomes according to the two parental consensus genomes (Supplementary Figure S2).

After calculating the $\Delta(\text{SNP/InDel-index})$ value using Goenbaek and Gomazou as references, 12 and 13 genomic regions were obtained as candidate QTLs controlling lignin at a 99% significance level, respectively, and a total of 10 common genetic regions were found (Table 2). Among them, three major QTLs with $\Delta(\text{SNP/InDel-index}) > 0.7$, which were expected to be closely associated with the trait, were identified on chromosomes 1, 6, and 11, and were namely: *qLignan1-1*, *qLignan6-1*, and *qLignan11-1* (Figure 2). The *qLignan6-1* region comprises 1.9 Mb (14,600,000–16,500,000 bp), and the $\Delta(\text{SNP/InDel-index})$ value was close to '1' when calculated on the basis of Goenbaek and close to '-1' when calculated on the basis of Gomazou (Table 2; Figure 2). This result implies that this QTL region consists of a pool (Lignan-H) of individuals with a region inherited from high lignin in Gomazou and a pool (Lignan-L) of individuals with a region inherited from low lignin in Goenbaek, and lignan content is partially regulated by underlying genetic factors in this region. In contrast, *qLignan1-1* and *qLignan11-1* comprise a region of 1.65 Mb (15,100,000–16,750,000 bp) and 1.6 Mb (500,000–2,100,000 bp), respectively, and showed the opposite pattern of index values to *qLignan6-1* (Table 2; Figure 2).

3.4 Functional classification of variants and candidate gene search

Variant analysis was performed to explore candidate SNPs/InDels or genes associated with sesame seed lignin content. First, we used the information from the reference genome to search for genes located in three QTL intervals, *qLignan1-1*, *qLignan6-1*, and *qLignan11-1*, and found 220, 195, and 146 genes, respectively (Table 3). A total of 2,272 variants were then selected, with counter-SNPs/InDels being defined as variants that were completely biased to one side between the bulks representing index values of 1 or -1. Using their location, the functional classification of the variant was performed according to the annotation and expected change in function within or near the gene. As a result, a total of 11 counter-SNPs/InDels in *qLignan1-1* were found to be in the intron region of one gene, 468 and 232 counter-SNPs/InDels in *qLignan6-1* and *qLignan11-1*, respectively, were associated with 60 and 18 genes, respectively (Table 3). To find genes that were previously reported to affect secondary metabolite synthesis, functional descriptions of the genes were obtained using BLASTP against *A. thaliana* protein sequences. The KEGG analysis revealed that out of a total of 73 genes, 18 were assigned KEGG orthology accession numbers and mapped to

