Identification of environment-insensitive genes for oil content by a combination of transcriptome and genome-wide association analysis in Brassica napus

Lunwen Qian¹, Min Yao¹, Dan He¹, Wen Li¹, Xinghua Xiong¹, Xin He¹, Zhongsong Liu¹, and Chunyun Guan¹

¹Hunan Agricultural University

October 20, 2023

Abstract

Seed oil content is an important agricultural trait in rapeseed breeding, which is controlled by multiple genes and influenced by environmental factors. In this study, we used transcriptome data from inbred lines with higher and lower oil contents at 35 days after pollination (DAP) to investigate environment-insensitive genes that exhibiting stable expression across three different environments. Consequently, a total of 405 environment-insensitive differentially expressed genes (DEGs) were identified, including 25 involved in lipid/fatty acid metabolism and 14 transcription factors. Among these genes, BnBZIP10-A09, BnaMYB61-A06, BnAPA1-A08, BnPAS2-A10, BnLCAT3-C05 and BnKASHI-C09 were located in previous reported confidence intervals and also found to exhibit significant associations with oil content in multiple different environments by GWAS of 50 re-sequenced Chinese semi-winter rapeseed accessions. Otherwise, we revealed the presence of additive effects among BnBZIP10-A09, BnKASHI-C09, and BnAPA1-A08, resulting in a significant increase in seed oil content. Meanwhile, co-expression network analysis revealed that most of these environment-insensitive DEGs are interconnected either directly or indirectly, thereby forming a molecular network implicated in the potential regulation of seed oil accumulation and stability. These results are important for us to further improve oil content accumulation and stability in rapeseed.

Introduction

With the development of society, the consumption demand for edible oil is increasing rapidly. Most vegetable oils are produced by four major crops, including soybean, oil palm, rape and sunflower (Dyer and Mullen, 2008). Rapeseed (Brassica napus L.) contributes more than 15% of the world’s edible oil supply (Usda, 2014). Rapeseed is the most significant oilseed crop and a primary source of edible oil in China, Canadian, and the European Union. As people’s demand for edible oil increases year by year, enhancing seed oil content and oil production per unit area of land is one of the most important goals in the breeding of rapeseed.

The seed oil content is a complicated quantitative trait that is easily affected by environmental factors and varies from 35% to 55% in rapeseed depending on ecological zones and climate conditions (Zhou et al., 2018). Low temperature increases polyunsaturated fatty acid content in plants, contributing to maintaining the fluidity of biological membranes (Los and Murata, 1998). Bellaloui et al. (2013) demonstrated that high temperature altered oil production and composition, which could be partially related to the limited availability and movement of carbohydrates from leaves to seeds. Zhou et al. (2018) suggested that temperature affects storage lipids and gene expression changes in lipid metabolism in low and high oil content rapeseed cultivars. Some research has suggested that temperature directly affects the expression instability of the FAD2 and FAD3 genes in the process of lipid accumulation (Román et al., 2012; Zhu et al., 2012; Li et al., 2015). Meanwhile, light intensity is an essential factor in determining the efficiency of photosynthesis. Light intensity affects gene expression in lipid metabolism, which regulates the seed oil content in developing seeds.
The expression of *WRINKLED1*, an important lipid synthesis regulatory gene in developing seeds, is associated with silique wall photosynthetic activity (Hua et al., 2012). Although the molecular mechanism by which temperature and light regulate fatty acid accumulation is not well known, uncovering environment-insensitive genes in lipid metabolism may be a better way to further improve the stability of oil content and fatty acid composition.

