



Article

Resequencing of 410 Sesame Accessions Identifies *SINST1* as the Major Underlying Gene for Lignans Variation

Senouwa Segla Koffi Dossou ^{1,2,†} , Shengnan Song ^{1,†}, Aili Liu ¹, Donghua Li ¹, Rong Zhou ¹, Muez Berhe ¹ , Yanxin Zhang ¹, Chen Sheng ¹, Zhijian Wang ¹, Jun You ^{1,*} and Linhai Wang ^{1,*}

¹ Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan 430062, China

² Laboratory of Plant Physiology and Biotechnologies, Faculty of Sciences, University of Lomé, Lomé 01BP 1515, Togo

* Correspondence: junyou@caas.cn (J.Y.); wanglinhai@caas.cn (L.W.);

Tel.: +86-18607147952 (J.Y.); +86-15926338805 (L.W.)

† These authors contributed equally to this work.

Abstract: Sesame is a promising oilseed crop that produces specific lignans of clinical importance. Hence, a molecular description of the regulatory mechanisms of lignan biosynthesis is essential for crop improvement. Here, we resequence 410 sesame accessions and identify 5.38 and 1.16 million SNPs (single nucleotide polymorphisms) and InDels, respectively. Population genomic analyses reveal that sesame has evolved a geographic pattern categorized into northern (NC), middle (MC), and southern (SC) groups, with potential origin in the southern region and subsequent introduction to the other regions. Selective sweeps analysis uncovers 120 and 75 significant selected genomic regions in MC and NC groups, respectively. By screening these genomic regions, we unveiled 184 common genes positively selected in these subpopulations for exploitation in sesame improvement. Genome-wide association study identifies 17 and 72 SNP loci for sesamin and sesamol variation, respectively, and 11 candidate causative genes. The major pleiotropic SNP^{C/A} locus for lignans variation is located in the exon of the gene *SiNST1*. Further analyses revealed that this locus was positively selected in higher lignan content sesame accessions, and the “C” allele is favorable for a higher accumulation of lignans. Overexpression of *SiNST1*^C in sesame hairy roots significantly up-regulated the expression of *SiMYB58*, *SiMYB209*, *SiMYB134*, *SiMYB276*, and most of the monolignol biosynthetic genes. Consequently, the lignans content was significantly increased, and the lignin content was slightly increased. Our findings provide insights into lignans and lignin regulation in sesame and will facilitate molecular breeding of elite varieties and marker-traits association studies.

Keywords: sesame; genome resequencing; lignan; genomic signatures; NST1; GWAS



Citation: Dossou, S.S.K.; Song, S.; Liu, A.; Li, D.; Zhou, R.; Berhe, M.; Zhang, Y.; Sheng, C.; Wang, Z.; You, J.; et al. Resequencing of 410 Sesame Accessions Identifies *SINST1* as the Major Underlying Gene for Lignans Variation. *Int. J. Mol. Sci.* **2023**, *24*, 1055. <https://doi.org/10.3390/ijms24021055>

Academic Editors: Andrés J. Cortés and Hai Du

Received: 9 November 2022

Revised: 12 December 2022

Accepted: 13 December 2022

Published: 5 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sesame (*Sesamum indicum* L., $2n = 26$) is a member of the Pedaliaceae family and one of the most important oilseed and industrial crops worldwide. It is primarily cultivated in subtropical and tropical regions in Asia, Africa, Latin America, and Central America [1]. Sesame seeds are promising sources of nutritional components and nutraceuticals for food and pharmacological industries [2,3]. It contains specific lignans, including sesamin, sesamol, sesaminol, sesamol, etc., that have shown diverse physiological abilities such as anti-cancers, antioxidative, anti-diabetes, anti-inflammatory, anti-hypertensive, anti-proliferative, anti-melanogenesis, anti-depression and memory loss, anti-osteonecrosis, and anti-osteoporosis [4–9]. Although it is documented that sesame is the one of the most ancient oilseed crops known to humans, its origins are still a subject of debate, and its progenitor also is not determined [1,10–12]. Yu et al. [13] found that the domestication of sesame in China and India might have occurred independently. Other reports indicate that sesame was introduced in China in 2200 BC during the Han dynasty [14]. China is

one of the leading countries in terms of sesame production and consumption (www.fao.org/statistics/en/, accessed on 22 November 2021) and has collected and preserved over 7000 accessions in its germplasms [15]. However, sesame's domestication, population structure, and genetic diversity remain largely unknown.

The release of the sesame reference genome via de novo assembly has allowed genomic studies on sesame biology [12]. The sesame reference genome quality has been improved via the resequencing of 430 recombinant inbred lines [16], which has provided a solid basis for investigating the molecular mechanisms controlling critical quality and agronomic traits of sesame such as high yield, high oil and lignans contents, seed coat color, plant architecture, and melatonin. Significant loci and candidate causative genes underlying these traits have been identified except for lignans variation [15,17–19]. The sesame-specific lignans seem to have evolved from defense and accommodation to biotic stresses [20–22], suggesting their underlying genes were positively selected during domestication and breeding [13]. Therefore, identifying adaptation-related genes may help understand sesame evolution [23]. The domestication and subsequent breeding of crops generate selective sweeps, eliminating or reducing variation amid the nucleotides adjoining a mutation in DNA [24,25]. Although large-scale genetic variations among diverse accessions of sesame have been reported [15], the detected genomic differentiation between the subpopulations was very low ($F_{ST} = 0.02$), and it was impossible to unravel regions under selective sweeps for exploitation in the crop improvement.

Lignans are a diversified group of phytochemicals that play essential roles in plants' interactions and adaptations to environmental conditions [26,27]. They are biosynthesized from phenylalanine via the phenylpropanoid pathway and mostly share the same precursors with lignin [26,28,29]. Most naturally occurring lignans are derived from the oxidative coupling of E-coniferyl alcohol moieties [26,28,30]. In sesame, the biosynthetic genes of the major lignans, sesamin and sesamol have been identified, and their mechanisms of action elucidated [30,31]. Recently, two laccase (*SiLAC1* and *SiLAC39*) and two peroxidase (*SiPOD52* and *SiPOD63*) family genes were identified as the potential pinosresinol synthase genes [32]. However, the regulatory mechanisms of lignan biosynthesis in sesame are largely unknown. The regulation of monolignol biosynthesis is a complex molecular mechanism that involves diverse transcription factors and phytohormones, among which NST1/2 detain pivotal roles [33–35]. NST1 and NST2 are master switch regulators of secondary cell wall formation through interaction with diverse family genes, mostly Myb transcription factors [33–35]. Therefore, deciphering the complex regulatory network of lignan biosynthesis in sesame is of particular interest for biotechnological applications.

Whole-genome resequencing and population genomic analyses have considerably enhanced the understanding of crops' genetic variation and domestication history and facilitated marker-traits association studies [36–38]. In the present study, we resequenced a worldwide collection of sesame, including 316 accessions from various geographical regions in China and 94 accessions from twenty-eight countries. We investigated the population structure and genetic diversity and revealed key candidate genes under selective sweeps in subpopulations. In addition, the master regulatory gene of lignan and lignin biosynthesis was discovered based on a genome-wide association study (GWAS) and gene expression analyses and validated in sesame hairy roots.

2. Results

2.1. Genome-Wide Variation and Population Structure

We resequenced 410 sesame germplasm samples, including 316 collected from southern, middle, and northern China and 94 collected from 28 countries worldwide representing various geographical regions (Table S1). The population included 329 landraces and 81 modern cultivars. The resequencing generated a total of 3.533 Tb of raw data, with an average of 8.617 Mb per accession (Table S2). The average sequencing depth was $24.93\times$, and the genome coverage 96.08%. The GC content varied from 36.41 to 40.98% (Table S2). After mapping against the sesame reference genome, we obtained a final set of 5,385,583 single-

nucleotide polymorphisms (SNPs) and 1,163,197 insertion-deletion (InDels). The SNPs and InDels are distributed unequally on the 13 chromosomes of the sesame genome (Figure 1a). Figures of 70.45, 10.70, and 4.45% of the SNPs were located in intergenic, intronic, and exonic regions, respectively (Figure 1c). Among the SNPs identified in coding sequences, 45.58% were synonymous SNPs and 52.79% non-synonymous SNPs (Figure 1b). The non-synonymous to synonymous SNPs ratio and the transition/transversion (Ts/Tv) ratio were 1.15 and 1.907, respectively (Table S3).

The SNP-based phylogenetic analysis classified the 410 sesame accessions into three major monophyletic clades which correlated with their geographical distribution (Figure 1d; Table S1). Group I (61 accessions), II (92 accessions), and III (257 accessions) included all the accessions from southern, northern, and central regions, respectively. Therefore, they were labeled SC (southern cluster, located south of 24° N latitude), NC (northern cluster, located north of 32–34° N latitude), and MC (central or middle cluster, located between SC and NC), respectively, to further identify selective sweeps. The landraces and modern cultivars exhibited admixture and were not separated on the phylogenetic tree (Figure 1d; Table S1). The principal component analysis (Figure 1e) and population structure analysis with different K-means levels (Figures 1f and S1) results were identical to the phylogenetic analysis results (Figures 1f and S1).

2.2. Genetic Diversity in Sesame Genomes

To explore genetic divergence among the sesame subpopulations, we performed a pairwise population differentiation (F_{ST}) across the genome between each group (Figure 2a). The F_{ST} value ($F_{ST} = 0.068$) between MC and NC groups is the smallest. The F_{ST} value between the SC and MC groups ($F_{ST} = 0.156$) was not too much different from that between the SC and NC groups ($F_{ST} = 0.159$). These results are consistent with the population structure and indicate that the MC and NC group populations have the closest distance and might evolve from the southern subpopulation. We then examined the genetic diversity by computing each group's nucleotide diversity (π). As we expected, the genetic diversity of the SC subpopulation ($\pi = 1.62 \times 10^{-3}$) is higher than that of the MC group ($\pi = 1.07 \times 10^{-3}$) and the NC group ($\pi = 1.09 \times 10^{-3}$) (Figure 2a). In addition, we explored Tajima's D values in each group and found a decrease in the average value from southern to northern regions (Figure 2c). We further investigated the linkage disequilibrium (LD) based on r^2 in the different groups (Figure 2b and Table S4). We found that the distance, when dropped to half of the maximum value, is 89.3 kb for the entire sesame population, which is slightly higher than that reported previously [15]. Finally, we investigated the directions of the gene flow between the subpopulations. The results indicated a "south to middle and north migration model" (Figure 2d,e).

2.3. Genomic Selected Signals in MC and NC Groups

To identify potential candidate genes for exploitation in sesame improvement, we searched for candidate selective genomic regions in MC and NC subpopulations using SC as a background population and adopting the criteria of $\log_2(\pi \text{ ratio}) > 2.5$ and $F_{ST} > 0.45$ (Figure 3a and Figure S3a). We identified 120 and 75 putative selective sweep regions comprising 968 and 571 genes for MC and NC groups, respectively (Tables S5 and S6).

