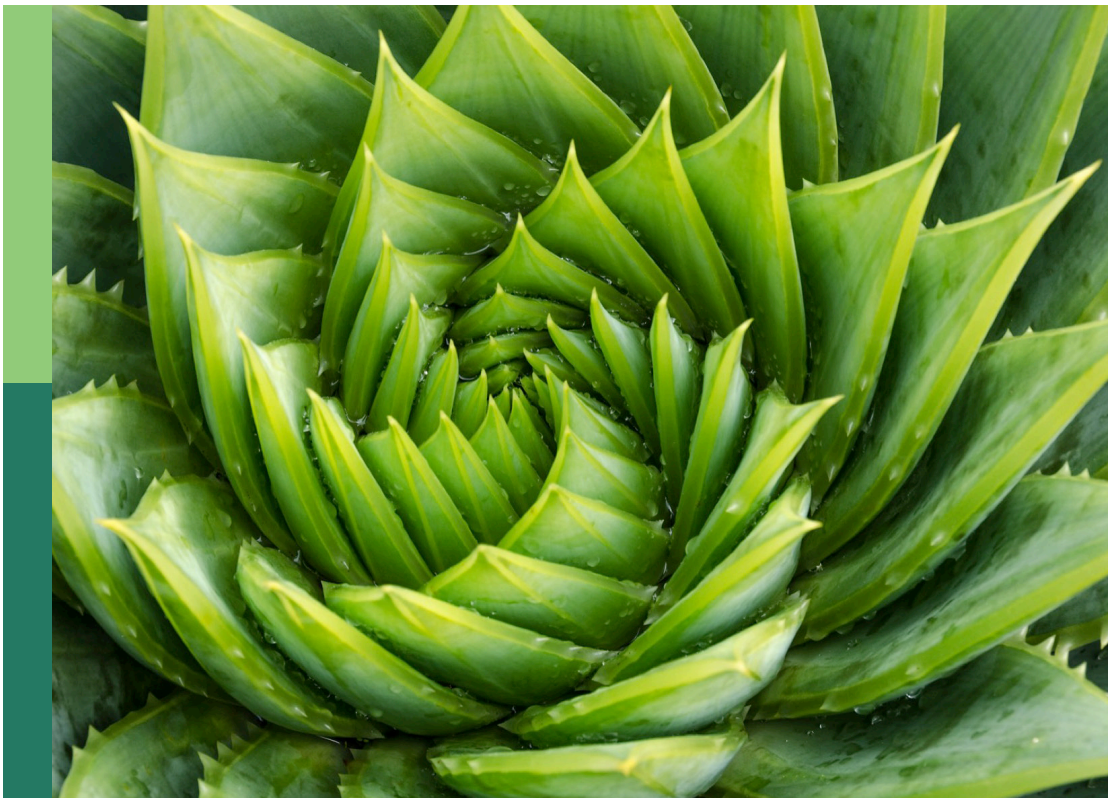
frontiers | Research Topics

Genetics, breeding and engineering to enhance oil quality and yield

Edited by
Hongbo Chao, Aruna Kilaru and Liezhao Liu

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3277-5
DOI 10.3389/978-2-8325-3277-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Genetics, breeding and engineering to enhance oil quality and yield

Topic editors

Hongbo Chao — Zhengzhou University, China

Aruna Kilaru — East Tennessee State University, United States

Liezhao Liu — Southwest University, China

Citation

Chao, H., Kilaru, A., Liu, L., eds. (2023). *Genetics, breeding and engineering to enhance oil quality and yield*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3277-5

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 04 August 2023
DOI 10.3389/fpls.2023.1265897

OPEN ACCESS

EDITED AND REVIEWED BY
Leo Marcelis,
Wageningen University and Research,
Netherlands

*CORRESPONDENCE
Aruna Kilaru
kilaru@etsu.edu
Liezhao Liu
liezhao@swu.edu.cn

RECEIVED 24 July 2023
ACCEPTED 24 July 2023
PUBLISHED 04 August 2023

CITATION
Chao H, Kilaru A and Liu L (2023) Editorial: Genetics, breeding and engineering to enhance oil quality and yield.
Front. Plant Sci. 14:1265897.
doi: 10.3389/fpls.2023.1265897

Editorial: Genetics, breeding and engineering to enhance oil quality and yield

Hongbo Chao[1], Aruna Kilaru[2*] and Liezhao Liu[3*]

[1]School of Agricultural Sciences, Zhengzhou University, Zhengzhou, China, [2]Department of Biological Sciences, East Tennessee State University, Johnson City, TN, United States, [3]College of Agronomy and Biotechnology, Southwest University, Beibei, Chongqing, China

KEYWORDS

oil content, oil quality, oil yield, oil crop, genetic improvement

Editorial on the Research Topic
Genetics, breeding and engineering to enhance oil quality and yield

Considering our rapidly increasing global population, the demand for food and edible oil continues to surge. Specifically, the rise in demand for vegetable oils, primarily composed of the storage lipid triacylglycerol, has driven extensive research efforts to enhance oil content and improve oil quality in oilseed crops. The quality of oil crops is important to consumers as they directly contribute to human health by providing essential nutrients along with oil. Vegetable oil quality is determined by its lipid category and fatty acid composition, while additional nutrients like vitamins, minerals, and bioactive compounds influence seed quality, nutritional value, and processing characteristics of oil crops.

Given the growing consumption of high-quality vegetable oils, substantial research has been directed towards enhancing oil content and quality over the years. Various approaches, including conventional breeding, molecular-assisted breeding, and targeted genetic manipulation, have been employed to achieve these goals. This Research Topic compiles a collection of ten articles that focus on increasing oil content and improving oil quality, ultimately deepening our understanding of the genetic control of these traits.

Key areas explored in these articles include:

Enhancing oil content

Increasing oil content is crucial for ensuring a steady and safe supply of global edible vegetable oil. The identification of important and stable quantitative trait loci (QTLs) sheds light on the genetic basis of oil content and provides a foundation for marker-assisted selection to increase oil content (Zhao et al.). Moreover, candidate genes identified through this research offer valuable genetic resources and targets for improving oil content via targeted genetic manipulation, such as transgenic and gene editing technologies (Xiao et al.; Zhang et al.).

Improving vegetable oil quality

Attention is also being given to enhancing the quality of vegetable oil. Fatty acid composition plays a pivotal role in determining oil quality. The development of molecular markers associated with fatty acid components facilitates genetic improvements in oil crop quality (Li et al.). Additionally, identifying important functional genes provides genetic resources for cultivating plant oils with specific fatty acid components through metabolic engineering (Behera et al.). Balancing high plant productivity with improved oil quality remains a challenge, as there is often a negative correlation between yield and quality traits in oil crops, such as between flavonoid content and oil content. Understanding the inheritance and regulation of quality traits could lead to further advancements in both yield and quality (Guan et al.).

Exploration through multiple means or omics

The search for essential genes related to oil content and quality improvement is conducted through various methods or omics. The verification of important homologous genes provides insights into their potential application for improving oil content and quality in oil crops (Behera et al.; Shen et al.; Lin et al.). Additionally, genome-wide data and gene function annotation enable the efficient mining of crucial functional genes (Cheng et al.; Yao et al.).

The research in this Research Topic encompasses various oil crops, such as rapeseed, peanut, sesame, oil palm, and avocado. This diversity of species provides a comprehensive understanding of the research progress in oil crops. Notably, the methodologies employed in these studies are equally diverse and offer a rich repository of references for oil crop research. Altogether, this work deepens our comprehension of the genetic control of oil content and quality in oil crops, paving the way for further genetic improvements in the future.

Author contributions

HC: Conceptualization, Data curation, Funding acquisition, Writing – original draft, Writing – review & editing. AK: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Supervision, Writing – review & editing. LL: Conceptualization, Supervision, Writing – review & editing.

Funding

HC is supported by the National Natural Science Foundation of China (32001583), and the National Natural Science Foundation of China (U2004149). AK received support from the U.S. Department of Agriculture, the National Institute of Food and Agriculture (GRANT13058738).

Acknowledgments

Thanks go to the many authors, reviewers, and editors who contributed to this effort.

Conflict of interest

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

Publisher's note

Author disclaimer

Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the funding agencies such as the National Natural Science Foundation of China and the U.S. National Institute of Food and Agriculture.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fpls.2022.885418

Development of Intron Polymorphism Markers and Their Association With Fatty Acid Component Variation in Oil Palm

Jing Li[1], Yaodong Yang[1], Xiwei Sun[1], Rui Liu[1], Wei Xia[2], Peng Shi[1], Lixia Zhou[1], Yong Wang[1], Yi Wu[1], Xintao Lei[3] and Yong Xiao[1]**

[1]Hainan Key Laboratory of Tropical Oil Crops Biology/Coconut Research Institute, Chinese Academy of Tropical Agricultural Sciences, Wenchang, China, [2]Institute of Tropical Agriculture and Forestry, Hainan University, Haikou, China, [3]Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

Oil palm (*Elaeis guineensis* Jacq.) is a tropical woody oil crop of the palm family and is known as "the oil king of the world," but its palm oil contains about 50% palmitic acid, which is considered unhealthy for humans. Intron polymorphisms (IP) are highly efficient and easily examined molecular markers located adjacent to exon regions of functional genes, thus may be associated with targeted trait variation. In order to speed up the breeding of oil palm fatty acid composition, the current study identified a total of 310 introns located within 52 candidate genes involved in fatty acid biosynthesis in the oil palm genome. Based on the intron sequences, 205 primer pairs were designed, 64 of which showed polymorphism among 70 oil palm individuals. Phenotypic variation of fatty acid content in the 70 oil palm individuals was also investigated. Association analysis revealed that 13 IP markers were significantly associated with fatty acid content variation, and these IP markers were located on chromosomes 2, 5, 6, 8, 9, and 10 of oil palm. The development of such IP markers may be useful for the genetic improvement of fatty acid composition in oil palm.

Keywords: ***Elaeis guineensis*, fatty acid, IP markers, candidate gene, association analysis**

OPEN ACCESS

Edited by:
Peter Poczai,
University of Helsinki, Finland

Reviewed by:
Wei Zhang,
Zhejiang Normal University, China
Robert Jarret,
United States Department of Agriculture, United States

***Correspondence:**
Xintao Lei
xtlei@263.net
Yong Xiao
xiaoyong_coconut@163.com

Specialty section:
This article was submitted to Plant Physiology, a section of the journal Frontiers in Plant Science

Received: *28 February 2022*
Accepted: *16 May 2022*
Published: *02 June 2022*

Citation:
Li J, Yang Y, Sun X, Liu R, Xia W, Shi P, Zhou L, Wang Y, Wu Y, Lei X and Xiao Y (2022) Development of Intron Polymorphism Markers and Their Association With Fatty Acid Component Variation in Oil Palm. Front. Plant Sci. 13:885418. doi: 10.3389/fpls.2022.885418

INTRODUCTION

Oil palm (*Elaeis guineensis*, 2n = 32) is an important tropical oil crop and is often referred to as "the oil king of the world" because it has the highest oil yield per unit area among all oil crops (Singh et al., 2013; Huang, 2017). Global production of palm oil in 2017 was approximately 75.70 million tons (Mozzon et al., 2020). The two oil storage tissues of oil palm are the mesocarp and kernel, each of which produces oil with a different fatty acid composition. Palmitic acid (16:0) is the major fatty acid (about 50%) in oil from the oil palm mesocarp, while lauric acid (12:0) is the major fatty acid (about 50%) in kernel oil. As for many important oil crops, improving fatty acid content is a major breeding objective for this tropical oil crop, especially aiming to decrease the palmitic acid content and increase the oleic acid content. However, the breeding scheme is slow due to the long life cycle of oil palms. Developing molecular markers associated

with fatty acid compositions could facilitate the breeding and selection of the tropical oil-seed crop.

Molecular markers such as amplified fragment length polymorphisms (AFLPs), random amplified polymorphic DNA (RAPD), and restriction fragment length polymorphisms (RFLPs) have been widely used for analyzing genetic diversity and population structure, identification of trait-associated markers, and genotype characterization in oil palm (Moretzsohn et al., 2000; Rance et al., 2001; Billotte et al., 2005; Cochard et al., 2009; Singh et al., 2009; Ting et al., 2013). In eukaryotic genomes, there are generally some introns across each gene (Hawkins, 1988; Deutsch and Long, 1999; Haas et al., 2005). Diversity of intron sequences between different individuals is abundant due to low selection pressure (Stoltzfus et al., 1997; Panjabi et al., 2008; Sharma et al., 2020). Therefore, numerous intron polymorphism (IP) markers are available in the eukaryotic genome (Badoni et al., 2016; Kita et al., 2016). Studies have identified different types of IP, including ILP (Intron Length Polymorphism; Wang et al., 2005; Sharma et al., 2020) and ISNP (Intron Single Nucleotide Polymorphism; Ferreira et al., 2009; Liu et al., 2018). IP markers are generally co-dominant and highly polymorphic, and are widely used for constructing genetic maps, diversity analysis, and quantitative trait locus mapping (Williams et al., 1990; Wei et al., 2005; Yang et al., 2007; Xia et al., 2017). In 2013, the genome sequence of oil palm was released, providing an opportunity to develop larger numbers of IP markers (Singh et al., 2013). Polymorphism markers located in the intron region can be efficient function markers in genic regions. Therefore, IP in targeted genes may be associated with targeted traits in *E. guineensis*.

In this study, candidate genes and their introns involved in the biosynthesis and metabolism of fatty acids were identified in the genome sequence of *E. guineensis*. Subsequently, IP markers were developed based on intron sequences among 70 oil palm individuals. Fatty acid composition was also investigated among the 70 oil palm individuals. Finally, associations between IP markers and fatty acid variation were analyzed. This study will provide a exhaustive understanding of fatty acid content in oil palm, and the IP makers and candidate genes detected will facilitate breeding for fatty acid content in oil palm. Provide a theoretical basis for increasing the content of oleic acid and reducing the content of palmitic acid in future oil palm breeding.

MATERIALS AND METHODS

Plant Materials and DNA Extraction

A total of 70 oil palm individuals were selected from the oil palm germplasm resources of the Coconut Research Institute of Chinese Academy of Tropical Agricultural Sciences, Wenchang town, Hainan province, China. Detailed information for these oil palm individuals is listed in **Supplementary Table 1**. DNA samples were prepared from spear leaves using the mini-CTAB method (Murray and Thompson, 1980).

Extraction and Measurement of Fatty Acid Contents in Oil Palm Mesocarp Tissues

Three fruits per oil palm individual (three biological replicates) were harvested, and fatty acid extraction and analyses for each mesocarp tissue were performed in triplicate (three different extractions as technical replicates). Approximately 60 mg mesocarp was used for extracting fatty acids according to methods described by Li-Beisson et al. (2013). Fatty acid composition was subsequently examined and measured using 7890A gas chromatograph equipped with a HP-5MS column (30 m by 250 μm, 0.25 μm). The heating procedure was as follows: initial temperature 180°C, followed by a temperature increase to 220°C at a rate of 10°C per min. The contents of decanoic acid (C10:0), lauric acid (12:0), myristic acid (C14:0), tripalmitelaidin acid (16:1), palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1), and linoleic acid (18:2) were determined with reference to an internal standard using the following calculation: (1) total fatty acid = (total area of all measured fatty acid peaks × quantity of heptadecanoic acid-methyl ester)/(peak area of heptadecanoic acid-methyl ester × quantity of the sample); (2) relative fatty acid percentage = peak area of a specific fatty acid /total area of all measured fatty acid peaks. The peak area was calculated using Agilent software.

Identification of Candidate Genes Involved in Fatty Acid Biosynthesis

The whole-genome sequence of *E. guineensis* was downloaded from the National Center for Biotechnology Information (NCBI). Protein sequences involved in fatty acid biosynthesis were downloaded from the Arabidopsis Information Resource (TAIR), available at http://www.arabidopsis.org. The protein sequences from Arabidopsis were used as queries for BLASTx searches against the coding sequence (CDS) database from *E. guineensis* to identify candidate genes involved in fatty acid biosynthesis. Conserved motifs were also predicted by alignment with Conserved Domains Database (CDD)[1] and PFAM databases.[2]

Identification of Intron Sequence and Primer Design

CDSs of candidate genes involving in fatty acid biosynthesis were used as queries for BLASTn searches against the genome sequence of *E. guineensis* to identify the boundary between exon and intron. The MEME program[3] was used to identify the gene structures of candidate genes. Primers were designed based on the intron sequences using Primer3 software (Rozen and Skaletsky, 2000). The list of PCR primers used in this study is provided in **Supplementary Table 2**.

PCR Amplification and Electrophoresis

PCR amplification was performed in 10 μl reaction mixtures containing 100 ng genomic DNA, 1× PCR buffer, 2 mM $MgCl_2$, 1 U Taq DNA polymerase (TaKaRa, China), 0.5 μM of each

[1]http://www.ncbi.nlm.nih.gov/cdd
[2]http://pfam.sanger.ac.uk
[3]http://meme-suit.org

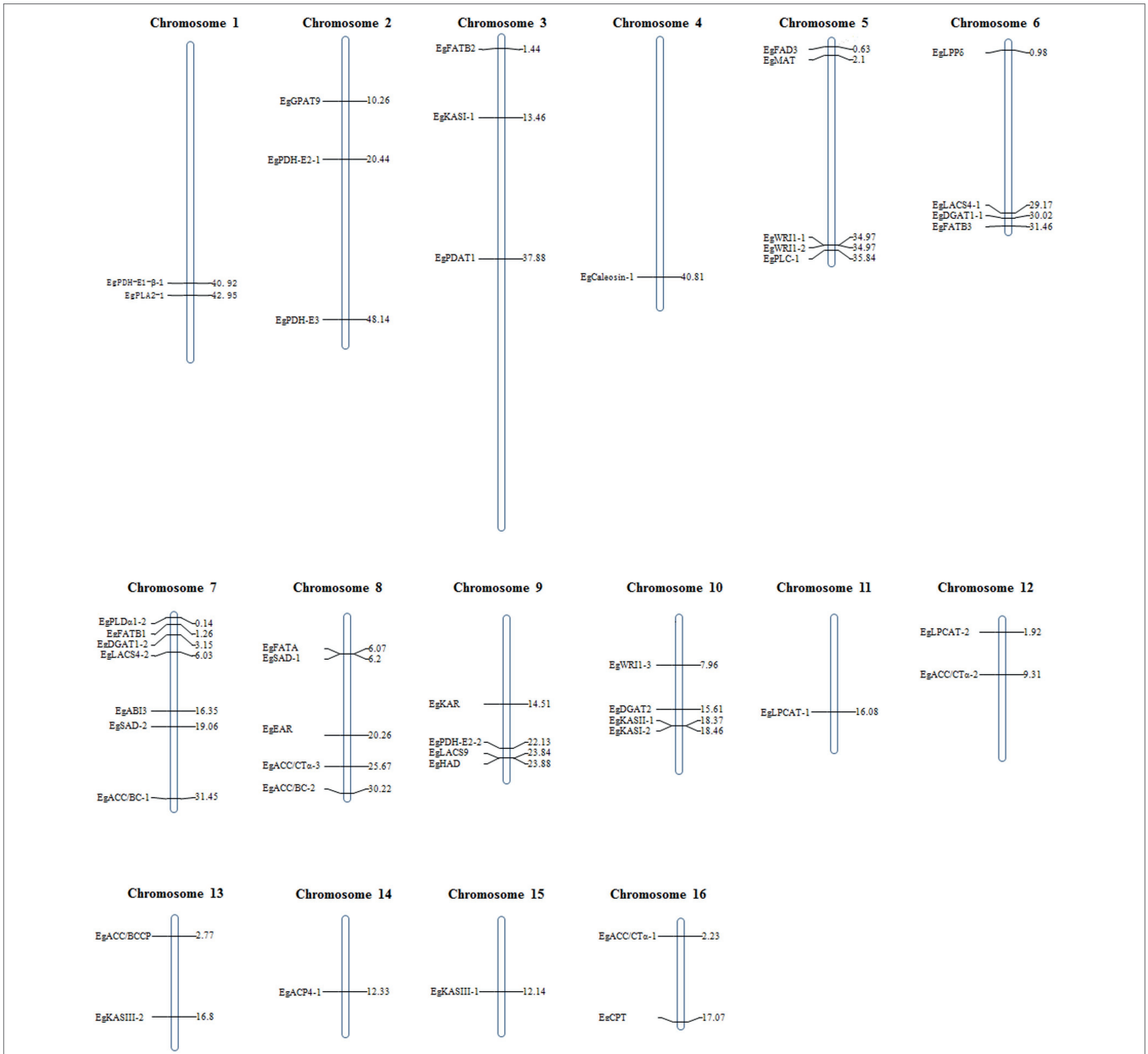


FIGURE 1 | Chromosomal positon of candidate genes involved in the biosynthesis of fatty acid identified in the genome of *E. guineensis*.

primer, and 0.2 mM dNTP mix. The PCR program comprised a denaturation step for 4 min at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at 54.7°C, and 30 s at 72°C. PCR products were electrophoretically separated on 8% polyacrylamide denaturing gels and visualized by silver staining. Product sizes were determined by comparison with a 100 bp DNA ladder.

Population Structure and Genetic Diversity Analysis

Bayesian clustering was employed to analyze the population structure of 70 oil palm individuals using the software STRUCTURE (Pritchard et al., 2000). Ten independent calculations were performed for K value (K set from 2 to 7). Burn-in time and replication number were both set to 100,000 in each run. The maximum-likelihood method was applied to assign every individual to a cluster, and the cut-off probability was set to 0.6. The most probable number of populations (K) was identified by plotting ΔK values of K from 1 to 10 in replicate runs for each K and corresponded to the peak of the ΔK graph. Association analysis was conducted using the software Tassel[4]; statistical significance (value of *p*) was determined by 100,000 permutations (Bradbury et al., 2007).

[4]http://www.maizegenetic.net/bioinforFormatics/tasselindex.htm

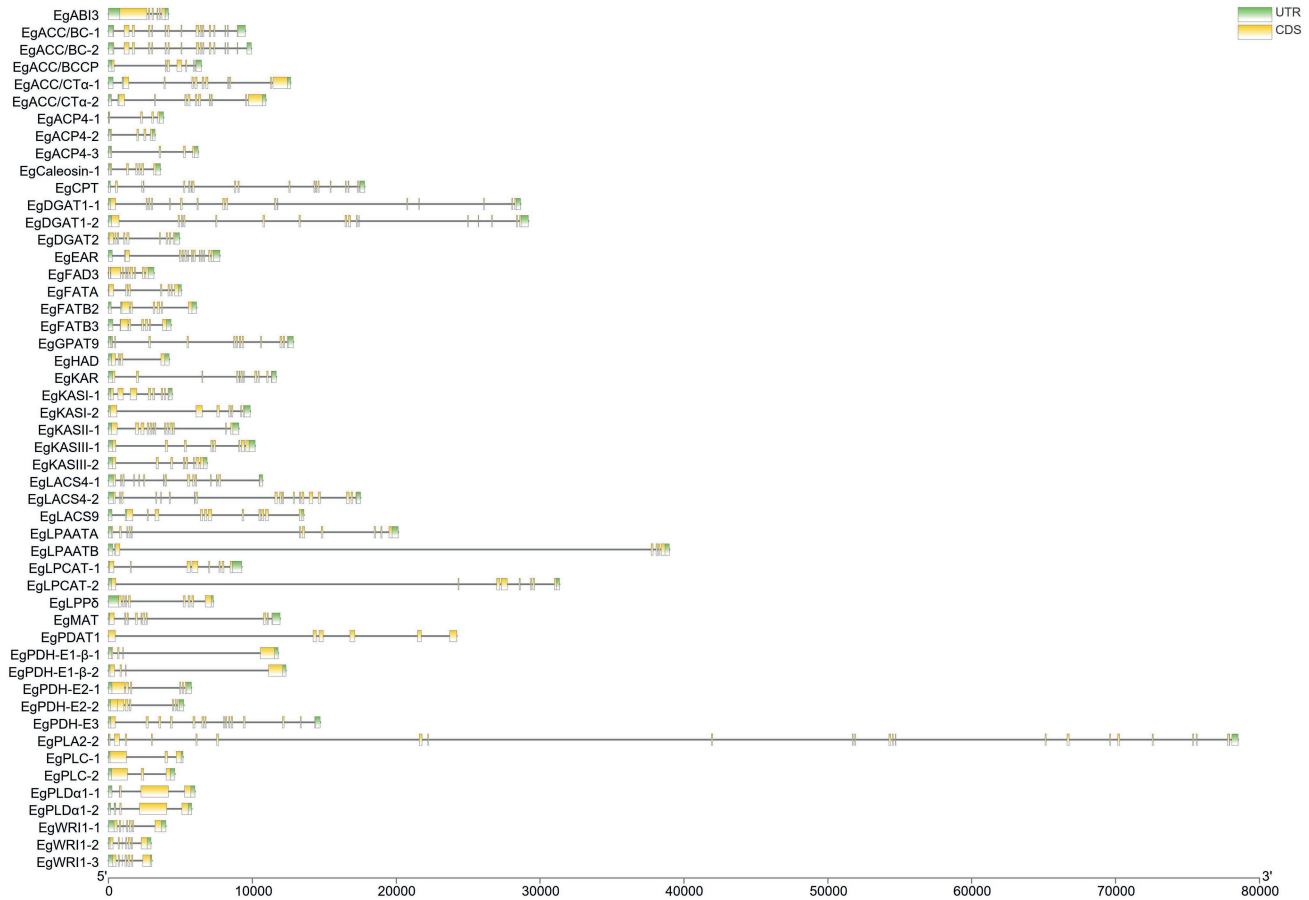


FIGURE 2 | The gene structure of enzyme genes related to the biosynthesis of fatty acid in *Elaeis guineensis*.

RESULTS

Identification of Candidate Genes Involved in *de novo* Synthesis of Fatty Acids

The protein sequences involved in *de novo* synthesis of fatty acids from *Arabidopsis thaliana* were used as queries to align against the protein database of *E. guineensis*. A total of 52 candidate genes that might participate in the *de novo* synthesis of fatty acids, carbon termination or dehydration were identified. These candidate genes were distributed across 16 different chromosomes. Among them, the maximum number of candidate genes were detected in chromosome 7, followed by chromosome 5, and then chromosome 9. Chromosomes 4 and 11 each contained only one candidate gene involved in fatty acid biosynthesis (**Figure 1**).

CDSs of 52 candidate genes involved in fatty acid biosynthesis were used as queries to align with the *E. guineensis* genome to identify the sequence boundary between intron and exon. Almost all candidate genes contained at least two intron sequences. Among all the candidate genes, the *EgPLA2-2* gene contained the maximum number of introns (22). A large proportion of candidate genes contained approximately 6–7 introns (**Figure 2**).

Development of IP Markers Based on the Intron Region of Candidate Genes

A total of 205 primer pairs were designed based on intron region of candidate genes and were used to amplify the DNA templates of different 70 oil palm individuals (**Figure 3**). Amplification products from 64 primer pairs showed polymorphisms between oil palm individuals; the observed heterozygosity varied from 0.0286 to 0.9714, with an average of 0.4514. Meanwhile, a total of 81 alleles were identified, with an average of 2.7 alleles per locus. Sizes of the amplification products varied from 120 bp to 1,300 bp, with an average of 424 bp per PCR product. The 81 alleles were located on the intron region of 18 candidate genes involved in fatty acid biosynthesis, were *ACP4*, *CPT*, *EAR*, *FAD3*, *KASII-1*, *LACS4-1*, *LACS4-2*, *LPAATA*, *DGAT1-1*, *PDH-E2*, *PDH-E1*, *WRI1*, *WRI2*, *FatA*, *LACS9*, *LACS4-2*, *LPCAT-1*, and *LPAATB* (**Table 1**).

Analysis on the Variation of Fatty Acid Content in 70 Oil Palm Individuals

Fatty acid components of the 70 oil palm individuals were analyzed. Myristic acid content varied from 0.54 to 1.99%, with an average of 1.03%; palmitic acid content varied from

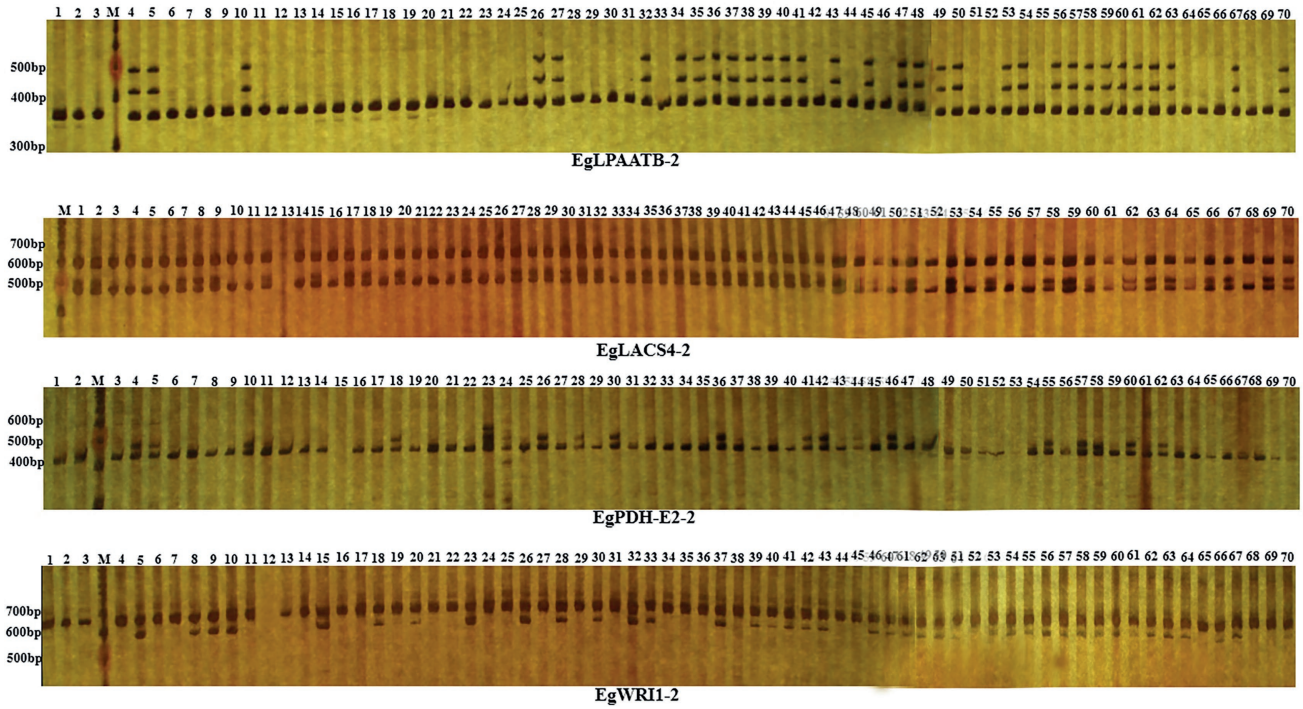


FIGURE 3 | PCR products of IP marker EgLPAATB-2, EgLACS4-2, EgPDH-E2-2 and EgWRI1-2 in 70 oil palm samlpes separated by electrophoresis on 6% non-denaturing PAGE. Lanes 1–70 represent the 70 oil palm samples. M represents 100 bp ladder.

TABLE 1 | Polymorphism information of IP markers developed in the study.

IP marker	Observed allele	Observed heterozygosity	Shannon diversity index
EgACP4-2	3	0.1	0.2727
EgCPT-1	3	0.3143	0.9594
EgCPT-2	3	0.0286	0.9314
EgEAR	3	0.7286	1.0333
EgFAD3-1	3	0.5429	1.0783
EgFAD3-2	2	0.6143	0.6634
EgKASII-1	4	0.0429	0.9785
EgLACS4-1	3	0.3	0.8568
EgLACS4-2-1	2	0.8	0.6895
EgLACS4-2-2	2	0.6286	0.6332
EgLPAATA	3	0.4571	1.0063
EgDGAT1-1	3	0.1	0.7405
EgFAD3-1	3	0.6857	1.0978
EgFAD3-2	2	0.8429	0.6807
EgPDH-E2-1	3	0.0714	0.3068
EgPDH-E1-1-1	2	0.9	0.6881
EgPDH-E1-1-2	2	0.1	0.6915
EgPDH-E2-2	4	0.2571	0.537
EgPDH-E2-2-1	3	0.4429	1.0061
EgPDH-E2-2-2	2	0.8429	0.6807
EgWRI1-1-1	3	0.3	0.9625
EgWRI1-1-2	2	0.1	0.6895
EgWRI1-2	3	0.5286	0.6479
EgFATA	3	0.9	0.9005
EgLPAATA-1	2	0.1	0.6429
EgLPAATA-2	3	0.2143	0.8535
EgLACS9	2	0.9714	0.6927
EgLACS4-2	3	0.8571	0.7497
EgLPCAT-1	3	0.4571	1.0063
EgLPAATB	2	0.3143	0.4349

32.7 to 44.32%, with an average of 39.21%; oleic acid content varied from 35.86 to 53.42%, with an average of 46.13%; linoleic acid content varied from 7.17 to 18.07%, with an average of 11.61%; stearic acid content varied from 0.24 to 1.26%, with an average of 0.65%, and lauric acid content varied from 0.03 to 1.97%, with an average of 0.44% (**Table 2**; **Figure 4**).

The relationships between different fatty acid components were analyzed using SPSS software. Oleic acid content was negatively correlated with myristic acid content ($r = -0.408^{**}$ and $p = 0.00$). Meanwhile, palmitic acid content showed significant negative correlation with oleic acid content ($r = -0.53^{**}$ and $p = 0.00$), indicating that decreasing palmitic acid content in oil palm may enhance oleic acid content. Palmitic acid content and oleic acid content both showed significant negative correlation with linoleic acid content ($r = -0.25^{*}$ and $p = 0.04$, and $r = -0.549^{**}$ and $p = 0.00$ for palmitic acid and oleic acid, respectively). Oleic acid content showed significant negative correlation with stearic acid content ($r = -0.369^{**}$ and $p = 0.002$), while linoleic acid content showed significant positive correlation with stearic acid content ($r = 0.345^{**}$ and $p = 0.003$; **Table 3**).

Association Between IP Markers and Fatty Acid Components

The developed IP markers were used to evaluate the population structure of 70 individuals of *E. guineensis*. When the method of Evanno et al. (2005) was applied to identify the most likely

TABLE 2 | Statistical values of different fatty acid components in the population of *Elaeis guineensis*.

Statistical values	Myristic acid	Palmitic acid	Oleic acid	Linoleic acid	Steric acid	Lauric acid
Maximum value	1.99%	44.32%	53.42%	18.07%	1.26%	1.97%
Minimum value	0.54%	32.70%	35.86%	7.17%	0.24%	0.03%
Average value	1.03%	39.21%	46.13%	11.61%	0.65%	0.44%
Standard deviation	0.33	2.44	3.08	2.12	0.33	0.43
Variation coefficient	0.32	0.06	0.06	0.18	0.52	0.97

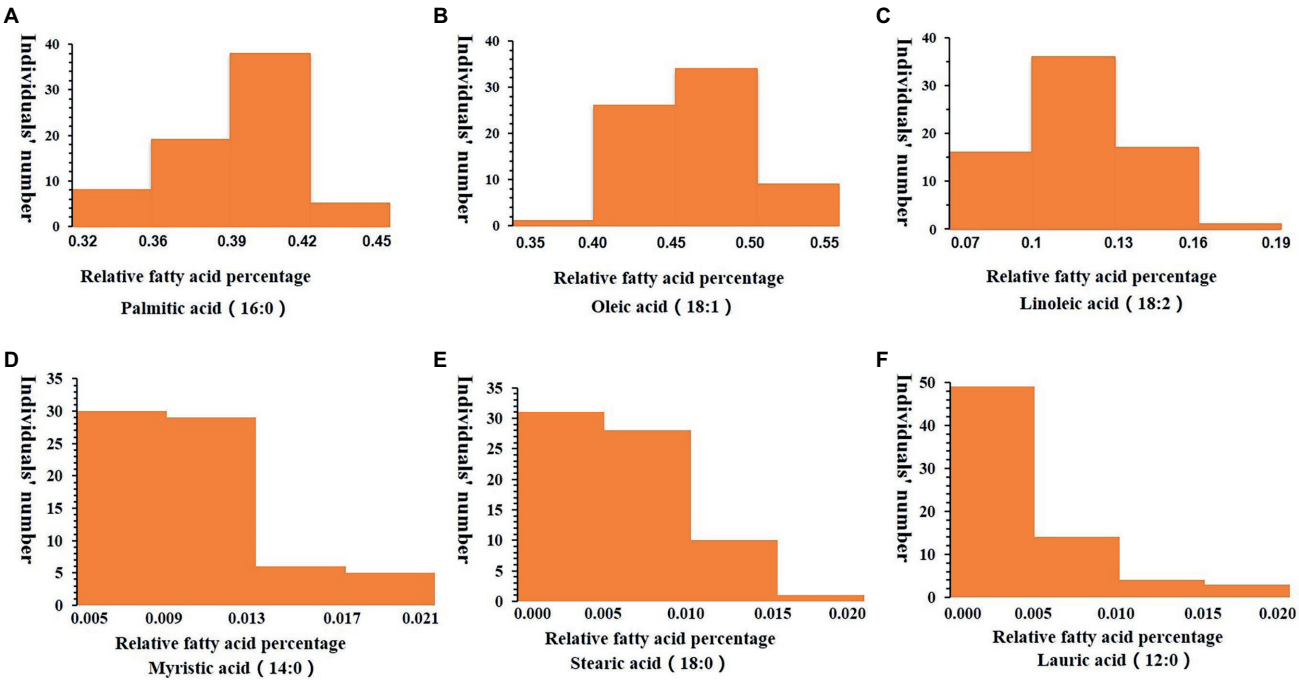


FIGURE 4 | Frequency distribution of relative fatty acid percentages in the mesocarp of 70 oil palm individuals, including palmitic acid (16:0; **A**), oleic acid (18:1; **B**), linoleic acid (18:2; **C**), Myristic acid (14:0; **D**), Stearic acid (18:0; **E**) and Lauric acid (12:0; **F**). The x-axis represents the trait value, and the y-axis represents the number of oil palm individuals.

TABLE 3 | Association between different fatty acid components.

	Myristic acid	Palmitic acid	Oleic acid	Linoleic acid	Stearic acid
Myristic acid	1	0.12	−0.408**	0.05	0.14
Palmitic acid		1	−0.53**	−0.25*	−0.086
Oleic acid			1	−0.549**	−0.369**
Linoleic acid				1	0.345**
Stearic acid					1

**and **represent significant at 5% and 1% probability levels respectively.*

number of 'true population', $K = 3$ genetic groups were found. The STRUCTURE assignment procedure revealed that the largest genetic group comprised 34 oil palm individuals, the second group comprised 19 oil palm individuals, and the remaining group only included 17 oil palm individuals (**Figure 5**).

A simple linear model (SLM) was used to identify the association between IP markers and fatty acid components. Five IP markers were significantly associated with the variation in myristic acid content, including EgLACS9-2 ($p = 0.0356$ and $r^2 = 0.0643$), EgPDH-E2-1-3 ($p = 0.0193$ and $r^2 = 0.079$), EgLACS4-2-3 ($p = 0.0103$ and $r^2 = 0.0943$), EgWRI1-1-2 ($p = 0.0475$ and $r^2 = 0.0574$) and EgLPAATB-2 ($p = 0.0329$ and $r^2 = 0.0662$). Three IP markers showed significant association with linoleic acid, including EgLACS4-1-2 ($p = 0.0386$ and $r^2 = 0.053$), EgPDH-E2-2-2 ($p = 0.0286$ and $r^2 = 0.0591$), and EgFATA-1($p = 0.0108$ and $r^2 = 0.0791$). One IP marker, EgPDH-E2-2, showed significant association with stearic acid content ($p = 0.0237$ and $r^2 = 0.0726$). Furthermore, four IP markers were significantly associated with palmitic acid content, including EgPDH-E2-2 -4 ($p = 0.0134$ and $r^2 = 0.0856$), EgPDH-E2-2 -5 ($p = 0.0261$ and $r^2 = 0.0699$), EgKASII-1-2 ($p = 0.0242$ and $r^2 = 0.0717$), and EgKASII-1-3 ($p = 0.0493$ and $r^2 = 0.055$; **Table 4**).

Expression Analysis of Candidate Genes

To analyze the expression patterns of the candidate genes in different tissues, transcriptomic raw reads data were downloaded

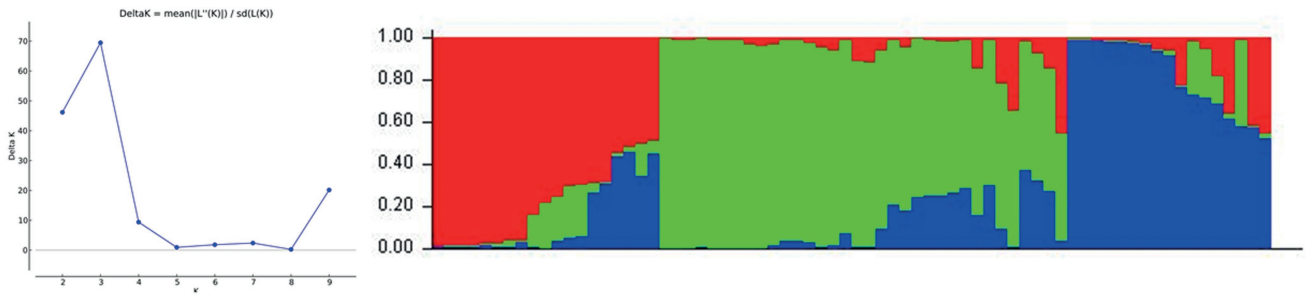


FIGURE 5 | Inferred population structure of 70 oil palm germplasm. Each oil palm individual is represented by a single vertical line. Each color represents one culster. The length of the colored segment indicates the proportion of an individual assigned into one genetic group. The left diagram indicates the true number of genetic clusters.

TABLE 4 | Association between IP markers and fatty acid components.

Trait	Locus	Position	Chromosome	Value of *p*	R^2 (%)
Myristic acid	EgLACS4-2-3	6,030,000	7	0.0103	0.0943
Myristic acid	EgPDH-E2-1-3	20,440,000	2	0.0193	0.079
Myristic acid	EgWRI1-1-2	34,970,000	5	0.0475	0.0574
Myristic acid	EgLACS9-2		–	0.0356	0.0643
Myristic acid	EgLPAATB-2	–	–	0.0329	0.0662
Palmitic acid	EgKASII-1-3	18,370,000	10	0.0493	0.055
Palmitic acid	EgPDH-E2-2-4	22,130,000	9	0.0134	0.0856
Palmitic acid	EgPDH-E2-2-5	22,130,000	9	0.0261	0.0699
Palmitic acid	EgKASII-1-2	18,370,000	10	0.0242	0.0717
Linoleic acid	EgLACS4-1-2	29,170,000	6	0.0386	0.053
Linoleic acid	EgPDH-E2-2-2	22,130,000	9	0.0286	0.0591
Linoleic acid	EgFATA-1	6,070,000	8	0.0108	0.0791
Stearic acid	EgPDH-E2-2-3	22,130,000	9	0.0237	0.0726

from the SRA (Short Read Archive) database of the NCBI website, including SSR851069 (mesocarp 10 weeks after anthesis), SRR190698 (mesocarp 15 weeks after anthesis), SRR190699 (mesocarp 17 weeks after anthesis), SRR190700 (mesocarp 19 weeks after anthesis), SRR190701 (mesocarp 21 weeks after anthesis), SRR190702 (mesocarp 23 weeks after anthesis), SSR851068 (kernel 10 weeks after anthesis), SSR851068 (kernel 15 weeks after anthesis), SSR851099 (pollen), SRR851103 (shoot), and SRR851110 (root). Results of the analysis revealed that almost all candidate genes had a high level of expression in mesocarp and kernel tissues compared with other tissues, except for the gene *EgLACS4-2*. Among them, *EgLPAATB* had higher expression in oil palm kernel compared with that in other tissues. However, seven other candidate genes had a higher expression level in mesocarp and kernel compared with the expression in three other tissues; these genes were *EgKASII-1*, *EgPDH-E2-2*, *EgLACS9*, *EgFATA*, *Eglacs4-1*, *EgWRI1-1*, and *EgPDH-E2-1* (**Figure 6**).

DISCUSSION

This study generated a total of 64 polymorphic IP markers, based on intron sequences of candidate genes involved in fatty acid biosynthesis, for use in molecular breeding of oil palm. These markers were screened across 70 oil palm individuals and are predicted to be suitable for widespread use in oil palm breeding, particularly association analysis. This premise was validated by identifying 13 IP markers linked to different fatty acid compositions; these IP markers will be used immediately for marker-assisted selection.

Molecular markers such as simple sequence repeat (SSR), RAPDs, and AFLPs have been widely used for analyzing genetic diversity and population structure, identification of trait-associated markers, and genotype characterization in *E. guineensis* (Moretzsohn et al., 2000; Rance et al., 2001; Billotte et al., 2005; Cochard et al., 2009; Seng et al., 2011; Premkrishnan and Arunachalam, 2012; Myint et al., 2021). The release of the whole genome sequence of *E. guineensis* has provided an opportunity to develop IP markers with regards to targeted candidate genes. Among the various molecular markers, identifying RFLP markers in a genetically diverse population can be cumbersome as it is mainly based on molecular hybridization (Maizura et al., 2006). Moreover, AFLP markers are based on enzyme digestion and PCR amplification with adapter primers and random primers, which is generally unstable (Barcelos et al., 2002). In recent years, a large number of SNP markers have been identified

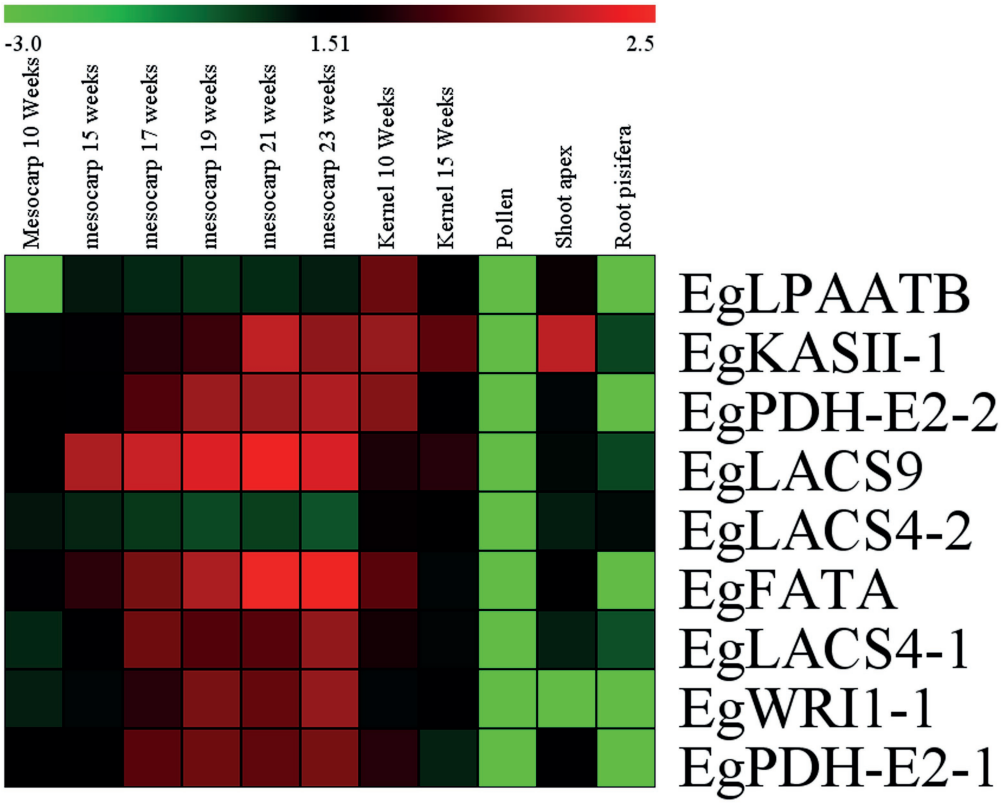


FIGURE 6 | Heat map of candidate genes expression in different tissues of *Elaeis guineensis*. Log_{10}^{RPKM} value were sued to construct the heat map with clustering.

by using next-generation sequencing technology. Xia et al. (2019) revealed 62 SNP markers that were signifcantly associated with fatty acid content, including palmitic acid content (32 SNPs), oleic acid content (4 SNPs), linoleic acid content (1 SNP), and total oil content (25 SNPs) in oil palm. However, SNP remains costly and consequently limits the broad applications of SNP markers for oil palm improvement using these strategies (Ithnin et al., 2021). In contrast, the polymorphisms obtained using IP markers were easily discernable on simple 1.5–2.0% agarose gels and this would enable rapid screening of a large diverse population. In this study, 64 IP markers were developed based on the intron sequences of candidate genes involved in fatty acid biosynthesis, 28.44% of which were polymorphic among different individuals in accordance with previous results (Xia et al., 2019). Moreover, among these IP markers, 20.31% were significantly associated with different fatty acid compositions, including EgLACS4-2-3, EgPDH-E2-1-3, EgWRI1-1-2, EgLACS9-2, EgLPAATB-2, EgFATA-1. The markers EgPDH-E2-2-4, EgPDH-E2-2-5, EgKASII-1-2, and EgKASII-1-3 were significantly associated with variation in palmitic acid content and located in the intron regions of different EgKASII genes that have an important role in carbon extension. Furthermore, EgLACS4-1-2, EgPDH-E2-2-2, and EgFATA-1 markers significantly associated with linoleic acid and located on the *FatA* gene, which is involved in carbon termination of unsaturated fatty acids.

In past several decades, several studies had been performed to identify and validate genes involved in fatty acid biosynthesis. For example, seed-specific RNAi-mediated down-regulation of KASII led to in dramatic increase of palmitic acid (Liu et al., 2017). Our research also showed the IP markers derived from EgKASII-1-2 and EgKASII-1-3 had significant association with palmitic acid composition ($p = 0.0242$ and 0.0493). Meanwhile, our study also indicated that the IP markers located on EgLACS4-1-2 were significantly associated with linoleic acid ($p = 0.0386$ and $r^2 = 0.053$). LACS has been validated to play a role in linoleic acid biosynthesis in a previous study (Jang et al., 2015). Our results also demonstrated that IP markers obtained from EgWRI1-4 and EgLPAAT were significantly associated with myristic acid. In previous studies, LPAAT catalyzed 14:0-acyl-carrier protein specifically and resulted in high myristic acid content in Cyanothece. However, no documents showed the relationship between EgWRI1-4 and myristic acid.

Therefore, it is possible that some IP markers are closely linked with targeted genes that govern fatty acid content and subsequently show significant associations with targeted traits. In future, these candidate markers could be ideal targets for further study and may have potential application in marker-assisted selection for fatty acid composition.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YX and WX participated in the design of the study. JL and YY performed the statistical analysis. JL and PS conducted the major experimental work including the extraction and measurement of oil content and relative fatty acid contents. YX and XL wrote the first draft of the manuscript. YWa, YWu, RL, LZ, and XS wrote sections of the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (no. 31870670).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.885418/full#supplementary-material

REFERENCES

Badoni, S., Das, S., Sayal, Y. K., Gopalakrishnan, S., Singh, A. K., Rao, A. R., et al. (2016). Genome-wide generation and use of informative intron-spanning and intron-length polymorphism markers for high-throughput genetic analysis in rice. *Sci. Rep.* 6:23765. doi: 10.1038/srep23765

Barcelos, E., Amblard, P., Berthaud, J., and Seguin, M. (2002). Genetic diversity and relationship in american and african oil palm as revealed by RFLP and AFLP molecular markers. *Pesquisa Agropecuária Brasileira* 37, 1105–1114. doi: 10.1590/S0100-204X2002000800008

Billotte, N., Marseillac, N., Risterucci, A. M., Adon, B., Brottier, P., Baurens, F. C., et al. (2005). Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 110, 754–765. doi: 10.1007/s00122-004-1901-8

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Cochard, B., Adon, B., Rekima, S., Billotte, N., de Desmier, C. R., and Koutou, A. (2009). Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. *Tree Genet. Genomes* 5, 493–504. doi: 10.1007/s11295-009-0203-3

Deutsch, M., and Long, M. (1999). Intron-exon structure of eukaryotic model organisms. *Nucleic Acids Res.* 27, 3219–3228. doi: 10.1093/nar/27.15.3219

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Ferreira, A. O., Cardoso, H. G., Macedo, E. S., Breviario, D., and Arnholdt-Schmitt, B. (2009). Intron polymorphism pattern in AOX1b of wild St John's wort (*Hypericum perforatum*) allows discrimination between individual plants. *Physiol. Plant.* 137, 520–531. doi: 10.1111/j.1399-3054

Haas, B. J., Wortman, J. R., Ronning, C. M., Hannick, L. I., Smith, R. K., Maiti, R., et al. (2005). Complete reannotation of the *Arabidopsis* genome: methods, tools, protocols and the final release. *BMC Biol.* 3:7. doi: 10.1186/1741-7007-3-7

Hawkins, J. D. A. (1988). A survey on intron and exon lengths. *Nucleic Acids Res.* 16, 9893–9908. doi: 10.1093/nar/16.21.9893

Huang, H. (2017). The industrial survey of *Elaeis guineensis* in Malaysia. *World Trop. Agric. Inf.* 7:7. doi: 10.3969/j.issn.1009-1726.2017.07.020

Ithnin, M., Vu, W. T., Shin, M. G., Suryawanshi, V., Sherbina, K., Zolkafli, S. H., et al. (2021). Genomic diversity and genome-wide association analysis related to yield and fatty acid composition of wild American oil palm. *Plant Sci.* 304:110731. doi: 10.1016/j.plantsci.2020.110731

Jang, Y. E., Kim, M. Y., Shim, S., Lee, J., and Lee, S. H. (2015). Gene expression profiling for seed protein and oil synthesis during early seed development in soybean. *Genes Genomics* 37, 409–418. doi: 10.1007/s13258-015-0269-2

Kita, T., Komatsu, K., Zhu, S., Iida, O., Sugimura, K., Kawahara, N., et al. (2016). Development of intron length polymorphism markers in genes encoding diketide-CoA synthase and curcumin synthase for discriminating *Curcuma* species. *Food Chem.* 194, 1329–1336. doi: 10.1016/j.foodchem.2015.08.034

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-lipid metabolism, Acyl-Lipid Metabolism. *Arabidopsis Book* 11:e0161. doi: 10.1199/tab. 0161

Liu, Q., Wu, M., Zhang, B., Shrestha, P., Petrie, J., Green, A. G., et al. (2017). Genetic enhancement of palmitic acid accumulation in cotton seed oil through RNAi down-regulation of ghKAS2 encoding β-ketoacyl-ACP synthase II (KASII). *Plant Biotechnol. J.* 15, 132–143. doi: 10.1111/pbi.12598

Liu, Y., Xu, A., Liang, F., Yao, X., Wang, Y., Liu, X., et al. (2018). Screening of clubroot-resistant varieties and transfer of clubroot resistance genes to *Brassica napus* using distant hybridization. *Breed. Sci.* 68, 258–267. doi: 10.1270/jsbbs.17125

Maizura, I., Rajanaidu, N., Zakri, A. H., and Cheah, S. C. (2006). Assessment of genetic diversity in oil palm (*Elaeis guineensis* Jacq.) using restriction fragment length polymorphism (RFLP). *Genet. Resour. Crop. Evol.* 53, 187–195. doi: 10.1007/s10722-004-4004-0

Moretzsohn, M. C., Nunes, C. D. M., Ferreira, M. E., and Grattapaglia, D. (2000). RAPD linkage mapping of the shell thickness locus in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 100, 63–70. doi: 10.1007/s001220050009

Mozzon, M., Foligni, R., and Mannozzi, C. (2020). Current knowledge on interspecific hybrid palm oils as food and food ingredient. *Foods* 9, 631–646. doi: 10.3390/foods9050631

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4326. doi: 10.1093/nar/8.19.4321

Myint, K. A., Yaakub, Z., Rafii, M. Y., Oladosu, Y., Samad, M., Ramlee, S. I., et al. (2021). Genetic diversity assessment of MPOB-Senegal oil palm germplasm using microsatellite markers. *Biomed. Res. Int.* 2021:6620645. doi: 10.1155/2021/6620645

Panjabi, P., Jagannath, A., Bisht, N. C., Padmaja, K. L., Sharma, S., Gupta, V., et al. (2008). Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using intron polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C Brassica genomes. *BMC Genomics* 9:113. doi: 10.1186/1471-2164-9-113

Premkrishnan, B. V., and Arunachalam, V. (2012). In silico RAPD priming sites in expressed sequences and iSCAR markers for oil palm. *Comp. Funct. Genomics* 2012:913709. doi: 10.1155/2012/913709

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Rance, K. A., Mayes, S., Price, Z., Jack, P. L., and Corley, R. H. V. (2001). Quantitative trait loci for yield components oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* 103, 1302–1310. doi: 10.1007/s122-001-8204-z

Rozen, S., and Skaletsky, H. (2000). Primer 3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365–386. doi: 10.1385/1-59259-192-2:365

Seng, T.-Y., Mohamed Saad, S. H., Chin, C.-W., Ting, N.-C., Harminder Singh, R. S., Qamaruz Zaman, F., et al. (2011). Genetic linkage map of a

high yielding FELDA deli×Yangambi oil palm cross. *PLoS One* 6:e26593. doi: 10.1371/journal.pone.0026593

Sharma, H., Bhandawat, A., Rahim, M. S., Kumar, P., Choudhoury, M. P., and Roy, J. (2020). Novel intron length polymorphic (ILP) markers from starch biosynthesis genes reveal genetic relationships in Indian wheat varieties and related species. *Mol. Biol. Rep.* 47, 3485–3500. doi: 10.1007/s11033-020-05434-2

Singh, R., Ong-Abdullah, M., Low, E. T., Manaf, M. A., Rosli, R., Nookiah, R., et al. (2013). Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature* 500, 335–339. doi: 10.1038/nature12309

Singh, R., Tan, S. G., Panandam, J. M., Rahman, R. A., Ooi, L. C., Low, E. T., et al. (2009). Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. *BMC Plant Biol.* 9:114. doi: 10.1186/1471-2229-9-114

Stoltzfus, A., Logsdon, J. M., Palmer, J. D., and Doolittle, W. F. (1997). Intron "sliding" and the diversity of intron position. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10739–10744. doi: 10.1073/pnas.94.20.10739

Ting, N. C., Jansen, J., Nagappan, J., Ishak, Z., Chin, C. W., Tan, S. G., et al. (2013). Identification of QTLs associated with callogenesis and embryogenesis in oil palm using genetic linkage maps improved with SSR markers. *PLoS One* 8:e53076. doi: 10.1371/journal.pone.0053076

Wang, X., Zhao, X., Zhu, J., and Wu, W. (2005). Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). *DNA Res.* 12, 417–427. doi: 10.1093/dnares/dsi019

Wei, H., Fu, Y., and Arora, R. (2005). Intron-flanking EST-PCR markers: from genetic marker development to gene structure analysis in Rhododendron. *Theor. Appl. Genet.* 111, 1347–1356. doi: 10.1007/s00122-005-0064-6

Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18, 6531–6535. doi: 10.1093/nar/18.22.6531

Xia, X., Luan, L. L., Qin, G., Yu, L. F., Wang, Z. W., Dong, W. C., et al. (2017). Genome-wide analysis of SSR and ILP markers in trees: diversity profiling, alternate distribution, and applications in duplication. *Sci. Rep.* 7:17902. doi: 10.1038/s41598-017-17203-6

Xia, W., Luo, T. T., Dou, Y. J., Zhang, W., Mason, A. S., Huang, D. Y., et al. (2019). Identification and validation of candidate genes involved in fatty acid content in oil palm by genome-wide association analysis. *Front. Plant Sci.* 10:1263. doi: 10.3389/fpls.2019.01263

Yang, L., Jin, G., Zhao, X., Zheng, Y., Xu, Z., and Wu, W. (2007). PIP: a database of potential intron polymorphism markers. *Bioinformatics* 23, 2174–2177. doi: 10.1093/bioinformatics/btm296

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fpls.2022.929197

Genome-Wide Association Analysis Combined With Quantitative Trait Loci Mapping and Dynamic Transcriptome Unveil the Genetic Control of Seed Oil Content in *Brassica napus* L.

Chuanji Zhao[1,2†], Meili Xie[2†], Longbing Liang[1], Li Yang[2,3], Hongshi Han[1], Xinrong Qin[1], Jixian Zhao[1], Yan Hou[1], Wendong Dai[1], Caifu Du[1], Yang Xiang[1], Shengyi Liu[2]* and Xianqun Huang[1]*

[1]Guizhou Rapeseed Institute, Guizhou Academy of Agricultural Sciences, Guiyang, China, [2]Key Laboratory of Biology and Genetic Improvement of Oil Crops, The Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China, [3]Biosystematics Group, Experimental Plant Sciences, Wageningen University and Research, Wageningen, Netherlands

Rapeseed, an allotetraploid oil crop, provides vegetable oil for human consumption. The growing demand for oilseeds has necessitated the development of rapeseed varieties with improved quality. Therefore, a clear understanding of the genetic basis underlying the seed oil content (SOC) is required. In this study, a natural population comprising 204 diverse accessions and recombinant inbred lines (RILs) derived from *Brassica napus* and *Sinapis alba via* distant hybridization were collected for genome-wide association analysis (GWAS) and quantitative trait loci (QTL) mapping of the SOC trait, respectively. The variable coefficient of the RIL and natural populations ranged from 7.43 to 10.43% and 8.40 to 10.91%. Then, a high-density linkage map was constructed based on whole genome re-sequencing (WGS); the map harbored 2,799 bin markers and covered a total distance of 1,835.21 cM, with an average marker interval of 0.66 cM. The QTLs for SOC on chromosome A07 were stably detected in both single and multiple environments. Finally, a novel locus *qA07.SOC* was identified as the major QTL for SOC based on the GWAS and RIL populations. In addition, the RNA-seq results showed that photosynthesis, lipid biosynthesis proteins, fatty acid metabolism, and unsaturated fatty acid biosynthesis were significantly different between the developed seeds of the two parents of the RIL population. By comparing the variation information and expression levels of the syntenic genes within *qA07.SOC* and its syntenic genomic regions, as well as through haplotype analysis *via* GWAS, *BnaA07.STR18*, *BnaA07.NRT1*, and *BnaA07g12880D* were predicted as candidate genes in the *qA07.SOC* interval. These stable QTLs containing candidate genes and haplotypes can potentially provide a reliable basis for marker-assisted selection in *B. napus* breeding for SOC.

Keywords: *Brassica napus*, seed oil content, distant hybridization, QTL mapping, syntenic gene, GWAS

OPEN ACCESS

Edited by:
Hongbo Chao,
Zhengzhou University, China

Reviewed by:
Kunjiang Yu,
Guizhou University, China
Chengming Sun,
Huazhong Agricultural University, China
Berisso Demo Kebede,
University of Alberta, Canada

****Correspondence:***
Yang Xiang
xiangyangcell@126.com
Shengyi Liu
liusy@oilcrops.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

Received: *26 April 2022*
Accepted: *13 June 2022*
Published: *01 July 2022*

Citation:
Zhao C, Xie M, Liang L, Yang L, Han H, Qin X, Zhao J, Hou Y, Dai W, Du C, Xiang Y, Liu S and Huang X (2022) Genome-Wide Association Analysis Combined With Quantitative Trait Loci Mapping and Dynamic Transcriptome Unveil the Genetic Control of Seed Oil Content in Brassica napus L. Front. Plant Sci. 13:929197. doi: 10.3389/fpls.2022.929197

INTRODUCTION

Rapeseed (*Brassica napus* L., 2n = 38, AACC) belongs to the family *Brassicaceae* and is one of the major oilseed crops that provide vegetable oil for human consumption worldwide. With an increase in global human population, improving seed oil content (SOC) has become an important goal in oilseed rape breeding programs (Sun et al., 2018); accordingly, understanding the genetic basis of SOC is necessary to improve the breeding value of rapeseed.

Most agronomic traits of crop plants are simultaneously controlled by multiple genes and influenced by the environment (Zhou et al., 2021). In the case of rapeseed, seed oil- contributing traits are of utmost importance, and they reflect the accumulated effect of a large number of genes expressed throughout the life span of the plant. However, the final plant phenotype is the result of interactions between environmental and genetic factors. Therefore, a genotype × environment interaction is an important aspect to consider when dissecting agronomic traits in *B. napus* (Xie et al., 2020).

Quantitative trait loci (QTL) mapping and genome-wide association analysis (GWAS) are effective strategies to dissect complicated genetic bases, and can help accelerate the breeding of rapeseed by marker-assisted selection. Several studies have reported multiple QTLs for the seed oil-related traits, distributed over all chromosomes in the rapeseed genome (Basunanda et al., 2010; Wang et al., 2010; Wei et al., 2011; Yang et al., 2012; Liu et al., 2015; Guo et al., 2017; Sun et al., 2018; Shi et al., 2019). Similarly, since the release of the *B. napus* reference genome, GWAS of seed oil content has substantially contributed to the related literature. A total of 52,157 SNP markers were obtained in 521 *B. napus* accessions using a 60 K chip array, and 50 QTLs were obtained by GWAS for oil content, of which 29 were newly identified (Liu et al., 2016).

The SOC of *B. napus* is mainly controlled by maternal effects, embryo gene effects, pollen instinct, cytoplasmic effects, and the corresponding gene-environmental interaction effects, which conforms to the additive-dominant-epistatic genetic model and are based on additive and dominant inheritance with high generalized heritability (Wang et al., 2010; Hua et al., 2012; Guo et al., 2017). Oil content has an obvious dynamic trend in the process of seed development and is closely related to multiple biological pathways, such as plant photosynthesis, seed development, material transport, lipid synthesis, accumulation and degradation, all of which form a regulatory network regulated by multiple genes (Liu et al., 2012; Chao et al., 2017; Shi et al., 2017). The biosynthesis of seed oil is divided into two stages: fatty acid synthesis and triacylglycerol synthesis, and is mainly stored as triacylglycerol, which accounts for 60% of the seed mass (Frentzen, 2010). The synthesis of triacylglycerol requires the interaction of many subcellular structures and multiple pathways, and many regulatory and enzyme genes are involved in the whole process (Hills, 2004; Bates et al., 2014). The key enzyme-encoding genes play an important role in the synthesis of rapeseed oil, and its effect on the SOC of *B. napus* mainly depends on the expression of the gene and the regulation the enzyme activity, The genes involved in this process mainly catalyze the synthesis of fatty acid chains and triacylglycerol, including *ACCase*, *Fat A*, *Fat B*, *GPAT*, *LPAAT*, and *DGAT* (Turnham and Northcote, 1983; Weselake et al., 2008; Lock et al., 2009). Problems in the expression or regulation of any kind of enzymes will affect the SOC. Key transcription factors also affect the SOC in *B. napus*, mainly regulating seed development and seed oil accumulation. Previous studies have shown that transcriptional factors, including *WRI1*, *LEC1*, *LEC2*, *FUS3*, and *ABI3* can increase seed oil content by activating or inhibiting the expression of genes related to seed oil synthesis (Kagaya et al., 2005; Wang et al., 2007; Wu et al., 2014; Elahi et al., 2015, 2016). In addition, miRNA may also play an important role in fatty acid and lipid metabolism in the seeds of *B. napus* (Wang et al., 2017). Therefore, a combination of QTL mapping and dynamic transcriptome analysis is an effective strategy for uncovering the genetic basis of the SOC trait in *B. napus*.

In this study, we collected 204 accessions of *B. napus* for GWAS and constructed a recombinant inbred lines (RIL) population of 158 individuals. We used this RIL population to construct a high-density genetic map of 2,799 bin markers, with a total coverage distance of 1,835.21 cM. We evaluated SOC traits across nine different environments for QTL mapping in the RIL population and 3 consecutive years for GWAS. A major QTL on chromosome A07 was detected. Combined dynamic transcriptome, synteny analysis, variant detection, and haplotype analysis were conducted in the major QTL region, and *BnaA07g12790D*, *BnaA07g12830D*, and *BnaA07g12880D* were predicted to be candidate genes. Our findings could contribute to improving the understanding of the genetic basis and breeding of the SOC trait in *B. napus*.

MATERIALS AND METHODS

Plant Materials and Phenotypic Evaluation

A *Sinapis alba* L. inbred line with a high 1,000-seed wight was selected as a male parent to cross with a *B. napus* L. inbred line "*Darmor*." Well-grown F_1 lines were obtained by *in vitro* culture of plant embryos for embryo rescue. Using the microspore culture method, GRG2462 was generated in $F_{2:3}$ lines with low SOC as female parent. We selected another high-SOC *B. napus* inbred line GRD328 as the male parent, in which the high-oil-content hybrid rapeseed Youyan2020 was developed using GRD328. Finally, a *B. napus* RIL population ($F_{2:11}$) consisting of 158 individuals was derived from the crossing of GRG2462 × GRD328, and used for QTL mapping study. A total of 204 *B. napus* inbred lines (**Supplementary Table 1**) collected worldwide were used for GWAS of the SOC trait. In the field, 120 individuals of each inbred line were planted in a 2.5 × 2.0 m^2 plot, and the field tests followed a randomized design. The field experiments for QTL mapping in the RIL population were replicated in six locations, namely Guiyang (GY), Qinghe (QH), Changshun (CS), Tangtou (TT), Jinsha (JS), and Badong (BD).

At maturity, the natural population, RIL population, and two parents (GRG2462 and GRD328) were evaluated for SOC in each environment. Foss NIR Systems 5000 Near-Infrared Reflectance Spectroscopy was performed to measure SOC (Tang et al., 2019). SPSS 22 (IBM SPSS, Armonk, NY, United States) was used to calculate the phenotypic variation and frequency distribution.

Whole-Genome Resequencing, SNP Calling, and Genotyping

All lines in the RIL and GWAS populations were sampled, and genomic DNA was extracted using the Hi-DNAsecure Plant Kit (TIANGEN, Beijing). Genomic DNA was diluted to a final concentration of 20 ng/μl and used to construct a DNA library. The DNA library was subjected to whole-genome resequencing (WGS) using the Illumina NovaSeq 6000 system. Clean reads were obtained by removing adaptor sequences, contamination, and low-quality reads from the raw reads. A Burrows-Wheeler Aligner (BWA; parameter: mem -t 4 -k 32 -M; http://bio-bwa.sourceforge.net/) was used to align the clean reads against the *B. napus* "*Darmor-bzh*" reference genome (https://www.genoscope.cns.fr/brassicanapus/data/; Li and Durbin, 2009; Chalhoub et al., 2014). Alignment duplication was removed using SAMTools (parameter: rmdup; http://samtools.sourceforge.net/; Li et al., 2009). SNP calling within the mapping population was performed by an in-house pipeline in the "Sentieon Genomics" tool (Freed et al., 2017). To exclude false variants, SNPs were filtered by GATK (version 4.1.4.0) based on the following parameters: QUAL < 30.0||MQ < 50.0||QD < 2.0, cluster size 3, and cluster window size 10 (Mckenna et al., 2010). Subsequently, 15 SNPs with a sliding window size and an SNP step size were used to screen the chromosomes. When the number of allelic SNP A:A (or B:B) in the sliding window was less than 11, allelic SNP A:B was used as the genotype to fill and correct this position (Huang et al., 2009).

For subsequent genetic analysis of the RIL population, the recombination of all offspring genotypes in the RIL population was analyzed to divide the bin markers. No recombination events were considered to be present in one Bin marker interval. To ensure the quality of the genetic map, bin markers were filtered according to the following criteria: (1) being less than 10 kb in length; (2) having severe partial separation (chi-square test, $^{***}p < 0.001$); (3) chromosomes with a small density of bin marker, add some Bin markers with unfixed parameters in terms of length and partial separation. The bin markers that met the criteria were used to construct a high-density genetic map. As per general genetic coding rules, the model "aa × bb" was used to analyze the genotype bin markers in the 158 RILs (Ren et al., 2020).

