Expression Profile And Genome-Wide Association Analyze of Silique Length In *Brassica Napus* L.

Jia wang
Southwest University https://orcid.org/0000-0002-4066-0201
Yueling Fan
Southwestern University
Lin Mao
Southwest University
Cunmin Qu
Southwest University
Kun Lu
Southwest University
Jiana Li
Southwest University
Liezhao Liu (liezhao@swu.edu.cn)
Southwest University

Research Article

Keywords: Brassica napus, silique length, QTL, GWAS, WGCNA, meta-GWAS

Posted Date: September 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-877275/v1

Abstract

Background: Silique length (SL) is an important trait tightly related to seed yield in *Brassica napus* (*B. napus*). Many studies related to SL have been reported in *B. napus*, but only a few candidate genes have been found and cloned, and the regulatory mechanism of SL is not clear.

Results: We identified QTL for SL by using a RIL population and two independent GWAS populations. Major QTL on A07, A09, and C08 chromosome were stably detected in all environments from all populations. As important candidate genes, several genes related to starch and sucrose metabolism, plant hormone signal transmission and phenylpropanoid biosynthesis were detected in the main QTL interval. Such as, *BnaA9.CP12-2*, *BnaA9.NST2*, *BnaA7.MYB63*, *BnaA7.ARF17*, etc. At the same time, the results of RNA-seq and WGCNA showed that starch and sucrose metabolism, photosynthesis, and secondary cell wall biosynthesis played an important role in the development of siliques.

Conclusions: we propose that photosynthesis, sucrose and starch metabolism, plant hormones, and lignin content play an important role in the development of rapeseed silique.

Introduction

Rapeseed (*Brassica napus* L.) is one of the oldest crops cultivated by human beings, and it is also an important oilseed crop with strong adaptability, fast growth and high economic value. As the second-largest oilseed crop in the world, *Brassica napus* (*B. napus*) is cultivated all over the world and is increasingly used for animal feed, vegetable oil and biodiesel [1]. Therefore, increasing rapeseed yield is one of the important goals of *B. napus* breeding and cultivation. Among the correlation traits of yield, silique length (SL) is an important trait tightly related to seed yield in *B. napus* [2, 3]. Silique plays an important role in the yield formation of *B. napus*. It is not only a sink organ for absorbing and accumulating photosynthetic products produced by leaves, but also a source organ for seed development [4]. In the late stage of seed development, the functional leaf area of *B. napus* decreased rapidly, and photosynthesis of green silique became the main source of nutrition for seed development [5]. Long silique generally has a large photosynthetic area to produce more energy potentially; it consumes much energy for its development. Therefore silique length is needed to balance the processes of producing, transferring, and consuming energy in silique in order to a trade-off between seed number and size [6]. In general, long siliques produce more and larger seeds than short siliques. Under the same planting density, silique number per plant, seed number per silique, and seed weight are the three direct components that determine the seed yield per plant [7]. Therefore, it is of great significance to clarify the genetic basis of long silique and cultivate long silique rapeseed varieties to improve the yield of rapeseed.

Most rapeseed agronomic traits are complex quantitative traits controlled by multiple genes and readily influenced by the environment, including silique length. Genome-wide association analysis (GWAS) and quantitative trait locus (QTL) mapping are effective method for dissecting complex traits. According to incomplete statistics, more than 100 QTL for SL have been identified by linkage and association

mapping, and QTL controlling SL is distributed on almost all chromosomes, though major QTL have been found mainly on chromosomes A07, A09, C02, C08, and C09 [1, 6, 8–15]. Based on these results, some candidate genes have been reported. Liu et al. (2015) successfully cloned the QTL gene *ARF18* that controls both silique length and seed weight in *B. napus* for the first time, which is a negative regulator. One 55-aa deletion prevented *ARF18* forming homodimer and inhibited the activity of downstream auxin genes, and promoted silique elongation by prolonging the length of pericarp cells [16]. *BnaA9.CYP78A9* was cloned in the major QTL region of A09 chromosome by fine mapping. A 3.7-kb insertion of a CACTA-like transposable element (TE) in the regulatory region functioned as an enhancer to stimulate the expression of *BnaA9.CYP78A9* and silique elongation [10]. The molecular regulation of silique development in *B. napus* is largely unknown, despite some important candidate genes were reported and their relationship with SL was analyzed. Therefore, further understanding the genetic architecture of rapeseed SL is necessary.

With the growing development of high-throughput sequencing technology and the decreasing continuously of high-throughput sequencing cost, GWAS is more and more widely used in the study of complex quantitative traits. Increasing the GWAS sample size is the most direct and effective way to improve the test efficiency. However, due to the huge workload and expensive experimental costs of GWAS, it may be difficult for a single GWAS sample size to meet the research needs. At the same time, the GWAS results for the same trait are often different due to different experimental populations, genotyping errors, and different analysis methods. Meta-analysis provides an attractive alternative to address the abovementioned challenges of individual GWAS, and the meta-GWAS studies have been utilized to detect genetic risk loci for various diseases in humans [17, 18]. In plants, using the meta-GWAS, Battenfield et al. (2018) identified marker-trait associations, allele effects, candidate genes, and can select using markers generated in bread wheat [19]. Another research group integrated genetic information from 73 published studies including 17556 soybean germplasm resources, and reported 393 unique peaks including 66 candidate genes across important traits and provide confirmation of many previously reported genes [20]. Fikere et al. (2020) identified 79 genomic regions (674 SNP) conferring potential resistance to canola blackleg by the meta-analysis GWAS [21]. Su et al. (2021) repoted that 3589 significant loci for three components traits and 3 loci for yield were detected by investigated the four traits of two rice hybrid populations in different environments and conducted meta-analyses of genome-wide association study (meta-GWAS) [22].

In this study, we carried out a QTL analysis with a population of 172 F9 recombinant inbred lines (RILs) and GWAS with two independent populations and identified significantly associated loci of SL. In addition, we performed a WGCNA with transcriptome sequencing of developing pericarp from short and long silique lines. Specifically, We aimed to detect candidate genes associated with SL, to analyze the genetic bases contributing to the relationship between short and long silique, and to investigate possible utilization patterns for selecting SL that is most conducive to increasing yield in breeding practice to further understand the genetic basis of SL and improve the rapeseed production.

Materials And Methods

Plant materials and trait measurement

The genetic map developed earlier by us using a recombinant inbred line (RIL) mapping population with 172 lines was used for QTL mapping in this study [29]. A previously reported GWAS population (referred to as 60K population) was used for association analysis of silique length [30], and another previously reported GWAS population (referred to as WGR population) was also used in this study [26]. The RILs population and the parents were grown in five environments, winter of 2016, 2017, and 2018 in Southwest University in Beibei, Chongqing, China (cq; 29.80°N, 106.40°E), and summer of 2018 and 2019 in Xian, Shanxi, China (xa; 34.27°N, 108.08°E). The 60K population and WGR population were cultivated under natural growing conditions in the experimental farm of Southwest University for 2015 and 2017, respectively. All lines were arranged in a randomized complete block design with three replicates, and each line was planted in two rows of 10 plants per row, with 30 cm between rows and a distance of 20 cm between plants within each row. After the silique is mature, use a ruler to measure the silique length (silique body length), and the silique beak length is not included in the silique length. Five siliques are randomly measured for each line of the RILs population and ten siliques are randomly measured for each accession of the GWAS population, and the average value is taken.

RNA-seq and WGCNA

Based on multi-year phenotypic observation, Z068 and Z191, from the RIL population, were used as representatives of the S- and L-silique. Silique pericarp of five lines with S- and L-silique were collected 9, 18, and 27 days after pollination (DAP) with two replications, and the samples were immediately frozen in liquid nitrogen and stored at -80°C for RNA-SEq. ST1, ST2, and ST3 represent S-silique at 9, 18 and 27 DAP, respectively; LT1, LT2, and LT3 represent L-silique at 9, 18 and 27 DAP, respectively. The libraries were sequenced on the Illumina HiSeq 4000 platform. After the removal of low-quality reads, the clean reads were aligned to the *B. napus* reference genome. Fragments per kilobase million (FPKM) values were calculated to estimate gene expression levels. Differentially expressed genes (DEGs) were identified in each pair of samples using the criteria listed in the Material and Methods section. The genes with reads per kilobase of transcript per million mapped reads (RPKM) value in more than 12 samples lower than 0.3 were filtered out.