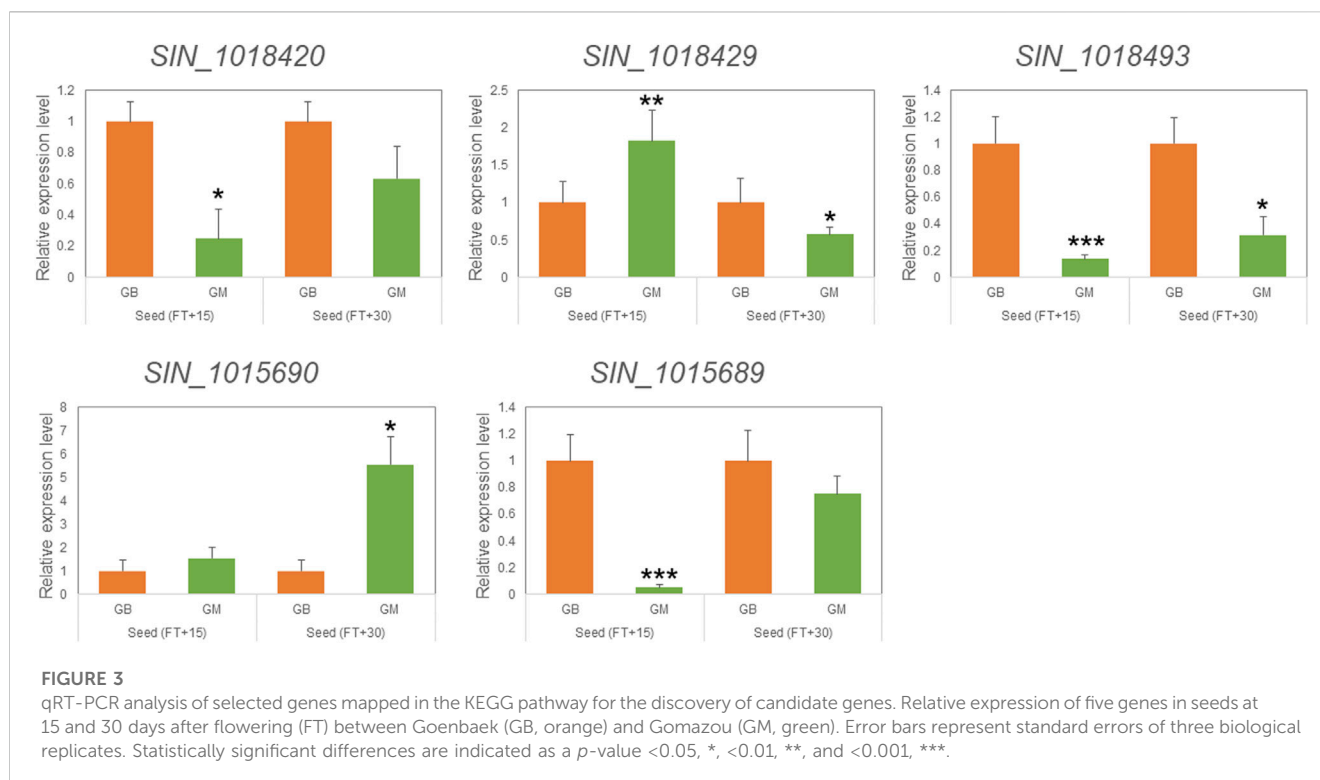


TABLE 5 DNA marker information to identify alleles for lignan content controlling in sesame.

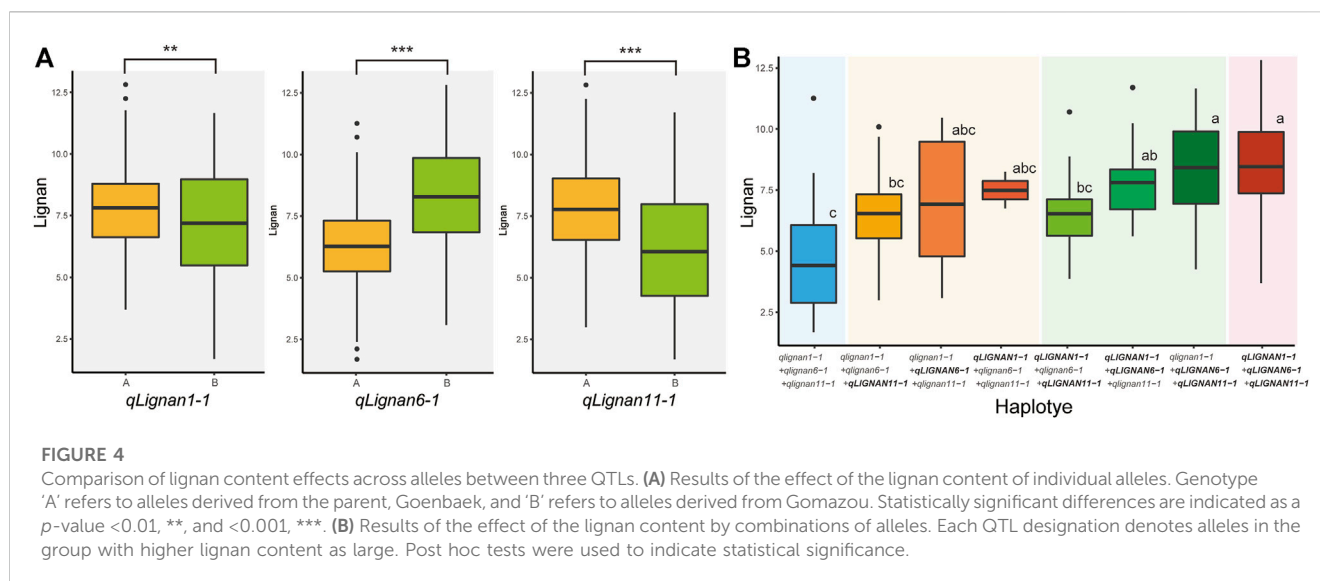
Marker	Position	Goenbaek	Gomazou	Primer sequence
InDel	Si1_15740491	TTTAATTTATTTCATTAATTTA	TTTAATTTA	F: TCGTACGCATGAGATGAAACA R: CATCCCAAATCCAAGTTGT
KASP	Si6_15239077	A	G	Primer allele X
				CCTTCACCAATTCCTCCTCG
				Primer allele Y
				CCCTTCACCAATTCCTCCTCA
				Primer common
				AACTGAAAGGCTCAAGCGCGTTT
CAPS (<i>Bsp</i> HI)	Si11_833990	T	C	F: ATTTCGATCGATGCCCCATAG
				R: TTCATGTAACGAGGAATCAATCA