Construction of a Genetic Linkage Map

The selected bin markers were assigned to 19 chromosomes in the *B. napus* reference genome according to their genetic position. Each chromosome was considered the corresponding linkage group. The HighMap software (http://highmap.biomarker.com.cn/; Liu et al., 2014) was used to combine co-segregating markers (SNP and/or InDel) into bin markers and estimate the genetic distance between adjacent markers in each linkage group. In the genetic map, centimorgan (cM) distances were calculated to map genetic distances (Kosambi, 2016). To evaluate the quality of the genetic map, we analyzed the pair-wise recombination values of all mapped markers by the Kosambi mapping function of RECORD_WIN (Truco et al., 2013), as well as the collinearity between the genetic map and the *B. napus* reference genome.

QTL Identification and Analysis

The QTL for SOC in a single environment was identified using the composite interval mapping (CIM) method (Zeng, 1994) in Windows QTL Cartographer software (WinQTLCart, version 2.5; https://brcwebportal.cos.ncsu.edu/qtlcart/WQTLCart.htm; Silva et al., 2012). The logarithm of the odds (LOD) threshold was calculated by a permutation test, with the following parameters: 1.0 cM intervals with a mapping window of 10 cM, five control markers, and a significance level of $p < 0.05$, $n = 1{,}000$ permutation (Zhang et al., 2019).

Quantitative trait loci × environment interaction effects (QTL by environment interaction in biparental population) were identified by the MET functionality and ICIM-ADD mapping method in the QTL IciMapping V4.1 software (http://www.isbreeding.net/; Li et al., 2008). Determination of the LOD threshold was followed by a significance test ($p < 0.05$) with 1,000 permutations.

GWAS for SOC

The SOC trait for three consecutive years (2014–2015 CS, 2015–2016 CS, and 2016–2017 CS) was measured, and the best linear unbiased prediction (BLUP) for each accession was obtained using an R script lme4 (CRAN-Package lme4 (r-project.org)) and lsmeans (Lu et al., 2019). The ADMIXTURE software (Version 1.3.0) was used to analyze the population structure in the natural populations (Alexander et al., 2009). The subgroup number (k) was set as 1–10, and the probability value (Q) of each material in each subgroup was calculated. The optimal number of subgroups was determined according to the cross-validation error (CV value), and each material was classified into the subgroup to which it belonged according to the Q value. The TASSEL 5.0 software was used to calculate the genetic relatedness between two specific accessions and the relative value of genetic relatedness between any accessions, which determined the kinship of the *B. napus* accessions in the natural population (Yu et al., 2006). The PopLDdecay and LDBlockShow softwares were used to analyze the linkage disequilibrium (LD) of the whole genome and the specific region (Zhang et al., 2018; Dong et al., 2020). The Efficient Mixed-Model Association eXpedited (EMMAX) was used to analyze the relative kinship of the natural population. A Mixed Linear Model (MLM) with EMMAX, and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) were adopted for association analysis (Huang et al., 2019). The significant p value thresholds of the GWAS were set as

$-\log_{10}P$ ($p = 1$/total SNPs). The quantile-quantile (Q-Q) plot was shown with the expected p value and $-\log_{10}P$ of each SNP, and the Manhattan plot was demonstrated using the R package qqman.

RNA-Seq and qRT–PCR Analysis

The developed seeds of GRD328 and GRG2462 at 35, 40, 45, and 50 days after flowering (DAF) were collected for RNA extraction and RNA-seq. High-quality RNA was extracted using the RNA Prep Pure Plant Kit (Tiangen, Beijing, China) and sequenced using the Illumina NovaSeq 6000 system. Clean reads were acquired, by filtering the raw reads, and then mapped to the *B. napus* "*Darmor-bzh*" reference genome[1] using the HISAT2 software (Chalhoub et al., 2014; Kim et al., 2015). The FPKM value was used to estimate gene expression levels (Trapnell et al., 2010). Pearson's correlation coefficient was calculated using the R package to detect correlations between biological replicates (Leng et al., 2013). The DESeq software was used to test the differentially expressed genes (DEGs) based on the selection criteria $|\log_2$ (Fold Change) $|>1$ and $p_{adj}<0.05$ (Anders and Huber, 2010). The Kyoto Encyclopedia of Genes and Genomes (KEGG) of DEGs and heatmaps were generated using the OmicShare tools.[2]

Total RNA was reverse transcribed using the HiScript III 1st Strand cDNA synthesis kit (Vazyme, Nanjing, China) following the manufacturer's instructions. Quantitative real-time PCR (qRT-PCR), with the *BnaActin* gene as an internal control, was performed in a CFX Connect Real-time PCR system (Bio-Rad, United States) using ChamQ™ SYBR qPCR Master Mix (Vazyem, Nanjing, China). Relative gene expression levels were determined using the $2^{(-\Delta\Delta CT)}$ method (Livak and Schmittgen, 2001). The primers used for qRT-PCR are listed in **Supplementary Table 2**.

[1]https://www.genoscope.cns.fr/brassicanapus/data/
[2]https://www.omicshare.com/tools/

Candidate Genes Analysis in the Major QTL Interval

The syntenic properties of the genomic regions were used to exclude genes while determining the candidates responsible for QTL-SOC (Zhang et al., 2019). The positional information of SNP and bin markers was used to determine the major QTL interval in the *B. napus* reference genome. The two parents were deeply sequenced on the Illumina NovaSeq 6000 system, and the homozygous polymorphic SNP and InDel between the two parents was used for variant analysis in SnpEff 5.1.[3] The genes within the major QTL and its syntenic block (or QTL) of another sub-genome in *B. napus* were selected to analyze the unique gene and syntenic gene, along with their variation information and expression levels, based on the deep WGS and RNA-seq data, respectively.

RESULTS

Phenotypic Variation of the GWAS and RIL Population

Seed oil content showed significant differences between the two parents in each environment (**Supplementary Table 3**). The frequency distributions of SOC in the RILs in the nine environments displayed normal distributions with considerable transgressive segregation (**Supplementary Figure 1**). In summary, combined with the SOC trait in nine environments, the minimum SOC was less than 35%, while the maximum was over 41% (**Table 1**), and the coefficient of variation ranged from 7.43 to 10.91% (**Table 1**). To test the effects of genotype (G), environment (E), and their interactions (G × E), a variance analysis (ANOVA) was conducted, and significant variation ($p < 0.01$) was observed among environments and genotypes, with the SOC having high narrow sense heritability of 78.54% (**Supplementary Table 4**).

[3]https://pcingola.github.io/SnpEff/

TABLE 1 | Phenotypic variations in the recombinant inbred line (RIL) and natural population.

Population Type	Environment	Min	Max	Mean	*SE*	*SD*	Var	Kurtosis	Skewness	CV (%)
RIL	2012–2013 GY	26.4	45.83	38.78	0.35	4.06	16.52	0.895	−0.883	10.47
	2013–2014 QH	27.93	45.42	36.49	0.27	3.31	10.95	−0.210	0.022	9.07
	2014–2015 CS	34.01	47.74	41.56	0.3	3.29	10.8	−0.592	−0.498	7.91
	2015–2016 GY	25.01	42.1	36.2	0.2	2.69	7.21	1.898	−0.791	7.43
	2015–2016 CS	28.72	41.71	35.53	0.24	2.72	7.4	−0.317	−0.245	7.66
	2016–2017 CS	24.07	48.12	39.84	0.29	3.49	12.2	1.872	−0.589	8.76
	2017–2018 TT	31.06	49.03	38.32	0.25	3.07	9.41	0.611	0.292	8.02
	2017–2018 JS	21.67	46.74	38.87	0.27	3.29	10.84	4.842	−1.237	8.46
	2018–2019 CS	29.94	43.52	37	0.23	2.81	7.89	−0.343	−0.123	7.59
GWAS	2014–2015 CS	32.02	50.77	42.24	0.27	3.82	14.62	−0.316	0.081	9.04
	2015–2016 CS	33.50	49.01	41.17	0.24	3.46	11.96	−0.733	0.092	8.40
	2016–2017 CS	29.24	53.39	41.53	0.32	4.53	20.53	−0.169	0.207	10.91
	BLUP	33.69	49.29	40.60	0.22	3.20	10.26	−0.286	0.547	7.88

Env, environment; Min, minimum value; Max, maximum value; Mean, mean value; SE, standard error; SD, standard deviation; Var, variance; and CV, coefficient of variation.

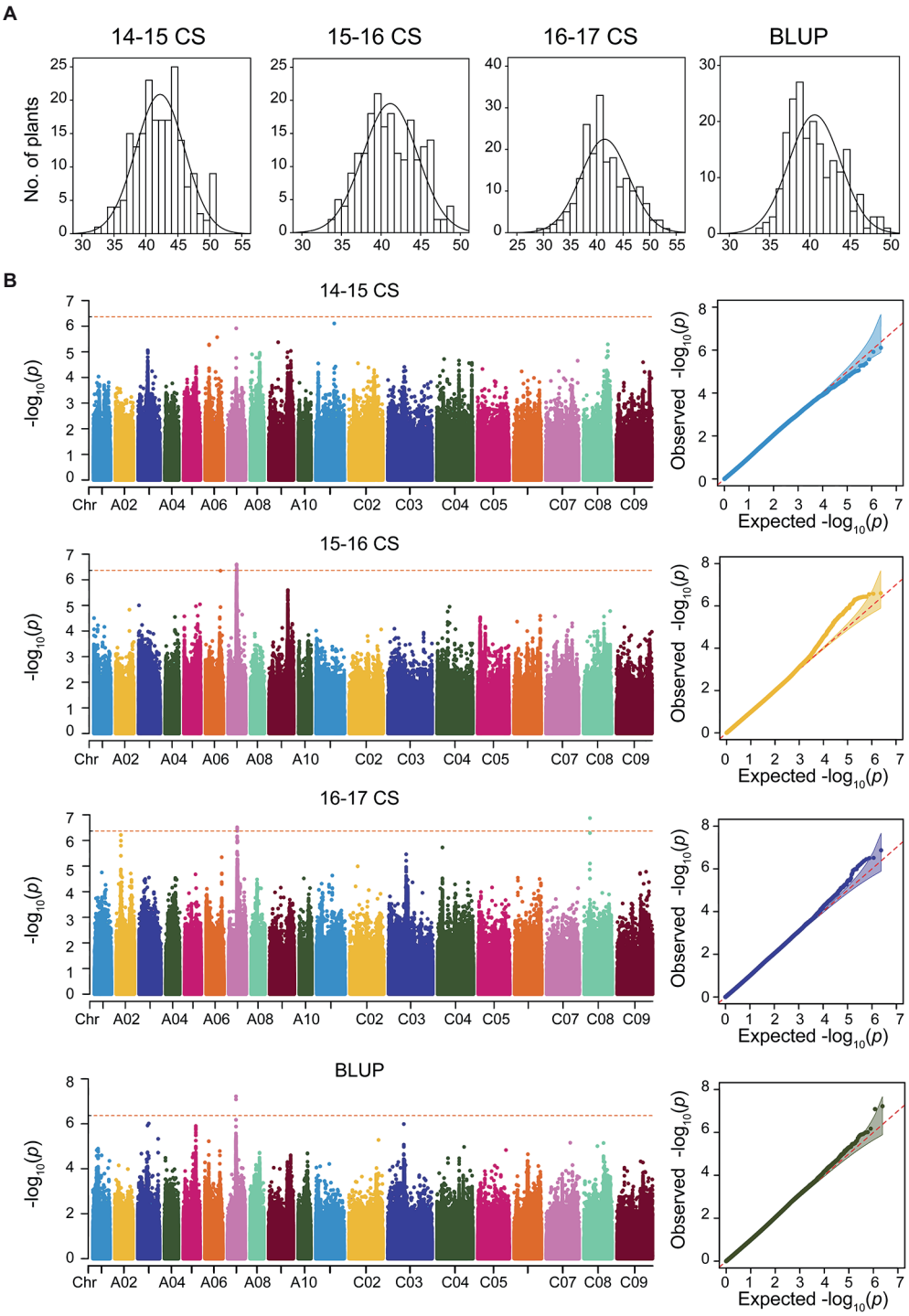


FIGURE 1 | Genome-wide association analysis (GWAS) for the seed oil content (SOC) trait in consecutive 3 years based on Mixed Linear Model (MLM) of Efficient Mixed-Model Association eXpedited (EMMAX). **(A)** The phenotype distributions of SOC in consecutive 3 years and best linear unbiased prediction (BLUP). **(B)** GWAS and Q-Q plot for SOC in consecutive 3 years and BLUP.

In the natural population, the frequency distributions for SOC in the consecutive 3 years and BLUP displayed normal distributions (**Figure 1A**). A significant variation in SOC was observed in the natural population; for example, SOC ranged from 33.69 to 49.29%, with an average of 40.60% in BLUP; the coefficient of variation was 7.88% (**Table 1**).

The correlations between environments for association and RIL population were, respectively, less than 75 and 45% (**Supplementary Figures 2A,B**), suggesting that the phenotypes were greatly influenced by the environments. Overall, these findings indicate that the SOC trait was stably inherited and was suitable for QTL mapping and GWAS.

Identification of Association Loci for Seed Oil Content

Based on the WGS data, a total of 2,340,881 SNPs were retained and unevenly distributed on all 19 *B. napus* chromosomes with minor allele frequency (MAF) > 0.05 in the association panel (**Supplementary Figure 3**; **Table 2**). The average SNP number is 123,204 on each chromosome (**Table 2**). According to the Q and K value, we confirmed that the 204 accessions were divided into four groups (**Figures 2A,B**). The linkage-disequilibrium (LD) distances of the A and C sub-genomes were 38 and 434 kb ($r^2 > 0.2$), respectively (**Figure 2C**). In the 204 accessions, the relative kinships among about 89% of the 204 accessions was zero (**Figure 2D**), suggesting that the genetic relationship between the accessions of natural population was weak and had little influence on further association analysis. According to the results of the MLM with EMMAX, a total of15 SNP loci on chromosomes A07 and C08 associated with SOC were identified at the suggestive threshold excepted 2014–2015 CS environment, and the SNP_11440137 on chromosomes A07 was identified repeatedly in 2016–2017 CS and BLUP (**Figure 1B**; **Table 3**). In additional, even if no significant SNPs were detected in 14–15 CS environment, A07_11440137 and A07_11440139 can be identified with $-\log_{10}(P) > 5.0$ (**Table 3**). Based on the BLINK model, we identified the significant SNP A07_ 11440137 in 2016_2017 CS, as well as A07_ 11440137 and A07_ 11440139 in BLUP (**Supplementary Figure 4**; **Supplementary Table 5**), which suggested that the BLINK results were as same as those of EMMAX in 2016_2017 CS and BLUP. In the BLINK model, the SNPs A07_11440137 and A07_11440139 were also identified with $\log_{10}(P) > 5.0$ in 2015_2016 CS (**Supplementary Table 5**).

TABLE 2 | The statistics of SNP number on each chromosome.

Chromosome	SNP-number	Chromosome length	SNP/kb
A01	107,089	23,267,856	4.6
A02	105,683	24,793,737	4.26
A03	165,992	29,767,490	5.58
A04	113,531	19,151,660	5.93
A05	147,616	23,067,598	6.4
A06	145,427	24,396,386	5.96
A07	142,496	24,006,521	5.94
A08	89,148	18,961,941	4.7
A09	158,319	33,865,340	4.67
A10	107,570	17,398,227	6.18
C01	128,195	38,829,317	3.3
C02	113,683	46,221,804	2.46
C03	185,302	60,573,394	3.06
C04	141,790	48,930,237	2.9
C05	85,975	43,185,227	1.99
C06	91,498	37,225,952	2.46
C07	129,098	44,770,477	2.88
C08	103,876	38,477,087	2.7
C09	78,593	48,508,220	1.62

Construction of a *Brassica napus* High-Density Linkage Map Based on Whole-Genome Resequencing

For linkage map construction, 158 RILs were performed for whole genome re-sequencing, and a total of 617,466 SNP markers were obtained to constitute 38,516 bin markers (**Supplementary Table 6**). Subsequently, with division and filtration, 2,799 Bin markers harboring 184,183 SNPs were successfully selected to genotype the 158 RILs (**Supplementary Figure 5**; **Supplementary Table 7**).

Based on the effective SNPs and bin markers, the map harbored 2,799 mapped bin markers, spanning a total distance of 1,835.21 cM with an average distance of 0.66 cM between adjacent markers (**Figure 3A**; **Table 4**). The A-sub-genome harbored 1,577 mapped bin markers with a total distance of 968.89 cM whereas the C-sub-genome harbored 1,222 mapped bin markers with a total distance of 866.32 cM (**Figure 3A**; **Table 4**). Among the 19 linkage groups (LG), the longest one was LG C06, reaching 145.79 cM with an average interval distance of 0.87 cM, and the shortest one was LG C04 at only 68.15 cM with an average interval distance of 0.59 cM (**Figure 3A**; **Table 4**). The linkage group with the largest number of mapped bin markers was LG A03, which harbored 240 markers, and the linkage group with the least number of bins was LG C09, with only 99 bin markers (**Table 4**). In the whole genetic map, about 97.35% of the gaps were less than 5 cM, and the max gap of each linkage group ranged from 2.07 to 21.7 cM (**Table 4**).

The collinearity analysis of the genetic map and physical map of the *B. napus* reference genome showed that the mapped bin markers were consistent with the *B. napus* genome (**Figure 3B**; **Supplementary Tables 8**, **9**). The average value of Spearman's correlation coefficient was 99.15%, which suggested that each linkage group had high collinearity with the corresponding chromosome (**Supplementary Table 10**). Collinearity analysis indicated that the genetic map was exhibiting a high quality.

QTL Mapping for SOC in the RIL Population

In the single environment, a total of nine QTLs with the PVE ranging from 7.97 to 15.07% were identified on chromosomes A02, A03, A05, A06, A07, A09, and C08 in six environments (**Figure 4A**; **Table 5**). Among these QTLs, both *qSOC-SE-5* with the highest LOD value of 6.9 and *qSOC-SE-4* with the highest PVE of 15.07% were identified on chromosome A07 (**Table 5**). In multiple environments,

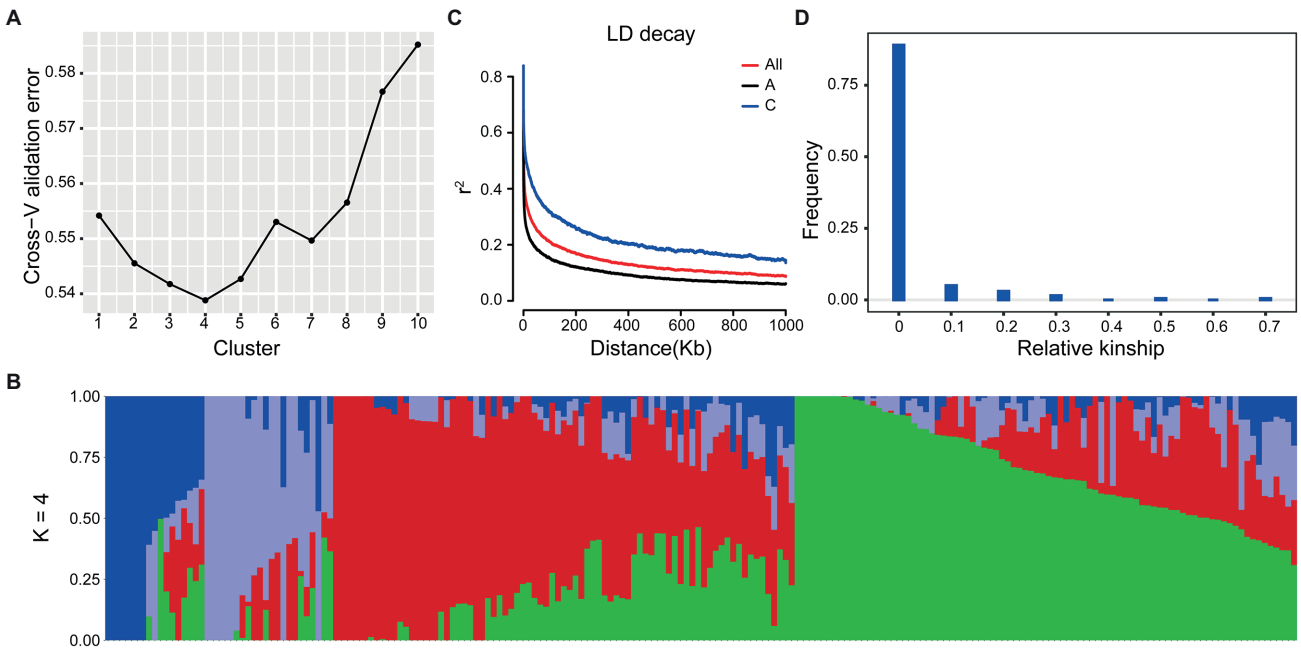


FIGURE 2 | Population structure of *Brassica napus* accessions. **(A)** The value of cross–validation error in different *K* value. **(B)** Model-based group analysis in $K = 4$. The *y* axis quantifies clusters membership, and the *x* axis lists the different *B. napus* accessions. **(C)** Genome-wide and sub-genome average LD decay estimated. **(D)** Relative kinship of 204 *B. napus* accessions.

TABLE 3 | SNP position with significant association for SOC in MLM of EMMAX.

Chromosome	Physical position	*p*-value	Environment
A07	11,440,139	1.21×10^{-6}	2014_2015 CS
A07	11,440,137	1.74×10^{-5}	2014_2015 CS
A07	11,343,133	4.21×10^{-5}	2014_2015 CS
A07	11,411,162	2.68×10^{-7}	2015_2016 CS
A07	11,630,281	3.85×10^{-7}	2015_2016 CS
A07	11,630,282	4.16×10^{-7}	2015_2016 CS
A07	11,630,292	2.82×10^{-7}	2015_2016 CS
A07	11,632,013	2.52×10^{-7}	2015_2016 CS
A07	11,632,726	4.04×10^{-7}	2015_2016 CS
A07	11,774,647	3.67×10^{-7}	2015_2016 CS
A07	11,778,813	3.71×10^{-7}	2015_2016 CS
A07	11,428,788	3.57×10^{-7}	2016_2017 CS
A07	11,428,860	3.07×10^{-7}	2016_2017 CS
A07	11,440,137	4.26×10^{-7}	2016_2017 CS
A07	11,575,763	3.16×10^{-7}	2016_2017 CS
C08	7,894,416	1.35×10^{-7}	2016_2017 CS
A07	11,440,137	6.02×10^{-7}	BLUP
A07	11,440,139	8.21×10^{-7}	BLUP

a total of 12 QTLs were detected using QTL IciMapping, and all these QTLs demonstrated a PVE range of 1.08–3.74% (**Figure 4B**; **Supplementary Table 11**). Among the 12 SOC QTLs, only *qSOC-ME-2*, *qSOC-ME-3*, *qSOC-ME-5*, *qSOC-ME-6*, and *qSOC-ME-8* were obtained from the high-seed oil parental line GRD328 (**Supplementary Table 11**). Considering PVE and the environmental factor, we concluded that *qSOC-ME-3* was the major QTL for the SOC trait, named *qA07.SOC*.

The major QTL *qA07.SOC* linked with the bin markers Block12623 and Block12635 was identified in the single environment of 2015–2016 GY (*qSOC-SE-5*; **Figure 4**; **Table 5**). Then, we divided the two groups (AA genotype and BB genotype) based on the genotypes of the twoflanking bin markers (**Supplementary Figure 6**); for example, the AA genotype represented the genotype of both Block12623 and Block12635, which were consistent with high-seed oil parental line GRD328. In the 158 RIL lines, the bin markers Block12623 and Block12635 divided the AA and BB groups with different significance in the SOC trait (**Supplementary Figure 6**), which suggested that the major and stable QTL-*qA07.SOC* was reliable for the marker-assisted selection of the SOC trait, even if other minor QTLs are present.

Global Transcriptome Analysis in the Developed Seeds

To study the DEGs and metabolic pathways contributing to the difference in SOC of the two parents, three biological replicates of the developed seeds, at 35, 40, 45, and 50 DAF, were collected for RNA sequencing. A total of 186.99 Gb clean data were obtained, and the average Q30 value was 94.84% (**Supplementary Table 12**). Spearman correlation based on the FPKM values in samples with high and low oil contents showed good consistency in biological replicates (**Figure 5A**). In addition, 18 genes differentially expressed between high- and low-oil content samples based on the RNA-seq data were verified by qRT-PCR (**Figures 5B,C**); results suggesting that the transcriptome data were reliable and could be used for subsequent analyses.

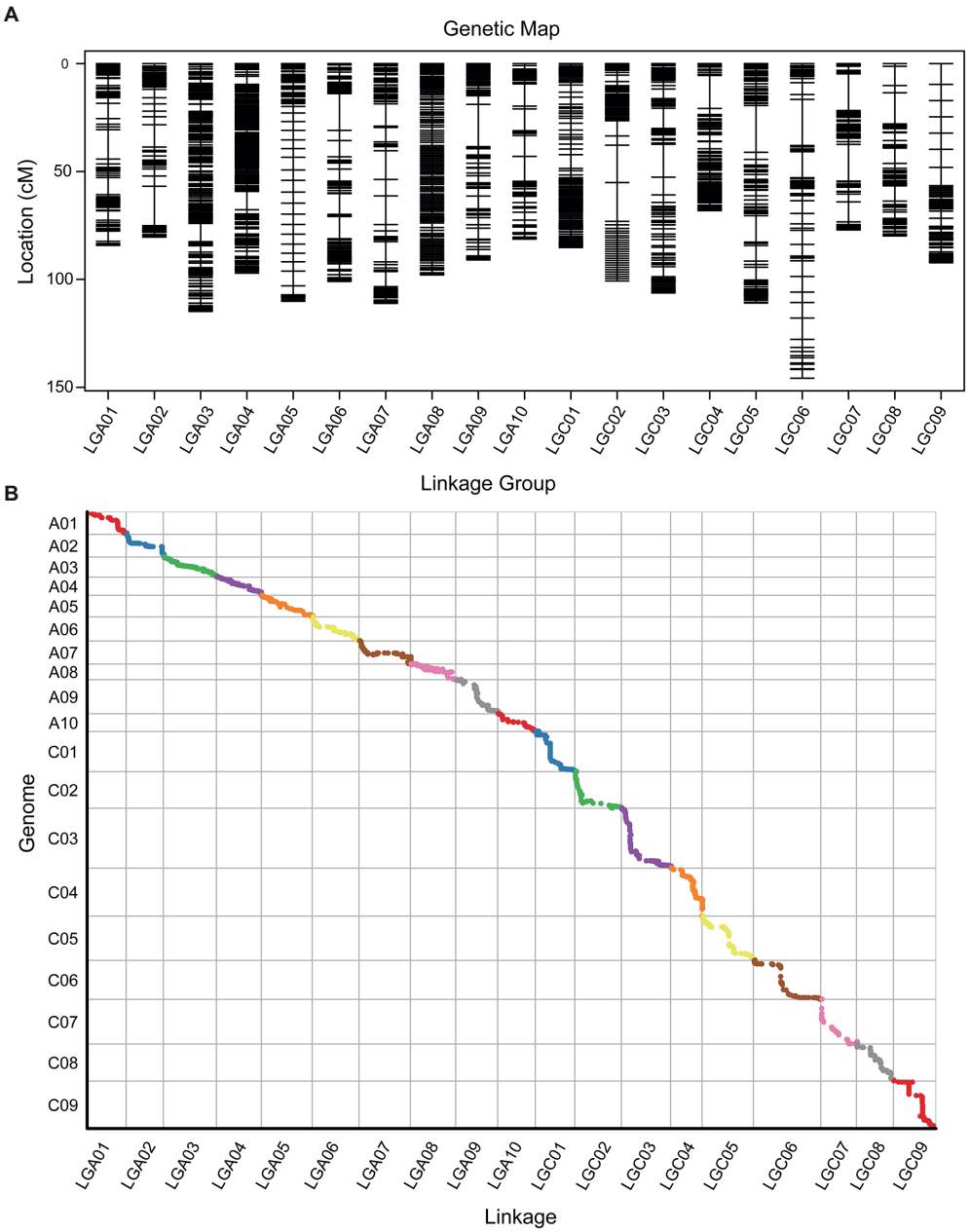


FIGURE 3 | High-density genetic map illustration and collinearity analysis between the genetic map and the genome. **(A)** Distribution of Bin markers on 19 linkage groups. Horizontal black lines on each linkage group represent mapped Bin markers. Vertical lines of each linkage group represent the total genetic distance (cM) of the linkage group. The linkage group ID is shown on the *x*-axis, and the genetic distance is shown on the *y*-axis. **(B)** The *x*-axis is the genetic distance of each linkage group. The *y*-axis is the physical distance of each chromosome, with the collinearity of genomic markers and genetic maps represented in scatter. Different colors represent different chromosomes or linkage groups.

Using the DEseq2 method, 3,086 genes were found to be differentially expressed between the two parents (**Figure 5D**). To discern the functional distribution of the 3,086 DEGs, KEGG pathway analysis was performed, which showed that the most significantly enriched pathways of the DEGs in the developed seeds were involved in lipid biosynthesis, fatty acid metabolism, unsaturated fatty acids biosynthesis, and photosynthesis (**Figure 5E**).

Prediction of Candidate Genes and Mining of Favorable Allelic Variants in the *qA07.SOC* Region

The *qA07.SOC* region was linked with markers Block12623 and Block12635, which covered 180 kb and consisted of 24 genes, of which eight genes (*BnaA07g12660D*, *BnaA07g12670D*, *BnaA07g12740D*, *BnaA07g12750D*, *BnaA07g12760D*, *BnaA07*

TABLE 4 | Characteristics of the high-density genetic map.

Linkage group	Total bin marker	Total distance (cM)	Average distance (cM)	Gaps < 5 cM	Max Gap (cM)
A01	128	84.23	0.66	97.64%	13.61
A02	178	80.41	0.45	98.87%	18.27
A03	240	114.82	0.48	99.58%	8.48
A04	231	97.09	0.42	99.57%	8.49
A05	103	110.12	1.07	95.10%	6.81
A06	140	100.92	0.72	96.40%	17.02
A07	137	111.1	0.81	94.85%	13.28
A08	127	97.91	0.77	100.00%	2.07
A09	173	90.96	0.53	98.26%	19.59
A10	120	81.33	0.68	96.64%	11.57
C01	155	85.23	0.55	99.35%	6.38
C02	125	100.7	0.81	97.58%	17.97
C03	122	106.29	0.87	97.52%	15.23
C04	115	68.15	0.59	99.12%	15.28
C05	169	110.94	0.66	97.62%	21.7
C06	167	145.79	0.87	93.98%	21.34
C07	131	77.07	0.59	97.69%	17.02
C08	139	79.85	0.57	97.10%	14.41
C09	99	92.3	0.93	92.86%	9.62
Total	2,799	1,835.21	0.66	97.35%	

g12840D, *BnaA07g12850D*, and *BnaA07g12870D*) were not expressed in the developed seeds of the two parents (**Figure 6C**; **Supplementary Table 13**). Four genes, *BnaA07g12790D*, *BnaA07g12810D*, *BnaA07g12830D*, and *BnaA07g12880D*, were differentially expressed between the two parents (**Figure 6C**; **Supplementary Table 13**). In DNA levels, the SNP and InDel variations between the two parents in the *qA07.SOC* were analyzed (**Figure 6A**), and most of the variations were identified in the upstream and downstream of the candidate genes, and only one missense mutation in the coding sequence of *BnaA07g12790D* (**Supplementary Tables 14**, **15**).

According to the syntenic analysis, the syntenic blocks of *qA07.SOC* ranged from 22,766,185-22,905,060 bp on chromosome C07 to 15,347,512-15,556,735 bp on chromosome C04 (**Figure 6B**). Based on the QTL mapping results in the RIL population, no QTLs for SOC were identified in the syntenic blocks (**Figure 4A**). After confirming the syntenic block regions, the SNP and InDel variations between the two parents in the syntenic blocks were also analyzed (**Figure 6A**). In the syntenic block regions, there was no syntenic gene for *BnaA07g12790D* and *BnaA07g12880D* (**Figure 6B**). The syntenic gene of *BnaA07g12810D* was *BnaC07g16850D*, while both *BnaC07g16870D* and *BnaC04g17300D* were syntenic genes of *BnaA07g12830D* (**Figure 6B**). As there were no QTLs for SOC in the syntenic blocks, the unique genes *BnaA07g12790D* and *BnaA07g12880D* in *qA07.SOC* could not be excluded. Based on the RNA-seq data, the expression pattern of *BnaA07g12810D* and its syntenic gene *BnaC07g16850D* were coincident in the two parents (**Figure 6C**; **Supplementary Table 16**), as well as no variation in the coding sequence of *BnaA07g12810D*, suggesting that *BnaA07g12810D* may be excluded from the *qA07.SOC* region. Consequently, *BnaA07g12790D* (*BnaA07.STR18*) and *BnaA07g12830D* (*BnaA07.NRT1*), which encode a thiosulfate sulfurtransferase and major facilitator superfamily protein, respectively, as well as the unknown functional gene *BnaA07g12880D* were the candidate genes.

According to the results of linkage and association analyses, the major QTL primarily responsible for SOC was located on chromosome A07 (**Figures 1**, **4**). While the mining of favorable allelic variants in the *qA07.SOC*, haplotype analysis in BLUP showed that the SOC phenotype with a significant difference was divided by haplotype I and haplotype II derived from SNP_11440137 to SNP_11440139, respectively (**Figure 6D**). The favorable allelic variants of SNP_11440137 (A) and SNP_11440139 (A) caused the amino acid to change from valine to isoleucine (**Figure 6D**). Overall, in this study, these stable QTLs containing candidate genes and haplotypes provided a reliable basis for marker-assisted selection in *B. napus* breeding for SOC.

DISCUSSION

Utility of a High-Density Genetic Map in the $F_{2:11}$ RIL Population

Most of the complex agronomic traits in *B. napus* are quantitative. Therefore, it is necessary to establish a genetic linkage map with a high accuracy and saturation for QTL mapping. In the early stages of genetic mapping, the stability and reliability of RFLP and SSR markers are relatively high and have been used to study the genetic linkage map of *B. napus* (Landry et al., 1991; Uzunova et al., 1995; Wang et al., 2013; Qi et al., 2014). With the announcement of the *B. napus* reference genome and the high-density of SNP in the genome, the SNP Chip Array became popular in plant research. A *B. napus* linkage map containing 2,771 bins with an average distance of 1.47 cM between bins was constructed for *Sclerotinia* resistance and flowering time QTL mapping based on a 60 K SNP Chip Array (Zhang et al., 2019). In recent years, simplified genome sequencing techniques, such as restriction site-associated DNA sequencing (RAD-seq), have been used to construct linkage maps with a total length of 1,610.4 cM in *B. napus* (Chen et al., 2017).

Previous studies have reported that the number of recombinants with close linkage markers in the RIL population is twice that in the F_2 population and that the genetic distance of linkage genes could be accurately estimated (Burr et al., 1988). In our study, after the introduction of *S. alba* L. inbred line, an $F_{2:11}$ RIL population was constructed. Then, we used the whole-genome resequencing method to genotype the RIL population and got a high-density genetic map that harbored 2,799 bin markers and covered a total distance of 1,835.21 cM (**Figure 3**; **Table 4**), which provided a super foundation for QTL mapping.

The Co-localization of the Major Locus for SOC by QTL Mapping and GWAS

There are many factors that affect the results of QTL mapping, such as the phenotype heredity, the type and size of mapping

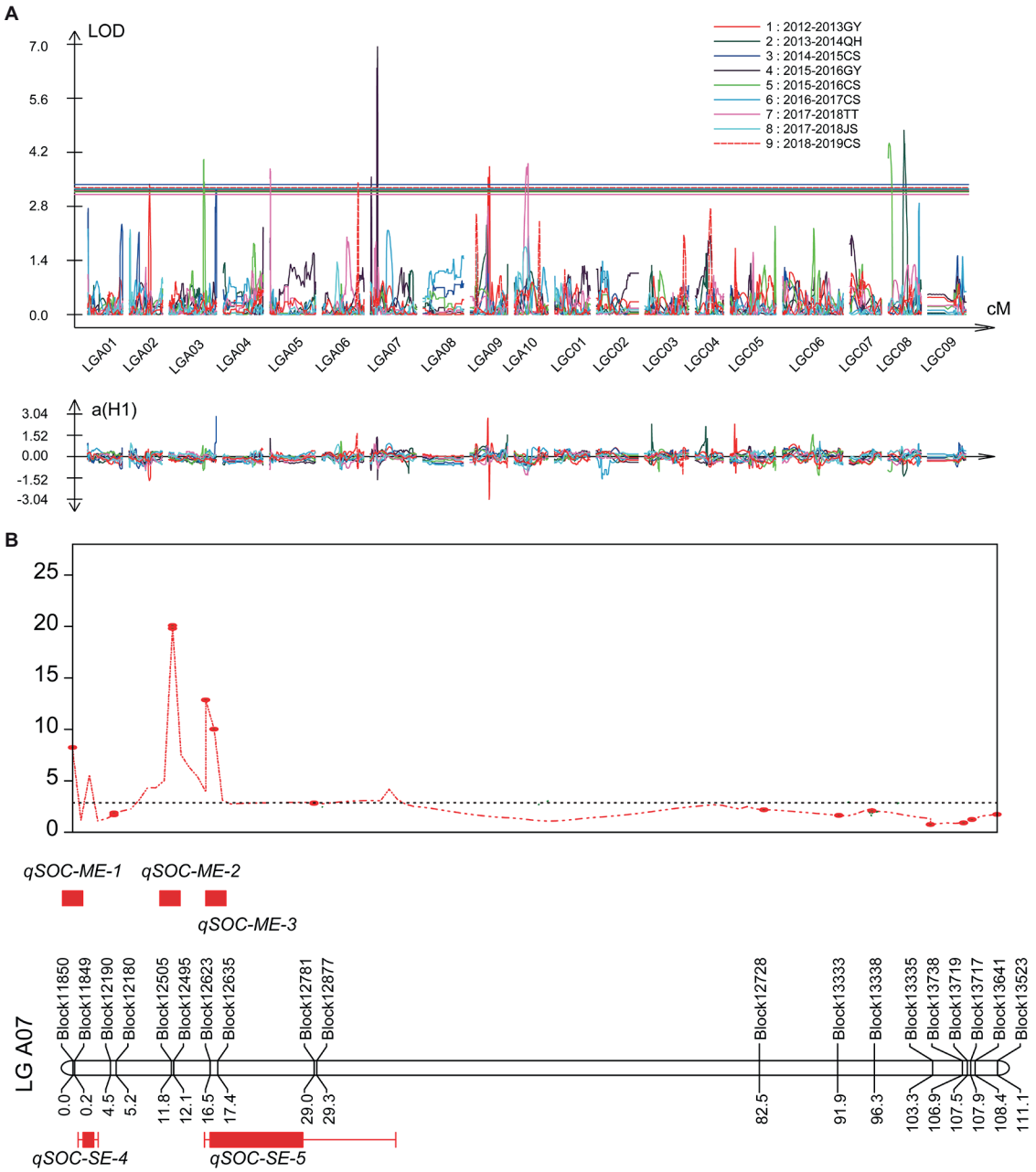


FIGURE 4 | Quantitative trait loci (QTLs) identification for SOC in single and multi-environments. **(A)** The various colors represented the QTL identified in different single environments. **(B)** Co-localization analysis of the major QTLs identified in single and multi-environments on linkage group A07. ME represented multi-environments; SE represented single environment.

population, and the type and density of mapping markers. QTL analysis is carried out on the basis of phenotypic analysis, and different traits have different heritability. Traits with lower heritability are greatly influenced by environment, and changes in environment will lead to changes in the corresponding relationship between genotype and phenotype, and then generate different QTL mapping results. In our study, the mapping population and markers were high quality, while the environment factor mainly effected the result of QTL mapping. In the nine environments, the correlations of the phenotypes were lower than 45% (**Supplementary Figure 2B**); therefore, the major QTL identified was different among each environment due to the environment factor. However, combined the single and multiple environments, we identified the major QTL was *qA07.SOC*. Similarly, the correlations of the association population were about 55–75% (**Supplementary Figure 2A**), suggesting that the environment factor also affected the GWAS results. Therefore, we compared the GWAS results between

TABLE 5 | Major QTLs for SOC identified in single environment.

QTL	Chr	Position (cM)	LOD	Interval (cM)	Add	PVE (%)	Environment
qSOC-SE-1	A02	48.8	3.4	46.8–51.9	−1.7007	8.49	2012–2013 GY
qSOC-SE-2	A09	46.2	3.8	46.1–47.6	−3.0434	9.72	2012–2013 GY
qSOC-SE-3	C08	38.2	4.7	37.4–44	−1.3912	11.23	2013–2014 QH
qSOC-SE-4	A07	2.1	3.6	0.6–3	−0.9701	15.07	2015–2016 GY
qSOC-SE-5	A07	16.5	6.9	15.8–38.8	1.404	11.72	2015–2016 GY
qSOC-SE-6	A03	83.8	4	82.4–86.7	−0.9741	11.07	2015–2016 CS
qSOC-SE-7	C08	2.3	4.4	0–9.5	−1.0376	13.79	2015–2016 CS
qSOC-SE-8	A05	0.3	3.7	0–1.3	−0.9162	8.11	2017–2018 TT
qSOC-SE-9	A06	86.5	3.4	85.3–86.9	−0.8491	7.97	2018–2019 CS

SOC, seed oil content; Chr, Chromosome; LOD, logarithm of the odds; PVE, phenotypic variation explained; and Add, additive effect. QTL was named by the consensus QTL nomenclature, and SE represented single environment.

the EMMAX and BLINK, suggesting the same significant SNPs, A07_11440137 and A07_11440139 were repeatedly identified in different environments by two software (**Table 3**; **Supplementary Table 5**). Finally, although the environment factor affected the phenotypes of association and RIL populations, the major QTL for SOC was reliable based on the multiple environments and methods.

Comparison Between Candidate Region With Syntenic Region to Predict Candidate Genes in Allotetraploid *Brassica napus*

Comparison between the syntenic block regions in the different genomes of closely related species can help to predict candidate genes The major QTL-*qSS.C9* for seeds number per silique in *B. napus* was fined mapped in a 50 kb region on chromosome C09. By comparison of the syntenic blocks in chromosome A10 of *B. rapa*, C09 of *B. oleracea* with the candidate region of *B. napus*, the deletion of *BnaC09g45890D* encoding the telomerase-activating protein resulted in a significant reduction in the number of seeds per silique (Li et al., 2015). For allotetraploid *B. napus*, the regions of the syntenic blocks between different sub-genomes can also be compared. In allotetraploid, syntenic blocks with syntenic QTLs may indicate the presence of syntenic genes with conserved functions, although their contributions to the corresponding phenotype may be different (Zhang et al., 2019). Regarding the absence of syntenic QTLs in the syntenic blocks, the unique gene in the major QTL region and the genes with expression patterns different from those of syntenic genes, as well as the genes with unique variations in the major QTL, were retained as candidate genes for further analysis. In our study, according to the expression levels in developed seeds, four genes, including *BnaA07g12790D* (*BnaA07.STR18*), *BnaA07g12810D*, *BnaA07g12830D*, and *BnaA07g12880D* were selected (**Figure 6C**). In addition, based on the syntenic properties of the genomic regions, *BnaA07g12810D* in the *qA07.SOC* region could be excluded. According to haplotype analysis, the favorable allelic variants located in *BnaA07.STR18* resulted in significant differences in phenotypic variation (**Figure 6D**). In addition, in *Arabidopsis thaliana*, *AT5G66170* is the orthologous gene of *BnaA07.STR18*, and the Gene Ontology biological process of *AT5G66170* is involved in homeostatic processes, lipid metabolic processes, monocarboxylic acid metabolic processes, pigment biosynthetic processes, and responses to inorganic substance.[4] Taken together, *BnaA07.STR18* may be a candidate gene conferring high SOC to *qA07.SOC*. These methods and principles will help elucidate polyploid genomics in future studies and provide useful information for *B. napus* breeding programs focusing in SOC.

qSOC.A07 Is Available for High-Yield and High-Oil Pyramid Breeding in *Brassica napus*

With an increase in the global human population and growing demand for oil, high yield and high oil content have become important goals in oilseed rape breeding programs (Sun et al., 2018). Thousand-seed weight (TSW) is the major factor affecting rapeseed yield. TSW and SOC are quantitative traits that are controlled by multiple genes and influenced by the environment. Therefore, aggregation of multiple QTLs for TSW and SOC is an important process and method to achieve high yield and high oil content in *B. napus*. Many QTLs for the SOC trait have been identified in previous studies, and a total of 59 QTLs can be mapped onto the *B. napus* reference genome (Zhao et al., 2005; Delourme et al., 2006; Qiu et al., 2006; Sun et al., 2012; Wang et al., 2013; Liu et al., 2016; Zou et al., 2016). While integrating these mapped QTLs to the *B. napus* reference, we found that there were fewer QTLs for SOC on chromosome A07 (**Supplementary Figure 7**; Fan et al., 2010; Shi et al., 2011; Wang et al., 2016; Zhao et al., 2016). In this study, the major QTL for SOC was located at 180 kb (11,378,925–11,559,658 bp; **Figure 4**; **Supplementary Table 11**). The major QTL *qSOC.A07* was found to be novel based on the integration of mapped QTLs (**Supplementary Figure 7**). In previous studies, QTLs for TSW identified on chromosome A07 overlapped with the region of *qSOC.A07* (**Supplementary Figure 7**),

[4]https://www.arabidopsis.org/

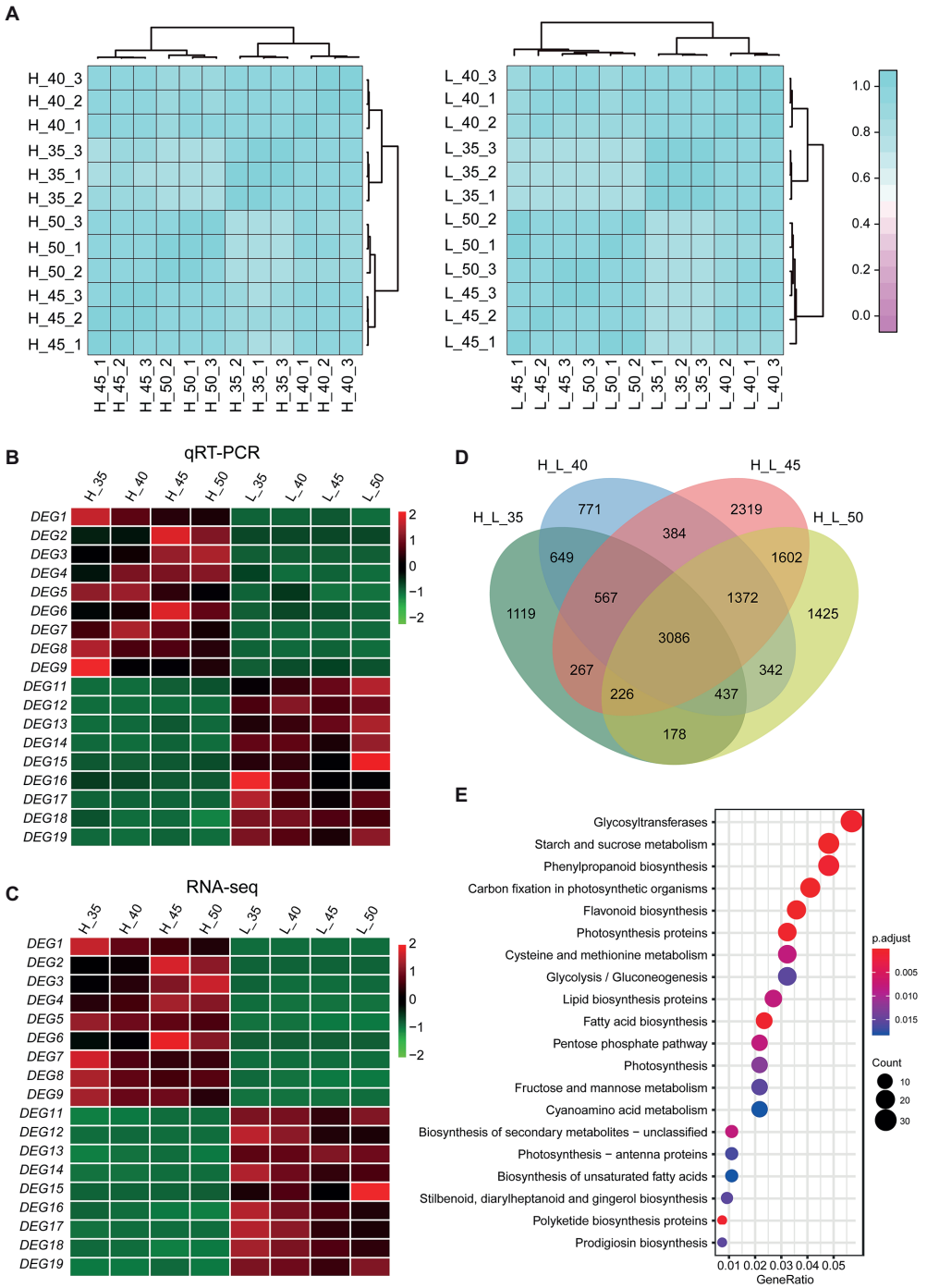


FIGURE 5 | RNA-seq analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of differentially expressed genes (DEGs). **(A)** Spearman correlation plot in developed seeds of GRD328 (high seed oil content) and GRG2462 (low seed oil content). **(B,C)** Heatmap represented that the expression levels of the DEGs based on the RNA-seq was verified by quantitative real-time PCR (qRT-PCR). DEGs 1–9 were upregulated in GRD328, and DEGs 11–19 were downregulated in GRD328. The bars represented the normalized transformed counts of qRT-PCR values **(B)** and FPKM **(C)**, respectively. **(D)** The statistical analysis of DEGs in developed seeds between two parents. **(E)** KEGG enrichment analysis of DEGs.

suggesting that this genomic region covering the loci for TSW and SOC can be used for pyramid breeding. The parent line GRG2462, with high TSW and low SOC, was derived from *B. napus* and *S. alba* by distant hybridization. The major QTL for SOC was identified based on the 158 RILs; using this information, and we hope to breed a new rapeseed germplasm with high yield and high oil content among the RILs in the future.

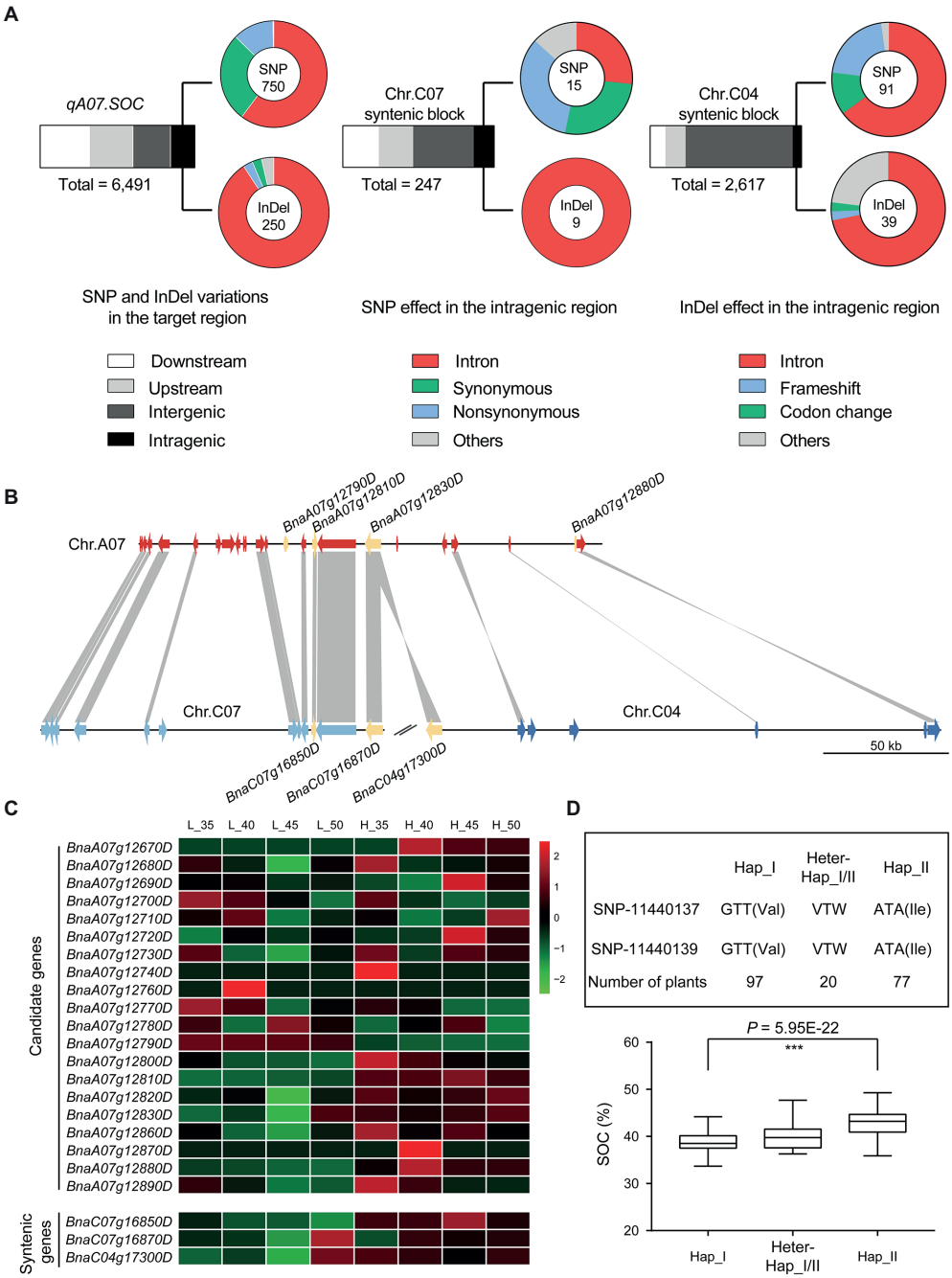


FIGURE 6 | The candidate genes prediction by comparative analysis. **(A)** The variations were identified in *qA07.SOC* and the syntenic regions. **(B)** The syntenic analysis for *qA07.SOC*. **(C)** Heatmap displays the expression levels of the candidate genes and the syntenic genes. The bars represented FPKM normalized transformed counts. **(D)** Haplotype analysis based on GWAS in BLUP.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The raw data generated in this study have been deposited into the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb; https://db.cngb.org/) with accession numbers CNP0002896 and CNP0002897.

AUTHOR CONTRIBUTIONS

YX, SL, CZ, and MX designed the research. CZ, MX, LL, LY, HH, XQ, JZ, YH, WD, CD, and YX performed the experiments. CZ, MX, LL, LY, and YX analyzed the data. XH provided the materials. CZ, MX, and YX wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by National Natural Science Foundation of China (32070217), Subsidy project from NSFC of Guizhou Academy of Agricultural Sciences [No. (2021)50], and the Agricultural Scientific and Technological Research Projects of Guizhou province [Nos. (2020)1Y106 and 1Y109].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.929197/full#supplementary-material

Supplementary Figure 1 | The frequency distributions of the SOC trait in the RIL population.

Supplementary Figure 2 | The correlations between environments for the association and RIL population. **(A)** The correlations for the association population. **(B)** The correlations for the RIL population.

Supplementary Figure 3 | The density of SNP covered in 19 *Brassica napus* chromosome for GWAS.

Supplementary Figure 4 | GWAS for the SOC trait in consecutive 3 years based on BLINK.

Supplementary Figure 5 | Genotype map based on the Bin markers of the total 158 RILs.

Supplementary Figure 6 | The effect analysis of the major QTL for SOC in 2015–2016 GY. AA and BB was the genotype of each flanking markers linked the major QTL in the two parents. The '*n*' represented the number lines in the RIL population conferring the corresponding genotype. *** represented significance level of the Student's *t*-test, $p < 0.001$.

Supplementary Figure 7 | QTLs for TSW and SOC identification in previous studies.

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Basunanda, P., Radoev, M., Ecke, W., Friedt, W., Becker, H. C., and Snowdon, R. J. (2010). Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 120, 271–281. doi: 10.1007/s00122-009-1133-z

Bates, P. D., Johnson, S. R., Cao, X., Li, J., Nam, J. W., Jaworski, J. G., et al. (2014). Fatty acid synthesis is inhibited by inefficient utilization of unusual fatty acids for glycerolipid assembly. *Proc. Natl. Acad. Sci. U. S. A.* 111, 1204–1209. doi: 10.1073/pnas.1318511111

Burr, B., Burr, F. A., Thompson, K. H., Albertson, M. C., and Stuber, C. W. (1988). Gene mapping with recombinant inbreds in maize. *Genetics* 118, 519–526. doi: 10.1093/genetics/118.3.519

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chao, H., Wang, H., Wang, X., Guo, L., Gu, J., Zhao, W., et al. (2017). Genetic dissection of seed oil and protein content and identification of networks associated with oil content in *Brassica napus*. *Sci. Rep.* 7:46295. doi: 10.1038/srep46295

Chen, J., Wang, B., Zhang, Y., Yue, X., and Liu, K. (2017). High-density ddRAD linkage and yield-related QTL mapping delimits a chromosomal region responsible for oil content in rapeseed (*Brassica napus* L.). *Breed* 67, 296–306. doi: 10.1270/jsbbs.16116

Delourme, R., Falentin, C., Huteau, V., Clouet, V., Horvais, R., Gandon, B., et al. (2006). Genetic control of oil content in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 113, 1331–1345. doi: 10.1007/s00122-006-0386-z

Dong, S.-S., He, W.-M., Ji, J.-J., Zhang, C., Guo, Y., and Yang, T.-L. (2020). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* 22:bbaa227. doi: 10.1093/bib/bbaa227

Elahi, N., Duncan, R. W., and Stasolla, C. (2015). Decreased seed oil production in *FUSCA3 Brassica napus* mutant plants. *Plant Physiol. Biochem.* 96, 222–230. doi: 10.1016/j.plaphy.2015.08.002

Elahi, N., Duncan, R. W., and Stasolla, C. (2016). Modification of oil and glucosinolate content in canola seeds with altered expression of *Brassica napus* LEAFY COTYLEDON1. *Plant Physiol. Biochem.* 100, 52–63. doi: 10.1016/j.plaphy.2015.12.022

Fan, C., Cai, G., Qin, J., Li, Q., Yang, M., Wu, J., et al. (2010). Mapping of quantitative trait loci and development of allele-specific markers for seed weight in *Brassica napus*. *Theor. Appl. Genet.* 121, 1289–1301. doi: 10.1007/s00122-010-1388-4

Freed, D., Aldana, R., Weber, J.A., and Edwards, J.S. (2017). The sentieon genomics tools—A fast and accurate solution to variant calling from next-generation sequence data. biorxiv [Preprint]. doi: 10.1101/115717

Frentzen, M. (2010). Acyltransferases from basic science to modified seed oils. *Eur. J. Lipid Sci. Technol.* 100, 161–166. doi: 10.1002/(SICI)1521-4133(19985)100%3A4/5<161%3A%3AAID-LIPI161>3.0.CO;2-P

Guo, Y., Si, P., Wang, N., Wen, J., Yi, B., and Ma, C. (2017). Genetic effects and genotype×environment interactions govern seed oil content in *Brassica napus* L. *BMC Genet.* 18:1. doi: 10.1186/s12863-016-0468-0

Hills, M. J. (2004). Control of storage-product synthesis in seeds. *Curr. Opin. Plant Biol.* 7, 302–308. doi: 10.1016/j.pbi.2004.03.003

Hua, W., Li, R. J., Zhan, G. M., Liu, J., Li, J., Wang, X. F., et al. (2012). Maternal control of seed oil content in *Brassica napus*: the role of silique wall photosynthesis. *Plant J.* 69, 432–444. doi: 10.1111/j.1365-313X.2011.04802.x

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res.* 19, 1068–1076. doi: 10.1101/gr.089516.108

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8:giy154. doi: 10.1093/gigascience/giy154

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., and Hattori, T. (2005). LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 46, 399–406. doi: 10.1093/pcp/pci048

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kosambi, D. D. (2016). "The estimation of map distances from recombination values," in *D.D. Kosambi: Selected Works in Mathematics and Statistics*. ed. R. Ramaswamy (New Delhi: Springer India), 125–130.

Landry, B. S., Hubert, N., Etoh, T., Harada, J. J., and Lincoln, S. E. (1991). A genetic map for *Brassica napus* based on restriction fragment length. *Genome* 34, 543–552. doi: 10.1139/g91-084

Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M. G., et al. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29:2073. doi: 10.1093/bioinformatics/btt337

Li, S., Chen, L., Zhang, L., Li, X., Liu, Y., Wu, Z., et al. (2015). *BnaC9.SMG7b* functions as a positive regulator of the number of seeds per Silique in *Brassica napus* by regulating the formation of functional female gametophytes. *Plant Physiol.* 169, 2744–2760. doi: 10.1104/pp.15.01040

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, H., Ribaut, J. M., Li, Z., and Wang, J. (2008). Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor. Appl. Genet.* 116, 243–260. doi: 10.1007/s00122-007-0663-5

Liu, S., Fan, C., Li, J., Cai, G., Yang, Q., Wu, J., et al. (2016). A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*. *Theor. Appl. Genet.* 129, 1203–1215. doi: 10.1007/s00122-016-2697-z

Liu, J., Hua, W., Hu, Z., Yang, H., Zhang, L., Li, R., et al. (2015). Natural variation in ARF18 gene simultaneously affects seed weight and silique length in polyploid rapeseed. *Proc. Natl. Acad. Sci.* 112, E5123–E5132. doi: 10.1073/pnas.1502160112

Liu, J., Hua, W., Yang, H. L., Zhan, G. M., Li, R. J., Deng, L. B., et al. (2012). The *BnGRF2* gene (GRF2-like gene from *Brassica napus*) enhances seed oil production through regulating cell number and plant photosynthesis. *J. Exp. Bot.* 63, 3727–3740. doi: 10.1093/jxb/ers066

Liu, D., Ma, C., Hong, W., Huang, L., Liu, M., Liu, H., et al. (2014). Construction and analysis of high-density linkage map using high-throughput sequencing data. *PLoS One* 9:e98855. doi: 10.1371/journal.pone.0098855

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lock, Y. Y., Snyder, C. L., Zhu, W., Siloto, R. M., Weselake, R. J., and Shah, S. (2009). Antisense suppression of type 1 diacylglycerol acyltransferase adversely affects plant development in *Brassica napus*. *Physiol. Plant.* 137, 61–71. doi: 10.1111/j.1399-3054.2009.01258.x

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., et al. (2019). Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* 10:1154. doi: 10.1038/s41467-019-09134-9

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Qi, L., Mao, L., Sun, C., Pu, Y., Fu, T., Ma, C., et al. (2014). Interpreting the genetic basis of silique traits in *Brassica napus* using a joint QTL network. *Plant Breed.* 133, 52–60. doi: 10.1111/pbr.12131

Qiu, D., Morgan, C., Shi, J., Long, Y., Liu, J., Li, R., et al. (2006). A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. *Theor. Appl. Genet.* 114, 67–80. doi: 10.1007/s00122-006-0411-2

Ren, H., Han, J., Wang, X., Zhang, B., Yu, L., Gao, H., et al. (2020). QTL mapping of drought tolerance traits in soybean with SLAF sequencing. *Crop J.* 8, 977–989. doi: 10.1016/j.cj.2020.04.004

Shi, J., Lang, C., Wang, F., Wu, X., Liu, R., Zheng, T., et al. (2017). Depressed expression of *FAE1* and *FAD2* genes modifies fatty acid profiles and storage compounds accumulation in *Brassica napus* seeds. *Plant Sci.* 263, 177–182. doi: 10.1016/j.plantsci.2017.07.014

Shi, J., Li, R., Zou, J., Long, Y., and Meng, J. (2011). A dynamic and complex network regulates the heterosis of yield-correlated traits in rapeseed (*Brassica napus* L.). *PLoS One* 6:e21645. doi: 10.1371/journal.pone.0021645

Shi, L., Song, J., Guo, C., Wang, B., Guan, Z., Yang, P., et al. (2019). A CACTA-like transposable element in the upstream region of *BnaA9.CYP78A9* acts as an enhancer to increase silique length and seed weight in rapeseed. *Plant J.* 98, 524–539. doi: 10.1111/tpj.14236

Silva, L. D. C. E., Wang, S., and Zeng, Z. B. (2012). "Composite interval mapping and multiple interval mapping: procedures and guidelines for using windows QTL cartographer," in *Quantitative Trait Loci (QTL): Methods and Protocols*. ed. S. A. Rifkin (Totowa, NJ: Humana Press), 75–119.

Sun, M., Hua, W., Liu, J., Huang, S., Wang, X., Liu, G., et al. (2012). Design of new genome- and gene-sourced primers and identification of QTL for seed oil content in a specially high-oil *Brassica napus* cultivar. *PLoS One* 7:e47037. doi: 10.1371/journal.pone.0047037

Sun, L., Wang, X., Yu, K., Li, W., and Zhang, J. (2018). Mapping of QTLs controlling seed weight and seed-shape traits in *Brassica napus* L. using a high-density SNP map. *J. Agric. Sci. Technol.* 214, 143–151. doi: 10.1007/s10681-018-2303-3

Tang, M., Zhang, Y., Liu, Y., Tong, C., Cheng, X., Zhu, W., et al. (2019). Mapping loci controlling fatty acid profiles, oil and protein content by genome-wide association study in *Brassica napus*. *Crop J.* 7, 217–226. doi: 10.1016/j.cj.2018.10.007

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Truco, M. J., Ashrafi, H., Kozik, A., Van Leeuwen, H., Bowers, J., Wo, S. R., et al. (2013). An ultra-high-density, transcript-based. *Genet. Map Lett.* 3, 617–631. doi: 10.1534/g3.112.004929

Turnham, E., and Northcote, D. H. (1983). Changes in the activity of acetyl-CoA carboxylase during rape-seed formation. *Biochem. J.* 212, 223–229. doi: 10.1042/bj2120223

Uzunova, M., Ecke, W., Weissleder, K., and Röbbelen, G. (1995). Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. *Theor. Appl. Genet.* 90, 194–204. doi: 10.1007/BF00222202

Wang, X., Chen, L., Wang, A., Wang, H., Tian, J., Zhao, X., et al. (2016). Quantitative trait loci analysis and genome-wide comparison for silique related traits in *Brassica napus*. *BMC Plant Biol.* 16:71.

Wang, H., Guo, J., Lambert, K. N., and Lin, Y. (2007). Developmental control of *Arabidopsis* seed oil biosynthesis. *Planta* 226, 773–783. doi: 10.1007/s00425-007-0524-0

Wang, X., Liu, G., Yang, Q., Hua, W., Liu, J., and Wang, H. (2010). Genetic analysis on oil content in rapeseed (*Brassica napus* L.). *Euphytica* 173, 17–24. doi: 10.1007/s10681-009-0062-x

Wang, Z., Qiao, Y., Zhang, J., Shi, W., and Zhang, J. (2017). Genome wide identification of microRNAs involved in fatty acid and lipid metabolism of *Brassica napus* by small RNA and degradome sequencing. *Gene* 619, 61–70. doi: 10.1016/j.gene.2017.03.040

Wang, X., Wang, H., Long, Y., Li, D., Yin, Y., Tian, J., et al. (2013). Identification of QTLs associated with oil content in a high-oil *Brassica napus* cultivar and construction of a high-density consensus map for QTLs comparison in *B. napus*. *PLoS One* 8:e80569. doi: 10.1371/journal.pone.0080569

Wei, H., Rong-Jun, L., Gao-Miao, Z., Jing, L., Jun, L., Wang, X. F., et al. (2011). Maternal control of seed oil content in *Brassica napus*: the role of silique wall photosynthesis. *Plant J.* 69, 432–444. doi: 10.1111/j.1365-313X.2011.04802.x

Weselake, R. J., Shah, S., Tang, M., Quant, P. A., Snyder, C. L., Furukawa-Stoffer, T. L., et al. (2008). Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. *J. Exp. Bot.* 59, 3543–3549. doi: 10.1093/jxb/ern206

Wu, X. L., Liu, Z. H., Hu, Z. H., and Huang, R. Z. (2014). *BnWRI1* coordinates fatty acid biosynthesis and photosynthesis pathways during oil accumulation in rapeseed. *J. Integr. Plant Biol.* 56, 582–593. doi: 10.1111/jipb.12158

Xie, Y., Xu, J., Tian, G., Xie, L., Xu, B., Liu, K., et al. (2020). Unraveling yield-related traits with QTL analysis and dissection of QTL×environment interaction using a high-density bin map in rapeseed (*Brassica napus*. L). *Euphytica* 216:171. doi: 10.1007/s10681-020-02708-5

Yang, P., Shu, C., Chen, L., Xu, J., and Liu, K. (2012). Identification of a major QTL for silique length and seed weight in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 125, 285–296. doi: 10.1007/s00122-012-1833-7

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2018). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875

Zhang, F., Huang, J., Tang, M., Cheng, X., Liu, Y., Tong, C., et al. (2019). Syntenic quantitative trait loci and genomic divergence for *Sclerotinia* resistance and flowering time in *Brassica napus*. *J. Integr. Plant Biol.* 61, 75–88. doi: 10.1111/jipb.12754

Zhao, J., Becker, H. C., Zhang, D., Zhang, Y., and Ecke, W. (2005). Oil content in a European × Chinese rapeseed population. *Crop Sci.* 45, 51–59. doi: 10.2135/cropsci2005.0051a

Zhao, W., Wang, X., Wang, H., Tian, J., Li, B., Chen, L., et al. (2016). Genome-wide identification of QTL for seed yield and yield-related traits and

construction of a high-density consensus map for QTL comparison in *Brassica napus*. *Front. Plant Sci.* 7:17. doi: 10.3389/fpls.2016.00017

Zhou, X., Dai, L., Wang, P., Liu, Y., Xie, Z., Zhang, H., et al. (2021). Mining favorable alleles for five agronomic traits from the elite rapeseed cultivar zhongshuang 11 by QTL mapping and integration. *Crop J.* 9, 1449–1459. doi: 10.1016/j.cj.2020.12.008

Zou, J., Zhao, Y., Liu, P., Shi, L., Wang, X., Wang, M., et al. (2016). Seed quality traits can be predicted with high accuracy in *Brassica napus* using genomic data. *PLoS One* 11:e0166624. doi: 10.1371/journal.pone.0166624

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

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 December 2022
DOI 10.3389/fpls.2022.1067121

OPEN ACCESS

EDITED BY
Hongbo Chao,
Zhengzhou University, China

REVIEWED BY
Jun Li,
Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, China
Weiguo Zhao,
Ankang University, China

*CORRESPONDENCE
Jiana Li
ljn1950@swu.edu.cn

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 11 October 2022
ACCEPTED 24 November 2022
PUBLISHED 08 December 2022

CITATION
Xiao Z, Zhang C, Qu C, Wei L, Zhang L, Yang B, Lu K and Li J (2022) Identification of candidate genes regulating seed oil content by QTL mapping and transcriptome sequencing in *Brassica napus*. *Front. Plant Sci.* 13:1067121. doi: 10.3389/fpls.2022.1067121

Identification of candidate genes regulating seed oil content by QTL mapping and transcriptome sequencing in *Brassica napus*

Zhongchun Xiao[1], Chao Zhang[2], Cunmin Qu[3], Lijuan Wei[3], Liyuan Zhang[3], Bo Yang[3], Kun Lu[3] and Jiana Li[3]*

[1]Key Laboratory of Biological Genetic Resources Mining and Molecular Breeding of Qianxinan Prefecture, College of Biology and Chemistry, Minzu Normal University of Xingyi, Xingyi, China, [2]Guizhou Oil Research Institute, Guizhou Academy of Agricultural Sciences, Guiyang, China, [3]Academy of Agricultural Sciences, Southwest University, Chongqing, China

Increasing oil production is a major goal in rapeseed (*Brassica napus*) molecular breeding programs. Identifying seed oil content (SOC)-related candidate genes is an important step towards achieving this goal. We performed quantitative trait locus (QTL) mapping of SOC in *B. napus* using a high-density SNP genetic map constructed from recombinant inbred lines and the Illumina Infinium^{TM} 60K SNP array. A total of 26 QTLs were detected in three years on A01, A03, A05, A06, A09, C01, C03 and C05, which accounted for 3.69%~18.47% of the phenotypic variation in SOC. Of these, 13 QTLs are reported here for the first time. 1713 candidate genes in the 26 QTLs confidence interval were obtained. We then identified differentially expressed genes (DEGs) between the high- and low-SOC accessions, to narrow down our focus to 21 candidate genes (Y1-Y21) related to SOC, and we will focus on 11 (Y1-Y11) candidate genes that contribute to the formation of high-SOC. In addition to providing insight into the genetic basis of SOC in *B. napus*, the loci identified and candidate genes in this study can be used in molecular breeding strategies to increase SOC in this important seed crop.