Environmental and genotypic interactions lead to gene expression pattern differences that result in phenotypic diversity. With the development of next-generation sequencing, gene expression variation can be measured quantitatively, and DEGs related to phenotypes and/or environments can be explored. Transcriptome analysis has been applied to uncover the DEGs involved in rapeseed oil content (Tan et al., 2019; Yu et al., 2023). A large number of DEGs could be identified in short times by this application. However, the transcriptome was only interpreted as phenotypic variation in terms of gene expression and failed to fully interpret genetic variations (van Dam et al., 2015). Meanwhile, numerous differentially expressed genes make explaining transcriptome results and enacting breeding strategies more difficult. GWAS is an application that studies complex phenotypes by investigating genetic variations in the whole genome and has been extensively applied in rapeseed (He et al., 2017; Zhou et al., 2018b; Chen et al., 2019). Therefore, combining transcriptome analysis and GWAS to identify differentially expressed genes and explore genetic variations is a novel strategy. For example, Zhang et al. (2018) used GWAS combined with transcriptome analysis to reveal that *HCTs* and *WRKYs* interact to regulate the defence response of poplar. Xiao et al. (2019) identified a few key genes of the lipid biosynthesis pathway controlling oil content by combining GWAS and transcriptome analysis in *Brassica napus*. A combination of genome-wide and transcriptome-wide association studies identified three thermal tolerance-associated loci that contained 75 protein-coding genes and 27 long noncoding RNAs (Ma et al., 2021).

In this study, we performed transcriptome sequencing of two HOC and LOC accessions at 35 DAP across three different environments. Meanwhile, a GWAS was performed to detect candidate genes significantly associated with seed oil content in three different environments. Our aimed to identity environment-insensitive genes in the process of oil accumulation. These results will provide useful molecular markers for the improvement of high and stability oil content varieties.

**Materials and Methods**

**Plant Materials**

Four Chinese semi-winter rapeseed inbred lines, XY777 and XY015 (LOC) and CS136 and CS511 (HOC), came from Hunan Agricultural University, China, and were planted in Changsha (CS: E112.938888, N28.228272), Hangzhou (HZ: E120.15358, N30.287458) and Kunming (KM; E102.71225, N25.040609) in China. These four accessions were sown on October 10, 2015, in CS and HZ and in KM on May 15, 2016. Seed tissues were sampled with 3 biological replicates at 20, 25, 30, 35, 40, and 45 DAP. Seed oil content was measured by the Soxhlet extraction method according to the determination of fat in foods (GB 5009.6-2016, standardization administration, China). A diverse panel of 50 homozygous Chinese semi-winter inbred lines (selfing for at least 5 generations) broadly encompassed allelic variability in the Chinese semi-winter rapeseed gene pool (Table S1). These accessions were grown in the experimental field of Southwest University, Chongqing, (China) during 2013 and 2014 (designated Field_2013 and Field_2014, respectively) using a randomized complete block design with two replications. Each plot consisted of 24 plants, with 30 cm between rows and 20 cm within rows. Glasshouse trials were completed at the University of Giessen in 2012 (designated GH_2012), each accession with 5 inbred lines. Dry mature seeds were harvested from each of the replicated plants, and the oil content was measured using near-infrared spectroscopy (NIR).

**Library construction and Data processing**

Seed tissues were sampled with 3 biological replicates at 20, 25, 30, 35, 40, and 45 DAP. Tissues were immediately frozen in liquid nitrogen and stored at -80 degC until total RNA extraction. RNA was ex-
Extracted by using a Plant Total RNA Extraction Kit (TransZol Plant, BioTeke, Beijing, China) following the manufacturer’s instructions. The RNA purity and concentration were checked using a NanoPhotometer(r) spectrophotometer (IMPLEN, CA, USA) and a Qubit(r)3.0 Fluorometer (Life Technologies, CA, USA), respectively. Then, an Agilent 2100 RNA Nano 6000 Assay Kit (Agilent Technologies, CA, USA) was used to assess the integrity of the RNA. After the total RNA sample was qualified, good quality and acceptable integrity RNA was further processed for poly-A-containing mRNA cleansing, fragmentation of mRNA, synthesis of double-stranded cDNA, and amplification through polymerase chain reaction (PCR). Samples with a cDNA sequencing library effective concentration of more than 10 nM were considered acceptable for sequencing. Sequencing was performed using the Illumina HiSeq X sequencing platform with sequencing strategy PE150 (a pair-end 2x150 bp mode), and various quality-controlling measures for raw data were conducted to obtain high-quality clean data.