WGCNA was performed using averaged FPKM values and the WGCNA package in R software, the analysis method based on tutorial of WGCNA official website (https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/) [31]. A total of 5809 putative candidate genes from RT-GWAS result were included in the WGCNA workflow. Briefly, cluster analysis was first carried out to remove outliers; then, the scale-free topology criterion was used to determine the softthreshold, which is defined as the similarity relationships between gene-pairs and obtained by computing the unsigned Pearson's correlation matrix, and threshold parameter beta selects the value at which the fitting curve first approaches 0.9; Subsequently, network was constructed using a step-by-step method by turning adjacency matrix into topological overlap matrix (TOM) and calling the

hierarchical clustering function, and network modules were identified using a dynamic tree cut algorithm with minimum cluster size of 30 and merging threshold function at 0.25. To identify hub genes within the modules, the module membership (MM) for each gene was calculated based on the Pearson correlation between the expression level and the module eigengene. Gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) were performed using the OmicShare tools, network depictions were visualized using Cytoscape v3.2.1.

QTL mapping, GWAS and meta-GWAS

QTL mapping were performed by composite interval mapping (CIM) using the software WinQTL Cartographer 2.5 [32], and LOD thresholds for QTL detection were set to $P = 0.05$ and were determined using permutation tests with 1000 permutations.

The GWAS was conducted by using the general linear model (GLM), mixed linear model (MLM) [33], Compressed mixed linear model (CMLM) [33], Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) [34], and Fixed and random model circulating probability unification (FarmCPU) [35] in GAPIT R package [36]. For meta-GWAS, we first divided the WGR population into three subgroups: spring, winter, and semi-winters, and then conducted GWAS independently. Then, pooled data from the GWAS results of three subgroups for meta-analysis. The meta-analysis was performed using METAL with the *P*-value, β-coefficients, and standard errors from single-subgroup GWAS [37]. In the single-population GWAS and meta-analysis, the genome-wide significant (1/N) thresholds by Bonferroni correction, in which N is the number of SNPs, were used in the analysis.

An LD block was generated using the Haploview v4.2 via the four-gamete rule [38]. The parameters were set as follows: the MAF was 0.05, the maximum number of Mendel errors was 1, the Hardy-Weinberg p-value cutoff was 0.001, and the minimum genotype was 75%.

Candidate genes analysis and qRT-PCR

The putative candidate genes were searched within the interval of the associated SNPs (± 200 kb) based on the Darmor-bzh *B. napus* reference genome v4.1. The functional annotation was implemented to predict the function of candidate genes using a Blastp program against to *Arabidopsis thaliana* TAIR10 protein database.

The expression level of silique length-related genes were validated by real-time PCR (qRT-PCR) using a CFX96 Real-time System (BIO-RAD, USA). In brief, 1 μg of RNA from each sample was used for cDNA synthesis; expression of the silique length-related genes in different rapeseed samples were evaluated using SYBR® Premix (TIANGEN, Beijing, China) and a CFX96 Real-time System (BIO-RAD, USA). Gene-specific primers were listed in Additional file 19: Table S9. Average Cq values were calculated from three replicates, and expression levels were normalized to reference gene Bna. Actin7 using the $2^{-\triangle\triangle Ct}$ method.

Statistical Analysis

The analysis of variance (ANOVA) of phenotypic data was conducted using “aov” function of R. The broad-sense heritability was estimated according to the following equation: $h^2=\sigma^2g/(\sigma^2g+\sigma^2ge/n+\sigma^2e/nr)\times100\%$, where σ^2g represents the genetic variance, σ^2ge represents the interaction variance between genotypes and environments, σ^2e represents the error variance, n represents the number of environments, and r represents the number of replicates within each environment, respectively [39]. To make the results more accurate, an lme4 package in R was used to estimate best linear unbiased predictions (BLUPs) and best linear unbiased estimates (BLUEs) across multi-environment on a per line basis for silique length [40]. The BLUPs and BLUEs value were used as a trait value for QTL analysis.

Results

Phenotypic evaluation showed that the SL varied widely

We measured the SL of 172 RILs from 2016 to 2019, with five replications performed each year, and measured the SL of 520 accessions during 2015 and 608 lines during 2017, with ten replications performed each year. The results showed that SL differs tremendously among rapeseed lines with a 76.26% broad-sense heritability, ranging from 4.81 to 10.89cm in the RIL population and ranging from 3.39 to 12.74cm in the GWAS population, 74.07% of lines concentrated in the range of 5.00–8.00cm (Fig. 1a, Additional file 1: Table S1, Additional file 2: Fig. S1a-b). A correlation analysis showed a strong correlation between SL and thousand seed weight (TSW) and yield per plant (YP), and seed number per silique (SN), seed number per plant (SNPP), siliques per the main inflorescence (SMI), harvest index (HI), and seed oil content (SOC) had a weak correlation with SL (Additional file 3: Fig. S2). The SL variation was further analyzed in Winter, Spring, and Semi-winter subgroups, respectively. The results indicated that there are differences among the three subgroups, but not significant. The SL of Semi-winter cultivars is 6.07 ± 1.23cm, while 5.51 ± 1.45cm and 5.31 ± 1.12cm in Winter and Spring accessions, respectively (Additional file 1: Table S2)

The dynamic observation of silique development was carried out using two extreme materials Z068 (5.45 ± 0.29cm) and Z191 (10.28 ± 1.01cm) who from the RIL population used as representatives of short and long silique, respectively. The results showed that there was no significant difference in SL between long and short siliques in the first 3 days after pollination (DAP). At 9 DAP, there was a significant difference in SL between long and short siliques. The length of the short silique tended to be stable at 18 DAP, but the length of the long silique did not elongate until 27 DAP (Fig. 1b). Further histological observation on the inner and outer epidermis of the long and short siliques of 27 DAP showed that the cell length of the long silique on the outer epidermis was significantly longer than that of the short silique, while the cell width of the long silique was significantly smaller than that of the short silique, and both the cell length and the cell width of the long silique on the inner epidermis were significantly larger than those of the short silique (Fig. 1c-j).

RNA-seq found the plant hormones and the transportation and synthesis of carbohydrates may affect the development of rapeseed silique

After a stringent quality filtering process, 79.37 Gb of clean data were obtained from 12 samples, with a Q30 percentage ≥ 94.80% (Additional file 4: Table S2). The obtained 12 samples of clean reads were mapped to the reference genome sequence of *B. napus*, and the percentages of mapped reads were similar among the 12 libraries (86.37–90.23%), and 82.71–86.29% of the reads were unique mapped (Additional file 5: Table S3). Based on the mapped results, the FPKM of all genes was counted and found that the $\log_{10}$FPKM value of all samples fluctuated slightly (-2.5–2.5) (Additional file 6: Fig. S3a), and the peak values of most genes were between 0 and 1 (Additional file 6: Fig. S3b). The results of correlation analysis showed that there was a high correlation and good repeatability between biological replicate (Additional file 6: Fig. S3c). After DEGs analysis, 21,482 DEGs were remained for further analysis (the number of DEGs is the sum of each DEG set) (Additional file 6: Fig. S3d-e, Additional file 7: Table S4).

To explore the metabolic pathways enriched for the DEGs, the related DEGs in 12 samples were subjected to KEGG metabolic pathway enrichment analysis. In "T1 vs T2", we found that the DEGs, whether from S-silique or L-silique, were significantly enriched in starch and sucrose metabolism, phenylalanine metabolism, phenylpropanoid biosynthesis, etc. pathways (Additional file 8: Fig. S4). The same enrichment result also appears in "T1 vs T3", and the difference is that the plant hormone signal transduction pathway, which was only enriched in the L-silique in "T1 vs T2", was also enriched in the S-silique in "T1 vs T3". In "T2 vs T3", the DEGs of L-silique were still enriched in plant hormone signal transduction, starch and sucrose metabolism, phenylalanine metabolism, etc (Additional file 8: Fig. S4). In the T3 stage, the origin of DEGs was significantly different from T1 and T2 stages, and the genes related to the cutin, suberine, and wax biosynthesis pathway were significantly up-regulated in this stage. Although there were significant differences in DEGs and enrichment pathways among the three stages, the genes related to starch and sucrose metabolism, plant hormone signal transduction, phenylpropanoid biosynthesis, etc. were differentially expressed in different stages of silique development. These results indicated that the genes related to starch and sucrose metabolism, plant hormone signal transduction, phenylpropanoid biosynthesis, etc. maybe play an important role in silique development.