KEYWORDS

***Brassica napus*, seed oil content, QTL mapping, differentially expressed genes (DEGs), candidate genes**

Introduction

Due to its low saturated fatty acid content, rapeseed (*Brassica napus*) oil is considered a healthy edible vegetable oil. Increasing oil production, by increasing seed yield and seed oil content (SOC), is a major goal of *B. napus* breeding. Tremendous progress has been made towards increasing *B. napus* oil production, especially in China, due to the

development of next-generation sequencing technology and SNP markers. *B. napus* breeders in China have produced germplasm with a SOC of 55–60%, with the potential to increase it to 75% (Hu et al., 2013; Hua et al., 2016). Therefore, there is a broad prospect for improving the SOC of *B. napus* in China.

The SOC of *B. napus* is a complex quantitative trait, controlled by multiple genes and environmental effects (Si et al., 2003; Liu et al., 2016). Previous studies have shown that the SOC of *B. napus* is mainly determined by the maternal genotype, and the photosynthesis of maternal silique pericarps plays an important role in regulating SOC (Hua et al., 2012; Wang et al., 2010). As for the QTL mapping of SOC in *B. napus*, the previous studies have mapped many related QTLs. However, due to differences in the subsets of accessions used, genetic maps, localization methods, and environmental differences in the populations used for QTL mapping, the QTLs associated with SOC in *B. napus* obtained by predecessors are also very different. Out of 14 and 10 SOC-related QTLs discovered in the DY and RNSL DH populations, respectively, only one was detected simultaneously in both populations (Delourme et al., 2006). Yan et al. (2009) detected 11 QTLs related to SOC in three different environments using the *B. napus* recombinant inbred line (RIL) population. These QTLs were mainly distributed on C05 and C06 chromosome, with a single QTL accounting for 5.19-13.57% of the phenotypic variation. Chen et al. (2010) detected 27 SOC-related QTLs distributed in 14 linkage groups under nine different environments, using a *B. napus* DH population, of which a single QTL explained 4.2-30.2% of the phenotypic variation (Chen et al., 2010). Zhao et al. (2012) reconstructed a map using the SG-DH population and detected nine SOC-related QTLs, located on chromosomes A01, A05, A07, A09, C02, C03, C06 and C08, under 11 different growth regimes, which together accounted for 57.79% of the phenotypic variation (Zhao et al., 2012). Wang et al. (2013) constructed a *B. napus* KNDH population using "KenC-8" and "N53-2" and identified 24 SOC-related QTLs (Wang et al., 2013). Sun et al. (2016) constructed two F_2 populations and obtained 40 SOC-related QTLs using linkage mapping under different environments (Sun et al., 2016).

In this study, we aimed to identify candidate genes associated with SOC by combining QTL mapping and transcriptome sequencing between extremely high- and low-SOC *B. napus* accessions. We created a high-generation RIL containing 186 strains, and constructed a high-density genetic linkage map, consisting of 8,575 SNP markers and 1,201 clusters across the 6140.2cM *B. napus* genome, using the 60K SNP chip. Finally, a total of 26 SOC-related QTLs were detected. Meanwhile, based on differential expression genes (DEGs) between the high- and low-SOC accessions from our previous study (Xiao et al., 2019), we identified 21 candidate genes, which could be useful for further gene cloning and molecular breeding for higher SOC in *B. napus*.

Materials and methods

Plant materials and phenotyping

The high-seed oil content (SOC; 44.57%) *B. napus* accession, GH06, was used as the female parent and a low-SOC (36.69%) accession, ZY821, as the male parent. Starting from the F_2 generation, plants were self-fertilized using the one grain transfer method. In the 10th generation, a high-generation recombinant inbred line (RIL) population containing 186 strains was constructed and cultivated in the rapeseed planting base of Xiema Town, Beibei District, Chongqing (29°45′39.99" N, 106°22′38.47"E, 238.57 m), for three consecutive years (2016-2018). All field experiments followed a randomized complete block design with two replicates. From each accession, 30 plants were grown in three rows per plot, with 10 plants per row, with a row spacing of 40 cm, and a plant spacing of 20 cm. The trial was managed in a conventional manner to ensure a consistent growth environment for all accessions. At maturity, five representative plants were collected from the middle of each plot. The oil content of the desiccated seeds was measured by near-infrared reflectance spectroscopy (NIRS DS2500).

Statistical analysis

The SOC data from three consecutive years were compared using a Student's *t*-test in Microsoft Excel 2013, and the normal distribution map was generated using Origin Pro 8.0. The correlation analysis of the RIL population was performed using DPS7.05 statistical analysis software. Best linear unbiased prediction (BLUP) for SOC was evaluated using an R script (http://www.eXtension.org/pages/61006). The coefficient of variation was calculated using the formula $CV = \sigma/\mu$, where σ is the standard deviation and μ is the average.

Genetic map construction

Five young leaves of each accession were mixed, and DNA was extracted for SNP marker analysis. DNA samples were pretreated, chip hybridized, eluted, single base extended, stained, and embedded according to Illumina's Infinium™ HD Assay Ultra instructions. The chip was scanned using Illumina HiSCAN, and the results were analyzed using GenomeStudio genotyping software v2011, to obtain the genotype of each accession. The SNP genetic map was

constructed using MSTmap software (Wu et al., 2008), and all markers were grouped using a minimum threshold LOD score of 5.0. The order of markers on each linkage group was calculated using the minimum recombination frequency between the markers. The SNP genetic map included 8,575 SNP markers and 1,201 bins, and covered 6340.2 cM of the *B. napus* genome.

QTL mapping

The QTL mapping of SOC was performed in WinQTLCart 2.5 software (Silva Lda et al., 2012) using the composite interval mapping (CIM) method. The LOD threshold was set to 2.0. Each QTL was named after the year and the italic lowercase "*q*", followed by the trait name, chromosome, and QTL serial number. For example, *2016-qOCA05-1* indicates the first QTL for SOC on chromosome A05 in 2016.

Differentially expressed genesin high- and low-SOC accessions

One low-SOC (CQ46: Ningyou12, 33.0% of SOC) and two high-SOC (CQ24: SWU47, 42.7% of SOC and CQ52: Zhongshuang11, 43.3% of SOC) lines were selected for transcriptome sequencing (RNA Seq), and DEGs were obtained in seeds 30 days after flowering on the main inflorescence (30SM) and on the primary branch (30SB) based on our previous research (Xiao et al., 2019), and the RNA-Seq datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: BIG Data Center under BioProject accession number PRJNA602979.

Screening of SOC–related candidate genes

To screen for candidate genes related to SOC, we combined all candidate genes in the QTL confidence interval, with DEGs between extremely high- and low- SOC accessions in 30SM and 30SB. Finally, candidate genes that are up-regulated and down-regulated in both 30SM and 30SB of high-SOC rapeseed (CQ24 and CQ52) are identified to be important candidate genes related to SOC, and candidate genes that contribute to the formation of high-SOC will be focused on in this study.

Quantitative real–time polymerase chain reaction analysis

Total cDNA was synthesized from 1,000 ng RNA according to the manufacturer's instructions (Perfect Real Time; TaKaRa Biotechnology, Dalian, China). qRT-PCR Primers are listed in Table S3. Each reaction contained 10 μL TBGreen II (TakaRa), 2.0 μL cDNA, 1.6 μL primer, 0.4 μL ROX Reference Dye II, and distilled water to a final volume of 20 μL. The PCR program was as follows: 95 °C for 30 s and 35 cycles of 95 °C for 5 s, followed by 56-62 °C (depending on the primers used) for 30 s. Each reaction including three biological replicates, relative expression levels were obtained using the $2^{-\Delta\Delta Ct}$ method, and *BnACTIN7* was used as internal control.

Results

Phenotypic variation of SOC

In current study, extensive phenotypic variations of seed oil content (SOC) were found (Table 1), and seed oil content (SOC, % of seed weight) phenotypes in three environments (2016CQ-2018CQ) of 186 lines for QTL mapping are shown in Table S1. SOC ranged from 31.32% to 42.6%, with an average of 37.2%, in 2016CQ, from 30.97% to 43.87%, with an average of 38.81%, in 2017CQ, and from 34.18% to 43.54%, with an average of 39.11%, in 2018CQ. The coefficient of variation (CV) was 6.59%, 5.95% and 5.09% in 2016CQ, 2017CQ, and 2018CQ, respectively. This indicated that the variation of SOC in the three environments is small and stable. The SOC varied continuously in the three environments and approached a normal distribution (Figure 1), indicating that this trait is controlled by multiple genes.

Correlation analysis between SOC and other traits

The phenotypic correlation analysis between SOC and 12 other traits indicated (Table 2) that there was positive correlation

TABLE 1 Phenotypic variation in seed oil content (SOC) in the recombinant inbred line (RIL) population.

Trait	Env.	Mean ±SD(%)	Min (%)	Max (%)	CV(%)
SOC	2016CQ	37.2±2.45	31.32	42.6	6.59
	2017CQ	38.81±2.31	30.97	43.87	5.95
	2018CQ	39.11±1.99	34.18	43.54	5.09

CQ, Chongqing; SD, standard deviation; CV, coefficient of variation.

10.3389/fpls.2022.1067121

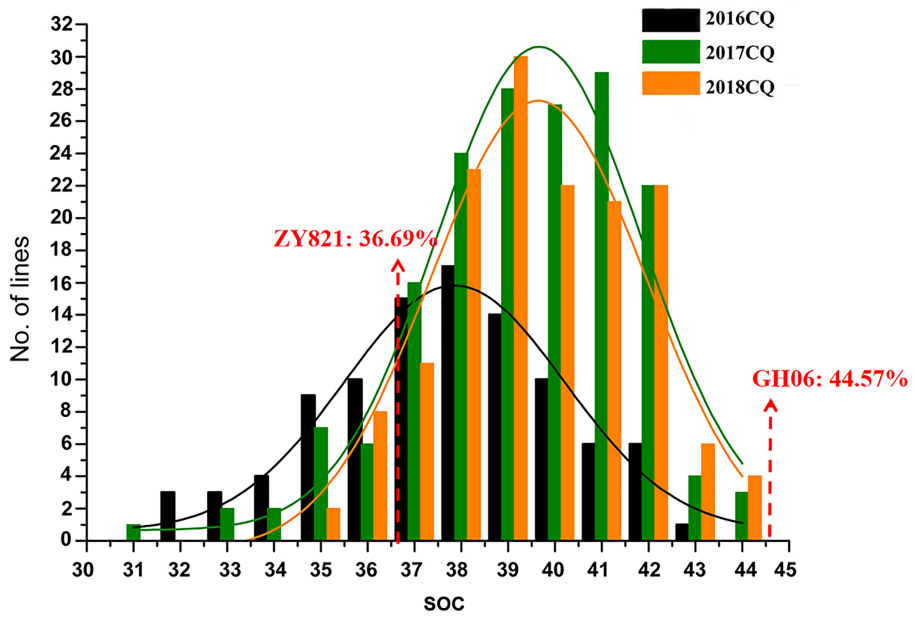


FIGURE 1
Frequency distribution of seed oil content (SOC, % of seed weight) of RIL populations in different environments. "CQ "refers to Chongqing and "No. of Lines" represents the number of accessions.

between SOC and oil + protein content, erucic acid, glucosinolate, linolenic acid, economic yield, biological yield, and harvest index. Besides, there was negative correlation between SOC and protein content, linoleic acid, stearic acid, oleic acid, and palmitic acid. And the significance of the correlation between SOC and other traits was shown in Table 2. Therefore, increasing the SOC has an extremely important role in production practice.

QTL mapping for SOC

QTL mapping for SOC was performed using the composite interval mapping (CIM) method. When the LOD value was≥2.0, we detected 26 SOC-related QTLs in three environments, located on chromosomes A01, A03, A05, A06, A09, C01, C03 and C05 (Table 3, Figure 2). These QTLs explained 3.69-18.47% of the phenotypic variation (R^2). The QTL *2017-qOCA09-2*, located at 30.08-30.31 Mb on chromosome A09, explained the highest proportion of the phenotypic variation (18.47%). We also identified several overlapping QTLs: *2016-qOCA09-1* and *2017-qOCA09-2*; *2016-qOCA09-2* and *BLUP-qOCA09-1*; *2018-qOCA09-1* and *BLUP-qOCA09-2*; *2016-qOCA05-1* and *2016-qOCA05-2*; *2018-qOCA03-1* and *2018-qOCA03-2*; and *2018-qOCA05-1* and *2018-qOCA05-2*. The additive effects of these overlapping QTLs are positive, indicating that the synergistic genes affecting SOC are mainly derived from the female parent (GH06), which is consistent with maternal control of SOC in *B. napus* (Hua et al., 2012).

TABLE 2 Correlation coefficients of seed oil content (SOC) with 12 other traits in a *B. napus* RIL population grown in different environments.

Trait	Env.	Protein content	Oil+Protein content	Erucic acid	Glucosinolate	Linolenic acid	Linoleic acid	Stearic acid	Oleic acid	Palmitic acid	Economic yield	Biological yield	Harvest index
SOC	2016CQ	-0.270**	0.857**	0.475**	0.148	0.412**	-0.639**	-0.442**	-0.406**	-0.730**	0.331**	0.221*	0.334**
	2017CQ	-0.346**	0.853**	0.505**	0.181*	0.058	-0.622**	-0.449**	-0.436**	-0.754**	0.409**	0.116	0.554**
	2018CQ	-0.565**	0.762**	0.200*	0.065	0.156*	-0.352**	-0.346**	-0.099	-0.454**	0.064	0.004	0.149

* and ** indicate significant differences at $p < 0.05$ and $p < 0.01$, respectively.

10.3389/fpls.2022.1067121

TABLE 3 QTL for seed oil content (SOC) detected from the RIL population in three environments.

QTL name	Env.	Chr.	LOD Score	SNP interval	Additive effect	R2 (%)	Physical interval (bp)	Candidate genes in confidence interval	Detected in previous studies
2016-*q*OCA05-1	2016CQ	ChrA05	3.71	SNP10512-SNP10774	1.43	11.10	9891125/13621300	BnaA05g15440D-BnaA05g18300D	
2016-*q*OCA05-2	2016CQ	ChrA05	2.11	SNP10749-SNP10827	1.23	6.56	10943170/13562855	BnaA05g16290D-BnaA05g18300D	
2016-*q*OCA05-3	2016CQ	ChrA05	2.96	SNP10556-SNP10574	-1.57	9.52	10799861/10877060	BnaA05g16210D-BnaA05g16250D	
2016-*q*OCA09-1	2016CQ	ChrA09	2.36	SNP21172-SNP21190	0.69	7.37	30078610/30361587	BnaA09g43400D-BnaA09g44150D	qOC-A9-4-TN (Jiang et al., 2014)
2016-*q*OCA09-2	2016CQ	ChrA09	4.52	SNP21199-SNP29923	0.93	13.44	30500499/30521739	BnaA09g44410D-BnaA09g44450D	cqOC-A9-9 (Chao et al., 2017); qOC-A9-4-TN (Jiang et al., 2014)
2017-*q*OCA06-1	2017CQ	ChrA06	2.62	SNP13773-SNP13754	0.52	4.91	20956439/21140416	BnaA06g31190D-BnaA06g31470D	
2017-*q*OCA09-1	2017CQ	ChrA09	4.66	SNP21160-SNP21167	0.79	10.81	29917465/30024926	BnaA09g43040D-BnaA09g43270D	qOC-A9-4-TN (Jiang et al., 2014)
2017-*q*OCA09-2	2017CQ	ChrA09	9.03	SNP21172-SNP21192	1.02	18.47	30078610/30386495	BnaA09g43400D-BnaA09g44200D	qOC-A9-4-TN (Jiang et al., 2014); Bn-A09-p32713083 (Liu et al., 2016)
2017-*q*OCC01-1	2017CQ	ChrC01	2.20	SNP38244-SNP38210	-0.47	3.94	11159198/11021503	BnaC01g15970D-BnaC01g16200D	Bn-scaff_17592_1-p654560 (Liu et al., 2016)
2017-*q*OCC03-1	2017CQ	ChrC03	2.82	SNP37315-SNP42418	0.52	4.73	47596910/46916297	BnaC03g57710D-BnaC03g58260D	
2017-*q*OCC05-1	2017CQ	ChrC05	2.23	SNP37339-SNP51044	0.47	3.97	40215021/40656747	BnaC05g43600D-BnaC05g44170D	cqOC-C5-7, cqOC-C5-8 (Chao et al., 2017)
2018-*q*OCA03-1	2018CQ	ChrA03	3.19	SNP7291-SNP7352	0.54	6.70	27827670/28701263	BnaA03g53180D-BnaA03g54210D	
2018-*q*OCA03-2	2018CQ	ChrA03	3.20	SNP7269-SNP7299	0.54	6.39	27627877/27900254	BnaA03g52920D-BnaA03g53200D	
2018-*q*OCA03-3	2018CQ	ChrA03	2.42	SNP7184-SNP7202	0.49	4.93	27277899/27337191	BnaA03g52340D-BnaA03g52400D	
2018-*q*OCA05-1	2018CQ	ChrA05	2.74	SNP11345-SNP11362	0.50	5.60	17522874/17737620	BnaA05g23110D-BnaA05g23420D	
2018-*q*OCA05-2	2018CQ	ChrA05	4.04	SNP11311-SNP11362	0.60	8.13	17277976/17737620	BnaA05g22780D-BnaA05g23420D	snp842906 (Wang et al., 2018)
2018-*q*OCA06-1	2018CQ	ChrA06	4.70	SNP13136-SNP13143	0.64	9.61	17149710/17214356	BnaA06g24680D-BnaA06g24800D	
2018-*q*OCA06-2	2018CQ	ChrA06	4.91	SNP13208-SNP13232	0.66	10.00	17811068/18056356	BnaA06g25740D-BnaA06g26190D	RNSL-qOC-A6 (Delourme et al., 2006); Bn-A06-p16689717 (Liu et al., 2016)
2018-*q*OCA06-3	2018CQ	ChrA06	2.43	SNP13320-SNP13329	0.47	5.19	18702433/18809438	BnaA06g27220D-BnaA06g27410D	
2018-*q*OCA09-1	2018CQ	ChrA09	4.80	SNP21215-SNP21326	0.68	10.45	30687778/31045659	BnaA09g44730D-BnaA09g45440D	cqOC-A9-9 (Chao et al., 2017); qOC-A9-4-TN (Jiang et al., 2014)
2018-*q*OCA09-2	2018CQ	ChrA09	4.22	SNP21369-SNP21385	0.61	8.61	31297302/31720463	BnaA09g45920D-BnaA09g46870D	cqOC-A9-10 (Chao et al., 2017)
BLUP-*q*OCA01-1	BLUP	ChrA01	4.51	SNP538-SNP620	-0.36	7.34	14133580/14587546	BnaA01g21840D-BnaA01g22210D	
BLUP-*q*OCA01-2	BLUP	ChrA01	5.40	SNP677-SNP837	-0.38	8.73	15577785/15683250	BnaA01g23190D-BnaA01g23330D	
BLUP-*q*OCA03-1	BLUP	ChrA03	2.53	SNP6846-SNP7283	0.25	3.69	24438882/27800984	BnaA03g47620D-BnaA03g53140D	*TN-qOC-A3-1* (Jiang et al., 2014); Bn-A03-p2948394 (Liu et al., 2016)
BLUP-*q*OCA09-1	BLUP	ChrA09	8.96	SNP21192-SNP21214	0.51	15.93	30386495/30681504	BnaA09g44210D-BnaA09g44720D	cqOC-A9-9 (Chao et al., 2017); qOC-A9-4-TN (Jiang et al., 2014); Bn-A09-p32713083,Bn-A09-p32864411 (Liu et al., 2016)
BLUP-*q*OCA09-2	BLUP	ChrA09	9.76	SNP21214-SNP21326	0.53	17.12	30681504/31045659	BnaA09g44720D-BnaA09g45440D	cqOC-A9-9 (Chao et al., 2017); qOC-A9-4-TN (Jiang et al., 2014)

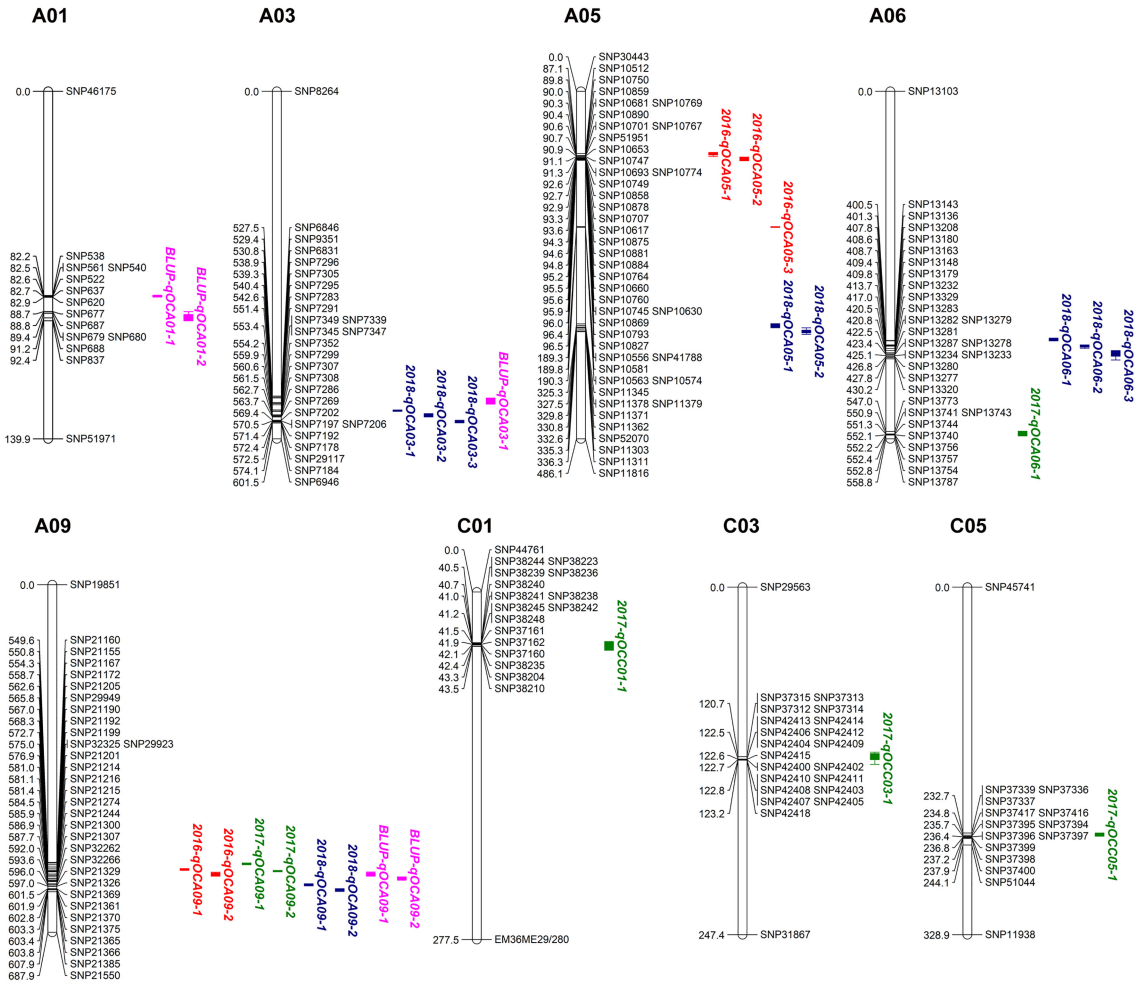


FIGURE 2
Putative QTL locations of seed oil content on the SNP genetic map. Red, green, blue and pink fonts represent QTL detected in 2016CQ, 2017CQ, 2018CQ and BLUP, respectively.

Screening of SOC–related candidate genes

We identified a total of 1,713 candidate genes within the mapping interval of the 26 SOC-related QTLs (Table S2). These were compared with the genes that were differentially expressed in 30SM and 30SB between extremely high- and low- oil content accessions. Finally, 11 genes that were up-regulated in both high-oil content (CQ24 and CQ52) accessions compared to low- oil content (CQ46) accession were screened as important candidate genes (Y1-Y11) that contribute to the formation of high SOC, and 10 genes that were down-regulated in both high- oil content (CQ24 and CQ52) accessions compared to low- oil content (CQ46) accession (Y12-Y21) were also considered candidate genes that affect SOC, and the details are listed in Table 4.

Validation of candidate genes by qRT –PCR analysis

To confirm the accuracy of the RNA-Seq results, 11 candidate genes (Y1-Y11) were performed qRT-PCR analysis, and the expression levels of these genes in 30SM and 30SB between high-SOC (CQ24 and CQ52) and low-SOC accessions (CQ46) by qRT-PCR and transcriptome sequencing (RNA-Seq) are shown in Figure 3. And the qRT-PCR results showed highly

10.3389/fpls.2022.1067121

TABLE 4 Identification of candidate genes by combining QTL mapping with RNA-Seq.

Code	Candidate genes	Physical position	*Arabidopsis*homologue	Functional description
Y1	BnaA01g22130D	ChrA01:14483566-14484250	Unknown	Unknown
Y2	BnaA03g48490D	ChrA03:24882893-24883369	AT4G27160	Seed storage albumin 3 (SESA3)
Y3	BnaA03g48800D	ChrA03:25043346-25045679	AT4G27760	FOREVER YOUNG (FEY)
Y4	BnaA03g50010D	ChrA03:25916589-25917524	AT4G30220	Small nuclear ribonucleoprotein F (RUXF)
Y5	BnaA03g50730D	ChrA03:26337386-26338898	AT3G48000	Aldehyde dehydrogenase 2B4 (ALDH2B4)
Y6	BnaA03g52350D	ChrA03:27303002-27304714	Unknown	Unknown
Y7	BnaA03g52640D	ChrA03:27453028-27455603	AT1G61180	LRR and NB-ARC domains-containing disease resistance protein
Y8	BnaA03g52660D	ChrA03:27458057-27461975	AT4G33630	EXECUTER1 (EX1)
Y9	BnaA03g53510D	ChrA03:28094719-28096583	AT4G36650	Plant-specific TFIIB-related protein (PBRP)
Y10	BnaA03g53860D	ChrA03:28425197-28427930	Unknown	Unknown
Y11	BnaA09g43150D	ChrA09:29959554-29960484	AT2G21490	Dehydrin LEA (LEA)
Y12	BnaA03g48780D	ChrA03:25039515-25040991	AT4G27680	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Y13	BnaA03g49200D	ChrA03:25322382-25322744	AT4G28365	Early nodulin-like protein 3 (ENODL3)
Y14	BnaA03g49250D	ChrA03:25378962-25382374	AT4G28410	Tyrosine transaminase family protein
Y15	BnaA03g52310D	ChrA03:27267291-27268611	AT4G33000	Calcineurin B-like protein 10 (CBL10)
Y16	BnaA06g31450D	ChrA06:21135320-21135661	AT3G28500	60S acidic ribosomal protein family
Y17	BnaA09g43660D	ChrA09:30169090-30169437	AT2G20619	Plant thionin family protein
Y18	BnaA09g45090D	ChrA09:30910078-30910815	AT1G17180	Glutathione S-transferase TAU 25 (GSTU25)
Y19	BnaA09g45320D	ChrA09:30987973-30988607	AT1G14980	Chaperonin 10 (CPN10)
Y20	BnaA09g46300D	ChrA09:31492745-31493868	AT2G37550	ARF-GAP domain 7 (AGD7)
Y21	BnaC05g43660D	ChrC05:40410175-40411373	AT3G08900	Reversibly glycosylated polypeptide 3 (RGP3)

consistent with RNA-Seq, which fully demonstrated the reliability and accuracy of the RNA-Seq data.

Expression patterns of candidate genes in seeds at different developmental stages

To explore the role of candidate genes in seed development, we investigated the expression profiles of all 21 genes in *B. napus ZS11* seeds at different developmental stages in the RNA-Seq dataset (BioProject ID PRJNA358784). The accuracy of the transcriptome sequencing has been verified in previous studies (Zhou et al., 2017; Di et al., 2018). 11 candidate genes (Y1-Y11) were expressed in seeds, but the levels often varied at different developmental stages (Figure 4; Table S4). For example, *BnaA03g48490D* (Y2) and *BnaA09g43150D* (Y11) were not expressed early in seed development, but were highly expressed in the middle and late stages. *BnaA01g22130D* (Y1), *BnaA03g50010D* (Y4), *BnaA03g52350D* (Y6), *BnaA03g53510D* (Y9), and *BnaA03g53860D* (Y10) showed little difference in expression over the course of seed development. *BnaA03g48800D* (Y3) was expressed in the middle stage of seed development, but was expressed at low levels in the early and late stages. *BnaA03g50730D* (Y5) exhibited relatively low expression across all stages. *BnaA03g52640D* (Y7) was expressed at higher levels in the middle and late stages than in the early stage, and the expression level of *BnaA03g52660D* (Y8) was higher in the late stage than in the early and middle stages of seed development. However, 5 candidate genes (Y12-Y15, Y20) were virtually unexpressed throughout seed development. Y16 and Y17 were only expressed early in seed development. The Y18 was expressed in small amounts throughout seed development, Y19 and Y21 genes were expressed during early and middle seed development. In general, the up-regulated candidate gene (Y1-Y11) in high-SOC accessions had higher expression levels than the down-regulated candidate gene (Y12-Y21) in the whole development process of seeds in *B. napus*.These results indicate that The up-regulated expression of 11 candidate genes may be more conducive to the formation of high-SOC. in *B. napus*.

Discussion

B. napus is an important oil crops around the world. Increasing SOC is a major goal of *B. napus* breeding (Mollers and Schierholt, 2002; Vigeolas et al., 2007; Tan et al., 2011; Fu et al., 2017). In current study, we found that there is a significant positive correlation between SOC and economic yield and harvest index (Table 2). In *B. napus* breeding, enhancing the SOC may increase both economic yield and harvest index. Therefore, discovering genetic loci that control SOC, and uncovering their genetic mechanism are key to developing

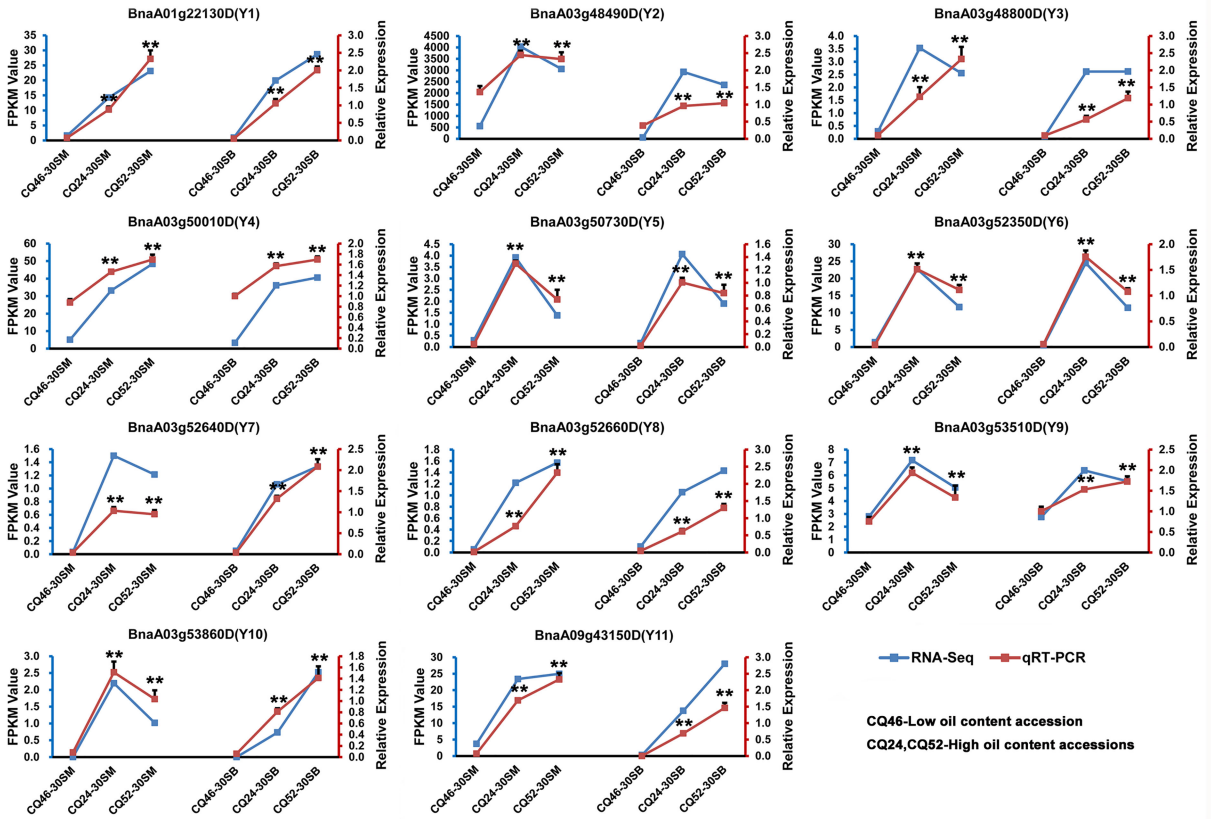


FIGURE 3
qRT-PCR analysis confirmed the accuracy of transcriptome sequencing by validating the expression patterns of 11 candidate genes in seeds 30 days after flowering on the main inflorescence (30SM) and on the primary branch (30SB). The blue line represents the RNA-Seq results and the red line represents the qRT-PCR results. **Denotes significance differences with $P < 0.01$, based on Student's *t*-test.

high-SOC varieties. Meanwhile, it is an effective way to identify candidate genes related to SOC and improve the SOC of *B. napus* by genetic engineering technology.

Although there have been many efforts to map QTLs controlling SOC, the loci discovered are different due to the use of different markers and populations. In current study, a new recombinant inbred population (RIL) containing 186 lines was constructed from a cross between 'GH06' and 'ZY821', two varieties with significant difference in SOC. A total of 26 QTLs were obtained, of which 13 are novel (Table 3). We were able to determine if they overlapped on the chromosomes by comparing the physical positions of QTLs obtained in this study with QTLs related to SOC obtained in previous studies using Darmor-bzh v4.1 reference genome of *Brassica napus* (Chalhoub et al., 2014). In this study, *2018-qOCA06-2* overlaps with the *RNSL-qOC-A6* obtained by Delourme et al (Delourme et al., 2006), *BLUP-qOCA03-1* overlaps with the *TN-qOC-A3-1*, and *2016-qOCA09-1*, *2016-qOCA09-2*, *2017-qOCA09-1*, *2017-qOCA09-2*, *2018-qOCA09-1*, *BLUP-qOCA09-1*, *BLUP-qOCA09-2* overlap with *qOC-A9-4-TN* obtained by Jiang et al (Jiang et al., 2014). *2016-qOCA09-2*, *2018-qOCA09-1*, *BLUP-qOCA09-1*, *BLUP-qOCA09-2* overlap with the *cqOC-A9-9*, *2018-qOCA09-2* overlaps with *cqOC-A9-10*, *2017-qOCC05-1* overlaps with cqOC-C5-7 and cqOC-C5-8 obtained in Chao's study (Chao et al., 2017). Besides, the *BLUP-qOCA03-1* overlaps with the Bn-A03-p2948394, *2018-qOCA06-2* overlaps with Bn-A06-p16689717, and *2017-qOCA09-2* overlaps with Bn-A09-p32713083, *BLUP-qOCA09-1* overlaps with Bn-A09-p32713083 and Bn-A09-p32864411, and *2017-qOCC01-1* overlaps with Bn-scaff_17592_1-p654560 detected by Liu et al (Liu et al., 2016). The *2018-qOCA05-2* obtained in this study overlaps with the snp842906 obtained from Wang et al (Wang et al., 2018). The detected QTLs overlapped within our study and with other QTLs identified by others, indicating high reliability for further analysis.

In this study, 21 candidate genes (Y1-Y21) were screened related to SOC. However, 11 up-regulated candidate genes (Y1-Y11) in high-SOC accessions were considered to be candidates for

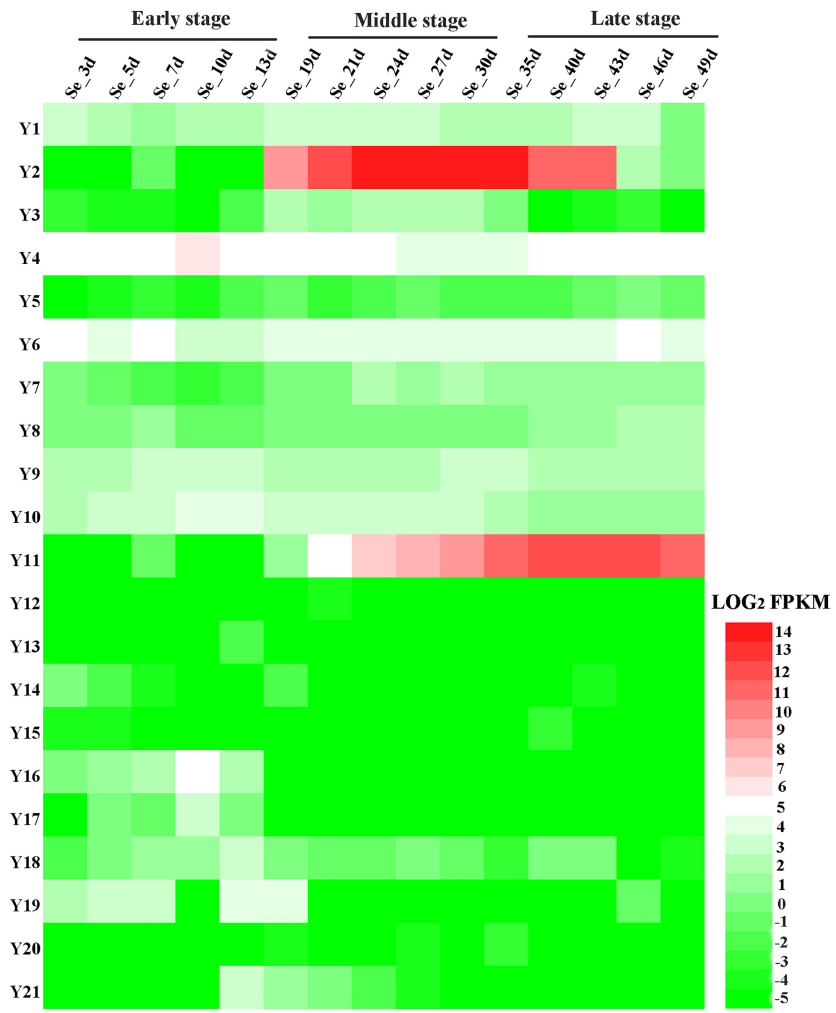


FIGURE 4
Expression patterns of 21 candidate genes in seeds at different developmental stages in *B. napus*. Se_3d, 5d, 7d, 10d, 13d, 19d, 21d, 24d, 27d, 30d, 35d, 40d, 43d, 46d, 49d represent seeds from *B. napus ZS11*, collected on the indicated number of days after pollination. The heatmap was drawn using Heatmap Illustrator (HemI) (Deng et al., 2014). The bar on the lower right corner represents LOG_2FPKM; green and red represent low and high expression levels, respectively.

the formation of high-SOC. Therefore, we will focus on these 11 candidate genes. Among all the 11 candidate genes (Y1-Y11) related to high-SOC formation, only Y1 (*BnaA01g22130D*), Y9 (*BnaA03g53510D*) and Y10 (*BnaA03g53860D*) were located in the novel QTL regions, and the remaining eight candidate genes were located in the overlapped QTLs detected previously. And there were seven candidate genes in the overlapped *BLUP-qOCA03-1* locus detected previously, which indicated that the *BLUP-qOCA03-1* may be a key locus associated with high SOC, which could provide a basis for marker-assisted breeding. And the homologous genes and functional descriptions of these genes in *Arabidopsis thaliana* were shown in Table 4. Among them, *BnaA01g22130D* (Y1), *BnaA03g52350D* (Y6) and *BnaA03g53860D* (Y10) are new proteins that have not been reported. The *Arabidopsis* homologs of other 8 candidate genes perform a variety of functions in growth, defense, and development (Table 4). Of these, we found *BnaA03g48490D* (Y2) is a putative seed storage albumin (*AtSESA3*). *BnaA03g48800D* (Y3) encodes an oxidoreductase required for proper development of the *Arabidopsis* vegetative shoot apex (Callos et al., 1994). *BnaA03g50010D* (Y4) is *AT4G30220* (*AtRUXF*) is a putative SmF component of Sm accessory ribonucleoprotein complex. *BnaA03g50730D* (Y5) encodes a putative (NAD+) aldehyde dehydrogenase. *BnaA03g52640D* (Y7) is LRR and NB-ARC domains-containing disease resistance protein. *BnaA03g52660D* (Y8) is a protein putatively involved in plastid to nucleus signaling (Lee et al.,

2007). *BnaA03g53510D* (Y9) is a general transcription factor for RNA polymerase I (Imamura et al., 2008). And we find *BnaA09g43150D* (Y11), the Late embryogenesis-abundant (LEA) homolog, to be the most compelling candidate because of its direct role in embryogenesis. LEA proteins are considered to be a large and highly diverse family involved in normal plant growth, seed development and the abiotic stress response (Chen et al., 2019; Jin et al., 2019). To date, LEA proteins have been identified in *Arabidopsis* (Hundertmark and Hincha, 2008), rice (*Oryza sativa*) (Wang et al., 2007), apple (*Malus domestica*) (Liang et al., 2012), tomato (*Solanum lycopersicum*) (Cao and Li, 2015), black popular (*Populus trichocarpa*) (Lan et al., 2013), and sweet orange (*Citrus sinensis* L. Osb.) (Pedrosa et al., 2015). Furthermore, Liang et al. (2016) identified 108 *BnLEA* genes in the *B. napus* genome and classified them into eight families based on their conserved domains (Liang et al., 2016). A study showed that overexpression (OE) of different copies of the drought response genes *LEA3* (not Y11) enhanced both drought tolerance and oil content in *Brassica napus* and *Arabidopsis*, and seed size, seed weight and membrane stability were also improved in OE lines. In contrast, oil content and drought tolerance were decreased in the *AtLEA3* mutant (*atlea3*) of *Arabidopsis* and in BnLEA-RNAi *B. napus* RNAi lines (Liang et al., 2019). Therefore, we speculate that the candidate gene Y11 obtained in this study has an important function in improving seed oil content and abiotic resistance in *B. napus*.

To date, the gene function in *B. napus* of the 11 candidate genes (Y1-Y11) that contribute to the formation of high-SOC were highlighted in this study has not been reported. We hypothesize that the 11 novel genes identified in our study influence the SOC of *B. napus*. Therefore, in the following research, we plan to explore these genes further to provide a basis for the molecular breeding of high-SOC *B. napus*.

Conclusions

We detected 26 QTLs associated with SOC in *B. napus*, that explained 3.69-18.47% of the phenotypic variation. Thirteen of these QTLs are reported here for the first time. And 1,713 candidate genes from the 26 QTLs mapping interval were obtained. Meanwhile, an analysis of DEGs between high- and low-SOC accessions revealed 21 candidate genes (Y1-Y21) related to SOC, and 11 candidate genes (Y1-Y11) contributing to the formation of high-SOC were highlighted. Current findings provide key information about the genetic loci associated with SOC, facilitating molecular breeding for higher SOC in *B. napus*.

Data availability statement

The RNA-Seq datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: BIG Data Center under BioProject accession number PRJNA602979. And CQ46-30SM corresponds to CQ46-30ZS, CQ24-30SM corresponds to CQ24-30ZS, CQ52-30SM corresponds to CQ52-30ZS, CQ46-30SB corresponds to CQ46-30CS, CQ24-30SB corresponds to CQ24-30CS, CQ52-30SB corresponds to CQ52-30CS.

Author contributions

JL and ZX conceived and designed the experiments. ZX, CZ, CQ, and LW performed the experiments. LZ, BY, and KL analyzed the data. ZX and JL wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the growth project of young scientific and technological talents in general colleges and universities in Guizhou Province (Qianjiaohe KY [2022] 101), Science and technology support project of Qianxinan Prefecture (2022QXN22728), Doctoral Research Fund project (20XYBS19), National Key R & D Program of China (2018YFD0200903; 2018YFD0100504), the "111"Project (B12006), Modern Agro-industry Technology Research System (CARS-13), Fundamental Research Funds for the Central Universities (XDJK2018C095; XDJK2017B030).

Acknowledgments

We extend our thanks to the reviewers for their careful reading and helpful comments on this manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1067121/full#supplementary-material

References

Callos, J. D., DiRado, M., Xu, B., Behringer, F. J., Link, B. M., and Medford, J. I. (1994). The forever young gene encodes an oxidoreductase required for proper development of the arabidopsis vegetative shoot apex. *Plant J.* 6 (6), 835–847. doi: 10.1046/j.1365-313x.1994.6060835.x

Cao, J., and Li, X. (2015). Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (Solanum lycopersicum). *Planta* 241 (3), 757–772. doi: 10.1007/s00425-014-2215-y

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., et al. (2014). Plant genetics. early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345 (6199), 950–953. doi: 10.1126/science.1253435

Chao, H. B., Wang, H., Wang, X. D., Guo, L. X., Gu, J. W., Zhao, W. G., et al. (2017). Genetic dissection of seed oil and protein content and identification of networks associated with oil content in *Brassica napus*. *Sci. Rep.* 7, 46295. doi: 10.1038/Srep46295

Chen, G., Geng, J. F., Rahman, M., Liu, X. P., Tu, J. X., Fu, T. D., et al. (2010). Identification of QTL for oil content, seed yield, and flowering time in oilseed rape (*Brassica napus*). *Euphytica* 175 (2), 161–174. doi: 10.1007/s10681-010-0144-9

Chen, Y., Li, C., Zhang, B., Yi, J., Yang, Y., Kong, C., et al. (2019). The role of the late embryogenesis-abundant (LEA) protein family in development and the abiotic stress response: A comprehensive expression analysis of potato (Solanum tuberosum). *Genes (Basel)* 10 (2), 148. doi: 10.3390/genes10020148

Delourme, R., Falentin, C., Huteau, V., Clouet, V., Horvais, R., Gandon, B., et al. (2006). Genetic control of oil content in oilseed rape (*Brassica napus* l.). *Theor. Appl. Genet.* 113 (7), 1331–1345. doi: 10.1007/s00122-006-0386-z

Deng, W. K., Wang, Y. B., Liu, Z. X., Cheng, H., and Xue, Y. (2014). HemI: A toolkit for illustrating heatmaps. *PloS One* 9 (11), e111988. doi: 10.1371/journal.pone.0111988

Di, F. F., Jian, H. J., Wang, T. Y., Chen, X. P., Ding, Y. R., Du, H., et al. (2018). Genome-wide analysis of the PYL gene family and identification of PYL genes that respond to abiotic stress in *Brassica napus*. *Genes* 9 (3), 156. doi: 10.3390/Genes9030156

Fu, Y., Zhang, D. Q., Gleeson, M., Zhang, Y. F., Lin, B. G., Hua, S. J., et al. (2017). Analysis of QTL for seed oil content in *Brassica napus* by association mapping and QTL mapping. *Euphytica* 213 (1), 17. doi: 10.1007/s10681-016-1817-9

Hua, W., Liu, J., and Wang, H. Z. (2016). Molecular regulation and genetic improvement of seed oil content in *Brassica napus* l. *Front. of Agric. Sci. and Eng.* 3 (3), 186–194. doi: 10.15302/J-Fase-2016107

Hua, W., Li, R. J., Zhan, G. M., Liu, J., Li, J., Wang, X. F., et al. (2012). Maternal control of seed oil content in *Brassica napus*: the role of silique wall photosynthesis. *Plant J.* 69 (3), 432–444. doi: 10.1111/j.1365-313X.2011.04802.x

Hu, Z. Y., Hua, W., Zhang, L., Deng, L. B., Wang, X. F., Liu, G. H., et al. (2013). Seed structure characteristics to form ultrahigh oil content in rapeseed. *PloS One* 8 (4), e62099. doi: 10.1371/journal.pone.0062099

Hundertmark, M., and Hincha, D. K. (2008). LEA (Late embryogenesis abundant) proteins and their encoding genes in arabidopsis thaliana. *BMC Genomics* 9, 118. doi: 10.1186/1471-2164-9-118

Imamura, S., Hanaoka, M., and Tanaka, K. (2008). The plant-specific TFIIB-related protein, pBrp, is a general transcription factor for RNA polymerase I. *EMBO J.* 27 (17), 2317–2327. doi: 10.1038/emboj.2008.151

Jiang, C. C., Shi, J. Q., Li, R. Y., Long, Y., Wang, H., Li, D. R., et al. (2014). Quantitative trait loci that control the oil content variation of rapeseed (*Brassica napus* l.). *Theor. and Appl. Genet.* 127 (4), 957–968. doi: 10.1007/s00122-014-2271-5

Jin, X. F., Cao, D., Wang, Z. J., Ma, L. L., Tian, K. H., Liu, Y. L., et al. (2019). Genome-wide identification and expression analyses of the LEA protein gene family in tea plant reveal their involvement in seed development and abiotic stress responses. *Sci. Rep.* 9 (1), 14123. doi: 10.1038/S41598-019-50645-8

Lan, T., Gao, J., and Zeng, Q. Y. (2013). Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in populus trichocarpa. *Tree Genet. Genomes* 9 (1), 253–264. doi: 10.1007/s11295-012-0551-2

Lee, K. P., Kim, C., Landgraf, F., and Apel, K. (2007). EXECUTER1- and EXECUTER2-dependent transfer of stress-related signals from the plastid to the nucleus of arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* 104 (24), 10270–10275. doi: 10.1073/pnas.0702061104

Liang, Y., Kang, K., Gan, L., Ning, S., Xiong, J., Song, S., et al. (2019). Drought-responsive genes, late embryogenesis abundant group3 (LEA3) and vicinal oxygen chelate, function in lipid accumulation in *Brassica napus* and arabidopsis mainly via enhancing photosynthetic efficiency and reducing ROS. *Plant Biotechnol. J.* 17 (11), 2123–2142. doi: 10.1111/pbi.13127

Liang, D., Xia, H., Wu, S., and Ma, F. W. (2012). Genome-wide identification and expression profiling of dehydrin gene family in malus domestica. *Mol. Biol. Rep.* 39 (12), 10759–10768. doi: 10.1007/s11033-012-1968-2

Liang, Y., Xiong, Z. Y., Zheng, J. X., Xu, D. Y., Zhu, Z. Y., Xiang, J., et al. (2016). Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*. *Sci. Rep.* 6, 24265. doi: 10.1038/Srep24265

Liu, S., Fan, C. C., Li, J. N., Cai, G. Q., Yang, Q. Y., Wu, J., et al. (2016). A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*. *Theor. and Appl. Genet.* 129 (6), 1203–1215. doi: 10.1007/s00122-016-2697-z

Mollers, C., and Schierholt, A. (2002). Genetic variation of palmitate and oil content in a winter oilseed rape doubled haploid population segregating for oleate content. *Crop Sci.* 42 (2), 379–384. doi: 10.2135/cropsci2002.3790

Pedrosa, A. M., Martins, C. D. S., Goncalves, L. P., and Costa, M. G. C. (2015). Late embryogenesis abundant (LEA) constitutes a Large and diverse family of proteins involved in development and abiotic stress responses in sweet orange (Citrus sinensis l. osb.). *PloS One* 10 (12), e0145785. doi: 10.1371/journal.pone.0145785

Silva Lda, C., Wang, S., and Zeng, Z. B. (2012). Composite interval mapping and multiple interval mapping: procedures and guidelines for using windows QTL cartographer. *Methods Mol. Biol.* 871, 75–119. doi: 10.1007/978-1-61779-785-9_6

Si, P., Mailer, R. J., Galwey, N., and Turner, D. W. (2003). Influence of genotype and environment on oil and protein concentrations of canola (*Brassica napus* l.) grown across southern Australia. *Aust. J. of Agric. Res.* 54 (4), 397–407. doi: 10.1071/AR01203

Sun, F. M., Liu, J., Hua, W., Sun, X. C., Wang, X. F., and Wang, H. Z. (2016). Identification of stable QTLs for seed oil content by combined linkage and association mapping in *Brassica napus*. *Plant Sci.* 252, 388–399. doi: 10.1016/j.plantsci.2016.09.001

Tan, H., Yang, X., Zhang, F., Zheng, X., Qu, C., Mu, J., et al. (2011). Enhanced seed oil production in canola by conditional expression of *Brassica napus* LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiol.* 156 (3), 1577–1588. doi: 10.1104/pp.111.175000

Vigeolas, H., Waldeck, P., Zank, T., and Geigenberger, P. (2007). Increasing seed oil content in oil-seed rape (*Brassica napus* l.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. *Plant Biotechnol. J.* 5 (3), 431–441. doi: 10.1111/j.1467-7652.2007.00252.x

Wang, X. F., Liu, G. H., Yang, Q., Hua, W., Liu, J., and Wang, H. Z. (2010). Genetic analysis on oil content in rapeseed (*Brassica napus* l.). *Euphytica* 173 (1), 17–24. doi: 10.1007/s10681-009-0062-x

Wang, X. D., Wang, H., Long, Y., Li, D. R., Yin, Y. T., Tian, J. H., et al. (2013). Identification of QTLs associated with oil content in a high-oil *Brassica napus* cultivar and construction of a high-density consensus map for QTLs comparison in b. napus. *PloS One* 8 (12), e80569. doi: 10.1371/journal.pone.0080569

Wang, B., Wu, Z. K., Li, Z. H., Zhang, Q. H., Hu, J. L., Xiao, Y. J., et al. (2018). Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. *Plant Biotechnol. J.* 16 (7), 1336–1348. doi: 10.1111/pbi.12873

Wang, X. S., Zhu, H. B., Jin, G. L., Liu, H. L., Wu, W. R., and Zhu, J. (2007). Genome-scale identification and analysis of LEA genes in rice (Oryza sativa l.). *Plant Sci.* 172 (2), 414–420. doi: 10.1016/j.plantsci.2006.10.004

Wu, Y. H., Bhat, P. R., Close, T. J., and Lonardi, S. (2008). Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PloS Genet.* 4 (10), e1000212. doi: 10.1371/journal.pgen.1000212

Xiao, Z., Zhang, C., Tang, F., Yang, B., Zhang, L., Liu, J., et al. (2019). Identification of candidate genes controlling oil content by combination of genome-wide association and transcriptome analysis in the oilseed crop *Brassica napus*. *Biotechnol. Biofuels* 12, 216. doi: 10.1186/s13068-019-1557-x

Yan, X. Y., Li, J. N., Fu, F. Y., Jin, M. Y., Liu, L. Z., et al. (2009). Co-location of seed oil content, seed hull content and seed coat color QTL in three different environments in *Brassica napus* L. *Euphytica* 170, 355–364. doi: 10.1007/s10681-009-0006-5

Zhao, J. Y., Huang, J. X., Chen, F., Xu, F., Ni, X. Y., Xu, H. M., et al. (2012). Molecular mapping of arabidopsis thaliana lipid-related orthologous genes in *Brassica napus*. *Theor. and Appl. Genet.* 124 (2), 407–421. doi: 10.1007/s00122-011-1716-3

Zhou, Y., Xu, D. X., Jia, L. D., Huang, X. H., Ma, G. Q., Wang, S. X., et al. (2017). Genome-wide identification and structural analysis of bZIP transcription factor genes in *Brassica napus*. *Genes* 8 (10), 288. doi: 10.3390/Genes8100288

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 December 2022
DOI 10.3389/fpls.2022.1098820

OPEN ACCESS

EDITED BY
Liezhao Liu,
Southwest University, China

REVIEWED BY
Jian Wu,
Yangzhou University, China
Jing Wen,
Huazhong Agricultural University, China
Yu Pan,
Southwest University, China

*CORRESPONDENCE
Maoteng Li
limaoteng426@hust.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 15 November 2022
ACCEPTED 08 December 2022
PUBLISHED 21 December 2022

CITATION
Cheng X, Liu X, He J, Tang M, Li H and Li M (2022) The genome wide analysis of *Tryptophan Aminotransferase Related* gene family, and their relationship with related agronomic traits in *Brassica napus*. *Front. Plant Sci.* 13:1098820. doi: 10.3389/fpls.2022.1098820

The genome wide analysis of *Tryptophan Aminotransferase Related* gene family, and their relationship with related agronomic traits in *Brassica napus*

Xin Cheng[1†], Xinmin Liu[1†], Jianjie He[1], Mi Tang[1], Huaixin Li[1] and Maoteng Li[1,2*]

[1]College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China, [2]Key Laboratory of Molecular Biophysics, the Ministry of Education of China, Wuhan, China

*Tryptophan Aminotransferase of Arabidopsis1/Tryptophan Aminotransferase-*Related (*TAA1/TAR*) proteins are the enzymes that involved in auxin biosynthesis pathway. The *TAA1/TAR* gene family has been systematically characterized in several plants but has not been well reported in *Brassica napus*. In the present study, a total of 102 *BnTAR* genes with different number of introns were identified. It was revealed that these genes are distributed unevenly and occurred as clusters on different chromosomes except for A4, A5, A10 and C4 in *B. napus*. Most of the these *BnTAR* genes are conserved despite of existing of gene loss and gene gain. In addition, the segmental replication and whole-genome replication events were both play an important role in the *BnTAR* gene family formation. Expression profiles analysis indicated that the expression of *BnTAR* gene showed two patterns, part of them were mainly expressed in roots, stems and leaves of vegetative organs, and the others were mainly expressed in flowers and seeds of reproductive organs. Further analysis showed that many of *BnTAR* genes were located in QTL intervals of oil content or seed weight, for example *BnAMI10* was located in *cqOC-C5-4* and *cqSW-A2-2*, it indicated that some of the *BnTAR* genes might have relationship with these two characteristics. This study provides a multidimensional analysis of the *TAA1/TAR* gene family and a new insight into its biological function in *B. napus*.

KEYWORDS

genome-wide identification, *TAR* gene family, agronomic traits, *Brassica napus*, phylogenetic analysis

Introduction

Auxin is an essential phytohormone and controls almost every aspect during plant development (Paque and Weijers, 2016). At cellular level, auxin could stimulate cell division of cambium and cell elongation of branch, promote cell differentiation of xylem and phloem, and regulate callus morphogenesis (Paque and Weijers, 2016). At the organ and whole plant level, auxin could affect almost all developmental steps in plants from early embryogenesis to fruit ripening, and control organogenesis at the meristems, which determines the plant structure (Paque and Weijers, 2016). In addition, the application of auxin could reduce the lipid peroxidation of plant cells, thereby reducing the consumption of unsaturated fatty acids (Dhindsa et al., 1984). Meanwhile, it was also found that the biomass and oil yield was increased by adding appropriate amount of auxin in microalgal (Chang et al., 2022).

Indoleacetic acid (IAA) is the best-studied naturally active auxin. IAA is biosynthesized through two major ways: tryptophan (Trp)-dependent and Trp-independent pathways (Woodward and Bartel, 2005). The indole-3-pyruvic acid (IPyA) pathway, which is a tryptophan (Trp)-dependent pathway, appears to be the main contributing factor to the formation of free IAA (Zhao, 2012). There are two steps in the IPyA pathway to converse the Trp to IAA: The first step is catalyzed by the *Tryptophan Aminotransferase of Arabidopsis 1* (*TAA1*), which belongs to *TAA1*-Related (*TAA1/TAR*) family of Trp aminotransferases that transfers the Trp to an alpha-keto acid, and to generate IPyA and another amino acid like L-glutamate (Zhao, 2012). The second step is transferring IPyA to IAA by an oxygen and NADPH-dependent reaction, which is catalyzed by the YUCCA (YUC) family of flavin monooxygenases (Zhao, 2012). Trp-independent pathways have not been fully explored yet, it was reported that some enzymes like Trp synthase α (TSA1) and indole synthase (INS) can catalyze the cleavage of indole 3-glycerol phosphate (IGP) to form indole and D-glyceraldehyde 3-phosphate (G_3P), and indole was transformed into IAA (Zhang R. et al., 2008). In Arabidopsis genome, four genes were found to have close relationship with *TAA1*, which were referred as *Tryptophan Aminotransferase Related 1-4* (*TAR1*, *TAR2*, *TAR3*, *and TAR4*) (Stepanova et al., 2008). It was revealed that the *BnTAR* proteins could classify into two families (Alliinase_C and Aminotran_1_2) according to the primary sequence and homology, which both belong to the superfamily of pyridoxal-5'-phosphate (PLP) dependent enzymes (Liepman and Olsen, 2004). The lack of the Epidermal Growth Factor (EGF) domain suggests that these proteins are not typical alliinases, although they share high sequence similarities with the EGF-alliinase group of C-S lyases (Stepanova et al., 2008).

B. napus is an important oil crop, it was an allotetraploid species that derived from the hybridization of *B. rapa* and *B. oleracea* in about 7500 years ago, its genome was contained many duplications as well as inversions and translocations (Chalhoub et al., 2014). Genomic differentiation events that resulted from whole-genome triplication (WGT) was occurred in Brassica species (Cheng et al., 2014). Structural and functional divergence of duplicate genes were also found in other gene families, such as the basic/helix-loop-helix (bHLH) gene family, the ionotropic glutamate receptor (iGluR) gene family, and the Receptor-like kinases (RLKs) gene family (Chiu et al., 2002; Shiu and Bleecker, 2003; Toledo-Ortiz et al., 2003). The characteristics of the duplicate genes suggested that the evolution could have caused an adaptive structural diversification, and this process was pervasive and could have contributed to the biological novelty in plants (Li et al., 2009). Thus, many duplicate genes have similar gene sequences but different functional performances (Liang et al., 2017).

Although there is one study on transaminases involved the *TAA1/TAR* protein of *B. napus* (Le Deunff et al., 2019), which suggests that *TAA1/TAR* not only play an important role in IAA biosynthesis but also in the plant nitrogen cycle. The previous studies have revealed that knock down of *TAA1/TAR* could lead to reduced plant height, decreased apical dominance, valveless gynoecia, vasculature defects, reduced organ number and complete sterility in *Arabidopsis thaliana* (Stepanova et al., 2008). Otherwise, over-expression of *TaTAR* could promote the root elongation (Shao et al., 2017). In addition, *TAA1/TAR* is also involved in ethylene response, it was revealed that the *TAA1/TAR* mutants were showed root-specific ethylene insensitivity in Arabidopsis (Stepanova et al., 2005). However, the gene evolution and structure of the *BnTAR* gene family in *B. napus* have not been reported.

Combined multiple methods to analyses the gene function was frequently used (Kim et al., 2016; Depuydt and Vandepoele, 2021). In our previous studies, many of the QTLs for the seed weight and oil content were identified (Raboanatahiry et al., 2018; Zhao et al., 2019; Yan et al., 2022), for instance, *cqSW-A2-1*, *cqSW-A7-2*, cq*SW-C1-2* and *cqSW-C9-2* for seed weight and *cqOC-A1-1*, *cqOC-A9-11*, *cqOC-C2-2* and *cqOC-C9-4* for oil content. These QTLs provided the convenience for the analysis of the *BnTAR* genes in QTL interval.

In this study, the systematical identification, structure and evolutionary analysis of the *BnTAR* gene family in *B. napus* were analyzed, and the expression patterns of several selected *BnTAR* members in different tissues of *B. napus* and their relationship with seed oil content and yield were also explored, which would provide informative clues to the future functional study.

Materials and methods

Identification of TAA1/TAR family genes in B. napus and other species

We used BLAST program to search Ensembl Plants (Ensembl Plants, RRID : SCR_008680) (http://plants.ensembl.org/index.html) (Yates et al., 2022) and 35 *AtTAR* paralogues, which are

homologous to *Arabidopsis thaliana AtTAR1* in phylogenetical tree, were identified. *TAA1/TAR* genes were identified in *B. napus* based on sequence similarity to the 36 *TAR* protein sequences of Arabidopsis by using the BLAT search program (BLAT, RRID : SCR_011919) in CNS-Genoscope database (CNS Genoscope, RRID : SCR_023020) (https://www.genoscope.cns.fr/brassicanapus/) (Chalhoub et al., 2014), and redundant sequences were removed manually. The *TAA1/TAR* genes in *B. rapa* and *B. oleracea* were obtained from the Brad database (Brassicaceae Database, RRID : SCR_023019) (http://brassicadb.cn) (Chen et al., 2022). All the candidate *BnTAR* genes were analyzed by using the Pfam database (Pfam, RRID : SCR_004726) (http://pfam.xfam.org/search) (El-Gebali et al., 2019), SMART database (SMART, RRID : SCR_005026) (http://smart.embl-heidelberg.de/) (Letunic et al., 2004) and NCBI Conserved Domain Search database (Conserved Domains Search, RRID : SCR_018729) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Lu et al., 2020) to confirm whether they are belonging to *TAA1/TAR* family or not. The categories of *TAA1/TAR* gene family in *B. napus* were using the Pfam analysis. For convenience, a univocal name is assigned to each *BnTAR* gene. For example, in the gene of '*BnAMI1*', the first two letters denote the source organism *Brassica napus*, following by the family name and the gene number index (Ostergaard and King, 2008).

The *TAR* genes were also identified in other higher plants in order to trace the evolutionary origin of this family in Phytozome (Phytozome, RRID : SCR_006507) (https://phytozome-next.jgi.doe.gov/) (Goodstein et al., 2012), including of *Amborella Trichopoda*, *Arabidopsis Thaliana*, *Brachypodium distachyon*, *Brassica oleracea*, *Brassica rapa*, *Chromochloris Zofingiensis*, *Glycine Max*, *Gossypium Hirsutum*, *Malus Domestica*, *Oryza Sativa* and *Zea Mays*. In addition, NCBI BLAST (NCBI BLAST, RRID : SCR_004870) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1990) was also used to search the *TAA1/TAR* family genes.

The chromosome locations, CDS lengths and the number of amino acids of the *BnTAR* genes were obtained from the CNS-Genoscope database (Chalhoub et al., 2014). The molecular weight (Mw) and isoelectric point (pI) were calculated by the Compute pI/Mw tool of ExPASy (ExPASy Bioinformatics Resource Portal, RRID : SCR_012880) (https://web.expasy.org/compute_pi/) (Bjellqvist et al., 1993; Bjellqvist et al., 1994). The grand average of hydropathy values (GRAVY) were calculated using the protein hydrophilicity analysis tool (protein hydrophilicity analysis tool, RRID : SCR_023015) (http://www.detaibio.com/sms2/protein_gravy.html) (Kyte and Doolittle, 1982). Subcellular location prediction was conducted using the Plant-mPLoc2.0 (Plant mPLoc, RRID : SCR_023014) (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) (Chou and Shen, 2010) and Multi Loc2 (Multi Loc2, RRID : SCR_023013) (https://abi-services.informatik.uni-tuebingen.de/multiloc2/webloc.cgi) (Blum et al., 2009).

Multiple sequence alignment, phylogenetic and structural analysis

Multiple sequence alignment of all selected *BnTAR*, *BrTAR* and *BoTAR* protein sequences was conducted by using ClustalW (ClustalW, RRID : SCR_017277) in MEGA-X (MEGA, RRID : SCR_023017) (Higgins and Sharp, 1988; Tamura et al., 2013). The unrooted phylogenetic tree of full-length *BnTAR* protein sequences was constructed by using the MEGA-X with Neighbor-Joining (NJ) method (Saitou and Nei, 1987), and the bootstrap analysis was performed using 1,000 replicates.

Based on the alignments of their coding sequences with the corresponding genomic sequences, the exon-intron structures of the *BnTAR* genes were determined and a diagram was drawn by using Gene structure display server (Gene Structure Display Server, RRID : SCR_023011) (GSDS; http://gsds.gao-lab.org/) (Hu et al., 2015). The conserved motif structures of *BnTAR* genes were searched by using Multiple Expectation Maximization for Motif Elicitation (MEME Suite - Motif-based sequence analysis tools, RRID : SCR_001783) (http://alternate.meme-suite.org/) (Bailey et al., 2009). Secondary structural analysis was carried out by using the two following tools: PBIL GOR4 (PBIL GOR4, RRID : SCR_023012) (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) (Jones, 1999) and PSIPRED (PSIPRED, RRID : SCR_010246) (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan et al., 2013). The Tertiary structures of the *BnTAR*s were predicted using Phyre2 (Phyre, RRID : SCR_010270) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley and Sternberg, 2009) and analyzed by VAST (Vector Alignment Search Tool, RRID : SCR_010655) (https://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html) (Gibrat et al., 1996).

Chromosomal location and gene duplication analysis

The chromosomal locations of the *BnTAR* genes were obtained from the CNS-Genoscope. Neighboring to homologous *BnTAR* genes on *B. napus* chromosomes or within a sequence distance of 50 kb were defined as tandemly duplicated *BnTAR* genes (Yu et al., 2014). According to the definition of tandemly duplicated genes, the genes with a distance of less than 50kb between the nearest genes as gene clusters. The synteny relationships between the *BnTAR*s and *AtTAR*s, *BrTAR*s, and *BoTAR*s were assessed by using the syntenic genes searching tool in BRAD database (http://brassicadb.org/brad/) (Wang et al., 2011) and the synteny tool in the *B. napus* Genome Browser (Chalhoub et al., 2014).

Calculation of the Ka/Ks

The paralogue pairs were determined according to the homology between *A. thaliana* and *B. napus*. The synonymous substitution rate and the non-synonymous substitution rate of all gene pairs were calculated by TBtools (TBtools, RRID : SCR_023018) (Toolbox for Biologists) (Chen et al., 2020).

Expression analysis of *TAA1/TAR* family genes in the intervals of QTLs for seed oil content and seed weight

The expression data of *TAR* family genes in various tissues and periods in *B. napus* could be accessed from BnTIR (BnTIR, RRID : SCR_023021) (http://yanglab.hzau.edu.cn/BnTIR) (Liu et al., 2021). The transcriptome data of seeds with different oil content or weight were obtained from the previous study of our group and can be found in the SRA database of NCBI (accessions for these SRA data: PRJNA661261 and PRJNA730112).

The QTLs from KN population for seed oil content and seed weight in *B. napus* were obtained from the previous published papers in our group (Raboanatahiry et al., 2018; Zhao et al., 2019; Yan et al., 2022), were used for determine the whether the *TAA1/TAR* family genes were located in the QTL interval or not. The KN population was derived from the microspore culture of F1 hybrids that derived from the hybridization between KenC-8 and N53-2, the oil content was about 40% and 50% in all cultivated environments, respectively (Wang et al., 2013), and the thousand seed weight was about 3.53 g and 3.81g in Dali of Shaanxi Province, respectively (Zhao et al., 2016). The genetic distances of the scanned QTLs are converted into physical distances according to the corresponding table (Table S8), in which the marker sequence is blasted on the genome to obtain the physical distance corresponding to the genetic distance. The start and end loci of the *BnTAR* genes were compared to the QTL interval to determine whether they were located within the QTL interval or not.