34 pathways, with 10 genes belonging to metabolic pathways, followed by seven genes belonging to secondary metabolite biosynthesis (Table 4). Based on the genes assigned to metabolic pathways, five genes were selected for transcriptional expression analysis (Figure 3). Among them, three genes (*SIN_1018420*, *SIN_1018429*, *SIN_1018493*) have CDS variants with altered amino acid sequences, and two genes (*SIN_1015690* and *SIN_1015689*) belonging to the tyrosine and phenylalanine pathway. The results revealed significantly lower levels of *SIN_1018420*, a homolog of *AT1G12240* (glycosylhydrolase family 32 protein), in immature seeds at the early stage (15 days after flowering) of Gomazou compared to those of Goenbaek. At both maturation stages, significant differences in gene expression were observed between the parental lines for *SIN_1018429*, a homolog of *AT2G30575* (adenine phosphoribosyltransferase 4), and *SIN_1018493*, a homolog of *AT2G30575* (lignin glycosyltransferase 5). Two gene

homologs of *AT4G12290*, *SIN_1015690*, and *SIN_1015689*, which encode proteins of the copper amine oxidase family, exhibited different patterns of gene expression. Specifically, *SIN_1015690* showed higher expression in the Gomazou compared to the Goenbaek during later stages of immature seeds (30 days after flowering). Conversely, *SIN_1015689* showed significantly lower gene expression in the Gomazou than in the Goenbaek during early stages of immature seeds.

3.5 Additive effect of alleles and QTL stacking

To investigate the allelic effect of the three QTL with high index values and to correlate the genotype and phenotype of individuals in the population, KASP, InDel, and CAPS markers were developed



using counter-SNP/InDel (Table 5). The allelic genotyping markers were used for variants located in the CDS region of *qLignan6-1*, in the intergenic region of *qLignan11-1*, and in the intron region of *qLignan1-1*. The differences in lignan content associated with each QTL allele in the population were statistically significant (Figure 4A). Specifically, *qLignan1-1* and *qLignan11-1* exhibited higher lignan content along the Goenbaek allele, while *qLignan6-1* showed higher lignan content along the Gomazou allele. The investigation of potential interactions among the QTLs revealed that specific combinations of alleles resulted in high-lignan content. Combinations of two or more alleles were found to contribute to an increase in lignan content in comparison to single alleles. The highest lignan content was observed when all three QTL alleles were incorporated (Figure 4B). Combinations including *Lignan6-1*, which showed significantly higher lignan content when combined with other QTLs, were particularly noticeable. This finding provides compelling evidence for the substantial contribution of the Gomazou-derived allele to lignan content.

4 Discussion

Identification of QTLs and the development of molecular markers that are closely linked to useful genes can be a useful tool in the breeding of new varieties by applying MAS, saving time and labor, and accelerating breeding cycles (Paran and Zamir, 2003; Collard et al., 2005). It is important that QTLs or markers are used to estimate phenotypic expression at the molecular level, particularly because direct measurement of quantitative traits is difficult (Nadeem et al., 2018). The SNP markers associated with the *ahFAD2* gene, responsible for the introgression of high oleic acid content in peanut and SSR markers for the selection of the *Gpc-B1* gene, responsible for the selection of high grain protein varieties, are examples of advancing the effectiveness of breeding using marker selection resulting from QTL studies (Vishwakarma et al., 2014; Bera et al., 2018). Recently, by fusing genome-level information on components such as secondary metabolites with information on

the biosynthesis of compounds using advanced NGS technology, research is expanding into a new area called the metabolome (Keurentjes, 2009; Carreno-Quintero et al., 2013). For example, 4,681 metabolic quantitative trait loci (mQTLs) were identified as profiling metabolites that accumulate in seeds at different developmental stages in rice (Li et al., 2019), and 42 loci were identified from 1,098 associations in a metabolite-based genome-wide association study (mGWAS), thus suggesting breeding strategies for flavonoid pathway metabolites in wheat (Chen et al., 2020).