All subsequent analyses are based on clean data. HISAT2 (Kim et al., 2015) was used to align the high-quality paired and clean reads to the B. napus reference genome. The read number mapped to each gene was counted using HTSeq v0.6.1 (Anders et al., 2015). Then, the fragments per kilobase of transcript per million reads (FPKM) of each gene were calculated based on the length of the gene and read count (Trapnell et al., 2010).

**Differentially expressed genes**

The differential expression between different samples was determined by the R package DESeq2 v1.24.0 (Love et al., 2014). Genes with an adjusted p-value (padj) < 0.05 and |log2 (fold change)| > 1 were considered differentially expressed genes (DEGs). The common DEGs of different comparisons which considered as environment-insensitive DEGs were displayed by Venny analysis (https://bioinfoo.p.cnbc.cscie.es/tools/venny/index.html). Gene sequences from the B. napus reference genome (http://www.genoscope.cns.fr/brassicanapus/) blast to the Arabidopsis genome database (http://www.arabidopsis.org) were used to assign putative functions of B. napus genes.

**KEGG enrichment analysis**

Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed using TBtools (Chen et al., 2020) and visualized with “ggplot2” (Villanueva and Chen, 2019).

**Real-Time Quantitative PCR Verification**

qRT-PCR was carried out to confirm the RNA-seq results. cDNAs were synthesized from the same RNAs as for RNA-seq. The results analysis was performed on LightCycler 480 SYBR Green I Mastermix and a LightCycler 480II real-time PCR system (Roche, Switzerland). Gene copy-specific primers were used for all selected environment-insensitive DEGs to verify the expression pattern by qRT-PCR, and data were normalized by BnEF (Nesi et al., 2009) independently (primer details are given in Supplementary Table. S5). The fold change was estimated using the 2^ΔΔCT method (Livak and Schmittgen, 2001). For direct comparison with RNA-Seq data, the relative expression quantified by qRT-PCR was transformed to log2-fold change (HOC/LOC).

**Genome-wide association analysis**

Whole-genome sequencing of 50 Chinese semi-winter rapeseed inbred lines was carried out using the Illumina HiSeqTM 4000 (Illumina Co., Inc., San Diego, CA, USA) with 5X sequencing depth and 125-bp paired-end sequencing length. Library preparation and sequencing were carried out at the Biomarker Technologies Corporation (Beijing, China). A total of 532,005 high-quality, single-locus single-nucleotide polymorphism (SNP) markers with minor allele frequency (MAF) >0.05 were detected among the accessions described by (Dong et al., 2018).

TASSEL 5.0 software was used to perform the relative kinship and principal component analysis (Bradbury et al., 2007). Based on PCA and relative kinship (P and K matrix), the calculation was performed with a mixed linear model (MLM) incorporated into TASSEL 5.0 software. The critical p value for assessing
the significance of SNP-trait associations was calculated for seed oil content based on the false discovery rate (FDR) (Storey, 2002). An FDR < 0.05 was used to identify significant marker-trait associations for oil content at cut-off values of $-\log_{10}(P) = 4$.

Additive effect analysis

The R package “SIPI” (Lin et al., 2017) was used to evaluate pairwise interactions between the candidate gene SNP markers. In this study, only additive-additive interactions were considered. Wald $p$ values < 0.01 were defined as significant SNP pairs.

Weighted gene co-expression network analysis (WGCNA) for candidate genes

The co-expression network was constructed by the R package WGCNA v1.69 (Langfelder and Horvath, 2008). A total of 14,019 genes had FPKM max and a mean value greater than 10 and 2, respectively. Module identification was implemented after merging modules whose expression profiles were similar with a merge Cut Height of 0.25. The interaction network of genes in the module was visualized using Cytoscape 3.5.1 (Shannon et al., 2003).