Co-expression network analysis reveals transcript level differences of photosynthesis and secondary cell wall biosynthesis in long- and short-silique

To identify genes related to SL, we performed a weighted gene co-expression network analysis (WGCNA) using non-redundant DEGs. After using a dynamic tree cutting algorithm, a total of 18 distinct co-expression modules containing 47 to 1787 genes per module were identified, and 1585 uncorrelated genes were assigned into a grey module which was ignored in the following study (Fig. 2a). An analysis of the module-trait relationships revealed that the “lightpink1”(r = 0.97, p = 9e-08), “chocolate3”（r=-0.86, p = 4e-04）, “darkgoldnrod4”（r=-0.75, p = 0.005）, and “lightblue2”（r = 0.73, p = 0.006） module were highly correlated with the SL in the 12 samples (Fig. 2b,). According to the heatmap of the top 20 genes with the high eigengene connectivity (KME) value in the four modules, the “lightpink1” module had an expression

peak in the T2 stage of long silique development, similar to the long silique development pattern, while genes in “chocolate3” module were highly expressed in the T3 stage of short silique development (Additional file 9: Fig. S5a-b). The expression level of top 20 genes in "darkgoldnrod4” module was the highest in the T1 stage of short silique development, while the expression level of “lightblue2” module was the highest in the T3 stage of long silique development (Additional file 9: Fig. S5c-d).

A Gene Ontology (GO) enrichment analysis of the “lightpink1” module genes identified ten significantly enriched GO terms, and most of them related to photosynthesis. Interestingly, photosynthesis-related pathway were also enriched in “lightpink1” module by KEGG pathway enrichment analysis (Fig. 2c, Additional file 10: Table S5). There was no significant enrichment of GO terms in “chocolate3” module. One and 13 GO terms were significantly enriched in "darkgoldnrod4” module and “lightblue2” module, respectively. Among them, most of the enriched terms belonged to “molecular function” and “biological process”, including “monosaccharide transmembrane transporter activity” and “sucrose transport” (Additional file 10: Table S5). *Psby* (KME = 0.993), encoding a protein in photosystem II, is one of the Hub genes in the “lightpink1” module. In photosystem II, *Psby* is in close contact with *Cytb559*, which can protect photosystem II from photoinhibition so that the silique can better for photosynthesis and provide more material and energy for cell proliferation and expansion of the silique pericarp [23]. *BnaC09g09210D* (KME = 0.982), another Hub gene of the “lightpink1” module, is the homologous gene of *AtKNAT7*. In *Arabidopsis thaliana*, *AtKNAT7* was a homologousdomain transcription factor of *TALE* gene family, which is involved in the regulation of secondary cell wall biosynthesis, and its expression is up-regulated by *SND1* and *MYB46* [24]. The genes with weights value between 0.8 and 1 were screened to construct part co-expression network around hub genes. In the network, multiple genes are involved in cell elongation and expansion, such as *DFL2*, *CSLD3*, *TCH4*, *CESA6*, etc (Fig. 2d). These results suggest that photosynthesis and secondary cell wall biosynthesis played an important role in the change of silique length during development.

QTL and GWAS Co-located several major loci related to SL

Using linkage mapping, we identified 95 QTLs associated with SL, with a logarithm of the odds (LOD) value above 3.0. Of these, 95 QTLs with phenotypic variation explained (PVE) ranging from 0.02–67.06% were identified on nine chromosomes using four years, BLUE, and BLUP data (Additional file 11: Table S6, Additional file 12: Fig. S6). Three Major QTL were located on A07, A09, and C08, respectively. *qSLA7-2* on A07 was a major locus that was stably detected across all environments and explained 56.24–67.06% of the phenotypic variation. *qSLA9-1* and *qSLA9-3* on A09 also were major loci that were stably detected across all environments but only explained 0.83–5.33% of the phenotypic variation. *qSLC8-2* on C08 was another major locus that was stably detected across all environments and explained 9.91–11.31% of the phenotypic variation. In addition, *qSLA6-1* and *qSLA6-2* on A06 were also detected simultaneously in all environments, but they had a low PVE.

Trait-marker associations were performed using the FarmCPU, Blink, CMLM, GLM, and MLM models in two GWAS mapping populations. Similar results were obtained in the two mapping populations with five models (Fig. 3a-b). In order to facilitate further analysis, combined with Q-Q diagram, we finally choose

MLM model as the follow-up analysis model (Additional file 13: Fig. S7a-b). In total, we identified 41 SNPs in 60K population on A06 (1), A07 (6), A09 (15), A10 (1), and C08 (18) chromosomes, whereas 181 SNPs in WRG population were identified on A01 (9), A03 (9), A04 (11), A06 (10), A07 (52), A08 (5), A09 (54), A10 (4), C01 (1), C04 (1), C05 (5), C06 (1), C07 (4), C08 (9), and C09 (6) chromosomes using a threshold of 5% after Bonferroni multiple test correction (Additional file 14: Table S7). The SL was associated with three common significant regions located on A07, A09, and C08 chromosomes, respectively. Among these, 40 SNPs forming a haplotype block on A09 (27.51–28.18Mb) were located in the interval of known QTL for SL [6].

Meta-GWAS and polymorphisms in the candidate region were associated with SL

The meta-analysis detected 85 SNPs associated with SL, of which nine SNPs were undetected in all of the single-population GWAS (Additional file 14: Table S7). 14 and 31 SNPs identified in the spring and semi-winters subgroups were confirmed by meta-analysis, respectively, but only three SNPs identified in the winter subgroup were confirmed by meta-analysis. On the A09 chromosome, the confidence interval of a stable major QTL overlapped with the highly associated region detected by GWAS (Fig. 3c-d). 40 SNPs identified in GWAS were confirmed by meta-analysis, and seven SNPs undetected in GWAS were mined by meta-analysis in this interval(Additional file 14: Table S7). 528 gene symbols were found in this region. Among them, 98 were DEGs in RNA-seq, which listed as candidate genes (Additional file 15: Table S8). LD analysis showed that most peak SNPs were mainly concentrated in block 2 and block 6 (Fig. 3e). The peak SNP (S9_28151819) was involved in a 35-kb LD block (block 6) that encompassed 19 SNP markers, and it was 0.91 kb away from the *BnaA9.CP12-2* related to carbohydrate anabolism (Additional file 15: Table S8). In addition, the gene *BnaA9.NST2* involved in secondary cell wall biogenesis and Aux/IAA family member *BnaA9.IAA30* related to auxin signal were also found in this block. The results of RNA-seq and qRT-PCR showed that the expression of *BnaA9.NST2* in short silique was significantly higher than that in long silique, especially in the T2 stage (18 DAP)(Fig. 3f).

In block 2, there are two peak SNPs located inside genes *BnaA9.SK21* and *BnaA9.TMP-C*, respectively. *BnaA9.SK21* and *BnaA9.TMP-C* have the same expression pattern in T1 and T2, but *BnaA9.TMP-C* has a higher expression level in long silique in T3 (Fig. 3g-h). Two nonsynonymous SNPs, S9_27782829 and S9_27788376, are wthin *BnaA9.SK21* and *BnaA9.TMP-C*, respectively. Further analysis of these two SNPs found that accessions with a AA genotype at the S9_27782829 displayed, on average, 18.56% increase silique length compared with the accessions with a TT genotype. The average silique length of AT genotypes was between the AA and TT genotype. The minor allele (A) was represented in only 17% of the 608 accessions (Additional file 16: Fig. S8a). There was a significant difference between GG and CC genotype at the S9_27788376 ($p < 0.01$), and the CC genotype will increase silique length (Additional file 16: Fig. S8b).