Finding a resource that is suitable for the target trait and generating a genetic population to find the genes that control the trait are necessary to achieve successful QTL results. In this study, to find the regulators of lignan content in sesame, an F7 RIL population was established using Goenbaek, an elite variety with stable cultivation but low-lignan content, and Gomazou, a variety with high-lignan content, as parental lines. We then compared sequence information from a pool of transgressive segregating individuals containing high and low-lignan content for the QTL analysis. Based on overlapped genetic regions from comparing two parental versions as a reference for the calculation of sequence frequencies to avoid errors caused by heterozygous or missing positions in the mapping process, three major QTLs (*qLignan1-1*, *qLignan6-1*, and *qLignan11-1*) were identified. In previous studies, QTL results using the highly biased variant frequency index in separated bulks have been shown to be reliable in comparison to traditional QTL analysis in rice, tomato, and maize (Chen et al., 2021; Topcu et al., 2021; Yang et al., 2021). The identified *qLignan11-1* is also located in close proximity to *qSim_11.1* and *qSmol_11.1*, which were reported to have high phenotypic effects (67.69% and 46.05%, respectively) on sesamin and sesamol content using F8 RILs in sesame (Xu et al., 2021). Although using the lignan with the combined content of sesamin and sesamol as an indicator is a potential limitation of this study, the strong correlation between the components (Supplementary Figure S1) supports our finding on a comprehensive locus that can be used for breeding. However, a genetic approach to the synthesis of sesamin and sesamol is needed for further functional analysis.

The relatively large size of the QTL region, which ranged from 1.6 to 1.9 Mb and contained 140–220 genes, limited the selection of specific candidate genes. To overcome this limitation, our study followed a two-step approach to candidate gene selection. First, some of the variants were extremely segregated within the region confined to the QTL region, whereas the SNP index values in most of the regions were heterozygous. These extreme variants were defined as counter-SNPs/InDels, with 11, 468, and 232 variants within each QTL region, spanning one, 60, and 18 genes, respectively (Table 3). Then, we conducted a comparative analysis based on the gene annotation of *A. thaliana*. By selecting genes belonging to secondary metabolic pathways, which are expected to be functionally altered due to variations in genomic regions affecting protein coding (Table 4), we validated their expression levels using parental lines with large differences in lignan content. As a result, All five genes, *SIN_1018420*, *SIN_1018429*, *SIN_1018493*, *SIN_1015690*, and *SIN_1015689*, were differentially expressed in the seeds (Figure 3), but further functional studies are needed to elucidate the trait-associated roles of these genes. Our results suggest that despite the large number of genes distributed among the potential loci, the gene and pathway databases can be helpful in the logical selection of promising candidate genes.

Quantitative traits are determined by multiple genes, each of which may have a large or small effect. Since the quantitative trait is often determined by the interaction of multiple genes with environmental influences, it is preferable to use methods such as QTL-seq to find useful genetic regions focusing on major QTL; although there are limitations in detecting minor QTL (Dong et al., 2021; Sheng et al., 2021). While the elucidation of gene functions at the molecular level is important, it is also important that QTL study results are practically applied in breeding through the use of major gene-induced markers for the efficient selection of pedigrees for time and labor reduction (Xu et al., 2018; Yang et al., 2019). This study verified the reliability of the QTL with KASP, InDel, and CAPS markers, showing that the allele derived from the lignan-rich Gomazou, *qLignan6-1*, and two alleles derived from Goenbaek, *qLignan1-1* and *qLignan11-1*, are dominant. The influence of the allele derived from the high-lignan Gomazou is higher compared to the two QTLs derived from the low-lignan Goenbaek, suggesting that the QTLs derived from the Goenbaek are likely to be minor QTLs. Interactions between alleles were also analyzed to exploit the additive allele effect and found that the dominant form of *qLignan6-1* had greater synergy when combined with other QTLs. It is therefore expected that *qLignan6-1* will play an important role in controlling lignan content, in addition to additive effects with other loci. The discovery of *qLignan6-1* will be useful in the breeding process as a marker for the introgression of useful genes from high-lignan resources into a new variety.