Results

Changes in oil content in developmental stages and different environments

To investigate the dynamics of oil accumulation between HOC and LOC during seed development, we measured the seed oil content at 20, 25, 30, 35, 40, and 45 DAP across three distinct environmental conditions. CS and HZ are characterized by low temperatures and high levels of precipitation, in contrast to KM (Fig. 1). The accumulation of seed oil content in CS and HZ occurred between 20 and 40 DAP, with the period from 30 to 35 DAF exhibiting the most rapid increase, followed by a decline observed at 40 to 45 DAP. While the seed oil content at 20 to 35 DAP showed a faster increase, the period at 35 to 45 DAP exhibited a downwards increase in KM (Fig. S1a). HOC and LOC reached significant differences at 35 DAP in the three different environments (Fig. S1a). A principal component analysis (PCA) was conducted to profile the expression patterns of all genes in HOC and LOC at 35 DAP in the three environments. The first PC1 accounted for 14.0% of the genetic variation and was clearly separated into HOC and LOC. (Fig. S1b). These results indicate that gene expression is more stable in different environments at 35 DAP. Therefore, we particularly aimed at transcriptome data from 35 DAP to explore the stable expression genes involved in the process of lipid accumulation in different environments.

Identification of environment-insensitive genes related to seed oil content

To investigate environment-insensitive genes associated with seed oil content, we analysed the stable differentially expressed genes (DEGs) between two HOCs and LOCs seeds at 35 DAP in three different environments: CS, HZ, and KM. The DEGs were defined as the fold change of FPKM expression values and were at least 2 in either direction when the $q$ -value or FDR < 0.001 and the absolute value of $\log_2$ (fold change) > 1. We identified 1738, 1535 and 3359 up-regulated genes in two HOCs compared with two LOCs under CS, HZ, and KM, respectively, and 2495, 1589, and 3291 down-regulated genes were discovered, respectively (Fig. S2b). As illustrated in Fig. 2a, 220 and 185 genes were stably up-regulated and down-regulated in the three environments, respectively, indicating that these genes were environment-insensitive genes. KEGG enrichment analysis showed that these 405 environment-insensitive DEGs were significantly enriched in lipid metabolism, flavonoid biosynthesis, energy metabolism and so on (Fig. S2c). Among them, we found 25 environment-insensitive DEGs involved in lipid/fatty acid metabolism, and 14 environment-insensitive DEGs were transcription factors (Fig. 2c, Table. S1).

A quantitative real-time polymerase chain reaction (qRT-PCR) assay was performed to assess the reliability and validity of the RNA-Seq data with 39 environment-insensitive DEGs from seeds of both inbred lines at 35 DAP. The expression pattern of selected co-DEGs by RNA-Seq and qRT-PCR was highly correlated ($R^2 = 0.70-0.77$), further confirming the reliability and consistency of our RNA sequencing data (Fig. 2d).

Identification of environment-insensitive genes for seed oil content
GWAS was performed to detect associations between SNPs and oil content in 50 Chinese semi-winter rapeseed accessions. Manhattan plots describing significant SNP associations for seed oil content in the glasshouse and field experiments are shown in Fig. 3d. A total of 60 SNPs were identified as being significantly associated with oil content using the significance threshold of $-\log_{10}(P) \geq 4.0$. These SNPs were distributed on 19 chromosomes, including 35 SNPs located in 23 QTLs previously reported (Fig. 3c and d; Table. S2). Meanwhile, combined with 39 environment-insensitive DEG gene analyses, 25 lipid/fatty acid metabolic and 14 transcription factor genes were located in these 23 QTL regions (Fig. 3b; Table. S2).

