Genome Biology

RESEARCH

Open Access

Integrative omics analysis reveals the genetic basis of fatty acid composition in *Brassica napus* seeds

Yuting Zhang[1†], Yunhao Liu[1†], Zhanxiang Zong[2], Liang Guo[1,2*], Wenhao Shen[2*] and Hu Zhao[2*]

[†]Yuting Zhang and Yunhao Liu contributed equally to this work.

*Correspondence: guoliang@mail.hzau.edu.cn; whshen@webmail.hzau.edu.cn; zhaohu@mail.hzau.edu.cn

[1] Yazhouwan National Laboratory, Sanya 572025, China
[2] National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China

Abstract

Background: The fatty acid content represents a crucial quality trait in *Brassica napus* or rapeseed. Improvements in fatty acid composition markedly enhance the quality of rapeseed oil.

Results: Here, we perform a genome-wide association study (GWAS) to identify quantitative trait locus (QTLs) associated with fatty acid content. We identify a total of seven stable QTLs and find two loci, *qFA.A08* and *qFA.A09.1*, subjected to strong selection pressure. By transcriptome-wide association analysis (TWAS), we characterize 3295 genes that are significantly correlated with the composition of at least one fatty acid. To elucidate the genetic underpinnings governing fatty acid composition, we then employ a combination of GWAS, TWAS, and dynamic transcriptomic analysis during seed development, along with the POCKET algorithm. We predict six candidate genes that are associated with fatty acid composition. Experimental validation reveals that four genes (*BnaA09.PYRD*, *BnaA08.PSK1*, *BnaA08.SWI3*, and *BnaC02.LTP15*) positively modulate oleic acid content while negatively impact erucic acid content. Comparative analysis of transcriptome profiles suggests that *BnaA09.PYRD* may influence fatty acid composition by regulating energy metabolism during seed development.

Conclusions: This study establishes a genetic framework for a better understanding of plant oil biosynthesis in addition to providing theoretical foundation and valuable genetic resources for enhancing fatty acid composition in rapeseed breeding.

Background

Rapeseed (*Brassica napus* L.) is an oilseed crop that is globally cultivated [1, 2]. Triacylglycerols, the primary form of oil in rapeseed, are composed of a glycerol backbone and fatty acid (FA) chains [3–5]. FAs are categorized into saturated and unsaturated types based on the hydrocarbon chain's degree of saturation [6]. Saturated FAs tend to accumulate on blood vessel walls and are less easily digested and absorbed by the human body [7–9], while unsaturated FAs are more beneficial to human health, helping reduce the risk of cardiovascular and cerebrovascular diseases [10, 11]. In rapeseed, breeding

efforts have focused on developing double-low (low erucic acid, low glucosinolate) and one-high (high oleic acid) varieties [12, 13].

The FA biosynthetic pathway in rapeseed is a quantitative trait regulated by QTLs, and its FA composition is controlled by many genes with additive and epistatic effects [14–17]. Among the genes and loci for FA breed improvement, researchers identified two major effector loci with additive effects through linkage mapping, which are located on rapeseed chromosomes A08 and C03 [18, 19]. The two homologs of *FAE1*, *BnaA08.FAE1* and *BnaC03.FAE1*, are key genes necessary for encoding the elongation of long-chain FAs and can control the synthesis of erucic acid in rapeseed [20]. Peng et al. silenced the *FAD2* and *FAE1* genes in rapeseed using RNAi technique and found a great increase in oleic acid and a decrease in erucic acid [21]. Wells et al. found that a significant increase in erucic acid content could be achieved by either repressing the expression of the *FAD2* gene in rapeseed or by directly mutating the gene [22, 23]. Additionally, it has been shown that the copy number of the *FAD3* gene affects the content of linolenic acid in seeds [24, 25]. Therefore, high linolenic acid breeding in rapeseed can be achieved by increasing the expression of *FAD3* gene. In rapeseed germplasm resources, the genetic regulatory mechanisms affecting FA compositions have been largely clarified [26, 27]. However, the proportions of FA composition in rapeseed are highly variable, and the construction of new genes affecting FA composition and their regulatory relationship networks requires further research and exploration.

Traditional reverse genetics strategies and map-based cloning approaches are time-consuming and labor-intensive, which makes it difficult to comprehensively explore the key variants and genes affecting FAs [28]. With the continuous development of sequencing technology, genome-wide association study (GWAS) and transcriptome-wide association study (TWAS) have been widely used in various fields of crop research [29–32]. In rapeseed, multi-omics data were used to analyze the genetic basis of seed oil content (SOC), seed coat content (SCC), and seed glucosinolate content (SGC) synthesis, resulting in the cloning of several key genes, such as *Bna.PMT6*, *Bna.CCRL*, *Bna.TT8*, *Bna.GTR2*, and *Bna.TT4* [33–37]. These genes have provided abundant genetic resources for the genetic improvement of rapeseed quality traits. In this study, we comprehensively analyzed the genetic mechanism of FA composition in rapeseed, combining multi-omics data analysis, co-expression network construction, and machine learning algorithms to predict and clone four new genes affecting FA composition. The results enriched the regulatory mechanisms affecting FA composition of rapeseed, as well as provided theoretical basis and genetic resources for rapeseed oil quality improvement.

Results

Fatty acid composition in rapeseed population

The FAs in rapeseed include various saturated and unsaturated types. To investigate the genetic basis of FA composition in rapeseed, this study analyzed the content of seven FAs—palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), linolenic acid (C18:3), eicosapentaenoic acid (C20:1), and erucic acid (C22:1)—across 505 rapeseed accessions in six locations (Wuhan, Ezhou, Chengdu, Hefei, Kunming, and Lanzhou) over 1 to 4 years. The results revealed that unsaturated FAs had a significantly higher average content than saturated FAs in rapeseed, with C18:1 having the highest

mean content (54.52%), followed by C18:2 at 16.59% (Fig. 1a). Pearson's correlation coefficient (PCC) analysis, following best linear unbiased prediction (BLUP) results, identified significant correlations among six of the FAs, excluding C18:3 (Fig. 1b; Additional file 1: Fig. S1). Notably, C22:1 showed a strong positive correlation with C20:1 ($R^2 = 0.94$) and strong negative correlation with C16:0, C18:0, C18:1, and C18:2 ($R^2 = 0.76$, 0.37, 0.97, and 0.67, respectively) (Additional file 1: Fig. S1). The high correlations between C18:1 and both C20:1 and C22:1 are likely due to the use of C18:1 as a substrate for carbon chain extension in the biosynthesis of C20:1 and C22:1. These findings confirm the reliability of the FA composition data and support the synthesis of multiple FA profiles for a comprehensive analysis of the regulatory network governing FA metabolism in rapeseed.

In our previous study, 505 rapeseed accessions were grouped into three subpopulations: semi-winter 1 (SW1), semi-winter 2 (SW2), and spring (SPR) [33]. FA composition analysis showed that SW1 and SPR had similar profiles, with higher levels of C16:0, C18:0, C18:1, and C18:2 compared to SW2, which had higher levels of C18:3, C20:1, and C22:1. Notably, C18:1 content was 40–45% higher in SW1 and SPR than in SW2, while C22:1 content was nearly absent (Additional file 1: Fig. S2).

GWAS reveals key QTLs influencing fatty acid composition

Seven co-localized QTLs were identified as controlling fatty acid composition

Using GWAS with 10,620,048 variants, we analyze seven FAs and their ratios (C18:1/C18:2, C18:1/C18:3, C18:1/C20:1, C18:2/C18:3, C22:1/C18:1, C22:1/C18:2, and C22:1/C18:3) (Additional file 1: Figs. S3, S4; Additional file 2: Table S2), and a total of 169 QTLs were identified. After combining the co-localized QTLs, seven QTLs (named as *qFA.A02*, *qFA.A08*, *qFA.A09.1*, *qFA.A09.2*, *qFA.C02*, *qFA.C03.1*, and *qFA.C03.2*) were consistently detected in GWAS across more than five FAs or six fatty acid ratios, which are distributed on chromosomes A02, A08, A09, C02, and C03, respectively (Fig. 1c, d; Additional file 1: Fig. S5; Additional file 2: Table S3).

(See figure on next page.)

Fig. 1 Co-located loci from the fatty acid genome-wide association analysis and locus selected analysis. **a** FA composition in 505 rapeseed accessions. Fatty acid phenotypic data (palmitic, stearic, oleic, linoleic, linolenic, eicosanoic, and erucic acids) were collected from seeds of 505 accessions at six sites (Wuhan, Ezhou, Chengdu, Hefei, Kunming, and Lanzhou) over 1 to 4 years. Data for each fatty acid were averaged across multiple years and locations for analysis. **b** Correlation of phenotypic variation of FA composition. The Pearson correlation coefficient was calculated using multi-year, multi-location BLUP values for FAs in rapeseed. Light blue lines indicate positive correlation, while dark blue lines indicate negative correlation. The thickness of the lines indicates the strength of the correlations between FAs. **c** Genome-wide association analysis and co-localized QTLs for the composition of six FAs (C16:0, C18:0, C18:1, C18:2, C20:1, and C22:1). **d** Genome-wide association analysis and co-localized QTLs for seven FA ratios (C18:1/C18:2, C18:1/C18:3, C18:1/C20:1, C18:2/C18:3, C22:1/C18:1, C22:1/C18:2, and C22:1/C18:3). The *x*-axis represents the rapeseed chromosomes, while the *y*-axis shows the $\log_{10}$-transformed *P* values obtained from GWAS. The dotted line marks the significance threshold. By integrating GWAS of different FA compositions and ratios, the co-localized QTLs *qFA.A02*, *qFA.A08*, *qFA.A09.1*, *qFA.A09.2*, *qFA.C02*, *qFA.C03.1*, and *qFA.C03.2* were identified. **e** The $\log_{10}$-transformed *P* values form C18:1 GWAS plotted against those form seed glucosinolate content (SGC) GWAS. **f** Signal detection intensity for *qFA.A09.1* in GWAS, showing nucleotide diversity (π) with LOWESS smoothed curves. **g** Manhattan plot showing 2 Mb region upstream and downstream of the *qFA.A09.1*. Colors indicate the LD (R^2) between the lead SNP and each variant in GWAS