RNA extraction and RT-qPCR analysis

An RNAprep Pure Plant Plus Kit (Tiangen) was used to separate the total RNA from each frozen sample, and then ReverTra Ace® qPCR RT Master Mix with gDNA Remover (TOYOBO) was used to synthesize the first strand cDNA from the RNA according to the manufacturer's instructions. Each reaction was repeated three times with a reaction volume of 20 µl containing 0.6 µl of gene-specific primers (1.0 µM), 1.0 µl of cDNA, 10 µl of SYBR green (TaKaRa), and 8.4 µl RNase-free double-distilled water. The PCR conditions were set as follows: Stage 1: 95°C for 5 min; stage 2: 40 cycles of 40 cycles, 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute. In the stage 2, the fluorescence intensity was measured. A housekeeping gene (*TIP41*) was used as a reference for normalization and analyzed by using a Real-Time PCR System StepOne Plus (ABI), and each reaction was performed in triplicate. The primers for mRNA RT–qPCR are listed in Table S6.

Results

Genome-wide identification of the *BnTAR* superfamily genes in the *B. napus* genome

To identify the possible candidate *BnTAR* genes, we searched the CNS Genoscope database with the *AtTAR* genes for their homologies in *B. napus*. As a result, a total of 102 genes as *BnTAR* gene family members were identified, and it could be classified into Alliinase_C and Aminotran_1_2 two families based on conserved structural domains in Pfam database. The Alliinase_C family includes 12 members (named as from *BnALL1* to *BnALL12*), and the other family includes 90 members (designated as *BnAMI1- BnAMI90*) (Table 1).

The properties of *BnTAR* were analyzed by Expasy database. All the *BnTAR* proteins had molecular weights less than 70 kDa, but the isoelectric points varied in a wide range, which are from 4.78 to 9.68 (Table S1). The hydrophilicity of *BnTAR*s was assessed based on the grand average of hydropathy (GRAVY) value (Table S1). On average, the GRAVY values of *BnALL* proteins are higher than those of *BnAMI* proteins, it indicated the higher hydrophilicity of *ALL* proteins. Meanwhile, almost all the *BnTAR*s with the GRAVY values less than zero, which demonstrated that most of *BnTAR* proteins are hydrophilic. The subcellular localization of 102 *BnTAR* proteins was predicted by using Plant-mPLoc and Multi Loc2 (Tables S1 and S2), it was revealed that most of them were located in the cytoplasm or chloroplast, only a few proteins were predicted in the secretory pathway or the nucleus.

Sequence alignment and phylogenetic analysis of *BnTAR* genes

We performed sequence alignment of *BnTAR* genes and constructed an unrooted phylogenetic tree of 102 *BnTAR* genes. This was used to determine the evolutionary relationships among *BnTAR* genes, and we found that most of the genes formed gene pairs with high Bootstrap support (Figure 1), it suggested that the chromosome duplication was contributed significantly to *BnTAR* gene formation. And most of the gene pairs have short branches (Figure 1), which might be separated in a more recent time. Aminotran_1_2 was generally divided into four large branches, among which *BnAMI8* and *BnAMI54* are in a separate clade (Figure 1). They are distant from the other members of the Aminotran_1_2 family, they might

10.3389/fpls.2022.1098820

TABLE 1 Genome identification and classification of *TARs* genes in *B. napus*.

Name	Gene ID	Family	Name	Gene ID	Family
BnALL1	BnaA01g14030D	Alliinase_C	*BnAMI40*	BnaA09g45030D	Aminotran_1_2
BnALL2	BnaA02g14990D	Alliinase_C	*BnAMI41*	BnaA10g00470D	Aminotran_1_2
BnALL3	BnaA08g06520D	Alliinase_C	*BnAMI42*	BnaAnng08090D	Aminotran_1_2
BnALL4	BnaC08g07070D	Alliinase_C	*BnAMI43*	BnaAnng22050D	Aminotran_1_2
BnALL5	BnaAnng22040D	Alliinase_C	*BnAMI44*	BnaAnng30490D	Aminotran_1_2
BnALL6	BnaC01g16530D	Alliinase_C	*BnAMI45*	BnaAnng31450D	Aminotran_1_2
BnALL7	BnaC02g19980D	Alliinase_C	*BnAMI46*	BnaC01g01270D	Aminotran_1_2
BnALL8	BnaC06g07950D	Alliinase_C	*BnAMI47*	BnaC01g04560D	Aminotran_1_2
BnALL9	BnaC06g43720D	Alliinase_C	*BnAMI48*	BnaC01g10020D	Aminotran_1_2
BnALL10	BnaC05g18610D	Alliinase_C	*BnAMI49*	BnaC01g10040D	Aminotran_1_2
BnALL11	BnaA09g31200D	Alliinase_C	*BnAMI50*	BnaC01g15060D	Aminotran_1_2
BnALL12	BnaA09g31180D	Alliinase_C	*BnAMI51*	BnaC01g18250D	Aminotran_1_2
BnAMI1	BnaA01g00270D	Aminotran_1_2	*BnAMI52*	BnaC02g13860D	Aminotran_1_2
BnAMI2	BnaA01g03290D	Aminotran_1_2	*BnAMI53*	BnaC02g15560D	Aminotran_1_2
BnAMI3	BnaA01g08470D	Aminotran_1_2	*BnAMI54*	BnaC02g16260D	Aminotran_1_2
BnAMI4	BnaA01g13270D	Aminotran_1_2	*BnAMI55*	BnaC02g20870D	Aminotran_1_2
BnAMI5	BnaA01g13280D	Aminotran_1_2	*BnAMI56*	BnaC02g23160D	Aminotran_1_2
BnAMI6	BnaA01g15380D	Aminotran_1_2	*BnAMI57*	BnaC03g26850D	Aminotran_1_2
BnAMI7	BnaA02g09960D	Aminotran_1_2	*BnAMI58*	BnaC03g28000D	Aminotran_1_2
BnAMI8	BnaA02g11830D	Aminotran_1_2	*BnAMI59*	BnaC03g45280D	Aminotran_1_2
BnAMI9	BnaA02g15650D	Aminotran_1_2	*BnAMI60*	BnaC03g66430D	Aminotran_1_2
BnAMI10	BnaA02g19790D	Aminotran_1_2	*BnAMI61*	BnaC03g76570D	Aminotran_1_2
BnAMI11	BnaA03g22790D	Aminotran_1_2	*BnAMI62*	BnaC04g35390D	Aminotran_1_2
BnAMI12	BnaA03g23600D	Aminotran_1_2	*BnAMI63*	BnaC05g00530D	Aminotran_1_2
BnAMI13	BnaA03g38440D	Aminotran_1_2	*BnAMI64*	BnaC05g13450D	Aminotran_1_2
BnAMI14	BnaA03g38470D	Aminotran_1_2	*BnAMI65*	BnaC05g18600D	Aminotran_1_2
BnAMI15	BnaA03g40560D	Aminotran_1_2	*BnAMI66*	BnaCnng71530D	Aminotran_1_2
BnAMI16	BnaA03g46060D	Aminotran_1_2	*BnAMI67*	BnaC06g19110D	Aminotran_1_2
BnAMI17	BnaA03g47880D	Aminotran_1_2	*BnAMI68*	BnaC06g20760D	Aminotran_1_2
BnAMI18	BnaA03g49250D	Aminotran_1_2	*BnAMI69*	BnaC06g32920D	Aminotran_1_2
BnAMI19	BnaA04g13310D	Aminotran_1_2	*BnAMI70*	BnaC06g38270D	Aminotran_1_2
BnAMI20	BnaA05g12790D	Aminotran_1_2	*BnAMI71*	BnaC06g40630D	Aminotran_1_2
BnAMI21	BnaA06g11610D	Aminotran_1_2	*BnAMI72*	BnaC06g43710D	Aminotran_1_2
BnAMI22	BnaA06g15630D	Aminotran_1_2	*BnAMI73*	BnaC07g13140D	Aminotran_1_2
BnAMI23	BnaA07g00460D	Aminotran_1_2	*BnAMI74*	BnaC07g31520D	Aminotran_1_2
BnAMI24	BnaA07g10020D	Aminotran_1_2	*BnAMI75*	BnaC07g40130D	Aminotran_1_2
BnAMI25	BnaA07g29720D	Aminotran_1_2	*BnAMI76*	BnaC07g41260D	Aminotran_1_2

(Continued)

TABLE 1 Continued

Name	Gene ID	Family	Name	Gene ID	Family
BnAMI26	BnaA07g35690D	Aminotran_1_2	*BnAMI77*	BnaC07g41280D	Aminotran_1_2
BnAMI27	BnaA07g38380D	Aminotran_1_2	*BnAMI78*	BnaC08g06270D	Aminotran_1_2
BnAMI28	BnaA08g10670D	Aminotran_1_2	*BnAMI79*	BnaC08g17310D	Aminotran_1_2
BnAMI29	BnaA08g11470D	Aminotran_1_2	*BnAMI80*	BnaC08g21300D	Aminotran_1_2
BnAMI30	BnaA08g20540D	Aminotran_1_2	*BnAMI81*	BnaC08g33730D	Aminotran_1_2
BnAMI31	BnaA08g23190D	Aminotran_1_2	*BnAMI82*	BnaC08g36300D	Aminotran_1_2
BnAMI32	BnaA09g07210D	Aminotran_1_2	*BnAMI83*	BnaC08g37860D	Aminotran_1_2
BnAMI33	BnaA09g10030D	Aminotran_1_2	*BnAMI84*	BnaC08g49010D	Aminotran_1_2
BnAMI34	BnaA09g13010D	Aminotran_1_2	*BnAMI85*	BnaC09g07070D	Aminotran_1_2
BnAMI35	BnaA09g21860D	Aminotran_1_2	*BnAMI86*	BnaC09g10110D	Aminotran_1_2
BnAMI36	BnaA09g31220D	Aminotran_1_2	*BnAMI87*	BnaC09g12920D	Aminotran_1_2
BnAMI37	BnaA09g39260D	Aminotran_1_2	*BnAMI88*	BnaC09g24050D	Aminotran_1_2
BnAMI38	BnaA09g41170D	Aminotran_1_2	*BnAMI89*	BnaCnng43090D	Aminotran_1_2
BnAMI39	BnaA09g43670D	Aminotran_1_2	*BnAMI90*	BnaCnng64190D	Aminotran_1_2

have diverged from other *BnAMI* genes earlier and their function might be more independently. Also, twelve genes of the Alliinase_C family are clustered together and share high similarities (Figure 1).

According to the sequence comparison results, protein sequences of members of *BnALL* family are very conserved, with about 30.5%-36.7% of amino acids having over 90% similarity (Figure S1). In contrast, sequence conservation of the members of *BnAMI* family is of low level. Among the four large branches of the Aminotran_1_2 family, the branch that contained *BnAMI31* and *BnAMI65* was had the highest internal conserved type (Figure S1). About 45.2%-61.5% of the amino acids in this branch were with more than 90% similarity across the sequences, which suggested that these genes were had a high level of functional redundancy. The branch that contained the *BnAMI2* and *BnAMI69* was the least conserved, with about 6.83%-10.0% of amino acids having high similarity except *BnAMI70* which is rather short (Figure S1). The large number of variation sites suggested that these genes have important roles in adaptation to ever-changing environment. The branch of *BnAMI39* and *BnAMI86* was with a close number of conserved amino acids to the branch of *BnAMI8* and *BnAMI80* (Figure S1), but the overall conservation was lower due to the presence of two genes with longer sequences, *BnAMI8* and *BnAMI54*.

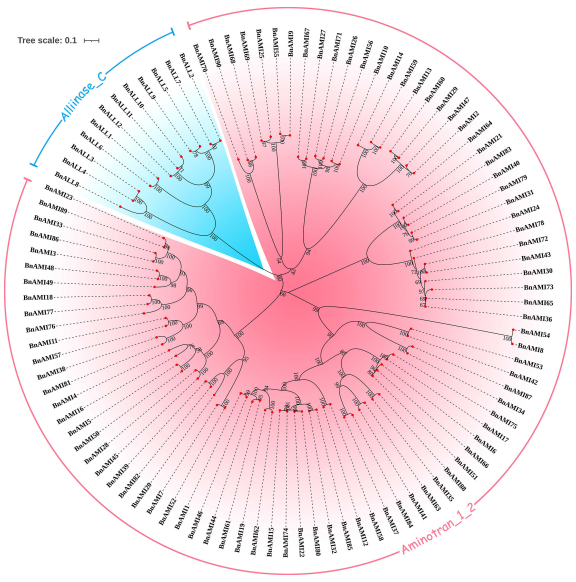


FIGURE 1
Phylogenetic analysis of the *BnTAR* genes. Different *BnTAR* gene families were marked in different colors.

Chromosomal location and the expansion pattern of the *BnTAR* Genes

The chromosomal locations of *BnTAR* genes were analyzed, and 102 *BnTAR* genes that located on the 19 chromosomes of *B. napus* were found (Figure 2). However, only one *BnTAR* gene was located on chromosomes A4, A5, A10 and C4, respectively. In contrast, chromosome A9 was with 11 *BnTAR* genes. The *BnTAR* gene arrangement is also relatively concentrated on different chromosomes and some of the *BnTAR* genes were

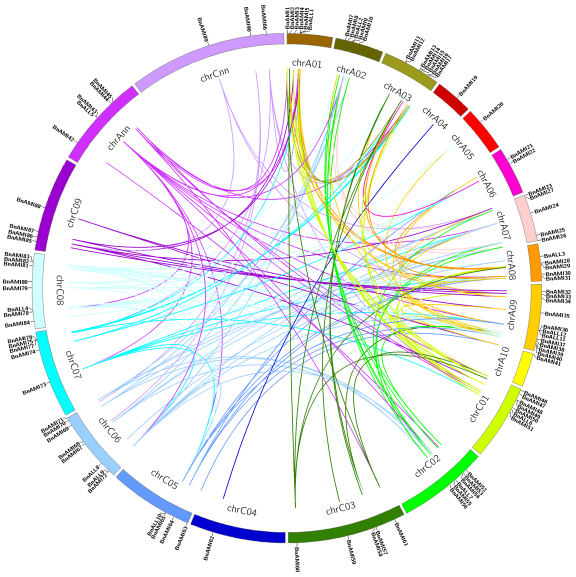


FIGURE 2
Distribution of *BnTAR* gene family among *B. napus* genome. The chromosomal information of the *BnTAR* genes was mapped to the *B. napus* chromosomes and the synteny relationship was highlighted in different colors.

existed as gene clusters, such as *BnAMI4* and *BnAMI5* on A1, *BnAMI48* and *BnAMI49* on C1 (Figure 2), which might be caused by the duplication of *BnTAR* genes during the evolutionary history.

To understand the expansion pattern of the *BnTAR* genes in *B. napus*, we investigated gene duplication events. Synteny analysis showed that the *BnTAR* genes are phylogenetically close to *TAR* genes in other three cruciferous species (*B. rapa*, *B. oleracea*, *Arabidopsis thaliana*). There are 36 *TAR* genes in *A. thaliana*, *AT1G23310* was contained eight copies, it was formed by fragment duplication. In addition, *AT1G17290*, *AT1G70580*, *AT1G72330* and *AT1G80360* were all have six copies, it was involved in replication formation. In the synteny analysis of *B. rapa*, *B. oleracea* and *B. napus*, 54 *BrTAR* genes were found in the genome of *B. rapa*, and it showed that at most 7 *BnTAR* genes were corresponding to the same *BrTAR* gene (such as BraA03g055430), and at least 1 *BnTAR* gene corresponding to the same *BrTAR* gene (such as BraA02g021330). 58 *BoTAR* genes were found in the genome of *B. oleracea*, it was also revealed that at most 7 *BnTAR* genes were corresponding to the same *BoTAR* gene (such as BolC07g053890), and at least 1 *BnTAR* gene corresponding to the same *BrTAR* gene (BolC08g046990). Further analysis revealed that those 54 *BrTAR* genes were corresponding to 80 *BnTAR* genes, and 41 and 39 of these *BnTAR* genes are located in A genome and C genome of *B. napus*, respectively. As for 58 *BoTAR* genes, which corresponding to 89 *BnTAR* genes, and 42 and 47 of them were located in A genome and C genome of *B. napus*, respectively. However, 13 *BnTAR* genes were not corresponding to any *BrTAR* genes or *BoTAR* genes. These results indicated that the *BnTAR* gene was highly conservative in the evolutionary process, but it also produces unique *BnTAR* genes, such as *BnAMI5*. These results above demonstrated that segment duplication was played an important role in the evolution of these gene loci.

By comparing the gene distribution of *TAR* genes in the genomes of *B. rapa*, *B. oleracea*, *A. thaliana* and *B. napus*, we found that the codominance of *AMI* and *ALL* families are maintained, but some genes are duplicated or lost. In addition, the synteny maps of *TAR* genes in *A. thaliana*, *B. rapa*, and A subgenome of *B. napus*, and *TAR* genes in *A. thaliana*, *B. oleracea*, and C subgenome of *B. napus* were analyzed separately (Figure 3). The results showed that *TAR* genes were almost evenly distributed in the Arabidopsis genome. These *AtTAR* genes are synteny to *BrTAR* and *BoTAR* genes, it indicated that *AtTAR* genes were replicated and formed into *BrTAR* or *BoTAR*, which was the result of genomic rearrangement after whole-genome triplication (WGT) in Brassicaceae species. In addition, almost all the *BrTAR* or *BoTAR* genes were maintained a synteny relationship with *BnTAR*s.

There are three mechanisms of gene family expansion: tandem duplication, segmental duplication and whole-genome duplication (WGD) (Xu et al., 2012). The original diploid genomes of Arabidopsis are ancient polyploids, which was undergo massive chromosomal rearrangements in the evolution (Schmidt et al., 2001). *BnAMI13*/*BnAMI14* on A3 chromosome cluster and *BnAMI76*/*BnAMI77* on C7 chromosome cluster are tandem duplicated genes (Table S1). Meanwhile, *BnAMI59* is clustered together with *BnAMI13*/*BnAMI14* and *BnAMI18* together with *BnAMI76*/*BnAMI77* (Figure 1), respectively, it suggested that these genes might come from the same ancestor. Interestingly, the two tandemly duplicated genes, *BnAMI13* was clustered together with *BnAMI59* but not with *BnAMI14*, and *BnAMI77* was clustered together with *BnAMI18* but not with *BnAMI76* (Figures 1 and 4A). This might indicate that these genes were experienced tandem duplication and then whole-genome duplication during the evolutionary process, and then caused the independent genes being more closely related to one of the tandem duplicated genes. In addition, *BnAMI74* and *BnAMI75* might be associated with segmental duplication, as they show synteny relationships with *BnAMI15* and *BnAMI14*, respectively. Phylogenetic analysis also indicated that *BnAMI74*/*BnAMI15* and *BnAMI75*/*BnAMI77* are homologous gene pairs, and they are respectively located in A and C subgenomes (Figure 4A), which might come from two different ancestors. Importantly, *BnAMI13*, *BnAMI14*, *BnAMI15*, *BnAMI17* and *BnAMI18* are all on the A3 chromosome cluster, and *BnAMI74*, *BnAMI75*, *BnAMI76* and *BnAMI77* are all on the C7 chromosome cluster, which indicates

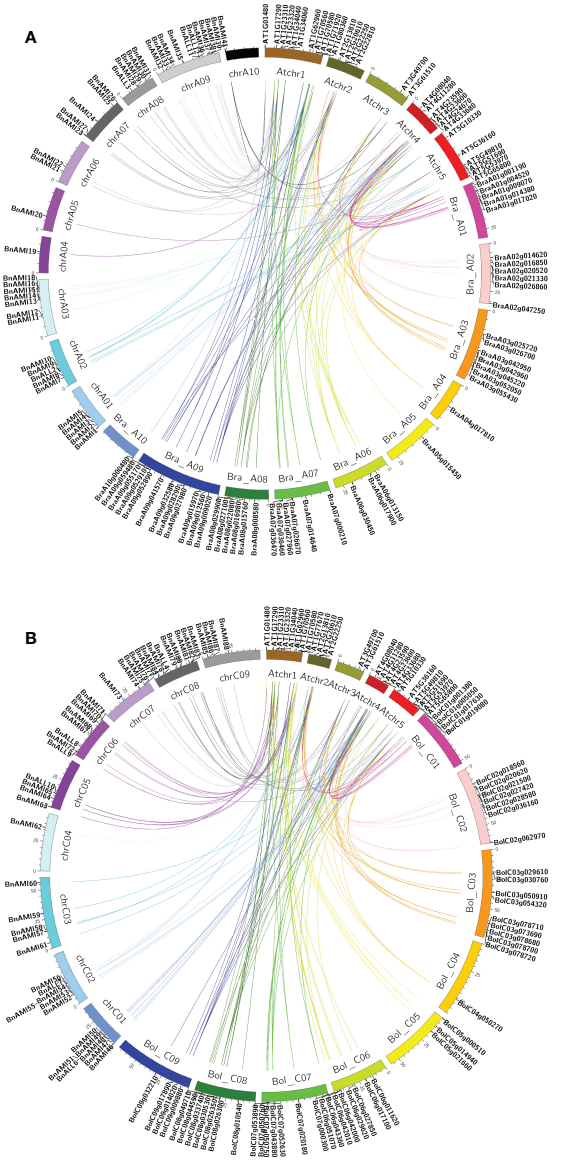


FIGURE 3
Synteny analysis map of *TAR* genes in *B. napus*, *B. rapa*, *B. oleracea*, and *A. thaliana* chromosomes. **(A)** The synteny of genes located on *B. napus* A subgenome, *B. rapa* genome and *A. thaliana* genome. **(B)** The synteny of genes located on *B. napus* C subgenome, *B. oleracea* genome and *A. thaliana* genome.

that the three duplication events play an important role in the formation of *BnTAR* gene clusters (Figure 4B).

Meanwhile, non-synonymous (Ka) and synonymous (Ks) values are used to analyze the selection pressure on *BnTAR* genes during the evolutionary process (Table S3). The results showed that most of *BnTAR* genes had the Ka/Ks ratio values between 0.1 and 1, it indicated that *BnTAR* has low selection pressure during evolution. We also found that there existed more than one Arabidopsis gene homologous to the same gene of *B. napus*. For instance, two Arabidopsis genes *AT1G70580* and *AT1G23310*, both homologous to *BnAMI78* of *B. napus*, and there are even three Arabidopsis genes, *AT1G01480*, *AT3G61510* and *AT4G11280* are all homologous to *BnAMI41* of *B. napus* (Table S3), this might be caused by synonymous substitution. Interestingly, four NaN values can be found in the Table S3, including of *AT3G61510*, *AT4G11280* and *BnAMI41* (Table S3). The NaN value appears because the pS is higher than 0.75, which means that there are more synonymous substitution sites resulting in higher sequence variability. These *TAR* genes themselves have high sequence similarity in *A. thaliana* and *B. napus*, most of the *TAR* family genes evolved neutrally, leading to a high rate of synonymous substitutions. Ultimately, this leads to a high level of synonymous substitution in *BnAMI41*, which resulted in its correspondence to multiple Arabidopsis duplicate genes. Therefore, there will be many Arabidopsis genes corresponding to the same *B. napus* gene together. These evolutionary relationships found through Ka/Ks ratio together with the homology homozygosity and chromosomal gene localization analyses above suggested that the WGD and segmental duplication play an important role in the evolution of the *TAR* gene family, which is consistent with the evolutionary process in Brassica species.

Gene structural analysis of *BnTAR*s

Most of *BnTAR* genes are less than 5 kb except for *BnALL10*, *BnAMI8* and *BnAMI54* (Table S1, Figure S2). It was revealed that the different branches were with different intron-exon structural features. Most of the *BnALL* genes were contained no more than five introns, while most of the *BnAMI* genes were with more than five introns (Figure S2). Particularly, the *BnAMI* genes in the fourth main branch were more distinct from other branches with more than ten introns (Figure S2). *BnAMI8* and *BnAMI54*, whose sequence length is larger than 6 kb, were found to have similar intron-exon structural features, it reflected their close phylogenetic relationship (Figure S2). And similarly, most of phylogenetical gene pairs have the similar intron-exon structures between each other (Figure S2).

Due to the low level of similarity of all *BnTAR* genes among the different branches, these genes were submitted to MEME for motif structural analysis, separately. The results showed that 9 conserved motifs were identified, of which the Motif 2 was observed in almost every branch (Figure 5). These conserved motifs were belonged to two Pfam domains, as shown in the Pfam codes (Pfam PF00155 and Pfam PF04864) and Web logo. In each phylogenetic branch, most of the gene pairs or genes in close branches have the similar motif composition. Sequence length of *BnALL5* is below 600bp, but it still has two motifs (motif 1 and motif 7) (Figure 5), it suggested that the *BnALL5* might had undergone fragment deletion or insertion during evolution, but the conserved motifs are still preserved, which

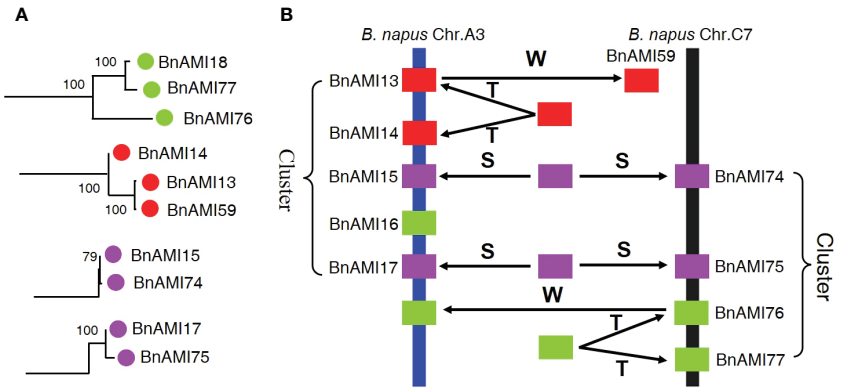


FIGURE 4
Phylogenetic relationships and hypothetical evolutionary progress of the cluster of *BnTAR* genes in *B. napus* chromosomes A3 and C7. **(A)** Phylogenetic relationships of selected *BnTAR* genes in the cluster. **(B)** Hypothetical mechanism of *BnTAR* gene cluster formation. The letters T, S, and W indicate putative tandem duplication, segmental duplication, and whole-genome duplication, respectively.

might be essential to its function. The sequence lengths of *BnAMI8* and *BnAMI54* are particularly long, which were with 6 and 8 motifs, respectively (Figure 5). The other genes do not differ much in length, and most of them were contained more than five motifs, and in each branch, the types of motifs are roughly similar. These results demonstrated that the different *TAR* genes have different structural compositions, but are similar within one phylogenetic branch and the motifs encoding *TAR* structural domains are conserved.

Secondary and tertiary structural analysis of *BnTAR*

The protein secondary structure of *BnTAR* was predicted by using the GOR4 and PSIPRED. It was revealed that the *BnTAR* proteins were mainly contained α-helixes, β-folds and random coils. Further analysis showed that the α-helixes, β-folds and random coils was account for about 35%, 20% and 45%, respectively (Figure S3). In addition, 3D structures of *BnTAR* proteins were also modeled and predicted by using the Phyre2. Firstly, we use VAST to analyze the protein sequences of all *BnTAR*, and then compared them with the structures that obtained from the PDB database. The predicted structural domains were displayed in the Cn3D macromolecular structure viewer (α-helix in green, β-fold in yellow and random coil in blue) (Figures 6A, C, E, G). And the structures of Alliinase from Allium Sativum (garlic) (2HOX) (Figures 6A, B) and the Crystal Structure of Alanine Aminotransferase from Hordeum vulgare (3TCM) (Figures 6C, D) were selected for homology modeling, as their structures have been fully studied. In the three-dimensional structure, the *BnTAR* proteins all appeared to be overlapping bowl-shaped, with a hollow in the center of the protein (Figures 6B, D, F, H), which is the active center of the enzyme and contains the pyridoxal binding site. The pyridoxal binding sites of *BnTAR*, 2HOX and 3TCM are all contain 10 amino acids, among which the first three amino acids are continuous amino acids, the seventh and eighth amino acids are separated by a random amino acid, the eighth and ninth amino acids are continuous and there are seven amino acids between the ninth and tenth amino acid (Table S4). In addition, the ninth amino acid Lys which is very conserved is also the catalytic residue (Table S4). These ten amino acids are formed into the conserved hole structure of proteins. At the same time, the difference of amino acid sequence in the protein also leads to a slight difference in the distance from each amino acid to the catalytic residue Lys in the 3D structure (Table S5). The distance from the pyridoxal binding site amino acid to the catalytic residue Lys in *BnALL* and 2HOX is farther. For example, the distance from fifth amino acid Asn to Lys of *BnALL10* and 2HOX are 6.1 nm and 6 nm, respectively, otherwise, the *BnAMI78* and 3TCM are 4.4 nm (Table S5). These results suggest that *BnTAR* proteins have a conserved structure, but also show some differences in the three-dimensional structural state and the number of secondary structures, which may help them to perform different functions in different environments.

Expression profiles analysis of *BnTAR* genes in different tissues

In order to fully understand the expression pattern of the *BnTARs*, the expression of *BnTAR* genes were performed by using the data in the BnTIR database (Figure S4). We found that about a third of the *BnTAR* genes were barely expressed in *B. napus*, but there were some genes (such as *BnAMI78*, *BnAMI36*, *BnAMI65*, *BnAMI23*, *BnAMI27*, *BnALL2*) that were highly expressed in specific tissues (Figure S4). It was revealed that a

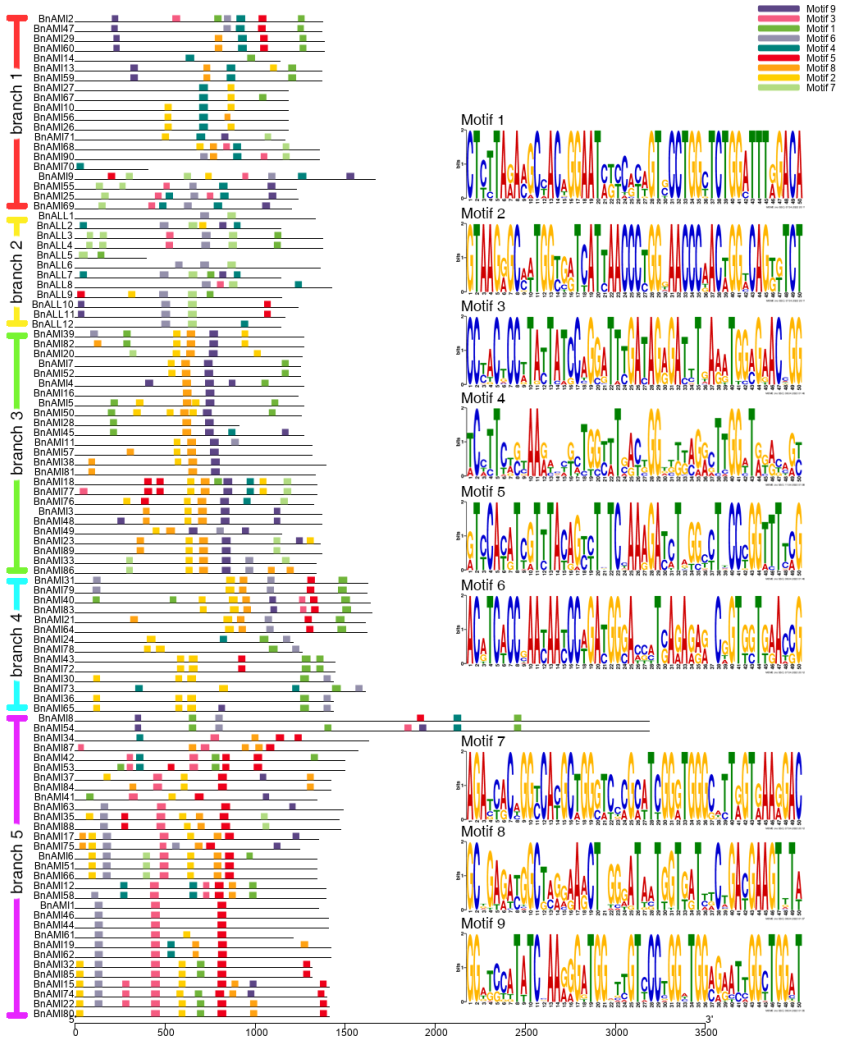


FIGURE 5
Motif patterns of different *BnTAR* branches and predicted consensus motifs in each *BnTAR* gene branch. Different motifs are shown in different colors. The lengths of the proteins and motifs can be estimated using the scale at the bottom. The motifs in Pfam of each branch are shown on the right.

part of *BnTAR* genes expressed higher in more actively growing tissues like root, stem and leaf, such as *BnAMI78*, *BnAMI36*, *BnAMI65* and *BnAMI23* (Figure S4). Another part of *BnTAR* genes were showed higher expression levels in early developed seeds, such as *BnAMI27*, *BnAMI67*, *BnALL2* and *BnALL7* (Figure S4).

To validate the data obtained from the database, 19 *BnTAR* genes were selected for RT-qPCR analysis in order to investigate the expression pattern of *BnTAR* genes in different tissues of *B. napus* (Tables S6 and S7). The results indicated that the expression of different *BnTAR* genes in each tissue was different, and the expression pattern of each gene also varies in different tissues (Figure 7). It was showed that the expression of *BnTAR* genes could be roughly divided into two main patterns: one is mainly expressed in nutritional organs, such as roots and stems (represented by *BnAMI71* and *BnAMI53*), while the other is mainly expressed in reproductive organs, such as flowers and seeds (characterized by *BnAMI37* and *BnAMI14*). This is consistent with the expression pattern available in the BnTIR database.

Relevance analysis of *BnTAR* genes and seed oil content and seed weight

Based on the transcriptome data of high and low oil content, and the transcriptome data of different seed weights in *B. napus*, we analyzed the contribution of the *BnTAR* genes to various

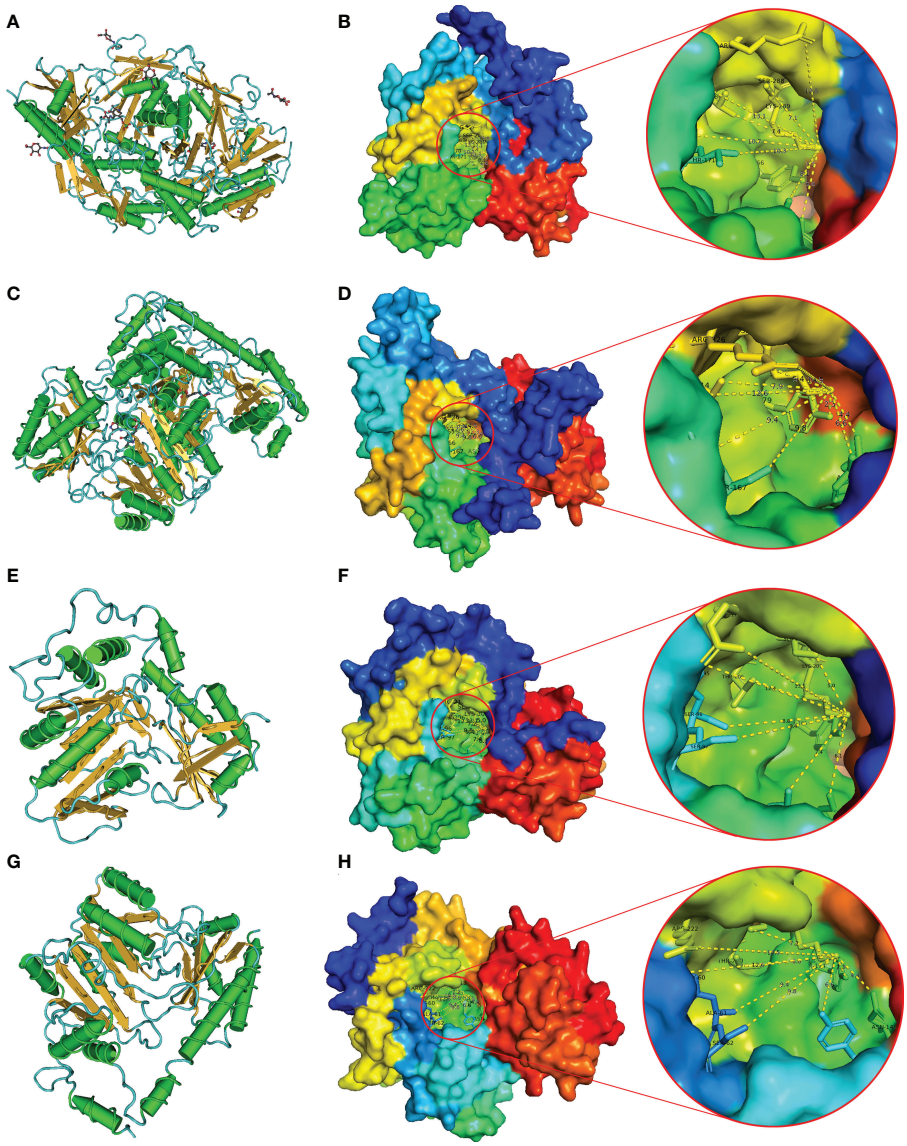


FIGURE 6
Predicted 3D structure of *BnTAR* genes. 2HOX and 3TCM were selected as examples. The models were predicted by Phyre2. Conserved domain analysis was highlighted using VAST. Red circles indicate the active center of the enzyme. **(A)** Predictive 3D structure of 2HOX in cylinder and plate mood. **(B)** The predicted 3D structure of 2HOX showed in the surface mood. **(C)** Predictive 3D structure of 3TCM in cylinder and plate mood. **(D)** The predicted 3D structure of 3TCM showed in the surface mood. **(E)** Predictive 3D structure of *BnALL10* in cylinder and plate mood **(F)** The predicted 3D structure of *BnALL10* showed in the surface mood. **(G)** Predictive 3D structure of *BnAMI78* in cylinder and plate mood. **(H)** The predicted 3D structure of *BnAMI78* showed in the surface mood.

agricultural traits. It was revealed that the expression of most *BnTAR* genes between samples has no obviously differences, and even many genes were barely expressed (Figure 8). However, four genes, *BnALL2*, *BnALL7*, *BnAMI27* and *BnAMI67*, were with higher expression, especially in the later seed development stages (36-42 days after flower). It also showed that the expression level in the *B. napus* with high oil content was higher than that of *B. napus* with the low oil content (Figure 8A). Further analysis of the expression of these four genes was also significantly different in *B. napus* with different seed weight. It was revealed that the expression was up-regulated at mid-seed development stage (20-26 days after flower) and down-regulated at late development stage (32 days after flower) in relatively large seed *B. napus* (Figure 8B). Also, some genes with a certain amount of expression were also found to be differentially expressed in different samples, for example, *BnAMI13* and *BnALL4* have differential expression in high and low oil samples, *BnAMI64* and *BnALL4* have differential

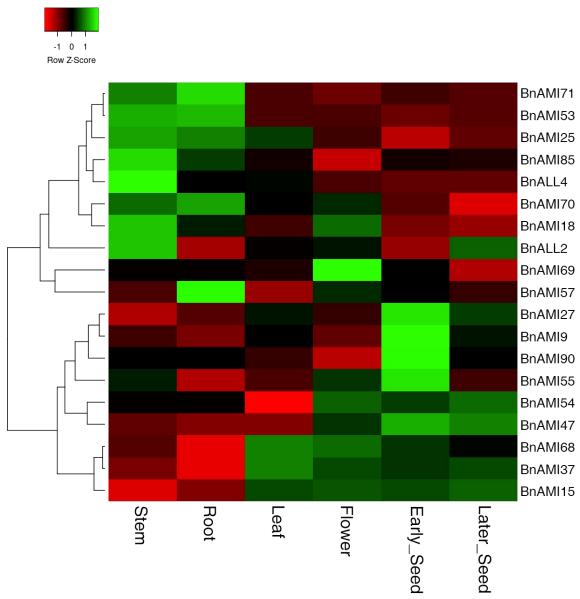


FIGURE 7
Hierarchical clustering of the expression profiles of *BnTAR* genes in different tissues. The color scale represents relative expression levels with high (green) or low (red). Early-developing seeds were got 14 days after flowering, and later-developing seeds were 27 days after flowering.

expression in different seed weight samples (Figure 8). These results suggest that most of the *BnTAR* genes might not contribute directly to seed oil or weight, but there are a number of genes (such as *BnALL2*, *BnALL7*, *BnAMI27* and *BnAMI67*) are involved in influencing seed oil content and seed weight.

In order further analysis the relationship between these *BnTAR* genes with seed oil content and seed weight, the positions of 102 *BnTAR* genes were analyzed in order to make clear whether they were located in the quantitative trait locus (QTL) for seed oil content (Yan et al., 2022) and seed weight (Raboanatahiry et al., 2018; Zhao et al., 2019) that obtained in our group or not. It was revealed that 21 and 12 *BnTAR* genes were found to be located within the QTL interval of seed oil content and seed weight, respectively (Table 2). Among them, the *BnAMI10* gene was located in both the *cqOC-A2-3* interval for oil content and the *cqSW-A2-2* interval for seed weight, which contribute to low oil content and small seed weight. It indicated that this gene might have the function for both seed weight and oil content. Also, many genes were found to be located in QTL intervals that obtained in multiple environments, such as interval of *cqOC-A9-16* where the *BnAMI40* gene is located, the phenotypic value to oil content was found to be significant with a maximum of 12.47% and a minimum of 9.84% and contribute to high oil content in multiple environments. Meanwhile, for seed weight, the interval of *cqSW-C6-4* and *cqSW-C9-6*, where *BnAMI69* and *BnAMI86*/*BnAMI87* are located, both with the phenotypic value more than 8% to seed weight. The *cqSW-C6-4* interval has an important contribution to seed weight enlargement, while *cqSW-C9-6* contributes to seed weight reduction. At the same time, during the comparison with the transcriptome data, we found that some genes that showed significant differential expression in the transcriptome (e.g., *BnALL2*, *BnALL3*, *BnAMI10*, *BnAMI67*, etc.) were also located in the corresponding QTL intervals, which corroborates our results for the transcriptome analysis (Figure 8). These results indicated that many of *BnTAR* genes were more or less involved in oil formation and seed development in *B. napus*.

Discussion

Structural characteristics of the *BnTAR* family

The *TAA1/TAR* gene family is a class of pyridoxal-5'-phosphate-dependent amino acid transaminases, which can catalyze the conversion of tryptophan into Indole-3-Pyruvate (Zhao, 2012). *TAA1/TAR* gene family has been found in many crops, such as rice and maize (Phillips et al., 2011; Kakei et al., 2017). However, the genome wide identification, expression pattern and its relationship with oil content and seed weight of the *TAA1/TAR* genes have not been reported in *B. napus*. In this study, 102 *TAA1/TAR* family genes were identified in *B. napus* by using the Darmor as reference genome (Chalhoub et al., 2014). At the same time, we also searched the corresponding *BnTAR* genes in ZS11 genome of *B. napus* (Liu et al., 2021), and 130 genes were found. However, some difference was found between the two genomes cannot be matched. For example, four *BnTAR* genes that corresponding to AT5G53970 were found in Darmor genome, while only two *BnTAR* genes were found in ZS11 genome. On the contrary, two *BnTAR* genes that corresponding to AT5G53970 were found in Darmor genome. Those differences might be due to the different genetic backgrounds of Darmor and ZS11.

Introns increase the length of transcripts and are prone to detrimental effects on gene expression during their splicing (Jeffares et al., 2008). Meanwhile, genes that respond to stress are contained fewer introns (Lan et al., 2013). In this study, *BnTAR* genes were found to have a large and variable number of introns, with only nine genes having fewer than three and a maximum of 14 (Figure S2). Perhaps it is because *TAR*s genes, as an important enzyme in the auxin synthesis pathway that is the basis of plant life, maintain high expression and have more introns number throughout the whole plant life cycle. Meanwhile, *BnTAR* genes are very long (all but *BnALL5* are >1000 bp), and genes length are two or three times longer than that of CDS length.

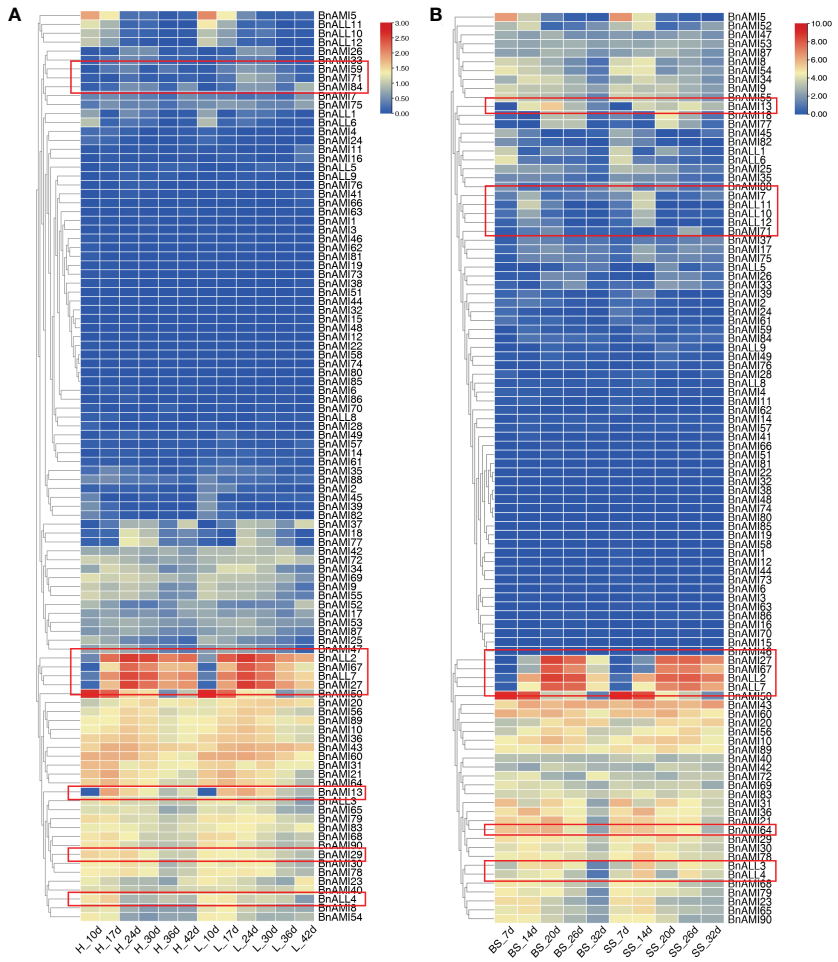


FIGURE 8
Expression of *BnTAR* gene in transcriptome data. The color scale represents relative expression levels with high (red) or low (blue). Red boxes indicate the genes that are differentially expressed. **(A)** The transcriptome data of high oil samples and low oil samples. 'H' means high oil samples while 'L' means low oil samples. **(B)** The transcriptome data of different seed weight simples. 'BS' means big seed simples and 'SS' means small seed simples. The letter 'd' all means how many days after flowering.

Motifs are evolutionarily conserved amino acid/nucleotide sequence regions that usually play an important role in structure and function (Semwal et al., 2022). Many amino acid transaminase genes have been found in Arabidopsis genome, and most of these amino acid transaminases contain similar motifs, such as Pfam aminotransferase I/II signature (Pfam 00155), Pfam alliinase signature (Pfam 04864), Pfam aminotransferase III signature (Pfam 00202), etc. (Liepman and Olsen, 2004). The type and number of motifs varied in each branch, but the conserved motifs of Pfam 00155 or Pfam 04864 could be observed in each member of the *BnTAR* gene family. The differences in the motifs between *BnALL*s and *BnAMI*s were likely due to their configurations in the functional form. The three-dimensional structural analysis showed that the *BnTAR* proteins were all appeared to be overlapping bowl-shaped with a hollow in the center of the protein (Figures 6B, D, F, H), which contains the pyridoxal binding site, moreover, the pyridoxal binding site of the *BnTAR* proteins contain ten amino acids, and the ninth binding site is very conserved as arginine, which is also a catalytic residue. In addition, *TAR*s proteins complete the amino acid transaminase function in dimer form, which could also be found in the aminotransaminase family in other species (Liepman et al., 2019).

Duplication patterns and synteny analysis of the *BnTAR* family

Gene duplication not only expands genome content but also diversifies gene function to ensure optimal adapt-ability and evolution of plants (Xu et al., 2012). Tandem duplication,

TABLE 2 Analysis of the QTL interval of the *BnTAR* genes.

QTL for oil content (OC)				QTL for thousand seed weight (TSW)			
Name	Consensus QTL*	LOD	R^2	Name	Consensus QTL**	LOD	R^2
BnALL1	*cqOC-A1-1*	3.00328-4.42897	3.3939-9.5762	*BnALL6*	*cqSW-C1-2*	3.90817	5.4284
BnALL2	*cqOC-A2-2*	7.62538	14.0549	*BnAMI10*	*cqSW-A2-2*	2.5172	4.574
BnALL3	*cqOC-A8-2*	2.95337-5.55303	3.473-7.3556	*BnAMI23*	*cqSW-A7-1*	2.55409	3.0467
BnALL8	*cqOC-C6-1*	2.46946	2.2607	*BnAMI34*	*cqSW-A9-4*	3.69334-5.23404	4.3782-5.5929
BnALL10	*cqOC-C5-4*	3.07706	4.1087	*BnAMI50*	*cqSW-C1-5*	4.0796	5.1477
BnAMI6	*cqOC-A1-1*	3.00328-4.42897	3.3939-9.5762	*BnAMI51*	*cqSW-C1-3*	3.57399-3.90817	5.2212-5.705
BnAMI7	*cqOC-A2-1*	2.95988	3.2563	*BnAMI69*	*cqSW-C6-4*	7.51905	8.9761
BnAMI9	*cqOC-A2-3*	3.03583-7.8337	3.5296-16.3876	*BnAMI70*	*cqSW-C6-7*	5.36207	7.3724
BnAMI10	*cqOC-A2-3*	3.03583-7.8337	3.5296-16.3876	*BnAMI85*	*cqSW-C9-2*	4.98232	6.2849
BnAMI13	*cqOC-A3-8*	2.98809-3.30057	2.8245-3.7503	*BnAMI86*	*cqSW-C9-6*	5.15809	8.7772
BnAMI14	*cqOC-A3-8*	2.98809-3.30057	2.8245-3.7503	*BnAMI87*	*cqSW-C9-6*	5.15809	8.7772
BnAMI28	*cqOC-A8-3*	5.29263-7.03948	6.1033-7.176	*BnAMI88*	*cqSW-C9-8*	3.53059	5.3449
BnAMI29	*cqOC-A8-3*	5.29263-7.03948	6.1033-7.176				
BnAMI32	*cqOC-A9-5*	2.82534-7.52773	3.6503-9.0369				
BnAMI33	*cqOC-A9-7*	6.45358	6.5969				
BnAMI35	*cqOC-A9-9*	3.76929	3.3955				
BnAMI40	*cqOC-A9-16*	8.03551-12.88112	9.8381-13.4651				
BnAMI64	*cqOC-C5-3*	3.66296	3.3156				
BnAMI65	*cqOC-C5-4*	3.07706	4.1087				
BnAMI67	*cqOC-C6-2*	2.61268-5.22536	2.7008-5.5518				
BnAMI68	*cqOC-C6-3*	2.76241	5.7748				

LOD indicates the significance of the QTL interval and R^2 indicates the contribution of the interval to the trait. Consensus QTL* were obtained from Wang et al. research (2013) and Consensus QTL** were obtained from Zhao et al. research (2016).

segmental duplication and whole-genome duplication (WGD) were the three mechanisms that could contribute to the gene family expansion (Xu et al., 2012). WGDs play a major role in Brassicas, particularly the mesopolyploidization events, which are simultaneously accompanied by extensive chromosomal and genetic diploidization processes (Hohmann et al., 2015). Many studies shown that Brassicaceae have experienced WGD events during their evolution (Rana et al., 2004; Parkin et al., 2005), nevertheless, little is known about the evolution and origin of the *TAA1/TAR* family in plants. *B. napus* was an allopolyploidy formed by the hybridization of *B. rapa* and *B. oleracea* in 7500 years ago (Chalhoub et al., 2014). In this study, it was revealed that the *BnTAR* family formation was associated with segmental duplications and WGD. The Arabidopsis genome contains 36 *TAR*s genes; therefore, more than 108 *TAR*s genes would be expected to produce through one WGT event in the *B. oleracea* or *B. rapa* genome, and finally leading to much more *TAR*s genes in *B. napus*. However, only 102 *TAR*s genes were observed in the *B. napus* genome in the present study. These results indicated that more than 50% of the duplicated *TAR*s genes were lost after WGT, which might be due to the extensive chromosome reshuffling during the diploidization after WGT (Cheng et al., 2014). It was found that some duplicated *TAR*s genes disappeared in the *B. napus* genome. For instance, certain genes that are homologous to Arabidopsis (*AT1G34060*, *AT1G01480*, *AT1G34040*, *AT2G22810*, *AT2G24850*, *AT2G24850*, *AT4G08040*, *AT5G49810*, *AT5G51690*, *AT1G23320*, *AT1G62960*, *AT1G77670*, *AT2G13810*, *AT3G61510*, *AT4G11280*, *AT4G24670*, *AT4G28410*, *AT5G36160*) had less than four copies in *B. napus* (Table S3), by simply calculating the Arabidopsis genes at three times after the WGD event, these Arabidopsis genes have thirty fewer homologous genes in *B. napus*, which indicates the thirty genes that were not detected might have been lost. Similar losses after WGT also can been observed in other gene families of Brassica (Yu et al., 2014).

Segmental duplication also plays an important role in the expansion of *BnTAR* family, and 71 *BnTAR* genes were determined to have one or two gene pairs in the corresponding regions. Therefore, the formation of 70% (71/102) of the *BnTAR* genes is related to segmental duplication. Synteny analysis demonstrated that most *BnTAR* gene family members are located in well-conserved synteny regions, and some genes were deleted or gained. Similarly, synteny genomic regions with some deleted genes have been identified in other gene families (Yu et al., 2014). Despite the loss of some genes after WGT, the present findings suggested that segmental replication and WGD might play an important role in the expansion of the *TAR*s family in *B. napus*, while tandem replication only played a minor role. The characteristics of the *BnTAR* family duplication patterns and synteny analysis were consistent with the Brassicaceae evolutionary history.

The origin and evolution of auxin biosynthetic genes in plants have been controversial for a long time (Wang et al., 2014; Yue et al., 2014). However, the recent studies have found that auxin biosynthesis was mainly originated in charophycean green algae, which was most closely related to terrestrial plants (Romani, 2017). The present results of the phylogenetic, evolutionary, and structural analysis show that the *BnTAR*s have a close phylogenetic relationship and similar structure (Figure 4, Figure S2, Table S3). As discussed above, *BnTAR* family genes undergo WGD and segmental replication, which was also observed in *Oryza sativa* and *Triticum aestivum* (Shao et al., 2017; Guo et al., 2020). *BnAMI8* and *BnAMI54* that belonged to a relatively unique phylogenetic branch with a different structure from other *BnTAR*, which may be caused by some additional gene replication and fusion events. Such events may produce more diverse monomers that combine to form more diverse amino transaminases.

The relationship between *BnTAR* genes and seed oil content and seed weight in *B. napus*

The expression patterns of *BnTAR* family genes are different in different tissues, it was found that about a third of the *BnTAR* genes were barely expressed in *B. napus*, further analysis showed that the expression of *BnTAR* genes were with two patterns, one was mainly expressed in vegetative organs (such as roots, stems and leaves), the other was mainly expressed in reproductive organs (such as flowers and seeds) (Figure 7, Figure S4). The similar *TAA1*/*TAR* genes expression patterns were also found in wheat and *A. thaliana* (Poulet and Kriechbaumer, 2017; Shao et al., 2017).

In recent studies, more attention has been paid to the effects of these genes on the formation of lateral roots (He et al., 2011; Zhou et al., 2011; Ma et al., 2014), while the effect of *TAR* gene on the yield of economic crops was ignored, and the role of the *TAR* gene in yield increase is only studied in wheat (Shao et al., 2017). By compared the transcriptomic results of high and low oil content samples, and the relatively high and low seed weight samples, a preliminary study on the effect of *BnTAR* genes on seed traits in *B. napus* was conducted. We found that the expression of most *BnTAR* genes was not significantly different between high and low oil samples and the relatively high and low seed weight samples, but there were still several genes, such as *BnALL2*, *BnALL7*, *BnAMI27* and *BnAMI67*, which were higher in the late stage of seed development in high oil samples, higher in the middle stage of seed development and lower in the late stage of seed development in high seed weight samples (Figure 8), it suggested their effects on seed oil content and weight regulation.

QTL mapping is an effective tool for the analysis of the genetic mechanism of complex quantitative traits (Mauricio, 2001), and a number of QTLs for oil content have been detected in diverse crops, for instance, *Zea mays* (Mangolin et al., 2004; Zhang J. et al., 2008; Yang et al., 2010), *Glycine max* (Orf et al., 1999; Kabelka et al., 2004; Zhang et al., 2004), *Arachis hypogaea* (Sarvamangala et al., 2011; Pandey et al., 2014) and *B. napus* (Gu]et al., 2016; Chao et al., 2022). On the basis of the results of the previous quantitative trait locus (QTL) results for oil content and seed weight in our group, we found that the 21 and 12 *BnTAR* genes were located in the QTL interval of oil content and seed weight, respectively (Table 2). And three major QTLs (*cqOC-A9-9*, *cqOC-C5-3* and *cqOC-C5-4*), where *BnAMI35*, *BnAMI64* and *BnAMI65* located were showed a large effect (phenotypic variation>10%) in at least two trials (Chao et al., 2017). Also, some *BnTAR* genes, for example, *BnALL2*, *BnALL3*, *BnAMI10*, and *BnAMI67*, had not only been mapped in QTL intervals, but also affected seed oil content in previous transcriptome analysis. These results indicated that some of the *BnTAR* genes play an important role in regulating the seed weight and seed oil formation in *B. napus*.

Conclusion

In this study, 102 *BnTAR* genes were identified and comprehensively analyzed. A conserved catalytic structural domain of the *TAR* superfamily was observed, and WGD and segmental replication were the main events for the *BnTAR* gene family formation. Expression patterns combined with QTL analysis provided the new insights into the biological functions of *BnTAR* proteins for oil content and seed weight.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

XC and XL did the identification of *TAR* family genes in *B. napus*, duplication pattern analysis, and drafted the manuscript. XL did multiple sequence alignments of the *BnTAR* genes and conducted the phylogenetic analysis. XC, XL, JH, and MT did the RT-qPCR experiment. XC, XL, and HL analyzed QTL and transcriptomic data. XC and XL collected *TAR* genes information of other species, and did the structural analysis and prepared Figures 1–8 and the Supplementary Information. ML designed and supervised the experiment. ML and JH revised the manuscript. All authors reviewed the manuscript.

Funding

This study is supported by the National Natural Science Foundation of China (32272067).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1098820/full#supplementary-material

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/s0022-2836(05)80360-2

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Bjellqvist, B., Basse, B., Olsen, E., and Celis, J. E. (1994). Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* 15, 529–539. doi: 10.1002/elps.1150150171

Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., et al. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 14, 1023–1031. doi: 10.1002/elps.11501401163

Blum, T., Briesemeister, S., and Kohlbacher, O. (2009). MultiLoc2: integrating phylogeny and gene ontology terms improves subcellular protein localization prediction. *BMC Bioinf.* 10, 274. doi: 10.1186/1471-2105-10-274

Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013). Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* 41, W349–W357. doi: 10.1093/nar/gkt381

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., et al. (2014). Plant genetics. early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chang, W., Li, Y., Qu, Y., Liu, Y., Zhang, G., Zhao, Y., et al. (2022). Mixotrophic cultivation of microalgae to enhance the biomass and lipid production with synergistic effect of red light and phytohormone IAA. *Renewable Energy* 187, 819–828. doi: 10.1016/j.renene.2022.01.108

Chao, H., Guo, L., Zhao, W., Li, H., and Li, M. (2022). A major yellow-seed QTL on chromosome A09 significantly increases the oil content and reduces the fiber content of seed in *Brassica napus*. *Theor. Appl. Genet.* 135, 1293–1305. doi: 10.1007/s00122-022-04031-0

Chao, H., Wang, H., Wang, X., Guo, L., Gu, J., Zhao, W., et al. (2017). Genetic dissection of seed oil and protein content and identification of networks associated with oil content in *Brassica napus*. *Sci. Rep.* 7, 46295. doi: 10.1038/srep46295

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of brassica plants. *Horticult. Res.* 1, 14024. doi: 10.1038/hortres.2014.24

Chen, H., Wang, T., He, X., Cai, X., Lin, R., Liang, J., et al. (2022). BRAD V3.0: an upgraded brassicaceae database. *Nucleic Acids Res.* 50, D1432–D1441. doi: 10.1093/nar/gkab1057

Chiu, J. C., Brenner, E. D., Desalle, R., Nitabach, M. N., Holmes, T. C., and Coruzzi, G. M. (2002). Phylogenetic and expression analysis of the glutamate-receptor-like gene family in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 19, 1066–1082. doi: 10.1093/oxfordjournals.molbev.a004165

Chou, K. C., and Shen, H. B. (2010). Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PloS One* 5, e11335. doi: 10.1371/journal.pone.0011335

Depuydt, T., and Vandepoele, K. (2021). Multi-omics network-based functional annotation of unknown arabidopsis genes. *Plant J.* 108, 1193–1212. doi: 10.1111/tpj.15507

Dhindsa, R. S., Amaral, A. C., and Cleland, R. E. (1984). Rapid reduction by IAA of malondialdehyde levels in avena coleoptiles, a possible effect on lipid peroxidation. *Biochem. Biophys. Res. Commun.* 125, 76–81. doi: 10.1016/s0006-291x(84)80336-8

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Gibrat, J. F., Madej, T., and Bryant, S. H. (1996). Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* 6, 377–385. doi: 10.1016/s0959-440x(96)80058-3

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

Gu, J., Chao, H., Wang, H., Li, Y., Li, D., Xiang, J., et al. (2016). Identification of the relationship between oil body morphology and oil content by microstructure comparison combining with QTL analysis in *Brassica napus*. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01989

Guo, T., Chen, K., Dong, N. Q., Ye, W. W., Shan, J. X., and Lin, H. X. (2020). Tillering and small grain 1 dominates the tryptophan aminotransferase family required for local auxin biosynthesis in rice. *J. Integr. Plant Biol.* 62, 581–600. doi: 10.1111/jipb.12820

He, W., Brumos, J., Li, H., Ji, Y., Ke, M., Gong, X., et al. (2011). A small-molecule screen identifies l-kynurenine as a competitive inhibitor of TAA1/TAR activity in

ethylene-directed auxin biosynthesis and root growth in arabidopsis. *Plant Cell* 23, 3944–3960. doi: 10.1105/tpc.111.089029

Higgins, D. G., and Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73, 237–244. doi: 10.1016/0378-1119(88)90330-7

Hohmann, N., Wolf, E. M., Lysak, M. A., and Koch, M. A. (2015). A time-calibrated road map of brassicaceae species radiation and evolutionary history. *Plant Cell* 27, 2770–2784. doi: 10.1105/tpc.15.00482

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly regulated genes are intron poor. *Trends Genet.* 24, 375–378. doi: 10.1016/j.tig.2008.05.006

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. doi: 10.1006/jmbi.1999.3091

Kabelka, E. A., Diers, B. W., Fehr, W. R., Leroy, A. R., Baianu, I. C., You, T., et al. (2004). Putative alleles for increased yield from soybean plant introductions. *Crop Sci.* 44, 784–791. doi: 10.2135/cropsci2004.7840

Kakei, Y., Nakamura, A., Yamamoto, M., Ishida, Y., Yamazaki, C., Sato, A., et al. (2017). Biochemical and chemical biology study of rice OsTAR1 revealed that tryptophan aminotransferase is involved in auxin biosynthesis: Identification of a potent OsTAR1 inhibitor, Pyruvamine2031. *Plant Cell Physiol.* 58, 598–606. doi: 10.1093/pcp/pcx007

Kelley, L. A., and Sternberg, M. J. (2009). Protein structure prediction on the web: a case study using the phyre server. *Nat. Protoc.* 4, 363–371. doi: 10.1038/nprot.2009.2

Kim, J., Woo, H. R., and Nam, H. G. (2016). Toward systems understanding of leaf senescence: An integrated multi-omics perspective on leaf senescence research. *Mol. Plant* 9, 813–825. doi: 10.1016/j.molp.2016.04.017

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132. doi: 10.1016/0022-2836(82)90515-0

Lan, T., Gao, J., and Zeng, Q. Y. (2013). Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in *Populus trichocarpa*. *Tree Genet. Genomes* 9, 253–264. doi: 10.1007/s11295-012-0551-2

Le Deunff, E., Beauclair, P., Deleu, C., and Lecourt, J. (2019). Inhibition of aminotransferases by aminoethoxyvinylglycine triggers a nitrogen limitation condition and deregulation of histidine homeostasis that impact root and shoot development and nitrate uptake. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01387

Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J., et al. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* 32, D142–D144. doi: 10.1093/nar/gkh088

Liang, Y., Wan, N., Cheng, Z., Mo, Y., Liu, B., Liu, H., et al. (2017). Whole-genome identification and expression pattern of the vicinal oxygen chelate family in rapeseed (*Brassica napus* l.). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00745

Liepman, A. H., and Olsen, L. J. (2004). Genomic analysis of aminotransferases in *Arabidopsis thaliana*. *Crit. Rev. Plant Sci.* 23, 73–89. doi: 10.1080/07352680490273419

Liepman, A. H., Vijayalakshmi, J., Peisach, D., Hulsebus, B., Olsen, L. J., and Saper, M. A. (2019). Crystal structure of photorespiratory Alanine:Glyoxylate aminotransferase 1 (AGT1) from *Arabidopsis thaliana*. *Front. Plant Sci.* 10, 1229. doi: 10.3389/fpls.2019.01229

Li, W., Liu, B., Yu, L., Feng, D., Wang, H., and Wang, J. (2009). Phylogenetic analysis, structural evolution and functional divergence of the 12-oxo-phytodienoate acid reductase gene family in plants. *BMC Evol. Biol.* 9, 90. doi: 10.1186/1471-2148-9-90

Liu, D., Yu, L., Wei, L., Yu, P., Wang, J., Zhao, H., et al. (2021). BnTIR: an online transcriptome platform for exploring RNA-seq libraries for oil crop *Brassica napus*. *Plant Biotechnol. J.* 19, 1895–1897. doi: 10.1111/pbi.13665

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Ma, W., Li, J., Qu, B., He, X., Zhao, X., Li, B., et al. (2014). Auxin biosynthetic gene *TAR2* is involved in low nitrogen-mediated reprogramming of root architecture in arabidopsis. *Plant J.* 78, 70–79. doi: 10.1111/tpj.12448

Mangolin, C. A., De Souza, C. L., Garcia, A., Garcia, A. F., Sibov, S. T., and De Souza, A. P. (2004). Mapping QTLs for kernel oil content in a tropical maize population. *Euphytica* 137, 251–259. doi: 10.1023/B:EUPH.0000041588.95689.47

Mauricio, R. (2001). Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. *Nat. Rev. Genet.* 2, 370–381. doi: 10.1038/35072085

Orf, J. H., Chase, K., Jarvik, T., Mansur, L. M., Cregan, P. B., Adler, F. R., et al. (1999). Genetics of soybean agronomic traits: I. comparison of three related recombinant inbred populations. *Crop Sci.* 39, 1642–1651. doi: 10.2135/cropsci1999.3961642x

Ostergaard, L., and King, G. J. (2008). Standardized gene nomenclature for the brassica genus. *Plant Methods* 4, 10. doi: 10.1186/1746-4811-4-10

Pandey, M. K., Wang, M. L., Qiao, L., Feng, S., Khera, P., Wang, H., et al. (2014). Identification of QTLs associated with oil content and mapping *FAD2* genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* l.). *BMC Genet.* 15, 133. doi: 10.1186/s12863-014-0133-4

Paque, S., and Weijers, D. (2016). Q&A: Auxin: the plant molecule that influences almost anything. *BMC Biol.* 14, 67. doi: 10.1186/s12915-016-0291-0

Parkin, I. A., Gulden, S. M., Sharpe, A. G., Lukens, L., Trick, M., Osborn, T. C., et al. (2005). Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171, 765–781. doi: 10.1534/genetics.105.042093

Phillips, K. A., Skirpan, A. L., Liu, X., Christensen, A., Slewinski, T. L., Hudson, C., et al. (2011). Vanishing tassel2 encodes a grass-specific tryptophan aminotransferase required for vegetative and reproductive development in maize. *Plant Cell* 23, 550–566. doi: 10.1105/tpc.110.075267

Poulet, A., and Kriechbaumer, V. (2017). Bioinformatics analysis of phylogeny and transcription of *TAA/YUC* auxin biosynthetic genes. *Int. J. Mol. Sci.* 18, 1791. doi: 10.3390/ijms18081791

Raboanatahiry, N., Chao, H., Dalin, H., Pu, S., Yan, W., Yu, L., et al. (2018). QTL alignment for seed yield and yield related traits in *Brassica napus*. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01127

Rana, D., Van Den Boogaart, T., O'neill, C. M., Hynes, L., Bent, E., Macpherson, L., et al. (2004). Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. *Plant J.* 40, 725–733. doi: 10.1111/j.1365-313X.2004.02244.x

Romani, F. (2017). Origin of TAA genes in charophytes: New insights into the controversy over the origin of auxin biosynthesis. *Front. Plant Sci.* 8, 1616. doi: 10.3389/fpls.2017.01616

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Sarvamangala, C., Gowda, M. V. C., and Varshney, R. K. (2011). Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea l.*). *Field Crops Res.* 122, 49–59. doi: 10.1016/j.fcr.2011.02.010

Schmidt, R., Acarkan, A., and Boivin, K. (2001). Comparative structural genomics in the brassicaceae family. *Plant Physiol. Biochem.* 39, 253–262. doi: 10.1016/S0981-9428(01)01239-6

Semwal, R., Aier, I., Raj, U., and Varadwaj, P. K. (2022). Pr[m]: An algorithm for protein motif discovery. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 19, 585–592. doi: 10.1109/tcbb.2020.2999262

Shao, A., Ma, W., Zhao, X., Hu, M., He, X., Teng, W., et al. (2017). The auxin biosynthetic TRYPTOPHAN AMINOTRANSFERASE RELATED TaTAR2.1-3A increases grain yield of wheat. *Plant Physiol.* 174, 2274–2288. doi: 10.1104/pp.17.00094

Shiu, S. H., and Bleecker, A. B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in arabidopsis. *Plant Physiol.* 132, 530–543. doi: 10.1104/pp.103.021964

Stepanova, A. N., Hoyt, J. M., Hamilton, A. A., and Alonso, J. M. (2005).). a link between ethylene and auxin uncovered by the characterization of two root-specific ethylene-insensitive mutants in arabidopsis. *Plant Cell* 17, 2230–2242. doi: 10.1105/tpc.105.033365

Stepanova, A. N., Robertson-Hoyt, J., Yun, J., Benavente, L. M., Xie, D. Y., Dolezal, K., et al. (2008). TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* 133, 177–191. doi: 10.1016/j.cell.2008.01.047

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Toledo-Ortiz, G., Huq, E., and Quail, P. H. (2003). The arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* 15, 1749–1770. doi: 10.1105/tpc.013839

Wang, C., Liu, Y., Li, S. S., and Han, G. Z. (2014). Origin of plant auxin biosynthesis in *charophyte algae*. *Trends Plant Sci.* 19, 741–743. doi: 10.1016/j.tplants.2014.10.004

Wang, X., Wang, H., Long, Y., Li, D., Yin, Y., Tian, J., et al. (2013). Identification of QTLs associated with oil content in a high-oil *Brassica napus* cultivar and construction of a high-density consensus map for QTLs comparison in *B. napus*. *PloS One* 8, e80569. doi: 10.1371/journal.pone.0080569

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919

Woodward, A. W., and Bartel, B. (2005). Auxin: Regulation, action, and interaction. *Ann. Bot.* 95, 707–735. doi: 10.1093/aob/mci083

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. U. States America* 109, 1187–1192. doi: 10.1073/pnas.1109047109

Yang, X., Guo, Y., Yan, J., Zhang, J., Song, T., Rocheford, T., et al. (2010). Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. *Theor. Appl. Genet.* 120, 665–678. doi: 10.1007/s00122-009-1184-1

Yan, S., Li, H., Chao, H., He, J., Ding, Y., Zhao, W., et al. (2022). Refinement of four major QTL for oil content in *Brassica napus* by integration of genome resequencing and transcriptomics. *Crop J.* 10, 627–637. doi: 10.1016/j.cj.2022.01.002

Yates, A. D., Allen, J., Amode, R. M., Azov, A. G., Barba, M., Becerra, A., et al. (2022). Ensembl genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res.* 50, D996–1003. doi: 10.1093/nar/gkab1007

Yue, J., Hu, X., and Huang, J. (2014). Origin of plant auxin biosynthesis. *Trends Plant Sci.* 19, 764–770. doi: 10.1016/j.tplants.2014.07.004

Yu, J., Tehrim, S., Zhang, F., Tong, C., Huang, J., Cheng, X., et al. (2014). Genome-wide comparative analysis of NBS-encoding genes between brassica species and *Arabidopsis thaliana*. *BMC Genomics* 15, 3. doi: 10.1186/1471-2164-15-3

Zhang, J., Lu, X. Q., Song, X. F., Yan, J. B., Song, T. M., Dai, J. R., et al. (2008). Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. *Euphytica* 162, 335–344. doi: 10.1007/s10681-007-9500-9

Zhang, W. K., Wang, Y. J., Luo, G. Z., Zhang, J. S., He, C. Y., Wu, X. L., et al. (2004). QTL mapping of ten agronomic traits on the soybean (*Glycine max* l. *Merr.*) genetic map and their association with EST markers. *Theor. Appl. Genet.* 108, 1131–1139. doi: 10.1007/s00122-003-1527-2

Zhang, R., Wang, B., Ouyang, J., Li, J., and Wang, Y. (2008). Arabidopsis indole synthase, a homolog of tryptophan synthase alpha, is an enzyme involved in the trp-independent indole-containing metabolite biosynthesis. *J. Integr. Plant Biol.* 50, 1070–1077. doi: 10.1111/j.1744-7909.2008.00729.x

Zhao, Y. (2012). Auxin biosynthesis: a simple two-step pathway converts tryptophan to indole-3-acetic acid in plants. *Mol. Plant* 5, 334–338. doi: 10.1093/mp/ssr104

Zhao, W., Wang, X., Wang, H., Tian, J., Li, B., Chen, L., et al. (2016). Genome-wide identification of QTL for seed yield and yield-related traits and construction of a high-density consensus map for QTL comparison in *Brassica napus*. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00017

Zhao, W., Zhang, L., Chao, H., Wang, H., Ta, N., Li, H., et al. (2019). Genome-wide identification of silique-related traits based on high-density genetic linkage map in *Brassica napus*. *Mol. Breed.* 39, 86. doi: 10.1007/s11032-019-0988-1

Zhou, Z. Y., Zhang, C. G., Wu, L., Zhang, C. G., Chai, J., Wang, M., et al. (2011). Functional characterization of the *CKRC1/TAA1* gene and dissection of hormonal actions in the arabidopsis root. *Plant J.* 66, 516–527. doi: 10.1111/j.1365-313X.2011.04509.x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fpls.2022.1056582

OPEN ACCESS

EDITED BY
Sangram K. Lenka,
Gujarat Biotechnology University, India

REVIEWED BY
Kristian Caldo,
University of Calgary, Canada
Kanti Kiran,
Gujarat Biotechnology University,
Gandhinagar, India
Rashmi Panigrahi,
University of Alberta, Canada

*CORRESPONDENCE
Aruna Kilaru
kilaru@etsu.edu
Jay Shockey
Jay.Shockey@usda.gov

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Plant Biotechnology, a section of the journal Frontiers in Plant Science

RECEIVED 29 September 2022
ACCEPTED 15 December 2022
PUBLISHED 11 January 2023

CITATION
Behera J, Rahman MM, Shockey J and Kilaru A (2023) Acyl-CoA-dependent and acyl-CoA-independent avocado acyltransferases positively influence oleic acid content in nonseed triacylglycerols. *Front. Plant Sci.* 13:1056582. doi: 10.3389/fpls.2022.1056582

Acyl-CoA-dependent and acyl-CoA-independent avocado acyltransferases positively influence oleic acid content in nonseed triacylglycerols

Jyoti Behera[1†], Md Mahbubur Rahman[1,2†], Jay Shockey[3*] and Aruna Kilaru[1*]

[1]Department of Biological Sciences, East Tennessee State University, Johnson City, TN, United States, [2]dNTP Laboratory, Teaneck, NJ, United States, [3]U.S. Department of Agriculture, Agricultural Research Service, Southern Regional Research Center, Commodity Utilization Research Unit, New Orleans, LA, United States

In higher plants, acyl-CoA:diacylglycerol acyltransferase (DGAT) and phospholipid:diacylglycerol acyltransferase (PDAT) catalyze the terminal step of triacylglycerol (TAG) synthesis in acyl-CoA-dependent and -independent pathways, respectively. Avocado (*Persea americana*) mesocarp, a nonseed tissue, accumulates significant amounts of TAG (~70% by dry weight) that is rich in heart-healthy oleic acid (18:1). The oil accumulation stages of avocado mesocarp development coincide with high expression levels for type-1 DGAT (*DGAT1*) and *PDAT1*, although type-2 DGAT (*DGAT2*) expression remains low. The strong preference for oleic acid demonstrated by the avocado mesocarp TAG biosynthetic machinery represents lucrative biotechnological opportunities, yet functional characterization of these three acyltransferases has not been explored to date. We expressed avocado *PaDGAT1, PaDGAT2*, and *PaPDAT1* in bakers' yeast and leaves of *Nicotiana benthamiana*. *PaDGAT1* complemented the TAG biosynthesis deficiency in the quadruple mutant yeast strain H1246, and substantially elevated total cellular lipid content. *In vitro* enzyme assays showed that *Pa*DGAT1 prefers oleic acid compared to palmitic acid (16:0). Both *PaDGAT1* and *PaPDAT1* increased the lipid content and elevated oleic acid levels when expressed independently or together, transiently in *N. benthamiana* leaves. These results indicate that *Pa*DGAT1 and *Pa*PDAT1 prefer oleate-containing substrates, and their coordinated expression likely contributes to sustained TAG synthesis that is enriched in oleic acid. This study establishes a knowledge base for future metabolic engineering studies focused on exploitation of the biochemical properties of *Pa*DGAT1 and *Pa*PDAT1.