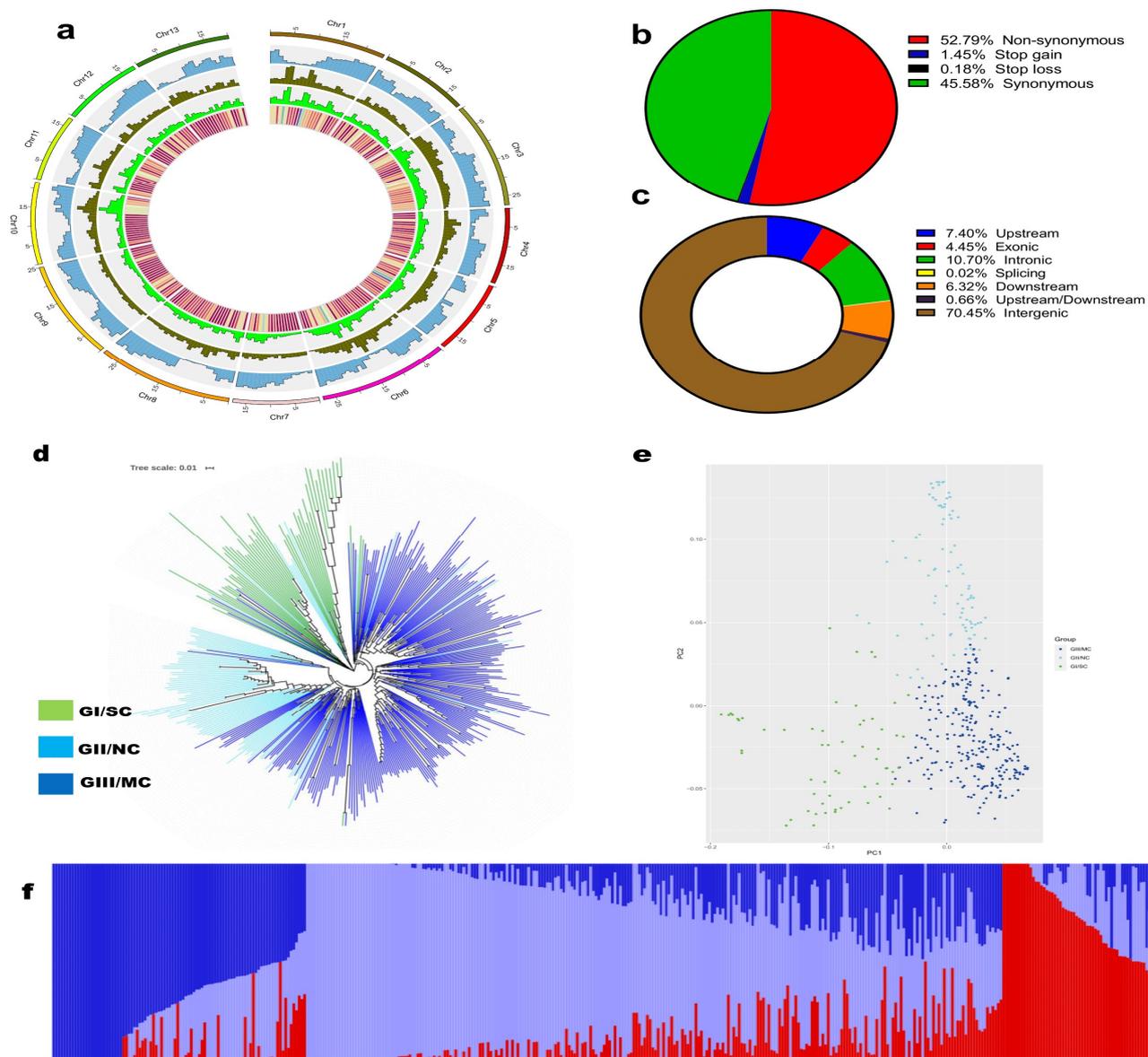


Figure 1. Population structure of sesame accessions in this study. **(a)** Chromosomal distribution of the genomic variants. From the outer to the inner are chromosomes, gene density, SNP density, InDels density, and nucleotide diversity. **(b,c)** Classification of the SNPs. **(d)** Phylogenetic tree of all accessions inferred from whole-genome SNPs filtered by linkage disequilibrium (LD) $r^2 < 0.05$; The three major groups (Southern cluster, SC; Northern cluster, NC; and middle or central cluster MC) are colored in purple, light blue, and blue, respectively. **(e)** Principal component analysis (PCA) plot for all the sesame accessions; dot colors correspond to the phylogenetic tree grouping. **(f)** Population stratification based on STRUCTURE for $K = 3$. Each color represents one ancestral population. Each accession is represented by a bar, and the length of each colored segment in the bar represents the proportion contributed by that ancestral population.

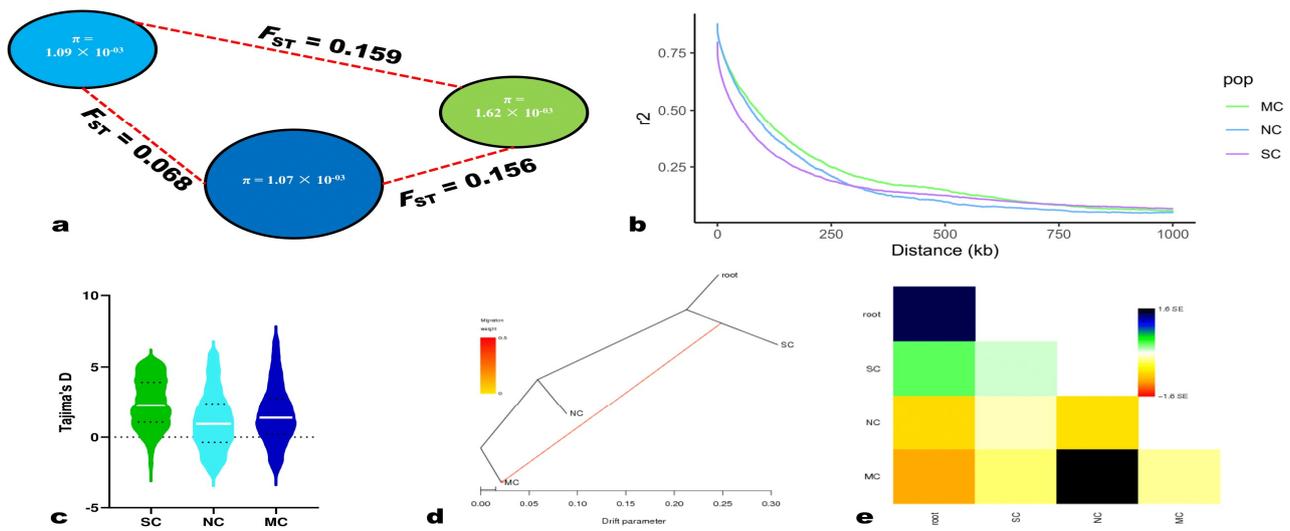


Figure 2. Genetic diversity and gene flow among the subpopulations. (a) The diversity (π) and genetic distance (F_{ST}) across the groups, where green, light blue, and blue colors indicate SC, NC, and MC, respectively; the radius of pie: genetic diversity; and dashed line length: F_{ST} value between two groups. SC, NC, and MC represent the subpopulations from the genetic diversity analysis. (b) Linkage disequilibrium (LD) differences between subpopulations. (c) Distribution of Tajima’s D values. The white line indicates the mean. (d) Most frequently found maximum-likelihood trees for gene flow among the subpopulations by Treemix. (e) The residual fit values obtained from the maximum likelihood tree (d). The plot was constructed using the normalized residual covariances by dividing the residual covariance value between each pair of populations by the standard deviation between all sample pairs. On the right is the color ruler.

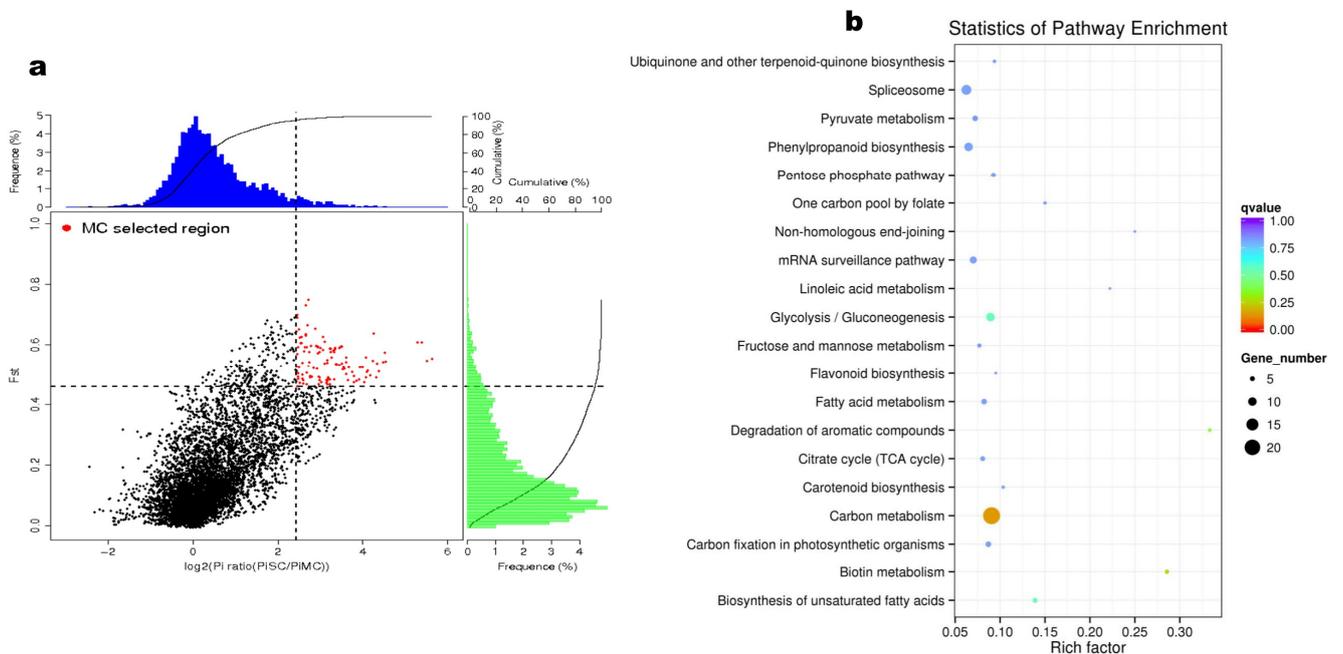


Figure 3. Selective sweeps and functional characterization of positively selected genes in MC group. (a) Selection signatures in MC subpopulation. Each red dot indicates a positive selective genomic region. (b) KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis and enrichment results of genes detected under selective genomic regions in MC subpopulation. The color ruler and p -values are shown on the right.