Furthermore, lignans in sesame have been shown to be functional components with antioxidant, anti-dementia, and antihypertensive activities, and the genetic mechanisms required for lignan synthesis have been reported using QTL, GWAS, and transcriptomics in several studies (Harada et al., 2020; Xu et al., 2021; Dossou et al., 2023). CYP92B14, an enzyme that converts sesamin to sesamolin and sesaminol, which was elucidated through functional characterization (Harada et al., 2020), and the *SiSNT1* gene located on chromosome 11, inferred through genetic mapping (Xu et al., 2021; Dossou et al., 2023),

may explain the genetic mechanisms of lignan synthesis in sesame. However, these are still poorly understood compared to other major crops. In this study, we found two novel genetic contributors that regulate lignan content, *qLignan1-1* and *qLignan6-1*, along with *qLignan11-1* at the same location as in previous study. By showing the utility of marker interaction that can be used in practical breeding strategies, we are confident that this will be useful as a basic resource for the development of high-functional lignan varieties.

Data availability statement

The data presented in the study are deposited in the National Center for Biotechnology Information (NCBI) repository (<https://www.ncbi.nlm.nih.gov/>) under BioProject accession number PRJNA1028140.

Author contributions

SK: Writing–original draft. EL: Writing–original draft. JL: Investigation. YA: Writing–original draft. EO: Investigation. JK: Investigation. SK: Investigation. MK: Investigation. ML: Project administration. K-SC: Conceptualization and Project administration.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Cooperative Research Program for Agriculture Science and Technology Development (Project PJ016076022023), Rural Development Administration, Republic of Korea.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1289793/full#supplementary-material>