KEYWORDS

triacylglycerol, avocado, nonseed, oleic acid, DGAT1, DGAT2, PDAT1

1 Introduction

In plants, storage lipids primarily consist of triacylglycerol (TAG) that are assembled through three sequential acylations of glycerol-3-phosphate (G3P), catalyzed by *sn*-specific membrane-bound acyltransferases (Shockey et al., 2006; McFie et al., 2010; Turchetto-Zolet et al., 2011) in the endoplasmic reticulum (ER) membrane. The first two acylations are catalyzed by glycerol-3-phosphate acyltransferase and lysophosphatidic acid acyltransferase to form an important metabolic intermediate phosphatidic acid (PA) (Murata and Tasaka, 1997; Shockey et al., 2016). After removal of a phosphate group from the *sn*-3 position of PA by phosphatidic acid phosphatase, diacylglycerol (DAG) is acylated by either diacylglycerol acyltransferase (DGAT) or phospholipid:diacylglycerol acyltransferase (PDAT) (Banaś et al., 2000; Dahlqvist et al., 2000) to yield TAG. At least two structurally unrelated polypeptides, namely DGAT1 and DGAT2, can catalyze the transfer of a fatty acid from acyl-coenzyme A (CoA) to DAG to generate TAG (Hobbs et al., 1999; Routaboul et al., 1999; Shockey et al., 2006; Xu et al., 2014). The specific roles of the two major DGAT types are not fully established, although existing studies suggest that DGAT1 plays a major role in seed and fruit oil accumulation in many plant species, whereas DGAT2 often plays a larger role in TAG production in certain novel oilseeds that produce high levels of unusual fatty acids (Burgal et al., 2008; Li et al., 2010a; Li et al., 2010b; Li et al., 2012; Chen et al., 2022). More recently, a third category of *DGAT* genes that encode soluble cytoplasmic activity, called DGAT3, were also shown to be associated with oil metabolism in plants (Saha et al., 2006; Hernández et al., 2012; Cao et al., 2013; Aymé et al., 2018; Xue et al., 2022). PDAT, on the other hand, catalyzes the transfer of an acyl group from phosphatidylcholine (PC) to the *sn*-3 position of DAG to form TAG (Dahlqvist et al., 2000). *In vitro* PDAT activity assays suggest a modest preference for the *sn*-2 position of acyl donor PC, although transfer from the *sn*-1 position also occurs (Ståhl et al., 2004). PC is of particular importance in oilseed metabolism; the ER membrane PC pool serves as the substrate for desaturation by fatty acid desaturase FAD2 and FAD3 in vegetative and reproductive organs of plants (Ohlrogge and Browse, 1995; Bates et al., 2007; Bates et al., 2009), as well as modifications such as fatty acid hydroxylation, conjugation, and epoxidation found in various exotic oilseeds (Bates et al., 2009). As such, direct metabolic utilization of PC by PDAT is a major determinant in the distinctive seed fatty acid profile of hydroxy-, and epoxy-fatty acid accumulating plant species such as *Ricinus communis* (castor bean), *Vernonia galamensis*, *Euphorbia lagascae*, and *Stokesia laevis* (Li et al., 2010a; van Erp et al., 2011). Along with DGAT, PDAT activity also plays an important complementary role in oil content and fatty acid composition in plants with otherwise 'normal' seed fatty acid profiles, such as *Arabidopsis thaliana* and *Camelina sativa* (Ståhl et al., 2004; Zhang et al., 2009; Marmon et al., 2017).

DGAT1 was first identified in mouse, based on sequence similarity with the mouse acyl CoA:cholesterol acyltransferase, and was shown to possess acyl-CoA-dependent DAG acyltransferase activity (Cases et al., 1998). Subsequently, DGAT1 homologs were identified in many plant species (Hobbs et al., 1999; Routaboul et al., 1999; Zou et al., 1999; Bouvier-Navé et al., 2000; He et al., 2004; Shockey et al., 2006; Aymé et al., 2015). DGAT1 proteins are integral ER membrane proteins containing between six and ten transmembrane-spanning domain (TMDs) (Yen et al., 2008). Both termini of the tung DGAT1 protein are cytosol-exposed, indicating that all or most DGAT1s (at least in plants) contain an even number of TMDs (Shockey et al., 2006; Zhang et al., 2009). The variable *N*-terminal region of DGAT1 is cytosol-localized and is associated with distinct functions in different organisms (Liu et al., 2012). Overall, DGAT1 orthologs (~490-530 amino acid residues) contain several conserved motifs including a DAG binding motif, an acyl-CoA binding motif, a hydrophobic *C*-terminal ER retrieval motif, and a fatty acid-binding protein signature (McCartney et al., 2004; Shockey et al., 2006; Guihéneuf et al., 2011). A relatively smaller (~320-350 amino acid residues) protein with two TMDs (Shockey et al., 2006), DGAT2 arose by convergent evolution, and is structurally unrelated to the DGAT1 family. First purified from the oleaginous fungus *Mortierella ramanniana* (Lardizabal et al., 2001), DGAT2 also has distinct effects on lipid content and composition in a broad array of organisms, from humans and mice (Cases et al., 2001; Stone et al., 2004) to yeast and castor (Oelkers et al., 2002; Burgal et al., 2008).

First identified in yeast (Dahlqvist et al., 2000), PDAT is a large (~650-680 amino acid residues) protein that is targeted to the ER membrane by a single TMD, rendering much of the protein facing the ER lumen. Overexpression of membrane anchor-deleted forms of yeast PDAT resulted in a soluble secreted protein that retains most of the biochemical properties of the native enzyme (Ghosal et al., 2007). The identity of the PDAT active site(s) remain unknown, but plant orthologs do show homology to the catalytic triad with nucleophile-acid-base (S-D-H) residues typically found in the alpha-beta hydrolase fold motif (Ollis et al., 1992).

Functional characterization of DGAT genes in mammalian systems has mostly outpaced similar research in plants. In mammals, both DGAT1 and DGAT2 contribute to TAG biosynthesis (Cases et al., 1998) and exhibit non-redundant physiological roles (Smith et al., 2000; Chen et al., 2002; Stone et al., 2004), while the specific roles and relative quantitative contributions of DGAT and PDAT to seed oil metabolism is less well-understood. At least six *PDAT*-like genes have been identified in *Arabidopsis* but only *PDAT1* accounted for most of the measurable acyltransferase activity (Ståhl et al., 2004). Disruption or overexpression of *PDAT1* had no significant

impact on TAG accumulation in *Arabidopsis* seeds (Ståhl et al., 2004; Mhaske et al, 2005). Loss of DGAT1 activity in *Arabidopsis* causes modest reductions in total seed oil content and a strong shift towards polyunsaturated fatty acids in seed oil (Routaboul et al., 1999; Zou et al., 1999). However, the phenotypic effect of a combined loss of DGAT1 and PDAT1 in *Arabidopsis* is severe; a *pdat1*/*dgat1* double mutant contained abnormal, sterile pollen, resulting in poorly developed nonviable embryos (Zhang et al., 2009). Seed-specific RNAi silencing of *PDAT1* expression in an *Arabidopsis dgat1* mutant background defined the outer limits of tolerable interference to seed oil metabolism, resulting in 70–80% reductions in seed TAG content (Zhang et al., 2009). The ability of *PDAT1* to compensate for the loss or suppression of *DGAT1* and *vice versa* suggests that the acyl-CoA-dependent and -independent pathways are cooperatively involved in TAG synthesis in many oil-rich tissues.

As such, DGAT and PDAT enzymes possess biotechnological potential as tools for metabolic engineering studies. In transgenic studies, *Arabidopsis DGAT1* enhanced TAG content when overexpressed in yeast and tobacco leaves (Bouvier-Navé et al., 2000). Some plant DGAT1 show preference for substrates containing distinct fatty acids (Li et al., 2010a; Li et al., 2010b; Li et al., 2012; Aznar-Moreno et al., 2015; McKeon and He, 2015). Arabidopsis DGAT1 prefers long-chain fatty acids, while oil palm (*Elaeis guineensis*) *EgDGAT1-1* prefers medium-chain fatty acids when expressed in yeast (Aymé et al., 2015). Enrichment of unusual fatty acids in engineered seed oils has been achieved with some plant DGAT2 enzymes. Tung tree (*Vernicia fordii*) DGAT2 preferentially incorporates α-eleostearic acid, a conjugated trienoic C18 fatty acid, in transgenic plants and yeast (Shockey et al., 2006; van Erp et al., 2015), while castor, *V. galamensis*, and *S. laevis* DGAT2s display strong selectivity towards epoxy and hydroxy fatty acids, both *in vitro* and *in vivo* (Kroon et al., 2006; Burgal et al., 2008; Li et al., 2010a; Li et al., 2010b). In addition to the striking differences in biochemical properties, tung and castor *DGAT2* are expressed at much higher levels than their cognate *DGAT1*s in developing seeds, suggesting strong positive selection for roles in selective accumulation of unusual fatty acids (Liu et al., 2012). However, in 'normal' seed tissues that accumulate only standard C16-C22 fatty acids, such as the embryos of *Arabidopsis*, rapeseed, and soybean, *DGAT1* was highly expressed, while *DGAT2* transcript was barely detectable (Li et al., 2010b; Troncoso-Ponce et al., 2011; Li et al., 2013; Zhang et al., 2016).

Avocado (*Persea americana*), a basal angiosperm species from the Lauraceae family, is an evolutionarily important cash crop that accumulates a significant amount (60-80% by dry weight) of nutritionally rich TAGs in its mesocarp (nonseed) tissue (Takenaga et al., 2008; Kilaru et al., 2015). In seeds, TAGs typically accumulate in the form of lipid droplets (LDs), which are ordered, spherical, neutral lipid aggregates, surrounded by an oleosin-rich phospholipid/protein monolayer (Huang, 1996). In avocado mesocarp, LDs are associated with a non-oleosin class of proteins and are predominantly composed of monounsaturated oleic acid (18:1) (Gidda et al., 2013; Horn et al., 2013). Although both *PaDGAT1* and *PaDGAT2* are expressed in mesocarp tissue, transcript levels for *PaDGAT1* were dominant, with ≥2-fold higher expression than *PaDGAT2* throughout fruit development (Kilaru et al., 2015). *PaPDAT1* expression was also detected at a comparable level to that of *PaDGAT1* during the early fruit developmental stages (Kilaru et al., 2015). Transcriptome studies of oil-rich nonseed tissues such as avocado, oil palm, and olive revealed several similarities in the expression pattern of genes involved in plastidial fatty acid synthesis, while unique variations were observed in the expression of genes involved in TAG assembly (Bourgis et al., 2011; Troncoso-Ponce et al., 2011; Kilaru et al., 2015; Rahman et al., 2016). The concomitant expression of both *DGAT1* and *PDAT1* in seed tissues is not evolutionarily maintained across all oilseeds; the mesocarp expression patterns might represent a less-common repertoire of coordinated TAG biosynthesis regulation in avocado. However, the acyltransferases from avocado that are involved in TAG assembly have not been identified and functionally characterized to date. Hence, we sought to study the properties of *PaDGAT1, PaDGAT2*, and *PaPDAT1* and assess their potential contributions to TAG biosynthesis in nonseed tissues.

2 Materials and methods

2.1 Identification and *in silico* analyses of putative *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1

Full-length complementary DNA (cDNA) sequences for *PaDGAT1* (OP727298), *PaDGAT2* (OP727299), and *PaPDAT1* (OP727300) were retrieved from the avocado transcriptome database (Kilaru et al., 2015) and verified against avocado genome sequence data (Ibarra-Laclette et al., 2015; Rendón-Anaya et al., 2019). The cDNA sequences were further analyzed using NetStart1.0 prediction software (http://www.cbs.dtu.dk/services/NetStart/) to check for the locations of probable start codons. The corresponding protein sequences were predicted using the ExPASy protein translation tool (https://web.expasy.org/translate/) and the coding sequence for the longest open reading frame was selected for further use. This resulted in 1608 bp, 996 bp, and 2049 bp long putative full-length protein coding sequences for *PaDGAT1, PaDGAT2*, and *PaPDAT1*, respectively (Table 1).

The predicted avocado protein sequences were compared to other plant orthologs by multiple sequence alignments using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Probable transmembrane spanning domains were predicted using TMHMM transmembrane domain prediction software (http://www.cbs.dtu.dk/services/TMHMM/) and Phobius

TABLE 1 Bioinformatic analytics for *Pa*DGAT1, *Pa*DGAT2 and *Pa*PDAT1.

In silico analyses	*Pa*DGAT1	*Pa*DGAT2	*Pa*PDAT1
Full length cDNA	**2079 bp**	**1385 bp**	**2079 bp**
Coding sequence length	**1608 bp**	**996 bp**	**2052 bp**
Probability of start codon	**78.6%**	**86.0%**	**89.80%**
Protein length	**535 aa**	**331 aa**	**682 aa**
Estimated molecular weight	**61.2 kDa**	**37.2 kDa**	**75.8 kDa**
No of transmembrane domains	**9**	**6**	**1**

(https://phobius.sbc.su.se/) and were compared with transmembrane domains of characterized DGAT1, DGAT2, and PDAT1 proteins, respectively. The phylogenetic tree was constructed by the maximum likelihood method based on the Poisson correction model (Zuckerkandl and Pauling, 1965), and the tree with the highest log likelihood was shown. The phylogenetic figure was drawn to scale, and evolutionary analyses were performed in Molecular Evolutionary Genetics Analysis (MEGA) X (Kumar et al., 2018). The name of each protein consists of the first letter of the species followed by the first letter of the genus.

The three-dimensional (3D) structures were predicted by SWISS-MODEL interactive workspace (https://swissmodel.expasy.org/interactive) using the 3D structure of proteins showing the highest sequence similarity as templates for each respective protein of interest. The predicted 3D models with full atomic coordination information were obtained as Protein Data Bank (.pdb) files and visualized using USCF CHIMERA 1.14 software (https://www.cgl.ucsf.edu/chimera/cgi-bin/secure/chimera-get.py?file=win64/chimera-1.16-win64.exe). For structure comparison, the 'MatchMaker > align' option was used and results were obtained as superimposed structures and structure-based sequence alignments. The positions of structurally important residues are shown in the figures with the amino acid single letter code followed by the corresponding position. The root mean square deviation (RMSD) values of C-alpha atoms were also calculated and shown along with the structure-based sequence alignment.

2.2 Cloning of putative *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1 for yeast expression

A bi-directional expression vector (pESC-URA, Agilent, USA) was used to direct galactose-inducible expression of *PaDGAT1*, *PaDGAT2* and *PaPDAT1* in bakers' yeast (*Saccharomyces cerevisiae*). The cDNAs were synthesized from avocado mesocarp total RNA using oligo (dT) primer and AMV Reverse Transcriptase (Promega). For PCR amplification, forward and reverse primers were designed to include *Not*I and *Pac*I restriction sites, located upstream of the start codon and downstream of the stop codon, respectively. Additionally, "ACC" nucleotides were added upstream of the restriction site in the forward primer to optimize translational efficiency (Kozak, 1989). The full-length cDNAs of *PaDGAT1*, *PaDGAT2*, and *PaPDAT1* were amplified using Advantage 2 PCR (clontech) kit. The resulting PCR products and the vector were digested with *Not*I and *Pac*I (New England Biolabs, USA). Agarose gel-purified digested amplicons and vector DNA were ligated using T4 DNA ligase (New England Biolabs, USA). Aliquots of the ligation reactions were transformed into chemically competent *E. coli* TOP10 cells and selected on solid agar media plates containing the appropriate antibiotic. Positive transformants were selected by colony PCR and further confirmed by sequencing. Plasmid DNA was extracted and used for yeast transformation.

The *S. cerevisiae* quadruple mutant strain H1246 (*Δdga1/Δlro1/Δare1/Δare2*) (Sandager et al., 2002) was transformed either with empty pESC-URA as a negative control or pESC-*VfDGAT*1-URA (*Vernicia fordii* DGAT1) for the positive control (Shockey et al., 2006), or each of pESC-*PaDGAT1*-URA, pESC-*PaDGAT2*-URA, and pESC-*PaPDAT1*-URA, using the S.c. EasyComp Transformation Kit (Invitrogen, USA). These transformed cells were plated on SD-URA (synthetic dextrose minus uracil) plates containing 2% glucose, 0.67% yeast nitrogen base. After 2-4 days, several colonies were selected and induced to grow in SRGG-URA (synthetic raffinose, glycerol, galactose minus uracil) plates containing 1% raffinose, 0.25% glycerol, 2% β-galactose (Difco), 0.67% yeast nitrogen base (Difco), and synthetic complete mixture of amino acids minus uracil for 24 hours. All liquid cultures were grown in similar media at 28°C on a shaker at 250 rpm.

2.3 Indirect enzyme activity analysis *via* lipotoxicity rescue assay

Lipotoxicity rescue assays were performed as described previously (Sandager et al., 2002; Garbarino and Sturley, 2009; Siloto et al., 2009). All transformed strains including negative and positive controls were grown on SD-URA plates at 28°C for

2-3 days. Positive colonies were first grown in a small volume (~5 mL) of SD-URA liquid media at 28°C. Cell optical densities (OD) were measured spectrophotometrically at 600 nm (OD_{600nm}). Starter cultures were back-diluted in SRGG-URA liquid media at 28°C with shaking overnight to induce transgene expression. Fresh SRGG-URA liquid media was prepared and split in half: one portion containing base components with 0.2% Tergitol detergent alone and one portion containing base components, 0.2% Tergitol, and 1.0 mM free linoleic (18:1) or linolenic (18:2) fatty acids (prepared as 0.5 M stock, in ethanol). Both types of media were warmed at 28°C at least 15-30 minutes prior to adding to the yeast cells to ensure even dispersal of fatty acid in the media, followed by addition of 0.1 OD of pelleted, washed cells into each media type. Cultures were incubated with vigorous shaking for 26 hours at 28°C and cell densities sampled at various time points (0, 12, 22, and 26 hours) until cultures reached late log or stationary phase (~8-12 OD). The cell density readings were plotted against various time points to generate yeast growth curves. Three technical replicates and two biological replicates were done to ensure reproducibility.

2.4 Microsome isolation

Microsomes were prepared from transgenic H1246 yeast strains as previously described (Shockey et al., 2006). Briefly, yeast cultures grown overnight in SRGG-URA liquid media were harvested by centrifugation, washed twice with sterile dH_20 and once with 1X phosphate buffer saline (PBS) respectively. After washing, the pellets were resuspended in cell lysis buffer I (1x PBS containing Mini EDTA-free protease cocktail tablets from Roche Diagnostics; 1 tablet/50 mL buffer) and transferred to a 15-mL glass tube. An equal volume of 0.50 mm acid-washed glass beads was added to the cells and were lysed by vortexing. Debris and unbroken cells were removed by low-speed centrifugation. The supernatant was recovered and centrifuged again at 100,000 x *g* for 1 h at 4°C. Microsomal pellets from the high-speed spin were collected and resuspended in 50 mM Tris-HCl and 25% glycerol, pH 8.0.

2.5 *In vitro* enzyme activity and substrate specificity

Enzyme activity and substrate specificity of DGAT1 were measured *in vitro* by the incubation of microsomal fractions with different combinations of radiolabeled fatty acyl-CoA substrate and DAG as described previously with some modifications (Shockey et al., 2006). Briefly, 25 µg of microsomal membrane proteins were incubated with 10 µM [1- ^{14}C] acyl-CoAs [palmitoyl (16:0) -CoA or oleoyl (18:1) -CoA from PerkinElmer, specific activity, 20,000 CPM/µmol] and 400 µM DAG [di-palmitin (di16:0) and diolein (di18:1); stock of 40 mM in methanol)] in a total reaction volume of 100 µL buffered with 100 µM Tris-HCl (pH 8.0). The reaction mixture was incubated at 37°C for one hour in a water bath. Lipids were extracted, concentrated under vacuum, and dissolved in 20 µL chloroform. Lipids were spotted on Silica Gel 60 plates (Whatman) thin layer chromatography (TLC) plates and separated as described previously (Janero and Barrnett, 1981). Developed TLC plates were dried for 10 min and placed on AR-2000 radio-TLC imaging scanner (Bioscan, Inc., USA) to detect and quantify the radioactive TAG spots. Assay mixtures including all components except microsomal proteins were used as a negative control. The *in vitro* assay was performed in triplicate for each acyl-CoA substrate and each enzyme. The obtained data are represented as mean ± SD.

2.6 Cloning for *in planta* expression

For *N. benthamiana* (benth) leaf expression, digested *Nco*I/*Sac*II amplicons for *PaDGAT1* and *PaPDAT1* were cloned into the corresponding sites in the pK34 cloning vector. Plasmid K34 contains a *CaMV35S* promoter and a *35S* terminator sequence (Shockey et al., 2015). The recombinant pK34 constructs were selected as described above, followed by *Asc*I digestion and gel-purification of the *CaMV35S promoter:gene:35S terminator* cassettes, and cloning into the corresponding site in the binary expression vector pB110. A viral silencing suppressor protein gene (*P19*) (Wood et al., 2009) was also cloned into pB110, either alone or along with the *DGAT/PDAT* expression cassettes, using the same series of cloning steps and was used for co-infiltration in the benth leaves. Prior to benth leaf infiltration, all positively confirmed plasmids were isolated from *E. coli* and transformed into *Agrobacterium tumefaciens* as described below.

2.7 *Agrobacterium*-mediated transformation of *N. benthamiana* leaves

Agrobacterium LBA4404 strain was used for transient gene expression in benth leaves. Approximately 100-1000 ng of finished binary plasmid DNA containing the genes of interest were transformed into competent cells, and selected on solid media containing kanamycin as described previously (Behera et al., 2021). Plates were incubated for 3-6 days at 28°C for colony formation. Positive colonies were used to grow O/N starter cultures at 28°C with shaking (250 rpm). Overnight cultures were supplemented with 100 mM acetosyringone and grown for another 2 hours. Pellets were collected by centrifugation at 400 x *g* for 5 min and resuspended in the infiltration buffer (5 mM $MgSO_4$, pH 5.7, 100 mM acetosyringone, and 5 mM methyl ethanesulfonate) to a final OD of 0.3.

N. benthamiana plants were grown in a 24°C plant growth room with overhead lighting using 9:15 light: dark cycle. Three

slightly aged leaves located slightly below the apical bud were selected from five/six-week-old *N. benthamiana*. Two leaves were infiltrated at high coverage with each transgenic *A. tumefaciens* strain. The last leaf was used for two controls-mock (containing only infiltration buffer) and negative control, which contains infiltration buffer and *Agrobacterium* cells expressing empty plasmid vector. To facilitate infiltration, several well-spaced small nicks were created with a needle in the abaxial leaf epidermis. After infiltration, plants were then transferred into the growth chamber and allowed to grow for 6-7 days to express the protein.

2.8 Yeast cell and benth leaf lipid extraction

Harvested yeast cells were washed twice with sterile distilled water and once with 1X PBS. For infiltrated *N. benthamiana* leaves, samples were collected at 6-7 days after infiltration and 200 mg fresh weight (FW) was used for lipid extraction. Cells/leaves were treated with isopropanol at 80°C for 10-15 min to stop lipolytic activity. Lipid extraction was carried out using a modified Hara and Radin method (Hara and Radin, 1978). Glyceryl triheptadecanoate (17:0) was added as an internal standard during lipid extraction. An equal volume of acid-washed glass beads (0.5 mm) was added to each sample and vortexed vigorously. For phase separation, water: isopropanol: hexane (1:4:6 v/v/v) were added to the sample. The sample was vortexed well and allowed to stand for a few minutes. Aqueous sodium sulfate solution (2.5 ml of 6.6 g anhydrous sodium sulfate in 100 mL) was added and vortexed vigorously, followed by centrifugation for 5 min at 5000 x *g* at room temperature. The upper hexane phase was carefully transferred to another pre-weighed glass tube and dried under nitrogen gas. The total lipid was quantified (mg/g of fresh tissue), resuspended in one mL hexane, and kept at -20°C for further analysis.

2.9 Triacylglycerol separation and quantification

TAGs were separated from the total lipid extracts by TLC on Silica Gel 60 plates (Whatman). Lipid extracts were spotted on the TLC plate and developed in a solvent tank containing hexane: diethyl ether: glacial acetic acid (70:30:1, v/v/v). TAGs were visualized by brief exposure to iodine vapor. For charring, plates were sprayed with a solution of 10% (v/v) H_2SO_4 in methanol and heated until spots appeared. TAGs were scratched from the TLC plate with a razor and were extracted from the silica plates using the lipid extraction procedure described previously. Lipids were then quantified gravimetrically and processed for composition analyses by gas chromatography (GC).

2.10 Determination of TAG fatty acid composition

The fatty acid content and composition of transformed yeast strains were determined by gas chromatography flame ionization detector (GC-FID) (Browse et al., 1986). Briefly, 100 µL of TAG extract was dried under liquid nitrogen and 2 mL of hexane was added. For transmethylation, 200 µL of KOH/MeOH (2M) was added and vortexed for 2 min. To acidify (~ pH 3-4), 400 µL of 2M HCl was added to the mixture. After adding 2 mL of hexane, the sample was centrifuged for 5 min at 3000 x *g* and the upper hexane phase was recovered into a new test tube and dried under nitrogen gas. The dried fatty acid methyl ester was resuspended in 1 mL hexane for analysis by GC-FID (Varian) (Han and Xianlin, 2010). A capillary column (DB-23; 30 m x 0.32 mm I.D., 0.25 µm) was used in the chromatography with the helium carrier gas (flow rate of 1.5 mL/min). The GC-FID column was equilibrated with 2 µL hexane (blank) injection before sample injection. The sample was injected for three min at a temperature of 150°C. Then samples were separated at 150°C for three minutes followed by 150-240°C at 6°C/min, where the detection temperature was 300°C. Each sample was analyzed in triplicate.

Fatty acids from extracted total lipids were converted to fatty acid methyl esters (FAMEs) by heating the samples with two mL of 1N methanolic HCl for 2 h at 85°C in a water bath. Samples were cooled after the transesterification reaction and FAMEs were extracted into the organic phase after vortexing with 1 mL aqueous 0.88% KCl and 1 mL hexane followed by centrifugation for 5 min at 5000 x *g*. The upper hexane phase containing FAMEs was collected into a new test tube, dried under nitrogen gas at 40°C, then resuspended in 100 µL of hexane. One µL of FAME sample was injected into the column for separation and detection by GC-FID. The retention time for each fatty acid was determined by comparison of coelution times of components of standard FAME mixes (Supelco, USA). Fatty acids were quantified relative to known quantities of triheptadecanoin (tri-17:0) internal standard added to each sample prior to lipid extraction.

2.11 Nile Red staining and visualization of neutral lipids

Transgenic yeast cells were harvested by centrifugation at 1300 x *g* for 10 min at room temperature, then washed twice each with 1X PBS and sterile distilled water. For sample preparation, 2 µL of Nile Red solution (0.8 mg/mL stock in methanol) was added to 200 µL of yeast cell suspension (1:100 ratio of Nile Red and yeast cells) and incubated at room temperature for 20 min. Portions of the stained cell suspensions were placed on glass slides with coverslips. Agroinfiltrated leaf discs were collected and placed in 50-mL

plastic tubes containing 4% paraformaldehyde in 1X PBS on a rotational shaker at 75 rpm for 1 h. Leaf discs were washed three times with 1X PBS and stained with 4 µg/mL Nile Red (4 mg/mL stock in DMSO) in 1X PBS for 15 min at room temperature in a rotational shaker at 100 rpm in the dark. Stained leaf samples were washed three times with 1X PBS and once with ddH_2O, then mounted on glass slides with coverslips. The microslides containing yeast cells or leaf discs were immediately observed using a Leica TCS SP8 confocal fluorescence microscope with an excitation wavelength of 488 nm and emission at 600-650 nm and 560-620 nm, respectively (n = 3). The total number of LDs was counted for each experiment by ImageJ software. At least three panels from each biological replicate, for a total of nine panels, were analyzed to quantify the LDs. Data were represented as mean ± SD.

2.12 Statistical analysis

Statistical analysis was conducted using Minitab statistical software. Data were expressed as their mean value and standard deviation (SD). Student's paired t-test, and/or one-way ANOVA was used to determine significant differences between the data sets. Pairwise comparisons were made using Tukey's post-test. The level of significance (p-value) for each analysis is mentioned in the text.

3 Results and Discussion

3.1 Putative *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1 contain conserved motifs and sequence similarity with plant orthologs

To gain insight into the structural and sequence features of avocado DGAT1, DGAT2, and PDATI, we conducted comprehensive *in silico* analyses using various bioinformatic tools. The predicted polypeptides of *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1 have 535, 331, and 682 amino acid residues, respectively (Table 1). The multiple sequence alignment comparisons of *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1 with other related classes of eukaryotic acyltransferases revealed high degree of sequence similarity (Supplementary Figure 1) and in some cases, conserved predicted and/or functionally characterized motifs (Figure 1). *Pa*DGAT1 is 95.5% similar to the putative DGAT1 from another Lauraceae species, Chinese spicebush (*Lindera communis*) (Dong et al., 2014), and 66% and 68% identical to biochemically characterized orthologs from olive (*Olea europaea*, *Oe*DGAT1) and canola (*Brassica napus*, *Bn*DGAT1), respectively (Supplementary Figure 1D). Although relatively few catalytic domains have been experimentally confirmed in DGATs, multiple sequence alignment of several plant DGAT1 protein sequences (including putative *Pa*DGAT1) reveals a set of predicted functional motifs, including putative binding sites for acyl-CoA, fatty acid, and DAG (Figure 1A), which were initially identified in *Arabidopsis* DGAT1 (Hobbs et al., 1999; Jako et al., 2001) and *T. majus* DGAT1 (Xu et al., 2008). A known *C*-terminal ER retrieval motif (Shockey et al., 2006) was also identified in the *C*-terminal region of *Pa*DGAT1. Previously, 55 DGAT1 sequences from plants, animals, and fungi were shown to contain 41 conserved amino acids (Cao, 2011), all of which were also retained in *Pa*DGAT1.

*Pa*DGAT2 also showed high similarity with biochemically characterized *Oe*DGAT2 (62.5%), *At*DGAT2 (57.8%), and *Vf*DGAT2 (58.6%) (Figure 1B; Supplementary Figure 1E). Like other plant DGAT2 sequences analyzed previously (Cao, 2011), *Pa*DGAT2 retained 16 conserved amino acids. Most of the conserved residues in both DGAT families are clustered in the *C*-terminal regions, while the *N*-termini showed greater degrees of variation. Interestingly, as a basal angiosperm, the evolutionary relationships between avocado proteins and other higher plant orthologs can be difficult to predict; in this case, *Pa*DGAT1 and *Pa*DGAT2 were most closely related to orthologs from the monocot species rice (*Oryza sativa*) (Supplementary Figure 1A) and maize (*Zea mays*) (Supplementary Figure 1B), respectively. The DAG-binding 'HPHG' motif first identified in murine DGAT2 (Stone et al., 2006) is replaced with 'EPHS' in *Pa*DGAT2 and other plant DGAT2s (Figure 1B) (Jeppson et al., 2020). Similarly, PDAT1 sequence alignment showed that avocado PDAT1 is more similar to *Zm*PDAT1 from monocot maize (78.5%) compared to the orthologs from the eudicots *Arabidopsis* (AtPDAT1; 75.2%) and *Cannabis sativa* (*Cs*PDAT1; 74.1%) (Figure 1C; Supplementary Figure 1F). The PDAT1 *N*-terminal region is more variable, while the *C*-terminal regions are more highly conserved. The lid region, salt bridge, and S-D-H catalytic triad previously shown in *At*PDAT1 are also conserved in all four sequences analyzed here, including *Pa*PDAT1 (Figure 1C) (Falarz et al., 2020). The structurally important methionine residue found near the catalytic triad serine is also conserved in *Pa*PDAT1 at position 263 (Falarz et al., 2020).

Membrane topology analyses for the avocado enzymes by TMHMM and Phobius suggested odd numbers of TMDs for both DGATs (Supplementary Figure 2). Previous studies demonstrated both even and odd number of TMDs for structurally related plant DGATs (Routaboul et al., 1999; Shockey et al., 2006; Wang et al., 2006). Only one TMD was predicted for *Pa*PDAT1 by both the prediction tools (Supplementary Figure 2), consistent with previous functional studies of yeast PDAT (Ghosal et al., 2007). The full summary of *in silico* analytics for avocado DGAT1, DGAT2, and PDAT1 is shown in Table 1. We further predicted a structurally disordered region in the *N*-terminal region of *Pa*DGAT1, upstream of predicted TMD1 (Supplementary Figure 3). Similar unstructured protein domains, called intrinsically disordered regions (IDRs) (Dyson and Wright, 2005), were also

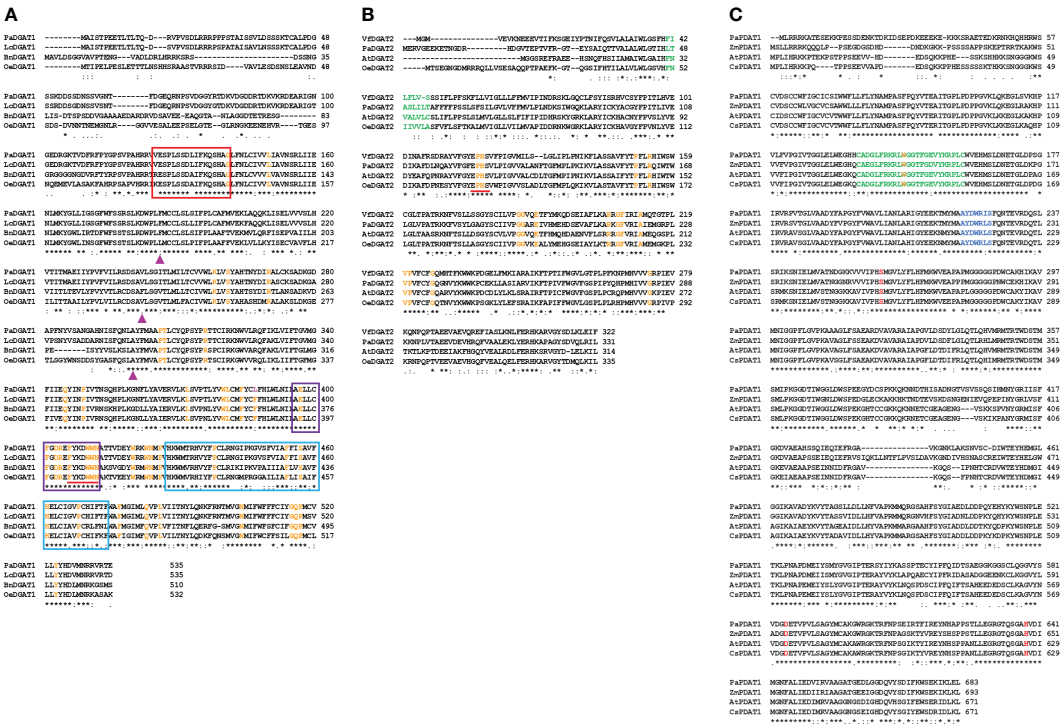


FIGURE 1
Sequence alignment of *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1 protein sequences to orthologs from different plant species. Orthologous sequences were collected from NCBI and subjected to sequence alignment using ClustalW online tool. Multiple sequence alignment and prediction of conserved motifs and amino acids in plant DGAT1 **(A)**, and DGAT2 **(B)** and PDAT1 **(C)** protein sequences are shown. The binding sites for acyl-CoA (Red), fatty acid (Purple) and DAG (Blue) conserved motifs of DGAT1 are shown in colored boxes. The 'FYxDWWN' domain is shown as underlined in red. The conserved 'L-V-V' residues among the high TAG plant groups is replaced by 'F-V-A' in *Pa*DGAT1 at positions 186, 242, and 300, respectively (▲). The 'EPHS' DAG-binding motif in DGAT2 orthologs is shown as underlined in red. The conserved amino acids identified by Cao. H., 2011 in DGAT1 and DGAT2 orthologs are shown in orange color and the substitution (F→L) in *Pa*DGAT1 at position 387 is shown in pink. The lid region, salt bridge and 'S-D-H' catalytic triad among the PDAT1 orthologs is shown in green, blue and red color, respectively. Positions having identical (the same) amino acids are indicated by *.

demonstrated in the *N*-terminal cytosolic domain of *Bn*DGAT1 and were implicated in oleoyl-CoA binding (Panigrahi et al., 2018). The IDR of *Bn*DGAT1was shown to undergo secondary structure changes upon ligand binding (Panigrahi et al., 2018) and is responsible for self-dimerization (Caldo et al., 2017). This plant DGAT1 domain also binds PA, a precursor of the DAG substrate and another key regulatory metabolite, leading to upregulation of DGAT activity (Caldo et al., 2017). Hence, it is possible that such regulatory features are intrinsic in *Pa*DGAT1 as well. Future work is necessary to confirm these various structural features and determine their functional role.

3.2 *Pa*DGAT1 and *Pa*PDAT1 showed structural conservation of active site residues

The 3D structure of the three avocado acyltransferases were predicted by SWISS-MODEL interactive workspace using closest available template (Figure 2; Supplementary Figure 4). The quality of predicted models based on structural assessment parameters are provided in Supplementary Figure 5. *Pa*DGAT1 was predicted using the recently solved 3D structure of human DGAT1 (*Hs*DGAT1, model 6VP0) as a template (Figures 2A, B) (Wang et al., 2020). The predicted structure was obtained as a dimer where the *N*-termini of each monomer interact with each other (Figure 2B). Structural comparison of the monomers showed that the universally conserved histidine residue (His461) in the MBOAT family proteins is also conserved in *Pa*DGAT1 (Supplementary Figure 6A) (Wang et al., 2020). An assessment of the predicted acyl-CoA binding site based on the information from *Hs*DGAT1 showed that the previously predicted loop region FYxDWWN domain is also conserved in *Pa*DGAT1 (Figure 1A; Supplementary Figure 6A) (Guo et al., 2001; Wang et al., 2020). The Gln465 residue in *Hs*DGAT1 that stabilizes the position of the acyl-CoA thioester for attack by DAG is also conserved in *Pa*DGAT1 at Gln516. The substrate binding pocket residues in *Hs*DGAT1 (Trp377, Asn378, His382, Ser411, His415, Glu416) are also conserved in *Pa*DGAT1 (Trp423, Asn424, His428, Ser457, His461, and Glu462,

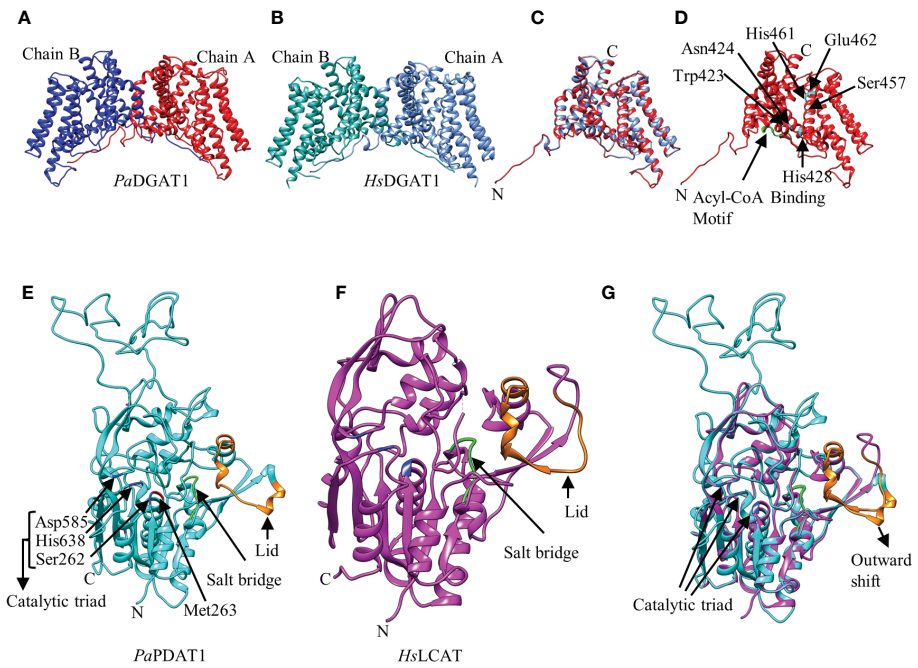


FIGURE 2
Predicted three-dimensional (3D) structures of *Pa*DGAT1 and *Pa*PDAT1 using SWISS-MODEL interactive workspace. The solved model for human DGAT1 (*Hs*DGAT1; PDB ID: 6VP0) and lecithin:cholesterol acyltransferase (*Hs*LCAT; PDB ID: 5txf) were used to predict the structures of *Pa*DGAT1 and *Pa*PDAT1, respectively. The dimeric structures of *Pa*DGAT1 **(A)** and *Hs*DGAT1 **(B)** are shown followed by structure alignment using chain A **(C)** of both. The predicted acyl-CoA binding motif is shown in green and the substrate binding pocket residues are indicated by amino acid three letter code followed by their respective positions **(D)**. Similarly, the 3D structures of *Pa*PDAT1 **(E)**, *Hs*LCAT **(F)** and their structure alignment **(G)** are shown. The salt bridge and the lid are shown in green and orange, respectively. Predicted 3D model for *Pa*DGAT2 is provided as Supplementary Figure 4 and the structural assessment parameters for all the three models are provided in Supplementary Figure 5.

respectively) (Figure 2D; Supplementary Figure 6A). The overall structural similarities between *Pa*DGAT1 and *Hs*DGAT1 support the likelihood that the avocado enzyme will effectively utilize oleic acid-containing substrates.

Prediction of the *Pa*DGAT2 3D structure by SWISS-MODEL interactive workspace identified bacterial enzyme *Thermotoga maritima* 1-acyl-*sn*-glycerol-3-phosphate acyltransferase (PlsC) as the best template (PDB ID: 5kym) (Supplementary Figures 4A, B). However, the *N*-terminal region (residues 1-93) could not be reliably predicted and thus was not included in the figures shown here. Further structural alignment showed only 15.15% identity (Supplementary Figures 4C, 5B). The catalytic domain 'HX$_4$D' identified in the PlsC structurally aligned with the 'PHSVLP' domain in the *Pa*DGAT2 structure (Supplementary Figure 4C) (Robertson et al., 2017), which overlaps with the 'EPHS' DAG binding domain (Supplementary Figure 6B). Furthermore, analysis of the putative active site and the acyl chain specificity tunnel showed very little similarity and the broad binding tunnel is masked by a 33-residue loop region (positions 241-273) (Supplementary Figure 4C). This degree of uncertainty precludes the identification of most of the important domains and key catalytic residues with confidence.

For prediction of the *Pa*PDAT1 3D structure, the human lecithin:cholesterol acyltransferase (*Hs*LCAT, PDB ID: 5txf) was used as template (Glukhova et al., 2015) and the predicted model includes residues 113-659, based on sequence similarities (Figures 2E, F). Overall, the structure alignment showed 21.09% identity (Supplementary Figure 6C). The lid region, salt bridge, and catalytic triad are shown in orange, green and blue, respectively (Figure 2E). The dynamic lid region, previously shown to play a role in controlling access of specific substrates to the active site (Manthei et al., 2017; Falarz et al., 2020) showed variation between the two structures (Figure 2G; Supplementary Figure 6C). However, such divergent structural variation among plant PDAT1s is common and likely contributes to variation in substrate specificities (Falarz et al., 2020). We hypothesize that the outward shifting of the lid in *Pa*PDAT1 (Figure 2G) may facilitate PC entry to the catalytic site. Moreover, the salt bridge, which is responsible for conformational specificity of the active site and substrate recognition, is structurally conserved in *Pa*PDAT1 (Donald et al., 2011; Falarz et al., 2020). Among the S-D-H catalytic triad (Ser262, Asp585, His638) present in avocado PDAT1, the serine residue showed structural variation compared to the human LCAT (Supplementary Figure 6C). However, the functionally important Met263 adjacent to the Ser262 in the catalytic triad is structurally conserved (Figure 2E; Supplementary Figure 6C), similar to that shown in predicted *Arabidopsis* PDAT1 structure

(Falarz et al., 2020). Hence, based on our structural analyses, we predict that *Pa*PDAT1 is a functional enzyme with different substrate preference than *Hs*LCAT.

3.3 *Pa*DGAT1 activity complements fatty acid lipotoxicity and TAG deficiency phenotypes in TAG-deficient yeast

Our previous bioinformatic and transcriptomic analyses suggested that TAG biosynthesis in avocado mesocarp is likely controlled in large part by *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1 activities (Kilaru et al., 2015), so we directly tested for TAG synthesis by avocado acyltransferase overexpression in yeast. We used the H1246 mutant strain of *S. cerevisiae* that lacks all four acyltransferases (Are1p and Are2p: sterol O-acyltransferases 1 and 2, Dga1p: diacylglycerol acyltransferase, and Lro1p: ortholog of yeast PDAT1) necessary for TAG and stearyl ester synthesis and thus is devoid of neutral lipids (Sandager et al., 2002). H1246 is sensitive to free fatty acids (FFAs) in the growth media, thus making it an ideal host for testing DGATs, PDATs, and other TAG biosynthetic genes for functional activity that will restore TAG production and rescue the cells from the FFA lipotoxic effects (Siloto et al., 2009; Pan et al., 2013). H1246 was transformed with expression plasmids bearing either *PaDGAT1*, *PaDGAT2*, or *PaPDAT1* and grown in the presence of either oleic acid or polyunsaturated (18:2) FFAs. Culture growth rates were compared to that of positive (*Vf*DGAT1) and negative (pESC-URA empty vector) controls (Shockey et al., 2006), with various time points sampled over a total of 26 hours (Figure 3). Growth curves of each strain were also compared to equivalent cultures grown in liquid media lacking fatty acids. Both oleic and linoleic acid completely inhibited growth of the negative control. On the other hand, growth of the positive control and the *Pa*DGAT1 strains were nearly identical in the presence and absence of 18:1 or 18:2 fatty acids (Figure 3). The ability to utilize FFAs from the media and prevent lipotoxic effects confirmed functional activity for *Pa*DGAT1 (Figures 3A, B). *PaDGAT2* however, was unable to complement H1246 in its native form (data not shown) or when expressed from a yeast codon-optimized synthetic gene (Figures 3C, D). *Pa*PDAT1 also failed to overcome the lipotoxic effect when expressed in H1246 yeast strain (Figures 3E, F). Though not a direct assessment of substrate specificity, the FFA lipotoxicity assay can provide some insight to enzyme selectivity. *Arabidopsis* DGAT1 successfully rescued H1246 when the growth media was supplemented with oleic acid (a major component of polar and neutral lipids in *Arabidopsis*) but not with medium-chain saturated fatty acids that are not typically produced in this plant (Iskandarov et al., 2017). Similarly, human DGAT2 reverted the lipotoxicity in H1246 in the presence of unsaturated fatty acids (Garbarino and Sturley, 2009). The results shown here establish that *Pa*DGAT1 is an active acyltransferase and may be a factor in the enrichment of oleic and linoleic acids in avocado TAG.

Previous studies indicate that DGAT2 enzymes may require as-yet unknown additional factors for optimal expression and *in vitro* enzyme assay conditions (Stone et al., 2004; Yen et al., 2008). Likewise, other studies suggest that low steady-state protein levels for transgenic DGAT2s may contribute to this phenomenon when expressed in yeast and other heterologous cell types. *Arabidopsis* DGAT2 (Aymé et al., 2014; Chen et al., 2015) and fungal *Umbelopsis ramanniana* DGAT2 (Lardizabal et al., 2008) showed almost no DGAT activity in yeast and/or were undetectable by western blotting. Codon optimization is a frequent adaptation used to overcome this issue, with mixed results. Optimized *At*DGAT2 gained orders of magnitude of protein stability relative to its native counterpart (Aymé et al., 2014), but optimization did not improve *Pa*DGAT2 expression enough to rescue the lipotoxic phenotype in H1246 yeast. The cause of these negative results is not clear, but it is also possible that *Pa*DGAT2 may play a highly specialized role in a limited number of cell or tissue types. We also questioned whether a mutation in *PaDGAT2* could have resulted in a loss, or reduction, of activity or protein stability. Previous studies showed that evolutionary drift at phenylalanine 469 in maize DGAT1-2 affected its activity in maize (Zheng et al., 2008). Interestingly, we found that among a large group of plant DGAT2s covering dozens of species, only *Pa*DGAT2 contains a tyrosine at position 314, while all other DGAT2s contain phenylalanine (Supplementary Figure 7). We suspect that *Pa*DGAT2 may have undergone a point mutation resulting in a loss of function. Future site-directed mutagenesis studies aimed at replacing Tyr314 with Phe could reveal if a reversion at this residue could favorably alter *Pa*DGAT2 enzyme activity.

3.4 *Pa*DGAT1 encodes a protein with TAG biosynthetic activity

The survival and growth of *PaDGAT1*-expressing yeast strain H1246 in the presence of FFAs should correlate with accumulation of neutral lipids, including TAG, which are stored in LDs. Microscopic analysis of H1246 cells expressing *Pa*DGAT1 or *Vf*DGAT1 (positive control) confirmed production of LDs. On the contrary, negative control mutant cells failed to accumulate any neutral lipids (Figure 4A). TLC analyses of total lipids extracted from these cells confirmed restoration of TAG biosynthesis in both the positive control and *Pa*DGAT1 strains, but not the negative control (Figures 4B, C). These results directly confirm that *Pa*DGAT1 encodes a protein with TAG biosynthetic activity. Total cellular fatty acid composition was analyzed in transgenic yeast cells either grown without fatty acid supplementation or with either 18:1 or 18:2 added to the media. After 48 hours, yeast cells were harvested, and fatty acid

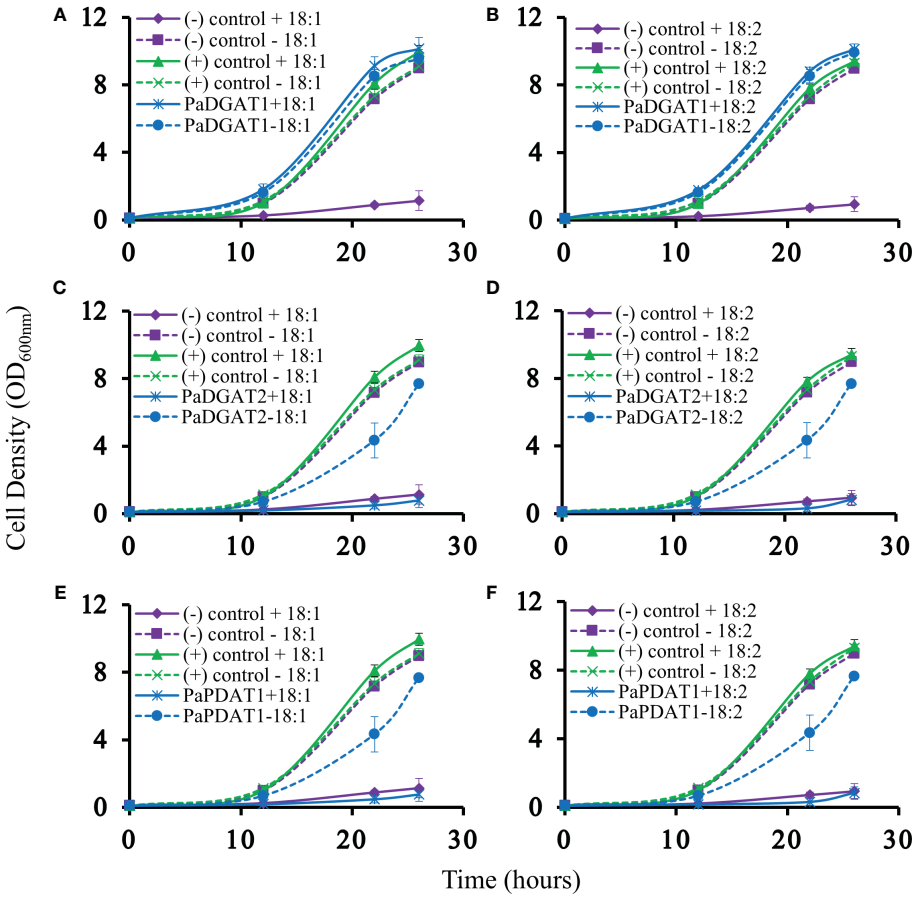


FIGURE 3
Lipotoxicity rescue complementation of H1246 yeast strain by *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1. Transformed TAG-deficient yeast strain expressing either *PaDGAT1*, *PaDGAT2*, or *PaPDAT1*, along with negative (H1246 containing empty pESC-URA vector) and positive (H1246 transformed with *VfDGAT1*) controls were challenged with 1 mM free oleic acid (18:1) (**A, C**, and **E,** respectively) and 1 mM free linoleic acid (18:2) (**B, D**, and **F**, respectively) in the growth media. Lipotoxicity assays were conducted as an indirect means to assess TAG biosynthetic competency of each enzyme. Only *Pa*DGAT1 complementation rescued H1246 yeast strain in the presence of 18:1 or 18:2, as indicated by growth curves similar to those of positive control.

composition was analyzed from total TAG content. In feeding studies, 16:0, 16:1, and 18:0 were produced by endogenous *de novo* fatty acid synthesis. Oleic acid can be produced directly by the yeast as well (in media lacking fatty acid supplementation) or may constitute a mixture of endogenous and exogenous oleate, in cultures fed with this fatty acid. Yeast does not have any intrinsic 18:2 biosynthetic capacity, therefore all 18:2 quantified from yeast comes only from the media supplement. In the absence of a fatty acid supplement, the fatty acid composition which is incorporated by *Pa*DGAT1 strains is dominated by 18:1, making up ~46% of the total fatty acids. Oleate-supplemented *Pa*DGAT1 cells contain ~70% 18:1. Interestingly, while linoleate supplementation increases 18:2 content to ~40%, *Pa*DGAT1 strains grown under this condition still retained nearly equivalent amount of 18:1, in this case drawn entirely from the endogenous fatty acid pool (Figure 4D).

3.5 *Pa*DGAT1 preferentially incorporates oleic acid (18:1) during TAG accumulation

The preferential accumulation of oleic acid in TAGs from transgenic *PaDGAT1*-expressing yeast was intriguing. To investigate the biochemical properties of this enzyme more directly, we performed *in vitro* enzyme assays using radiolabeled 16:0- or 18:1-CoAs as substrates (Figures 5A, B). Our results suggest that *Pa*DGAT1 utilizes both 16:0 and 18:1, but it showed a two-fold higher preference for 18:1-CoA than 16:0-CoA (Figure 5C). The substrate preference for DGAT1 in plants differs depending on the species and tissue types and on the availability of the substrates in that species (Shockey et al., 2006). Plant DGAT1 has a wide range of substrate specificities from medium to long-chain, and saturated, monounsaturated, or

polyunsaturated fatty acids. Ancestral maize DGAT1-2 had a high preference for oleic acid but partially lost specificity and enzyme efficiency upon loss of phenylalanine-469 during the breeding and domestication of this crop (Zheng et al., 2008). *Brassica napus* DGAT1 also had a higher preference for 18:1 compared to 16:0. The high preference of *Bn*DGAT1 for 18:1 positively correlates with the seed fatty acid composition (Aznar-Moreno et al., 2015). Similarly, in avocado, the preference of DGAT1 towards 18:1 correlates with the fatty acid composition of avocado mesocarp (Kilaru et al., 2015).

3.6 Transient expression of *PaDGAT1* and *PaPDAT1* in *N. benthamiana* leaves leads to TAG accumulation

Although our yeast transgene expression studies using *Pa*PDAT1 failed to confirm a functional role in TAG biosynthesis, based on its high expression level in the mesocarp (Kilaru et al., 2015), we still suspected that it might be involved in avocado oil metabolism. Hence, we further assessed *PaDGAT1* and *PaPDAT1* for their functional role through *in planta* expression of these genes in *N. benthamiana* leaves. The 'benth leaf system', as it is commonly known, has become a valuable experimental system, due to its speed and ease of use, thus providing a plant cell environment that can be more conducive to studies of plant enzymes (Grimberg et al., 2015; Reynolds et al., 2015a; Delatte et al., 2018). Among many other plant DGATs and other lipid metabolic enzymes, *Arabidopsis* DGAT1 was expressed in this system and dramatically increased the number of LDs produced (Vanhercke et al., 2014; Vanhercke et al., 2017). Co-expression of *Arabidopsis* DGAT1 with the *WRINKLED1* transcription factor increased TAG content in benth leaves up to 15% of dry weight (Vanhercke et al., 2013; Reynolds et al., 2015b). Thus, we transiently expressed *PaDGAT1* and *PaPDAT1* in benth leaves, individually or combined with the viral silencing suppressor protein P19 (Wood et al., 2009), under the control of the *CaMV35S* promoter (Figure 6A) and observed accumulated LDs by confocal fluorescence microscopy. We observed that *Pa*DGAT1 produced elevated LD counts compared to wild-type, which provided further evidence for TAG biosynthesis by this enzyme (Figure 6B). The number of LDs produced by *Pa*DGAT1 in *N. benthamiana* leaves was almost 10-fold and 2-fold higher than those in mock and P19 negative controls (Students' T-test, $P< 0.01$ and $P<0.05$, respectively) (Figure 6C). The combination of *Pa*DGAT1 with P19 protein had a more robust effect, where the number of LDs increased by 37-fold and 7-fold compared to the wild-type ($P<0.001$) and P19 control ($P<0.001$), respectively (Figure 6C). A similar trend was observed when *PaPDAT1* was transiently expressed in benth leaves, where the number of LDs increased by 4-fold and 2-fold compared to wild-type control and *P19* control, respectively. Co-infiltration of *PaPDAT1* with *P19* resulted in an elevation of LDs by 22-fold and 4-fold compared to wild-type ($P<0.001$) and P19 control ($P<0.001$), respectively (Figure 6D). The P19 viral silencing suppressor is included here to help suppress transgene silencing, a common problem in both transiently and stably transformed plant cells (Naim et al., 2016). It is not completely clear why *P19* expression alone can increase LD production in benth leaves, but it is possible that P19 protein affected the expression levels of endogenous enzymes to a certain extent. In any case, the increase in LD count induced by *Pa*DGAT1 and *Pa*PDAT1, relative to P19 control confirm acyltransferase activity and strongly suggest that *Pa*DGAT1 and *Pa*PDAT1 are functional and thus, responsible for TAG production in nonseed avocado tissues.

In addition to the observed changes in LD numbers in *N. benthamiana* cells expressing avocado acyltransferases, we directly determined the change in total lipid content in benth leaves expressing *PaDGAT1* or *PaPDAT1*. We extracted total lipid from infiltrated leaves expressing *PaDGAT1, PaPDAT1* or controls and separated them by TLC (Figures 7A, C). The expression of *PaDGAT1* significantly increased (>2-fold) total lipid content compared to either wild-type or P19 controls ($P<0.01$) (Figure 7B). Likewise, *PaPDAT1* expression increased total lipid content by ~2-fold (Figure 7B). These results indicate that transient overexpression of *PaDGAT1* and *PaPDAT1* increased total lipid content in benth leaves specifically by making more TAGs.

3.7 *Pa*DGAT1 and *Pa*PDAT1 show a preference for oleic acid when expressed in *N. benthamiana* leaves

To gain more insight into the substrate specificities of *Pa*DGAT1 and *Pa*PDAT1 expressed in plant vegetative tissues, we examined the fatty acid profile in *N. benthamiana* leaves by GC-FID. The content of 16:0, 16:1, and 18:3 did not change significantly in response to *PaDGAT1* or *PaPDAT1* expression (Supplementary Table 1), compared to the wild-type or P19 controls. However, *Pa*DGAT1 activity significantly increased oleic acid content compared to wild-type ($P<0.05$); viral silencing suppressor protein *P19* expression enhanced this effect (vs WT controls, $P<0.001$; vs P19 controls, $P<0.05$) (Figure 7D). *PaPDAT1* produced similar results (Figure 7D). Expression of both avocado acyltransferases significantly increased 18:2 levels over wild-type controls, but neither *Pa*DGAT1 nor *Pa*PDAT1 activity elevated linoleic acid content over the P19 background levels (Figure 7E).

Our results for *Pa*DGAT1 were consistent with those of Vanhercke et al. (2014), who reported that *Arabidopsis* DGAT1 enhanced oleic acid levels in *N. tabacum* leaves. Given the unsuccessful detection of *Pa*PDAT1 activity in yeast cells, transient expression of this enzyme in *N. benthamiana* leaves

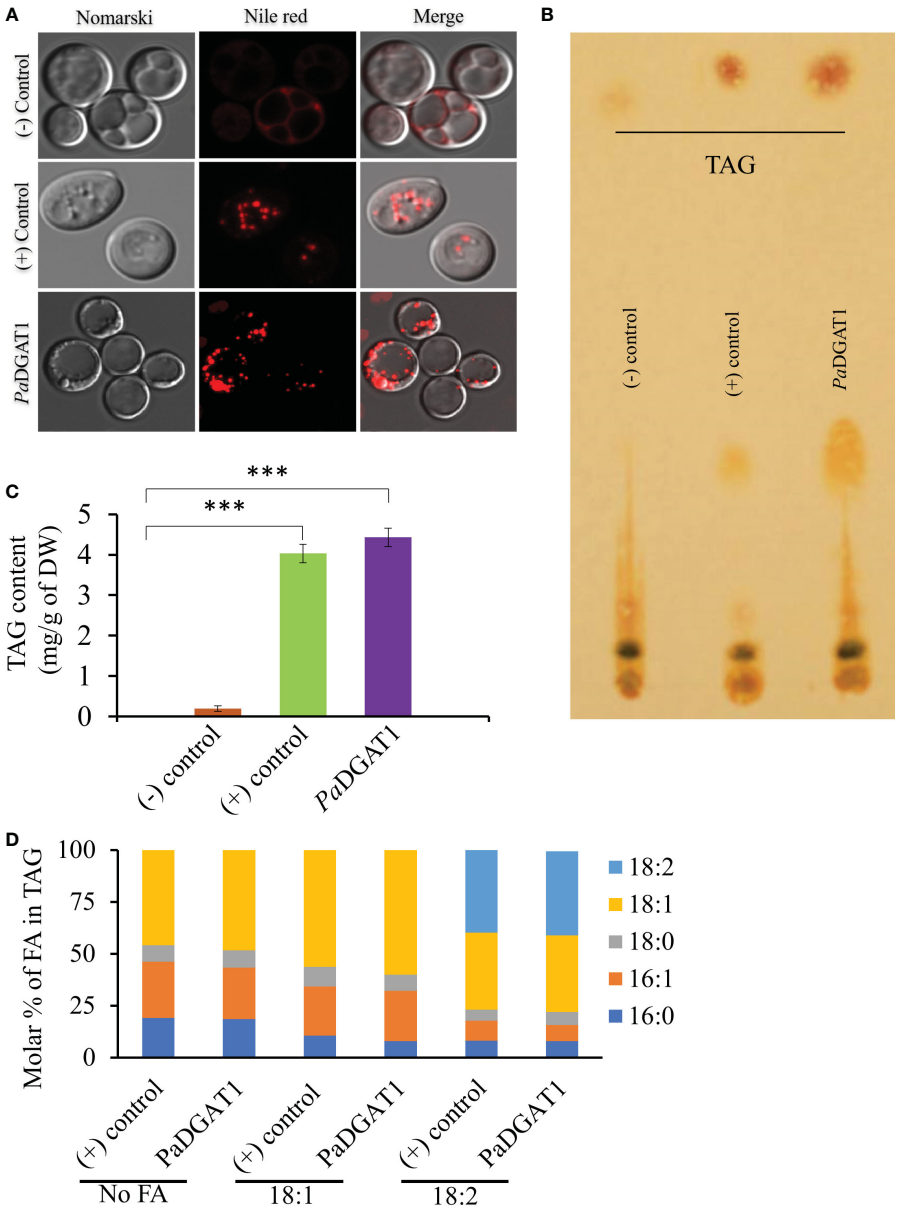


FIGURE 4
Observation of newly synthesized lipid droplets (LD) in complemented yeast strain and determination of TAG content and composition. TAG-deficient yeast cells expressing *PaDGAT1* restored LD formation, as visualized by staining the cells with the lipid-specific fluorescent dye Nile Red **(A)**. The neutral lipid deficient quadruple mutant H1246 strain harboring *VfDGAT1* was used as a positive control. The quadruple mutant H1246 strain containing empty plasmid was used as a negative control. Nile Red fluorescence was observed with an excitation of 488 nm and an emission of 600-650 nm wavelength. Lipids were extracted from yeast cultures grown under protein expression inducing conditions for 24 h, then separated by TLC **(B)**. Fatty acid content **(C)** and composition **(D)** were quantified by GC-FID for cultures grown in media containing either no fatty acid supplementation, or 1 mM oleic acid (18:1) or linoleic acid (18:2). Experiments were done in triplicate and the data are mean ± SD, n=3. ***P<0.001.

provided the first direct evidence of TAG biosynthetic activity and suggested a similar preference for oleic acid-containing substrates as seen with *Pa*DGAT1; further validation by *in vitro* PDAT enzyme activity assays are still warranted. Although *Arabidopsis* PDAT was active with a broad spectrum of substrates containing acyl chains of varying lengths, it showed strong preferences for hydroxy, epoxy, and polyunsaturated fatty acids (Ståhl et al., 2004), suggesting that this class of enzymes may also play a role in membrane surveillance and turnover of oxidized or otherwise damaged fatty acids. For example, *Camelina sativa* PDAT1-A increased polyunsaturated fatty acids (18:2 and 18:3) compared to wild-type control in two

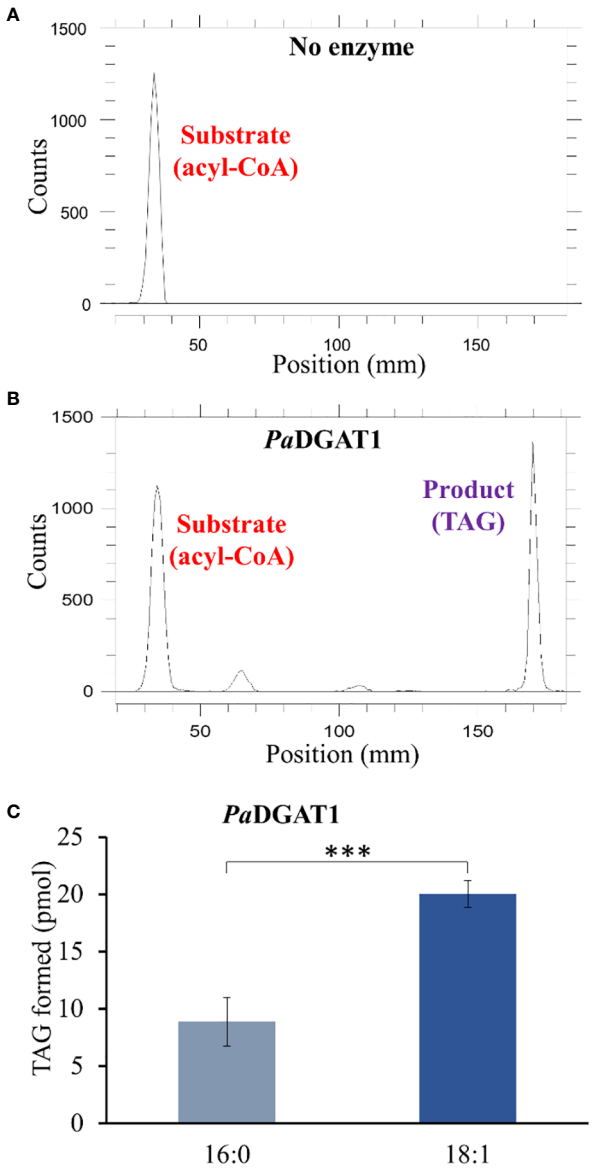


FIGURE 5
In vitro enzyme activity and specificity assays for yeast expressing PaDGAT1. PaDGAT1 with an N-terminal HA epitope tag was expressed in the H1246 yeast strain and was induced for 24 h. Microsomal fractions were prepared and incubated with radiolabeled fatty acyl-CoA and DAG. Saturated (palmitoyl-CoA, 16:0-CoA) and monounsaturated (oleoyl-CoA, 18:1-CoA) acyl donor substrates were tested. Representative chromatogram generated by Radio-TLC imaging scanner for negative control with only acyl-CoA substrate and no enzyme **A)**, and PaDGAT1 microsomes incubated with 16:0 or 18:0 acyl-CoA and diolein that resulted in TAG synthesis **(B)**. Quantification of TAG product peaks revealed that PaDGAT1 prefers 18:1-CoA over 16:0-CoA **(C)**. ***P<0.001.

independent studies (Marmon et al., 2017; Yuan et al., 2017). Similarly, castor bean *Rc*PDAT1A showed specificity toward the hydroxy fatty acid ricinoleic acid (van Erp et al., 2011). Interestingly, while flax PDAT1 showed strong preference for 18:3, PDAT2 preferred other polyunsaturated fatty acids (Pan et al., 2013). These camelina, castor, and flax results show positive correlation between PDAT activity and the dominant fatty acid present in their respective seed oils. The different substrate preferences by various PDAT paralogs and orthologs indicate its divergent functional role in plants. In the case of avocado, the coordinate evolution towards oleic acid preference by *Pa*DGAT1 and *Pa*PDAT1 strongly suggests a role for both enzymes in controlling the final fatty acid profile of its mesocarp oils.