Gene ontology (GO) enrichment analysis for genes selected in the MC subpopulation were mostly related to “structural molecule activity” (GO:0005198), “iron-sulfur cluster binding” (GO:0051536), “metal cluster binding” (GO:0051540), and “protein folding” (GO:0006457) (Figure S2). For the selected genes in NC, the GO terms “transferase activity, transferring phosphorus-containing groups” (GO:0016772), “hydrolase activity, acting on glycosyl bonds” (GO:0016798), “hydrolase activity, hydrolyzing O-glycosyl compounds” (GO:0004553), and “carbohydrate metabolic process” (GO:0005975) were predominant (Figure S3b). KEGG enrichment analysis assigned the MC-selected genes mainly to carbon metabolism, glycolysis/gluconeogenesis, and phenylpropanoid biosynthesis (Figure 3b). Meanwhile, the NC-selected genes were primarily assigned to ribosome, carbon metabolism, glycolysis/gluconeogenesis, phenylpropanoid biosynthesis, and purine and pyruvate metabolism (Figure S3c). We then constructed a Venn diagram among the two sets of selected genes and identified 184 genes (Figure S4). These genes encode diverse enzymes essential for plant development (Table S7).

2.4. Genome-Wide Associations for Lignans Variation in Sesame

Sesame contains two major and clinically important lignans, sesamin and sesamol. To dissect the genetic basis of lignans variation in sesame, we cultivated all the 410 accessions in six different environments in Guangxi (GX, 2014), Wuhan (WH, 2014 and 2018), Hunan (HN, 2014), Shandong (SD, 2018), and Zhumadian (ZM, 2019). Seed samples from each location were analyzed by HPLC (high-performance liquid chromatography). The summary of the phenotypic data is presented in Table S8. In general, the sesamin and sesamol contents varied from 0.06 to 10.65 mg/g and 0.05 to 4.77 mg/g, respectively. The coefficient of variation of sesamin ranged from 33.92 to 45.18%, while that of sesamol ranged from 21.41 to 28.88%. The frequency distribution of the two traits and their correlations across the six environments are shown in Figure S5. As sesamin and sesamol contents showed considerable variation, we fitted their values in the six environments to obtain a BLUP (best linear unbiased predictor) value, using a linear mixed model in the lme4 R package. To examine the influence of genetic variation on sesamin and sesamol contents, we compared their respective values in the subpopulations (Figure 4a,b). We found no significant difference in the sesamol content of the three groups (Figure 4b). However, the sesamin content in MC and SC was significantly ($p < 0.001$) higher than in NC (Figure 4a). We further investigated the population differentiation (F_{ST}) between high-sesamin (HS, value > 6 mg/g) and low-sesamin (LS, value < 1 mg/g) accessions. The F_{ST} analysis revealed significant divergent genomic regions comprising 2119 genes (Figure 4c; Tables S9 and S10). These genes were mainly related to metabolic processes, especially oxidoreductase activities and cellular processes (Table S10 and Figure S6). We searched for candidate selective regions for HS and LS content. We identified 41 and 65 putative selective genomic regions harboring 363 and 343 genes for HS and LS content, respectively (Tables S11–S14). KEGG annotation and enrichment analysis assigned the HS accessions selective genes mainly to plant-pathogen interaction, peroxisome, pentose and glucuronate interconversions, mRNA surveillance pathway, glycerolipid metabolism, and carbon metabolism (Figure S7a). While the LS accessions selective genes were principally assigned to metabolic pathways and biosynthesis of secondary metabolites (Figure S7b).

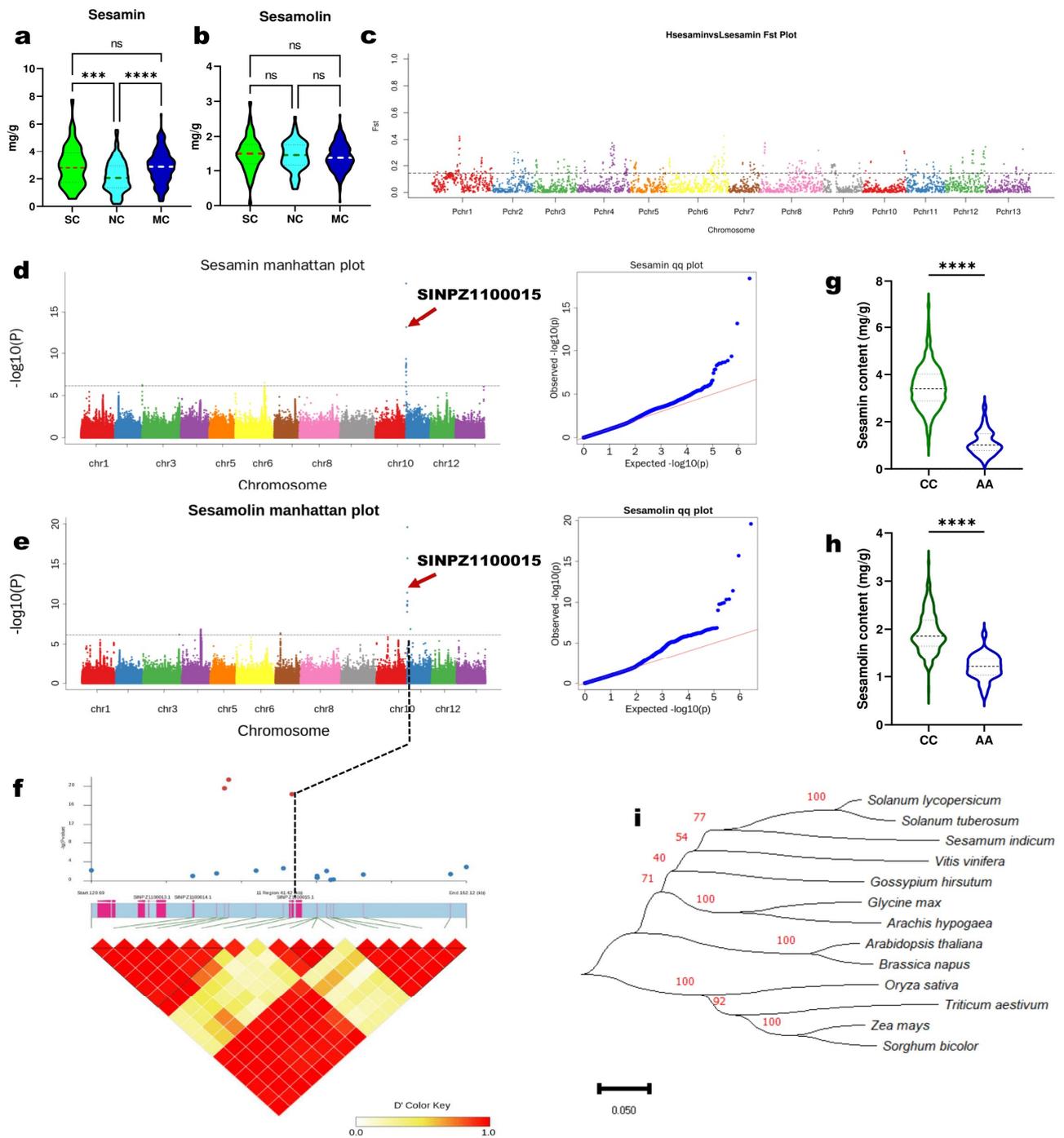


Figure 4. Genome-wide association for sesamin and sesamolins variation. (a,b) Boxplot of sesamin and sesamolins contents variation among the subpopulations, respectively. MC, NC, and MC indicate the different subpopulations. (c) Candidate selective genomic regions in high-sesamin content sesame accessions. F_{ST} stands for genetic differentiation. (d,e) Manhattan and QQ plots for GWAS analysis for sesamin and sesamolins, respectively. The dashed blue line indicates the threshold $-\log_{10}(p) = 6$. (f) LD (linkage disequilibrium) heatmap around *SiNST1* (*SINPZ1100015*). The dotted line from subfigures e and d to f show the position of *SiNST1* on the sesame genome. (g,h) Mining of favorable allele for sesamin and sesamolins contents, respectively; accessions with “C” allele, 362 and “A” allele, 34. (i) Phylogenetic analysis of *SiNST1* and *NST1* from other species. *** $p < 0.001$; **** $p < 0.0001$, *t*-test.

We performed a GWAS on the two traits via the MLM (mixed linear model) implemented in the EMMAX (Efficient Mixed-Model Association eXpedited) software. The results showed that the model was adequate for analyzing these important traits (Figure 4d,e). In total, 17 and 72 significant SNP loci (threshold $-\log(p)$ value > 6) were identified for sesamin and sesamolin, respectively (Table S15). Interestingly, most of the loci locate at our previously detected QTL positions of the two lignans [39]. The major association signal (Chr11-142842) was pleiotropic for the two traits (Figure 4d,e) and is located in the exon of the gene *SINPZ1100015* (*SiNST1/SIN_1005755*), which encodes an NAC domain-containing protein 43. It is homologous to the *Arabidopsis* NST1, which is a master regulator of secondary cell wall deposition through the control of monolignol biosynthesis [33–35]. *SiNST1* has shown significant associations with variations in oil, seed coat thickness, and phytosterol contents in sesame [15,40]. In addition, we screened for candidate genes within the LD windows of the other associated loci. By combining gene function annotation and expression patterns in two contrasting sesamin content varieties, we selected ten additional candidate causative genes for lignans variation in sesame (Figure S8 and Table S16).

Interestingly, the locus of *SiNST1* exhibited significant divergence between HS and LS sesame accessions (Figure 4c and Table S10). The results of the regional association (LD) analysis of *SiNST1* ± 10 kb corroborate with the GWAS analysis (Figure 4f). *SiNST1* is primarily expressed in sesame seeds and capsules (Figure S9A). The non-synonymous SNP locus (Chr11-142842) consists of “C” allele change into “A” conducting amino acid change from Thr (threonine) to Lys (lysine). Superior allele analysis showed that the common allele “C” significantly increased ($p < 0.0001$) the sesamin and sesamolin contents than the variant allele “A” (Figure 4g,h). qRT-PCR analysis also showed that the “C” allele is favorable for lignans biosynthesis during seed development (Figure S9B). In sesame accessions with the “C” allele, *SiNST1* is highly expressed at the key stages (10 to 20 DPA) of sesamin and sesamolin biosynthesis [12,30,41]. While in sesame accessions with the “A” allele, *SiNST1* is mostly expressed at the early stages of seed development. Further, we searched for the coding sequence (CDS) of homologs of *SiNST1* in other crops via NCBI. The CDS of *SiNST1* consists of 376 amino acids (aa). Phylogenetic analysis clearly separated NST1 from monocots and dicots. On the phylogenetic tree, *SiNST1* clustered closely to NST1 in *Solanum tuberosum* and *S. lycopersicum* (Figure 4i).