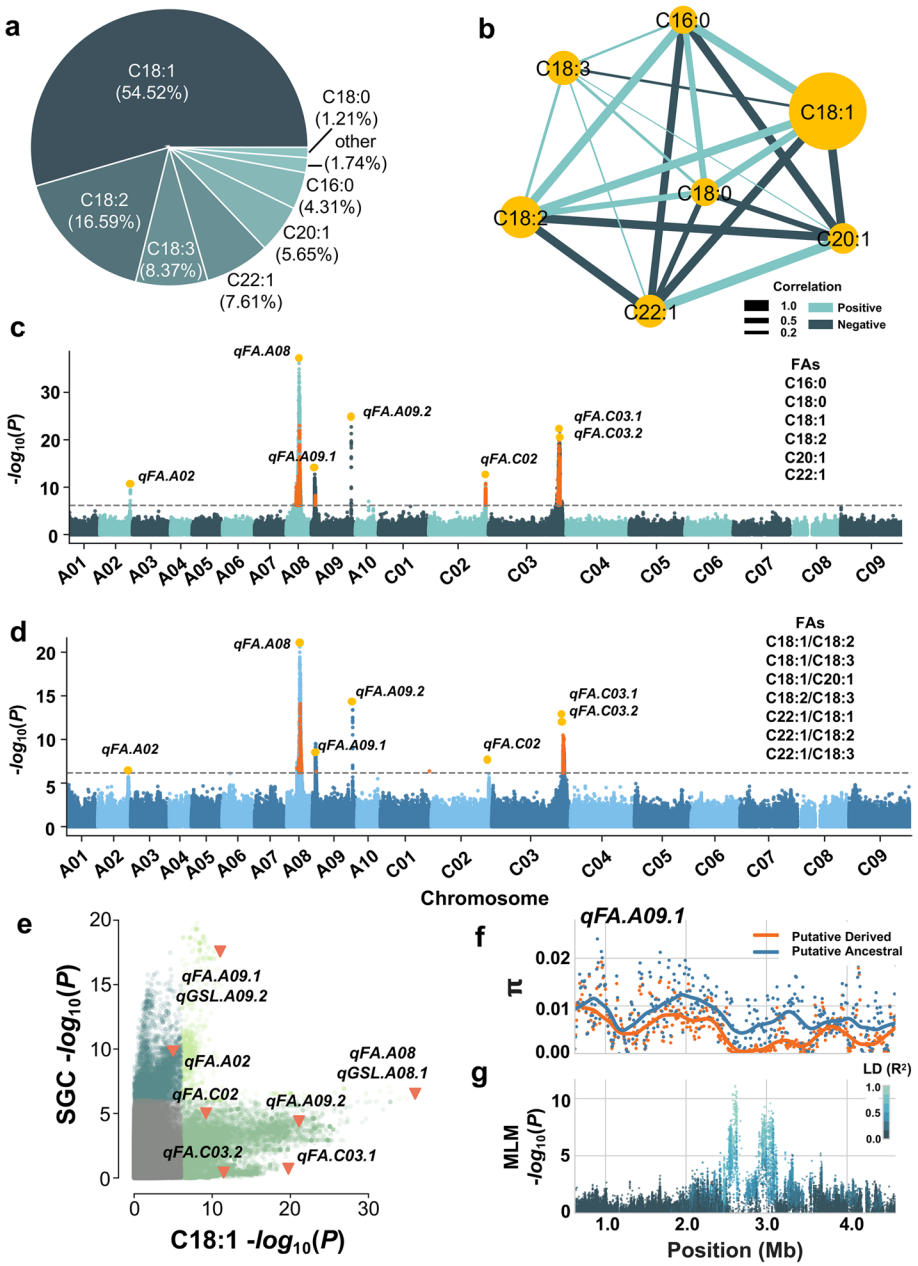

Fig. 1 (See legend on previous page.)

Candidate genes located within 100 kb of the lead SNPs in the seven stable QTLs were identified, yielding 1412 potential GWAS candidate genes. These included key genes involved in FA synthesis, such as those encoding 3-ketoacyl-CoA synthase (KCS) enzymes (*Bna.FAE1* and *Bna.KCS3*) for FA elongation, *Bna.LACS* for long-chain FA synthesis, *Bna.FAD3* for linolenic acid synthesis, and *Bna.LPAAT5* for triacylglycerol formation (Table 1). Interestingly, several flavonoid metabolism-related genes were also identified, including *Bna.CCoAOMT*, which is involved in lignin biosynthesis, and genes from transparent seed coat family (*Bna.TT6* and *Bna.TT16*), which influence seed coat color. These findings suggest a complex regulatory network

Table 1 Seven co-localized QTLs and FA-related genes identified by genome-wide association analysis

QTL	Position[a]	Gene ID[b]	Name	Description	Reference
qFA.A02	24,099,591	BnaA02g33410D	*MYB96*	Myb-related protein	[38]
qFA.A08	10,151,436	BnaA08g11130D	*FAE1*	3-ketoacyl-CoA synthase 18	[20]
		BnaA08g11140D	*KCS17*	3-ketoacyl-CoA synthase 17	[39]
qFA.A09.1	2,612,389	BnaA09g04580D	*SHN3*	Ethylene-responsive factor SHINE 3	[40]
		BnaA09g06090D	*MYB96*	Myb-related protein	[38]
qFA.A09.2	31,542,153	BnaA09g46210D	*AAPT1*	Ethanolaminephosphotransferase 1	[41]
qFA.C02	44,879,480	BnaC02g42240D	*TT16*	Transparent testa 16	[42]
qFA.C03.1	55,619,752	BnaC03g65980D	*FAE1*	3-ketoacyl-CoA synthase 18	[20]
qFA.C03.2	56,398,549	BnaC03g66040D	*KCS17*	3-ketoacyl-CoA synthase 17	[39]
		BnaC03g77180D	*SUD1*	E3 ubiquitin ligase SUD1	[43]

[a] Physical position of the lead SNP

[b] Gene ID associated with FAs in QTL

in rapeseed seed development and underscore the genetic diversity underlying FA composition.

Strong selection pressure of *qFA.A09.1* and *qFA.A08* during rapeseed domestication.

In the rapeseed breeding, seeds with low C22:1, high C18:1, and low glucosinolate content are key targets. Our study has identified two co-localized QTLs in the GWAS results for seed glucosinolate content (SGC) [35] and FAs [$-\log_{10}(P_{qGSC.A08.1}) = 6.53, -\log_{10}(P_{qFA.A08}) = 35.98, -\log_{10}(P_{qGSC.A09.2}) = 17.56, -\log_{10}(P_{qFA.A09.1}) = 10.99$] (Fig. 1e); this suggests a co-selection for glucosinolate and FA traits.

To further explore the selection of loci controlling FAs, ancestral alleles in the rapeseed genome were inferred using resequencing data from *B. oleracea* and *B. rapa*. Previous studies confirmed that *qFA.A08* (which corresponds to *qOC.A08.1* in our previous research) [33], including *BnaA08.FAE1*, a key regulator of C22:1 synthesis in rapeseed, has undergone significant selection. Comparative analysis of the nucleotide diversity (π) between the ancient and derived haplotypes of the remaining six co-localized QTLs revealed that *qFA.A09.1* experienced strong selection during domestication (Fig. 1f, g). These findings suggest that *qFA.A08* and *qFA.A09.1* were crucial for breeding high oleic and low erucic acid rapeseed. However, other QTLs associated with erucic and oleic acid content have remained underutilized in rapeseed breeding (Additional file 1: Fig. S6). Therefore, further exploration of QTLs and new genes involved in fatty acid composition is essential to enhance rapeseed improvement.

TWAS revealed the molecular basis of seed fatty acid composition

TWAS correlates gene expression with phenotype at the population level. To identify genes influencing the FA content, we conducted TWAS on seven FA compositions using expression data from two time points. A total of 3295 genes were significantly associated with the FAs (FDR < 0.05), with 1576 genes identified at 20 days after flowering (DAF) and 2196 at 40 DAF (Fig. 2a). The highest number of significantly related genes was found in C18:1, followed closely by C20:1 and C22:1. C18:3 had almost no significant associations, while the other six FAs had 30 overlapping significant genes identified by TWAS (Fig. 2b).

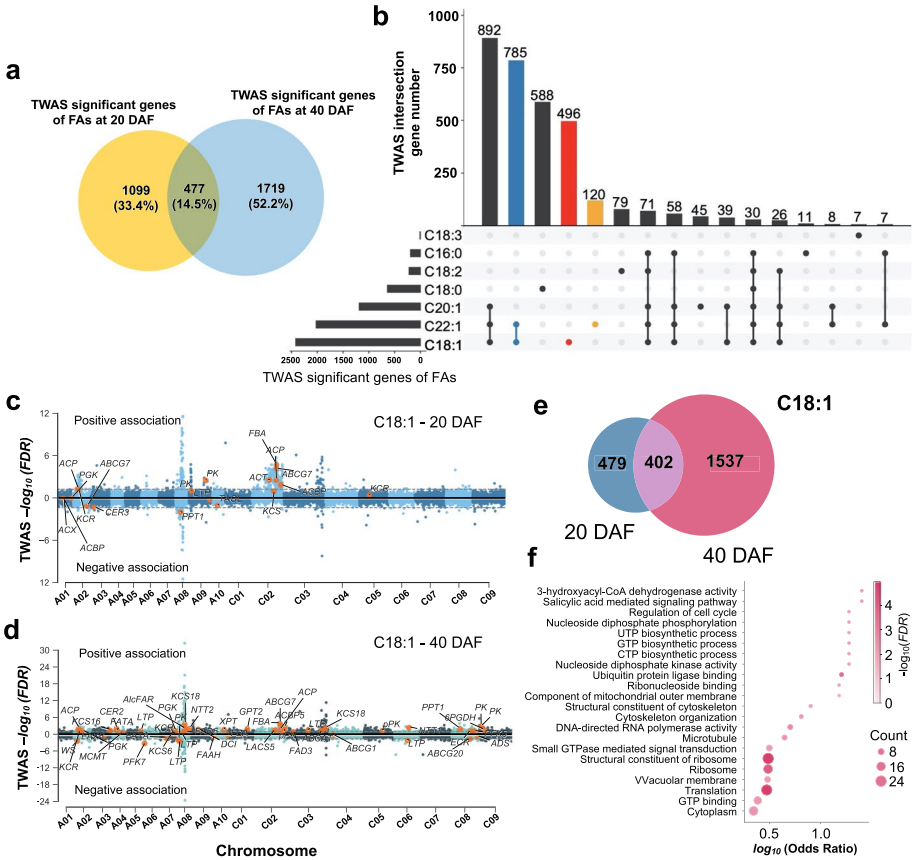

Fig. 2 Transcriptome-wide association studies of fatty acids. **a** Venn diagram shows TWAS significant genes for FAs at 20 and 40 days after flowering (DAF). **b** An upset plot displays TWAS significant gene sets associated with the FAs. The bar on the left represents the original number of the significant genes for seven FAs, while the bar above shows the number of intersecting genes. Circles and connecting lines below provide the phenotypic information for calculating the intersecting genes. **c** Manhattan plot of C18:1 at 20 DAF. **d** Manhattan plot of C18:1 at 40 DAF. Each point represents a gene, with those positively or negatively associated with C18:1 depicted above or below the thick black line, respectively. Gold dots indicate significant lipid metabolism genes previously identified in *Arabidopsis*. **e** Venn diagram shows the significant C18:1 associated genes from TWAS at 20 and 40 DAF. **f** GO enrichment analysis of significant overlapping C18:1 associated genes at 20 and 40 DAF. The *y*-axis represents the names of the enriched pathways. The count value indicates the number of overlapped genes in each pathway, while the *FDR* value reflects the significance of enrichment