Similar results were found on chromosome A07, where a stable major QTL Co-located with GWAS. 14 SNPs identified in GWAS were confirmed by meta-analysis, and four SNPs undetected in GWAS were

mined by meta-analysis in this interval (Additional file 17: Fig. S9). 102 gene symbols were found in this region, of which 46 were DEGs in RNA-seq, which listed as candidate genes (Additional file 15: Table S8). *BnaA7.MYB63*, homologous gene *MYB63* is a transcriptional regulators specifically activating lignin biosynthetic genes during secondary wall formation in *Arabidopsis thaliana* [25], was only 0.36 kb away from the significant SNP S7_16015077 with a low expression level during the whole silique development (Fig. 3i). The lignin biosynthesis key gene *BnaA7.CCR2* was also detected in this QTL region and *BnaA7.CCOAMT*, another key gene of lignin biosynthesis, was detected in another highly associated region on A07. Three significant SNPs S7_16214445, S7_16214995, and S7_16215169) are located inside *BnaA7.ARF17*, and *BnaA7.ARF17* is a negative auxin response factor that inhibits downstream auxin-related genes. Further haplotype analysis was focused on the gene *BnaA7.ARF17*, and four major haplotypes were observed, with low frequency haplotypes (less than five accessions) being omitted (Fig. 4a). We conducted multiple comparison tests of SL, and the results showed that Hap2 and Hap3 had shorter SL than Hap4 ($P < 0.05$), while Hap1 was an intermediate type (Fig. 4b).

We further analyzed the sequence diversity of the major QTL region on chr A07, A09, and C08 among landraces and pseudo-wild ancestral (European turnip and *B. oleracea* subspecies) genomes based on previously published data [26] (Fig. 4c-d). The landraces had a lower π value than pseudo-wild ancestral in the major loci region of A09 (27.50–29.40 Mb). On A07, the highest π value of the major QTL region (from approx. 15.80 to 16.4 Mb) was in landraces (1.18×10^{-3}) and pseudo-wild ancestral (1.36×10^{-3}). The sequence diversity of the major loci region on chromosome C08 is consistent with A07 and A09 (Additional file 16: Fig. S8c). These results suggested that these major QTLs might be domesticated and selected during the process of rapeseed domestication from wild type to cultivated rapeseed, resulting in a decrease in sequence diversity.

Lignin biosynthesis plays an important role in silique elongation

The important candidate gene *BnaA9.NST2*, *BnaA7.MYB63, BnaA7.CCR2, and BnaA7.CCOAMT*, all of them are lignin biosynthesis related genes, found in the major QTL region of A07 and A09. Meanwhile, the pathway of “Phenylpropanoid biosynthesis” was significantly enriched in different developmental stages of long silique and short silique (Additional file 18: Fig. S10a-c). These results indicate that lignin may play an important role in the development of silique. Thus, we compared the expression levels of genes related to the lignin biosynthesis pathway in different developmental stages of long silique and short silique. An interesting phenomenon is that the expression level of lignin biosynthesis related genes in different stages of the long and short siliques development are significantly different, especially in the T2 stage when the length difference between long and short silique increased sharply, the lignin biosynthesis related genes are highly expressed in short silique (Additional file 18: Fig. S10d).

Similar results were obtained from the observation of the silique pericarp tissue section and the determination of lignin content. The lignified degree increased gradually with the silique developmental process. Especially at the T2 stage, the lignification degree of short silique pericarp was significantly

higher than that of long silique pericarp (Fig. 5a-f). Correspondingly, the lignin content of the short silique was significantly higher than that of the long silique at the T1 and T2 stages, especially T2. However, there was no significant difference in lignin content between long and short siliques at the T3 stage (Fig. 5g). Consistent with the results of RNA-seq, the expression levels of the four genes in the long silique were significantly lower than those in the short silique, especially the T2 stage (Fig. 5h). All these results suggest that lignin plays an important role in the formation of long and short silique, especially in the rapid elongation period of the silique.

Discussion

Population stratification is a common source of false positives in GWAS studies. When the population is stratified, the conventional method is to use fixed effect covariate matrix (Q matrix) or PCA analysis to control the false positive caused by population structure. Meta-analysis can increase power, reduce false-positive findings, and even identify some new genetic loci, so it can solve the shortcomings of single-population GWAS [27]. In this study, we tried to perform independent GWAS by subgroups, and then perform meta-analysis on the significant SNPs obtained by independent analysis. The results show that the meta-analysis confirmed 71 SNPs identified in the single-population GWAS, of which 59 SNPs had greater *p*-values than in the single-population GWAS. It is worth noticing that 14 SNPs undetected in GWAS were identified by meta-analysis, and seven SNPs identified in the single-population GWAS were unconfirmed by meta-analysis. S9_27787423 is closely interlocked with S9_27788376 and located inside *BnaA9.TMP-C*. The presence of S9_27788376 may mask the effect of S9_27787423, resulting in a small contribution to the phenotype. Meta-analysis showed that S9_27787423 was a significant locus. Similar to this are S7_15927280 and S9_27929254. In brief, using the meta-analysis, missing SNPs that were undetected by the single-population GWAS were retrieved and false-positive SNPs that were identified by the single-population GWAS were filtered, and showed the advantage of meta-analysis that it can not only integrate the results of different research groups to increase the sample size of GWAS analysis, but also can integrate results across subgroups to avoid the influence of population stratification.

Silique is an important photosynthetic product storage organ in rapeseed. Moderately increasing the length of silique can increase the capacity of the silique sink, which is conducive to the transportation of filling materials to seeds. RNA-seq data display that photosynthesis maybe plays an important role during silique development. An interesting phenomenon is that *CP12-2*, a gene related to carbohydrate synthesis and metabolism [28], as an important candidate gene, was detected simultaneously within the main QTL confidence interval on A09 and C08. *BnaA9.CP12-2* and *BnaC8.CP12-2* have the same expression pattern that there is no significant difference between long and short silique in the T1 stage. In T2 / T3 stage, the expression in the long silique is significantly higher than that in the short silique, especially in the T2 stage when the difference in silique expands sharply. At the same time, KEGG enrichment analysis found that the synthesis and transport pathways of starch and sucrose were significantly enriched in both vertical and horizontal development stages. These results indicate that carbohydrate synthesis and metabolism are directly involved in the regulation of silique development.

In this study, both linkage analysis and association analysis detected major QTL loci on A09. The QTL of SL on the A09 chromosome has always been the focus of research. Although major QTLs have been detected on A09 in many studies, there are differences in mapping results, indicating that there may be multiple genes controlling SL on A09. At present, two SL-related genes have been cloned on A09, and *ARF18* was the first cloned gene. This gene regulates the SL by regulating the auxin signal pathway and can change the seed weight without changing the number of seeds per silique [16]. After that, Shi et al. cloned *BnaA9.CYP78A9* using map-based cloning [10]. This gene is widely expressed in rapeseed tissues and regulates silique development by affecting the auxin content. Its target gene *ARF10/16/17* is similar to the mechanism of *ARF18* reported by predecessors that they are auxin negative response factors and inhibit downstream auxin-related genes. In this study, in the RNA-seq data, we found that the plant hormone signal transduction pathway was significantly enriched at different stages of silique development. At the same time, linkage analysis and GWAS detected several candidate genes related to plant hormones, such as *BnaA7.ARF17*, *BnaA9.IAA30*, etc. These results further showed that plant hormones played an important role in silique development.

In previous studies, long and short siliques were often used as parental materials for gene mapping. Few studies have analyzed the characteristics of long and short siliques. In this study, through dynamic measurement, we found that there was a significant difference in the extension length between long and short siliques at 15 DAP, and the elongation time of short silique fruit was shorter. The results showed that the genes related to the lignin synthesis pathway were significantly differentially expressed between long and short siliques in the early, middle, and late stages of development. At the same time, *BnaA7.MYB63* and *BnaA9.NST2*, the key gene of lignin synthesis, as candidate genes detected by linkage analysis and GWAS. Further determination of lignin content also showed that there was a significant difference in lignin content between long and short siliques. These pieces of evidence suggest that lignin may be one of the key factor for the difference in silique length. The differential expression of lignin biosynthesis-related genes leads to the difference in lignin content. Lignin accumulation inhibits the expansion of pericarp cells and finally affects the silique length.