2.5. *SiNST1* Is a Master Regulator of Lignans and Lignin Biosynthesis in Sesame

Since NST1 is reported as a master regulator of lignans and lignin biosynthesis in plants by interacting with hormonal signals and MYB transcription factors, we explored the cis-acting elements in the promoter region of *SiNST1* and *PSS* (piperitol/sesamin synthase). The results showed that *SiNST1* contains mainly MYB binding, light responsiveness, and hormone-related elements (Table S17). *PSS* primarily contains MYB recognition elements and MeJA (methyl jasmonate) responsive elements (Table S18). We further performed a WGCNA (weighted gene correlation analysis) using the expression values (FPKM) of *SiNST1*, the 287 sesame MYB genes (*SiMYBs*) [42], and the 25 predicted CYP81Q1 genes [12], including *PSS*. As shown in Figure 5a, these genes were classified into four subgroups: blue; brown; grey; and turquoise modules. The blue module exhibited the highest positive correlation ($r = 0.31$) with sesamin and included *SiNST1* (Figure 5b). *SiNST1* appeared as the hub gene in this module and showed co-expression relationships with 85 *SiMYBs* and four CYP81Q1-related genes (Table S19). We found no direct interaction between *SiNST1* and *PSS*. *PSS* was classified into the turquoise module, which also exhibited a positive correlation ($r = 0.19$) with sesamin content (Figure 5b). The top ten genes in this module included seven MYBs and three CYP81Q1 genes (Table S20).

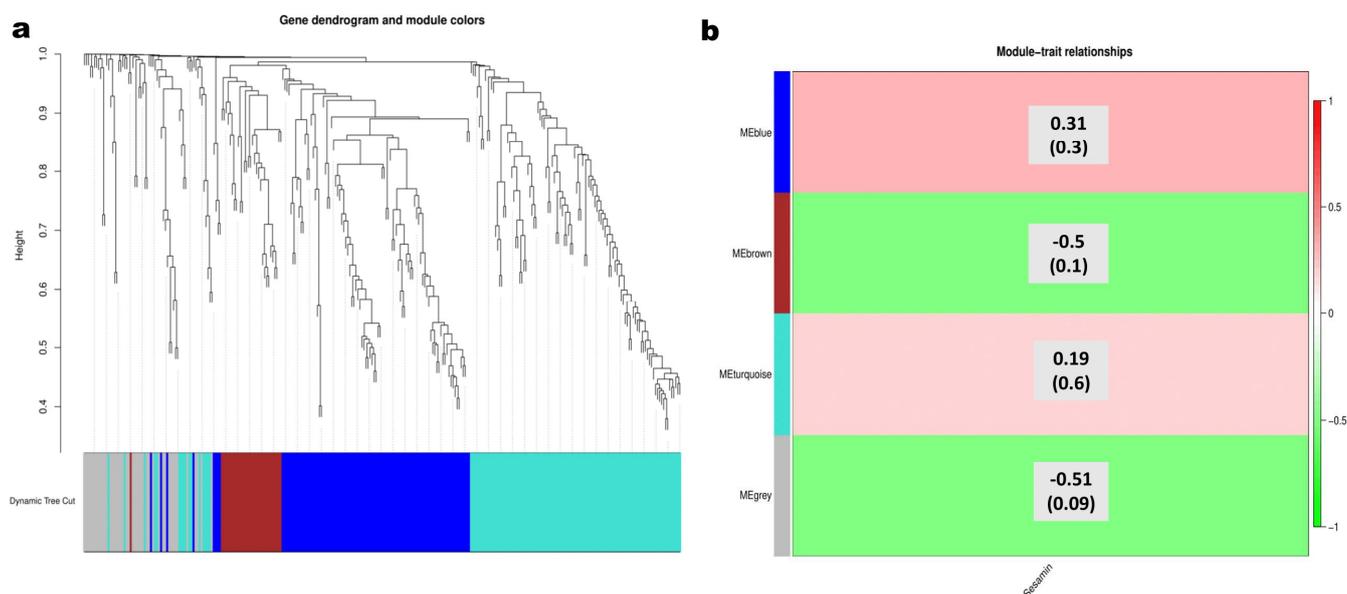


Figure 5. WGCNA results of *SiNST1*, *SiMYBs*, and predicted CYP81Q1-related genes in developing sesame seeds. **(a)** Classification of the analyzed genes in four modules represented by the blue, brown, grey, and turquoise colors. The genes with higher weight values have the most extended height on the dendrogram. **(b)** Heatmap of the correlation of WGCNA modules with sesamin content. The ordinate and abscissa indicate modules and the trait (sesamin), respectively. The green or red colors indicate the correlation coefficients (the top number in each cell). On the right is the color ruler. The bottom numbers (numbers in brackets) indicate the *p*-values.

To validate the potential function of *SiNST1*, we overexpressed *SiNST1^C* in sesame hairy roots (Figure S10). The *SiNST1^C*-overexpressing (OE) sesame hairy roots exhibited a significant increase ($p < 0.01$) in the content of sesamin and sesamol (Figure 6e). As lignans and lignin share the same biosynthetic pathway, we also evaluated the lignin content in the sesame hairy roots. The results indicated a slight increase in the lignin content of OE compared to the control, but the difference was not significant (Figure 6d). To acquire insights into the regulatory mechanism of *SiNST1*, we investigated via qRT-PCR the expression levels of sesame monolignol biosynthetic genes, co-expressed *SiMYBs* predicted to involve lignin or phenylpropanoid biosynthesis, and *PSS*. The results showed that the *SiNST1* overexpression significantly influenced the expression of monolignol biosynthetic genes (Figure 6a). The expression levels of *SiPAL* (phenylalanine ammonia-lyase), *SiC4H* (cinnamate 4-hydroxylase), *SiC3H* (coumarate 3-hydroxylase), *SiCCR* (cinnamoyl-CoA reductase), *SiCOMT* (caffeic acid 3-O-methyltransferase), and *SiCCaAOMT* (caffeoyl-CoA O-methyltransferase), were significantly increased. Among the investigated MYB genes, *SiMYB58*, *SiMYB209*, *SiMYB134*, and *SiMYB276* exhibited a significant increase in their expression levels in OE (Figure 6b). The expression level of *PSS* in OE was also significantly increased (Figure 6c).

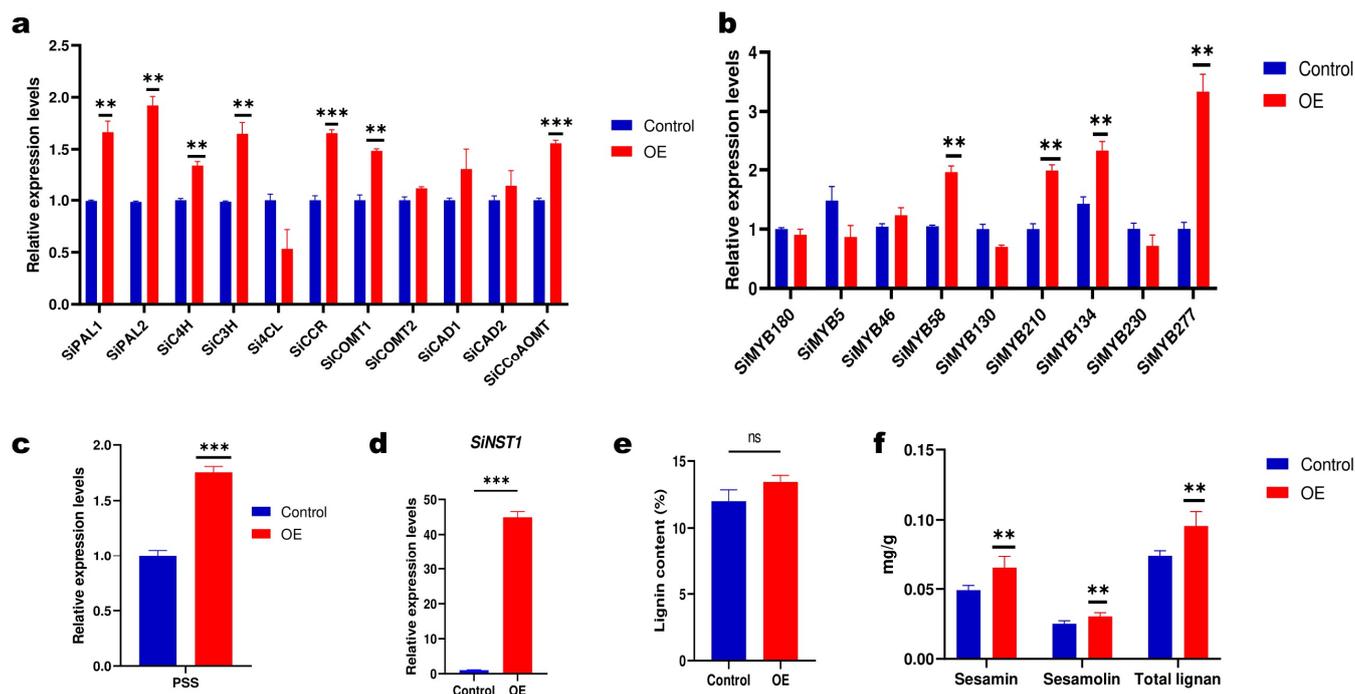


Figure 6. Functional validation of *SiNST1* in sesame hairy roots. (a) Expression analysis of monolignol biosynthetic genes. (b) Co-expressed *SiMYBs* with *SiNST1* predicted to involve in lignin biosynthesis. (c) piperitol/sesamin synthase, *PSS*. (d) *SiNST1*. (e) Lignin content. (f) Lignans content in transgenic sesame hairy roots. *** $p < 0.001$; ** $p < 0.01$; t -test.

3. Discussion

We provide comprehensive large-scale genome-wide polymorphism data in sesame, one of the most ancient oilseed crops with high economic, cultural, nutritional, and therapeutical values worldwide, especially in East Asia. The 410 sesame accessions resequencing data provide insights into the population structure and high throughput markers (5.38 and 1.16 million SNPs and InDels, respectively) for genotype selection and association studies to unveil key regulatory genes and dissect complex molecular mechanisms underlying agronomical and quality traits. The non-synonymous to synonymous SNPs ratio was 1.15, which is smaller than in soybean (*Glycine max* L., 1.61) [38], castor (*Ricinus communis* L., 1.39) [43], Tartary buckwheat (*Fagopyrum tataricum* L., 1.69) [44], and chickpea (*Cicer arietinum* L., 1.20) [45]. The transition/transversion (Ts/Tv) ratio of 1.907 for sesame is smaller than that for Tartary buckwheat (2.175) [44] but higher than that for maize (*Zea mays* L., 1.02) [46], tomato (*Solanum lycopersicum* L., 1.75) [47], and black gram (*Vigna mungo* L., 1.58) [48].