Next, two of 14 environment-insensitive transcription factor DEGs, namely, $BnaMYB61$ -A06 and $BnBZIP10$ -A09, were detected to be located in QTL-A06 (2,847,280-3,181,818 bp) and QTL-A09.1 (142,667-1,638,715 bp), respectively (Fig. 4). These two genes showed a significant association with oil content (Fig. 4a and d). In addition, two and three haplotype alleles were identified within these gene regions (Fig. 4c and f). Comparative analysis of these haplotype alleles related to the oil content phenotype revealed that $BnaMYB61$ -A06-Hap1 and $BnBZIP10$ -A09-Hap1 corresponded to accessions with relatively high oil content (Fig. 4c and f; Table. S3).

Among the 25 environment-insensitive DEGs of lipid/fatty acid metabolic processes, namely, $Bn APA1$ -A08, $BnPAS2$ -A10, $BnLCAT3$ -C05 and $BnKASIII$ -C09 were found to be located within the significant interval QTLs QTL-A08 (16,454,214-17,628,998 bp), QTL-A10 (14,388,914-15,045,397 bp), QTL-C05 (41,242,900-42,812,067 bp) and QTL-C09.2 (6,631,263-7,574,558 bp), respectively (Fig. S3 and Fig. S4). These four genes showed a significant association with oil content (Fig. S3a and d). Two, four, and three haplotype (Hap) alleles were found in these four gene regions, and $BnAPA1$ -A08-Hap1, $BnPAS2$ -A10-Hap1, $BnLCAT3$ -C05-Hap1, and $BnKASIII$ -C09-Hap1 had higher oil contents than the other haplotype alleles, respectively (Fig. S3c and f; Fig. S4c and f; Table. S3).

Analysis of additive effects for environment-insensitive genes

To perform a single-variant-additive-effect analysis in GWAS, we investigated the interaction effects on oil content among both significant and nonsignificant SNP markers within these six candidate gene regions. Our findings revealed that the combination of $BnKASIII$ -C09 and $BnAPA1$ -A08, as well as $BnKASIII$ -C09 and $BnBZIP10$ -A09, exhibited an additive effect (Fig. 5a, Table. S6). Additionally, we identified two combinations of haplotype alleles (BnKASIII -C09-Hap1+$BnAPA1$ -A08-Hap1, $BnPAS2$ -A10-Hap1, $BnLCAT3$ -C05-Hap1, and $BnKASIII$ -C09-Hap1+$BnBZIP10$ -A09) that correspond to accessions with relatively high oil content compared to single haplotype alleles (BnKASIII -C09-Hap1, $BnAPA1$ -A08-Hap1, and $BnBZIP10$ -A09-Hap1) (Fig. 5b). These results suggest potential additive effects of these three candidate genes on oil content.

Co-expression network analysis of environment-insensitive genes

To provide additional context for the proposed functions of $BnKASIII$-C09, $BnBZIP10$ -A09, $BnAPA1$ -A08, $BnLCAT3$ -C05, $BnPAS2$ -A10 and $BnMYB61$ -A06, we used transcriptome data from four accessions across three different environments at 35 DAP to construct co-expression networks. This analysis yielded 12 gene modules. The $BnKASIII$-C09, $BnBZIP10$ -A09, $BnAPA1$ -A08, and $BnLCAT3$ -C05 genes fell in the blue module, which showed a significant positive correlation with oil content ($r$, 0.84; Fig. S5b). The turquoise module including the $BnPAS2$ -A10 and $BnMYB61$ -A06 genes showed a significant positive correlation with oil content ($r$, 0.53; Fig. S5b, Fig. S6a). KEGG enrichment analysis of module genes was performed. The blue module was enriched in fatty acid degradation, unsaturated fatty acid biosynthesis and flavonoid biosynthesis, while the turquoise module was enriched in fatty acid biosynthesis, photosynthesis and flavonoid biosynthesis (Fig. S6b).