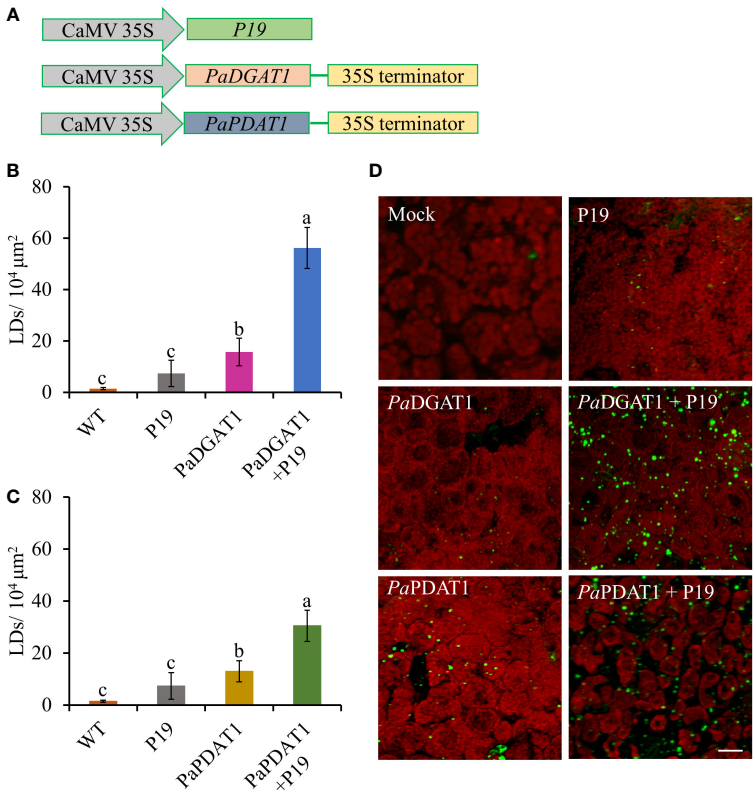


FIGURE 6
Vector constructs, LD visualization, and quantification of LDs in *N. benthamiana* leaves expressing *Pa*DGAT1 and *Pa*PDAT1. Schematic diagrams for P19, *PaDGAT1*, and *PaPDAT1* vector constructs used for transient *Agrobacterium* transformation of *N. benthamiana* leaves **(A)**. Quantification of LDs in different *N. benthamiana* leaves expressing *PaDGAT1* **(B)** and *PaPDAT1* **(C)**. Data represent mean ± SD of three independent experiments and different letters indicate significant differences ($P < 0.05$), as determined by ANOVA with Tukey's post-test. Phenotypic visualization of LDs (in green color) in *N. benthamiana* leaves expressing either *PaDGAT1* or *PaPDAT1* alone or with P19 **(D)**. LDs were stained with Nile Red then visualized by confocal microscopy; leaves with mock infiltration or p19 expression alone are used as negative controls. Scale bar represents 20 µm.

3.8 Transient co-expression of *PaDGAT1* and *PaPDAT1* further improved TAG production in *N. benthamiana* leaves

Since both DGAT1 and PDAT1 in avocado are functional in TAG biosynthesis, we hypothesized that the continuous high expression of both genes in mesocarp tissue cooperatively achieves high oleic acid accumulation in TAG. Although DGAT1 and PDAT1 have overlapping functions in at least some seeds (Zhang et al., 2009), their relative contribution to the TAG pool is highly variable and is species-specific (Banaś et al., 2013). Here we assessed the combined effect of *Pa*DGAT1 and *Pa*PDAT1 on TAG biosynthesis in nonseed tissue, which, to our knowledge, has not been explored for other plant species. We co-expressed *PaDGAT1* and *PaPDAT1* along with *P19* in *N. benthamiana* leaves; *P19* expression alone was used as the control. Interestingly, *DGAT1 +PDAT1+P19* co-expression increased the LD accumulation in the leaf tissue by ~3-fold and ~4-fold compared to independent expression of *DGAT1+P19* and *PDAT1+P19*, respectively (Figures 8A, C). The increase in LD count was more than additive, i.e., the co-expression resulted in a greater LD accumulation compared to the sum of LDs shown by DGAT1 and PDAT1. The LD accumulation in animals and plants is a dynamic process and may be triggered to maintain energy homeostasis, regulate membrane remodeling and as a response to stress (Chitraju et al., 2017; Pyc et al., 2017; Welte and Gould, 2017; Yang and Benning, 2018; Fernández-Santos et al., 2020). It is possible that the elevated LD count occurs in part due to the stress induced by the *DGAT1*/*PDAT1* co-expression, as a reduction in chlorophyll was noticeable (Figure 8A). Co-expression also increased average LD size (area) increase by >2-fold and >4-fold compared to single expression of *DGAT1* and *PDAT1*, respectively (Figure 8D), suggesting improved LD production and packaging efficiency. However, there was no significant improvement in the total lipid content in leaves (Figure 8B), which might suggest that co-expression did not increase *de novo* fatty acid biosynthesis but contributed to elevated TAG assembly. Previously, an increase in the PC pool upon *At*DGAT1 expression in tobacco leaves was

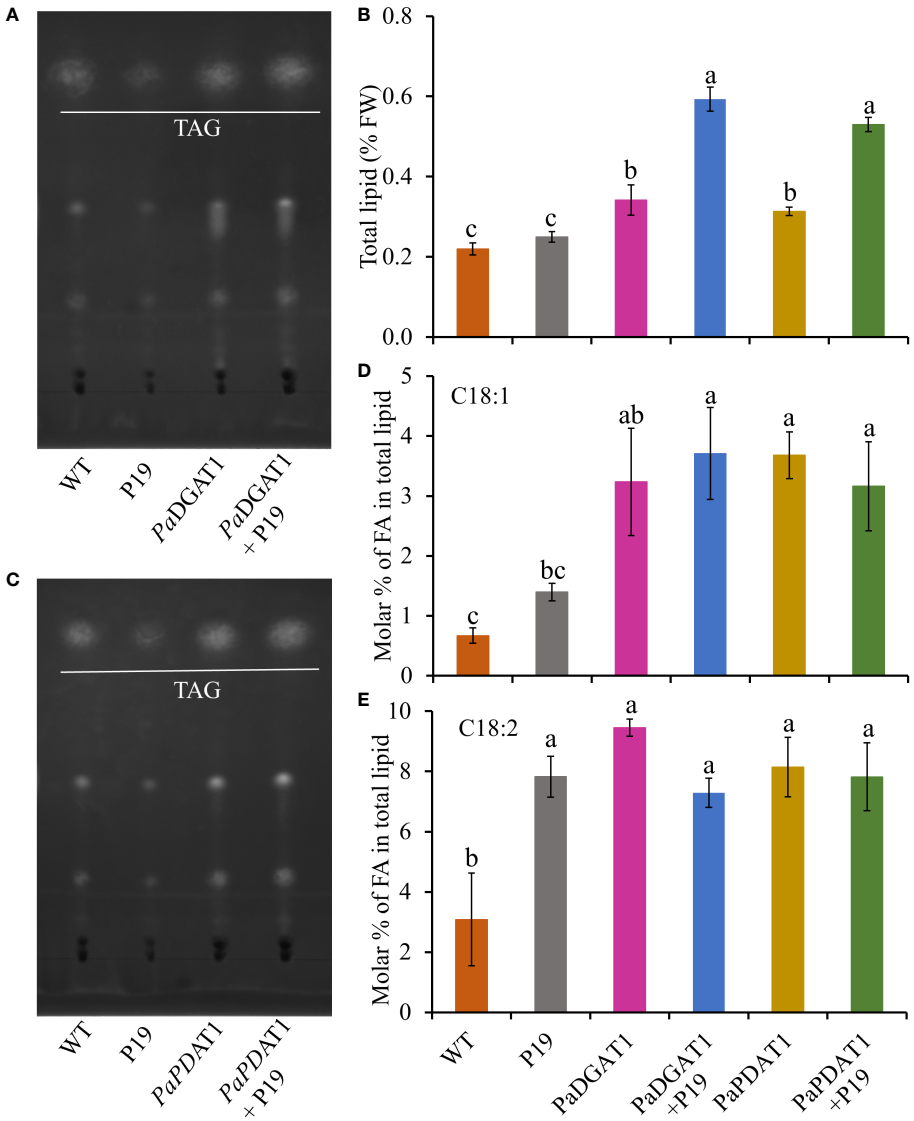


FIGURE 7
Quantification of total lipid content in *N. benthamiana* leaves expressing *PaDGAT1 or PaPDAT1*. Separation of total lipid by TLC from *N. benthamiana* leaves expressing *PaDGAT1* **(A)** or *PaPDAT1* **(C)**. Quantification of total lipid from benth leaves expressing *PaDGAT1* or *PaPDAT1*, **(B)**. Fatty acid composition of total lipid content was analyzed and quantified by GC-FID in benth leaves expressing *PaDGAT1* or *PaPDAT1* with or without P19 coexpression. The molar percentage composition for C18:1 **(D)** and C18:2 **(E)** is shown. Data represent mean ± SD of three independent experiments and different letters indicate significant differences ($P < 0.05$), as determined by ANOVA with Tukey's post-test.

reported (Andrianov et al., 2010; Vanhercke et al., 2013). Similarly, such an increase in PC might be used as a substrate for utilization by *Pa*PDAT1 contributing to TAG assembly. Recently it was shown that *At*DGAT1 and *At*PDAT1 physically interact with each other both *in planta* and in yeast two hybrid system to maximize TAG biosynthesis (Lee and Seo, 2019; Regmi et al., 2020). Possibly, such interaction between *Pa*DGAT1 and *Pa*PDAT1 might result in a rapid carbon flux into TAG in this micro-environment, which further leads to LD size increase. Further experimental verification is needed to confirm avocado DGAT1/PDAT1 interactions.

Since both *Pa*DGAT1 and *Pa*PDAT1 preferred oleic acid as substrate in our previous experiments, we further analyzed the fatty acid profile of TAG to see if there is any further improvement in oleic acid content. TAG was separated from the total lipid pool by TLC (Figure 8B), methylated and analyzed by GC-FID. Our results suggest an increase in the proportion of oleic acid in the TAG by >30% and >50% in the *DGAT1+PDAT1+P19*-expressed leaves compared to the leaves expressing *DGAT1+P19* and *PDAT1+P19*, respectively (Figure 8E; Supplementary Table 2). However, the improvement was statistically significant compared to only the *PDAT1+P19*.

Nonetheless, the results suggest that the synchronous activity of DGAT1 and PDAT1 in avocado mesocarp is responsible for the high proportion of oleic acid in the synthesized TAG since it was significantly (by ~2.4 fold) increased compared to the P19 control. Interestingly, 18:2 content in the TAG from *DGAT1+PDAT1+P19*-expressed leaves is similar to that from *P19* control leaves, although it is significantly reduced in *PDAT1+P19*-expressed leaves (Figure 8E). This suggests that combination of *Pa*DGAT1 and *Pa*PDAT1 is a better strategy to specifically improve oleic acid content without affecting other aspects of fatty acid composition.

4 Conclusions

We have identified and functionally characterized *PaDGAT1* and *PaPDAT1* by complementation of yeast quadruple mutant H1246

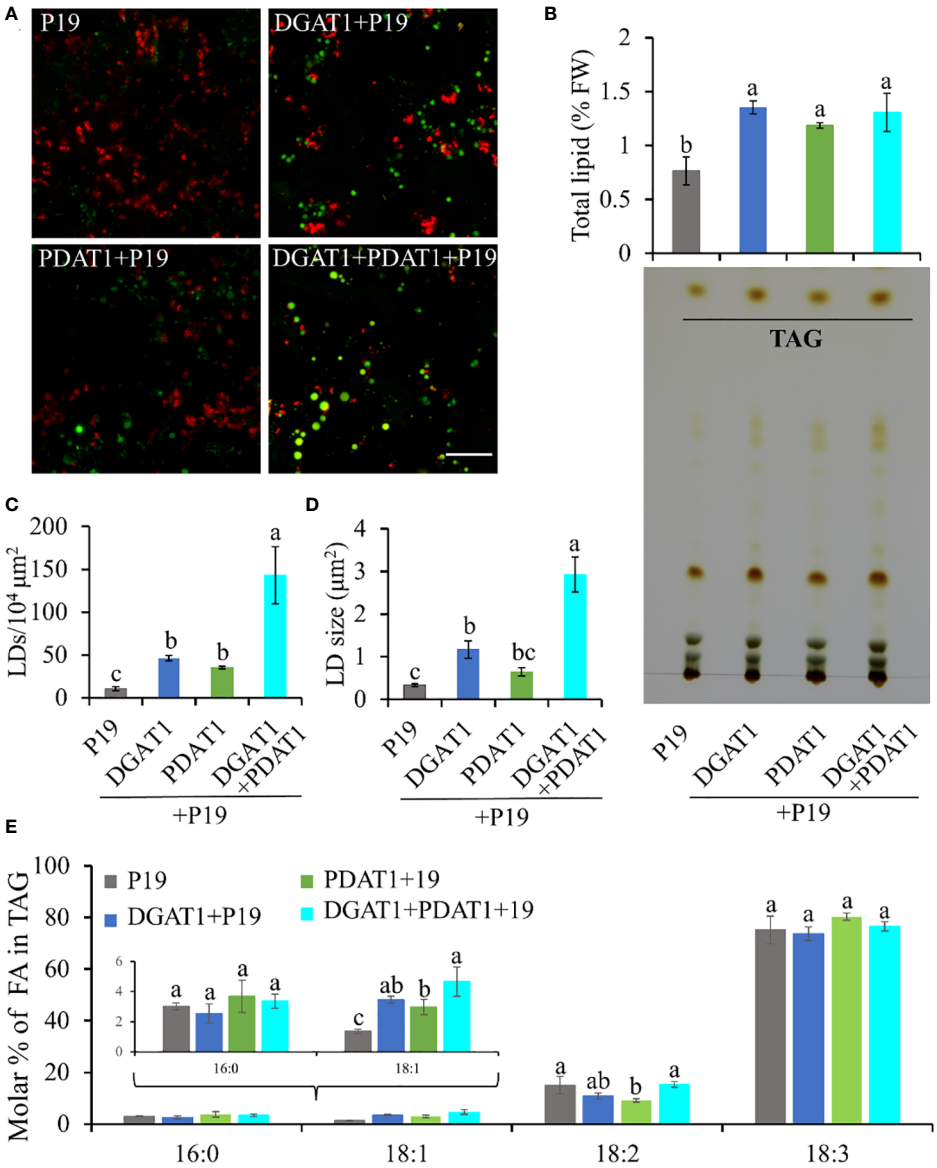


FIGURE 8
Quantification of LD accumulation and fatty acid profile in TAG in N. benthamiana leaves co-expressing PaDGAT1 and PaPDAT1. Confocal images of the accumulated LDs stained with Nile Red (green) in N. benthamiana leaves expressing control (+P19), DGAT1 (+P19), PDAT1 (+P19), and DGAT1 + PDAT1 (+P19) **(A)**. The scale bar represents 20 µm. Quantification of total lipid extracted from the leaves and separation of TAG by TLC **(B)**. Number of accumulated LDs per unit area **(C)** and their average size (area) **(D)**. Fatty acid profile of extracted TAG from the TLC plates **(E)**. Data represent mean ± SD of three independent experiments and different letters indicate significant differences ($P < 0.05$), as determined by ANOVA with Tukey's post-test.

and/or transient expression in *N. benthamiana* leaves. Heterologous expression and *in vitro* enzyme activity confirmed the diacylglycerol acyltransferase activity of *PaDGAT1* and its preference for oleic acid-containing substrates. Transient expression of *PaDGAT1* and *PaPDAT1* in *N. benthamiana* leaves also revealed the acyltransferase activity and substrate specificity for oleic acid. The present findings suggest that *PaDGAT1* and *PaPDAT1* cooperatively play roles in TAG assembly and provide new insight into the underlying mechanism of lipid biosynthesis in avocado mesocarp. The results obtained from this study will enrich our knowledge towards an understanding of lipid biosynthesis in nonseed tissue and can be translated into other plant species or less complex organisms, such as oleaginous microbes to produce oil enriched in oleic acid.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: GenBank OP727298, OP727299 and OP727300.

Author contributions

MR, JS, JB and AK designed the research. MR, JB and JS conducted the experiments in yeast and *N. benthamiana*. JB and MR conducted the bioinformatics analyses. JB, MR, JS, and AK conducted the data analysis. JB and MR wrote the manuscript and JS and AK edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

MR received Sigma Xi GIAR Award. Both JB and MR received East Tennessee State University School of Graduate Studies Research Grant. AK was supported in part by minor and major grants from the Research and Development Committee, East Tennessee State University and the National Institute of Food and Agriculture (GRANT13058738).

Acknowledgments

The authors are thankful to Prof. Sten Stymne (Swedish University of Agricultural Sciences) for providing the H1246 yeast strain. We also thank Ms. Catherine Mason, U.S. Department of Agriculture, Agricultural Research Service, Southern Regional Research Center, for technical assistance.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1056582/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Phylogenetic analysis and percentage sequence similarities of avocado DGAT1, DGAT2, and PDAT1 to orthologous plant proteins. Phylogenetic analysis using MEGA X software.

SUPPLEMENTARY FIGURE 2
Predicted avocado DGAT1, DGAT2, and PDAT2 membrane topologies by TMHMM and Phobius.

SUPPLEMENTARY FIGURE 3
Predicted intrinsically disordered region (IDR) in the *N*-terminal region of *Pa*DGAT1.

SUPPLEMENTARY FIGURE 4
Predicted three-dimensional (3D) structures of *Pa*DGAT2 by the SWISS-MODEL interactive workspace tools, using the solved or predicted models for *Thermotoga maritima* 1-acyl-sn-glycerol-3-phosphate acyltransferase (PlsC) (PDB Id: 5kym). The 3D structures of *Pa*DGAT2 **(A)** and PlsC **(B)** followed by their structural comparison **(C)** showing the *Pa*DGAT2 in cyan, PlsC in red and Hx4D domain in blue, respectively.

SUPPLEMENTARY FIGURE 5
The quality of the predicted 3D structures based on the structure assessment. The Ramachandran plot showing the bond angles (Φ and Ψ) for amino acids at each position, the quality of the predicted structure assessed by QMEANDisCo and MolProbity for *Pa*DGAT1 **(A)**, *Pa*DGAT2 **(B)**, and *Pa*PDAT1 **(C)** are shown.

SUPPLEMENTARY FIGURE 6
RMSD values of structure-based sequence alignments of *Pa*DGAT1, *Pa*DGAT2, and *Pa*PDAT1 with their respective templates.

SUPPLEMENTARY FIGURE 7
Sequence alignment of *C*-terminal regions of additional plant DGAT2 proteins. The invariant Phe residue that has been changed to a Tyr residue in avocado DGAT2 (shown in red) is boxed.

References

Andrianov, V., Borisjuk, N., Pogrebnyak, N., Brinker, A., Dixon, J., Spitsin, S., et al. (2010). Tobacco as a production platform for biofuel: overexpression of arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. *Plant Biotechnol. J.* 8, 277–287. doi: 10.1111/j.1467-7652.2009.00458.x

Aymé, L., Arragain, S., Canonge, M., Baud, S., Touati, N., Bimai, O., et al. (2018). Arabidopsis thaliana DGAT3 is a [2Fe-2S] protein involved in TAG biosynthesis. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-35545-7

Aymé, L., Baud, S., Dubreucq, B., Joffre, F., and Chardot, T. (2014). Function and localization of the arabidopsis thaliana diacylglycerol acyltransferase DGAT2 expressed in yeast. *PloS One* 9, e92237. doi: 10.1371/journal.pone.0092237

Aymé, L., Jolivet, P., Nicaud, J. M., and Chardot, T. (2015). Molecular characterization of the elaeis guineensis medium-chain fatty acid diacylglycerol acyltransferase DGAT1-1 by heterologous expression in yarrowia lipolytica. *PloS One* 10, e0143113. doi: 10.1371/journal.pone.0143113

Aznar-Moreno, J., Denolf, P., Van Audenhove, K., De Bodt, S., Engelen, S., Fahy, D., et al. (2015). Type 1 diacylglycerol acyltransferases of brassica napus preferentially incorporate oleic acid into triacylglycerol. *J. Exp. Bot.* 66, 6497–6506. doi: 10.1093/jxb/erv363

Banaś, A., Dahlqvist, A., Ståhl, U., Lenman, M., and Stymne, S. (2000). The involvement of phospholipid:diacylglycerol acyltransferases in triacylglycerol production. *Biochem. Soc Trans.* 28, 703–705. doi: 10.1042/bst0280703

Banaś, W., Sanchez Garcia, A., Banaś, A., and Stymne, S. (2013). Activities of acyl-CoA:Diacylglycerol acyltransferase (DGAT) and phospholipid:Diacylglycerol acyltransferase (PDAT) in microsomal preparations of developing sunflower and safflower seeds. *Planta* 237, 1627–1636. doi: 10.1007/s00425-013-1870-8

Bates, P. D., Durrett, T. P., Ohlrogge, J. B., and Pollard, M. (2009). Analysis of acyl fluxes through multiple pathways of triacylglycerol synthesis in developing soybean embryos. *Plant Physiol.* 150, 55–72. doi: 10.1104/pp.109.137737

Bates, P. D., Ohlrogge, J. B., and Pollard, M. (2007). Incorporation of newly synthesized fatty acids into cytosolic glycerolipids in pea leaves occurs *via* acyl editing*. *J. Biol. Chem.* 282, 31206–31216. doi: 10.1074/jbc.M705447200

Behera, J. R., Rahman, M. M., Bhatia, S., Shockey, J., and Kilaru, A. (2021). Functional and predictive structural characterization of WRINKLED2, a unique oil biosynthesis regulator in avocado. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.648494

Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G. F., Drira, N., Ohlrogge, J. B., et al. (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning (Proceedings of the national academy of sciences of the united states of America, (2011) 108, 30 (12527-12532) DOI. *Proc. Natl. Acad. Sci. U. S. A.* 108, 18186. doi: 10.1073/pnas.1115243108

Bouvier-Navé, P., Benveniste, P., Oelkers, P., Sturley, S. L., and Schaller, H. (2000). Expression in yeast and tobacco of plant cDNAs encoding acyl CoA: diacylglycerol acyltransferase. *Eur. J. Biochem.* 267, 85–96. doi: 10.1046/j.1432-1327.2000.00961.x

Browse, J., McCourt, P., and Somerville, C. (1986). A mutant of arabidopsis deficient in c 18:3 and c 16:3 leaf lipids. *Plant Physiol.* 81, 859–864. doi: 10.1104/pp.81.3.859

Burgal, J., Shockey, J., Lu, C., Dyer, J., Larson, T., Graham, I., et al. (2008). Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. *Plant Biotechnol. J.* 6, 819–831. doi: 10.1111/j.1467-7652.2008.00361.x

Caldo, K. M. P., Acedo, J. Z., Panigrahi, R., Vederas, J. C., Weselake, R. J., and Lemieux, M. J. (2017). Diacylglycerol acyltransferase 1 is regulated by its n-terminal domain in response to allosteric effectors. *Plant Physiol.* 175, 667–680. doi: 10.1104/pp.17.00934

Cao, H. (2011). Structure-function analysis of diacylglycerol acyltransferase sequences from 70 organisms. *BMC Res. Notes* 4, 249. doi: 10.1186/1756-0500-4-249

Cao, H., Shockey, J. M., Klasson, K. T., Chapital, D. C., Mason, C. B., and Scheffler, B. E. (2013). Developmental regulation of diacylglycerol acyltransferase family gene expression in tung tree tissues. *PloS One* 8, e76946. doi: 10.1371/journal.pone.0076946

Cases, S., Smith, S. J., Zheng, Y. W., Myers, H. M., Lear, S. R., Sande, E., et al. (1998). Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 95, 13018–13023. doi: 10.1073/pnas.95.22.13018

Cases, S., Stone, S. J., Zhou, P., Yen, E., Tow, B., Lardizabal, K. D., et al. (2001). Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members *. *J. Biol. Chem.* 276, 38870–38876. doi: 10.1074/jbc.M106219200

Chen, Y., Cui, Q., Xu, Y., Yang, S., Gao, M., and Wang, Y. (2015). Effects of tung oilseed FAD2 and DGAT2 genes on unsaturated fatty acid accumulation in rhodotorula glutinis and arabidopsis thaliana. *Mol. Genet. Genomics* 290, 1605–1613. doi: 10.1007/s00438-015-1011-0

Chen, G., Harwood, J. L., Lemieux, M. J., Stone, S. J., and Weselake, R. J. (2022). Acyl-CoA:diacylglycerol acyltransferase: Properties, physiological roles, metabolic engineering and intentional control. *Prog. Lipid Res.* 88, 101181. doi: 10.1016/j.plipres.2022.101181

Chen, H. C., Smith, S. J., Tow, B., Elias, P. M., and Farese, J. R.V. (2002). Leptin modulates the effects of acyl CoA:diacylglycerol acyltransferase deficiency on murine fur and sebaceous glands. *J. Clin. Invest.* 109, 175–181. doi: 10.1172/JCI13880

Chitraju, C., Mejhert, N., Haas, J. T., Diaz-Ramirez, L. G., Grueter, C. A., Imbriglio, J. E., et al. (2017). Triglyceride synthesis by DGAT1 protects adipocytes from lipid-induced ER stress during lipolysis. *Cell Metab.* 26, 407–418.e3. doi: 10.1016/j.cmet.2017.07.012

Dahlqvist, A., Ståhl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., et al. (2000). Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6487–6492. doi: 10.1073/pnas.120067297

Delatte, T. L., Scaiola, G., Molenaar, J., de Sousa Farias, K., Alves Gomes Albertti, L., Busscher, J., et al. (2018). Engineering storage capacity for volatile sesquiterpenes in nicotiana benthamiana leaves. *Plant Biotechnol. J.* 16, 1997–2006. doi: 10.1111/pbi.12933

Donald, J. E., Kulp, D. W., and DeGrado, W. F. (2011). Salt bridges: geometrically specific, designable interactions. *Proteins* 79, 898–915. doi: 10.1002/prot.22927

Dong, S., Huang, J., Li, Y., Zhang, J., Lin, S., and Zhang, Z. (2014). Cloning, characterization, and expression analysis of acyl-acyl carrier protein (ACP)-thioesterase b from seeds of Chinese spicehush (Lindera communis). *Gene* 542, 16–22. doi: 10.1016/j.gene.2014.03.028

Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208. doi: 10.1038/nrm1589

Falarz, L. J., Xu, Y., Caldo, K. M. P., Garroway, C. J., Singer, S. D., and Chen, G. (2020). Characterization of the diversification of phospholipid:diacylglycerol acyltransferases in the green lineage. *Plant J.* 103, 2025–2038. doi: 10.1111/tpj.14880

Fernández-Santos, R., Izquierdo, Y., López, A., Muñiz, L., Martínez, M., Cascón, T., et al. (2020). Protein profiles of lipid droplets during the hypersensitive defense response of arabidopsis against pseudomonas infection. *Plant Cell Physiol.* 61, 1144–1157. doi: 10.1093/pcp/pcaa041

Garbarino, J., and Sturley, S. L. (2009). Saturated with fat: new perspectives on lipotoxicity. *Curr. Opin. Clin. Nutr. Metab. Care* 12, 110–116. doi: 10.1097/MCO.0b013e32832182ee

Ghosal, A., Banas, A., Ståhl, U., Dahlqvist, A., Lindqvist, Y., and Stymne, S. (2007). Saccharomyces cerevisiae phospholipid:diacylglycerol acyl transferase (PDAT) devoid of its membrane anchor region is a soluble and active enzyme retaining its substrate specificities. *Biochim. Biophys. Acta* 1771, 1457–1463. doi: 10.1016/j.bbalip.2007.10.007

Gidda, S. K., Watt, S. C., Collins-Silva, J., Kilaru, A., Arondel, V., Yurchenko, O., et al. (2013). Lipid droplet-associated proteins (LDAPs) are involved in the compartmentalization of lipophilic compounds in plant cells. *Plant Signal. Behav.* 8(11), e27141. doi: 10.4161/psb.27141

Glukhova, A., Hinkovska-Galcheva, V., Kelly, R., Abe, A., Shayman, J. A., and Tesmer, J. J. G. (2015). Structure and function of lysosomal phospholipase A2 and lecithin:cholesterol acyltransferase. *Nat. Commun.* 6, 6250. doi: 10.1038/ncomms7250

Grimberg, Å., Carlsson, A. S., Marttila, S., Bhalerao, R., and Hofvander, P. (2015). Transcriptional transitions in nicotiana benthamiana leaves upon induction of oil synthesis by WRINKLED1 homologs from diverse species and tissues. *BMC Plant Biol.* 15, 192. doi: 10.1186/s12870-015-0579-1

Guihéneuf, F., Leu, S., Zarka, A., Khozin-Goldberg, I., Khalilov, I., and Boussiba, S. (2011). Cloning and molecular characterization of a novel acyl-CoA: diacylglycerol acyltransferase 1-like gene (PtDGAT1) from the diatom phaeodactylum tricornutum. *FEBS J.* 278, 3651–3666. doi: 10.1111/j.1742-4658.2011.08284.x

Guo, Z., Cromley, D., Billheimer, J. T., and Sturley, S. L. (2001). Identification of potential substrate-binding sites in yeast and human acyl-CoA sterol acyltransferases by mutagenesis of conserved sequences. *J. Lipid Res.* 42, 1282–1291. doi: 10.1016/s0022-2275(20)31579-0

Han, C., and Xianlin, W. W. (2010) *Lipid analysis : isolation, separation, identification and lipidomic analysis*. Available at: http://www.sciencedirect.com/science/book/9780955251245.

Hara, A., and Radin, N. S. (1978). Lipid extraction of tissues with a low-toxicity solvent. *Anal. Biochem.* 90, 420–426. doi: 10.1016/0003-2697(78)90046-5

Hernández, M. L., Whitehead, L., He, Z., Gazda, V., Gilday, A., Kozhevnikova, E., et al. (2012). A cytosolic acyltransferase contributes to triacylglycerol synthesis in sucrose-rescued arabidopsis seed oil catabolism mutants. *Plant Physiol.* 160, 215–225. doi: 10.1104/pp.112.201541

He, X. H., Turner, C., Chen, G. Q., Lin, J. T., and McKeon, T. A. (2004). Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. *Lipids* 39(4), 311–318. doi: 10.1007/s11745-004-1234-2

Hobbs, D. H., Lu, C. F., and Hills, M. J. (1999). Cloning of a cDNA encoding diacylglycerol acyltransferase from arabidopsis thaliana and its functional expression. *FEBS Lett.* 452, 145–149. doi: 10.1016/S0014-5793(99)00646-8

Horn, P. J., James, C. N., Gidda, S. K., Kilaru, A., Dyer, J. M., Mullen, R. T., et al. (2013). Identification of a new class of lipid droplet-associated proteins in plants. *Plant Physiol.* 162, 1926–1936. doi: 10.1104/pp.113.222455

Huang, A. H. C. (1996). Oleosins and oil bodies in seeds and other organs. *Plant Physiol.* 110, 1055–1061. doi: 10.1104/pp.110.4.1055

Ibarra-Laclette, E., Méndez-Bravo, A., Pérez-Torres, C. A., Albert, V. A., Mockaitis, K., Kilaru, A., et al. (2015). Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. *BMC Genomics* 16, 1–18. doi: 10.1186/s12864-015-1775-y

Iskandarov, U., Silva, J. E., Kim, H. J., Andersson, M., Cahoon, R. E., Mockaitis, K., et al. (2017). A specialized diacylglycerol acyltransferase contributes to the extreme medium-chain fatty acid content of cuphea seed oil. *Plant Physiol.* 174, 97–109. doi: 10.1104/pp.16.01894

Jako, C., Kumar, A., Wei, Y., Zou, J., Barton, D. L., Giblin, E. M., et al. (2001). Seed-specific over-expression of an arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. *Plant Physiol.* 126, 861–874. doi: 10.1104/pp.126.2.861

Janero, D. R., and Barrnett, R. (1981). Analytical separation of green-plant and animal neutral lipids by thin-layer chromatography. *J. Chromatogr. A* 216, 417–422. doi: 10.1016/S0021-9673(00)82379-1

Jeppson, S., Mattisson, H., Demski, K., and Lager, I. (2020). A predicted transmembrane region in plant diacylglycerol acyltransferase 2 regulates specificity toward very-long-chain acyl-coas. *J. Biol. Chem.* 295, 15398–15406. doi: 10.1074/jbc.RA120.013755

Kilaru, A., Cao, X., Dabbs, P. B., Sung, H. J., Rahman, M. M., Thrower, N., et al. (2015). Oil biosynthesis in a basal angiosperm: Transcriptome analysis of persea Americana mesocarp. *BMC Plant Biol.* 15, 203. doi: 10.1186/s12870-015-0586-2

Kozak, M. (1989). The scanning model for translation: An update. *J. Cell Biol.* 108, 229–241. doi: 10.1083/jcb.108.2.229

Kroon, J. T. M., Wei, W., Simon, W. J., and Slabas, A. R. (2006). Identification and functional expression of a type 2 acyl-CoA:diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. *Phytochemistry* 67, 2541–2549. doi: 10.1016/j.phytochem.2006.09.020

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lardizabal, K., Effertz, R., Levering, C., Mai, J., Pedroso, M. C., Jury, T., et al. (2008). Expression of umbelopsis ramanniana DGAT2A in seed increases oil in soybean. *Plant Physiol.* 148, 89–96. doi: 10.1104/pp.108.123042

Lardizabal, K. D., Mai, J. T., Wagner, N. W., Wyrick, A., Voelker, T., and Hawkins, D. J. (2001). DGAT2 is a new diacylglycerol acyltransferase gene family: PURIFICATION, CLONING, AND EXPRESSION IN INSECT CELLS OF TWO POLYPEPTIDES FROM MORTIERELLA RAMANNIANA WITH DIACYLGLYCEROL ACYLTRANSFERASE ACTIVITY *. *J. Biol. Chem.* 276, 38862–38869. doi: 10.1074/jbc.M106168200

Lee, H. G., and Seo, P. J. (2019). Interaction of DGAT1 and PDAT1 to enhance TAG assembly in arabidopsis. *Plant Signal. Behav.* 14, 1–3. doi: 10.1080/15592324.2018.1554467

Li, R., Hatanaka, T., Yu, K., Wu, Y., Fukushige, H., and Hildebrand, D. (2013). Soybean oil biosynthesis: role of diacylglycerol acyltransferases. *Funct. Integr. Genomics* 13, 99–113. doi: 10.1007/s10142-012-0306-z

Liu, Q., Siloto, R. M. P., Lehner, R., Stone, S. J., and Weselake, R. J. (2012). Acyl-CoA:diacylglycerol acyltransferase: Molecular biology, biochemistry and biotechnology. *Prog. Lipid Res.* 51, 350–377. doi: 10.1016/j.plipres.2012.06.001

Li, R., Yu, K., Hatanaka, T., and Hildebrand, D. F. (2010a). Vernonia DGATs increase accumulation of epoxy fatty acids in oil. *Plant Biotechnol. J.* 8, 184–195. doi: 10.1111/j.1467-7652.2009.00476.x

Li, R., Yu, K., and Hildebrand, D. F. (2010b). DGAT1, DGAT2 and PDAT expression in seeds a nd other tissues of epoxy and hydroxy fatty acid accumulating plants. *Lipids* 45, 145–157. doi: 10.1007/s11745-010-3385-4

Li, R., Yu, K., Wu, Y., Tateno, M., Hatanaka, T., and Hildebrand, D. F. (2012). Vernonia DGATs can complement the disrupted oil and protein metabolism in epoxygenase-expressing soybean seeds. *Metab. Eng.* 14, 29–38. doi: 10.1016/j.ymben.2011.11.004

Manthei, K. A., Ahn, J., Glukhova, A., Yuan, W., Larkin, C., Manett, T. D., et al. (2017). A retractable lid in lecithin: Cholesterol acyltransferase provides a structural mechanism for activation by apolipoprotein a-I. *J. Biol. Chem.* 292, 20313–20327. doi: 10.1074/jbc.M117.802736

Marmon, S., Sturtevant, D., Herrfurth, C., Chapman, K., Stymne, S., and Feussner, I. (2017). Two acyltransferases contribute differently to linolenic acid levels in seed oil. *Plant Physiol.* 173, 2081–2095. doi: 10.1104/pp.16.01865

McCartney, A. W., Dyer, J. M., Dhanoa, P. K., Kim, P. K., Andrews, D. W., McNew, J. A., et al. (2004). Membrane-bound fatty acid desaturases are inserted co-translationally into the ER and contain different ER retrieval motifs at their carboxy termini. *Plant J.* 37, 156–173. doi: 10.1111/j.1365-313X.2004.01949.x

McFie, P. J., Stone, S. L., Banman, S. L., and Stone, S. J. (2010). Topological orientation of acyl-CoA:Diacylglycerol acyltransferase-1 (DGAT1) and identification of a putative active site histidine and the role of the n terminus in dimer/tetramer formation. *J. Biol. Chem.* 285, 37377–37387. doi: 10.1074/jbc.M110.163691

McKeon, T. A., and He, X. (2015). Castor diacylglycerol acyltransferase type 1 (DGAT1) displays greater activity with diricinolein than arabidopsis DGAT1. *Biocatal. Agric. Biotechnol.* 4, 276–278. doi: 10.1016/j.bcab.2015.01.005

Mhaske, V., Beldjilali, K., Ohlrogge, J., and Pollard, M. (2005). Isolation and characterization of an arabidopsis thaliana knockout line for phospholipid: diacylglycerol transacylase gene (At5g13640). *Plant Physiol. Biochem.* 43, 413–417. doi: 10.1016/j.plaphy.2005.01.013

Murata, N., and Tasaka, Y. (1997). Glycerol-3-phosphate acyltransferase in plants. *Biochim. Biophys. Acta - Lipids Lipid Metab.* 1348, 10–16. doi: 10.1016/S0005-2760(97)00115-X

Naim, F., Shrestha, P., Singh, S. P., Waterhouse, P. M., and Wood, C. C. (2016). Stable expression of silencing-suppressor protein enhances the performance and longevity of an engineered metabolic pathway. *Plant Biotechnol. J.* 14, 1418–1426. doi: 10.1111/pbi.12506

Oelkers, P., Cromley, D., Padamsee, M., Billheimer, J. T., and Sturley, S. L. (2002). The DGA1 gene determines a second triglyceride synthetic pathway in yeast*. *J. Biol. Chem.* 277, 8877–8881. doi: 10.1074/jbc.M111646200

Ohlrogge, J., and Browse, J. (1995). Lipid biosynthesis. *Plant Cell* 7, 957–970. doi: 10.1105/tpc.7.7.957

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., et al. (1992). The α/β hydrolase fold. *Protein Eng. Des. Sel.* 5, 197–211. doi: 10.1093/protein/5.3.197

Panigrahi, R., Matsui, T., Song, A. H., Caldo, K. M. P., Young, H. S., Weselake, R. J., et al. (2018). Intrinsic disorder in the regulatory n-terminal domain of diacylglycerol acyltransferase 1 from brassica napus. *Sci. Rep.* 8, 1–13. doi: 10.1038/s41598-018-34339-1

Pan, X., Siloto, R. M. P., Wickramarathna, A. D., Mietkiewska, E., and Weselake, R. J. (2013). Identification of a pair of phospholipid: Diacylglycerol acyltransferases from developing flax (Linum usitatissimum l.) seed catalyzing the selective production of trilinolenin. *J. Biol. Chem.* 288, 24173–24188. doi: 10.1074/jbc.M113.475699

Pyc, M., Cai, Y., Greer, M. S., Yurchenko, O., Chapman, K. D., Dyer, J. M., et al. (2017). Turning over a new leaf in lipid droplet biology. *Trends Plant Sci.* 22, 596–609. doi: 10.1016/j.tplants.2017.03.012

Rahman, M., Liu, Q., Zhou, X.-R., and Singh, S. P. (2016). Oil-rich nonseed tissues for enhancing plant oil production. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 11, 1–11. doi: 10.1079/PAVSNNR201611021

Regmi, A., Shockey, J., Kotapati, H. K., and Bates, P. D. (2020). Oil-producing metabolons containing DGAT1 use separate substrate pools from those containing DGAT2 or PDAT. *Plant Physiol.* 184, 720–737. doi: 10.1104/pp.20.00461

Rendón-Anaya, M., Ibarra-Laclette, E., Méndez-Bravo, A., Lan, T., Zheng, C., Carretero-Paulet, L., et al. (2019). The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 116, 17081–17089. doi: 10.1073/pnas.1822129116

Reynolds, K. B., Taylor, M. C., Zhou, X.-R., Vanhercke, T., Wood, C. C., Blanchard, C. L., et al. (2015a). Metabolic engineering of medium-chain fatty acid biosynthesis in nicotiana benthamiana plant leaf lipids. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00164

Reynolds, K. B., Taylor, M. C., Zhou, X. R., Vanhercke, T., Wood, C. C., Blanchard, C. L., et al. (2015b). Metabolic engineering of medium-chain fatty acid biosynthesis in nicotiana benthamiana plant leaf lipids. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00164

Robertson, R. M., Yao, J., Gajewski, S., Kumar, G., Martin, E. W., Rock, C. O., et al. (2017). A two-helix motif positions the lysophosphatidic acid acyltransferase

active site for catalysis within the membrane bilayer. *Nat. Struct. Mol. Biol.* 24, 666–671. doi: 10.1038/nsmb.3436

Routaboul, J.-M., Benning, C., Bechtold, N., Caboche, M., and Lepiniec, L. (1999). The TAG1 locus of arabidopsis encodes for a diacylglycerol acyltransferase. *Plant Physiol. Biochem.* 37, 831–840. doi: 10.1016/S0981-9428(99)00115-1

Saha, S., Enugutti, B., Rajakumari, S., and Rajasekharan, R. (2006). Cytosolic triacylglycerol biosynthetic pathway in oilseeds. molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. *Plant Physiol.* 141, 1533–1543. doi: 10.1104/pp.106.082198

Sandager, L., Gustavsson, M. H., Ståhl, U., Dahlqvist, A., Wiberg, E., Banas, A., et al. (2002). Storage lipid synthesis is non-essential in yeast. *J. Biol. Chem.* 277, 6478–6482. doi: 10.1074/jbc.M109109200

Shockey, J., Gidda, S. K., Chapital, D. C., Kuan, J. C., Dhanoa, P. K., Bland, J. M., et al. (2006). Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *Plant Cell* 18, 2294–2313. doi: 10.1105/tpc.106.043695

Shockey, J., Mason, C., Gilbert, M., Cao, H., Li, X., Cahoon, E., et al. (2015). Development and analysis of a highly flexible multi-gene expression system for metabolic engineering in arabidopsis seeds and other plant tissues. *Plant Mol. Biol.* 89, 113–126. doi: 10.1007/s11103-015-0355-5

Shockey, J., Regmi, A., Cotton, K., Adhikari, N., Browse, J., and Bates, P. D. (2016). Identification of arabidopsis GPAT9 (At5g60620) as an essential gene involved in triacylglycerol biosynthesis1[OPEN]. *Plant Physiol.* 170, 163–179. doi: 10.1104/pp.15.01563

Siloto, R. M. P., Truksa, M., He, X., McKeon, T., and Weselake, R. J. (2009). Simple methods to detect triacylglycerol biosynthesis in a yeast-based recombinant system. *Lipids* 44, 963. doi: 10.1007/s11745-009-3336-0

Smith, S. J., Cases, S., Jensen, D. R., Chen, H. C., Sande, E., Tow, B., et al. (2000). Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking dgat. *Nat. Genet.* 25, 87–90. doi: 10.1038/75651

Ståhl, U., Carlsson, A. S., Lenman, M., Dahlqvist, A., Huang, B., Banas, W., et al. (2004). Cloning and functional characterization of a phospholipid:diacylglycerol acyltransferase from arabidopsis. *Plant Physiol.* 135, 1324–1335. doi: 10.1104/pp.104.044354

Stone, S. J., Levin, M. C., and Farese, R. V. (2006). Membrane topology and identification of key functional amino acid residues of murine acyl-CoA:diacylglycerol acyltransferase-2. *J. Biol. Chem.* 281, 40273–40282. doi: 10.1074/jbc.M607986200

Stone, S. J., Myers, H. M., Watkins, S. M., Brown, B. E., Feingold, K. R., Elias, P. M., et al. (2004). Lipopenia and skin barrier abnormalities in DGAT2-deficient mice *. *J. Biol. Chem.* 279, 11767–11776. doi: 10.1074/jbc.M311000200

Takenaga, F., Matsuyama, K., Abe, S., Torii, Y., and Itoh, S. (2008). Lipid and fatty acid composition of mesocarp and seed of avocado fruits harvested at northern range in Japan. *J. Oleo Sci.* 57, 591–597. doi: 10.5650/jos.57.591

Troncoso-Ponce, M. A., Kilaru, A., Cao, X., Durrett, T. P., Fan, J., Jensen, J. K., et al. (2011). Comparative deep transcriptional profiling of four developing oilseeds. *Plant J.* 68, 1014–1027. doi: 10.1111/j.1365-313X.2011.04751.x

Turchetto-Zolet, A. C., Maraschin, F. S., de Morais, G. L., Cagliari, A., Andrade, C. M. B., Margis-Pinheiro, M., et al. (2011). Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. *BMC Evol. Biol.* 11, 263. doi: 10.1186/1471-2148-11-263

van Erp, H., Bates, P. D., Burgal, J., Shockey, J., and Browse, J. (2011). Castor phospholipid:Diacylglycerol acyltransferase facilitates efficient metabolism of hydroxy fatty acids in transgenic arabidopsis. *Plant Physiol.* 155, 683–693. doi: 10.1104/pp.110.167239

van Erp, H., Shockey, J., Zhang, M., Adhikari, N. D., and Browse, J. (2015). Reducing isozyme competition increases target fatty acid accumulation in seed triacylglycerols of transgenic arabidopsis. *Plant Physiol.* 168, 36–46. doi: 10.1104/pp.114.254110

Vanhercke, T., Divi, U. K., El Tahchy, A., Liu, Q., Mitchell, M., Taylor, M. C., et al. (2017). Step changes in leaf oil accumulation *via* iterative metabolic engineering. *Metab. Eng.* 39, 237–246. doi: 10.1016/j.ymben.2016.12.007

Vanhercke, T., El Tahchy, A., Liu, Q., Zhou, X. R., Shrestha, P., Divi, U. K., et al. (2014). Metabolic engineering of biomass for high energy density: Oilseed-like triacylglycerol yields from plant leaves. *Plant Biotechnol. J.* 12, 231–239. doi: 10.1111/pbi.12131

Vanhercke, T., El Tahchy, A., Shrestha, P., Zhou, X. R., Singh, S. P., and Petrie, J. R. (2013). Synergistic effect of WRI1 and DGAT1 coexpression on triacylglycerol biosynthesis in plants. *FEBS Lett.* 587, 364–369. doi: 10.1016/j.febslet.2012.12.018

Wang, L., Qian, H., Nian, Y., Han, Y., Ren, Z., Zhang, H., et al. (2020). Structure and mechanism of human diacylglycerol O-acyltransferase 1. *Nature* 581, 329–332. doi: 10.1038/s41586-020-2280-2

Wang, H.-W., Zhang, J.-S., Gai, J.-Y., and Chen, S.-Y. (2006). Cloning and comparative analysis of the gene encoding diacylglycerol acyltransferase from wild type and cultivated soybean. *Theor. Appl. Genet.* 112, 1086–1097. doi: 10.1007/s00122-006-0210-9

Welte, M. A., and Gould, A. P. (2017). Lipid droplet functions beyond energy storage. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1862, 1260–1272. doi: 10.1016/j.bbalip.2017.07.006

Wood, C. C., Petrie, J. R., Shrestha, P., Mansour, M. P., Nichols, P. D., Green, A. G., et al. (2009). A leaf-based assay using interchangeable design principles to rapidly assemble multistep recombinant pathways. *Plant Biotechnol. J.* 7, 914–924. doi: 10.1111/j.1467-7652.2009.00453.x

Xue, J., Gao, H., Xue, Y., Shi, R., Liu, M., Han, L., et al. (2022). Functional characterization of soybean diacylglycerol acyltransferase 3 in yeast and soybean. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.854103

Xu, J., Francis, T., Mietkiewska, E., Giblin, E. M., Barton, D. L., Zhang, Y., et al. (2008). Cloning and characterization of an acyl-CoA-dependent diacylglycerol acyltransferase 1 (DGAT1) gene from tropaeolum majus, and a study of the functional motifs of the DGAT protein using site-directed mutagenesis to modify enzyme activity and oil content. *Plant Biotechnol. J.* 6, 799–818. doi: 10.1111/j.1467-7652.2008.00358.x

Xu, R., Yang, T., Wang, R., and Liu, A. (2014). Characterisation of *DGAT1* and *DGAT2* from *Jatropha curcas* and their functions in storage lipid biosynthesis. *Funct. Plant Biol.* 41, 321–329. doi: 10.1071/FP12388

Yang, Y., and Benning, C. (2018). Functions of triacylglycerols during plant development and stress. *Curr. Opin. Biotechnol.* 49, 191–198. doi: 10.1016/j.copbio.2017.09.003

Yen, C.-L. E., Stone, S. J., Koliwad, S., Harris, C., and Farese, J. R.V. (2008). Thematic review series: Glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. *J. Lipid Res.* 49, 2283–2301. doi: 10.1194/jlr.R800018-JLR200

Yuan, L., Mao, X., Zhao, K., Ji, X., Ji, C., Xue, J., et al. (2017). Characterisation of phospholipid: diacylglycerol acyltransferases (PDATs) from camelina sativa and their roles in stress responses. *Biol. Open* 6, 1024–1034. doi: 10.1242/bio.026534

Zhang, M., Fan, J., Taylor, D. C., and Ohlrogge, J. B. (2009). DGAT1 and PDAT1 acyltransferases have overlapping functions in arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. *Plant Cell* 21, 3885–3901. doi: 10.1105/tpc.109.071795

Zhang, L., Wang, S. B., Li, Q. G., Song, J., Hao, Y. Q., Zhou, L., et al. (2016). An integrated bioinformatics analysis reveals divergent evolutionary pattern of oil biosynthesis in high- and low-oil plants. *PloS One* 11, 1–24. doi: 10.1371/journal.pone.0154882

Zheng, P., Allen, W. B., Roesler, K., Williams, M. E., Zhang, S., Li, J., et al. (2008). A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat. Genet.* 40, 367–372. doi: 10.1038/ng.85

Zou, J., Wei, Y., Jako, C., Kumar, A., Selvaraj, G., and Taylor, D. C. (1999). The arabidopsis thaliana TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene. *Plant J.* 19, 645–653. doi: 10.1046/j.1365-313X.1999.00555.x

Zuckerkandl, E., and Pauling, L. (1965). *Evolutionary divergence and convergence in proteins*. Eds. V. Bryson and P. Vogel (Academic Press), 97–166. doi: 10.1016/B978-1-4832-2734-4.50017-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 February 2023
DOI 10.3389/fpls.2023.1144306

OPEN ACCESS

EDITED BY
Hongbo Chao,
Zhengzhou University, China

REVIEWED BY
Chuan Tang Wang,
Shandong Peanut Research Institute, Qingdao, China
Yongtai Yin,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Yi Shen
shenyi1202@163.com
Zhide Chen
chen701865@aliyun.com

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 14 January 2023
ACCEPTED 31 January 2023
PUBLISHED 10 February 2023

CITATION
Shen Y, Shen Y, Liu Y, Bai Y, Liang M, Zhang X and Chen Z (2023) Characterization and functional analysis of *AhGPAT9* gene involved in lipid synthesis in peanut (*Arachis hypogaea* L.). *Front. Plant Sci.* 14:1144306. doi: 10.3389/fpls.2023.1144306

Characterization and functional analysis of *AhGPAT9* gene involved in lipid synthesis in peanut (*Arachis hypogaea* L.)

Yue Shen[1], Yi Shen[1]*, Yonghui Liu[1], Yang Bai[2], Man Liang[1], Xuyao Zhang[1] and Zhide Chen[1]*

[1]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [2]Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (Nanjing Botanical Garden Mem. Sun Yat-Sen), Nanjing, China

GPAT enzymes (glycerol-3-phosphate 1-O-acyltransferase, EC 2.3.1.15) catalyze the initial and rate-limiting step of plant glycerolipid biosynthesis for membrane homeostasis and lipid accumulation, yet little research has been done on peanuts. By reverse genetics and bioinformatics analyses, we have characterized an AhGPAT9 isozyme, of which the homologous product is isolated from cultivated peanut. QRT-PCR assay revealed a spatio-temporal expression pattern that the transcripts of *AhGPAT9* accumulating in various peanut tissues are highly expressed during seed development, followed by leaves. Green fluorescent protein tagging of *AhGPAT9* confirmed its subcellular accumulation in the endoplasmic reticulum. Compared with the wild type control, overexpressed *AhGPAT9* delayed the bolting stage of transgenic *Arabidopsis*, reduced the number of siliques, and increased the seed weight as well as seed area, suggesting the possibility of participating in plant growth and development. Meanwhile, the mean seed oil content from five overexpression lines increased by about 18.73%. The two lines with the largest increases in seed oil content showed a decrease in palmitic acid (C16:0) and eicosenic acid (C20:1) by 17.35% and 8.33%, respectively, and an increase in linolenic acid (C18:3) and eicosatrienoic acid (C20:3) by 14.91% and 15.94%, respectively. In addition, overexpressed *AhGPAT9* had no significant effect on leaf lipid content of transgenic plants. Taken together, these results suggest that *AhGPAT9* is critical for the biosynthesis of storage lipids, which contributes to the goal of modifying peanut seeds for improved oil content and fatty acid composition.

KEYWORDS

***Arachis hypogaea*, *AhGPAT9*, evolution analysis, triacylglycerol, oil content, fatty acid**

Introduction

Peanut (*Arachis hypogaea* L.) or groundnut is an important self-pollinated legume crop widely cultivated around the world for edible oil, food and feed use. Peanut kernels are comprised of over 50% oil and about 30% protein, as well as many minerals and vitamins (Janila et al., 2013). The fatty acid (FA) composition of peanut is unique, in that the total unsaturated fatty acid content exceeds 80%. The variation of oleic (O) and linoleic (L) fatty acids represents the most important quality traits for evaluating stability and nutrition, and high O/L ratio can increase the shelf life (Jung et al., 2000). China's annual peanut production ranks first in the world, and more than 50% of its total output is used for oil extraction, which plays an important strategic role in ensuring the safety of edible oil. However, constraints including limited acreage and the increasing demand for lipid consumption means that improvement of peanut oil content and quality remains the focus of current scientific work.

Triacylglycerol (TAG) represents the major storage reserve in oilseeds, usually esterified by 3 fatty acids and 1 glycerol. It can provide energy support for germinating plants during the non-autotrophic stage before photosynthesis, and is also involved in plant development and stress resistance (Graham, 2008; Zhang et al., 2009; Cui et al., 2016). Many studies have demonstrated the existence of several parallel TAG biosynthetic pathways in plants, including the acyl-CoA-dependent Kennedy pathway (*de novo* DAG/TAG synthesis), the acyl-CoA-independent PC pathway (PC-derived DAG/TAG synthesis), and the monoacylglycerol (MAG) pathway, showing that plant lipid synthesis involves a complex metabolic network with multiple regulatory pathways and genes (Tumaney et al., 2001; Shi and Cheng, 2009; Chen et al., 2015; Bates, 2016).

Our focus on glycerol-3-phosphate acyltransferase (GPAT) bridges the two compartmentalization pathways of fatty acid synthesis in plastid and glycerolipid synthesis in the endoplasmic reticulum, and the first step of the acylation reaction catalyzed by it is considered to be the key rate-limiting step of the Kennedy pathway. As the carbon chain skeleton for TAG synthesis, glycerol-3-phosphate (G3P) harbors three fatty acid binding sites, *sn*-1, *sn*-2 and *sn*-3. GPATs mainly transfer the acyl groups on acyl-CoA to the *sn*-1 or *sn*-2 position hydroxyl groups of G3P to produce lysophosphatidic acid (LPA), a crucial intermediate for the formation of several acyl-lipids. In plants, different subcellular localization of GPATs may determine the metabolites of LPAs, such as extracellular lipid polyesters, membrane and storage lipids (Lee et al., 2016).

In *Arabidopsis*, ten GPAT homologs have been identified, of which *GPAT1~8* are the land-plant-specific *sn*-2-GPAT that is mainly involved in the biosynthesis of lipid polyesters such as cutin and suberin, and is associated with plant flower development and stress response (Li et al., 2007; Li-Beisson et al., 2009; Chen et al., 2011; Yang et al., 2012). Plastid-localized *ATS1* also can catalyze the acylation reaction at the *sn*-1 site of G3P using acyl-ACP as an acyl donor, and may be related to the mechanism of plant cold tolerance response (Payá-Milans et al., 2015; Sun et al., 2015). Loss of function of *GPAT1* located in mitochondria alters the fatty acid composition in floral tissues and seeds, but its effect on seed oil content is still controversial (Zheng et al., 2003; Bai et al., 2021). Undoubtedly, *Arabidopsis GPAT9*, which is similar to mammalian *GPAT3/4* function in fat synthesis, can be directly involved in the membrane lipids and TAG biosynthesis for plants as the *sn*-1 bifunctional enzyme gene localized in endoplasmic reticulum (Gidda et al., 2009; Shockey et al., 2016; Singer et al., 2016). In addition, *GPATs* involved in TAG accumulation have been cloned in *Lepidium latifolium* (Mohan et al., 2013), *Brassica napus* (Chen et al., 2014; Liu et al., 2015), *Helianthus annuus* (Payá-Milans et al., 2016), *Jatropha curcas* (Misra et al., 2017), *Physcomitrella patens* (Yang et al., 2019) and other plants.

In this study, a homologous transcript of *Arabidopsis GPAT9* gene (At5g60620), *AhGPAT9*, isolated from cultivated peanut, which encodes a fragment of 1131bp in length. The Conserved Domain Databases (CDD) was used to predict that its encoded protein possessed typical glycerol acyltransferase activity. The expression characteristics of *AhGPAT9* were analyzed to elucidate its biological function involved in plant lipid synthesis.

Materials and methods

Plant materials and growth conditions

Cultivated peanut (*Arachis hypogaea* cv. Tifrunner) seeds were sterilized in 2% (v/v) sodium hypochlorite for 10 min, rinsed thoroughly with deionized water, sown into nutrient-enriched soil, and grown in the phytotron at 25°C with a 16:8 h light: dark (L:D) photoperiod. *A. thaliana* ecotype Col-0 seeds were pre-grown in MS basal medium (Murashige and Skoog, 1962) for one week, and then transferred to nutrient-enriched soil in a growth chamber at 23°C with a 14:10 h L:D photoperiod. Hoagland's nutrient solution was used for water and fertilizer management.

Data mining of peanut GPATs

The sequence data of GPAT conserved domain (PF01553) was downloaded from Pfam database (http://pfam.xfam.org/), and the protein library of cultivated peanut and its two diploid progenitors (*Arachis duranensis* and *Arachis ipaensis*) were downloaded from PeanutBase database (https://www.peanutbase.org). The hidden Markov model was constructed by HMMER3.0 software to obtain the preliminary screened ID information of peanut GPAT. The protein sequences containing PF01553 domain were anchored and extracted by SeqHunter1.0 software, and the integrity of conserved domain of above sequences was analyzed by Pfam and SMART databases, while repetitive sequences and redundant transcripts were removed to screen the candidate genes of peanut *GPATs*.

Phylogenetic and chromosome mapping analyses of peanut GPATs

A. thaliana GPAT sequences was retrieved from the Arabidopsis Information Resource (https://www.arabidopsis.org). The multiple sequence alignment analysis of GPAT proteins from *Arachis* and *Arabidopsis* were analyzed by Clustal-X software, then phylogenetic tree was constructed by maximum likelihood method with 1000 bootstraps using MEGA-X software. The genome sequence and

genome structure annotation data of cultivated peanut were downloaded, from which the chromosomal location data of *AhGPATs* were extracted, and physical location diagram of candidate genes on peanut chromosome were drawn by TBtools software (Chen et al., 2020).

Cloning of *AhGPAT9*

The coding sequence of *AhGPAT9* (1131bp) was amplified from reverse-transcribed RNA isolated from peanut seedlings, the sequence data with the database locus number Arahy.5QGNM was obtained from PeanutBase database. Primers were designed by PrimerPrimer5.0 software, PCR amplification was performed using KOD-Plus-Neo high-fidelity enzyme (TOYOBO, Shanghai, China), and the clone was ligated into the pEASY vector (TransGen, Beijing, China). The primers used for amplification of CDS fragment are listed in Supplementary Table S1.

Bioinformatics analysis

We chose the CDD database (https://www.ncbi.nlm.nih.gov/cdd) for initial identification of conserved domain in AhGPAT9 protein sequence. Then the prediction analyses of physicochemical parameters, hydrophobicity, secondary structure, transmembrane helices, signal peptides, functional domains and phosphorylation sites of the protein encoded by *AhGPAT9* were performed by ProtParam (https://web.expasy.org/protparam), ProtScale (https://web.expasy.org/protscale), SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html), TMHMM (https://services.healthtech.dtu.dk/service.php?TMHMM), SignalIP (https://services.healthtech.dtu.dk/service.php?SignalP), SMART (http://smart.embl-heidelberg.de) and NetPhos (https://services.healthtech.dtu.dk/service.php?NetPhos) databases, respectively. Phylogenetic analysis of AhGPAT9 homologs from different species was performed by screening homologous sequences through the BLASTp non-redundant protein database (https://blast.ncbi.nlm.nih.gov), and then constructing a phylogenetic tree by maximum likelihood method with 1000 bootstraps using MEGA-X software.

RNA isolation and QRT-PCR analysis

Total RNA was extracted from various tissues of peanut with plant RNA isolation kit (Sangon, Shanghai, China) following the manufacturer's instructions. First-strand cDNA was synthesized using PrimeScript^{TM} RT reagent Kit with gDNA eraser (TaKaRa, Dalian, China). PCR amplification was performed using SYBR Premix Ex Taq^{TM} kit (TaKaRa, Dalian, China) with *AhACT11* (Xu et al., 2021) as the internal reference, each template was repeated three times, and the data were analyzed by comparative cycle threshold method ($2^{-\Delta\Delta CT}$). Quantitative real-time (qRT) PCR assay was performed in 20μl reaction volumes using a QuantStudio5 real-time PCR system (Applied Biosystems, California, USA). The primers used for qRT-PCR are listed in Supplementary Table S1.

Subcellular localization assay

Prediction analysis of the subcellular location of *AhGPAT9* was performed by ProtCompVersion9.0 (http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc) software. Meanwhile, *AhGPAT9* and *HDEL* coding sequence (excluding the stop codon) were amplified and fused with *EGFP* and *mCherry* reporter genes, respectively, at the C-terminal in the pSuper1300^{+} vector. The recombinant plasmids *AhGPAT9-EGFP* and *HDEL-mCherry* were transferred into *Agrobacterium tumefaciens* strain *GV3101* by freeze-thaw method. Then tobacco leaves were transfected by infiltration using injection method, and incubated for 2-3 d under normal cultural conditions before performing fluorescence assays using an UltraVIEW VoX laser confocal imaging analysis system (PerkinElmer, Massachusetts, USA). The primers used for subcellular location are listed in Supplementary Table S1.

Generation of transgenic *Arabidopsis* expressing *AhGPAT9*

The coding sequence of *AhGPAT9* was built into the modified binary vector pSuper1300^{+} containing the *Super* promoter (Shen et al., 2015). Then the recombinant plasmid was transferred into *Agrobacterium tumefaciens* strain *GV3101* to infect inflorescences of *Arabidopsis* using the floral-dipping method. Mendelian inheritance of hygromycin resistance was showed as single dominant locus in T1 and T2 generation of all transgenic plants, and the homozygous lines of T3 generation were used for phenotype analysis and determination of oil content and FA composition. The primers are used for overexpression are listed in Supplementary Table S1.

Lipid extraction and analysis

The extraction of seed oil (Li et al., 2006) and total leaf lipid (Kim et al., 2015) was extracted using micro-extraction method, 10-20 mg freeze-dried samples were taken and repeated five times for each treatment. TAG component of total leaf lipid was separated using thin-layer chromatography method (TLC). Fatty acid composition of TAG in seeds and leaves was quantitatively analyzed by internal standard method using TAG-17:0 for normalization. The methyl esterification products were determined using gas chromatography method with Thermo TRACE GC (ThermoFisher, Massachusetts, USA), and 37 fatty acid methyl ester mixture (Sigma, Missouri, USA) was used as the standard substance. According to the peak time of different fatty acids in the standard, the fatty acid components of each sample were analyzed, and then the relative molar content of each fatty acid was calculated based on the peak area.

Results

Identification and chromosomal localization of peanut GPATs

To identify peanut *GPAT* genes and their homologs, the PlsC acyltransferase domain of Pfam (PF01553) was used as a probe to

screen 56, 25 and 31 proteins, respectively, with complete domain motifs from the protein sequence data of cultivated peanut and its two diploid progenitors. Based on the ploidy relationship among them, it was found that the quantity of GPATs in cultivated peanut was consistent with the sum of GPAT members in wild peanut. Phylogenetic analysis of GPAT proteins in peanut and *Arabidopsis* was performed by maximum likelihood method using MEGA software, and the results showed that they were divided into five subfamilies. As shown in Figure 1, sequences homologous to AtATS1 and AtGPAT9 constituted A and C independent groups, respectively, with *sn*-1 acyltransferase properties (Nishida et al., 1993; Singer et al., 2016). Group B contained all the land-plant-specific *sn*-2-CPAT, which mainly performed the function of a membrane-bound protein and could be further classified into three distinct conserved sub-clades based on the key stages of the morphological and functional evolution: cutin synthesis (II), suberin synthesis (III) and other vascular ancient branches (I) (Yang et al., 2012). In addition, there was no clustering of GPAT for *A. thaliana* within group D and group E, which suggests that they might be glycerol-3-phosphate acyltransferases specific to peanut. The 56 *AhGPAT*s members were found to be unevenly distributed across 18 chromosomes of cultivated peanut genome except for chromosomes 10 and 11 (Figure 2). Chromosomes 02, 06, 12, 16 and 17 contained five genes, chromosome 7 contained four genes, chromosomes 03, 04, 13, 14 and 20 contained three genes, chromosomes 01, 08, 09, 18 and 19 contained two genes, and chromosomes 05 and 15 contained only one gene.

Cloning and bioinformatics analysis of *AhGPAT9*

In this study, a fragment (Figure S1) homologous to *AtGPAT9* for *Arabidopsis* was obtained by PCR amplification. After sequencing and alignment, it was found that this gene was consistent with the *Arahy.5QGNM* reference sequence, named *AhGPAT9* (Genebank accession number: MN124513) and located on chromosome 13 in peanut (Figure 2). The length of the coding sequence for *AhGPAT9* was 1131bp, which encoded 376 amino acids. Physicochemical property analysis showed that AhGPAT9 was an unstable protein with molecular formula $C_{1983}H_{3083}N_{533}O_{529}S_{21}$, molecular weight 43.528kD, isoelectric point 9.09, fat coefficient 91.97 and stability coefficient 46.56. Meanwhile, the total average hydrophobicity coefficient was -0.110, making it a hydrophilic protein (Figure S2A). Secondary structure prediction results showed that α-helix accounted for 48.40%, β-turn accounted for 3.46%, random coil accounted for 33.24%, and extended strand accounted for 14.89%, indicating that α-helix and random coil are the main components of the secondary structure for AhGPAT9 (Figure S2B). Phosphorylation site prediction results showed that the protein had 25 phosphorylation sites, including 5 threonine, 17 serine and 3 tyrosine (Figure S2C), suggesting that phosphorylation modification may be involved in regulating the activity of AhGPAT9 protein, and are dominated by the serine sites. In addition, it was found that the protein contains no signal peptide but three transmembrane domains and a PlsC functional domain (Figure S2D, E), suggesting that AhGPAT9 may be associated with glycerolipid synthesis given the phosphor-acyltransferase activity of the PlsC domain.