Analyzing the functions of the 3295 genes significantly associated with FAs, we identified several genes related to lipid synthesis. These include ABC transporter family genes (*Bna.ABCG1*, *Bna.ABCG7*, *Bna.ABCG10*, and *Bna.ABCG25*), acyl carrier proteins (*Bna.ACP*), acyl CoA thioesterase genes (*Bna.ACT*), FA desaturases (*Bna.FAD3* and *Bna.FAD6*), keto-CoA synthase and reductase genes (*Bna.KCS6*, *Bna.KCS16*, *Bna.FAE1*, and *Bna.KCR*), long-chain acyl-CoA synthetase gene (*Bna.LACS5*), pyruvate kinase gene (*Bna.PK*), and lipid transfer proteins (*Bna.LTP*, *Bna.PPT1*, *Bna.NTT1*, and *Bna.NTT2*) (Fig. 2c, d; Additional file 1: Fig. S8a, b; Additional file 2: Tables S8, S9). These findings confirm the accuracy of the TWAS approach in identifying candidate genes. GO enrichment analysis of the 477 overlapping significant genes at both time points revealed involvement in processes such as

3-hydroxyacetyl-coenzyme, dehydrogenase activity, salicylic acid-mediated signaling, and cation transport (Additional file 1: Fig. S7a). At 20 DAF, the 1576 significant genes were mainly enriched in pathways related to nucleotide biosynthesis, salicylic acid signaling, protein kinase regulation, translation initiation factors, and fatty acid acyl-coenzyme A binding (Additional file 1: Fig. S7b). At 40 DAF, the 2196 significant genes were associated with nucleotide binding, glycolysis, energy metabolism, and other nucleotide-related processes (Additional file 1: Fig. S7c). These results suggest that fatty acid biosynthesis is closely linked with energy metabolism [44, 45].

C18:1 and C22:1 are key targets in rapeseed breeding. We identified 881 and 1939 significant genes by TWAS for C18:1 at 20 DAF and 40 DAF, respectively, with 402 overlapping genes (Fig. 2e; Additional file 2: Tables S4, S5). These genes were primarily enriched in pathways related to nucleotide phosphorylation, hydroxyacyl-coenzyme A dehydrogenase activity, salicylic acid signaling, and nucleotide biosynthesis (Fig. 2f). For C22:1, 2386 significantly genes were identified, with 364 genes overlapping between the two periods, which were involved in processes such as phosphotransferase activity, starch catabolism, and energy metabolism (Additional file 1: Fig. S8c, d; Additional file 2: Tables S6, S7). The TWAS results for oleic and erucic acids showed a similar distribution pattern across the two time points, with a high correlation between gene associations for both FAs at 20 DAF and 40 DAF (Additional file 1: Fig. S9). Genes such as *Bna.ACP*, *Bna.KCS*, *Bna.LTP*, *Bna.FATA*, *Bna.LACS5*, and *Bna.PPT1* were significantly associated with both FAs, showing positive correlations with C18:1 and negative correlations with C22:1 (Fig. 2c, d; Additional file 1: Fig. S8a, b; Additional file 2: Tables S8, S9).

Prediction of six key candidate genes affecting fatty acids

To identify genes influencing FA composition, we analyzed the dynamic gene expression profiles of the rapeseed cultivar Zhongshuang11 (ZS11) from 2 to 60 DAF, which resulted in the classification of eight expression modules (Fig. 3a). Enrichment analysis of significant genes from GWAS and TWAS for seed FA composition and SGC revealed that four modules—C1, C3, C6, and C8—were enriched with significant TWAS genes of FA and SGC (Fig. 3b).

In order to further predict candidate genes affecting the FA composition in rapeseed, we prioritized genes within 100 kb upstream and downstream of seven FA-related QTLs using POCKET [33]. By integrating variation effect (VE), haplotype effect (HE), expression effect (EE), and gene function prediction, we identified *BnaA09.PYRD*, *BnaA08.SWI3*, *BnaA09.PDR2*, and *BnaC03.SLOMO*, as top candidates based on POCKET scoring (Fig. 4a; Additional file 1: Figs. S13a, S15a, S16a), Additionally, *BnaA08.PSK1*, a TWAS-significant gene, located in the C1 or C6 module, exhibits similar expression patterns with several fatty acid synthesis-related genes (Fig. 3c, d). And *BnaC02.LTP15*, found within the *qFA.C02* locus and part of the C3 module (Additional file 1: Fig. S14a), shows expression patterns similar to genes involved in seed coat development and lipid synthesis, such as *TT9*, *LPAAT*, and *DGD2*, indicating its possible involvement in lipid metabolism. Taken together, we hypothesize that *BnaA09.PYRD*, *BnaA08.PSK1*, *BnaA08.SWI3*, *BnaC02.LTP15*, *BnaA09.PDR2*, and *BnaC03.SLOMO* are key candidates influencing FA composition in rapeseed.

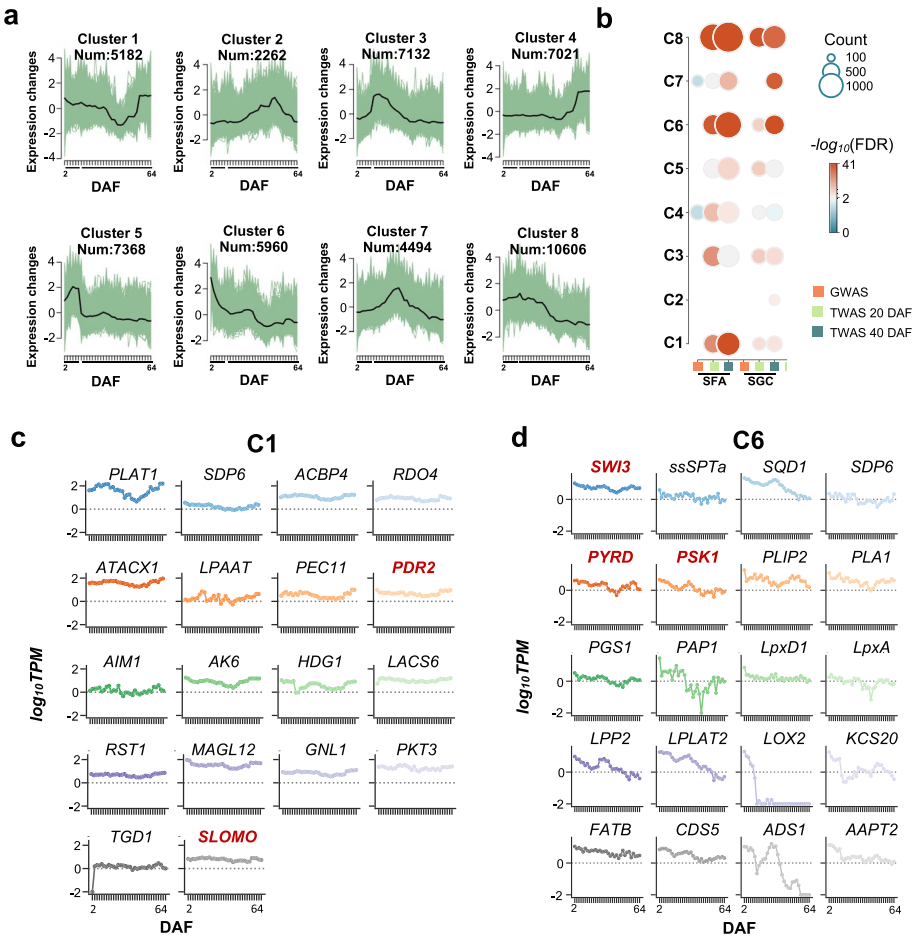

Fig. 3 Identification of candidate gene expression patterns during ZS11 seed development. **a** Expression modules of genes across ZS11 seed development stages from 2 to 64 DAF. **b** Bubble diagram of ZS11 development from 2 to 64 DAF, including modules and seed FA (SFA) and glucosinolate content (SGC). **c** Expression patterns of candidate genes and FA associated genes in module C1. **d** Expression patterns of candidate genes and FA associated genes in module C6