Conclusions

In the present study, we identified QTL for SL by using a RIL population and two independent GWAS populations. Major QTL on A07, A09, and C08 chromosomes were stably detected in all environments from all populations. Combining RNA-seq and WGCNA, we found that carbohydrate synthesis and metabolism and plant hormones are directly involved in the regulation of silique development, and the difference of lignin accumulation may also be a key factor for the difference in silique length. The study will provide theoretical guidelines for enhancing rapeseed yield potential by selecting SL that is most conducive to increasing yield.

Declarations

Acknowledgements

We are grateful to the reviewers and editors for their constructive review and suggestions for this paper.

Authors' contributions

LZL designed study; YLF collected the phenotypic data; JW and YLF conducted study, JW, YLF, and LM analyzed data; CMQ, KL, and JNL provided resources; JW wrote manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (31771830, 31971902) and the "111" Project (B12006).

Availability of data

The datasets supporting the conclusions of this article are included within the article and its additional files, and the raw RNA-Seq data have been deposited in the NCBI database under BioProject accession codes PRJNA752474.

Competing interests

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 interesting.

References

1. Wang X, Chen L, Wang A, Wang H, Tian J, Zhao X, et al. Quantitative trait loci analysis and genome-wide comparison for silique related traits in Brassica napus. BMC Plant Biol [Internet]. BMC Plant Biology; 2016;16:1–15. Available from: http://dx.doi.org/10.1186/s12870-016-0759-7.
2. Diepenbrock W. Yield analysis of winter oilseed rape (Brassica napus L.): A review. F Crop Res. 2000;67:35–49.
3. Chay P, Thurling N. Identification of Genes Controlling Pod Length in Spring Rapeseed, Brassica napus L., and their Utilization for Yield Improvement. Plant Breed. 1989;103:54–62.
4. Bennett EJ, Roberts JA, Wagstaff C. The role of the pod in seed development: Strategies for manipulating yield. New Phytol. 2011;190:838–53.
5. King SP, Lunn JE, Furbank RT. Carbohydrate content and enzyme metabolism in developing canola siliques. Plant Physiol. 1997;114:153–60.
6. Dong H, Tan C, Li Y, He Y, Wei S, Cui Y, et al. Genome-wide association study reveals both overlapping and independent genetic loci to control seed weight and silique length in Brassica napus. Front Plant Sci. 2018;9:1–9.
7. Wang S, Cao L, Wang H. Arabidopsis ubiquitin-conjugating enzyme UBC22 is required for female gametophyte development and likely involved in Lys11-linked ubiquitination. J Exp Bot.

2016;67:3277–88.

8. Fu Y, Wei D, Dong H, He Y, Cui Y, Mei J, et al. Comparative quantitative trait loci for silique length and seed weight in Brassica napus. Sci Rep Nature Publishing Group. 2015;5:1–9.
9. Li N, Shi J, Wang X, Liu G, Wang H. A combined linkage and regional association mapping validation and fine mapping of two major pleiotropic QTLs for seed weight and silique length in rapeseed (Brassica napus L.). BMC Plant Biol. 2014;14:114.
10. Shi L, Song J, Guo C, Wang B, Guan Z, Yang P, et al. A CACTA-like transposable element in the upstream region of BnaA9.CYP78A9 acts as an enhancer to increase silique length and seed weight in rapeseed. Plant J. 2019;98:524–39.
11. Wang H, Zaman QU, Huang W, Mei D, Liu J, Wang W, et al. QTL and Candidate Gene Identification for Silique Length Based on High-Dense Genetic Map in Brassica napus L. Front Plant Sci. 2019;10:1579.
12. Yang P, Shu C, Chen L, Xu J, Wu J, Liu K. Identification of a major QTL for silique length and seed weight in oilseed rape (Brassica napus L.). Theor Appl Genet. 2012;125:285–96.
13. Yang Y, Shen Y, Li S, Ge X, Li Z. High density linkage map construction and QTL detection for three silique-related traits in orychophragmus violaceus derived Brassica napus population. Front Plant Sci. 2017;8:1512.
14. Zhao W, Zhang L, Chao H, Wang H, Ta N, Li H, et al. Genome-wide identification of silique-related traits based on high-density genetic linkage map in Brassica napus. Mol Breed. 2019;39:86.
15. Zhou X, Dai L, Wang P, Liu Y, Xie Z, Zhang H, et al. Mining favorable alleles for five agronomic traits from the elite rapeseed cultivar Zhongshuang 11 by QTL mapping and integration. Crop J [Internet]. Crop Science Society of China and Institute of Crop Science; 2021; Available from: https://doi.org/10.1016/j.cj.2020.12.008.
16. Liu J, Hua W, Hu Z, Yang H, Zhang L, Li R, et al. Natural variation in ARF18 gene simultaneously affects seed weight and silique length in polyploid rapeseed. Proc Natl Acad Sci U S A. 2015;112:E5123–32.
17. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008;40:638–45.
18. De Jager PL, Jia X, Wang J, De Bakker PIW, Ottoboni L, Aggarwal NT, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet. 2009;41:776–82.
19. Battenfield SD, Sheridan JL, Silva LDCE, Miclaus KJ, Dreisigacker S, Wolfinger RD, et al. Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. PLoS One. 2018;13:e0204757.
20. Shook JM, Zhang J, Jones SE, Singh A, Diers BW, Singh AK. Meta-GWAS for quantitative trait loci identification in soybean. G3 Genes|Genomes|Genetics. 2021;11.

21. Fikere M, Barbulescu DM, Malmberg MM, Spangenberg GC, Cogan NOI, Daetwyler HD. Meta-analysis of GWAS in canola blackleg (Leptosphaeria maculans) disease traits demonstrates increased power from imputed whole-genome sequence. Sci Rep [Internet]. Nature Publishing Group UK; 2020;10:14300. Available from: https://doi.org/10.1038/s41598-020-71274-6.
22. Su J, Xu K, Li Z, Hu Y, Hu Z, Zheng X, et al. Genome-wide association study and Mendelian randomization analysis provide insights for improving rice yield potential. Sci Rep [Internet]. Nature Publishing Group UK; 2021;11:6894. Available from: https://doi.org/10.1038/s41598-021-86389-7.
23. Von Sydow L, Schwenkert S, Meurer J, Funk C, Mamedov F, Schröder WP. The PsbY protein of Arabidopsis Photosystem II is important for the redox control of cytochrome b559. Biochim Biophys Acta - Bioenerg [Internet]. Elsevier B.V.; 2016;1857:1524–33. Available from: http://dx.doi.org/10.1016/j.bbabio.2016.05.004.
24. Liu Y, You S, Taylor-Teeples M, Li WL, Schuetz M, Brady SM, et al. BEL1-LIKE HOMEODOMAIN6 and KNOTTED ARABIDOPSIS THALIANA7 interact and regulate secondary cell wall formation via repression of REVOLUTA. Plant Cell. 2014;26:4843–61.
25. Zhou J, Lee C, Zhong R, Ye ZH. MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. Plant Cell. 2009;21:248–66.
26. Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, et al. Whole-genome resequencing reveals Brassica napus origin and genetic loci involved in its improvement. Nat Commun. 2019;10:1–12.
27. Zhou S, Ding R, Meng F, Wang X, Zhuang Z, Quan J, et al. A meta-analysis of genome-wide association studies for average daily gain and lean meat percentage in two Duroc pig populations. BMC Genomics BMC Genomics. 2021;22:12.
28. Fermani S, Trivelli X, Sparla F, Thumiger A, Calvaresi M, Marri L, et al. Conformational selection and folding-upon-binding of intrinsically disordered protein CP12 regulate photosynthetic enzymes assembly. J Biol Chem [Internet]. Â© 2012 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology.; 2012;287:21372–83. Available from: http://dx.doi.org/10.1074/jbc.M112.350355.
29. Liu L, Qu C, Wittkop B, Yi B, Xiao Y, He Y, et al. A high-density snp map for accurate mapping of seed fibre qtl in brassica napus l. PLoS One. 2013;8:e83052.
30. Wang J, Jian H, Wei L, Qu C, Xu X, Lu K, et al. Genome-Wide Analysis of Seed Acid Detergent Lignin (ADL) and hull content in rapeseed (Brassica napus L.). PLoS One. 2015;10:e0145045.
31. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
32. Wang S, Basten CJ, Zeng Z-B. Windows QTL Cartographer 2.5 [Internet]. Department of Statistics, North Carolina State University: Raleigh; 2012. Available from: http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.
33. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. Nat Genet [Internet]. 2010;42:355–60. Available from:

http://www.nature.com/articles/ng.546.

34. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. 8. Gigascience: Oxford University Press; 2019. pp. 1–12.
35. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet [Internet]. 2016;12:e1005767. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26828793.
36. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, et al. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. Plant Genome. 2016;9.
37. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.
38. Barrett JC, Fry B, Maller J, Daly MJ. Haploview. Analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–5.
39. Cai D, Xiao Y, Yang W, Ye W, Wang B, Younas M, et al. Association mapping of six yield-related traits in rapeseed (Brassica napus L.). Theor Appl Genet. 2014;127:85–96.
40. Merk HL, Yarnes SC, Van Deynze A, Tong N, Menda N, Mueller LA, et al. Trait diversity and potential for selection indices based on variation among regionally adapted processing Tomato Germplasm. J Am Soc Hortic Sci. 2012;137:427–37.

Figures

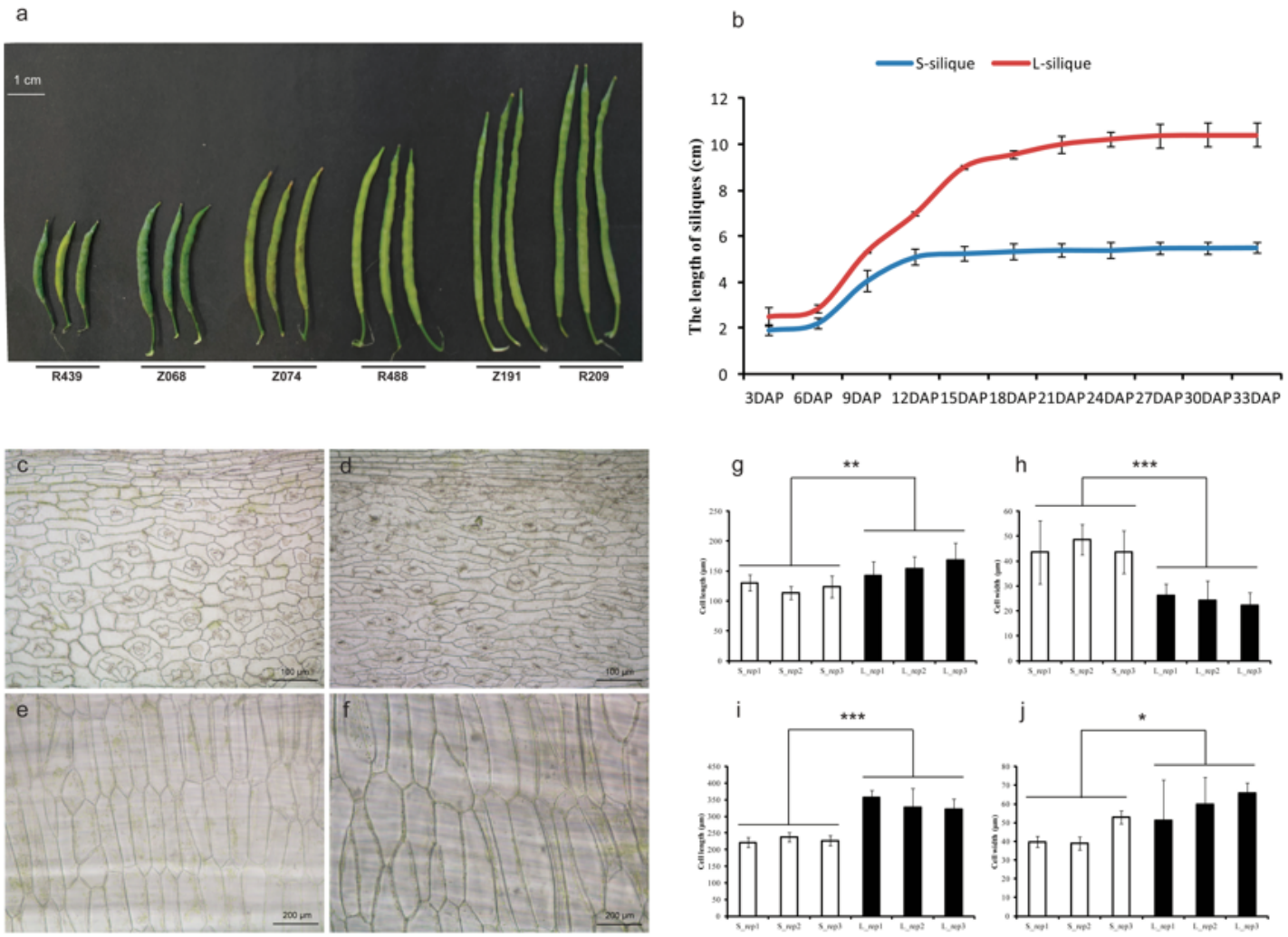

Figure 1

Phenotypic characterization of the short- and long-silique of repaseed. a, silique length of different lines, Bar = 1cm. b, dynamic changes of silique lengths of S- and L-silique in different development stages. c-d, microstructure observation of outer pericarp of the S-silique materials and L-silique materials, bar = 100μm. e-f, microstructure observation of endocarp of the S-silique materials and L-silique materials, bar = 200μm. g, cell length measurement of outer pericarp. h, cell width measurement of outer pericarp. i, cell length measurement of endocarp. j, cell width measurement of endocarp. Statistically significant differences were revealed using a Student's t-test: *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$.