Fourteen protein sequences with high similarity to AhGPAT9 were selected from the non-redundant protein database (sequence identity 84.51~88.82%). Protein functional annotation showed that they were GPAT9 proteins from different species. Multiple sequence alignment analysis revealed that all these protein sequences had a complete PlsC functional domain. Furthermore, the phylogenetic tree constructed by the maximum likelihood method of MEGA software (Figure 3), shows eight legume sources of GPATs clustered into the same group, of which peanut AhGPAT9 has the closest molecular

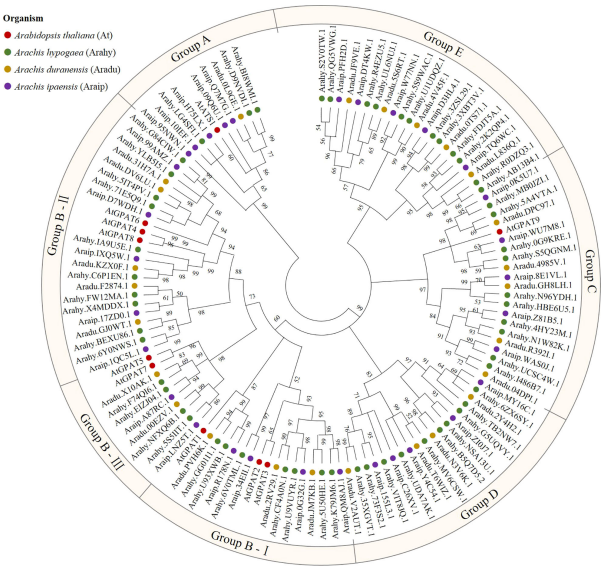


FIGURE 1
Phylogenetic analysis of glycerol-3-phosphate acyltransferase (GPAT) family in peanut. GPATs in cultivated peanut (*Arachis hypogaea*) and its two diploid progenitors (*Arachis duranensis* and *Arachis ipaensis*), as well as *A. thaliana* were used to construct a phylogenetic tree by maximum likelihood method with 1000 bootstraps using MEGA software (version X). Red = *A. thaliana*; Green = *Arachis hypogaea*; Yellow = *Arachis duranensis*; Purple = *Arachis ipaensis*.

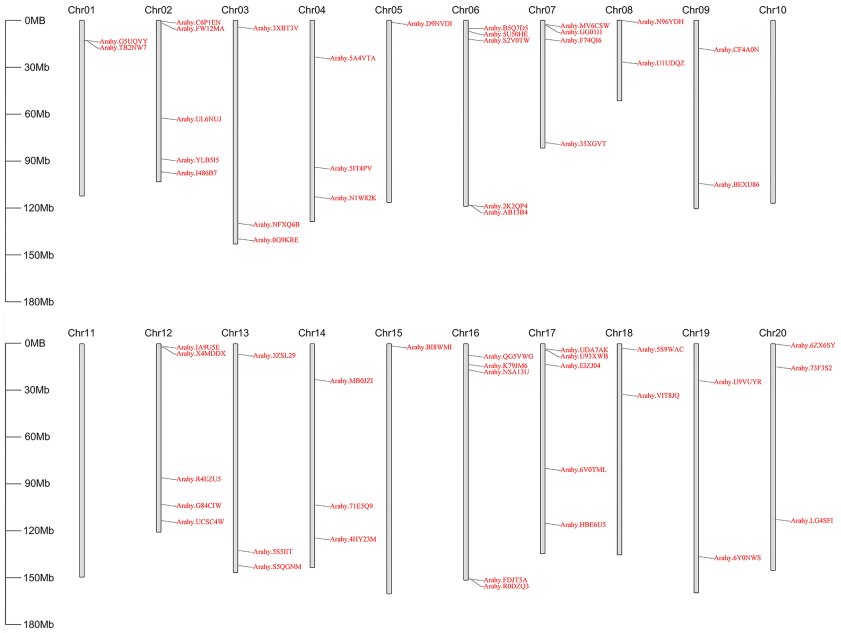


FIGURE 2
Physical location of *AhGPATs* genes on peanut chromosomes. The 56 predicted *AhGPATs* genes from cultivated peanut (*Arachis hypogaea*) were used to construct a physical location distribution map on chromosomes using TBtools software. Chromosome numbers are indicated on the top of each scaffold. Chromosome size is shown on the left vertical scale.

distance to dhal CcGPAT9 (XP_020232467.1) and soybean GmGPAT9 (XP_003524805.1), suggesting that AhGPAT9 may be functionally similar to both of them.

Expression pattern and subcellular localization of *AhGPAT9*

To investigate the tissue-specific expression pattern of *AhGPAT9*, the relative expression of *AhGPAT9* in root, shoot, stem, mature leaf, flower, gynophore and seeds at different developmental stages were analyzed by quantitative real-time PCR. The results showed that AhGPAT9 was differentially expressed in various tissues, with the highest expression in seeds, followed by leaves and shoots, and relatively low expression in other tissues (Figure 4). In addition, the expression level of *AhGPAT9* showed a trend of increasing and then decreasing during seed development. These results indicate that the tissue expression of *AhGPAT9* is spatio-temporal specific, and also imply that *AhGPAT9* may play a positive role in the accumulation of substances during seed development.

We used ProtCompVersion9.0 software, an online tool for subcellular localization, to predict that AhGPAT9 was localized in the endoplasmic reticulum (ER). On that base, we expressed *AhGPAT9* fused to an *EGFP* at the C-terminal in tobacco leaves to confirm the ER localization of AhGPAT9, and used endoplasmic reticulum retention signal (HDEL) as an ER marker for co-localization (Song et al., 2021). As shown in Figure 5, when *AhGPAT9-EGFP* and *HDEL-mCherry* recombinants were co-expressed in tobacco mesophyll cells, the green and red fluorescence signals overlapped considerably (Figure 5B), and further magnification showed that the two fluorescence signals completely overlapped (Figure 5C), whereas unfused *EGFP* and *mCherry* was localized in cytoplasm and nucleus, as expected (Figure 5A). Overall, the transient expression assay turned up strong evidence that AhGPAT9 is a functional protein localized in the endoplasmic reticulum.

AhGPAT9 contributes to TAG accumulation and FAs alteration in seeds

To investigate the physiological functions of AhGPAT9 involved in plant oil synthesis, the coding sequence of *AhGPAT9* was introduced into a modified constitutive-expression vector for transfection using the *Arabidopsis* floral-dipping method. In this study, 37 T1 generation individual plants that overexpressed *AhGPAT9* were identified, from which 5 homozygous lines were randomly selected for tissue qRT-PCR. The results showed that the foreign gene had been successfully inserted into the genome of *Arabidopsis* and obtained stable generation (Figure S3).

We extracted the seed lipid of the above transgenic lines, and found that the seed oil content of them was significantly higher than that of WT control, with an increase in range from 14.67% to 22.83%, and the mean oil content increased by about 18.73% (Figure 6A). Two overexpression lines of OE1 and OE32 were selected to analyze the changes in fatty acid composition for seed oil. Compared with the WT, the average content of palmitic acid (C16:0) and eicosenoic acid (C20:1) in overexpression seeds decreased by about 17.35% and 8.33%, respectively, while the average content of linolenic acid (C18:3) and eicosatrienoic acid (C20:3) increased by about 14.91% and 15.94%, respectively

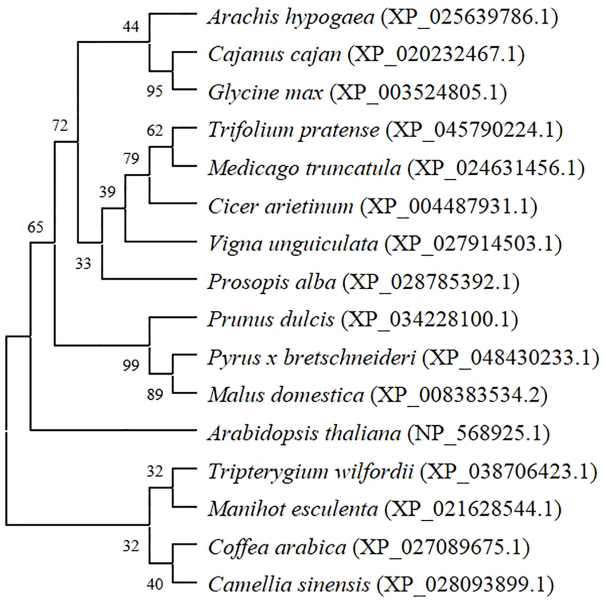


FIGURE 3
Phylogenetic tree of peanut AhGPAT9 proteins with related GPATs from other plant species. Species names are in italics, and accession numbers of reference sequences in NCBI are in parentheses.

(Figure 6B). The content of 16C/18C/20C/22C FAs were then summed separately, and it was found that the content of 16C FAs in overexpression lines significantly reduced, while the content of 18C FAs significantly increased (Figure 6C). Similarly, the content of monounsaturated FAs significantly decreased and the content of triunsaturated FAs significantly increased in overexpression lines, while the content of saturated FAs remained stable (Figure 6D). These results indicate that overexpressed *AhGPAT9* not only significantly increases the oil content of transgenic seeds, but also significantly changes their FA composition ratio, suggesting that it may be preferable to utilize linolenic acid and eicosatrienoic acid as substrates during TAG assembly.

Different from the mass accumulation of storage TAG in seeds, the glycerolipid in leaves are more complex and contain several polar photosynthetic membrane lipids, while TAG in mesophyll cells store in the cytosolic lipid droplets. We isolated the component of TAG from 35-40 DAG* rosette leaves (Figure 7A), and found that there was no significant difference in TAG content between transgenic lines and WT (Figure 7B). Further analysis of FA composition showed that the average content of oleic acid (C18:1) and docosahexaenoic acid (C22:6) in overexpression leaves decreased by about 59.60% and 57.43%, respectively, while the average content of stearic acid (C18:0) and erucic acid (C22:1) increased by about 20.75% and 23.27%, respectively (Figure 7C). In addition, the content of

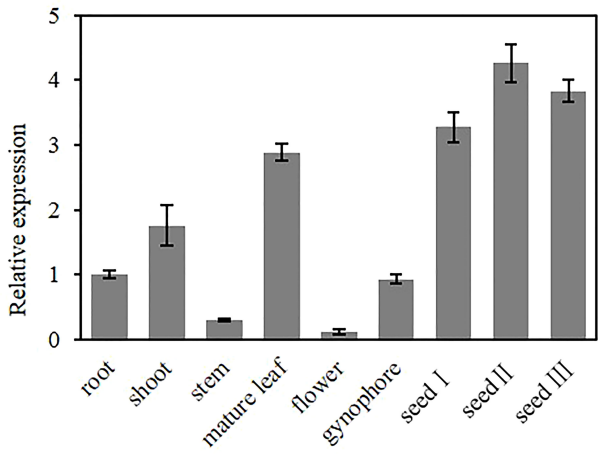


FIGURE 4
Expression analysis of *AhGPAT9* using qRT-PCR in diverse tissues of peanut. The root and shoot samples were taken at seedling stage, the stem, leaf, flower, gynophore and seed samples were taken at adult-plant stage. Seed I, seed II, and seed III represent the early, middle and late stages of seed development, respectively. Relative mRNA abundance was normalized with respect to that of peanut *AhACT11*, and data are shown as means ± standard deviation (n = 3).

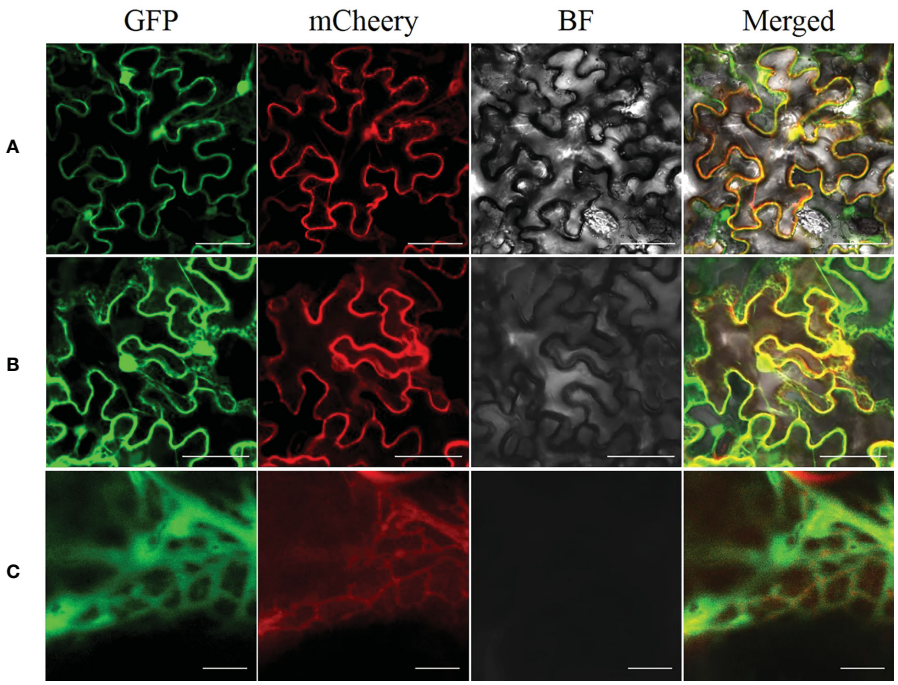


FIGURE 5
Subcellular localization of AhGPAT9 in tobacco leaves. **(A)** Representative laser-scanning images of tobacco mesophyll cells co-expressing EGFP and mCherry empty vectors (bars = 50 mm). **(B)** Representative laser-scanning images of tobacco mesophyll cells co-expressing AhGPAT9-EGFP and HDELmCherry fusion proteins (bars = 50 mm). **(C)** Enlargement of Figure 5B (bars = 5 mm).

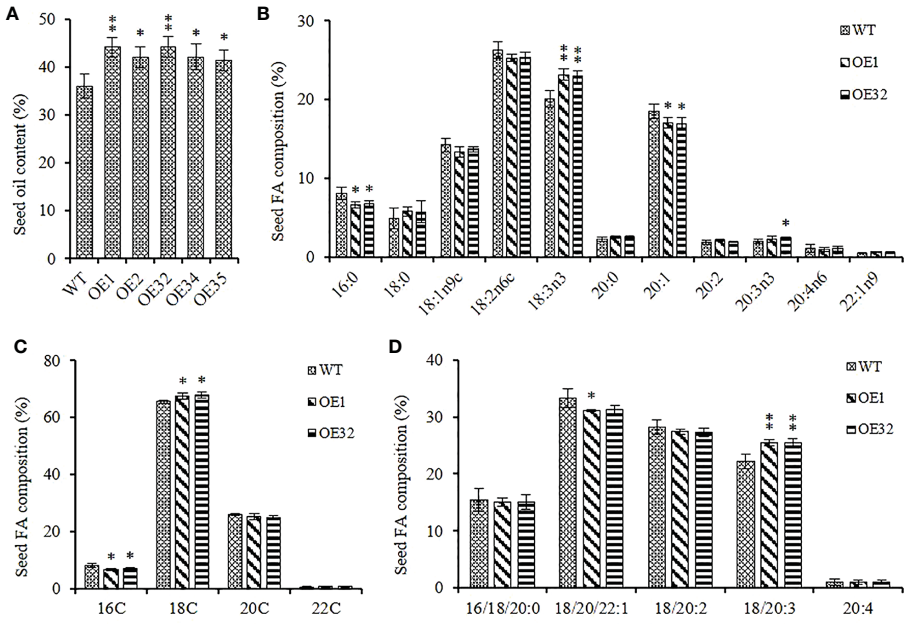


FIGURE 6
Seed oil content and composition of homozygous AhGPAT9 overexpression lines. **(A)** Oil content in seeds from wild type (WT) and AhGPAT9 overexpression lines (OE). **(B)** FA composition of seed TAG. **(C)** 16C/18C/20C/22C FAs content of seed TAG. **(D)** Saturated and unsaturated FAs content of seed TAG. Asterisks indicate significant differences from the WT: *P < 0.05; **P < 0.01 (Student's t-test). Data are shown as means ± standard deviation (n = 5).

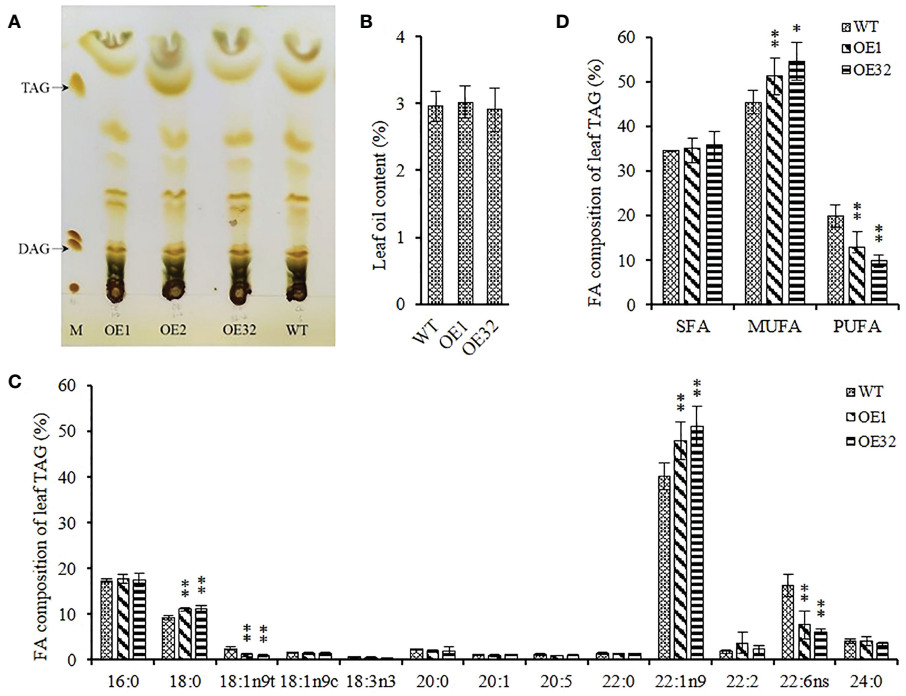


FIGURE 7
Leaf oil content and composition of homozygous AhGPAT9 overexpression lines. (A) Separation of lipid components from 35-40 DAG* rosette leaves by TLC. M represents a lipid marker: TAG = triacylglycerol; DAG = diacylglycerol. DAG* = days after germination. (B) TAG content in leaves from wild type (WT) and AhGPAT9 overexpression lines (OE). (C) FA composition of leaf TAG. (D) Saturated and unsaturated FAs content of leaf TAG. SFA = saturated fatty acid; MUFA = monounsaturated fatty acid; PUFA = polyunsaturated fatty acid. Asterisks indicate significant differences from the WT: **P < 0.01 (Student's t-test). Data are shown as means ± standard deviation (n = 5).

monounsaturated FAs significantly increased and the content of polyunsaturated FAs significantly decreased in overexpressed lines, while the content of saturated FAs remained stable (Figure 7D). These results indicate that overexpressed *AhGPAT9* cannot significantly affect the TAG content of transgenic leaves, but does alter the FAs composition for TAG and make it inclined to recombine FAs with lower saturation, which is contrary to that in seeds.

AhGPAT9 affects plant growth and seed development

Under the same culture conditions, we found that overexpressed *AhGPAT9* hardly affected the seed germination and seedling growth of transgenic lines, but their bolting time of rosette was generally delayed by more than one week compared with that of the wild type control, with a corresponding delay in first-flowering (Figures 8A and S4A). The bolting rate of transgenic lines was found to variation range from 27.76% to 80.56% when grew to 5-week-old, of which four lines were significantly lower than that of wild type in the same period (Figure S4B). Agronomic trait analysis showed that the plant height of transgenic lines was generally shorter than that of WT at maturity, but there was no statistically significant difference. Meanwhile, the number of siliques per plant was significantly reduced, and the mean weight per silique showed an increasing trend with no statistically significant difference (Figures 8B, C). Compared with WT, the mean weight and area per seed for transgenic lines were significantly increased by about 5.85% and 2.15%, respectively, and other indices such as length, width, diameter and roundness of transgenic seeds also changed significantly (Figure 8D). These results suggest that *AhGPAT9* may be closely related to plant growth and development, especially in the reproductive growth stage, which ultimately affects the yield and quality formation of oilseeds.

Discussion

The currently available information relevant to the potential involvement of peanut *GPATs* in glycerolipid biosynthesis is mainly derived from transcription analysis. They have spatio-temporal expression differences in various tissues, and can be induced by abiotic stress such as salt, drought and ABA treatment (Chi et al., 2015; Hao et al., 2018; Lv et al., 2020). Here, we provide genetic evidence to demonstrate that *AhGPAT9* is essential for TAG accumulation and alters the FA composition.

In this study, the homologous transcript of *GPAT9* in peanut, *AhGPAT9*, was isolated and cloned using bioinformatics methods. It was located on the 13th chromosome of the peanut genome and encoded a fragment of 1131bp in nucleotide length. The polypeptide sequence encoded by *AhGPAT9* had multiple-types of phosphorylation sites and a typical PLsC domain, indicating that it belongs to the acyltransferase family. In addition, it shared 79.58% identity with the homologous polypeptide sequence in *Arabidopsis*, as well as being phylogenetically most closely related to pigeonpea (*Cajanus cajan*) and soybean (*Glycine max*) crops. Recent studies

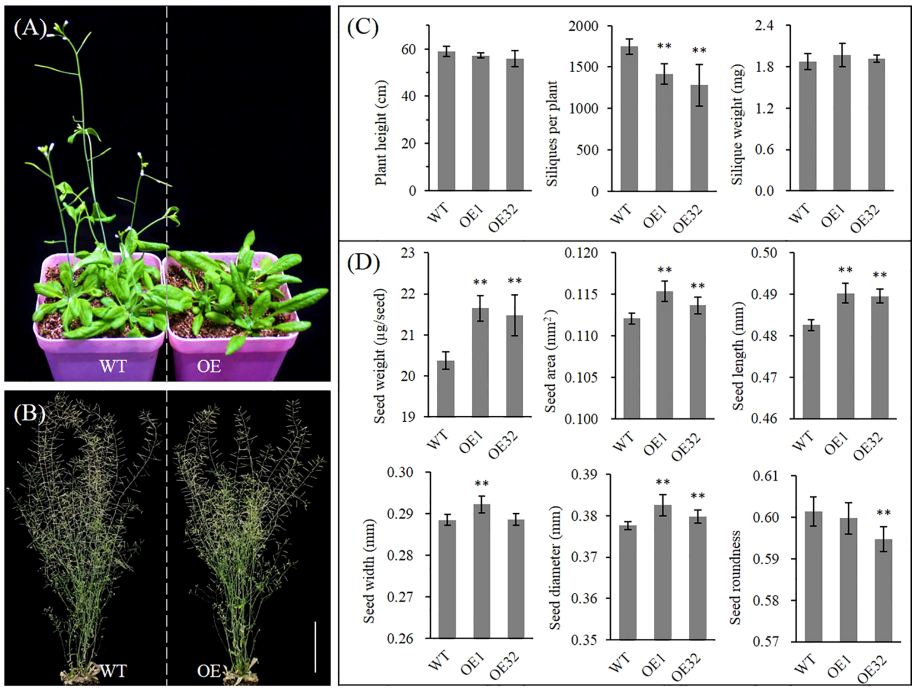


FIGURE 8
Phenotype analysis of homozygous *AhGPAT9* overexpression lines in *Arabidopsis*. **(A)** Wild type (WT) and *AhGPAT9* overexpression lines (OE) at 5-week-old under normal conditions. **(B)** Wild type (WT) and *AhGPAT9* overexpression lines (OE) at maturity under normal conditions. **(C)** Agronomic trait analysis of WT and OE lines at maturity, including plant height, siliques per plant and mean silique weight (n = 6, bars = 10 cm). **(D)** Dry seed weight and seed-shape trait analyses of WT and OE lines, including mean seed area, seed length, seed width, seed diameter and seed roundness. About 3000 seeds harvested from 12 individual plants were tested in each sample (n = 12). Asterisks indicate significant differences from the WT: $^{**}P < 0.01$ (Student's *t*-test). Data are shown as means ± standard deviation.

have shown that *GmGAPT9* in soybean exhibits significant acyltransferase activity by yeast genetic complementation assay, and can elevate the proportion of arachidic and erucic acids in *Arabidopsis* transgenic seeds with no changes in oil content (Liu et al., 2022). Furthermore, *in vivo* and *in vitro* experiments showed that *AtGAPT9* is highly specific for acyl-CoA and contributes to the biosynthesis of both leaf polar lipids and seed oil, as well as lipid droplet production in developing pollen grains. Loss of function results in a lethal phenotype of male and female gametophytes (Shockey et al., 2016; Singer et al., 2016). Given that GPAT9 is highly conserved throughout evolution and is largely present as one single copy in most plants, we suggest GPAT9 in peanut is very likely to encode essential housekeeping functions similar to those in *Arabidopsis* and soybean, implying the possibility that *AhGPAT9* is essential for TAG biosynthesis.

We found that *AhGPAT9* was significantly more abundantly expressed in seeds compared with other peanut tissues, with a bell-shaped expression pattern at seed developmental stage. This was consistent with the conclusion that *AhGAPT9* transcripts reached the maximum value at 42DAP (Lv et al., 2020), during which the embryo morphology was close to maturation and the seed oil accumulated rapidly. Our previous GUS fusion results indicated that *AhGPAT9* displayed the strongest activity at podding stage, especially in the developing siliques and the corresponding walking-stick embryos, which was similar to the expression pattern of *AtGAPT9* (Singer et al., 2016; Shen et al., 2022). The walking-stick embryo is in the late stage of embryonic development, during which the storage compounds are mainly accumulated to prepare for seed dormancy and germination. Taken together, the spatio-temporal expression pattern of *AhGPAT9* exhibits tissue specificity in seeds/embryos at various developmental stages, and its expression abundance is consistent with the rate of oil accumulation, suggesting that *AhGPAT9* may be a limiting factor affecting seed glycerolipid biosynthesis. The ER-bound properties of AhGPAT9 demonstrates its participation in the ER-localized glycerolipid compartmentalized biosynthesis in peanut (Gidda et al., 2009; Chapman and Ohlrogge, 2012).

In plants, TAG accumulate mainly in seeds, pericarps, leaves and flowers as neutral storage lipids (Yang and Benning, 2018). Knockdown or overexpression of *Arabidopsis* GPAT9 had been shown to affect the levels of TAG both in leaves and seeds (Shockey et al., 2016; Singer et al., 2016), which hinted at the potential functional properties of peanut homologs. Previous studies indicated that *AhGPAT9A/B* alleles indeed affect the accumulation of seed oil in peanut, and its higher oil content can be achieved not only by over-expressing *AhGPAT9*, but also may be associated with the polymorphic combination of the two alleles (Lv et al., 2020). In this study, constitutive *AhGPAT9-OE* resulted in large seed oil enhancement, while the lipid content of rosette leaves was not affected, and the TAG accumulation only accounted for about 3% of dry weight, indicating that *AhGPAT9* might be a major contributor to TAG biosynthesis in oilseeds. We further analyzed the FA composition of TAG in overexpressed seeds, and found that the proportion of linolenic acid (C18:3) and eicosatrienoic acid (C20:3) were significantly elevated compared with wild type lines, whereas

palmitic acid (C16:0) and eicosenoic acid (C20:1) decreased notably, suggesting that *AhGPAT9* may display preferences for some specific PUFA substrates *in vivo*. Specially, α-linolenic acid (ALA/C18:3) is an essential omega-3 fatty acid and dietary component for human health, which can be converted into eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) for the development of nerve cells and brain, as well as anti-allergy and anti-lipemic (Lee et al., 2018; Zhang et al., 2018). Peanut kernels are mainly rich in oleic and linoleic acids, while the content of linolenic acid is less than 1% or even undetectable (Artemis, 2001). Thus, Overexpressed *AhGPAT9* to increase linolenic acid content will be helpful to improve the fatty acid composition of peanut and expand its functional health effects. There has been a view that the substrate preferences of endogenous acyltransferases play an important role in the utilization of unusual fatty acids in transgenic oilseeds (Snyder et al., 2009). Our research helps to support the contribution of the potential role of GPAT in TAG biosynthesis and its FA recombination.

Overexpressed *AhGPAT9* delayed the bolting time of transgenic lines and reduced the siliques number, but the weight and size of progeny seeds increased significantly, speculating that *AhGPAT9* might affect florescence and seed development of plant. Combined with the contribution of *AhGPAT9* to the biosynthesis and accumulation of oil in transgenic seeds, all these results point to a positive regulatory role for *AhGPAT9* in glycerolipid metabolism, and it may run through the whole process of seed development, which has great significance for the rational design of peanut lipid traits by biotechnology in the future. Of course, peanut oil synthesis is a quantitative trait genetically controlled by multiple genes and easily affected by the environment, so it is difficult to achieve precise regulation by changing the expression of one single gene, especially given that GPAT9 catalyzes the first step in the acylation reaction of the Kennedy pathway. TAG biosynthesis needs to undergo synergistic catalysis by three acyl-assembling enzymes (Chapman and Ohlrogge, 2012; Bates et al., 2013), and the effective flux and their cumulative effects in glycerolipid metabolism need to be further explained.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

YueS wrote the manuscript. YiS and YueS conceived the study and analyzed the data. YueS, YL, ML, and XZ performed experiments. YB provided valuable references and revision ideas. ZC revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the National Natural Science Foundation of China (Grant No. 31701461 to YueS), the Jiangsu Agricultural Science and Technology Innovation Fund (Grant No. CX(20)3121 to YueS), the Jiangsu Open Competition Project (Grant No. JBGS(2021)062 to YiS), and the earmarked fund for CARS-13 to ZC.

Acknowledgments

We profoundly thank Professor Wenhua Zhang (Nanjing Agricultural University) for providing ER marker.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1144306/full#supplementary-material

SUPPLEMENTARY TABLE 1
Primers sequences for PCR

SUPPLEMENTARY FIGURE 1
PCR amplification of peanut *AhGPAT9* gene. M represents a 2000 bp DNA marker, and the single band in right lane represents the target fragment of *AhGPAT9*.

SUPPLEMENTARY FIGURE 2
Structure analysis of peanut AhGPAT9 protein. (A) Hydrophobic prediction in ProtScale. (B) Secondary structure prediction in SOPMA. Blue = α-helix; Green = β-turn; Red = extended strand; Yellow = random coil. (C) Phosphorylation sites prediction in NetPhos (threshold > 0.5). (D) Transmembrane helices prediction in TMHMM. (E) Functional domain prediction in SMART.

SUPPLEMENTARY FIGURE 3
Expression of *AhGPAT9* in overexpression lines of *Arabidopsis*. Transcript level of *AhGPAT9* in 35-40 DAG* rosette leaves and 12-15 DAF siliques of *Arabidopsis* wild type (WT) and *AhGPAT9*-overexpressed lines (OE). DAG* = days after germination; DAP = days after pollination. The relative mRNA abundance was normalized with respect to that of peanut *AhACT11*, and data are shown as means ± standard deviation (n = 3).

SUPPLEMENTARY FIGURE 4
Bloting phenotype of homozygous *AhGPAT9* overexpression lines in *Arabidopsis*. (A) Wild type (WT) and *AhGPAT9* overexpression lines (OE) at 5-week-old under normal conditions. (B) Bloting rate statistics of WT and OE lines. Asterisks indicate significant differences from the WT: $^{**}P < 0.01$ (Student's *t*-test). Data are shown as means ± standard deviation (n = 12).

References

Artemis, P. S. (2001). N-3 fatty acids and human health: defining strategies for public policy. *Lipids* 36, S83–S89. doi: 10.1007/s11745-001-0687-7

Bai, Y., Shen, Y., Zhang, Z. Q., Jia, Q. R., Xu, M. Y., Zhang, T., et al. (2021). A GPAT1 mutation in *Arabidopsis* enhances plant height but impairs seed oil biosynthesis. *Int. J. Mol. Sci.* 22, 785. doi: 10.3390/ijms22020785

Bates, P. D. (2016). Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis. *BBA-Mol. Cell Biol. L.* 1861, 1214–1225. doi: 10.1016/j.bbalip.2016.03.021

Bates, P. D., Stymne, S., and Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Curr. Opin. Plant Biol.* 16, 358–364. doi: 10.1016/j.pbi.2013.02.015

Chapman, K. D., and Ohlrogge, J. B. (2012). Compartmentation of triacylglycerol accumulation in plants. *J. Biol. Chem.* 287, 2288–2294. doi: 10.1074/jbc.R111.290072

Chen, X., Chen, G. Q., Truksa, M., Snyder, C. L., Shah, S., and Weselake, R. J. (2014). Glycerol-3-phosphate acyltransferase 4 is essential for the normal development of reproductive organs and the embryo in *Brassica napus*. *J. Exp. Bot.* 65, 4201–4215. doi: 10.1093/jxb/eru199

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, X., Snyder, C. L., Truksa, M., Shah, S., and Weselake, R. J. (2011). Sn-Glycerol-3-phosphate acyltransferases in plants. *Plant Signal. Behav.* 6, 1695–1699. doi: 10.4161/psb.6.11.17777

Chen, G. Q., Woodfield, H. K., Pan, X., Harwood, J. L., and Weselake, R. J. (2015). Acyl-trafficking during plant oil accumulation. *Lipids* 50, 1057–1068. doi: 10.1007/s11745-015-4069-x

Chi, X., Yang, Q., Pan, L., Chen, N., Wang, T., Wang, M., et al. (2015). Isolation and expression analysis of glycerol-3-phosphate acyltransferase genes from peanuts (*Arachis hypogaea* l.). *Grasas Aceites* 66, e093. doi: 10.3989/gya.1190142

Cui, S. K., Hayashi, Y., Otomo, M., Mano, S., Oikawa, K., Hayashi, M., et al. (2016). Sucrose production mediated by lipid metabolism suppresses the physical interaction of peroxisomes and oil bodies during germination of *Arabidopsis thaliana*. *J. Biol. Chem.* 291, 19734–19745. doi: 10.1074/jbc.M116.748814

Gidda, S. K., Shockey, J. M., Rothstein, S. J., Dyer, J. M., and Mullen, R. T. (2009). *Arabidopsis thaliana* GPAT8 and GPAT9 are localized to the ER and possess distinct ER retrieval signals: functional divergence of the dilysine ER retrieval motif in plant cells. *Plant Physiol. Bioch.* 47, 867–879. doi: 10.1016/j.plaphy.2009.05.008

Graham, I. A. (2008). Seed storage oil mobilization. *Annu. Rev. Plant Biol.* 59, 115–142. doi: 10.1146/annurev.arplant.59.032607.092938

Hao, C. C., Liang, C. W., Shi, L., Li, H. Y., Chen, M. N., Pan, L. J., et al. (2018). Cloning and expression analysis of glycerol-3-phosphate acyltransferase (*GPAT*) genes in peanut. *J. Peanut Sci.* 47, 1–10. doi: 10.14001/j.issn.1002-4093.2018.01.001

Janila, P., Nigam, S. N., Pandey, M. K., Nagesh, P., and Varshney, R. K. (2013). Groundnut improvement: use of genetic and genomic tools. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00023

Jung, S., Swift, D., Sengoku, E., Patel, M., Teulé, F., Powell, G., et al. (2000). The high oleate trait in the cultivated peanut [Arachis hypogaea L.]. I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. *Mol. Gen. Genet.* 263, 796–805. doi: 10.1007/s004380000244

Kim, H. U., Lee, K. R., Jung, S. J., Shin, H. A., Go, Y. S., Suh, M. C., et al. (2015). Senescence-inducible LEC2 enhances triacylglycerol accumulation in leaves without negatively affecting plant growth. *Plant Biotechnol. J.* 13, 1346–1359. doi: 10.1111/pbi.12354

Lee, B. H., Choi, S. H., Kim, H. J., Jung, S. W., Kim, H. K., and Nah, S. Y. (2016). Plant lysophosphatidic acids: a rich source for bioactive lysophosphatidic acids and their pharmacological applications. *Biol. Pharm. Bull.* 39, 156–162. doi: 10.1248/bpb.b15-00575

Lee, A. Y., Choi, J. M., Lee, M. H., Lee, J., Lee, S., and Cho, E. J. (2018). Protective effects of perilla oil and alpha linolenic acid on SH- SY5Y neuronal cell death induced by hydrogen peroxide. *Nutr. Res. Pract.* 12, 93–100. doi: 10.4162/nrp.2018.12.2.93

Li, Y. H., Beisson, F., Koo, A. J. K., Molina, I., Pollard, M., and Ohlrogge, J. (2007). Identification of acyltransferases required for cutin biosynthesis and production of cutin with suberin-like monomers. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18339–18344. doi: 10.1073/PNAS.0706984104

Li, Y. H., Beisson, F., Pollard, M., and Ohlrogge, J. (2006). Oil content of *Arabidopsis* seeds: the influence of seed anatomy, light and plant-to-plant variation. *Phytochemistry* 67, 904–915. doi: 10.1016/j.phytochem.2006.02.015

Li-Beisson, Y., Pollard, M., Sauveplane, V., Pinot, F., Ohlrogge, J., and Beisson, F. (2009). Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. *Proc. Natl. Acad. Sci. U.S.A.* 106, 22008–22013. doi: 10.1073/pnas.0909090106

Liu, H. B., Wei, L. Y., Zhu, J. B., Zhang, B. X., Gan, Y., and Zheng, Y. P. (2022). Identification of GmGPATs and their effect on glycerolipid biosynthesis through seed-specific expression in soybean. *Mol. Biol. Rep.* 49, 9585–9592. doi: 10.1007/S11033-022-07852-W

Liu, F., Xia, Y., Wu, L., Fu, D., Hayward, A., Luo, J., et al. (2015). Enhanced seed oil content by overexpressing genes related to triacylglyceride synthesis. *Gene* 557, 163–171. doi: 10.1016/j.gene.2014.12.029

Lv, Y. Y., Zhang, X. R., Luo, L., Yang, H., Li, P. H., Zhang, K., et al. (2020). Characterization of glycerol-3-phosphate acyltransferase 9 (*AhGPAT9*) genes, their allelic polymorphism and association with oil content in peanut (*Arachis hypogaea* l.). *Sci. Rep.* 10, 14648. doi: 10.1038/s41598-020-71578-7

Misra, A., Khan, K., Niranjan, A., Kumar, V., and Sane, V. A. (2017). Heterologous expression of two GPATs from *Jatropha curcas* alters seed oil levels in transgenic *Arabidopsis thaliana*. *Plant Sci.* 263, 79–88. doi: 10.1016/j.plantsci.2017.07.003

Mohan, G. S., Pankaj, P., Atul, G., Yadav, P. V., Sadhana, S., and Zakwan, A. (2013). Cloning and characterization of GPAT gene from *Lepidium latifolium* l.: a step towards translational research in agri-genomics for food and fuel. *Mol. Biol. Rep.* 40, 4235–4240. doi: 10.1007/s11033-013-2505-7

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plantarum* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Nishida, I., Tasaka, Y., Shiraishi, H., and Murata, N. (1993). The gene and the RNA for the precursor to the plastid-located glycerol-3-phosphate acyltransferase of *Arabidopsis thaliana*. *Plant Mol. Biol.* 21, 267–277. doi: 10.1007/bf00019943

Payá-Milans, M., Aznar-Moreno, J. A., Balbuena, T. S., Haslam, R. P., Gidda, S. K., Pérez-Hormaeche, J., et al. (2016). Sunflower HaGPAT9-1 is the predominant GPAT during seed development. *Plant Sci.* 252, 42–52. doi: 10.1016/j.plantsci.2016.07.002

Payá-Milans, M., Venegas-Calerón, M., Salas, J. J., Garcés, R., and Martínez-Force, E. (2015). Cloning, heterologous expression and biochemical characterization of plastidial sn-glycerol-3-phosphate acyltransferase from *Helianthus annuus*. *Phytochemistry* 111, 27–36. doi: 10.1016/j.phytochem.2014.12.028

Shen, Y., Shen, Y., Liu, Y. H., Liang, M., Zhang, X. Y., and Chen, Z. D. (2022). Cloning and functional analysis of peanut *AhGPAT9* promoter. *Chin. J. Oil Crop Sci. Online.* doi: 10.19802/j.issn.1007-9084.2022126

Shen, Y., Shen, L. K., Shen, Z. X., Jing, W., Ge, H. L., Zhao, J. Z., et al. (2015). The potassium transporter OsHAK21 functions in the maintenance of ion homeostasis and tolerance to salt stress in rice. *Plant Cell Environ.* 38, 2766–2779. doi: 10.1111/pce.12586

Shi, Y. G., and Cheng, D. (2009). Beyond triglyceride synthesis: the dynamic functional roles of MGAT and DGAT enzymes in energy metabolism. *Am. J. Physiol. Endoc. M.* 297, E10–E18. doi: 10.1152/ajpendo.90949.2008

Shockey, J., Regmi, A., Cotton, K., Adhikari, N., Browse, J., and Bates, P. D. (2016). Identification of *Arabidopsis* GPAT9 (At5g60620) as an essential gene involved in triacylglycerol biosynthesis. *Plant Physiol.* 170, 163–179. doi: 10.1104/pp.15.01563

Singer, S. D., Chen, G. Q., Mietkiewska, E., Tomasi, P., Jayawardhane, K., Dyer, J. M., et al. (2016). *Arabidopsis* GPAT9 contributes to synthesis of intracellular glycerolipids but not surface lipids. *J. Exp. Bot.* 67, 4627–4638. doi: 10.1093/jxb/erw242

Snyder, C. L., Yurchenko, O. P., Siloto, R. M. P., Chen, X., Liu, Q., Mietkiewska, E., et al. (2009). Acyltransferase action in the modification of seed oil biosynthesis. *New Biotechnol.* 26, 11–16. doi: 10.1016/j.nbt.2009.05.005

Song, T. Z., Shi, Y. Y., Shen, L. K., Cao, C. J., Shen, Y., Jing, W., et al. (2021). An endoplasmic reticulum-localized cytochrome b_5 regulates high-affinity k^+ transport in response to salt stress in rice. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2114347118. doi: 10.1073/pnas.2114347118

Sun, S. K., Yang, N. N., Chen, L. J., Irfan, M., Zhao, X. H., and Li, T. L. (2015). Characterization of LpGPAT gene in *Lilium pensylvanicum* and response to cold stress. *BioMed. Res. Int.* 2015, 792819. doi: 10.1155/2015/792819

Tumaney, A. W., Shekar, S., and Rajasekharan, R. (2001). Identification, purification, and characterization of monoacylglycerol acyltransferase from developing peanut cotyledons. *J. Biol. Chem.* 276, 10847–10852. doi: 10.1074/jbc.M100005200

Xu, J., Pan, L. J., Li, H. Y., Wang, T., Chen, N., Chen, M. N., et al. (2021). Expression pattern analysis of genes related to lipid synthesis in peanut. *Acta Agron. Sin.* 47, 1124–1137. doi: 10.3724/SP.J.1006.2021.04105

Yang, Y., and Benning, C. (2018). Functions of triacylglycerols during plant development and stress. *Curr. Opin. Biotech.* 49, 191–198. doi: 10.1016/j.copbio.2017.09.003

Yang, S. U., Kim, J., Kim, H., and Suh, M. C. (2019). Functional characterization of physcomitrellapatens glycerol-3-phosphate acyltransferase 9 and an increase in seed oil content in *Arabidopsis* by its ectopic expression. *Plants* 8, 284. doi: 10.3390/plants8080284

Yang, W. L., Simpson, J. P., Li-Beisson, Y., Beisson, F., Pollard, M., and Ohlrogge, J. B. (2012). A land-plant-specific glycerol-3-phosphate acyltransferase family in *Arabidopsis*: substrate specificity, sn-2 preference, and evolution. *Plant Physiol.* 160, 638–652. doi: 10.1104/pp.112.201996

Zhang, M., Fan, J. L., Taylor, D. C., and Ohlrogge, J. B. (2009). DGAT1 and PDAT1 acyltransferases have overlapping functions in *Arabidopsis* triacylglycerol biosynthesis and are essential for normal pollen and seed development. *Plant Cell* 21, 3885–3901. doi: 10.1105/tpc.109.071795

Zhang, Q. Y., Yu, R., Xie, L. H., Rahman, M. M., Kilaru, A., Niu, L. X., et al. (2018). Fatty acid and associated gene expression analyses of three tree peony species reveal key genes for α-linolenic acid synthesis in seeds. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00106

Zheng, Z. F., Xia, Q., Dauk, M., Shen, W. Y., Selvaraj, G., and Zou, J. T. (2003). *Arabidopsis* AtGPAT1, a member of the membrane-bound glycerol-3-phosphate acyltransferase gene family, is essential for tapetum differentiation and male fertility. *Plant Cell* 15, 1872–1887. doi: 10.1105/tpc.012427

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 March 2023
DOI 10.3389/fpls.2023.1154208

OPEN ACCESS

EDITED BY
Hongbo Chao,
Zhengzhou University, China

REVIEWED BY
Libin Zhang,
Huazhong University of Science and Technology, China
Yu Liang,
Guangxi Normal University, China
Raheel Shahzad,
Universitas Muhammadiyah Bandung, Indonesia

*CORRESPONDENCE
Jiana Li
ljn1950@swu.edu.cn
Cunmin Qu
drqucunmin@swu.edu.cn

SPECIALTY SECTION
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

RECEIVED 30 January 2023
ACCEPTED 17 February 2023
PUBLISHED 09 March 2023

CITATION
Guan M, Shi X, Chen S, Wan Y, Tang Y, Zhao T, Gao L, Sun F, Yin N, Zhao H, Lu K, Li J and Qu C (2023) Comparative transcriptome analysis identifies candidate genes related to seed coat color in rapeseed. *Front. Plant Sci.* 14:1154208. doi: 10.3389/fpls.2023.1154208

Comparative transcriptome analysis identifies candidate genes related to seed coat color in rapeseed

Mingwei Guan[1,2,3], Xiangtian Shi[1,2,3], Si Chen[1,2,3], Yuanyuan Wan[1,2,3], Yunshan Tang[1,2,3], Tian Zhao[1,2,3], Lei Gao[1,2,3], Fujun Sun[1,2,3], Nengwen Yin[1,2,3], Huiyan Zhao[1,2,3], Kun Lu[1,2,3], Jiana Li[1,2,3]* and Cunmin Qu[1,2,3]*

[1]Integrative Science Center of Germplasm Creation in Western China (CHONGQING) Science City and Southwest University, College of Agronomy and Biotechnology and Academy of Agricultural Sciences, Southwest University, Chongqing, China, [2]Academy of Agricultural Sciences, Southwest University, Chongqing, China, [3]Affiliation Engineering Research Center of South Upland Agriculture, Ministry of Education, Chongqing, China

Yellow seed coat in rapeseed (*Brassica napus*) is a desirable trait that can be targeted to improve the quality of this oilseed crop. To better understand the inheritance mechanism of the yellow-seeded trait, we performed transcriptome profiling of developing seeds in yellow- and black-seeded rapeseed with different backgrounds. The differentially expressed genes (DEGs) during seed development showed significant characteristics, these genes were mainly enriched for the Gene Ontology (GO) terms carbohydrate metabolic process, lipid metabolic process, photosynthesis, and embryo development. Moreover, 1206 and 276 DEGs, which represent candidates to be involved in seed coat color, were identified between yellow- and black-seeded rapeseed during the middle and late stages of seed development, respectively. Based on gene annotation, GO enrichment analysis, and protein–protein interaction network analysis, the downregulated DEGs were primarily enriched for the phenylpropanoid and flavonoid biosynthesis pathways. Notably, 25 transcription factors (TFs) involved in regulating flavonoid biosynthesis pathway, including known (e.g., KNAT7, NAC2, TTG2 and STK) and predicted TFs (e.g., C2H2-like, bZIP44, SHP1, and GBF6), were identified using integrated gene regulatory network (iGRN) and weight gene co-expression networks analysis (WGCNA). These candidate TF genes had differential expression profiles between yellow- and black-seeded rapeseed, suggesting they might function in seed color formation by regulating genes in the flavonoid biosynthesis pathway. Thus, our results provide in-depth insights that facilitate the exploration of candidate gene function in seed development. In addition, our data lay the foundation for revealing the roles of genes involved in the yellow-seeded trait in rapeseed.

KEYWORDS

***Brassica napus* L., yellow trait, transcriptome, flavonoid pathway, protein-protein interaction**

Introduction

Rapeseed (*Brassica napus* L., AACC, 2n = 38) is an important oilseed crop worldwide, serving as a source of edible vegetable oil and feed meal (Saeidnia and Gohari, 2012; Carruthers et al., 2017; Kaur et al., 2020). Increasing seed oil content is an important breeding objective in rapeseed. Improving the seed oil content of rapeseed by 1% is equivalent to a 2.3–2.5% increase in seed yield (Wang, 2004). The vegetable oil from rapeseed is excellent for human health (Lu et al., 2011; Liu et al., 2016). Therefore, the demand for high-quality rapeseed has risen sharply. Yellow-seeded rapeseed has many favorable qualities over black-seeded varieties, such as higher oil and protein contents, lower fiber content, and fewer pigments and polyphenols (Tang et al., 2010; Chao et al., 2022). Therefore, the yellow-seeded trait is a major breeding objective for rapeseed.

Seeds are specific sites of oil accumulation in *B. napus*. Seed development in rapeseed, like the model plant Arabidopsis (*Arabidopsis thaliana*), can divided into two phases: embryogenesis and seed maturation (Ruuska et al., 2002; Locascio et al., 2014). Carbohydrates, proteins, triacylglycerol (TAG), and pigments are important reserves in developing seeds. Many genes involved in the biosynthetic pathways of these compounds have been identified in Arabidopsis (White et al., 2001; Zhang M. et al., 2009; Lu et al., 2020; Seok et al., 2021). Among these pathways, the flavonoid biosynthesis pathway is a 'model' for secondary metabolite production. Flavonoids perform different functions in plants, including color formation and resistance to abiotic and biotic stresses (Nishihara and Nakatsuka, 2011; Chezem and Clay, 2015; Davies et al., 2018). Much effort has focused on revealing the flavonoid biosynthesis pathway involved in seed color formation using *Arabidopsis TRANSPARENT TESTA* (*TT*) mutants, such as mutants of *TT1-TT19*, *TTG1*, and *TTG2* (Albert et al., 1997; Dong et al., 2001; Bharti and Khurana, 2003; Kerhoas et al., 2006; Fu et al., 2007; Stracke et al., 2007; Yoshida et al., 2008; Cutanda-Perez et al., 2009; Baudry et al., 2010; Appelhagen et al., 2011; Chen et al., 2013; Pang et al., 2013; Stein et al., 2013; Zhang et al., 2013; Gesell et al., 2014; Ichino et al., 2015; Gonzalez et al., 2016; Zhang et al., 2017). Numerous genes involved in the formation of seed coat color have also been identified in *Brassica* species. For example, *TRANSPARENT TESTA GLABRA 1* (*TTG1*) shares the same functions in *B. rapa* as its Arabidopsis counterpart, controlling both hairiness and seed coat color traits in this crop (Zhang J.F. et al., 2009). The bHLH transcription factor (TF) gene *TT8* regulates the accumulation of proanthocyanidins (PAs) and controls seed coat color in *Brassica* species (Li et al., 2012); silencing of *TT1* genes and knockout of *TT2* homologs altered seed coat color in *B. napus* (Lian et al., 2017; Xie et al., 2020). In *B. napus*, flavonoid biosynthesis pathway genes, including the P-type H^+-ATPase genes (*AHA10*), *TT10*, *TT12*, and *TTG1*, are located within the quantitative trait locus (QTL) region for seed coat color on chromosome A09 (Liu et al., 2005; Badani et al., 2006; Fu et al., 2007; Xiao et al., 2007; Yan et al., 2009; Zhang et al., 2011; Stein et al., 2013; Zhang et al., 2013; Qu et al., 2015). In addition, several candidate genes (e.g., MYB2, MYB3, MYB4, TT8, and MYBL2.1) involved in PA accumulation have been identified in *Brassica* species using transcriptomic analysis, providing important information for identifying candidate genes for seed coat color (Hong et al., 2017; Jiang et al., 2019; He et al., 2022). However, the molecular mechanisms controlling the yellow-seeded trait in rapeseed remain largely unexplored.

To explore the dynamic regulation of seed coat color formation, we comprehensively investigated the dynamic changes in gene expression during seed development in yellow- and black-seeded rapeseed by performing multistage comparative transcriptomic analysis. Many candidate genes were uncovered, such as *KNOTTED-like homeobox of Arabidopsis thaliana 7* (*KNAT7*), *SEPALLATA 2* (*SEP2*), *G-box binding factor 6* (*GBF6*), *SEEDSTICK* (*STK*), and *TTG2*, which play important roles in regulating seed development and seed coat color formation in rapeseed. Notably, some candidate genes, including *BnaA02g17180D*, *BnaC06g22430D*, *BnaA07g21710D*, *BnaA05g00070D*, and *BnaA04g01810D*, were differentially expressed in black-seeded vs. yellow-seeded *B. napus* during seed development. These results increase our understanding of the gene networks involved in yellow trait formation during seed development in rapeseed, which will facilitate breeding efforts for this trait.

Materials and methods

Plant materials

Seven inbred rapeseed (*Brassica napus*) lines, four yellow-seeded lines (GH06, 16G15, 16G16, and 16G47), and three black-seeded lines (ZY821, 16G48, and 17G56) were grown in experimental field in Beibei (106.38°E, 29.84°N), Chongqing, China. Inbred lines GH06 (Y1) and ZY821 (B1) are typical yellow- and black-seeded *B. napus* lines (Qu et al., 2013), respectively. Two groups of near-isogenic lines developed by Chongqing Rapeseed Technology Research Center (CRTRC) were planted during 2019–2020 in the field in CRTRC in Beibei, Chongqing, China: 16G15 (Y2), 16G16 (Y3) and 16G48 (B2); and 17G56 (B3), and 16G47 (Y4). At the flower initiation stage, individual flowers were marked with different colored wool to ensure that seed development could be dated exactly. Subsequently, seeds were sampled at 15, 30, and 50 DAP (days after pollination). At each stage, samples from five individual plants were pooled as a biological replicate, immediately frozen in liquid nitrogen, and stored at −80°C until total RNA extraction.

RNA extraction, cDNA library construction, and sequencing

For each sample, 3 μg RNA was used as input material for cDNA library preparation. Total RNA was extracted from the samples using an RNAprep Pure Plant Kit (Tiangen Biotech, Beijing, China) following the manufacturer's instructions, and index codes were added to attribute sequences to each sample. First-strand cDNA was synthesized using random hexamer primers. Second-strand cDNA

synthesis was then performed using DNA Polymerase I and RNase H. The remaining overhangs were converted into blunt ends using a MicroPoly (A) Purist Kit (Ambion, USA). The library fragments were purified with an AMPure XP system (Beckman Coulter, Beverly, MA, USA). Size-selected, adaptor-ligated cDNA was digested with 3 µl USER Enzyme (NEB, USA) at 37°C for 15 min and then at 95°C for 5 min. The fragments were amplified by PCR using Phusion High-Fidelity DNA Polymerase, Universal PCR Primers, and Index (X) Primer. Following quality checking, the RNA was used for library construction using a NEB Next Ultra RNA Library Prep Kit for Illumina with an insert size of 300 bp. The Illumina HiSeq X was used, which generates short reads of 150 bp in PE mode.

Identification of differential gene expression (DEGs)

After removing low-quality reads, Illumina sequencing reads were mapped to the *B. napus* Darmor-bzh reference genome (http://www.genoscope.cns.fr/brassicanapus) (Chalhoub et al., 2014) using HISAT with default settings (Kim et al., 2015). The bam files of uniquely mapped reads were used as inputs for HTseq, and FPKM (fragments per kilobase of transcript per million mapped reads) values were calculated to measure the expression levels of genes (Roberts et al., 2011; Anders et al., 2015). The Pearson correlation coefficient between biological replicates was calculated based on the normalized expression levels of log2(FPKM + 1). DEGs were recognized by Cuffdiff with a cut-off of log2 fold change (FC) ≥ 1 and a false discovery rate (FDR) ≤ 0.05 (Wang et al., 2009; Love et al., 2014).

Enrichment analysis of DEGs

Clustering analysis was performed using the TCseq clustering package, and PCA was performed using the Pheatmap package. Transformed and normalized gene expression values with log2 (FPKM + 1) were used for hierarchical clustering.

GOseq was used to obtain significant GO terms for the DEGs in the significant modules (Young et al., 2010), and the clusterProfiler package in R was used to perform KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis of DEGs in the modules: $P < 0.02$ was selected as the cut-off value.

Identification of hub genes and sub-networks associated with seed color

Hub genes, which are highly interconnected with nodes in a module, are functionally significant. The top 25% of genes evaluated by Cytoscape_v3.7.2 with the CytoNCA (BC, Betweenness centrality; CC, Closeness centrality; DC, Degree centrality) application of the STRING database (http://www.string-db.org) were considered to be hub genes. Hub genes were identified among DEGs between black- and yellow-seeded rapeseed. Protein–protein interaction (PPI) networks were established with STRING, with a cut-off confidence score > 0.4 (medium confidence). Cytoscape_v3.7.2 was then used to calculate the sub-modules of the DEG interaction network. Hub genes were identified according to BC, CC, and DC (Shannon et al., 2003; Chin et al., 2014). The top 25% of genes were ultimately identified as crucial genes based on the centrality values of genes in the PPI network.

Weight gene co-expression networks analysis (WGCNA) and integrative gene regulatory network (iGRN) analysis to identify target genes

The RNA-seq data were analyzed to construct gene co-expression networks using the R package WGCNA (Langfelder and Horvath, 2008). To reduce noise, genes with criterion FPKM ≤ 3 in each sample were excluded. In total, 21 samples were used for analysis, including 7 lines each at the early, middle, and late stages of seed development. The following parameters were used to identify each gene module: weighted network, unsigned; hierarchical clustering tree, dynamic hybrid tree cut algorithm; power, 30; and minimum module size, 250.

Gene regulation is a dynamic process in which TFs play an important role in controlling spatiotemporal gene expression (De Clercq et al., 2021). To identify TFs closely related to seed color, iGRN was used to predict candidate genes. iGRN is used to determine the enrichment of TFs associated with (i.e., regulating) a set of input genes. Enrichment statistics were computed using hypergeometric distribution combined with Benjamini–Hochberg correction for multiple hypotheses testing (with a q-value cut-off of 1e-3). Hub genes confirmed by PPI and Cytoscape_v3.7.2 were used to predict candidate TFs. The results of analysis, combined with the classification of hub genes by WGCNA, were jointly used to predict TFs that are closely associated with the formation of seed color in rapeseed. The expression values of the TFs were calculated using FPKM of materials with the same seed color. A heatmap was generated using the R package Pheatmap (scale = "row", cluster_row = T) (Diao et al., 2018).

Results

Overview of sequencing data analysis

Through library construction and sequencing, 158.95 million reads in yellow- and black-seeded *B. napus* during seed development were generated using the Illumina sequencing platform, and deposited in the NCBI database (Accession No., PRJNA931458). An average of 89.26% of the reads were mapped to the *B. napus* Darmor-*bzh* reference genome (version 4.1, http://www.genoscope.cns.fr/brassicanapus/data/; Chalhoub et al., 2014) using HTseq version 2.2, and the expression profiles of the genes were quantified in terms of FPKM. Of these mapped reads, an average of 83.32% were uniquely mapped and used to calculate normalized gene expression levels. Detailed information about the transcriptome sequencing data is given in Table 1. Based on the *B. napus* reference genome (http://www.genoscope.cns.fr/brassicanapus), 89,136 expressed genes were detected during seed development. After eliminating 31,089 genes with lower expression levels (FPKM < 1) in all samples, the 58,047 remaining genes were used for further analysis.

Identification and annotation of differentially expressed genes during seed development

In this study, 31,843, 32,538, and 24,944 genes with high expression levels (FPKM ≥ 1) were identified at 15, 30, and 50 days after pollination (DAP), respectively (Figure 1A). Of these genes, 21,563 were expressed throughout seed development, and 3,120, 2,321, and 1,427 were specifically expressed at 15, 30, and 50 DAP, respectively. Based on the clustering analysis and principal component analysis (PCA) of the differentially expressed genes (DEGs), these DEGs were divided into three groups, which are consistent with the stages of seed development: early (E, 15 DAP), middle (M, 30 DAP), and late (L, 50 DAP) stages (Figure 1B).

Moreover, the expression levels of the DEGs during the three stages of seed development were investigated *via* pairwise comparisons. Using DEGseq and DESeq, the DEGs with adjusted log2 fold change (FC) > 1 and q-value ≤ 0.05 were further investigated. The results of DEG analysis between materials are shown in Table S1. The results showed that 3,240, 847, and 1,343 DEGs were specifically expressed during the early, middle, and late stages of seed development, respectively, in all materials (Figure 1C). However, other genes did not show stable differences in expression among materials during seed development. In addition, a total 1,224 DEGs (FPKM ≥ 1), including 438 genes with high expression levels (FPKM ≥ 10) during seed development, had stable expression levels throughout development. These genes could be core genes for seed development in rapeseed. These DEGs were grouped into six clusters by K-means clustering (TCseq package; Figure 2). The expression of DEGs in clusters 1 and 5 peaked during early seed development, while the expression of DEGs in clusters 2 and 6 peaked during late seed development, and genes in the other clusters were expressed at their highest levels during the middle stage of seed development (Figure 2).

Besides, the functions of these DEGs were predicted using Gene Ontology (GO) and KEGG enrichment analysis. The GO terms were divided into the biological process (BP), cellular component (CC), and molecular function (MF) categories (Figure 3). As shown in Figure 3, the DEGs were enriched in different GO terms during the three stages of development. For example, the most highly enriched terms during early seed development included carbohydrate metabolic process in the BP category and catalytic activity and enzyme inhibitor activity in the MF category. During the middle stage of seed development, the

TABLE 1 Summary statistics of RNA-Seq results during seed development in *B. napus*.

Sample	Total reads	Total mapped reads	Uniquely mapped reads	Mapped paired-end reads
Y1-15DAP	117,534,018	90.23%	84.51%	81.27%
Y1-30DAP	130,975,684	90.84%	84.67%	83.05%
Y1-50DAP	135,283,922	91.15%	80.63%	82.78%
B1-15DAP	116,634,942	89.24%	83.21%	80.03%
B1-30DAP	132,568,802	90.56%	84.38%	82.85%
B1-50DAP	127,441,128	91.28%	80.88%	82.52%
Y2-15DAP	64,070,908	89.75%	85.41%	82.96%
Y2-30DAP	49,766,158	87.05%	82.77%	78.96%
Y2-50DAP	46,453,248	87.32%	79.97%	77.04%
Y3-15DAP	53,759,758	89.41%	84.79%	83.29%
Y3-30DAP	54,646,278	84.22%	78.81%	75.99%
Y3-50DAP	49,294,400	88.66%	81.73%	79.32%
B2-15DAP	48,872,142	88.60%	84.02%	81.77%
B2-30DAP	47,936,028	86.30%	80.82%	77.79%
B2-50DAP	56,744,446	86.91%	80.03%	76.88%
Y4-15DAP	55,991,424	91.34%	86.76%	84.54%
Y4-30DAP	67,242,064	91.30%	86.91%	84.31%
Y4-50DAP	47,889,758	91.39%	85.76%	84.61%
B3-15DAP	58,565,488	89.82%	85.26%	82.66%
B3-30DAP	61,333,478	90.17%	85.63%	83.13%
B3-50DAP	66,537,336	89.01%	82.69%	82.01%

Total reads indicates all clean reads in every sample; total mapped reads indicates the ratio of mapped reads in every sample; uniquely mapped reads indicates the ratio of reads in each sample that mapped to a single location; mapped paired-end reads indicates the ratio of paired-end reads that mapped to the genome. DAP, days after pollination.

10.3389/fpls.2023.1154208

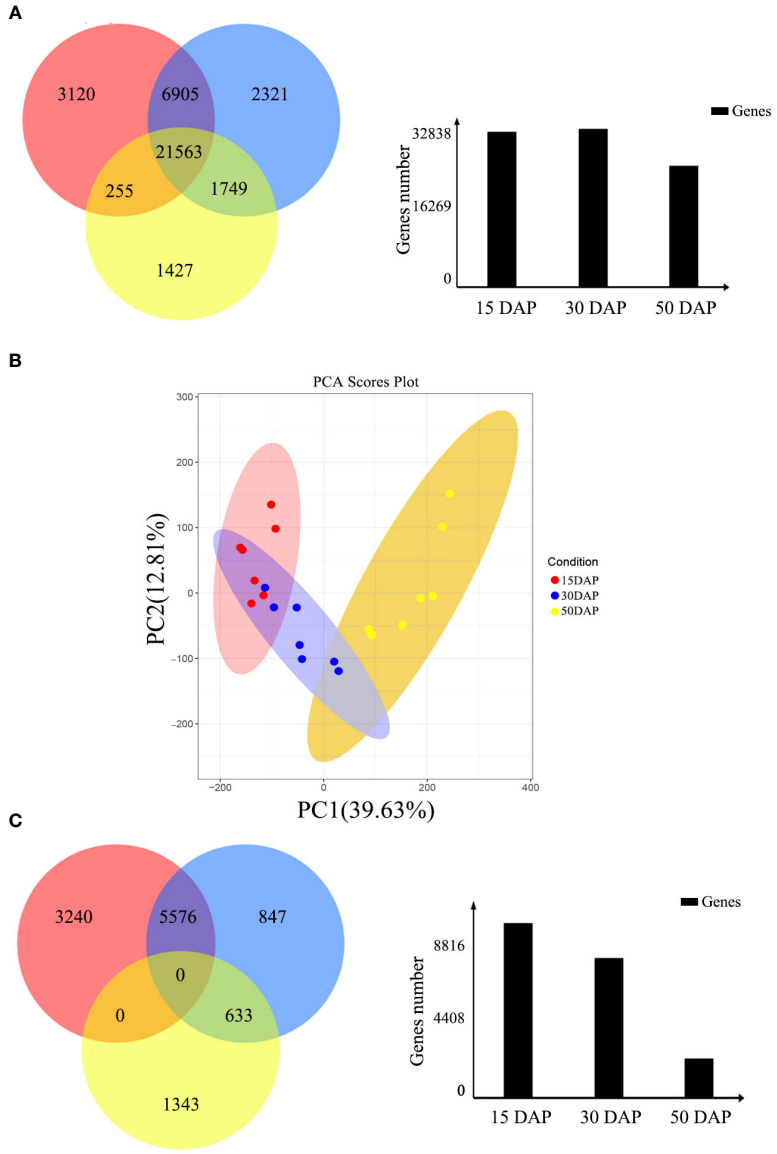


FIGURE 1
Identification of differentially expressed genes during seed development. **(A)** Venn diagram of the number of genes expressed during seed development. The red, blue, and yellow circles represent the early, middle, and late stages of seed development, respectively. The numbers represent the number of genes. The bar chart shows the total number of genes expressed during the early (15 DAP), middle (30 DAP), and late (50 DAP) stages of seed development. **(B)** Cluster dendrogram showing three distinct developmental stages: the early, middle, and late stages of seed development. **(C)** Venn diagram of DEGs expressed during seed development. The red, blue, and yellow circles represent the early, middle, and late stages of seed development, respectively. The numbers represent the number of DEGs. The bar chart shows the total number of DEGs expressed during the early (15 DAP), middle (30 DAP), and late (50 DAP) stages of seed development.

most highly enriched GO terms included photosynthesis in the BP category and photosynthetic membrane and photosystem in the CC category. During late seed development, the most highly enriched GO terms included embryo development in the BP category, lipid droplet and monolayer-surrounded lipid storage body in the CC category, and nutrient reservoir activity in the MF category (Table S2). KEGG enrichment result showed that the pathways involved mainly carbohydrate, starch and sucrose metabolism in the early seed development; photosynthesis proteins and photosynthesis metabolism in the middle stage; and cutin, suberine and wax biosynthesis, transporters, cytochrome P450 and lipid in the late stage, etc. (Table S3). These results suggest that these DEGs play different roles during seed development in *B. napus*.

GO enrichment analysis of the 438 core genes with stable, high expression levels during seed development revealed that these genes were clustered into 89 GO terms. The top three terms were translation regulator activity, translation factor activity, and RNA binding and translation regulator activity in the MF category; intracellular anatomical structure, organelle, and intracellular organelle in the CC category; and gene expression, translational initiation, and cellular nitrogen compound metabolic process in the BP category (Table S2). KEGG enrichment result showed the most highly enriched in the pathway of translation factors, protein families: genetic information

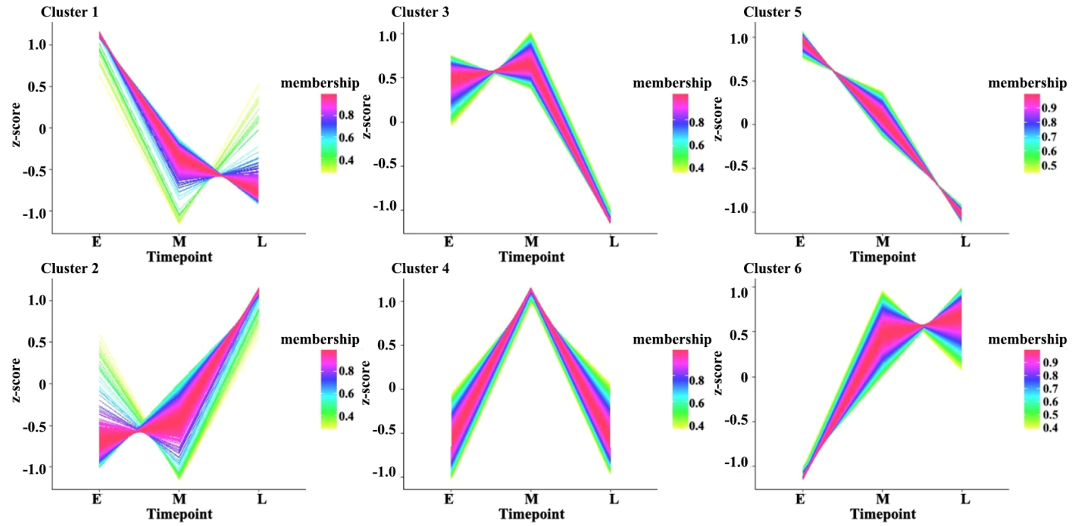


FIGURE 2
Clustered gene expression profiles in developing seeds. The clusters were defined based on the temporal expression profiles of genes using K-means clustering in R. The Y-axis represents the standardized FPKM values of genes, and E, M, and L on the X-axis represent seed samples at 15, 30, and 50 DAP, respectively.

processing and spliceosome, etc. (Table S3). These results suggest that these DEGs may play crucial roles in translation and the regulation of seed development in *B. napus*.

Identification of DEGs between the yellow- and black-seeded rapeseed

To identify changes in gene expression related to the yellow-seeded trait, the expression patterns of DEGs were further investigated between yellow- and black-seeded lines during seed development (15, 30, and 50 DAP). In total, 516, 1,206, and 276 DEGs between yellow- and black-seeded rapeseed were obtained in the early, middle, and late stages of seed development, respectively (Figure 4). The highest number of DEGs was observed in the middle stage of seed development (1,206 DEGs, including 982 upregulated and 224 downregulated DEGs in yellow-seeded materials), while fewer DEGs were detected in the late stage of seed development (276 DEGs, including 96 upregulated and 180 downregulated DEGs in yellow-seeded materials). The DEGs in the middle stage of seed development may play important roles in seed color formation. In addition, we identified 47 common DEGs during all three stages of seed development, including 27 upregulated and 20 downregulated genes (Table S4).