A total of 286 genes were identified within the subnetwork, including 125 environment-insensitive DEGs (Fig. 6, Table. S4). The subnetwork included 30, 34, 15 and 46 genes related to lipid/fatty acid metabolic processes, carbohydrate metabolic processes, and flavonoid and plant hormone metabolic pathways, respectively. Further analysis of candidate gene subnetworks revealed that $BnBZIP10$ -A09 is directly linked to $BnKASIII$ -C09, $BnAPA1$ -A08, $BnLCAT3$ -C05 and $BnPAS2$ -A10, while $BnMYB61$ -A06 is directly linked to $BnPAS2$ -A10 (Fig. 6). These results suggest that these genes establish a potential molecular network.
that impacts the accumulation and stability of oil content in rapeseed.

Discussion

Increasing seed oil content is one of the most important targets for rapeseed breeding. Oil content is a complex quantitative trait controlled by multiple genes and easily influenced by environmental factors (Boem et al., 1996; Jensen et al., 1996; Si et al., 2003). To date, extensive investigations have been conducted on numerous QTLs associated with seed oil and fatty acids in rapeseed, including some environment-insensitive QTLs identified across different environments (Zhao et al., 2005; Li et al., 2011; Sun et al., 2016; Teh and Möllers, 2016). Based on the results of previous studies, environment-insensitive genes might exist and regulate oil synthesis in rapeseed.

In this study, we initially investigated the dynamics of oil accumulation at six different developmental stages (20, 25, 30, 35, 40 and 45 DAP) under three distinct environmental conditions. Our findings revealed key stages of oil synthesis occurring between 30 and 35 DAP, which is consistent with the results obtained from previous studies (Yu et al., 2023). To identify genes that exhibit insensitivity to environmental changes through transcriptome analysis of HOC and LOC inbred lines at 35 DAP across three distinct environments. As a result, we discovered 27 environment-insensitive DEGs involved in lipid/fatty acid metabolism, while 23 environment-insensitive DEGs were identified as transcription factors (Fig. 2c), indicating that these genes are environment-insensitive and associated with the accumulation of oil content in rapeseed.

Meanwhile, six out of 50 environment-insensitive DEGs, namely, BnAPA1 -A08, BnPAS2 -A10, BnLCAT3 -C05, Bn KASIII -C09, BnaMYB61 -A06, and BnBZIP10- A09, were also significantly associated with oil content by GWAS. APA1 and LCAT3 are involved in lipid metabolic processes (Noiriel et al., 2004; Gaudet et al., 2011). PAS2 interacts with CER10, a component of the microsomal fatty acid elongase complex, suggesting a role in the very long-chain fatty acid biosynthetic process (Bach et al., 2008). Dehesh et al. (2001) suggested that BnKASIII affects oil content accumulation in rapeseed. BZIPI10 forms a ternary complex with BZIP53 and ABI3 to promote the expression of seed maturation genes and affect seed oil accumulation (Alonso et al., 2009; Fatihi et al., 2016; Kumar et al., 2020). MYB61 positively regulating GL2 inhibits seed oil biosynthesis partly by influencing the formation of mucilage in the seed coat (Shi et al., 2012; Matías-Hernández et al., 2017). Moreover, the co-expression network revealed that BnBZIP10 -A09 is directly linked to BnKASIII -C09, BnAPA1 -A08, BnLCAT3 -C05 and BnPAS2 -A10. BnMYB61 -A06 is directly linked to BnPAS2 -A10 and BnGL2-A07. These results suggest that fatty acid/lipid metabolism and transcription factors potentially interact with and are involved in the co-regulation of the accumulation and stability of seed oil in rapeseed.

The identification of shifts in allele frequencies within the genome can be important information for breeders since it alerts them to monitor specific haplotypes and can be used to design appropriate breeding strategies (Collard and Mackill, 2008). Recent research has demonstrated that favourable haplotypes are associated with improved cold tolerance in rice (Zhang et al., 2017), head blight resistance in wheat (Hao et al., 2012), and drought tolerance in maize (Wang et al., 2016). In this study, we identified six favourable haplotypes corresponding to accessions with relatively high oil content, three of which have additive effects with each other. Voss-Fels et al. (2017) revealed the presence of an additive × additive epistasis effect between the two haplotypes, resulting in a significant increase in root biomass in wheat. An additive effect was found between the two haplotypes, which significantly increased the chlorophyll content (Qian et al., 2016). Multiple interacting genes may change in the same fitness direction, at a similar evolutionary rate and across the same timescale to achieve a common phenotypic outcome (Buntjer et al., 2005). The results of the present study underline the benefit of combining haplotype analysis with GWAS to dissect additive effects of quantitative trait loci in crops. These findings provide a theoretical basis to further improve the accumulation and stability of seed oil in rapeseed.