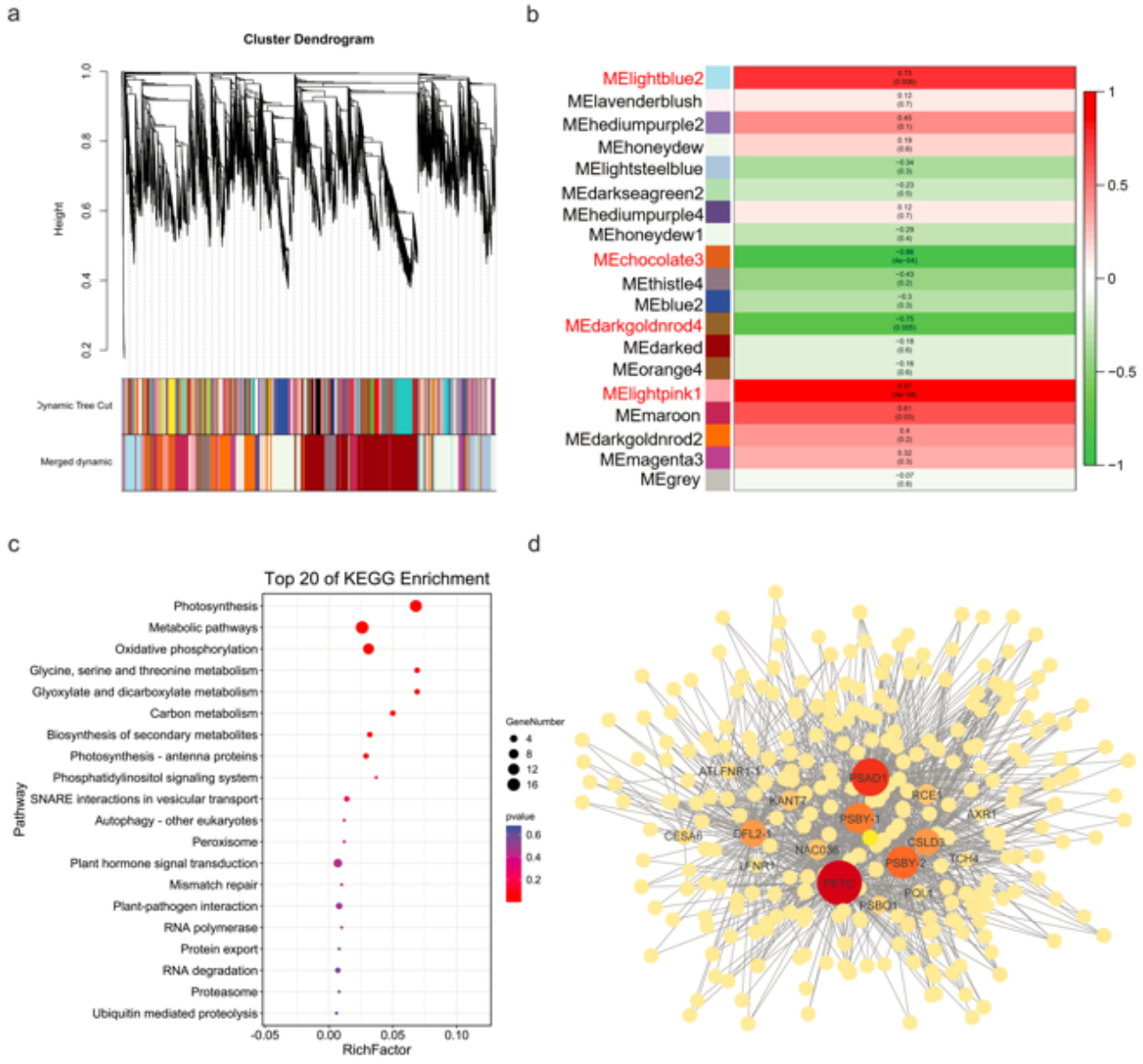

Figure 2

WGCNA coexpression network composed of DEGs. a, clustering dendrogram of genes and construction of modules. b, phenotype and module correlation analysis heat map. Red indicates that the correlation between module eigengenes and silique length is high. c, top 20 KEGG enriched pathways in the set of the lightpink1 module. d, the gene co-expression network of the lightpink1 module. The node and edge size is proportional to the core.

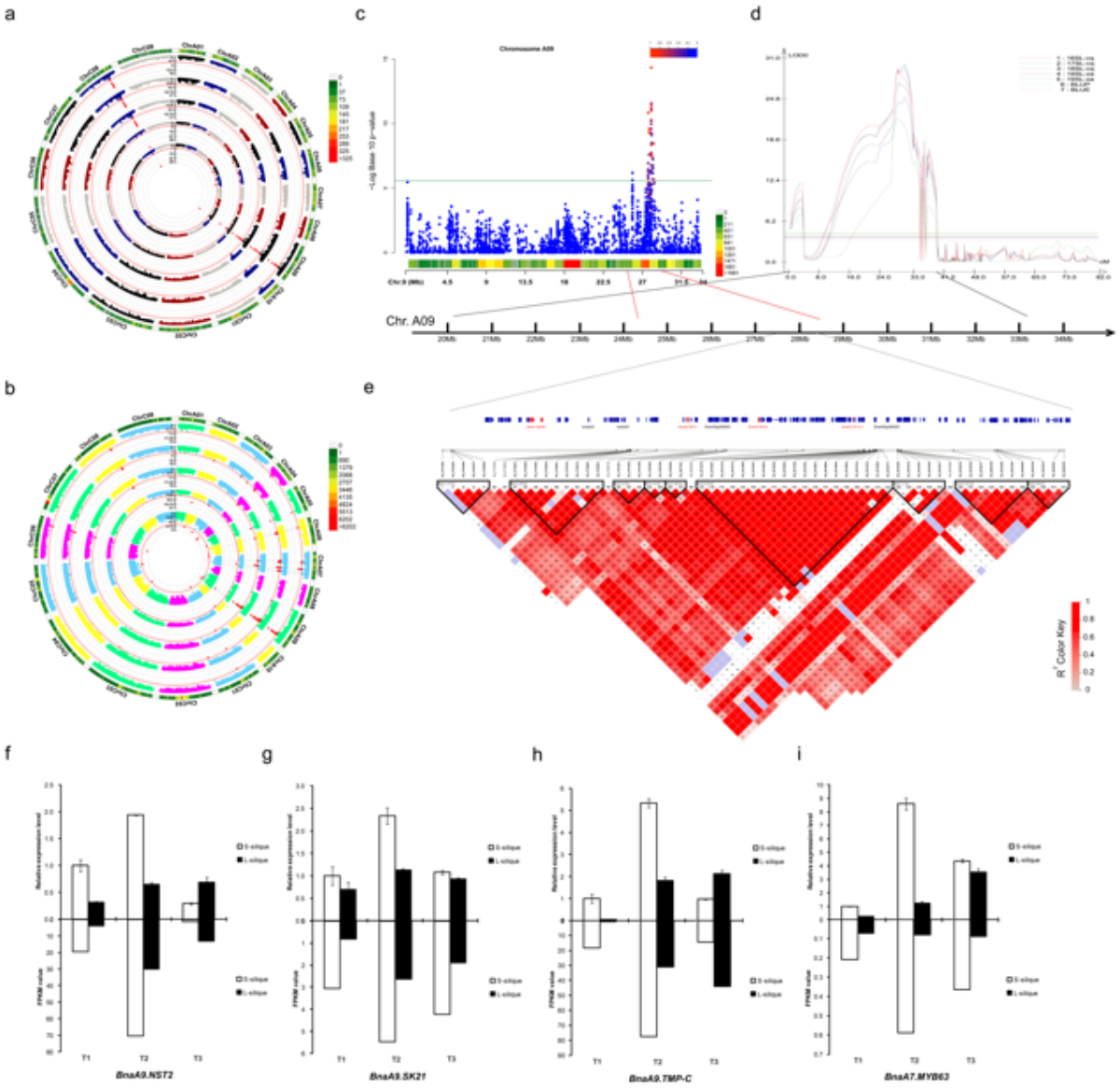

Figure 3

GWAS and QTL co-located major loci on A09 chromosome. a-b, circular Manhattan plots of 60K population and WGR population. From inner ring to outer ring are results of FarmCPU, Blink, CMLM, GLM, and MLM models. C, Scatterplot of association results from a MLM-model analysis of SL on chromosome A09. Negative log10-transformed P values from the GWAS analysis are plotted against the genomic physical position. The green line indicated the threshold level log(1/N) = 5.58. d, the major QTL loci on the A09 chromosome were repeatedly detected in multiple environments by QTL mapping in a RIL population. 16SL-cq, 17SL-cq, and 18SL-cq represent silique length from Chongqing in 2016, 2017, and 2018, respectively; 18SL-xa and 19SL-xa represent silique length from Xi'an in 2018 and 2019, respectively; BLUP and BLUE represent best linear unbiased predictions and best linear unbiased estimates, respectively. e, the location of the reference genome region on A09 corresponding with the major effective loci, and LD block analysis of this region. The red gene ID represents that the gene is an important candidate gene. f-i, FPKM value and qRT-PCR validation of candidate gene on A09. The upper

half represents qRT-PCR results, and the normalized levels ST1 were arbitrarily set to 1. The lower part is the FPKM value obtained by RNA-seq.