The domestication and evolution histories of sesame are unclear and are expected to be explored after the whole-genome sequencing of wild sesame plants [1]. Here, our investigations (phylogenetic and PCA analyses) show that sesame accessions were geographically different and exhibited a pattern of SC, MC, and NC. The landraces and modern cultivars exhibited admixture, which is consistent with the previous report by Wei et al. [15]. The population differentiation (F_{ST}), genetic diversity, and gene flow analyses indicated that the SC subpopulation is the oldest sesame population, and the MC and NC clades emanated from the SC group. The genetic distance (F_{ST}) between NC and MC groups was very low compared to their respective genetic distance with the SC group. The MC and NC subpopulations have relatively low genetic diversity (π) compared with the SC group. These results suggest that the SC subpopulation might be relatively close to wild sesame species. Moreover, the variation in genetic diversity suggests different intensities of genes selection during geographical differentiation. We identified 184 common genes under significant selective genomic regions in MC and NC groups. These genes encode diverse enzymes essential for plant development and may represent target genes for improving

sesame quality traits and environmental adaptation. However, these potential environmental adaptation-related genome markers and candidate genes need to be confirmed via a Genome-Environment Association (GEA) and subsequently validated through functional genomics studies [49,50]. Otherwise, as the studied population contained majority accessions from China, our findings could imply that sesame introduction and domestication in China commenced in a locality in southern China and later extended to middle and northern China. Supportively, similar domestication events have been reported in soybean and Chinese castor [38,43]. This extensive genomic variation and substantiation of genetic differentiation amid regional subpopulations will favorize structured association mapping and the design of breeding programs that will seize the entire diversity in sesame [51–53].

Sesame lignans have become compounds of considerable scientific research interest due to their biological properties and huge importance [54–57]. Accordingly, the main current objective in sesame breeding is to create elite varieties containing high oil and lignans. The genetic variants governing natural variation in oil content and composition have been explored [15,58,59]. Selective sweeps analysis identified 363 positive selected genes in high-sesamin content sesame accessions. Functional analysis assigned these genes mostly in plant-pathogen interaction, peroxisome, pentose and glucuronate interconversions, mRNA surveillance pathway, glycerolipid metabolism, and carbon metabolism. These findings indicate that sesamin and sesamolins might be primarily involved in sesame biotic stress tolerance mechanisms. We found that the sesamin and sesamolins contents varied considerably among the six environments, and the content of sesamin significantly decreased from SC to NC. These results show genetic selection, domestication, and environmental conditions have significant impacts on sesame agronomic traits. We mainly focused on association analysis for sesamin and sesamolins variation and identified significantly associated loci and major candidate effect genes. The genetic polymorphism within the exonic region of the gene *SiNST1* may contribute substantially to lignan variation in sesame. Favorable allele mining provides important resources for marker-assisted selection. *SiNST1* encodes for an NAC domain-containing protein 43 and was positively selected in the high-sesamin content sesame accessions. Its homolog *AtNST1* in *Arabidopsis* is a master switches transcriptional regulator of SCW (secondary cell wall) formation by interacting with some MYB genes and hormonal signals [33–35]. Phylogenetic analysis clearly separated *NST1* from monocots and dicots, indicating the gene might play different functions in the two categories of Angiosperm. On the phylogenetic tree, *SiNST1* clustered within dicots closely to *NST1* in *Solanum tuberosum* and *S. lycopersicum*, indicating it might fulfill similar functions in sesame. *Cis*-acting elements analysis revealed that the promoter region of *SiNST1* mainly contains MYB binding, light responsiveness, and hormone-recognizing elements. These results suggest that *SiNST1* might regulate sesamin and sesamolins biosynthesis through interplays with MYB genes and phytohormones as per its homologous. Consistently, WGCNA revealed that *SiNST1* might interact with diverse *SiMYBs* in developing sesame seeds, confirming it might regulate diverse developmental processes in sesame. However, no direct interaction between *SiNST1* and *PSS* was noticed. *SiNST1* is also significantly associated with seed coat thickness, phytosterol content, and oil content in sesame [15,40]. The non-synonymous SNP locus in *SiNST1* consists of “C” allele change into “A” resulting in an amino acid change from Thr (threonine) to Lys (lysine). Sesame accession with the “C” allele accumulated significantly higher sesamin and sesamolins compared to accessions with the “A” allele. Supportively, gene expression analysis showed that *SiNST1^C* is mainly expressed during lignan biosynthesis stages of 10 to 20 DPA, while *SiNST1^A* is primarily expressed at seed coat development stages (early stages) [12,15,30]. These findings show that *SiNST1^C* is a potential target for improving nutrient content in sesame, while *SiNST1^A* may favorize seed coat thickness [15]. Together, our findings suggest that the thicker the sesame seed, the less oil and lignan may it contain. However, it should be noted that only 35 accessions with the “A” allele were contained in the studied population. Therefore, a future study is required to analyze the impact of nucleotide variability of *SiNST1* on sesame seed quality traits.

Since an efficient genetic transformation method for the sesame plant is not yet established, we overexpressed *SiNST1^C* in sesame hairy roots to investigate its regulatory mechanisms. The overexpression of *SiNST1^C* significantly induced the expression of *SiMYB58*, *SiMYB209*, *SiMYB134*, *SiMYB276*, and most monolignol biosynthetic genes. Consequently, the *SiNST1^C*-overexpressing sesame hairy roots exhibited an increase in lignan and lignin contents. However, the increase in the lignin content was not significant. These findings indicate that *SiNST1* plays a master switches function in regulating SCW formation and lignan biosynthesis in sesame by interacting with these four *SiMYBs*. Moreover, we identified ten other candidate genes for sesamin and sesamol variation and seven *SiMYBs* that showed significant co-expression correlation with PSS in the turquoise module. These seven *SiMYBs* might be the key regulator of PSS expression. Therefore, further functional characterization of *SiNST1* and these genes in developing sesame seeds via subsequent knockout and overexpression is required when a genetic transformation protocol will be established to dissect this complex molecular regulatory network for biotechnological applications. Lignans and lignin are principally biosynthesized from the oxidative coupling of E-coniferyl alcohol, a critical reaction catalyzed by dirigent proteins (DIRs) [26,28,60,61]. Therefore, genome-wide characterization of the DIR gene family and identification of *SiDIRs* that may promote lignan accumulation in sesame are required for targeted balancing of the oxidation coupling in favor of lignan biosynthesis. In addition, identifying key quality-related candidate genes in wild sesame accessions will offer supplementary resources for sesame improvement. For instance, Wang et al. have found that the gene CYP92B14 promotes high sesamol accumulation in wild allotetraploid sesame (*S. schinzianum*) [62]. This gene has to be functionally characterized for exploitation in genomic-assisted breeding of high sesamin and sesamol content sesame varieties.

4. Materials and Methods

4.1. Plant Materials and Genome Sequencing

The 410 sesame accessions (329 landraces and 81 modern cultivars) were given by the National Mid-term Gene Bank, Oil crops Research Institute, Chinese Academy of Agricultural Sciences (OCRI, CAAS), Wuhan, China. The population is composed of 316 accessions from diverse geographical regions in China and 94 accessions from twenty-eight other countries (Table S1). Related to the type of variety, there were 329 and 81 landraces (local varieties) and modern cultivars (improved via breeding), respectively. Each accession sample was maintained via cultivation and self-pollination for over three generations prior to sequencing and phenotypes evaluation. A representative plant of each accession was chosen for the genome sequencing. Genomic DNA was extracted from leaves at the seedling stage via the CTAB method [63] and following the manufacturer's instructions. The sequencing library was constructed from high-quality DNA with an insert size of ~350 bp. The paired-end sequencing of each accession library was carried out on the Illumina HiSeqTM/MiSeqTM platform (Novogene Co., Ltd., Beijing, China).

4.2. Sequence Alignment, Variants Detection, and SNP Annotation

All the paired-end reads were mapped onto the sesame reference genome (version 2.0) [16] using the BWA (Burrows-Wheeler Aligner) software [64] with default parameters in order to locate the position of each read on the sesame genome. Next, we filtered out the low-quality (MQ < 20) reads with the SAMtools (v1.1) [65] program. We sorted and discarded PCR duplicates to construct a bam file index using Picard Tools (<http://broadinstitute.github.io/picard/>; v1.118, accessed on 2 March 2021). Finally, we called the genome variants (SNPs and InDels) with GATK (Genome Analysis Toolkit) [66], as described by Fan et al. [43]. To evaluate the quality of the identified SNPs, we randomly selected fifteen sesame accessions and repeated the whole-genome sequencing. The consistency of the results was approximately 99.8%.

The SNP annotation was conducted according to the sesame reference genome with ANNOVAR [67]. First, the SNPs were classified into upstream or downstream regions,

intronic, exonic, intergenic, and splicing sites. Further, SNPs located in the exonic areas were grouped into synonymous and non-synonymous SNPs.

4.3. Population Genetic Analyses

The effective, high-quality SNPs were sorted using the criteria of Min depth $\geq 8x$, MAF (Minor allele frequency) ≥ 0.05 , and missing rate ≤ 0.2 . Phylogenetic, PCA, structure, and LD analyses allow the examination of the evolutionary history or relationship between individuals of a population and characterize the population stratification and genetic variation. We constructed the approximate maximum-likelihood tree using the general time-reversible model in the TreeBeST (<http://treesoft.sourceforge.net/treebest.shtml>, accessed on 12 June 2021) software [68]. The population structure analysis was conducted in the Admixture program [69] with various K values. PCA (principal component analysis) was performed in the GCTA software [70]. The dataset used to conduct the PCA was screened for quality using the criteria of scaffold length >100 kb and (QUAL in the VCF file) <2000 . Linkage disequilibrium (LD) decay analysis was performed in the PopLDdecay software [71].

To gain insight into the genetic diversity of the sesame population, we computed the genetic distance (F_{ST}) and nucleotide diversity (π) using VCFtools [72] with a 10 kb nonoverlapping window slid across each scaffold. The F_{ST} between clades was evaluated using the average value of all 10 kb sliding windows. The selective sweeps of geographical differentiation of SC subpopulation against other groups were revealed by considering both the F_{ST} value ($F_{ST} > 0.45$) and the π ratio ($\log_2(\pi \text{ ratio}) > 2.5$) in each sliding window, with a 5% cutoff [43]. The intergroup gene flow analysis was performed in Treemix [73], using a composite-likelihood approach. We defined the root via the root parameter and tested for many migration events to select the most frequent ones with significant ($p < 0.05$) gene flow.

4.4. Gene Ontology (GO) and KEGG Analyses

The gene annotation was conducted with PFAM terms using InterProScan 5 (<http://www.ebi.ac.uk/Tools/pfa/iprscan5/>, accessed on 19 September 2020) as per Hazzouri et al. [52]. Functional annotation and categorization of genes were achieved through gene ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analyses. The GO terms and the KEGG analyses were carried out using AutoFact [74] and the Automated Annotation Server (KAAS; <http://www.genome.jp/tools/kaas/>, accessed on 9 August 2021), respectively. The ggplot2 package in R was used to sort the significant terms (p -value < 0.05) [75].