Acknowledgements

Not applicable.
Author Contributions
All authors planned and supervised the research. MY performed data curation, investigation, and visualization and wrote the original draft. DH, WL, XHX and XH assisted with the data for the investigation. ZSL and CYG provided the resources. LWQ provided funding and performed supervision and led the conceptualization, project administration, resources, writing, review, and editing.

Funding
This study was funded by the National Key Research and Development Program of China (grant no. 2022YFD1200404), the Science Foundation for Distinguished Youth Scholars of Hunan Province, China (grant number: 2022JJ10027) and the Research Foundation of Education Bureau of Hunan Province, China (grant number: 21A0135).

Availability of data and materials
All data generated or analysed during this study are included in this published article and its additional files.

Declarations
Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Reference


Figures

Fig 1: Climate data for three environments tested. The graphs show sunlight (monthly mean values), temperature (monthly mean values), and precipitation (monthly sums) for each environment, the environments are abbreviated as follows: Changsha (CS), Hangzhou (HZ) and Kunming (KM) in China.

Fig 2: Summary DEGs between extremely high- and low-oil content inbred lines. (a) and (b) Number of common up- and down-regulated DEGs in CS, HZ and KM. (c) The heatmap of 39 selected environment-insensitive DEGs from the comparison between HOC and LOC inbred lines in three different environments. The expression values of the candidate genes were calculated using three biological replicates with three technical replicates and normalized by log_{10} of mean expression values. (d) Validation of RNA sequencing data by using quantitative real-time polymerase chain reaction (qRT-PCR). Expression abundance of 39 selected differentially expressed genes at 35 days after pollination is presented as the ratio of HOC and LOC inbred lines. BnEF genes were used as an internal control for data normalization. Normalized expression (2^{[-\Delta \Delta Ct]} of HOC inbred line was divided by normalized expression in LOC inbred line and log_{10} transformed. The correlation coefficient ($R^2$) was calculated for comparison. Orange, green and turquoise line represent CS, HZ and KM environments fitted curve, respectively.*p $\leq$ 0.05, **p $\leq$ 0.01.

Fig 3: The circos plot of 39 selected environment-insensitive DEGs, previous reported QTLs regions and GWAS for oil content. (a) Chromosomes; (b) 39 selected environment-insensitive DEGs; (c) QTLs and association region genetic intervals from published studies; (d) The scatter plot of GWAS for oil content in 50 Chinese semi-winter rapeseed accessions, red point and blue line represent -log_{10} p-value $\leq$ 4.0. Red fill represents significant QTL.

Fig 4: The analysis of candidate genes in the significant associated QTL-A06 and QTL-A09 regions. Regional Manhattan plot surrounding the peak signals on QTL-A06 (a) and QTL-A09 (d). Green dot indicates SNPs located in BnaMYB61-A06 (a) and BnBZIP10-A09 (d) gene regions which significant associated with oil content. Genetic structure variations of BnaMYB61-A06 (b) and BnBZIP10-A09 (e). (c) and (f) Boxplots showed comparative analysis between haplotypes related to oil content phenotype by t-test, *p $\leq$ 0.05, **p $\leq$ 0.01, ***p $\leq$ 0.001.

Fig 5: Additive effect analysis between candidate genes. (a) Circos plot showed candidate genes interaction analysis by R package “SIPT”. The dot in the outer arcs represent -log_{10} (p-value) of GWAS, secondary arcs represent candidate gene structure. Links in the center of the circle represent the additive-by-additive between two SNPs. (b) Boxplots showing the oil content of additive-by-additive haplotypes which is higher than single haplotype.