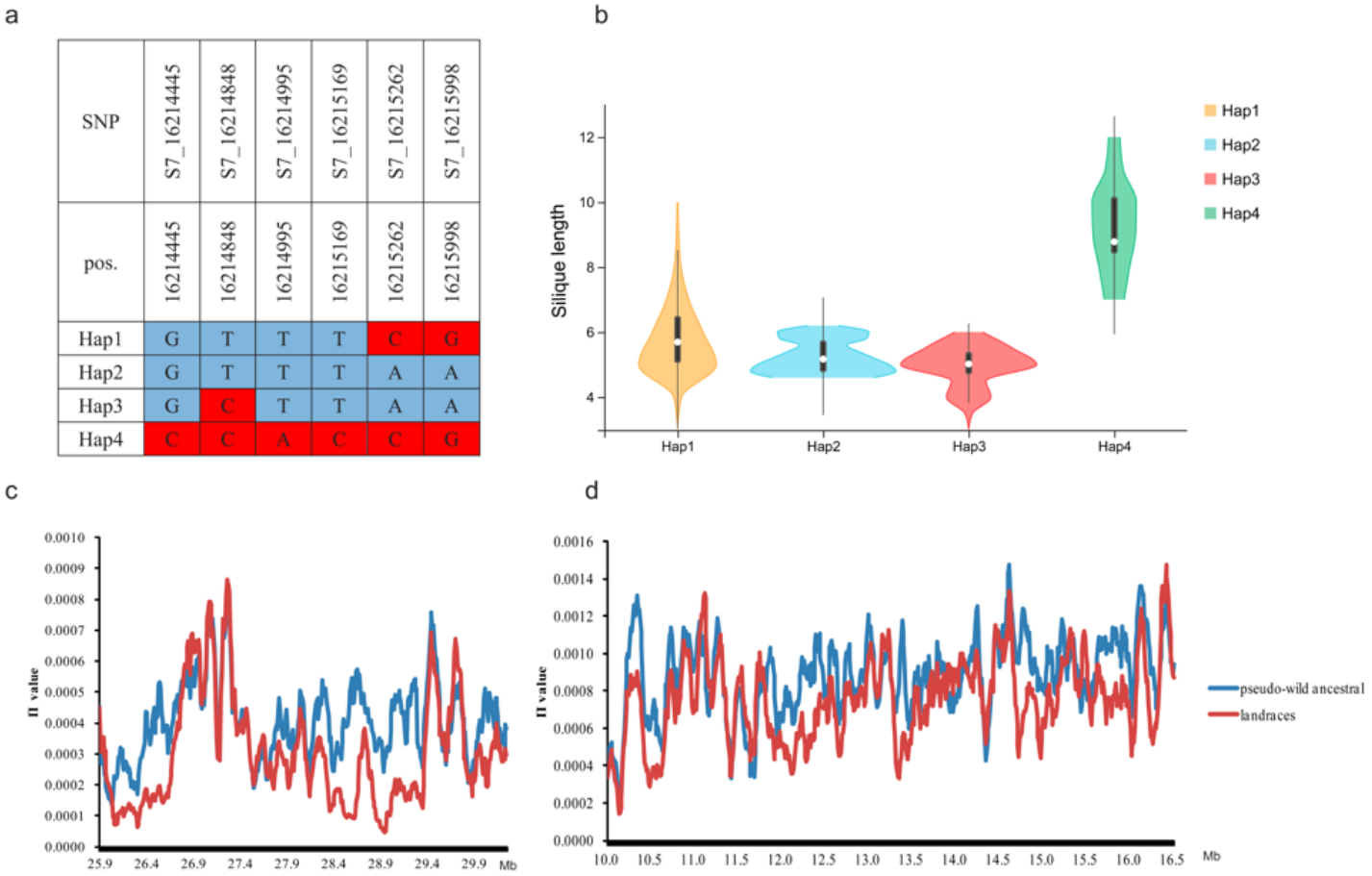

Figure 4

Haplotype analysis and polymorphisms in the candidate region were associated with SL. a, Haplotypes in 608 accessions (haplotypes with less than five accessions were omitted) according to SNP data from WGR population. b, Violin plots showing levels of SL from four haplotypes. c, Genomic diversity of chr A09. d, Genomic diversity of chr A07. The blue line represent pseudo-wild ancestral and the red line represent landrace rapeseed, respectively.

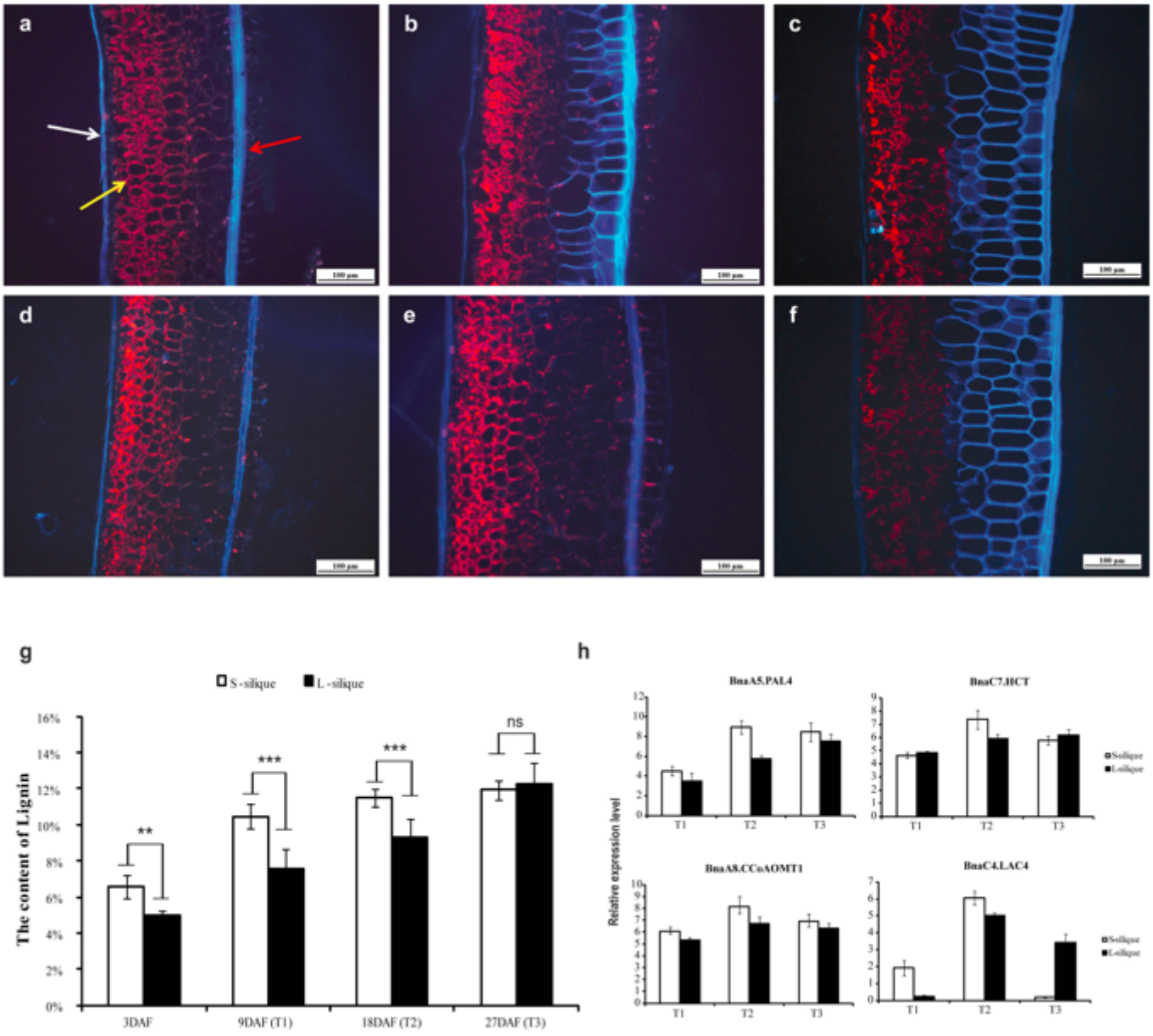

Figure 5

Evaluation of phenotypic contribution of lignin to silique length. a-f, Microstructure observation of the outer, middle and inner pericarp of the short and long silique. White arrows point to the outer pericarp, yellow arrows point to the middle pericarp, and red arrows point to the inner pericarp. a and d represent microscopic observation of short and long silique at T1 stage; b and e represent microscopic observation of short and long silique at T2 stage; c and f represent microscopic observation of short and long silique at T3 stage. Bar = 100μm. g, determination of total lignin content in silique wall of the short and long silique. h, qRT-PCR verification of DEGs of key enzymes in phenylpropanoid-lignin pathway in the short and long silique at T1, T2, and T3 stage.

Supplementary Files

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

- FigS1.tif
- FigS2.tif
- FigS3.tif
- FigS4.tif
- FigS5.tif
- FigS6.tif
- FigS7.tif
- FigS8.tif
- FigS9.tif
- FigS10.tif
- TableS1.docx
- TableS29.xlsx