References

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi:10.1038/nbt.2095
- Alseekh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., et al. (2015). Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell* 27, 485–512. doi:10.1105/tpc.114.132266
- Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., and Edwards, D. (2020). Plant pan-genomes are the new reference. *Nat. Plants* 6, 914–920. doi:10.1038/s41477-020-0733-0
- Bera, S. K., Kamdar, J. H., Kasundra, S. V., Dash, P., Maurya, A. K., Jasani, M. D., et al. (2018). Improving oil quality by altering levels of fatty acids through marker-assisted selection of ahfad2 alleles in peanut (*Arachis hypogaea* L.). *Euphytica* 214, 162–215. doi:10.1007/s10681-018-2241-0
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170
- Carreno-Quintero, N., Bouwmeester, H. J., and Keurentjes, J. J. (2013). Genetic analysis of metabolome–phenotype interactions: from model to crop species. *Trends Genet.* 29, 41–50. doi:10.1016/j.tig.2012.09.006
- Chen, J., Hu, X., Shi, T., Yin, H., Sun, D., Hao, Y., et al. (2020). Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnol. J.* 18, 1722–1735. doi:10.1111/pbi.13335
- Chen, Z., Tang, D., Hu, K., Zhang, L., Yin, Y., Ni, J., et al. (2021). Combining QTL-seq and linkage mapping to uncover the genetic basis of single vs. paired spikelets in the advanced populations of two-ranked maize x teosinte. *BMC Plant Biol.* 21, 572. doi:10.1186/s12870-021-03353-3
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *SnpEff. Fly.* 6, 80–92. doi:10.4161/fly.19695
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169–196. doi:10.1007/s10681-005-1681-5
- Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 557–572. doi:10.1098/rstb.2007.2170
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330
- Dar, A. A., and Arumugam, N. (2013). Lignans of sesame: purification methods, biological activities and biosynthesis – a review. *Bioorg. Chem.* 50, 1–10. doi:10.1016/j.bioorg.2013.06.009
- Dong, Z., Alam, M. K., Xie, M., Yang, L., Liu, J., Helal, M. M. U., et al. (2021). Mapping of a major QTL controlling plant height using a high-density genetic map and QTL-seq methods based on whole-genome resequencing in *Brassica napus*. *G3 (Bethesda)* 11, jkab118. doi:10.1093/g3journal/jkab118
- Dossou, S. S. K., Song, S., Liu, A., Li, D., Zhou, R., Berhe, M., et al. (2023). Resequencing of 410 sesame accessions identifies SINST1 as the major underlying gene for lignans variation. *Int. J. Mol. Sci.* 24, 1055. doi:10.3390/ijms24021055
- Harada, E., Murata, J., Ono, E., Toyonaga, H., Shiraiishi, A., Hideshima, K., et al. (2020). (+)-Sesamin-oxidising CYP92B14 shapes specialised lignan metabolism in sesame. *Plant J.* 104, 1117–1128. doi:10.1111/tpj.14989
- Keurentjes, J. J. (2009). Genetical metabolomics: closing in on phenotypes. *Curr. Opin. Plant Biol.* 12, 223–230. doi:10.1016/j.pbi.2008.12.003
- Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D., and Varshney, R. K. (2020). Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci.* 25, 148–158. doi:10.1016/j.tplants.2019.10.012
- Kim, M. Y., Kim, S., Lee, J., Kim, J. I., Oh, E., Kim, S. W., et al. (2023). Lignan-rich sesame (*Sesamum indicum* L.) cultivar exhibits *in vitro* anti-cholinesterase activity, anti-neurotoxicity in amyloid- β induced SH-SY5Y cells, and produces an *in vivo* nootropic effect in scopolamine-induced memory impaired mice. *Antioxidants (Basel)* 12, 1110. doi:10.3390/antiox12051110
- Kim, S.-U., Lee, M.-H., Pae, S.-B., Oh, E.-Y., Kim, J.-I., and Ha, T.-J. (2018). A sesame variety “Goenbaek” with *Phytophthora* Blight disease resistance and high yield. *Korean J. Breed. Sci.* 50, 256–260. doi:10.9787/KJBS.2018.50.3.256
- Langyan, S., Yadava, P., Sharma, S., Gupta, N. C., Bansal, R., Yadav, R., et al. (2022). Food and nutraceutical functions of sesame oil: an underutilized crop for nutritional and health benefits. *Food Chem.* 389, 132990. doi:10.1016/j.foodchem.2022.132990
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Li, K., Wang, D., Gong, L., Lyu, Y., Guo, H., Chen, W., et al. (2019). Comparative analysis of metabolome of rice seeds at three developmental stages using a recombinant inbred line population. *Plant J.* 100, 908–922. doi:10.1111/tpj.14482
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262
- Majdalawieh, A. F., Massri, M., and Nasrallah, G. K. (2017). A comprehensive review on the anti-cancer properties and mechanisms of action of sesamin, a lignan in sesame seeds (*Sesamum indicum*). *Eur. J. Pharmacol.* 815, 512–521. doi:10.1016/j.ejphar.2017.10.020
- Michailidis, D., Angelis, A., Aliagiannis, N., Mitakou, S., and Skaltsounis, L. (2019). Recovery of sesamin, sesamol, and minor lignans from sesame oil using solid support-free liquid–liquid extraction and chromatography techniques and evaluation of their enzymatic inhibition properties. *Front. Pharmacol.* 10, 723. doi:10.3389/fphar.2019.00723
- Monteiro, E. M., Chibli, L. A., Yamamoto, C. H., Pereira, M. C., Vilela, F. M., Rodarte, M. P., et al. (2014). Antinociceptive and anti-inflammatory activities of the sesame oil and sesamin. *Nutrients* 6, 1931–1944. doi:10.3390/nu6051931
- Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yildiz, M., et al. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Amp Biotechnol. Equip.* 32, 261–285. doi:10.1080/13102818.2017.1400401
- Noguchi, A., Fukui, Y., Iuchi-Okada, A., Kakutani, S., Satake, H., Iwashita, T., et al. (2008). Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1 glucosyltransferases. *Plant J.* 54, 415–427. doi:10.1111/j.1365-313X.2008.03428.x
- Paran, I., and Zamir, D. (2003). Quantitative traits in plants: beyond the QTL. *Trends Genet.* 19, 303–306. doi:10.1016/S0168-9525(03)00117-3
- Ray, S., and Satya, P. (2014). Next generation sequencing technologies for next generation plant breeding. *Front. Plant Sci.* 5, 367. doi:10.3389/fpls.2014.00367
- Sheng, C., Song, S., Zhou, R., Li, D., Gao, Y., Cui, X., et al. (2021). QTL-seq and transcriptome analysis disclose major QTL and candidate genes controlling leaf size in Sesame (*Sesamum indicum* L.). *Front. Plant Sci.* 12, 580846. doi:10.3389/fpls.2021.580846
- Sugihara, Y., Young, L., Yaegashi, H., Natsume, S., Shea, D. J., Takagi, H., et al. (2022). High-performance pipeline for MutMap and QTL-seq. *PeerJ* 10, e13170. doi:10.7717/peerj.13170
- Sun, Y., Shang, L., Zhu, Q. H., Fan, L., and Guo, L. (2022). Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci.* 27, 391–401. doi:10.1016/j.tplants.2021.10.006
- Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi:10.1111/tpj.12105
- Topcu, Y., Sapkota, M., Illa-Berenguer, E., Nambesan, S. U., and Van Der Knaap, E. (2021). Identification of blossom-end rot loci using joint QTL-seq and linkage-based QTL mapping in tomato. *Theor. Appl. Genet.* 134, 2931–2945. doi:10.1007/s00122-021-03869-0
- Uzun, B., Arslan, Ç., and Furat, Ş. (2008). Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (*Sesamum indicum* L.). *J. Am. Oil Chem. Soc.* 85, 1135–1142. doi:10.1007/s11746-008-1304-0
- Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M. E. (2021). Designing future crops: genomics-assisted breeding Comes of Age. *Trends Plant Sci.* 26, 631–649. doi:10.1016/j.tplants.2021.03.010
- Vishwakarma, M. K., Mishra, V. K., Gupta, P. K., Yadav, P. S., Kumar, H., and Joshi, A. K. (2014). Introgression of the high grain protein gene Gpc-B1 in an elite wheat variety of Indo-Gangetic Plains through marker assisted backcross breeding. *Curr. Plant Biol.* 1, 60–67. doi:10.1016/j.cpb.2014.09.003
- Wan, Y., Li, H., Fu, G., Chen, X., Chen, F., and Xie, M. (2015). The relationship of antioxidant components and antioxidant activity of sesame seed oil. *J. Sci. Food Agric.* 95, 2571–2578. doi:10.1002/jsfa.7035
- Wang, L., Xia, Q., Zhang, Y., Zhu, X., Zhu, X., Li, D., et al. (2016). Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. *BMC Genomics* 17, 31. doi:10.1186/s12864-015-2316-4
- Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., et al. (2014). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 15, R39. doi:10.1186/gb-2014-15-2-r39
- Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., et al. (2015). Genetic discovery for oil production and quality in sesame. *Nat. Commun.* 6, 8609. doi:10.1038/ncomms9609