4.5. Phenotyping

The two primary lignans in sesame, sesamin and sesamol, were investigated in six environmental conditions located in China, including Guangxi (GX, 2014), Wuhan (WH, 2014 and 2018), Hunan (HN, 2014), Shandong (SD, 2018), and Zhumadian (ZM, 2019). Wuhan belongs to the Yangtze river valley (southern China) with a warm and temperate climate. Zhumadian belongs to the Huang-Huai river valley (northern China) with a humid subtropical climate. Hunan and Guangxi are located at southern China with a subtropical humid monsoon and tropical monsoon climates, respectively. Shandong has a temperate monsoon climate. HPLC analysis of the sesamin and sesamol contents in seed samples was carried out as previously reported [39,76]. The phenotypic data were analyzed with GraphPad.prism V. 9.0.0121 (GraphPad Software Inc., La Jolla, CA, USA).

4.6. GWAS Analysis and Candidate Genes Identification

The association analysis was carried out with the EMMAX (Efficient Mixed-Model Association eXpedited) program via the mixed linear model (MLM) [77], in order to uncover genetic markers that drive lignan variation in sesame. The significant association threshold was set at $-\log_{10}p > 6$. The Manhattan and QQ plots were constructed in R using the package “qq-man” [78]. The candidate genes in LD windows (± 89 Kb) were selected

through the integration of gene function annotation, non-synonymous SNPs, and gene expression analysis. Finally, we performed pairwise LD correlation analysis to confirm the major significant associated locus. The pairwise LD correlations analysis of the major candidate causative gene was carried out with LDBLockShow software [79].

4.7. Cis-Acting Elements and Phylogeny Analyses

Cis-acting elements prediction in the 2000 bp upstream promoter region of the targeted genes was achieved with PLANTCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>, accessed on 21 November 2021), in order to predict their respective potential functions. The protein sequence of *SiNST1* homolog in *Glycine max* (XM_003548125), *Arabidopsis thaliana* (NM_130243), *Brassica napus* (XM_013895547), *Triticum aestivum* (XM_044585881), *Oryza sativa* (XM_015793488), *Sorghum bicolor* (XM_002436372), *Zea mays* (XM_008650556), *Arachis hypogaea* (XM_025776021), *Gossypium hirsutum* (XM_016853342), *Vitis vinifera* (XM_002279509), *Solanum lycopersicum* (XM_004248375), and *S. tuberosum* (XM_015310959) were downloaded by blast from NCBI. These species were selected to represent monocots and dicots and based on previous studies [33–35]. The alignment and the phylogenetic analysis were performed with the MEGA X software [80].

4.8. WGCNA Analysis

Weighted gene co-expression network analysis (WGCNA) is a system biology method widely used to reveal correlation patterns among genes and identify candidate genes for specific agronomic traits. The WGCNA was conducted using the expression values (FPKM) of *SiMYBs* [42], predicted CYP81Q1 genes [12], and *SiNST1*. The expression values were extracted from available RNA-seq data of two sesame accessions (Zhongzhi16 and ZZM2748) with contrasting lignans content. The co-expression analysis was performed with the WGCNA package in R (version 4.0.2) [81].

4.9. Hairy Roots Transformation and Culture

The coding sequence of *SiNST1* was cloned (primers: F-AGAAGAGGGCTGGGTGGTT and R-GGATGACGAGTTGGGGCTAT) from Zhongfengzhi1 (harboring the favorable allele “C” at the locus Chr11-142842) and inserted into the pCAMBIA 1301S vector. The recombinant vector pCAMBIA 1301S-*SiNST1* was transformed to *Escherichia coli* DH5 α (Weidi Biotechnology, Shanghai, China). After confirming that the transformation sequence was correct, a suitable amount of plasmid DNA was extracted from the previously transformed *E. coli* and transferred to *Agrobacterium rhizogenes* K599 for subsequent generating transgenic sesame hairy roots following previously described methods [82,83] with some modifications. Sesame sterile seedlings at two weeks old were wounded and used for the *Agrobacterium* infection for 10 min. Next, co-culture in the dark on MS solid medium for 48 h at 25 °C, followed by explants washing by MS liquid medium (containing 300 mg/mL cefotaxime) and sterile water. Then the explants were cultured on MS solid medium containing 200 mg/L timentin in a growth chamber to induce the hairy roots. The induced single sesame hairy root lines were discarded from explants after two weeks and grown on MS solid medium containing 200 mg/L timentin for detoxification. The positively transgenic lines were screened via PCR and grown on MS solid medium for two weeks under a 16/8 h light/dark photoperiod at 28 °C. Hairy roots were harvested, washed, and dried for sesamin, sesamol, and lignin evaluation.

4.10. Lignans and Lignin Contents Analyses in the Sesame Hairy Roots

The extraction method of sesamin and sesamol from the hairy roots was different. First, 0.3 g of crushed hairy roots was extracted with 5 mL absolute ethanol by shaking for 2 h. After centrifugation at 5000 \times g for 5 min, the supernatant was collected, and the ethanol was evaporated completely at 60 °C. Then, the residue was dissolved in 1 mL ethanol 80% by vortexing for 2 to 3 min. Finally, the extract was filtered and subjected to

HPLC analysis. The lignin (acid-soluble) content was evaluated as per Wu et al. via the Klason method [84].

4.11. qRT-PCR Analysis

Total RNA from the hairy root samples was extracted with the EASYspin Plus plant RNA Kit (Aidlab, Beijing, China) and, thereafter, reverse transcribed using the HiScript II first-strand cDNA synthesis Kit (Vazyme Biotechnology, Nanjing, China) following the manufacturer's instructions. The qRT-PCR was carried out with ChamQ™ SYBR1 qPCR Master Mix (Vazyme Biotech, Nanjing, China) on LightCycler480 (Roche, Switzerland) real-time PCR system. The histone H3.3 gene (*SINPZ1301428*) was used as an internal control to normalize the transcript levels [85]. Each gene was analyzed using three biological replicates under the same conditions. The expression levels of the genes were analyzed using the $2^{-\Delta\Delta CT}$ method [86]. The primers were designed with Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>), accessed on 18 February 2022) and are listed in Additional file 1: Table S21.

4.12. Statistical Analysis

GraphPad Prism v9.0.0121 (GraphPad 159 Software Inc., La Jolla, CA, USA) was used for statistical analysis and graphs' construction. Statistical differences were performed by independent *t*-test at $p < 0.05$.

5. Conclusions

Overall, this study upgraded knowledge of genetic polymorphism in sesame through whole-genome resequencing of 410 worldwide accessions. Population structure and genetic diversity analyses classified sesame population into different geographic patterns, consistent with reported potential domestication of the crop from the southern to the middle and northern regions. The key selected genes in MC and NC subgroups were identified for exploitation in sesame improvement. Genotype and environmental conditions have significant impacts on lignan variation in sesame. We identified key loci and 11 candidate genes governing the major sesame lignans (sesamin and sesamol) variation. Of them, *SiNST1* is the major underlying gene via a non-synonymous SNP (C/A). The "C" allele may favorize lignan and other quality traits accumulation, while the "A" may promote seed coat thickness. Functional characterization in sesame hairy roots revealed that *SiNST1* might function in synergetic action with *SiMYB58*, *SiMYB209*, *SiMYB134*, and *SiMYB276* to control monolignol biosynthetic gene expression at transcriptional levels. Our results may considerably contribute to sesame quality improvement via genomic-assisted breeding.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24021055/s1>.

Author Contributions: Conceptualization, L.W. and J.Y.; methodology, formal analysis and data curation, S.S. and S.S.K.D.; validation and investigation, A.L., R.Z., C.S. and Z.W.; resources and visualization, Y.Z., D.L. and M.B.; writing—original draft preparation, S.S.K.D.; writing—review and editing, L.W., J.Y. and S.S.K.D.; supervision, project administration and funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Agricultural Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2016-OCRI), the Key Research Projects of Hubei province (2020BBA045), the Science and Technology Innovation Project of Hubei province (2021-620-000-001-035), the Fundamental Research Funds for Central Non-profit Scientific Institution (Y2022XK11, 1610172022010), the Open Project of Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, P.R. China (KF2022002), and China Agriculture Research System (CARS-14).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw resequencing data reported in this paper have been deposited at the Genome Sequence Archive [87], National Genomics Data Center [88], China National Center for Bioinformatics/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA006452), publicly accessible at <https://ngdc.cnbc.ac.cn/gsa>, accessed on 22 November 2021. The RNA-seq data is available at the National Center for Biotechnology Information (NCBI), Sequence Read Archive (SRA), accession PRJNA806786. Other datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