BnaA09.PYRD and *BnaA09.PDR2* were located within *qFA.A09.1* locus (Fig. 4b; Additional file 1: Fig. S15b), while *BnaA08.PSK1* and *BnaA08.SWI3* showed the strongest GWAS signals from *qFA.A08* locus (Additional file 1: Figs. S12a, S13b). Additionally, *BnaC02.LTP15* was found within *qFA.C02* locus (Additional file 1: Fig. S14b), and *BnaC03.SLOMO* within the *qFA.C03.1* locus (Additional file 1: Fig. S16b). To assess the impact of SNP on protein function, we identified three nonsynonymous variants in both *BnaA09.PYRD* (BnvaA0902629811, C/T, $P=1.95\times10^{-12}$; BnvaA0902628936, G/C, $P=3.77\times10^{-7}$; BnvaA0902628958, T/C, $P=3.80\times10^{-7}$) and *BnaA09.PDR2* (BnvaA0902561985, A/G, $P=6.66\times10^{-9}$; BnvaA0902556267, G/C, $P=3.65\times10^{-8}$; BnvaA0902561869, C/T, $P=2.02\times10^{-5}$) genes (Fig. 4c; Additional file 1: Fig. S15c). In contrast, *BnaC03.SLOMO* has 14 nonsynonymous variants (Additional file 1: Fig. S16c), while *BnaA08.PSK1* and *BnaA08.SWI3* each have 11 nonsynonymous variants (Additional file 1: Figs. S12b, S13d). Notably, a frameshift mutation (BnvaA0810225954, ATC/A, $P=1.56\times10^{-21}$) was detected in exon of *BnaA08.SWI3*, potentially causing a significant alteration in protein function (Additional file 1: Fig. S13c).

a

Gene	Symbol	VE	EE	HE	GF	SS
BnaA09g05380D	PYRD	0.92	1.00	0.95	0.64	2.46
BnaA09g05370D	SOD3	0.20	0.01	1.00	0.69	1.53
BnaA09g05200D	PDR2	0.58	0.47	0.58	0.38	0.84
BnaA09g05190D	SWEET12	0.47	0.01	0.57	0.40	0.64
BnaA09g05310D	HMGB6	0.65	0.06	0.51	0.25	0.43
BnaA09g05270D	CEX18	0.00	0.01	0.28	0.52	0.29
BnaA09g05280D	/	1.00	0.09	0.63	0.11	0.25
BnaA09g05170D	RUP2	0.30	0.03	0.57	0.11	0.16
BnaA09g05230D	/	0.34	0.02	0.46	0.13	0.16
BnaA09g05180D	/	0.47	0.00	0.45	0.11	0.14

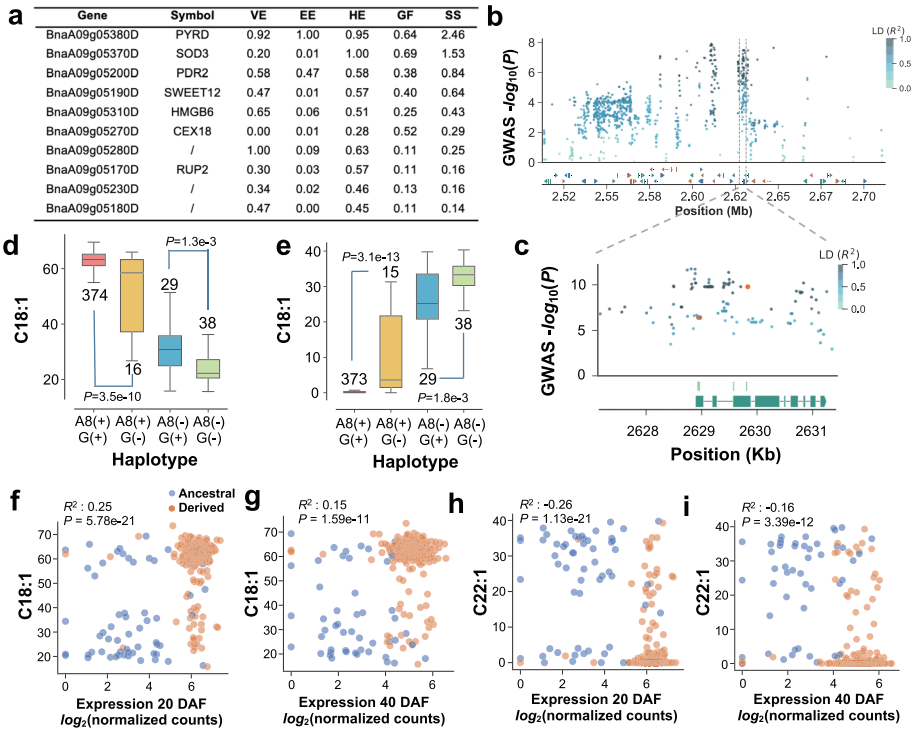

Fig. 4 Characterization of *BnaA09.PYRD* at the population level. **a** List of top ten genes ranked within the QTL *qFA.A09.1*. **b** Manhattan plot and gene information for the 100 kb genomic region upstream and downstream of the *qFA.A09.1* locus. The dots indicate variants within the region, with the color of each dot indicating the LD (R^2) between the variants and the lead SNP. The gray dashed line marks the position of the candidate gene's position, and the gene structure in the locus is displayed at the bottom. **c** GWAS results around *BnaA09.PYRD*, with nonsynonymous variants marked by gold dots. **d**–**e** Box plots of the effects of haplotypes on C18:1 and C22:1 constructed from the variants within the *BnaA09.PYRD* and the upstream 2 kb region. (+) indicates the presence of effects at the *qFA.A08* locus or gene, (-) indicates the absence of effects at the *qFA.A08* locus or gene (*BnaA09.PYRD*). The center line of the box plot is the median C18:1 or C22:1 content in each haplotype, with the upper and lower edges of the box showing the respective quartiles. Dots represent the values of C18:1 or C22:1 content for each variety. The MLM module from EMMAX software was applied to calculate the effect of haplotypes. **f**–**i** Correlation analysis of gene expression values at 20 and 40 DAF with C18:1 and C22:1 content. Accessions were grouped by the genotypes of the lead SNP of *qFA.A09.1*, colored by the two accession groups. The *x*-axis indicates the normalized expression value, while the *y*-axis represents phenotype value

To examine the association between haplotypes and fatty acid composition, we classified haplotypes based on candidate genes and their 2 kb promoter region variations, and using the lead SNP of *qFA.A08* as a covariate to exclude the effect of *FAE1*. The results showed that, with the *qFA.A08* locus fixed, different haplotypes of *BnaA09.PYRD*, *BnaC02.LTP15*, *BnaA09.PDR2*, and *BnaC03.SLOMO* had significant effects on C18:1 and C22:1 (Fig. 4d, e; Additional file 1: Figs. S14c–d, S15d–e, S16d–e). Previous analyses indicated that *qFA.A08* and *qFA.A09.1* were significantly selected during fatty acid domestication. Under the condition of consistent variation at the *qFA.A08* locus, variations at the *qFA.A09.1* locus resulted in significant differences in C18:1 and C22:1 (Additional file 1: Fig. S11). Further observation revealed that during the selection process, the expression level of the key gene *BnaA09.PYRD* at the *qFA.A09.1* locus was increased. In derived varieties, high expression level of *BnaA09.PYRD* is often associated with higher C18:1 and lower C22:1, while the opposite trend is observed in ancient

varieties (Fig. 4f–i). This suggests that the changes in fatty acid composition at the *qFA.A09.1* locus are mediated through the regulation of *BnaA09.PYRD* gene expression.

Transcriptional impact of candidate genes on fatty acid composition

To assess the effect of the six candidate genes on FA composition, we analyzed the gene expression at 20 DAF and 40 DAF in the population, correlating them with C18:1 and C22:1. Four genes showed a positive correlation with C18:1 and a negative correlation with C22:1. *BnaA08.SWI3* exhibited the strongest correlation, with R^2 values of 0.53 and 0.55 for C18:1 at 20 DAF and 40 DAF, respectively ($P_{20\ DAF} = 2.87 \times 10^{-46}$; $P_{40\ DAF} = 6.54 \times 10^{-49}$) (Additional file 1: Fig. S13e, f), and for C22:1 were 0.55 and 0.58 ($P_{20\ DAF} = 3.\ 34 \times 10^{-48}$; $P_{40\ DAF} = 3.75 \times 10^{-52}$) (Additional file 1: Fig. S13g, h). *BnaA09.PYRD*, *BnaC02.LPT15*, and *BnaA09.PDR2* showed higher correlations at 20 DAF (Additional file 1: Figs. S10, S14e–h, S15f–i), while *BnaA08.PSK1* and *BnaC03.SLOMO* were more significant at 40 DAF (Additional file 1: Figs. S12c–f, S16f–i).

BnaPYRD, *BnaPSK1*, *BnaSWI3*, and *BnaLTP15* act as negative regulator of erucic acid.

To verify the functions of six genes, transgenic rapeseed plants were generated using cv. WH3411, characterized by low C18:1 and high C22:1, as the receptor material. *BnaPYRD* and *BnaPSK1* mutants (CRISPR/Cas9) and overexpression lines, along with *BnaSWI3* and *BnaLTP15* mutant lines (CRISPR/Cas9), were successfully constructed (Additional file 1: Figs. S17–S20). Compared with the wild type, C22:1 content in the seeds of *BnaPYRD* overexpression lines OE-19 and OE-5 decreased by 9.75% and 23.15%, respectively, while C18:1 content increased by 6.0% and 21.68% (Fig. 5a). In contrast, C22:1 content in the seeds of the quadruple mutant line *pyrd-L35* and the triple mutant line *pyrd-L1* increased by 11.08% and 6.52%, with a corresponding decrease in C18:1 content by 11.99% and 6.71% (Fig. 5a). The C22:1 content in the seeds of *BnaPSK1* overexpression lines OE-12 and OE-14 decreased by 15.68% and 7.79%, respectively, with an increase in C18:1 content of 9.94% and 3.40% (Fig. 5b). Conversely, C22:1 content in the seeds of the double mutant line *psk1-L5* and the triple mutant *psk1-L26* increased by 14.82% and 14.01%, respectively, accompanied by a decrease in C18:1 content of 23.02% and 16.57% (Fig. 5b). The *BnaSWI3* quadruple mutant strains *swi3-L3*, *swi3-L10*, and double mutant strain *swi3-L27* exhibited an increase in C22:1 content of 8.46 to 11.24% and a decrease in C18:1 content of 8.01 to 15.85% (Fig. 5c). In *BnaLTP15* double mutant lines *ltp15-L3*, *ltp15-L11*, and *ltp15-L20* C22:1 content increased by 7.19 to 16.03%, and the C18:1 content decreased of 6.79 to 18.58% (Fig. 5d). In addition, the content of C20:1 in the seeds of these four gene mutant lines was reduced to varying degrees (Fig. 5a–d). Furthermore, the transgenic plants, both gain-of-function and loss-of-function mutants, showed no significant differences in overall growth compared with control plants (Additional file 1: Fig. S21). Collectively, our findings support the potential of these genes as promising targets for breeding programs aimed at improving seed oil quality in *B. napus*.