- Xu, F., Zhou, R., Dossou, S. S. K., Song, S., and Wang, L. (2021). Fine mapping of a major pleiotropic QTL associated with sesamin and sesamol variation in sesame (*Sesamum indicum* L.). *Plants (Basel)* 10, 1343. doi:10.3390/plants10071343
- Xu, P., Cai, F., Liu, X., and Guo, L. (2015). Sesamin inhibits lipopolysaccharide-induced proliferation and invasion through the p38-MAPK and NF- κ B signaling pathways in prostate cancer cells. *Oncol. Rep.* 33, 3117–3123. doi:10.3892/or.2015.3888
- Xu, Y., Zhang, X.-Q., Harasymow, S., Westcott, S., Zhang, W., and Li, C. (2018). Molecular marker-assisted backcrossing breeding: an example to transfer a thermostable β -amylase gene from wild barley. *Mol. Breed.* 38, 63. doi:10.1007/s11032-018-0828-8
- Yang, G., Chen, S., Chen, L., Gao, W., Huang, Y., Huang, C., et al. (2019). Development and utilization of functional KASP markers to improve rice eating and cooking quality through MAS breeding. *Euphytica* 215, 66–12. doi:10.1007/s10681-019-2392-7
- Yang, L., Wang, J., Han, Z., Lei, L., Liu, H. L., Zheng, H., et al. (2021). Combining QTL-seq and linkage mapping to fine map a candidate gene in qCTS6 for cold tolerance at the seedling stage in rice. *BMC Plant Biol.* 21, 278. doi:10.1186/s12870-021-03076-5
- Yasumoto, S., and Katsuta, M. (2006). Breeding a high-lignan-content sesame cultivar in the prospect of promoting metabolic functionality. *Jpn. Agric. Res. Q. JARQ.* 40, 123–129. doi:10.6090/jarq.40.123
- Yokota, T., Matsuzaki, Y., Koyama, M., Hitomi, T., Kawanaka, M., Enoki-Konishi, M., et al. (2007). Sesamin, a lignan of sesame, down-regulates cyclin D1 protein expression in human tumor cells. *Cancer Sci.* 98, 1447–1453. doi:10.1111/j.1349-7006.2007.00560.x