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

References

1. Miao, H.; Zhang, H. *The Sesame Genome*; Miao, H., Zhang, H., Kole, C., Eds.; Compendium of Plant Genomes; Springer International Publishing: Cham, Switzerland, 2021; ISBN 978-3-319-98097-3.
2. Pathak, N.; Rai, A.K.; Kumari, R.; Bhat, K.V. Value addition in sesame: A perspective on bioactive components for enhancing utility and profitability. *Pharmacogn. Rev.* **2014**, *8*, 147–155. [[CrossRef](#)] [[PubMed](#)]
3. Dossou, S.S.K.; Xu, F.; You, J.; Zhou, R.; Li, D.; Wang, L. Widely targeted metabolome profiling of different colored sesame (*Sesamum indicum* L.) seeds provides new insight into their antioxidant activities. *Food Res. Int.* **2022**, *151*, 110850. [[CrossRef](#)] [[PubMed](#)]
4. Majdalawieh, A.F.; Massri, M.; Nasrallah, G.K. A comprehensive review on the anti-cancer properties and mechanisms of action of sesamin, a lignan in sesame seeds (*Sesamum indicum*). *Eur. J. Pharmacol.* **2017**, *815*, 512–521. [[CrossRef](#)] [[PubMed](#)]
5. Majdalawieh, A.F.; Dalibalta, S.; Yousef, S.M. Effects of sesamin on fatty acid and cholesterol metabolism, macrophage cholesterol homeostasis and serum lipid profile: A comprehensive review. *Eur. J. Pharmacol.* **2020**, *885*, 173417. [[CrossRef](#)]
6. Wu, M.S.; Aquino, L.B.B.; Barbaza, M.Y.U.; Hsieh, C.L.; De Castro-Cruz, K.A.; Yang, L.L.; Tsai, P.W. Anti-inflammatory and anticancer properties of bioactive compounds from *Sesamum indicum* L.—A review. *Molecules* **2019**, *24*, 4426. [[CrossRef](#)]
7. Deng, S.; Zhou, J.-L.; Fang, H.-S.; Nie, Z.-G.; Chen, S.; Peng, H. Sesamin Protects the Femoral Head from Osteonecrosis by Inhibiting ROS-Induced Osteoblast Apoptosis in Rat Model. *Front. Physiol.* **2018**, *9*, 1787. [[CrossRef](#)]
8. Ma, Z.P.; Zhang, Z.F.; Yang, Y.F.; Yang, Y. Sesamin promotes osteoblastic differentiation and protects rats from osteoporosis. *Med. Sci. Monit.* **2019**, *25*, 5312–5320. [[CrossRef](#)]
9. Zhao, Y.; Wang, Q.; Jia, M.; Fu, S.; Pan, J.; Chu, C.; Liu, X.; Liu, X.; Liu, Z. (+)-Sesamin attenuates chronic unpredictable mild stress-induced depressive-like behaviors and memory deficits via suppression of neuroinflammation. *J. Nutr. Biochem.* **2019**, *64*, 61–71. [[CrossRef](#)]
10. Bedigian, D. Evolution of sesame revisited: Domestication, diversity and prospects. *Genet. Resour. Crop Evol.* **2003**, *50*, 779–787.
11. Bedigian, D.; Seigler, D.S.; Harlan, J.R. Sesamin, sesamol and the origin of sesame. *Biochem. Syst. Ecol.* **1985**, *13*, 133–139. [[CrossRef](#)]
12. Wang, L.; Yu, S.; Tong, C.; Zhao, Y.; Liu, Y.; Song, C.; Zhang, Y.; Zhang, X.; Wang, Y.; Hua, W.; et al. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* **2014**, *15*, R39. [[CrossRef](#)] [[PubMed](#)]
13. Yu, J.; Golicz, A.A.; Lu, K.; Dossa, K.; Zhang, Y.; Chen, J.; Wang, L.; You, J.; Fan, D.; Edwards, D.; et al. Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. *Plant Biotechnol. J.* **2019**, *17*, 881–892. [[CrossRef](#)] [[PubMed](#)]
14. Qiu, Z.; Zhang, Y.; Bedigian, D.; Li, X.; Wang, C.; Jiang, H. Sesame Utilization in China: New Archaeobotanical Evidence from Xinjiang. *Econ. Bot.* **2012**, *66*, 255–263. [[CrossRef](#)]
15. Wei, X.; Liu, K.; Zhang, Y.; Feng, Q.; Wang, L.; Zhao, Y.; Li, D.; Zhao, Q.; Zhu, X.; Zhu, X.; et al. Genetic discovery for oil production and quality in sesame. *Nat. Commun.* **2015**, *6*, 8609. [[CrossRef](#)] [[PubMed](#)]
16. Wang, L.; Xia, Q.; Zhang, Y.; Zhu, X.; Zhu, X.; Li, D.; Ni, X.; Gao, Y.; Xiang, H.; Wei, X.; et al. Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. *BMC Genom.* **2016**, *17*, 31. [[CrossRef](#)]
17. Cui, C.; Liu, Y.; Liu, Y.; Cui, X.; Sun, Z.; Du, Z.; Wu, K.; Jiang, X.; Mei, H.; Zheng, Y. Genome-wide association study of seed coat color in sesame (*Sesamum indicum* L.). *PLoS ONE* **2021**, *16*, e0251526. [[CrossRef](#)]
18. Zhou, R.; Dossa, K.; Li, D.; Yu, J.; You, J.; Wei, X.; Zhang, X. Genome-wide association studies of 39 seed yield-related traits in sesame (*Sesamum indicum* L.). *Int. J. Mol. Sci.* **2018**, *19*, 2794. [[CrossRef](#)]
19. Wang, X.; You, J.; Liu, A.; Qi, X.; Li, D.; Zhao, Y.; Zhang, Y.; Zhang, L.; Zhang, X.; Li, P. Variation in Melatonin Contents and Genetic Dissection of Melatonin Biosynthesis in Sesame. *Plants* **2022**, *11*, 2005. [[CrossRef](#)]
20. Horii, S.; Ishii, T. Effect of Arbuscular Mycorrhizal Fungi and Their Partner Bacteria on the Growth of Sesame Plants and the Concentration of Sesamin in the Seeds. *Am. J. Plant Sci.* **2014**, *05*, 3066–3072. [[CrossRef](#)]
21. Radhakrishnan, R.; Pae, S.B.; Shim, K.B.; Baek, I.Y. *Penicillium* sp. mitigates Fusarium-induced biotic stress in sesame plants. *Biotechnol. Lett.* **2013**, *35*, 1073–1078. [[CrossRef](#)]
22. Radhakrishnan, R.; Kang, S.M.; Baek, I.Y.; Lee, I.J. Characterization of plant growth-promoting traits of *Penicillium* species against the effects of high soil salinity and root disease. *J. Plant Interact.* **2014**, *9*, 754–762. [[CrossRef](#)]

23. Liu, Z.; Zhu, H.; Zhou, J.; Jiang, S.; Wang, Y.; Kuang, J.; Ji, Q.; Peng, J.; Wang, J.; Gao, L.; et al. Resequencing of 296 cultivated and wild lotus accessions unravels its evolution and breeding history. *Plant J.* **2020**, *104*, 1673–1684. [[CrossRef](#)] [[PubMed](#)]
24. Wu, D.; Liang, Z.; Yan, T.; Xu, Y.; Xuan, L.; Tang, J.; Zhou, G.; Lohwasser, U.; Hua, S.; Wang, H.; et al. Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence. *Mol. Plant* **2019**, *12*, 30–43. [[CrossRef](#)] [[PubMed](#)]
25. Maynard Smith, J.; Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res. (Camb)* **2008**, *89*, 391–403. [[CrossRef](#)]
26. Hano, C.F.; Dinkova-Kostova, A.T.; Davin, L.B.; Cort, J.R.; Lewis, N.G. Lignans: Insights into Their Biosynthesis, Metabolic Engineering, Analytical Methods and Health Benefits. *Front. Plant Sci.* **2021**, *11*, 2020–2022. [[CrossRef](#)]
27. Markulin, L.; Drouet, S.; Corbin, C.; Decourtil, C.; Garros, L.; Renouard, S.; Lopez, T.; Mongelard, G.; Gutierrez, L.; Auguin, D.; et al. The control exerted by ABA on lignan biosynthesis in flax (*Linum usitatissimum* L.) is modulated by a Ca²⁺ signal transduction involving the calmodulin-like LuCML15b. *J. Plant Physiol.* **2019**, *236*, 74–87. [[CrossRef](#)]
28. Davin, L.B.; Lewis, N.G. An historical perspective on lignan biosynthesis: Monolignol, allylphenol and hydroxycinnamic acid coupling and downstream metabolism. *Phytochem. Rev.* **2003**, *2*, 257–288. [[CrossRef](#)]
29. Andargie, M.; Vinas, M.; Rathgeb, A.; Möller, E.; Karlovsky, P. Lignans of sesame (*Sesamum indicum* L.): A comprehensive review. *Molecules* **2021**, *26*, 883. [[CrossRef](#)]
30. Ono, E.; Nakai, M.; Fukui, Y.; Tomimori, N.; Fukuchi-Mizutani, M.; Saito, M.; Satake, H.; Tanaka, T.; Katsuta, M.; Umezawa, T.; et al. Formation of two methylenedioxy bridges by a *Sesamum* CYP81Q protein yielding a furofuran lignan, (+)-sesamin. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10116–10121. [[CrossRef](#)]
31. Murata, J.; Ono, E.; Yoroizuka, S.; Toyonaga, H.; Shiraiishi, A.; Mori, S.; Tera, M.; Azuma, T.; Nagano, A.J.; Nakayasu, M.; et al. Oxidative rearrangement of (+)-sesamin by CYP92B14 co-generates twin dietary lignans in sesame. *Nat. Commun.* **2017**, *8*, 2155. [[CrossRef](#)]
32. Zoclanclounon, Y.A.B.; Rostás, M.; Chung, N.J.; Mo, Y.; Karlovsky, P.; Dossa, K. Characterization of Peroxidase and Laccase Gene Families and In Silico Identification of Potential Genes Involved in Upstream Steps of Lignan Formation in Sesame. *Life* **2022**, *12*, 1200. [[CrossRef](#)] [[PubMed](#)]
33. Mitsuda, N.; Ohme-Takagi, M. NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. *Plant J.* **2008**, *56*, 768–778. [[CrossRef](#)] [[PubMed](#)]
34. Zhong, R.; Ye, Z.H. The *Arabidopsis* NAC transcription factor NST2 functions together with SND1 and NST1 to regulate secondary wall biosynthesis in fibers of inflorescence stems. *Plant Signal. Behav.* **2015**, *10*, 37–41. [[CrossRef](#)]
35. Zhong, R.; Richardson, E.A.; Ye, Z.H. Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. *Planta* **2007**, *225*, 1603–1611. [[CrossRef](#)]
36. Gouin, A.; Legeai, F.; Nouhaud, P.; Whibley, A.; Simon, J.C.; Lemaitre, C. Whole-genome re-sequencing of non-model organisms: Lessons from unmapped reads. *Heredity* **2015**, *114*, 494–501. [[CrossRef](#)] [[PubMed](#)]
37. Martin, J.; Schackwitz, W.; Lipzen, A. Genomic sequence variation analysis by resequencing. *Methods Mol. Biol.* **2018**, *1775*, 229–239. [[CrossRef](#)]
38. Zhou, Z.; Jiang, Y.; Wang, Z.; Gou, Z.; Lyu, J.; Li, W.; Yu, Y.; Shu, L.; Zhao, Y.; Ma, Y.; et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **2015**, *33*, 408–414. [[CrossRef](#)]
39. Xu, F.; Zhou, R.; Dossou, S.S.K.; Song, S.; Wang, L. Fine mapping of a major pleiotropic qtl associated with sesamin and sesamol variation in sesame (*Sesamum indicum* L.). *Plants* **2021**, *10*, 1343. [[CrossRef](#)]
40. Wang, Z.; Zhou, Q.; Dossou, S.S.K.; Zhou, R.; Zhao, Y.; Zhou, W.; Zhang, Y.; Li, D.; You, J.; Wang, L. Genome-Wide Association Study Uncovers Loci and Candidate Genes Underlying Phytosterol Variation in Sesame (*Sesamum indicum* L.). *Agriculture* **2022**, *12*, 392. [[CrossRef](#)]
41. Dossou, S.S.K.; Xu, F.; Dossa, K.; Zhou, R.; Zhao, Y.; Wang, L. Antioxidant lignans sesamin and sesamol in sesame (*Sesamum indicum* L.): A comprehensive review and future prospects. *J. Integr. Agric.* **2021**. *in press*.
42. Mmadi, M.A.; Dossa, K.; Wang, L.; Zhou, R.; Wang, Y.; Cisse, N.; Sy, M.O.; Zhang, X. Functional characterization of the versatile MYB gene family uncovered their important roles in plant development and responses to drought and waterlogging in sesame. *Genes* **2017**, *8*, 362. [[CrossRef](#)]
43. Fan, W.; Lu, J.; Pan, C.; Tan, M.; Lin, Q.; Liu, W.; Li, D.; Wang, L.; Hu, L.; Wang, L.; et al. Sequencing of Chinese castor lines reveals genetic signatures of selection and yield-associated loci. *Nat. Commun.* **2019**, *10*, 3418. [[CrossRef](#)] [[PubMed](#)]
44. Zhang, K.; He, M.; Fan, Y.; Zhao, H.; Gao, B.; Yang, K.; Li, F.; Tang, Y.; Gao, Q.; Lin, T.; et al. Resequencing of global Tartary buckwheat accessions reveals multiple domestication events and key loci associated with agronomic traits. *Genome Biol.* **2021**, *22*, 23. [[CrossRef](#)] [[PubMed](#)]
45. Varshney, R.K.; Thudi, M.; Roorkiwal, M.; He, W.; Upadhyaya, H.D.; Yang, W.; Bajaj, P.; Cubry, P.; Rathore, A.; Jian, J.; et al. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* **2019**, *51*, 857–864. [[CrossRef](#)]
46. Batley, J.; Barker, G.; O'Sullivan, H.; Edwards, K.J.; Edwards, D. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol.* **2003**, *132*, 84–91. [[CrossRef](#)]
47. Bhardwaj, A.; Dhar, Y.V.; Asif, M.H.; Bag, S.K. In Silico identification of SNP diversity in cultivated and wild tomato species: Insight from molecular simulations. *Sci. Rep.* **2016**, *6*, 38715. [[CrossRef](#)] [[PubMed](#)]