BnaA09.PYRD may influence fatty acid composition by modulating energy metabolism during rapeseed development.

In rapeseed fatty acid breeding, *qFA.A09.1* locus is under strong selection, making it crucial to identify genes affecting FA content within this segment. We predicted that *BnaA09.PYRD* might be a causal gene influencing FA composition. To further

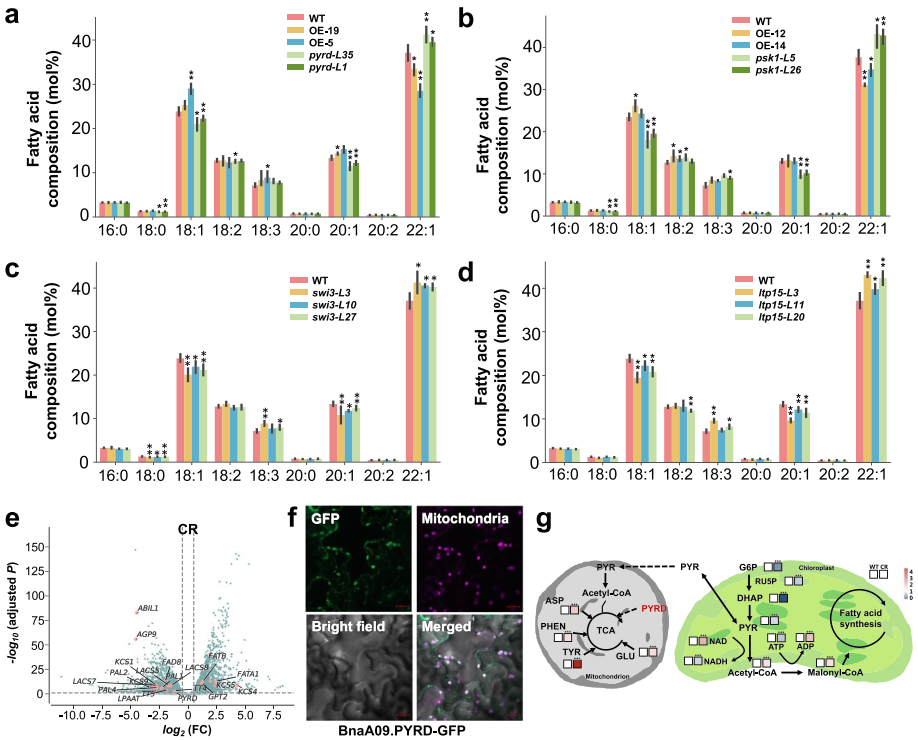

Fig. 5 Functional characterization of four genes in rapeseed. FAs composition of *BnaPYRD* (**a**), *BnaPSK1* (**b**), *BnaSWI3* (**c**), and *BnaLTP15* (**d**) in wild-type (WT) and mutant lines ($n = 3$–6). * indicates $P < 0.05$, ** indicates $P < 0.01$ in Student's *t*-test. **e** Volcano plot of differential expression genes (DEGs) in *bna.pyrd* mutants compared to WT. The *x*-axis indicates the fold change (FC) of gene expression after log2-transformation and the *y*-axis indicates the adjusted *P* value. The gray line indicates log2(FC) = 1, and red points highlight DEGs related to lipid metabolism. **f** Subcellular localization of BnaA09.PYRD protein using BnaA09.PYRD-GFP in tobacco leaves. Green fluorescence indicates the target gene localization, red fluorescence indicates the mitochondrial marker, and the merged image shows co-localization. **g** A proposed working model for the function of *BnaA09.PYRD* in the regulation of FA composition. Heatmaps illustrating the levels of carbon metabolites in WT and *bna.pyrd* mutants. Statistical significance was calculated using Student's *t*-tests, * indicates $P < 0.05$ and ** indicates $P < 0.01$. Abbreviations include G6P (glucose-6-phosphate), RU5P (D-ribulose-5-phosphate), DHAP (dihydroxyacetone phosphate), PYR (pyruvate), ATP (adenosine triphosphate), ADP (adenosine diphosphate), NAD (nicotinamide adenine dinucleotide), NADH (β-nicotinamide adenine dinucleotide), Acetyl-CoA (acetyl-coenzyme A), Malonyl-CoA (malonyl-coenzyme A), ASP (aspartic acid), PHEN (phenylalanine), TYR (threonine), and GLU (glutamic acid)

understand how *BnaA09.PYRD* modulates FA synthesis, transcriptome sequencing was conducted on seeds of WT and *pyrd-L35* at 35 DAF. A total of 51,276 expressed genes (TPM > 1) and 5398 differentially expressed genes (DEGs) ($|\log_2$ (fold change)$| > 1$ and $P_{adj} < 0.05$) were identified (Additional file 2: Table S10). These encompassed numerous genes associated with FA synthesis, including FA desaturase genes (*FAD8*, *FATA1*, and *FATB*), 3-ketoacyl-CoA synthetase genes (*KCS1*, *KCS4*, and *KCS5*), and several long-chain acyl-CoA synthetase genes (*LACS5*, *LACS7*, and *LACS8*) (Fig. 5e). GO enrichment analysis indicated that these DEGs linked to lipid synthesis were mainly enriched in flavonoid biosynthesis pathways, galactolipid and phospholipid biosynthesis, FA elongation, wax biosynthesis, and lignin mucin-associated pathways (Additional file 1: Fig. S22). The combined experimental results suggest that *BnaA09.*

PYRD may modify FA compositions by modulating the conversion of C18:1 to C18:2 and the elongation of C18:1 to long-chain unsaturated FAs.

To investigate the localization of BnaA09.PYRD protein in cells, GFP-fused expression vectors containing the gene were constructed and transformed into tobacco lower epidermal cells for visualization. Under confocal microscopy, BnaA09.PYRD localized in mitochondria (Fig. 5f). Moreover, primary metabolite analysis of the *pyrd-L35* mutant revealed significant changes in the levels of several metabolites and amino acids in the tricarboxylic acid cycle (TCA) and the Calvin-Benson cycle (Fig. 5g). Based on these results, we hypothesize that *BnaA09.PYRD* regulates energy metabolism during rapeseed development, thereby affecting FA composition.

The optimal haplotype of *BnaA08.PSK1* remains underutilized in *B. napus* breeding.

We categorized the accessions from the previous study into three breeding periods: "Before 1980," "1980–2000," and "After 2000" [46, 47]. Haplotypes for four functional genes were classified based on variants within the 2 kb promoter and gene region. The haplotype associated with the highest C18:1 and the lowest C22:1 levels was considered the optimal haplotype (Additional file 1: Fig. S23). The proportion of optimal haplotypes in *BnaA08.SWI3*, *BnaA09.PYRD*, and *BnaC02.LTP15* increased significantly over time. After 2000, the optimal haplotype for all three genes become the major haplotype (Fig. 6a–c). In contrast, *BnaA08.PSK1* haplotype exhibited a different distribution. The proportion of optimal haplotypes slightly increased from "Before 1980" to "1980–2000." Compared to "1980–2000," the proportion of optimal haplotypes in the "After 2000" remained nearly unchanged. The proportion of suboptimal haplotypes (Hap B) of *BnaA08.PSK1* gradually increased becoming the major haplotype (Fig. 6d, e).

In addition to temporal changes in haplotypes over time, we also categorized the accessions by geographical region. The optimal haplotypes of *BnaA08.SWI3* and

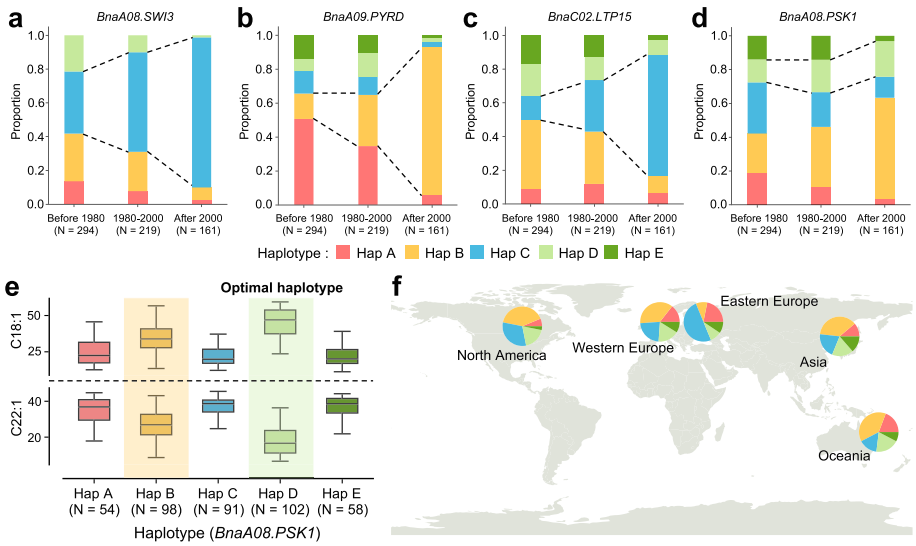

Fig. 6 Haplotype frequency variation of four genes across breeding periods. Proportion of *BnaA08.SWI3* (**a**), *BnaA09.PYRD* (**b**), *BnaC02.LTP15* (**c**), and *BnaA08.PSK1* (**d**) haplotypes across breeding periods. Distinct colors indicate specific haplotypes. Dashed lines link optimal haplotypes for C18:1 and C22:1. Haplotype effects on C18:1 and C22:1 (**e**) and geographic distribution (**f**) across 418 accessions, derived from variants within *BnaA08.PSK1* and its upstream 2 kb region

BnaA09.PYRD were predominant across regions, except in Eastern Europe (Additional file 1: Fig. S24a, b). The optimal haplotype of *BnaC02.LTP15* was less frequent in North America and Oceania (Additional file 1: Fig. S24c). The optimal haplotype of *BnaA08.PSK1* was not widely adopted in any region (Fig. 6f). These results suggest that *BnaA08.SWI3*, *BnaA09.PYRD*, and *BnaC02.LTP15* have been actively utilized for improvement in *B. napus* breeding. However, the optimal haplotype of *BnaA08.PSK1* remained underutilized.