Fig 6: Co-expression network analysis. Hexagon nodes represent eight candidate genes, rhombus nodes represent environment-insensitive DEGs in three environments. Based on the functional annotation, co-expression network was classified into the following groups: Lipid/fatty acid biosynthetic process (red nodes), transcription factors (Turquoise nodes), flavonoid metabolic pathway (wathet blue nodes), carbohydrate metabolic pathway (brown nodes), photosynthesis (green nodes) and plant hormone (purple nodes).
Supplementary Figures

Fig S1: Oil content phenotype of development period and PCA distribution for LOC and HOC inbred lines using 35 DAPs FPKM. (a) The seed oil content of LOC and HOC inbred lines at different environments in different growth stage, the blue and gray, red and orange lines represent LOC and HOC inbred lines, respectively. (b) PCA distribution for LOC and HOC inbred lines using 35 DAPs FPKM. CS, HZ and KM was signed by cycle, tangle and rhombus, respectively.

Fig S2: Overview of DEGs at 35 DAP seed of the HOC compared to LOC inbred lines. (a) DEGs number of different comparisons at 35 DAP seed. (b) DEGs overlapped in same environment under study. (c) Top 20 KEGG enhancement of common DEGs in three environments.

Fig S3: The analysis of candidate genes in the significant associated QTL-A08 and QTL-A10 regions. Regional Manhattan plot surrounding the peak signals on QTL-A08 (a) and QTL-A10 (d). Green dot indicates the SNPs located in *BnAPA1*-A08 (a) and *BnPAS2*-A10 (d) gene region which associated with oil content. Genetic structure variations of *BnAPA1*-A08 (b) and *BnPAS2*-A10 (e). (c) and (f) Boxplots showed comparative analysis between haplotypes related to oil content phenotype by t-test, *p \[0.05\], **p \[0.01\], ***p \[0.001\].

Fig S4: The analysis of candidate genes in the significant associated QTL-C05 and QTL- QTL-C09.2 regions. Regional Manhattan plot surrounding the peak signals on QTL-C05 (a) and QTL-C09.2 (d). Green dot indicates the SNPs located in *BnLCAT3*-C05 (a) and *BnKASIII*-C09 (d) which associated with oil content. Genetic structure variations of *BnLCAT3*-C05 (b) and *BnKASIII*-C09 (e), numbers indicate the SNP positions from gene start site. (c) and (f) Boxplots showed comparative analysis between haplotypes related to oil content phenotype by t-test, *p \[0.05\], **p \[0.01\], ***p \[0.001\].

Fig S5: The result of co-expression network analysis. (a) Cluster dendrogram of WGCNA gene modules. (b) The information of module-trait coefficient and module gene numbers.

Fig S6: Co-expression network analysis. (a) Whole co-expression network exhibit, hexagon nodes represent eight candidate genes, triangle nodes represent genes directly linked to the candidate gene. The blue and turquoise nodes represent blue and turquoise module genes. (b) Top 20 KEGG enhancement of blue and turquoise module genes.

Supplementary Tables

Table S1: The detailed information of 39 environment-insensitive DEGs.

Table S2: Phenotypic characteristics for oil content in 50 Chinese semi-winter rapeseed accessions.

Table S3: Detailed analysis of haplotypes in candidate genes.

Table S4: Gene information in the co-expression network.

Table S5: Primers used for qRT-PCR.

Table S6: The detailed information of additive affect between haplotypes in 50 Chinese semi-winter rapeseeds accessions.
Hosted file

Table S1.xlsx available at https://authorea.com/users/675333/articles/673396-identification-
of environment-insensitive genes for oil content by a combination of transcriptome and genome-wide association analysis in Brassica napus

Hosted file


Hosted file


Hosted file


Hosted file


Hosted file