48. Raizada, A.; Souframanien, J. Transcriptome sequencing, de novo assembly, characterisation of wild accession of blackgram (*Vigna mungo* var. *silvestris*) as a rich resource for development of molecular markers and validation of SNPs by high resolution melting (HRM) analysis. *BMC Plant Biol.* **2019**, *19*, 358. [[CrossRef](#)]
49. Cortés, A.J.; López-Hernández, F.; Blair, M.W. Genome–Environment Associations, an Innovative Tool for Studying Heritable Evolutionary Adaptation in Orphan Crops and Wild Relatives. *Front. Genet.* **2022**, *13*, 910386. [[CrossRef](#)]
50. Kaushik, S.; Kaushik, S.; Sharma, D. Functional genomics. *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.* **2018**, *1–3*, 118–133. [[CrossRef](#)]
51. Hamblin, M.T.; Buckler, E.S.; Jannink, J.L. Population genetics of genomics-based crop improvement methods. *Trends Genet.* **2011**, *27*, 98–106. [[CrossRef](#)]
52. Hazzouri, K.M.; Flowers, J.M.; Visser, H.J.; Khierallah, H.S.M.; Rosas, U.; Pham, G.M.; Meyer, R.S.; Johansen, C.K.; Fresquez, Z.A.; Masmoudi, K.; et al. Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nat. Commun.* **2015**, *6*, 8824. [[CrossRef](#)]
53. Teklu, D.H.; Shimelis, H.; Abady, S. Genetic Improvement in Sesame (*Sesamum indicum* L.): Progress and Outlook: A Review. *Agronomy* **2022**, *12*, 2144. [[CrossRef](#)]
54. Zhang, Y.; Liu, F.; Lin, Y.; Li, L.; Chen, M.; Ni, L. A Comprehensive Review on Distribution, Pharmacological Properties, and Mechanisms of Action of Sesamin. *J. Chem.* **2022**, *2022*, 4236525. [[CrossRef](#)]
55. Wei, P.; Zhao, F.; Wang, Z.; Wang, Q.; Chai, X.; Hou, G.; Meng, Q. Sesame (*Sesamum indicum* L.): A Comprehensive Review of Nutritional Value, Phytochemical Composition, Health Benefits, Development of Food, and Industrial Applications. *Nutrients* **2022**, *14*, 4079. [[CrossRef](#)] [[PubMed](#)]
56. Mohammed, S. An Overview on Nutritional Composition and Therapeutic Benefits of Sesame Seeds (*Sesamum indicum*). *Int. J. Res. Appl. Sci. Eng. Technol.* **2022**, *10*, 1119–1127. [[CrossRef](#)]
57. Kim, M.Y.; Kim, S.; Kim, J.I.; Oh, E.; Kim, S.W.; Lee, J.; Lee, E.; Lee, M.H. Antioxidant and Hepatoprotective Effects of Different Varieties of Sesame (*Sesamum indicum* L.) with Variation in Lignan Content. *J. Korean Soc. Food Sci. Nutr.* **2022**, *51*, 660–670. [[CrossRef](#)]
58. Wang, L.; Zhang, Y.; Li, D.; Dossa, K.; Wang, M.L.; Zhou, R.; Yu, J.; Zhang, X. Gene expression profiles that shape high and low oil content sesames. *BMC Genet.* **2019**, *20*, 45. [[CrossRef](#)] [[PubMed](#)]
59. Song, S.; You, J.; Shi, L.; Sheng, C.; Zhou, W.; Dossou, S.S.K.; Dossa, K.; Wang, L.; Zhang, X. Genome-wide analysis of ns1tp gene family and identification of siltps contributing to high oil accumulation in sesame (*Sesamum indicum* L.). *Int. J. Mol. Sci.* **2021**, *22*, 5291. [[CrossRef](#)]
60. Vanholme, R.; De Meester, B.; Ralph, J.; Boerjan, W. Lignin biosynthesis and its integration into metabolism. *Curr. Opin. Biotechnol.* **2019**, *56*, 230–239. [[CrossRef](#)]
61. Paniagua, C.; Bilkova, A.; Jackson, P.; Dabravolski, S.; Riber, W.; Didi, V.; Houser, J.; Gigli-Bisceglia, N.; Wimmerova, M.; Budínská, E.; et al. Dirigent proteins in plants: Modulating cell wall metabolism during abiotic and biotic stress exposure. *J. Exp. Bot.* **2017**, *68*, 3287–3301. [[CrossRef](#)]
62. Wang, X.; Wang, S.; Lin, Q.; Lu, J.; Lv, S.; Zhang, Y.; Wang, X.; Fan, W.; Liu, W.; Zhang, L.; et al. The wild allotetraploid sesame genome provides novel insights into evolution and lignan biosynthesis. *J. Adv. Res.* **2022**, *in press*. [[CrossRef](#)]
63. Murray, M.G.; Thompson, W.F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 2589–2598. [[CrossRef](#)]
64. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
65. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
66. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytzky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)] [[PubMed](#)]
67. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [[CrossRef](#)] [[PubMed](#)]
68. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)]
69. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [[CrossRef](#)]
70. Reverter, A.; Fortes, M.R.S. *Genome-Wide Association Studies and Genomic Prediction*; Gondro, C., van der Werf, J., Hayes, B., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2013; Volume 1019, ISBN 978-1-62703-446-3.
71. Zhang, C.; Dong, S.S.; Xu, J.Y.; He, W.M.; Yang, T.L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **2019**, *35*, 1786–1788. [[CrossRef](#)]
72. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]
73. Pickrell, J.K.; Pritchard, J.K. Inference of population splits and mixtures from genome-wide allele frequency data. *Nat. Publ. Gr.* **2012**, *8*, e1002967.

74. Koski, L.B.; Gray, M.W.; Lang, B.F.; Burger, G. AutoFACT: An automatic functional annotation and classification tool. *BMC Bioinform.* **2005**, *6*, 151. [[CrossRef](#)]
75. Sun, X.; Wu, K.; Cook, D. PKgraph: An R package for graphically diagnosing population pharmacokinetic models. *Comput. Methods Programs Biomed.* **2011**, *104*, 461–471. [[CrossRef](#)]
76. Wang, L.; Zhang, Y.; Li, P.; Wang, X.; Zhang, W.; Wei, W.; Zhang, X. HPLC analysis of seed sesamin and sesamolin variation in a sesame germplasm collection in China. *JAOCS J. Am. Oil Chem. Soc.* **2012**, *89*, 1011–1020. [[CrossRef](#)]
77. Wang, S.-B.; Feng, J.-Y.; Ren, W.-L.; Huang, B.; Zhou, L.; Wen, Y.-J.; Zhang, J.; Dunwell, J.M.; Xu, S.; Zhang, Y.-M. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* **2016**, *6*, 19444. [[CrossRef](#)] [[PubMed](#)]
78. Turner, S. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **2018**, *3*, 731. [[CrossRef](#)]
79. Dong, S.S.; He, W.M.; Ji, J.J.; Zhang, C.; Guo, Y.; Yang, T.L. LDBlockShow: A fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* **2021**, *22*, bbaa227. [[CrossRef](#)]
80. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)] [[PubMed](#)]
81. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)] [[PubMed](#)]
82. Jin, U.H.; Chun, J.A.; Han, M.O.; Lee, J.W.; Yi, Y.B.; Lee, S.W.; Chung, C.H. Sesame hairy root cultures for extra-cellular production of a recombinant fungal phytase. *Process Biochem.* **2005**, *40*, 3754–3762. [[CrossRef](#)]
83. Chun, J.A.; Lee, W.H.; Han, M.O.; Lee, J.W.; Yi, Y.B.; Goo, Y.M.; Lee, S.W.; Bae, S.C.; Cho, K.J.; Chung, C.H. Molecular and biochemical characterizations of dehydroascorbate reductase from sesame (*Sesamum indicum* L.) hairy root cultures. *J. Agric. Food Chem.* **2007**, *55*, 6067–6073. [[CrossRef](#)]
84. Wu, R.; Wang, L.; Wang, Z.; Shang, H.; Liu, X.; Zhu, Y.; Qi, D.; Deng, X. Cloning and expression analysis of a dirigent protein gene from the resurrection plant *Boea hygrometrica*. *Prog. Nat. Sci.* **2009**, *19*, 347–352. [[CrossRef](#)]
85. Wei, L.; Miao, H.; Zhao, R.; Han, X.; Zhang, T.; Zhang, H. Identification and testing of reference genes for Sesame gene expression analysis by quantitative real-time PCR. *Planta* **2013**, *237*, 873–889. [[CrossRef](#)] [[PubMed](#)]
86. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]
87. Chen, T.; Chen, X.; Zhang, S.; Zhu, J.; Tang, B.; Wang, A.; Dong, L.; Zhang, Z.; Yu, C.; Sun, Y.; et al. The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. *Genom. Proteom. Bioinforma.* **2021**, *19*, 578–583. [[CrossRef](#)] [[PubMed](#)]
88. Xue, Y.; Bao, Y.; Zhang, Z.; Zhao, W.; Xiao, J.; He, S.; Zhang, G.; Li, Y.; Zhao, G.; Chen, R. Database Resources of the National Genomics Data Center, China National Center for Bioinformatics in 2022. *Nucleic Acids Res.* **2022**, *50*, D27–D38. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.