Discussion

Mining fatty acid related QTLs in rapeseed

FA metabolism in rapeseed is highly complex. To address this complexity, integrating bioinformatics with genetic tools and molecular breeding design can generate new genetic resources for FA improvement. Our study localized QTLs using GWAS, and the results align with previous research, showing that QTLs were primarily distributed on chromosomes A08, A09, C02, and C03 (*qFA.A08*, *qFA.C09.1*, *qFA.C02*, *qFA.C03.1*, *qFA.C03.2*, respectively) [48, 49]. Notably, we identified two novel QTLs (*qFA.A02* and *qFA.A09.2*) that had not been previously reported (Fig. 1c, d; Table 1). However, some key enzymes affecting FA metabolism, such as FAD2, FAD3, and SAD, were not detected in the GWAS, probably due to limitations in population diversity, which hindered the identification of relevant variants. Analysis of the seven fatty acid-related QTLs revealed that *qFA.A08* and *qFA.A09.1* were strongly selected during breeding (Fig. 1e–g). Although the genes affecting fatty acids at *qFA.A09.1* have not been reported, it is hypothesized that selection of the *BnaA09.PYRD* gene at this locus may reduce erucic acid content and increase oleic acid content. In contrast, major variants in the other five QTLs were not significantly selected (Additional file 1: Fig. S6), suggesting that further exploration of QTLs and novel genes regulating FA composition is essential in rapeseed.

Integrating key candidate genes affecting fatty acids through multi-omics approaches

During GWAS based on 505 accessions of *B. napus*, LD decay was slow, and the QTL identified often had dozens of candidate genes [50]. There were always many false-positive loci in the results. A single GWAS cannot satisfy the construction of gene networks. To overcome these challenges, we integrated GWAS, TWAS, and transcriptomic profiles from seed development to comprehensively construct a multi-omics gene network influencing FA composition in rapeseed. This network comprised seven independent QTLs and 3295 TWAS-significant genes (Additional file 1: Fig. S25), including key lipid synthesis genes such as *Bna.ACP*, *Bna.KCS*, *Bna.LTP*, *Bna.FATA*, *Bna.LACS5*, and *Bna.PPT1* (Additional file 2: Tables S8, S9). This network plays a crucial role in regulating FA composition in rapeseed, contributing to a deeper understanding of the regulatory mechanisms governing FA metabolism.

Mining the regulatory mechanism of candidate genes affecting fatty acid composition

Our study successfully identified four genes influencing FA composition in rapeseed through functional validation (Fig. 5a–d). Mutations in these genes can significantly enhance C22:1 content. Given C22:1's substantial industrial value, increasing its content is essential to addressing the supply shortage of high-erucic-acid rapeseed

varieties. Research on these genes (*BnaA08.SWI3*, *BnaA09.PYRD*, *BnaC02.LTP15*, and *BnaA08.PSK1*) offers insights into the biosynthetic mechanisms of C22:1. One of these genes, *PYRD*, encodes a dehydrogenase involved in mitochondrial energy metabolism and the fourth step of pyrimidine biosynthesis [51]. Previous studies have shown that *PYRD* mutants display reduced pyrimidine metabolites, lower energy status, decreased photosynthetic capacity, and accumulation of reactive oxygen species. In this study, we hypothesized that the *BnaA09.PYRD* gene may affect FA composition by inhibiting FA elongation. In addition, seed transcriptome analysis indicated that *BnaA09.PYRD* may influence FA metabolism and flavonoid metabolism pathways, both of which are crucial for carbon allocation in rapeseed (Fig. 5g). However, no current evidence suggests that *BnaA09.PYRD* interacts with key transcription factors (such as *LEC1*, *LEC2*, and *WRI1*) or protein complexes to affect FA synthesis. Future research could leverage machine learning and deep learning models, along with chromatin accessibility data and eQTLs, to identify significant genomic variants and explore transcription factor sequences for conserved or specific coregulatory elements, with experimental validation [52–54].

Breeding applications of the optimal haplotypes in four functional genes

Breeding goals for *B. napus* have shifted across periods. In the 1970s–1980s, double-low breeding was the primary objective. From the 1990s to the 2000s, the emphasis transitioned to achieving high yield and resistance [55]. Throughout the breeding process, the optimal haplotype of *BnaA08.SWI3* in "Before 1980" varieties emerged as predominant (Fig. 6a). The proportion of optimal haplotypes has gradually increased with ongoing breeding efforts, demonstrating selection of the optimal *BnaA08.SWI3* haplotype in double-low breeding. Conversely, the optimal haplotypes of *BnaA09.PYRD*, *BnaC02.LTP15*, and *BnaA08.PSK1* were underrepresented in "Before 1980" varieties, suggesting that these optimal haplotypes were lack of selection during double-low breeding (Fig. 6b–d). In contrast, the rapid increase in the proportion of optimal haplotypes in *BnaA09.PYRD* and *BnaC02.LTP15* post-2000 indicates incidental selection during the breeding process aimed at high yield and high resistance. The optimal haplotypes of *BnaA08.PSK1* did not show significant variation over time, which implies that the optimal haplotypes of *BnaA08.PSK1* retain significant breeding potential. These findings offer valuable genetic resources for understanding the genetic basis of fatty acid composition and for breeding double-low rapeseed.

Conclusions

This study unravels the genetic regulatory network underlying fatty acid biosynthesis in rapeseed using an integrative multi-omics approach and identifies four key genes that modulate the balance between oleic acid and erucic acid. These findings offer both a theoretical framework and valuable gene targets for the precise enhancement of oil quality. Beyond advancing our understanding of plant lipid metabolism, this work lays a solid foundation for molecular design breeding, facilitating the development of high-quality rapeseed cultivars for agricultural application.

Methods

Plant materials and trait determination

The genetic transformation material used in this study was the rapeseed inbred line WH3411, characterized by high erucic acid content (34.9%) and high seed oil content (51.28%); it is one of the 505 rapeseed accessions included in the resequencing project with ID X1062 (Additional file 2: Table S1). This line was sourced from the Wuhan National Engineering Research Center for Rapeseed, China. The data used in the study included resequencing data of 505 rapeseed accessions, 309 seed transcriptome data at 20 days after flowering (DAF), and 274 seed transcriptome data at 40 DAF [33]. Additionally, time series transcriptome data were collected during ZS11 seed development (2 DAF to 60 DAF) [56]. For transgenic material subjected to transcriptome sequencing, 50 mg of developing seeds (35 DAF) were collected from both WT and *bna.pyrd-L35* lines, with three biological replicates per genotype. The transcriptome data are available under the BioProject ID PRJNA1214085.

FA phenotypes were collected from multiple locations over various years: Wuhan (2016–2019, 4 years), Ezhou (2017–2019, 3 years), Hefei (2017–2018, 2 years), Lanzhou (2018–2019, 2 years), Chengdu (2017), and Kunming (2018). Mature, open-pollinated seeds from the population were harvested and dried. FA composition was analyzed using a Foss NIRSystems 5000 near-infrared reflectance spectrometer, measuring palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), linolenic acid (C18:3), eicosenoic acid (C20:1), and erucic acid (C22:1). Six biological were identified for each germplasm for analysis.

Variant identification and genotype imputation

The rapeseed reference genome Darmor-bzh was obtained from Genoscope (http://www.genoscope.cns.fr/brassicanapus/) [57]. Sequence alignments were conducted using BWA software [58], and duplicate PCR fragments were removed using SAMTools markdup [59]. Variants from the 505 rapeseed accessions were identified using GATK [60]. SNPs and InDels with low mapping quality (MQ < 20) or shallow sequencing depth (DP < 50) were filtered out. Missing genotypes were imputed using the LD-KNN algorithm [61], resulting in final sequencing data with an accuracy greater than 99.7%.

Genome-wide association analysis

In the ALL, SWR, and SWR1 subgroups, 10,620,048, 9,783,864, and 8,609,979 high-quality SNPs (with MAF > 0.05) were identified, respectively. The significance thresholds for association were calculated by GEC software [62] and were determined to be 7.96×10^{-7}, 9.38×10^{-7}, and 1.69×10^{-7}, for each subgroup, respectively. Association analyses of multiple FA compositions and their ratios were conducted using a mixed linear model implemented in GEMMA software [63, 64].

Transcriptome-wide association analysis

The quality of transcriptome data was assessed using FastQC software [65], and filtered sequences were aligned to the reference genome using STAR software [66].

Sequence quantification was performed with Salmon [67], and normalization was conducted using Tximport software [68], and low-expression genes (with TPM < 1 in 95% of the accessions) in the population were removed in subsequent analyses. Association analysis was performed using the EMMAX mixed linear model [69], with an FDR-corrected *P* value ≤ 0.5 as the threshold for significance in TWAS.

Significant QTLs selection analysis

High-coverage resequencing data from *B. oleracea* and *B. rapa* were utilized to infer ancestral alleles [70]. The data were obtained from the NCBI database (with BioProject ID PRJNA312457), with 17 samples excluded due to low sequence quality. Sequences were aligned to the rapeseed *An* and *Cn* subgenomes using BWA software [58], with average alignment rates of 70.9% for the *An* subgenome and 81.1% for the *Cn* subgenome. Genotypes were extracted using SAMtools [59] and BCFtools [71] to prepare data for selection analysis.

Delineation of gene expression modules during seed development

The expression module analysis based on transcriptome data collected from 2 to 64 DAF during ZS11 seed development. Extreme values of average gene expression (TPM < 1 or TPM > 10,000) were filtered out. Using the default parameters of the Mfuzz package [72], 50,025 genes were classified into eight expression modules.

Transcriptome data analysis

Libraries were sequenced using the Illumina HiSeq platform with paired-end reads. Quality control was conducted using FastQC [65], and low-quality sequences were filtered out using Trimmomatic [73]. Filtered data were aligned to the rapeseed reference genome using HISAT2 software [74], and gene expression values were quantified with featureCounts [75]. Differentially expressed genes (DEGs) were identified with DESeq2 [76], with significance thresholds set at *Padj* < 0.05 and $|\log_2 (\text{fold change})| > 1$.

Vector construction and plant transformation

DNA from 7-day-old WH3411 seedlings was extracted using the CTAB method. The gDNA of *BnaPYRD* and *BnaPSK1* was cloned and linked to pCAMBIA2306 with the 35S promoter. CRISPR targets were selected, and primers were designed using CRISPR-P v2.0 (http://cbi.hzau.edu.cn/CRISPR2/) (Additional file 2: Table S11). The sgRNA-Cas9 system was utilized for vector construction, with DNA sequence templates for CRISPR vectors sourced from pCBC-DT1T2. Purified PCR products were ligated into *pKSE401* vector. The high-erucic-acid rapeseed germplasm WH3411 was transformed via *Agrobacterium*-mediated techniques using the hypocotyl and tissue culture system [77].

Subcellular localization

Subcellular localization was conducted using cDNA synthesized from RNA extracted from the leaves of the WH3411 variety as a template, full-length CDS of rapeseed genes were amplified with gene-specific primers. The target fragments were then cloned into the GFP fusion vector PMDC83, using *Bam*HI and *Kpn*I restriction sites. After sequencing and verification, the constructs were transformed into *Agrobacterium* strain GV3101.

Single colonies of *Agrobacterium* containing the PMDC83-GFP vector were expanded and collected. The bacterial pellets were resuspended in buffer solution (50 mM MES, pH 5.6; 5 mM Na3PO4; 1 mM acetosyringone) and infiltrated into tobacco leaves via syringe injection. Green fluorescence signals were observed under a fluorescence microscope (Olympus BX35) 2–5 days post-infiltration. The excitation wavelength for the GFP was set to 488 nm, and the emission filter wavelength was 500–530 nm.

Fatty acid analysis

FAs were extracted from mature seeds using the gas chromatography (GC) FA methyl ester method, following Lu et al. [78]. Various FA species were measured with an Agilent 6890 GC.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03558-x.

Additional file 1: Supplementary Figures S1 to S25.

Additional file 2: Supplementary Tables S1–S11.

Acknowledgements
Computations in this study were conducted on the bioinformatics computing platform of the National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University. This work is supported by the Biological Breeding-National Science and Technology Major Project (2023ZD04069), National Science Fund for Distinguished Young Scholars (32225037)，National Natural Science Foundation of China (U2102217), Hubei Hongshan Laboratory Fund (2021HSZD004), Project X2662024ZKPY001 supported by the Fundamental Research Funds for the Central Universities and HZAU-AGIS Cooperation Fund (SZYJY2021004).

Peer review information
Qingxin Song and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
H.Z., S.W. and L.G. designed and supervised this study. Y.Z. performed the bioinformatics analysis. Y.L. performed the related experiments. Y.Z. and Y.L. prepared the manuscript. H.Z., S.W. and L.G. revised the manuscript. All authors read and approved the final manuscript.

Funding
This work is supported by the Biological Breeding-National Science and Technology Major Project (2023ZD04069), National Science Fund for Distinguished Young Scholars (32225037), National Natural Science Foundation of China (U2102217), Hubei Hongshan Laboratory Fund (2021HSZD004), Project X2662024ZKPY001 supported by the Fundamental Research Funds for the Central Universities, and HZAU-AGIS Cooperation Fund (SZYJY2021004).

Data availability
Resequencing data for 505 B. napus accessions are available at the Genome Sequence Archive (https://bigd.big.ac.cn/gsa/) under BioProject ID PRJCA002835 [79], with corresponding transcriptome data under BioProject ID PRJCA002836 [80]. PYRD-related transcriptome data can be accessed under BioProject ID PRJNA1214085 [81]. Time-series transcriptome data are deposited in NCBI under BioProject ID PRJNA722877 [82]. Additionally, resequencing data for B. oleracea and B. rapa are available from the NCBI database with BioProject ID PRJNA312457 [83]. The phenotypic data used in this study can be obtained from the website http://rgmi.hzau.edu.cn/phenotype and have also been deposited in Zenodo at https://zenodo.org/records/15048705 [84].All software and tools used in this study are publicly available as described in the Methods section. The code for POCKET can be accessed on GitHub at https://github.com/zhaouu/POCKET under the BSD 3-Clause license [85]. The customized scripts used in the present study are also available via Zenodo at https://doi.org/10.5281/zenodo.14842123 [86].Any additional information required to reanalyze the data reported in this paper can be provided upon request.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 17 November 2024 Accepted: 25 March 2025
Published online: 02 April 2025

References

1. Friedt W, Tu J, Fu T. Academic and economic importance of Brassica napus rapeseed. The Brassica napus genome; 2018. p. 1–20.
2. Goyal A, Tanwar B, Sihag MK, Kumar V, Sharma V, Soni S. Rapeseed/canola (Brassica napus) seed. Oilseeds Health Food Appl; 2021. p. 47–71.
3. Gurr M. The biosynthesis of triacylglycerols. Lipids: structure and function; 1980. p. 205–248.
4. Gunstone, Frank D., ed. Rapeseed and canola oil production, processing, properties and uses. CRC Press; 2004. p. 79–105.
5. Stymne S, Stobart AK. Triacylglycerol biosynthesis. Lipids: structure and function; 1987. p. 175–214.
6. De Carvalho CC, Caramujo MJ. The various roles of fatty acids. Molecules. 2018;23:2583.
7. Costa MC, Rolemberg MP, Boros LA, Krähenbühl MA, de Oliveira MG, Meirelles AJ. Solid–liquid equilibrium of binary fatty acid mixtures. J Chem Eng Data. 2007;52:30–6.
8. Orsavova J, Misurcova L, Vavra Ambrozova J, Vicha R, Mlcek J. Fatty acids composition of vegetable oils and its contribution to dietary energy intake and dependence of cardiovascular mortality on dietary intake of fatty acids. Int J Mol Sci. 2015;16:12871–90.
9. Liu S, Fan C, Li J, Cai G, Yang Q, Wu J, Yi X, Zhang C, Zhou Y. A genome-wide association study reveals novel elite allelic variations in seed oil content of Brassica napus. Theor Appl Genet. 2016;129:1203–15.
10. Karantonis HC, Zabetakis I, Nomikos T, Demopoulos CA. Antiatherogenic properties of lipid minor constituents from seed oils. J Sci Food Agric. 2003;83:1192–204.
11. Messina M, Shearer G, Petersen K. Soybean oil lowers circulating cholesterol levels and coronary heart disease risk, and has no effect on markers of inflammation and oxidation. Nutr. 2021;89: 111343.
12. Shen J, Liu Y, Wang X, Bai J, Lin L, Luo F, Zhong H. A comprehensive review of health-benefiting components in rapeseed oil. Nutr. 2023;15:999.
13. Priyamedha P, Singh BK, Lijo Thomas LT, Manju Bala MB, Singh VV, Dhiraj Singh DS. Status and perspective of canola quality rapeseed-mustard cultivation in India: a review. J Oilseed Brassica. 2016;1:142–51.
14. Yang X, Guo Y, Yan J, Zhang J, Song T, Rocheford T, Li J-S. Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. Theor Appl Genet. 2010;120:665–78.
15. Coonrod D, Brick MA, Byrne PF, DeBonte L, Chen Z. Inheritance of long chain fatty acid content in rapeseed (Brassica napus L.). Euphytica. 2008;164:583–92.
16. Li H, Zhao T, Wang Y, Yu D, Chen S, Zhou R, Gai J. Genetic structure composed of additive QTL, epistatic QTL pairs and collective unmapped minor QTL conferring oil content and fatty acid components of soybeans. Euphytica. 2011;182:117–32.
17. Otyama PI, Chamberlin K, Ozias-Akins P, Graham MA, Cannon EK, Cannon SB, MacDonald GE, Anglin NL. Genome-wide approaches delineate the additive, epistatic, and pleiotropic nature of variants controlling fatty acid composition in peanut (Arachis hypogaea L.). G3. 2022;12:jkab382.
18. Downey R, Harvey B. Methods of breeding for oil quality in rape. Can J Plant Sci. 1963;43:271–5.
19. Harvey B, Downey R. The inheritance of erucic acid content in rapeseed (Brassica napus). Can J Plant Sci. 1964;44:104–11.
20. Ecke W, Uzunova M, Weissleder K. Mapping the genome of rapeseed (Brassica napus L.). II. Localization of genes controlling erucic acid synthesis and seed oil content. Theor Appl Genet. 1995;91:972–7.
21. Peng Q, Hu Y, Wei R, Zhang Y, Guan C, Ruan Y, Liu C. Simultaneous silencing of FAD2 and FAE1 genes affects both oleic acid and erucic acid contents in Brassica napus seeds. Plant Cell Rep. 2010;29:317–25.
22. Okuzaki A, Ogawa T, Koizuka C, Kaneko K, Inaba M, Imamura J, Koizuka N. CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in Brassica napus. Plant Physiol Biochem. 2018;131:63–9.
23. Wells R, Trick M, Soumpourou E, Clissold L, Morgan C, Werner P, Gibbard C, Clarke M, Jennaway R, Bancroft I. The control of seed oil polyunsaturate content in the polyploid crop species Brassica napus. Mol Breed. 2014;33:349–62.
24. Yeom WW, Kim HJ, Lee KR, Cho HS, Kim JY, Jung HW, Oh SW, Jun SE, Kim HU, Chung YS. Increased production of α-linolenic acid in soybean seeds by overexpression of Lesquerella FAD3-1. Front Plant Sci. 2020;10:1812.
25. Muthu RK, Kumar A, Venugopal BP, Shanmugam K. IAA combine with kinetin elevates the α-linolenic acid in callus tissues of soybean by stimulating the expression of FAD3 gene. Plant Gene. 2021;28:100336.
26. Shi J, Ni X, Huang J, Fu Y, Wang T, Yu H, Zhang Y. CRISPR/Cas9-mediated gene editing of BnFAD2 and BnFAE1 modifies fatty acid profiles in Brassica napus. Genes. 2022;13: 1681.
27. Porokhovinova E, Matveeva T, Khafizova G, Bemova V, Doubovskaya A, Kishlyan N, Podolnaya L, Gavrilova V. Fatty acid composition of oil crops: genetics and genetic engineering. Genet Resour Crop Evol. 2022;69:2029–45.
28. Peters JL, Cnudde F, Gerats T. Forward genetics and map-based cloning approaches. Trends Plant Sci. 2003;8:484–91.
29. Wu D, Li X, Tanaka R, Wood JC, Tibbs-Cortes LE, Magallanes-Lundback M, Bornowski N, Hamilton JP, Vaillancourt B, Diepenbrock CH. Combining GWAS and TWAS to identify candidate causal genes for tocochromanol levels in maize grain. Genetics. 2022;221: iyac091.
30. Li D, Liu Q, Schnable PS. TWAS results are complementary to and less affected by linkage disequilibrium than GWAS. Plant Physiol. 2021;186:1800–11.
31. Ferguson JN, Fernandes SB, Monier B, Miller ND, Allen D, Dmitrieva A, Schmuker P, Lozano R, Valluru R, Buckler ES. Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. Plant Physiol. 2021;187:1481–500.
32. Han B, Huang X. Sequencing-based genome-wide association study in rice. Curr Opin Plant Biol. 2013;16:133–8.

33. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang Q-Y, Zhou Y, Wang X, Ma W, et al. Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in Brassica napus. Mol Plant. 2021;14:470–87.
34. Tan Z, Peng Y, Xiong Y, Xiong F, Zhang Y, Guo N, Tu Z, Zong Z, Wu X, Ye J, et al. Comprehensive transcriptional variability analysis reveals gene networks regulating seed oil content of Brassica napus. Genome Biol. 2022;23:233.
35. Tan Z, Xie Z, Dai L, Zhang Y, Zhao H, Tang S, Wan L, Yao X, Guo L, Hong D. Genome-and transcriptome-wide association studies reveal the genetic basis and the breeding history of seed glucosinolate content in Brassica napus. Plant Biotechnol J. 2022;20:211–25.
36. Zhang Y, Zhang H, Zhao H, Xia Y, Zheng X, Fan R, Tan Z, Duan C, Fu Y, Li L, et al. Multi-omics analysis dissects the genetic architecture of seed coat content in Brassica napus. Genome Biol. 2022;23:86.
37. Li L, Tian Z, Chen J, Tan Z, Zhang Y, Zhao H, Wu X, Yao X, Wen W, Chen W. Characterization of novel loci controlling seed oil content in Brassica napus by marker metabolite-based multi-omics analysis. Genome Biol. 2023;24:141.
38. Lee HG, Park B-Y, Kim HU, Seo PJ. MYB96 stimulates C18 fatty acid elongation in Arabidopsis seeds. Plant Biotechnol Rep. 2015;9:161–6.
39. Kim RJ, Han S, Kim HJ, Hur JH, Suh MC. Arabidopsis 3-ketoacyl-CoA synthase 17 produces tetracosanoic acids required for synthesizing seed coat suberin. J Exp Bot. 2024;75:1767–80
40. Sadler C, Schroll B, Zeisler V, Waßmann F, Franke R, Schreiber L. Wax and cutin mutants of Arabidopsis: quantitative characterization of the cuticular transport barrier in relation to chemical composition. Biochim Biophys Acta - Mol Cell Biol Lipids. 2016;1861:1336–44.
41. Qi Q, Huang YF, Cutler AJ, Abrams SR, Taylor DC. Molecular and biochemical characterization of an aminoalcoholphosphotransferase (AAPT1) from Brassica napus. Effects of low temperature and abscisic acid treatments on AAPT expression in Arabidopsis plants and effects of over-expression of BnAAPT1 in transgenic Arabidopsis. Planta. 2003;217:547–58.
42. Deng W, Chen G, Peng F, Truksa M, Snyder CL, Weselake RJ. Transparent testa16 plays multiple roles in plant development and is involved in lipid synthesis and embryo development in canola. Plant Physiol. 2012;160:978–89.
43. Wu P, Gao H, Liu J, Kosma DK, Lü S, Zhao H. Insight into the roles of the ER-associated degradation E3 ubiquitin ligase HRD1 in plant cuticular lipid biosynthesis. Plant Physiol Biochem. 2021;167:358–65.
44. Rawsthorne S. Carbon flux and fatty acid synthesis in plants. Prog Lipid Res. 2002;41:182–96.
45. Ohlrogge JB, Jaworski JG. Regulation of fatty acid synthesis. Annu Rev Plant Biol. 1997;48:109–36.
46. Hu J, Chen B, Zhao J, Zhang F, Xie T, Xu K, Gao G, Yan G, Li H, Li L. Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. Nat Genet. 2022;54:694–704.
47. Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H. Whole-genome resequencing reveals Brassica napus origin and genetic loci involved in its improvement. Nat Commun. 2019;10:1154.
48. Qu C, Jia L, Fu F, Zhao H, Lu K, Wei L, Xu X, Liang Y, Li S, Wang R. Genome-wide association mapping and identification of candidate genes for fatty acid composition in Brassica napus L. using SNP markers. BMC Genomics. 2017;18:1–17.
49. Guan M, Huang X, Xiao Z, Jia L, Wang S, Zhu M, Qiao C, Wei L, Xu X, Liang Y. Association mapping analysis of fatty acid content in different ecotypic rapeseed using mrMLM. Front Plant Sci. 1872;2019:9.
50. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang Q-Y, Zhou Y, Wang X, Ma W. Genome-and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in Brassica napus. Mol Plant. 2021;14:470–87.
51. Bellin L, Melzer M, Hilo A, Garza Amaya DL, Keller I, Meurer J, Möhlmann T. Nucleotide limitation results in impaired photosynthesis, reduced growth and seed yield together with massively altered gene expression. Plant Cell Physiol. 2023;64:1494–510.
52. Kari H, Bandi SMS, Kumar A, Yella VR. DeePromClass: delineator for eukaryotic core promoters employing deep neural networks. IEEE/ACM Trans Comput Biol Bioinform. 2022;20:802–7.
53. Akagi T, Masuda K, Kuwada E, Takeshita K, Kawakatsu T, Ariizumi T, Kubo Y, Ushijima K, Uchida S. Genome-wide cis-decoding for expression design in tomato using cistrome data and explainable deep learning. Plant Cell. 2022;34:2174–87.
54. Toneyan S, Tang Z, Koo PK. Evaluating deep learning for predicting epigenomic profiles. Nat Mach Intell. 2022;4:1088–100.
55. Lin L, Zhang X, Fan J, Li J, Ren S, Gu X, Li P, Xu M, Xu J, Lei W. Natural variation in BnaA07.MKK9 confers resistance to Sclerotinia stem rot in oilseed rape. Nat Commun. 2024;15:5059.
56. Liu D, Yu L, Wei L, Yu P, Wang J, Zhao H, Zhang Y, Zhang S, Yang Z, Chen G. BnTIR: an online transcriptome platform for exploring RNA-seq libraries for oil crop Brassica napus. Plant Biotechnol J. 1895;2021:19.
57. Chalhoub B, France D, Shengyi L, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science. 2014;345:950–3.
58. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv. 2013;00:1–3.
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing S. The sequence alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
60. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
61. Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, Li Y, Liu X, Zhang H, Dong H, et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat Genet. 2014;46:714–21.
62. Li M-X, Yeung JMY, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet. 2011;131:747–56.
63. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4.

64. Goldstein H. Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika. 1986;73:43–56.
65. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8.
66. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
67. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417–9.
68. Charlotte S, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]. F1000 Res. 2016;4:1521.
69. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010;42:348–54.
70. Cheng F, Sun R, Hou X, Zheng H, Zhang F, Zhang Y, Liu B, Liang J, Zhuang M, Liu Y, et al. Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in Brassica rapa and Brassica oleracea. Nat Genet. 2016;48:1218–24.
71. Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics. 2017;33:2037–9.
72. Kumar L, Futschik ME. Mfuzz: a software package for soft clustering of microarray data. Bioinformation. 2007;2: 5.
73. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
74. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
75. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
76. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:1–21.
77. Dai C, Li Y, Li L, Du Z, Lin S, Tian X, Li S, Yang B, Yao W, Wang J. An efficient Agrobacterium-mediated transformation method using hypocotyl as explants for Brassica napus. Mol Breed. 2020;40:1–13.
78. Lu S, Yao S, Wang G, Guo L, Zhou Y, Hong Y, Wang X. Phospholipase Dε enhances B. napus growth and seed production in response to nitrogen availability. Plant Biotechnol J. 2016;14:926–37.
79. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang QY, Zhou Y, Wang X, Ma W, et al. Genome-wide re-sequencing data of *Brassica napus*. Genome Sequence Archive. BioProject accession: PRJCA002835. 2021. https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA002835.
80. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang QY, Zhou Y, Wang X, Ma W, et al. Transcriptome-wide data of seed in Brassica napus. Genome sequence archive. BioProject accession: PRJCA002836. 2021. https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA002836.
81. Zhang Y, Liu Y, Zong Z, Guo L, Shen W, Zhao H. RNA-Seq analyses of differentially expressed genes in the seeds of WT and pyrd-L35 mutant. NCBI. BioProject accession: PRJNA1214085. 2025. https://www.ncbi.nlm.nih.gov/bioproject?term=PRJNA1214085.
82. Liu D, Yu L, Wei L, Yu P, Wang J, Zhao H, Zhang Y, Zhang S, Yang Z, Chen G, et al. Brassica napus raw sequence reads. NCBI. BioProject accession: PRJNA722877. 2021. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA722877.
83. Cheng F, Sun R, Hou X, Zheng H, Zhang F, Zhang Y, Liu B, Liang J, Zhuang M, Liu Y, Liu D, Wang X, Li P, Liu Y, et al. Genome resequencing of Brassica rapa and Brassica oleracea accessions. NCBI. BioProject accession: PRJNA312457. 2016. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA312457.
84. Zhang Y, Liu Y, Zong Z, Guo L, Shen W, Zhao H. The extensive data on the oil content and fatty acid composition of rapeseed populations from multiple regions and various time periods. Zenodo. 2025. https://zenodo.org/records/15048705.
85. Zhao H. Prioritizing the candidate genes by incorporating information of knowledge-based gene sets, effects of variants, GWAS and TWAS. Github. 2021. https://github.com/zhaouu/POCKET.
86. Zhang Y, Liu Y, Zong Z, Guo L, Shen W, Zhao H. Elucidation of the genetic basis of seed fatty acids in Brassica napus through integrative omics analysis. Zenodo. 2025. https://doi.org/10.5281/zenodo.14842123.

Publisher's Note