Identification of hub genes associated with seed color *via* PPI network analysis

To explore the molecular mechanism of seed color formation in *B. napus*, protein–protein interaction (PPI) networks were constructed using the STRING online database, with parameters including a minimum required interaction score > 0.4 (medium confidence). The upregulated and downregulated DEGs in yellow-seeded *B. napus* seeds at three stages of development were submitted to the STRING database for predicting PPIs and construct PPI networks (Figure S1; Table S5). Strikingly, the PPI networks of downregulated DEGs were highly consistent, especially during the middle and late stages of seed development; these genes are mainly associated with the proanthocyanidin and flavonoid biosynthesis pathways (Table S5). To further elucidate the interactions of these genes, these downregulated and repeatedly detected DEGs in the middle and late stages of seed development were submitted to the STRING database and reconstructed the PPI networks. We then selected the top 25% of genes as hub genes following evaluation by Cytoscape_v3.7.2 with the CytoNCA (BC, CC, DC) application (Figure 5). These hub genes, including *TT* (*TT1*, *TT4*, *TT8*, *TT12*, and *TT16*), *BANYULS* (*BAN*), *Cinnamate-4-hydroxylase* (*C4H*), *AHA10*, *Leucoanthocyanidin dioxygenase* (*LDOX*), *MYB-LIKE 2* (*MYBL2*), and *Phenylalanine ammonia-lyase* (*PAL*), were downregulated in yellow- vs. black-seeded rapeseed (Table S5), suggesting they might play important roles in seed color formation in rapeseed.

Weight gene co-expression network analysis (WGCNA) to identify DEGs for seed color in rapeseed

To further identify the genes/functional pathways associated with seed color, the weighted gene co-expression networks of yellow- and black-seeded rapeseed were constructed using RNA-seq data in this study. To reduce noise, genes with FPKM ≤ 3 in each sample were excluded. The co-expression networks were constructed based on pairwise correlations (>0.8) and minModuleSize = 250. A total of 15 distinct modules labeled with

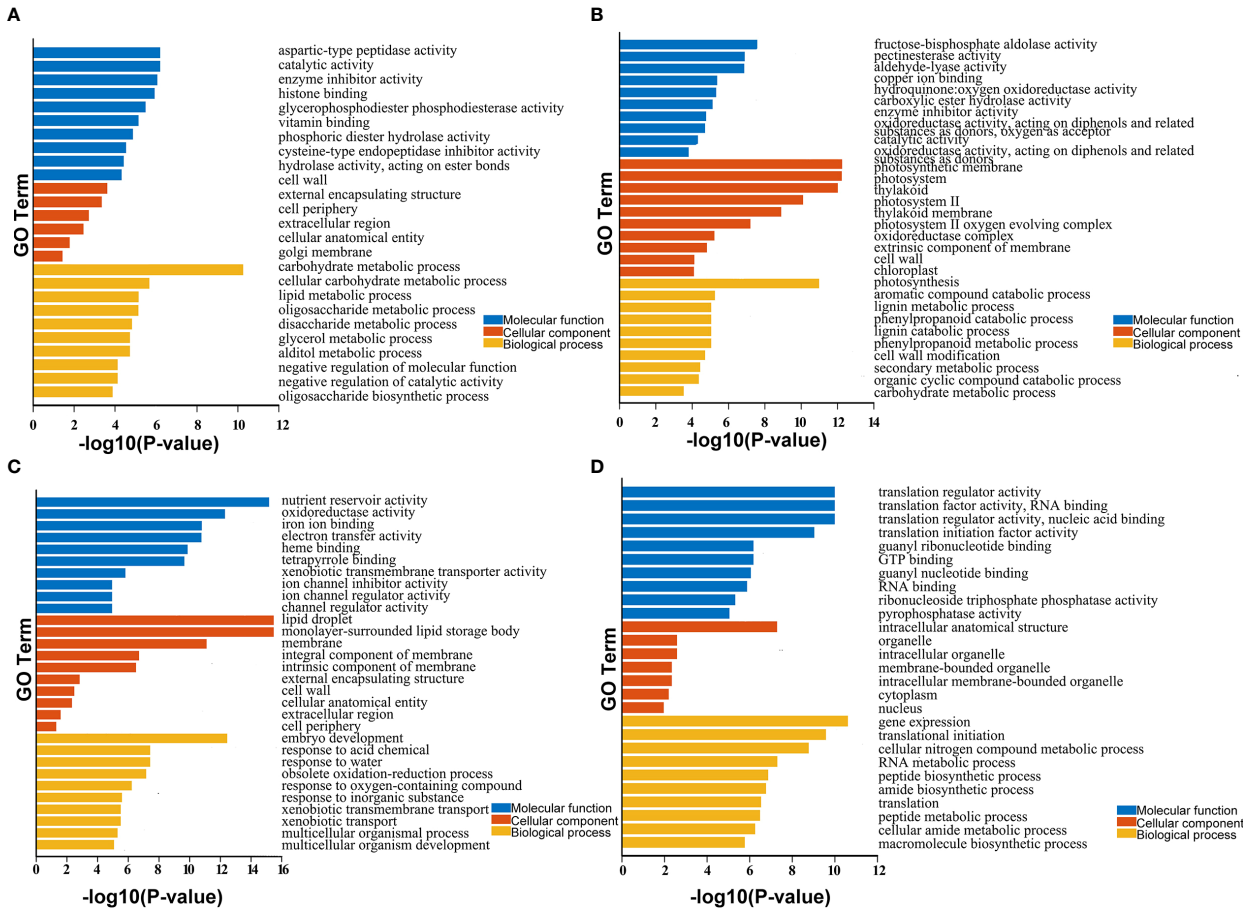


FIGURE 3
GO enrichment analysis of identified DEGs and core genes in developing seeds. **(A–D)** show enriched GO terms in seeds at 15, 30, and 50 DAP and 438 core genes with stable, high expression levels during seed development. Only significant categories (top 10 sorted by P-value in MF, CC, and BP) are displayed.

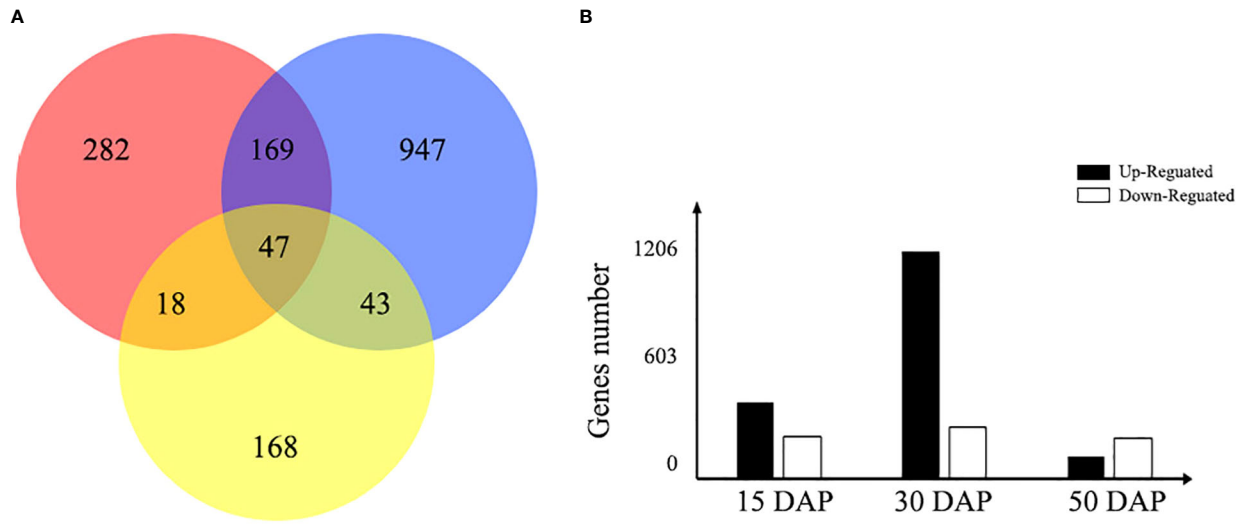


FIGURE 4
Venn diagram of the number of DEGs expressed during seed development. **(A)** The red, blue, and yellow circles represent the early, middle, and late stages of seed development, respectively. The number represents the number of DEGs between yellow- and black-seeded rapeseed. **(B)** The bar chart shows the total number of DEGs expressed during early (15 DAP), middle (30 DAP), and late (50 DAP) seed development. Black bars indicate upregulated DEGs in yellow seeds, and white bars indicate downregulated DEGs in yellow seeds.

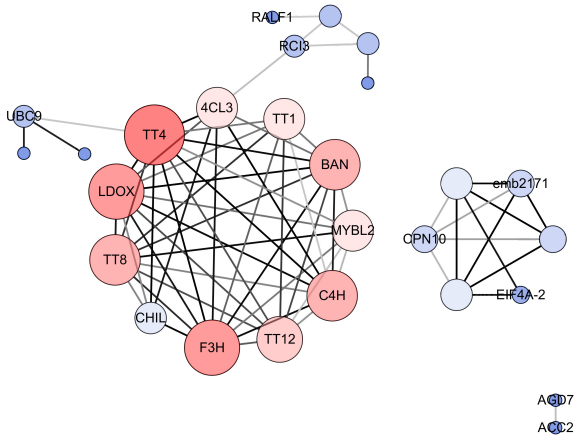


FIGURE 5
PPI networks for hub genes involved in *B. napus* seed color. The size and color depth of the circle represent the degree of center, and the color depth of the line represents the degree of correlation. Crucial genes were identified based on eigenvector centrality (EGC), degree centrality (DC), and closeness centrality (CC), which were calculated using the Cytoscape_v3.7.2 plug-in CytoNCA.

different colors were identified, with the number of genes per module ranging from 258 (orange) to 13,402 (gray) (Figure 6).

The hub genes detected by PPI network analysis belonged to four modules (Table 2). Most hub genes (23/28) were classified into the orange module, which were significantly enriched in the GO terms sulfur compound biosynthetic process, aromatic amino acid family metabolic process, and sulfur compound metabolic process. The blue and red modules each contained two hub genes, which were strongly enriched in the GO terms nuclear processes and macromolecule biosynthetic process, respectively. The salmon module, with one hub gene, was significantly enriched in protein transport, cellular macromolecule localization, and protein localization (Table S6). These results indicate that the modules containing hub genes are most strongly correlated with seed color, thus representing suites of interconnected genes underlying the regulation of seed color formation.

The integrated gene regulatory network is linked to the flavonoid biosynthesis pathway in rapeseed

Flavonoids play important roles in seed color formation and are much more complex in *Brassica* species than in the model plant Arabidopsis (Qu et al., 2016). To enhance our understanding of the regulatory gene network of the flavonoid biosynthesis pathway, an iGRN associated with *B. napus* seed color were constructed using the DEGs between the yellow- and black-seeded rapeseed. Twenty-five TFs highly interacted with hub genes involved in the flavonoid biosynthesis pathway (Table 3). Among genes known to be involved in the flavonoid biosynthesis pathway, the homologous genes of *KNAT7* (*BnaA09g52990D*), *NAC2* (*NAC domain containing protein 2, BnaC08g43050D* and *BnaCnng64100D*), *STK* (*BnaA03g24210D*, *BnaAnng39120D* and *BnaCnng46740D*), and *SEP2* (*BnaC05g48320D*, *BnaA01g33070D* and *BnaC09g42060D*) were expressed at higher levels in black-seeded materials, whereas TTG2 (*BnaA03g17120D* and *BnaC03g20650D*) showed the opposite pattern (Figure 7A). We also predicted novel TF genes, including *Integrase-type DNA-binding superfamily* (*BnaA10g00620D*, *BnaC05g00680D*, and *BnaCnng08620D*), *bZIP44* (*basic leucine-zipper 44*; *BnaA02g17180D*, *BnaC06g22430D*, and *BnaA07g21710D*), *Exonuclease family* (*BnaA05g00070D*), *SHP1* (*BnaA04g01810D*, and *BnaA07g18050D*), which showed differential expression levels between yellow- and black-seeded rapeseed (Figure 7A). Notably, five TF genes that were categorized in the orange and red modules by WGCNA (Figure 7B) had similar expression trends to most hub genes, which were closely related to the flavonoid pathway or proanthocyanidin pathway. Six TFs in the salmon module regulate the expression of the hub genes *via Bn_MYBL2* (*BnaC06g32180D*), and other TFs in the blue module play important roles in regulating the flavonoid biosynthesis pathway *via Bn_TT1a* (*BnaC06g08390D*) and *Bn_TT16a* (*BnaA03g39500D*) (Figure 7). These findings suggest that these TFs might be involved in regulating flavonoid pathways, thus affecting seed color in rapeseed.

Discussion

In rapeseed, the yellow-seeded trait is desirable because it offers many advantages, including higher oil and protein content, lower pigment and polyphenol contents, and lower fiber content, compared to black-seeded varieties (Abraham and Bhatia, 1986; Simbaya et al., 1995; Tang et al., 2010). To date most yellow-seeded varieties were developed *via* the interspecific hybridization of *Brassica* species (Chen and Heneen., 1992; Liu et al., 2010; Qu et al., 2013), but the genetic stability and underlying mechanism remain unclear. As known, seed color is considered to be a typical quantitative trait. Despite numerous studies involving QTL analysis, GWAS, and candidate gene identification have been performed to elucidate the molecular mechanism underlying yellow seed coat

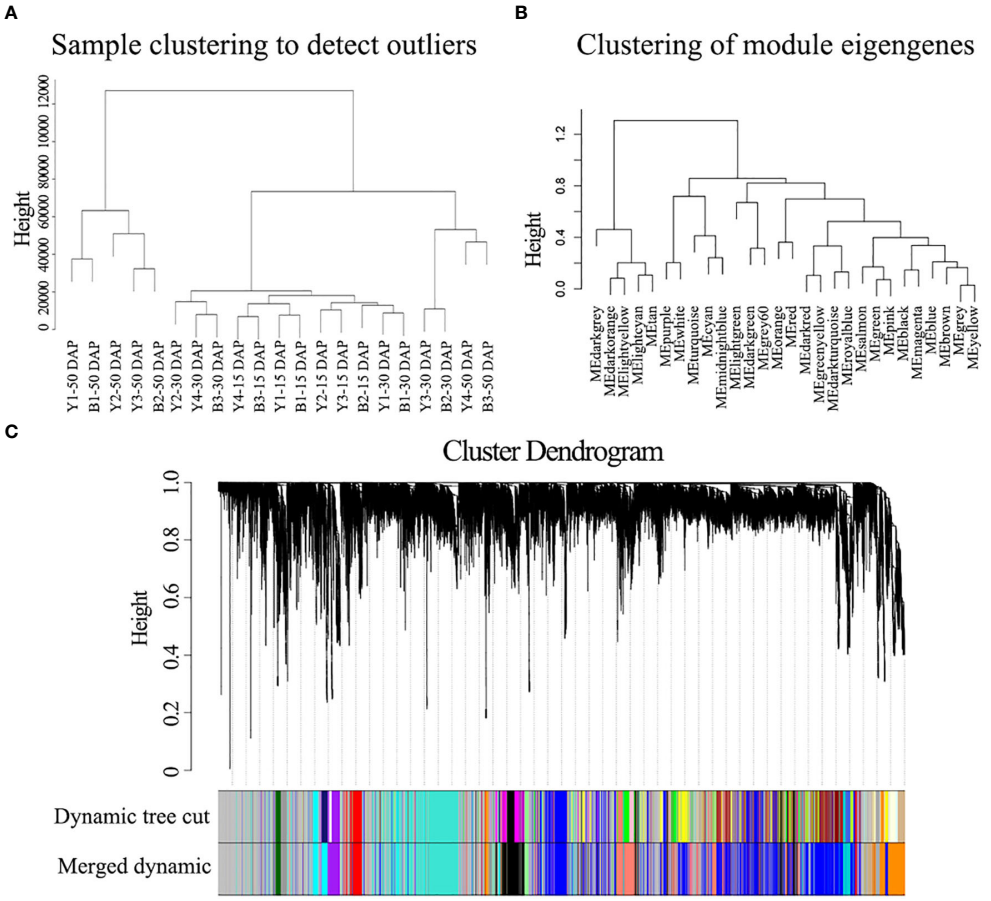


FIGURE 6
WGCNA of the gene expression matrix in seeds. **(A)** Clustering tree analysis of different samples. **(B)** Clustering tree analysis of different modules. **(C)** Clustering dendrogram of the average network adjacency for the identification of gene co-expression modules.

formation in rapeseed (Liu et al., 2005; Badani et al., 2006; Xiao et al., 2007; Yan et al., 2009; Zhang et al., 2011; Stein et al., 2013; Zhang et al., 2013; Lian et al., 2017; Xie et al., 2020; Chao et al., 2022), few major QTL and candidate genes involved in seed color have been successfully cloned in rapeseed. Recently, a major yellow-seed QTL on chromosome A09 was reported to control seed color, oil content and fiber content, which includes important candidate genes (e.g., *BnaA09.JAZ1*, *BnaA09.GH3.3*, and *BnaA09.LOX3*) (Chao et al., 2022).

Numerous studies have revealed that the phenolic compounds cyanidin and procyanidins play dominant roles in seed color formation in rapeseed (Auger et al., 2009; Zhai et al., 2019). The flavonoid biosynthesis pathway is a major factor determining seed color in rapeseed. In the current study, GO enrichment and PPI analyses showed that down-regulated DEGs in yellow-seeded rapeseed were mainly associated with the proanthocyanidin biosynthetic pathway and the flavonoid biosynthetic pathway. The hub genes *TT16*, *BAN*, *TT4*, *LDOX*, *CHIL*, *F3H*, *TT12*, *4CL*, and *C4H* genes were located in four modules (Table 2) and were expressed at higher levels in black- vs. yellow-seeded rapeseed, especially during the middle and late stages of seed development (Figure 5). *BnTT12* was previously identified as a candidate gene involved in seed color formation (Chai et al., 2009) and was detected within QTL intervals for seed color in rapeseed (Wang et al., 2017); *TT16* is involved in proanthocyanidin accumulation and is expressed in endothelial cells (Chen et al., 2013). However, compared to black-seeded rapeseed, photosynthesis-related genes were more active in yellow-seeded rapeseed, which were clustered into the GO terms xenobiotic transmembrane transport (GO:0006855), xenobiotic transport (GO:0042908), transport (GO:0006810), and amino acid metabolism (GO:1901605) (Table S7). It appears that yellow-seeded rapeseed has more advantageous traits compared to black-seeded varieties, such as higher oil and protein contents; this notion requires further study.

The MYB-bHLH-WD40 (MBW) complex plays crucial roles in determining seed color by regulating the flavonoid biosynthesis pathway in various plant species (Appelhagen et al., 2010; Xu et al., 2014; Xu et al., 2015), but little is known about this process in rapeseed. In this study, *BnTT1* and *BnTT8* were identified as hub genes (Table 2). Indeed, the recently generated rapeseed mutant *tt8* and *tt2* (created by CRISPR/Cas9-mediated gene editing) has yellow seeds (Zhai et al., 2019; Xie et al., 2020). In addition, downregulating *BnTT1 via* RNA interference (RNAi) reduced the accumulation of flavonoids in seeds (Lian et al., 2017). Moreover, the ternary MBW complex (TT2-TT8-TTG1) was previously shown to function in the flavonoid biosynthesis pathway by regulating *DFR*, *LDOX*/*TT18*,

10.3389/fpls.2023.1154208

TABLE 2 WGCNA of DEGs associated with the flavonoid biosynthesis pathway in rapeseed.

Gene ID	Annotation	WGCNA
BnaA03g39500D	Bn_TT16a	blue
BnaC06g08390D	Bn_TT1a	blue
BnaC06g26980D	Bn_4CL3a	orange
BnaA07g25210D	Bn_4CL3b	orange
BnaA03g60670D	Bn_BANb	orange
BnaC01g29820D	Bn_BANc	orange
BnaC04g18950D	Bn_BANd	orange
BnaA03g14010D	Bn_C4Ha	orange
BnaA04g17570D	Bn_C4Hb	orange
BnaA05g11950D	Bn_C4Hc	orange
BnaC03g16960D	Bn_C4Hd	orange
BnaC04g14330D	Bn_C4He	orange
BnaC09g50050D	Bn_CHILa	orange
BnaA10g25120D	Bn_CHILb	orange
BnaA09g31780D	Bn_F3Ha	orange
BnaC01g14310D	Bn_LDOXa	orange
BnaC07g37670D	Bn_LDOXb	orange
BnaA03g45610D	Bn_LDOXc	orange
BnaC06g17050D	Bn_TT12a	orange
BnaA03g04590D	Bn_TT4a	orange
BnaA10g19670D	Bn_TT4b	orange
BnaC02g05070D	Bn_TT4c	orange
BnaC03g06120D	Bn_TT4d	orange
BnaC09g43250D	Bn_TT4e	orange
BnaA09g22810D	Bn_TT8a	orange
BnaA01g36200D	Bn_BANa	red
BnaC09g24870D	Bn_TT8b	red
BnaC06g32180D	Bn_MYBL2	salmon

WGCNA, the module to which the gene belongs based on WGCNA.

BAN, and *TT12* expression, leading to changes in seed color (Lepiniec et al., 2006; Hichri et al., 2011; Xu et al., 2015). We also determined that *BAN*, *LDOX*, and *TT12* were more highly expressed in black- vs. yellow-seeded rapeseed. These genes shared similar expression patterns with *BnTT8* (Figure 7; Table 2), suggesting they might function together to regulate the flavonoid biosynthesis pathway.

With the development of high-throughput technologies such as genomic, transcriptomic, and proteomic profiling, numerous databases have been used to elucidate the complex networks involved in plant development and molecular responses to environmental cues. The release of *B. napus* datasets (Shen et al., 2015; Song et al., 2020) has made it possible to reveal the flavonoid biosynthetic pathway in rapeseed at the genome-wide level. iGRN analysis is helpful for uncovering the regulatory mechanisms of TFs in the flavonoid biosynthetic pathway in rapeseed using a network-based approach based on supervised learning for large-scale functional data integration (De Clercq et al., 2021). In the current study, 25 TF genes were identified that could be involved in regulating the genes associated with the flavonoid biosynthesis pathway based on iGRN (Figure 7; Table 3). Of these genes, *KNAT7*, *NAC2*, and *SEP2* are associated with the development of secondary cell walls in the seed coat (Grallert et al., 1997; Kunieda et al., 2008; Li et al., 2011), *KNAT7* is related to anthocyanin and proanthocyanidin biosynthesis (Bhargava et al., 2013), and *SEP2* is co-expressed with multiple MYB- and bHLH-related genes (Ahmad et al., 2022). Furthermore, *SHP1* and *CUC1* are involved in seed development (Kusumanjali et al., 2012; Mizzotti et al., 2012;

10.3389/fpls.2023.1154208

TABLE 3 The predicted 25 transcription factors by iGRN.

Gene ID	AGI No.	*P*-value	Target genes	Functional description
BnaA01g33070D	AT3G02310	4.06E-05	*TT1*, *TT8*, *TT16*, *BAN*	SEPALLATA 2 (SEP2)
BnaA02g33680D	AT5G63280	8.94E-07	*TT8*, *LDOX*	C2H2-like zinc finger protein
BnaA03g17120D	AT2G37260	6.35E-06	*LDOX*	TRANSPARENT TESTA GLABRA 2 (TTG2)
BnaC03g20650D	AT2G37260	6.35E-06	*LDOX*	TTG2
BnaC08g43050D	AT3G15510	5.32E-05	*TT8*, *BAN*, *LDOX*	NAC domain containing protein 2 (NAC2)
BnaCnng64100D	AT3G15510	5.32E-05	*TT8*, *BAN*, *LDOX*	NAC2
BnaA02g17180D	AT1G75390	6.07E-05	*TT1*, *TT12*, *TT8*, *TT16*, *BAN*, *LDOX*	basic leucine-zipper 44 (bZIP44)
BnaA04g01810D	AT3G58780	1.91E-06	*TT1*, *TT12*, *TT8*, *TT16*, *BAN*	SHATTERPROOF 1 (SHP1)
BnaA05g00070D	AT2G48100	1.65E-07	*TT8*, *C4H*, *LDOX*, *PAL2*	Exonuclease family protein
BnaC05g48320D	AT3G02310	4.06E-05	*TT1*, *TT8*, *TT16*, *BAN*	SEP2
BnaA09g52990D	AT1G62990	0.00004	*TT8*, *BAN*, *LDOX*	KNOTTED-like homeobox of *Arabidopsis thaliana* 7 (KNAT7)
BnaA03g24210D	AT4G09960	0.000044	*TT1*, *TT16*, *BAN*, *F3H*	SEEDSTICK (STK)
BnaA03g33770D	AT3G15170	1.09E-05	*TT12*, *TT8*, *LDOX*	CUP-SHAPED COTYLEDON1 (CUC1)
BnaA03g52920D	AT4G34590	2.58E-06	*TT1*, *TT12*, *TT8*, *TT16*, *BAN*, *LDOX*, *F3H*	G-box binding factor 6 (GBF6)
BnaA07g18050D	AT3G58780	1.91E-06	*TT1*, *TT12*, *TT8*, *TT16*, *BAN*	SHP1
BnaA07g21710D	AT1G75390	6.07E-05	*TT1*, *TT12*, *TT8*, *TT16*, *BAN*, *LDOX*	bZIP44
BnaA10g00620D	AT1G01250	4.76E-05	*TT1*, *TT12*, *TT8*, *BAN*, *C4H*, *LDOX*	Integrase-type DNA-binding superfamily protein
BnaAnng39120D	AT4G09960	0.000044	*TT1*, *TT16*, *BAN*, *F3H*	STK
BnaC02g42440D	AT5G63280	8.94E-07	*TT8*, *LDOX*	C2H2-like zinc finger protein
BnaC05g00680D	AT1G01250	4.76E-05	*TT1*, *TT12*, *TT8*, *BAN*, *C4H*, *LDOX*	Integrase-type DNA-binding superfamily protein
BnaC06g22430D	AT1G75390	6.07E-05	*TT1*, *TT12*, *TT8*, *TT16*, *BAN*, *LDOX*	bZIP44
BnaC08g35830D	AT2G21320	4.75E-05	*TT8*, *LDOX*	B-box zinc finger family protein
BnaC09g42060D	AT3G02310	4.06E-05	*TT1*, *TT8*, *TT16*, *BAN*	SEP2
BnaCnng08620D	AT1G01250	4.76E-05	*TT1*, *TT12*, *TT8*, *BAN*, *C4H*, *LDOX*	Integrase-type DNA-binding superfamily protein
BnaCnng46740D	AT4G09960	0.000044	*TT1*, *TT16*, *BAN*, *F3H*	STK

Cucinotta et al., 2018). In addition, *GBF6*, *STK*, and *TTG2* are candidate genes involved in seed coat color. The expression pattern of *GBF6* is similar to that of *PAP1*, which regulates the anthocyanin pathway in Arabidopsis (Gonzalez et al., 2008; Adamiec et al., 2011). *STK* is a key gene involved in proanthocyanidin production (Mizzotti et al., 2014), and TTG2 regulates genes in the PA biosynthetic branch of the flavonoid pathway (Gonzalez et al., 2016). Therefore, we demonstrated that iGRN can provide accurate and reliable information about protein structure and function.

In addition, some novel regulators could be involved in seed coat color, including *BnbZIP44* (*BnaA02g17180D*, *BnaC06g22430D*, and *BnaA07g21710D*), Exonuclease family protein (*BnaA05g00070D*), and *SHP1* (*BnaA04g01810D*), which were more highly expressed in black-vs. yellow-seeded *B. napus* (Figure 7). *bZIP44* is closely related to external environmental stimuli; its knockout lines showed slower germination during seed coat rupture (Weltmeier et al., 2009; Iglesias-Fernandez et al., 2013). Recently, *bZIP44* was considered as key candidate transcription factors associated with anthocyanin biosynthesis in *C. maximowiczii* fruits (Zhang et al., 2023) *SHP1* is a key gene that regulates inner seed coat development in Arabidopsis (Ehlers et al., 2016). Therefore, these identified TFs can be the targets for further molecular characterization.

In conclusion, the present study is the comprehensive report on transcriptome profiling of developing seeds in yellow- and black-seeded rapeseed. A total of 1206 and 276 DEGs involved in seed coat color could be identified in the middle and late stages, and the downregulated DEGs are mainly enriched for the phenylpropanoid and flavonoid biosynthesis pathways. Moreover, twenty-five TFs involved in seed color are predicted by iGRN and WGCNA that are associated with regulating flavonoid biosynthesis pathway. Our findings provide many clues for elucidating the regulatory

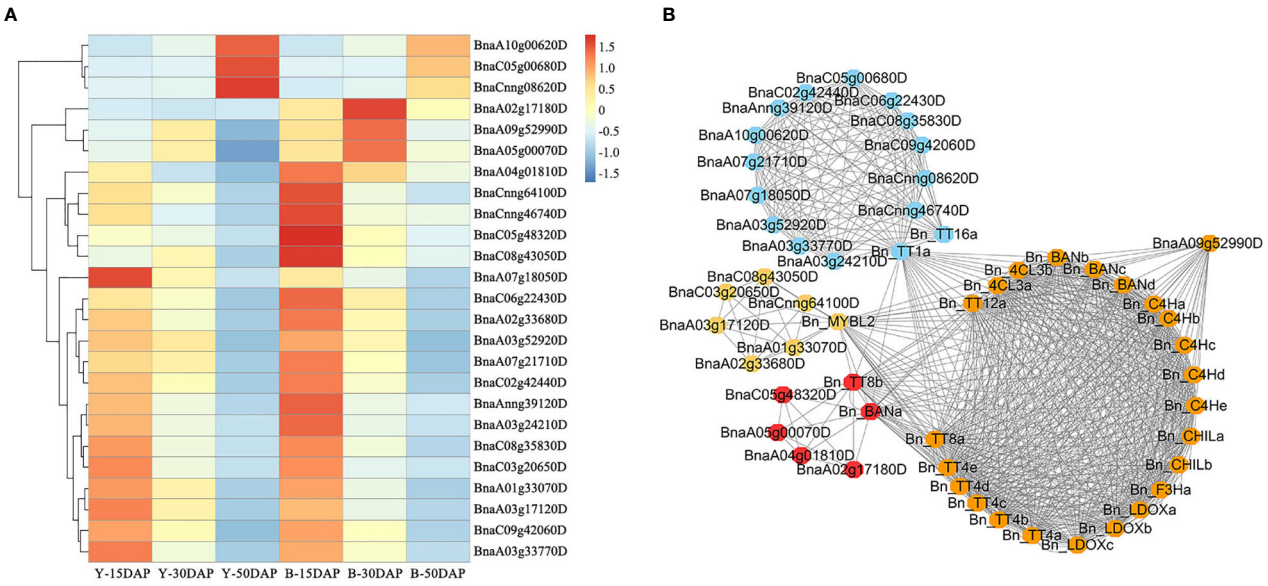


FIGURE 7
Expression profiles of 25 candidate TF genes. **(A)** Comparative expression analysis of 25 TF genes associated with seed color between yellow- and black-seeded rapeseed. The expression levels of the TF genes were calculated using FPKM values of materials with the same seed color. The heatmap was generated using the R package Pheatmap (scale = "row", cluster_row = T). **(B)** Co-expression network of 25 TFs and hub genes. The same color indicates that genes are located in the same module, as determined by WGCNA.

networks of the flavonoid biosynthesis pathway and for better understanding the molecular mechanism of the yellow-seeded trait in rapeseed.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI Bioproject PRJNA931458. (Pending release on acceptance).

Author contributions

JL and CQ conceived the project and designed the experiment plans; MG, XS, SC, YW, YT, TZ, LG, and FS conducted the experiments; MG, HZ, KL, and NY analyzed the data and wrote the article; MG, HZ, NY, JL and CQ reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the National Science Foundation of China (32272150, 31830067), Natural Science Foundation of Chongqing (CSTB2022NSCQ-LZX0034), China Agriculture Research System of MOF and MARA, Innovation and Entrepreneurship Training Program for graduates (CYS21124), the 111 Project (B12006).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1154208/full#supplementary-material

References

Abraham, V., and Bhatia, C. R. (1986). Development of strains with yellow-seed coat in Indian mustard (*Brassica juncea* czern & coss). *Plant Breed.* 97, 86–88. doi: 10.1111/j.1439-0523.1986.tb01307.x

Adamiec, M., Luciński, R., and Jackowski, G. (2011). The irradiance dependent transcriptional regulation of *AtCLPB3* expression. *Plant Sci.* 181, 449–456. doi: 10.1016/j.plantsci.2011.07.004

Ahmad, S., Chen, J. L., Chen, G. Z., Huang, J., Hao, Y., Shi, X. L., et al. (2022). Transcriptional proposition for uniquely developed protocorm flowering in three orchid species: Resources for innovative breeding. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.942591

Albert, S., Delseny, M., and Devic, M. (1997). *BANYULS*, a novel negative regulator of flavonoid biosynthesis in the *Arabidopsis* seed coat. *Plant J.* 11, 289–299. doi: 10.1046/j.1365-313X.1997.11020289.x

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1101/002824

Appelhagen, I., Huep, G., Lu, G. H., Strompen, G., Weisshaar, B., and Sagasser, M. (2010). Weird fingers: Functional analysis of *WIP* domain proteins. *FEBS Lett.* 584, 3116–3122. doi: 10.1016/j.febslet.2010.06.007

Appelhagen, I., Lu, G. H., Huep, G., Schmelzer, E., Weisshaar, B., and Sagasser., M. (2011). *TRANSPARENT TESTA 1* interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of *Arabidopsis thaliana* seeds. *Plant J.* 67, 406–419. doi: 10.1111/j.1365-313X.2011.04603.x

Auger, B., Baron, C., Lucas, M. O., Vautrin, S., Berges, H., Chalhoub, B., et al. (2009). Brassica orthologs from *BANYULS* belong to a small multigene family, which is involved in procyanidin accumulation in the seed. *Planta* 230, 1167–1183. doi: 10.1007/s00425-009-1017-0

Badani, A. G., Snowdon, R. J., Wittkop, B., Lipsa, F. D., Baetzel, R., Horn, R., et al. (2006). Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). *Genome* 49, 1499–1509. doi: 10.1139/G06-091

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2010). TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 39, 366–380. doi: 10.1111/j.1365-313X.2004.02138.x

Bhargava, A., Ahad, A., Wang, S. C., Mansfield, S. D., Haughn, G. W., Douglas, C. J., et al. (2013). The interacting *MYB75* and *KNAT7* transcription factors modulate secondary cell wall deposition both in stems and seed coat in *Arabidopsis*. *Planta* 237, 1199–1211. doi: 10.1007/s00425-012-1821-9

Bharti, A. K., and Khurana, J. P. (2003). Molecular characterization of *transparent testa* (*tt*) mutants of *Arabidopsis thaliana* (ecotype estland) impaired in flavonoid biosynthetic pathway. *Plant Sci.* 165, 1321–1332. doi: 10.1016/S0168-9452(03)00344-3

Carruthers, J. M., Cook, S. M., Wright, G. A., Osborne, J. L., Clark, S. J., Swain, J. L., et al. (2017). Oilseed rape (*Brassica napus*) as a resource for farmland insect pollinators: quantifying floral traits in conventional varieties and breeding systems. *Global Change Biol. Bioenergy* 9, 1370–1379. doi: 10.1111/gcbb.12438

Chai, Y. R., Lei, B., Huang, H. L., Li, J. N., Yin, J. M., Tang, Z. L., et al. (2009). *TRANSPARENT TESTA 12* genes from brassica napus and parental species: cloning, evolution, and differential involvement in yellow seed trait. *Mol. Genet. Genomics* 281, 109–123. doi: 10.1007/s00438-008-0399-1

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H. B., Wang, X. Y., et al. (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chao, H., Guo, L., Zhao, W., Li, H., and Li, M. (2022). A major yellow-seed QTL on chromosome A09 significantly increases the oil content and reduces the fiber content of seed in *Brassica napus*. *Theor. Appl. Genet.* 135, 1293–1305. doi: 10.1007/s00122-022-04031-0

Chen, G. Q., Deng, W., Peng, F., Truksa, M., Singer, S., Snyder, C. L., et al. (2013). *Brassica napus TT16* homologs with different genomic origins and expression levels encode proteins that regulate a broad range of endothelium-associated genes at the transcriptional level. *Plant J.* 74, 663–677. doi: 10.1111/tpj.12151

Chen, B. Y., and Heneen, W. K. (1992). Inheritance of seed colour in *Brassica campestris* l. and breeding for yellow-seeded *B. napus* l. *Euphytica* 59, 157–163. doi: 10.1007/BF00041268

Chezem, W. R., and Clay, N. K. (2015). "Regulatory and biosynthetic mechanisms underlying plant chemical defense responses to biotic stresses," in Laitinen R. A. E Ed. *Molecular mechanisms in plant adaptation*. (John Wiley & Sons, Inc). doi: 10.1002/9781118860526.ch5

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. T. (2014). cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8, S11. doi: 10.1186/1752-0509-8-S4-S11

Cucinotta, M., Manrique, S., Cuesta, C., Benkova, E., Novak, O., and Colombo, L. (2018). *CUP-SHAPED COTYLEDON1* (*CUC1*) and *CUC2* regulate cytokinin homeostasis to determine ovule number in *Arabidopsis*. *J. Exp. Bot.* 69, 5169–5176. doi: 10.1093/jxb/ery281

Cutanda-Perez, M. C., Ageorges, A., Gomez, C., Vialet, S., Terrier, N., Romieu, C., et al. (2009). Ectopic expression of *VlmybA1* in grapevine activates a narrow set of genes involved in anthocyanin synthesis and transport. *Plant Mol. Biol.* 69, 633–648. doi: 10.1007/s11103-008-9446-x

Davies, K. M., Albert, N. W., Zhou, Y., and Schwinn, K. E. (2018). Functions of flavonoid and betalain pigments in abiotic stress tolerance in plants. *Am. Cancer Soc.* 1, 21–61. doi: 10.1002/9781119312994.apr0604

De Clercq, I., Van de Velde, J., Luo, X., Liu, L., Storme, V., Van Bel, M., et al. (2021). Integrative inference of transcriptional networks in arabidopsis yields novel ROS signalling regulators. *Nat. Plants* 7, 1–14. doi: 10.1038/s41477-021-00894-1

Diao, C., Xi, Y., and Xiao, T. (2018). Identification and analysis of key genes inosteosarcoma using bioinformatics. *Oncol. Lett.* 15, 2789–2794. doi: 10.3892/ol.2017.7649

Dong, X. Y., Braun, E. L., and Grotewold, E. (2001). Functional conservation of plant secondary metabolic enzymes revealed by complementation of *Arabidopsis* flavonoid mutants with maize genes. *Plant Physiol.* 127, 46–57. doi: 10.1104/pp.127.1.46

Ehlers, K., Bhide, A. S., Tekleyohans, D. G., Wittkop, B., Snowdon, R. J., and Becker, A. (2016). The MADS box genes ABS, SHP1, and SHP2 are essential for the coordination of cell divisions in ovule and seed coat development and for endosperm formation in *Arabidopsis thaliana*. *PloS One* 11, e0165075. doi: 10.1371/journal.pone.0165075

Fu, F. Y., Liu, L. Z., Chai, Y. R., Chen, L., Yang, T., Jin, M. Y., et al. (2007). Localization of QTLs for seed color using recombinant inbred lines of *Brassica napus* in different environments. *Genome* 50, 840–854. doi: 10.1139/G07-068

Gesell, A., Yoshida, K., Tran, L. T., and Constabel, C. P. (2014). Characterization of an apple TT2-type R2R3 MYB transcription factor functionally similar to the poplar proanthocyanidin regulator *PtMYB134*. *Planta* 240, 497–511. doi: 10.1007/s00425-014-2098-y

Gonzalez, A., Brown, M., Hatlestad, G., Akhavan, N., Smith, T., Hembd, A., et al. (2016). *TTG2* controls the developmental regulation of seed coat tannins in *Arabidopsis* by regulating vacuolar transport steps in the proanthocyanidin pathway. *Dev. Biol.* 419, 54–63. doi: 10.1016/j.ydbio.2016.03.031

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in arabidopsis seedlings. *Plant J.* 53, 814–827. doi: 10.1111/j.1365-313X.2007.03373.x

Grallert, A., Miklos, I., and Sipiczki, M. (1997). Division-site selection, cell separation, and formation of anucleate minicells in schizosaccharomyces pombe mutants resistant to cell-wall lytic enzymes. *Protoplasma* 198, 218–229. doi: 10.1007/BF01287571

He, Q., Xue, Y. H., Wang, Y. X., Zhang, N. A., and Zhang, L. G. (2022). Metabolic profiling and transcriptomic data providing critical flavonoid biosynthesis mechanisms disclose color differences of purple heading Chinese cabbages (*Brassica rapa* l.). *Lebensmittel-Wissenschaft. Und-Technol.* 168, 113885. doi: 10.1016/j.lwt.2022.113885

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J. Exp. Bot.* 62, 2465–2483. doi: 10.1093/jxb/erq442

Hong, M. Y., Hu, K. N., Tian, T. T., Li, X., Chen, L., Zhang, Y., et al. (2017). Transcriptomic analysis of seed coats in yellow-seeded *Brassica napus* reveals novel genes that influence proanthocyanidin biosynthesis. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01674

Ichino, T., Fuji, K., Ueda, H., Takahashi, H., Koumoto, Y., Takagi, J., et al. (2015). *GFS9/TT9* contributes to intracellular membrane trafficking and flavonoid accumulation in *Arabidopsis thaliana*. *Plant J.* 80, 410–423. doi: 10.1111/tpj.12637

Iglesias-Fernandez, R., Barrero-Sicilia, C., Carrillo-Barral, N., Onate-Sanchez, L., and Carbonero, P. (2013). *Arabidopsis thaliana bZIP44*: A transcription factor affecting seed germination and expression of the mannanase-encoding gene *AtMAN7*. *Plant J.* 74, 767–780. doi: 10.1111/tpj.12162

Jiang, J., Zhu, S., Yuan, Y., Wang, Y., Zeng, L., Batley, J., et al. (2019). Transcriptomic comparison between developing seeds of yellow- and black-seeded *Brassica napus* reveals that genes influence seed quality. *BMC Plant Biol.* 19, 203. doi: 10.1186/s12870-019-1821-z

Kaur, H., Wang, L. H., Stawniak, N., Sloan, R., van Erp, H., and Eastmond, P. (2020). The impact of reducing fatty acid desaturation on the composition and thermal stability of rapeseed oil. *Plant Biotechnol. J.* 18, 983–991. doi: 10.1111/pbi.13263

Kerhoas, L., Aouak, D., Cingoz, A., Routaboul, J. M., Lepiniec, L., and Einhorn, J. (2006). Structural characterization of the major flavonoid glycosides from *Arabidopsis thaliana* seeds. *J. Agric. Food Chem.* 54, 6603–6612. doi: 10.1021/jf061043n

Kim, D., Landmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kunieda, T., Mitsuda, N., Ohme-Takagi, M., Takeda, S., Aida, M., Tasaka, M., et al. (2008). NAC family proteins *NARS1/NAC2* and *NARS2/NAM* in the outer integument regulate embryogenesis in *Arabidopsis*. *Plant Cell* 20, 2631–2642. doi: 10.1105/tpc.108.060160

Kusumanjali, K., Kumari, G., Srivastava, P. S., and Das, S. (2012). Sequence conservation and divergence in *miR164C1* and its target *CUC1* in *Brassica* species. *Plant Biotechnol. Rep.* 6, 149–163. doi: 10.1007/s11816-011-0208-x

Langfelder, P., and Horvath, S. (2008). WGCNA: An r package for weighted correlation network analysis. *BMC Bioinf.* 9, 559–571. doi: 10.1186/1471-2105-9-559

Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252

Li, X., Chen, L., Hong, M. Y., Zhang, Y., Zu, F., Wen, J., et al. (2012). A Large insertion in bHLH transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa*. *PloS One* 7, e44145. doi: 10.1371/journal.pone.0044145

Li, E., Wang, S., Liu, Y., Chen, J. G., and Douglas, C. (2011). *OVATE FAMILY PROTEIN 4* (*OFP4*) interaction with *KNAT7* regulates secondary cell wall formation in arabidopsis thaliana. *Plant J. Cell Mol. Biol.* 67, 328–341. doi: 10.1111/j.1365-313X.2011.04595.x

Lian, J. P., Lu, X. C., Yin, N. W., Ma, L. J., Lu, J., Liu, X., et al. (2017). Silencing of *BnTT1* family genes affects seed flavonoid biosynthesis and alters seed fatty acid composition in *Brassica napus*. *Plant Sci.* 254, 32–47. doi: 10.1016/j.plantsci.2016.10.012

Liu, S., Fan, C. C., Li, J. N., Cai, G. Q., Yang, Q. Y., Wu, J., et al. (2016). A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*. *Theor. Appl. Genet.* 129, 1203–1215. doi: 10.1007/s00122-016-2697-z

Liu, Z. W., Fu, T. D., Tu, J. X., and Chen, B. Y. (2005). Inheritance of seed colour and identification of RAPD and AFLP markers linked to the seed colour gene in rapeseed (*Brassica napus* l.). *Theor. Appl. Genet.* 110, 303–310. doi: 10.1007/s00122-004-1835-1

Liu, X. P., Tu, J. X., Chen, B. Y., and Fu, T. D. (2010). Identification and inheritance of a partially dominant gene for yellow seed colour in *Brassica napus*. *Plant Breed.* 124, 9–12. doi: 10.1111/j.1439-0523.2004.01051.x

Locascio, A., Roig-Villanova, I., Bernardi, J., and Varotto, S. (2014). Current perspectives on the hormonal control of seed development in *Arabidopsis* and maize: A focus on auxin. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00412

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/PREACCEPT-8897612761307401

Lu, C. F., Napier, J. A., Clemente, T. E., and Cahoon, E. B. (2011). New frontiers in oilseed biotechnology: Meeting the global demand for vegetable oils for food, feed, biofuel, and industrial applications. *Curr. Opin. Biotechnol.* 22, 252–259. doi: 10.1016/j.copbio.2010.11.006

Lu, C. Q., Xie, Z. J., and Feng, Y. (2020). Mitochondrial ribosomal protein s9m is involved in male gametogenesis and seed development in *Arabidopsis*. *Plant Biol.* 22, 655–667. doi: 10.1111/plb.13108

Mizzotti, C., Ezquer, I., Paolo, D., Rueda-Romero, P., Guerra, R. F., Battaglia, R., et al. (2014). *SEEDSTICK* is a master regulator of development and metabolism in the *Arabidopsis* seed coat. *PloS Genet.* 10, e1004856. doi: 10.1371/journal.pgen.1004856

Mizzotti, C., Mendes, M. A., Caporali, E., Schnittger, A., Kater, M. M., Battaglia, R., et al. (2012). The MADS box genes *SEEDSTICK* and *ARABIDOPSIS* bsister play a maternal role in fertilization and seed development. *Plant J.* 70, 409–420. doi: 10.1111/j.1365-313X.2011.04878.x

Nishihara, M., and Nakatsuka, T. (2011). Genetic engineering of flavonoid pigments to modify flower color in floricultural plants. *Biotechnol. Lett.* 33, 433–441. doi: 10.1007/s10529-010-0461-z

Pang, Y. Z., Cheng, X. F., Huhman, D. V., Ma, J. Y., Peel, G. J., Yonekura-Sakakibara, K., et al. (2013). Medicago glucosyltransferase *UGT72L1*: Potential roles in proanthocyanidin biosynthesis. *Planta* 238, 139–154. doi: 10.1007/s00425-013-1879-z

Qu, C. M., Fu, F. Y., Lu, K., Zhang, K., Wang, R., Xu, X. F., et al. (2013). Differential accumulation of phenolic compounds and expression of related genes in black- and yellow-seeded *Brassica napus*. *J. Exp. Bot.* 64, 2885–2898. doi: 10.1093/jxb/ert148

Qu, C. M., Hasan, M., Lu, K., Liu, L. Z., Zhang, K., Fu, F. Y., et al. (2015). Identification of QTL for seed coat colour and oil content in *Brassica napus* by association mapping using SSR markers. *Can. J. Plant Sci.* 95, 387–395. doi: 10.4141/CJPS2013-411

Qu, C. M., Zhao, H. Y., Fu, F. Y., Wang, Z., Zhang, K., Zhou, Y., et al. (2016). Genome-wide survey of flavonoid biosynthesis genes and gene expression analysis between black- and yellow-seeded *Brassica napus*. *Front. Plant Sci.* 7, 1775. doi: 10.3389/fpls.2016.01755

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L., and Pachter, L. (2011). Improving RNA-seq expression estimates by correcting for fragment bias. *Genome Biol.* 12, R22. doi: 10.1186/gb-2011-12-3-r22

Ruuska, S., Girke, A. T., and Ohlrogge, J. B. (2002). Contrapuntal networks of gene expression during arabidopsis seed filling. *Plant Cell* 14, 1191–1206. doi: 10.1105/tpc.000877

Saeidnia, S., and Gohari, A. R. (2012). Importance of *Brassica napus* as a medicinal food plant. *J. Med. Plants Res.* 6, 2700–2703. doi: 10.5897/JMPR11.1103

Seok, H. Y., Bae, H., Kim, T., Mehdi, S. M. M., Nguyen, L. V., Lee, S. Y., et al. (2021). Non-TZF protein *AtC3H59/ZFWD3* is involved in seed germination, seedling development, and seed development, interacting with PPPDE family protein Desi1 in *Arabidopsis*. *Int. J. Mol. Sci.* 22, 4738. doi: 10.3390/ijms22094738

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecularinteraction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shen, E. H., Zou, J., Behrens, F. H., Chen, L., Ye, C. Y., Dai, S. T., et al. (2015). Identification, evolution, and expression partitioning of miRNAs in allopolyploid *Brassica napus*. *J. Exp. Bot.* 66, 7241–7253. doi: 10.1093/jxb/erv420

Simbaya, J., Slominski, B., Rakow, G., Campbell, L. D., Downey, R. K., and Bell, J. M. (1995). Quality characteristics of yellow-seeded brassica seed meals: Protein, carbohydrate, and dietary fiber components. *J. Agric. Food Chem.* 43, 2062–2066. doi: 10.1021/jf00056a020

Song, J. M., Liu, D. X., Xie, W. Z., Yang, Z. Q., Guo, L., Liu, K., et al. (2020). BnPIR: *Brassica napus* pan-genome information resource for 1,689 accessions. *Plant Biotechnol. J.* 19, 412–414. doi: 10.1111/pbi.13491

Stein, A., Wittkop, B., Liu, L. Z., Obermeier, C., Friedt, W., and Snowdon, R. J. (2013). Dissection of a major QTL for seed colour and fibre content in *Brassica napus* reveals colocalization with candidate genes for phenylpropanoid biosynthesis and flavonoid deposition. *Plant Breed.* 132, 382–389. doi: 10.1111/pbr.12073

Stracke, R., Ishihara, H., Barsch, G. H. A., Mehrtens, F., Niehaus, K., and Weisshaar, B. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 50, 660–677. doi: 10.1111/j.1365-313X.2007.03078.x

Tang, Z. L., Li, J. N., Zhang, X. K., Chen, L., and Wang, R. (2010). Genetic variation of yellow-seeded rapeseed lines (*Brassica napus* l.) from different genetic sources. *Plant Breed.* 116, 471–474. doi: 10.1111/j.1439-0523.1997.tb01033.x

Wang, H. Z. (2004). Strategy for rapeseed genetic improvement in China in the coming fifteen years. *Chin. J. Oil Crop Sci.* 26, 98–101. doi: 10.3321/j.issn:1007-9084.2004.02.025

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Wang, J., Xian, X. H., Xu, X. F., Qu, C. M., Lu, K., Li, J. N., et al. (2017). Genome-wide association mapping of seed coat color in *Brassica napus*. *J. Agric. Food Chem.* 65, 5229–5237. doi: 10.1021/acs.jafc.7b01226

Weltmeier, F., Rahmani, F., Ehlert, A., Dietrich, K., Schutze, K., Wang, X., et al. (2009). Expression patterns within the *Arabidopsis C/S1 bZIP* transcription factor network: Availability of heterodimerization partners controls gene expression during stress response and development. *Plant Mol. Biol.* 69, 107–119. doi: 10.1007/s11103-008-9410-9

White, J. A., Todd, T., Newman, T., Focks, N., Girke, T., de Ilarduya, O. M., et al. (2001). A new set of arabidopsis expressed sequence tags from developing seeds. the metabolic pathway from carbohydrates to seed oil. *Plant Physiol.* 124, 1582–1594. doi: 10.1104/pp.124.4.1582

Xiao, S. S., Xu, J. S., Li, Y., Zhang, L., Shi, S. J., Shi, S. W., et al. (2007). Generation and mapping of SCAR and CAPS markers linked to the seed coat color gene in *Brassica napus* using a genome-walking technique. *Genome* 50, 611–618. doi: 10.1139/g07-044

Xie, T., Chen, X., Guo, T. L., Rong, H., Chen, Z. Y., Sun, Q. F., et al. (2020). Targeted knockout of *BnTT2* homologues for yellow-seeded *Brassica napus* with reduced flavonoids and improved fatty acid composition. *J. Agric. Food Chem.* 68, 5676–5690. doi: 10.1021/acs.jafc.0c01126

Xu, W. J., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 20, 176–185. doi: 10.1016/j.tplants.2014.12.001

Xu, W. J., Grain, D., Bobet, S., Le Gourrierec, J., Thevenin, J., Kelemen, Z., et al. (2014). Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol.* 202, 132–144. doi: 10.1111/nph.12620

Yan, X. Y., Li, J. N., Fu, F. Y., Jin, M. Y., Chen, L., and Liu, L. Z. (2009). Co-Location of seed oil content, seed hull content and seed coat color QTL in three different environments in *Brassica napus* l. *Euphytica* 170, 355–364. doi: 10.1007/s10681-009-0006-5

Yoshida, K., Iwasaka, R., Kaneko, T., Sato, S., Tabata, S., and Sakuta, M. (2008). Functional differentiation of lotus japonicus *TT2s*, R2R3-MYB transcription factors comprising a multigene family. *Plant Cell Physiol.* 49, 157–169. doi: 10.1093/pcp/pcn009

Young, M. D., Wakefield., M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* 11, R14. doi: 10.1186/gb-2010-11-2-r14

Zhai, Y. G., Yu, K. D., Cai, S. L., Hu, L. M., Amoo, O., Xu, L., et al. (2019). Targeted mutagenesis of *BnTT8* homologs controls yellow seed coat development for effective oil production in *Brassica napus* l. *Plant Biotechnol. J.* 18, 1153–1168. doi: 10.1111/pbi.13281

Zhang, X. B., Abrahan, C., Colquhoun, T. A., and Liu, C. J. (2017). A proteolytic regulator controlling chalcone synthase stability and flavonoid biosynthesis in *Arabidopsis*. *Plant Cell* 29, 1157–1174. doi: 10.1105/tpc.16.00855

Zhang, M., Fan, J. L., Taylor, D. C., and Ohlrogge, J. B. (2009). *DGAT1* and *PDAT1* acyltransferases have overlapping functions in *Arabidopsis* triacylglycerol biosynthesis and are essential for normal pollen and seed development. *Plant Cell* 21, 3885–3901. doi: 10.1105/tpc.109.071795

Zhang, Y., Li, X., Chen, W., Yi, B., Wen, J., Shen, J. X., et al. (2011). Identification of two major QTL for yellow seed color in two crosses of resynthesized *Brassica napus* line no. 212717. *Mol. Breed.* 28, 335–342. doi: 10.1007/s11032-010-9486-1

Zhang, K., Lu, K., Qu, C. M., Liang, Y., Wang, R., Chai, Y. R., et al. (2013). Gene silencing of *BnTT10* family genes causes retarded pigmentation and lignin reduction in the seed coat of *Brassica napus*. *PloS One* 8, e61247. doi: 10.1371/journal.pone.0061247

Zhang, J. F., Lu, Y., Yuan, Y. X., Zhang, X. W., Geng, J. F., Chen, Y., et al. (2009). Map-based cloning and characterization of a gene controlling hairiness and seed coat color traits in *Brassica rapa*. *Plant Mol. Biol.* 69, 553–563. doi: 10.1007/s11103-008-9437-y

Zhang, X., Wang, J., Li, P., Sun, C., and Dong, W. (2023). Integrative metabolome and transcriptome analyses reveals the black fruit coloring mechanism of *Crataegus maximowiczii* c. k. schneid. *Plant Physiol. Biochem.* 194, 111–121. doi: 10.1016/j.plaphy.2022.11.008

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 March 2023
DOI 10.3389/fpls.2023.1118339

OPEN ACCESS

EDITED BY
Liezhao Liu,
Southwest University, China

REVIEWED BY
Maoteng Li,
Huazhong University of Science and Technology, China
Zhihua Hua,
Ohio University, United States

*CORRESPONDENCE
Chaobo Tong
tongchaobo@126.com
Xiaoli Yu
yxll268@126.com

SPECIALTY SECTION
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

RECEIVED 07 December 2022
ACCEPTED 07 March 2023
PUBLISHED 20 March 2023

CITATION
Yao S, Xie M, Hu M, Cui XB, Wu H, Li X, Hu P, Tong C and Yu X (2023) Genome-wide characterization of ubiquitin-conjugating enzyme gene family explores its genetic effects on the oil content and yield of *Brassica napus*. *Front. Plant Sci.* 14:1118339. doi: 10.3389/fpls.2023.1118339

Genome-wide characterization of ubiquitin-conjugating enzyme gene family explores its genetic effects on the oil content and yield of *Brassica napus*

Shengli Yao[1], Meili Xie[2], Ming Hu[2], XiaoBo Cui[2], Haoming Wu[1], Xiaohua Li[1], Peng Hu[1], Chaobo Tong[2]* and Xiaoli Yu[1]*

[1]School of Life Science and Technology, Wuhan Polytechnic University, Wuhan, Hubei, China, [2]The Key Laboratory of Biology and Genetic Improvement of Oil Crops, the Ministry of Agriculture and Rural Affairs of the PRC, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China

Ubiquitin-conjugating enzyme (UBC) is a critical part of the ubiquitin–proteasome pathway and plays crucial roles in growth, development and abiotic stress response in plants. Although *UBC* genes have been detected in several plant species, characterization of this gene family at the whole-genome level has not been conducted in *Brassica napus*. In the present study, 200 putative *BnUBCs* were identified in *B. napus*, which were clustered into 18 subgroups based on phylogenetic analysis. *BnUBC*s within each subgroup showed relatively conserved gene architectures and motifs. Moreover, the gene expression patterns in various tissues as well as the identification of *cis*-acting regulatory elements in *BnUBC* promoters suggested further investigation of their potential functions in plant growth and development. Furthermore, three *BnUBC*s were predicted as candidate genes for regulating agronomic traits related to oil content and yield through association mapping. In conclusion, this study provided a wealth of information on the *UBC* family in *B. napus* and revealed their effects on oil content and yield, which will aid future functional research and genetic breeding of *B. napus*.

KEYWORDS

***Brassica napus*, the *UBC* family, evolution, association mapping analysis, oil content and yield**

Introduction

Ubiquitination is a crucial regulatory process for the selective protein degradation mechanism which regulates cell physiology through the ubiquitin-26S proteasome pathway in plants (Pickart, 2001; Smalle and Vierstra, 2004; Vierstra, 2009; Ye and Rape, 2009). Ubiquitination regulates a wide range of various plant growth and developmental processes

including photomorphogenesis, protein translocation within cells, flower development, cell cycle control, both abiotic and biotic stress responses, phytohormone, regulation of proteome homeostasis as well as light signaling (Welchman et al., 2005; Dreher and Callis, 2007; Vierstra, 2009; Sadanandom et al., 2012; Doroodian and Hua, 2021; Yu and Hua, 2022). In addition, the ubiquitin-26S proteasome system (UPS) plays critical role in plant adaption (Hua and Yu, 2019). Protein ubiquitination involves the covalent attachment of a 76-amino acid (aa) sequence, through one of the encompassing seven lysine residues (Lys-6, Lys-11, Lys-27, Lys-29, Lys-33, Lys-48, Lys-63), to substrate protein (Welchman et al., 2005; Kim et al., 2013). Moreover, the fate of the ubiquitinated substrate is determined by the type of ubiquitination as well as the choice of the Lys residue for the modification, resulting in different linkages with various functions (Fang and Weissman, 2004; Sun and Chen, 2004).

The protein ubiquitin-proteasome system (UPS) is a multistep reaction mediated by three enzymes, E1 (ubiquitin-activating enzyme, UBA), E2 (ubiquitin-conjugating enzyme, UBC) and E3 (Ubiquitin-ligase enzyme) (Glickman and Ciechanover, 2002). Ubiquitin is initially activated by E1 through the ATP-dependent reaction, which results in the formation of a thioester-linked ubiquitin (Ramadan et al., 2015). Then, E1 transfers the thioester-linked ubiquitin to the cysteine (Cys) residue (active site) of the UBC domain through the passing of E2 *via* transesterification (Hershko et al., 1983). Subsequently, E2 transfers the ubiquitin moiety to the substrate protein with the help of E3 directly *via* a second *ans*-thiolation reaction, which mediates the formation of polyubiquitin chains on target proteins and determines the specificity of the substrate in the ubiquitination system (Bae and Kim, 2013; Bae and Kim, 2014). Finally, the 26S proteasome degrades the target protein. The family of UBCs (E2s) is central to this enzymatic cascade, which offers a platform for the attachment of ubiquitin to the target proteins (Hershko et al., 1983; Burroughs et al., 2008).

Most ubiquitin-like conjugating enzymes (UBLs) and E2s comprise a conserved catalytic core consisting of approximately 140-150 aa known as the UBC domain harboring the active site cysteine residue required for enzyme-ubiquitin (Criqui et al., 2002; Michelle et al., 2009; Schumacher et al., 2013). Moreover, several studies indicated that the UBC domain contributes to the mediation of the interaction between E2 and E3 (Huang et al., 1999; Christensen et al., 2007; Poyurovsky et al., 2007). The UBC domains in various members of the E2 family is highly conserved, both in terms of amino acid sequence and three-dimensional structure, which possessed four α-helices, an anti-parallel β-sheet formed by four strands, and a short 310–helix (Lin et al., 2002; Ozkan et al., 2005; Wenzel et al., 2011). Among the multiple binding sites, one highly conserved active-site (Cys residue) is located at a shallow groove formed by a short loop connecting α-helix 2 with α-helix 3 as well as a proximal long loop (Ye and Rape, 2009; van Wijk and Timmers, 2010). The formation of E2-ubiquitin thioester requires E2s to interact with ubiquitin as well as with E1 (or its cognates) which results in considerable evolutionary constraints on E2 structure and the formation of the conserved active site (Silver et al., 1992). Aside from the core E2 domain, some detected E2 enzymes comprise diverse N- and C-terminal extensions that are thought to contribute to the intracellular localization of the enzyme and its substrate specificity (Arrigoni et al., 2012; Hodson et al., 2014). Furthermore, UBC enzymes are divided into four classes, based on the N- and C-terminal extensions as well as the UBC domain. Class I UBCs contain only the UBC domain with a region of approximately 150 conserved residues, whereas Class II UBCs comprise the N-terminal extension and the UBC domain; Class III E2s harbor the C-terminal extension as well as the UBC domain; and Class IV UBCs possess both the N- and C-terminal extensions and the UBC domain (Papaleo et al., 2012; Schumacher et al., 2013).

The genes encoding *UBCs* usually form a multigene family and the number of *UBC*s in the multigene family is greater in the higher eukaryotes than in the lower eukaryotes, due to the expansion during the process of evolution (Jue et al., 2015). Moreover, UBLs, including the SUMO-conjugating enzymes, ubiquitin (RUB) conjugating enzymes and ubiquitin-conjugating enzyme variants (UEVs) also belong to the E2 category. Among the 48 UBC domain-containing proteins identified in *Arabidopsis* (Kraft et al., 2005), except 37 potential E2s, three are thioester-linked UBLs, one is a SUMO-conjugating enzyme [AtSCE1, At3g57870), and two are RUB-conjugating related enzymes (At4g36800 (RCE1) and At2g18600 (RCE2)]. These UBL-specific enzymes perform the same function as E2s, but they do not belong to the *UBC* family. In addition, eight UBC proteins are termed as ubiquitin-conjugating enzyme variants (UEVs), since they lack the active-site Cys residue, which is not active on their own (Callis, 2014). Based on sequence homology, the 48 *Arabidopsis UBCs* were classified into 16 groups (Kraft et al., 2005). Plant UBC proteins, as previous studies reported, play a crucial role in regulating plant development, growth and stress response (Dreher and Callis, 2007; Sadanandom et al., 2012; E et al., 2015; Millyard et al., 2016; Jue et al., 2018; Kravic et al., 2020; Liu et al., 2020). For example, functional analyses have revealed that *AtUBC1* and *AtUBC2* are involved in leaf development, tolerance response to UV stress, as well as activation of the floral repressor gene (Xu et al., 2009). *AtUBC13* is involved in the plant response to DNA damage, iron deficiency and epidermal cell differentiation (Welchman et al., 2005; Wen et al., 2006; Wen et al., 2008; Li and Schmidt, 2010; Sadanandom et al., 2012). *AtUBC19* and *AtUBC20*, which belong to the *E2-C* gene family, potentially contribute to protein ubiquitination reactions and play key functional roles in the cell cycle in differentiating or differentiated cells (Criqui et al., 2002). However, *AtUBC21* (*AtPEX4*) specializes in protein for ubiquitination in peroxisome maintenance (Zolman et al., 2005). *AtUBC22* has a function in female gametophyte development and potentially in Lys11-linked ubiquitination (Wang et al., 2016). *AtUBC24* regulates the uptake, allocation as well as remobilization of inorganic phosphate (Dong et al., 1998; Lin et al., 2009), and is also the target gene of miR399s (Aung et al., 2006). *AtUBC32* is a part of the endoplasmic reticulum-associated protein degradation (ERAD) complex, which is involved in salt stress tolerance mediated by brassinosteroid (BR) (Cui et al., 2012) and in plant growth promotion in *Arabidopsis*. Moreover, the viral pathogen *Zymoseptoria tritici* induced silence of *Triticum aestivum*

ubiquitin-conjugating enzyme 4 (*TaU4*) gene, resulting in postponed progression of disease symptoms, and limited reproduction as well as the growth of Septoria in wheat leaves (Millyard et al., 2016). In *Vitis vinifera*, the *UBC* family is involved in the berry ripening process and cold and heat stress responses (Gao et al., 2017). In conclusion, *UBC* genes in plants play various important roles related to stress response and plant growth.

Allotetraploid *Brassica napus* ($A_nA_nC_nC_n$, 2n = 38) is a globally important oilseed crop, which is derived from the hybridization between *Brassica rapa* (A_rA_r, 2n=20) and *Brassica oleracea* (C_oC_o, 2n=18) followed by chromosome doubling approximately 7500 years ago (Rana et al., 2004; Lu et al., 2011; Chalhoub et al., 2014; Sun et al., 2017; Lu et al., 2019; Wu et al., 2019). During evolution, *B. napus* and other *Brassica* species underwent multiple rounds of whole-genome duplication events (Chalhoub et al., 2014; Liu et al., 2014), and thus serve as an ideal polyploid model for studying polyploid genome evolution including gene family expansion and gene sequence and function divergence (Liu et al., 2014; Cheng et al., 2018). High-quality whole genome sequences of *B. rapa* (Zhang et al., 2018), *B. oleracea* (Belser et al., 2018) and *B. napus* (Chalhoub et al., 2014) have been published, which provide an opportunity to systematically investigate a specific gene family in *B. napus*.

So far, several studies have been performed to detect and analyze the UBC proteins on a genome-wide level in different plant species including 53 *UBC*s in sorghum (Jia et al., 2019), 57 in potato (Liu et al., 2019), 40 in longan (Jue et al., 2018), 39 in rice (E et al., 2015), 75 in maize (Jue et al., 2015), 34 in papaya (Jue et al., 2017), 59 in tomato (Sharma and Bhatt, 2017) and 74 in banana (Dong et al., 2016). However, systematic investigation of the *UBC*s in *B. napus* has been lacking. In this research, 200 *BnUBC*s were detected in *B. napus*, and their phylogenetic relationship, gene structure, conserved motif, *cis*-acting regulatory element, duplication pattern and target genes of miRNA were also systematically analyzed. The evolutionary history of the *UBC* family was explored through synteny analysis between *B. napus*, *B. rapa*, *B. oleracea*, and *Arabidopsis*. Additionally, the RNA-seq data of various tissues of *B. napus* were collected from publicly available databases for exploring the expression patterns of *BnUBC* genes. Furthermore, the genetic variations of *UBC* genes in germplasm from a global core collection of *B. napus* (Tang, 2019) were also investigated. The analysis of association mapping uncovered that some *UBC* genes were significantly associated with agronomic traits related to oil content and yield in *B. napus*. The results of this study will help us better understand the *BnUBC*s and lay the groundwork for future research on gene function and genetic breeding.

Materials and methods

Genome-wide identification and characterization of *BnUBC*s

Genome sequence and annotation data of *B. napus* cv. 'Darmor-*bzh*' (v_5.0) were obtained from the Genoscope database (http://www.genoscope.cns.fr/brassicanapus/) (Chalhoub et al., 2014). *B. rapa* 'Chiifu' (v3.0) datasets were obtained from the database (https://bigd.big.ac.cn/gwh/Assembly/134/show) (Zhang et al., 2018) and *B. oleracea* 'HDEM' (broccoli) datasets were acquired from Bolbase database (http://www.ocri-genomics.org/bolbase/index.html) (Belser et al., 2018). The ubiquitin-conjugating enzyme (UQ_con) domain (PF00179) annotation file obtained from the Pfam database (https://www.ebi.ac.uk/interpro/) was applied as queries to perform HMM search against protein sequences of annotated genes using HMMER version 3.1 (http://hmmer.org/) (Eddy, 2011) with E-value cutoff of 1e-5 (Finn et al., 2011). The amino acid sequences of *UBC*s (predicted above) were utilized as queries to carry out BLASTP searches against the full-length protein sequences of 48 *AtUBC*s from The Arabidopsis Information Resource (TAIR) database (http://www.arabidopsis.org/) and 39 *OsUBC*s from Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/downloads_gad.shtml) with E-value < 1e-5. The putative *BnUBCs* with hits of both AtUBCs and OsUBCs were deployed for further confirming the existence of the UQ_con domain using the Pfam (https://www.ebi.ac.uk/interpro/), SMART (http://smart.embl-heidelberg.de/) and CDD (https://www.ncbi.nlm.nih.gov/Structure/cdd) databases. The molecular weight (MW), isoelectric point (PI), instability index, aliphatic index and grand average of hydropathicity (GRAVY) of UBC proteins, were calculated using the ProtParam tool ExPASy (https://web.expasy.org/protparam/). The subcellular location of *BnUBCs* was predicted using Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/#) (Chou and Shen, 2010).

Analyses of the chromosomal localization, gene structure, conserved motif and *cis*-acting regulatory element for *BnUBC*s

The chromosomal location, coding sequence (CDS) and amino acid sequences were determined based on the genome annotation of *B. napus* in the Genoscope database (http://www.genoscope.cns.fr/brassicanapus/). The chromosomal distribution of *BnUBC*s was graphically represented using the RIdeogram package of the R software (https://github.com/TickingClock1992/RIdeogram) (Hao et al., 2020). Multiple alignments of the BnUBC protein sequences were performed using CLUSTAL v2.1, with default parameters. A schematic diagram of *BnUBC*s gene structure was conducted by Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2015). Conserved motifs in the BnUBC proteins were investigated using the online MEME server (http://meme-suite.org/tools/meme) with the following parameters settings: the maximum number of motifs, 10; minimum width of motifs, 6; maximum width of motifs, 100 aa; and E-value < 1e-10 (Bailey et al., 2009). Moreover, the 2-kb upstream sequence of the *BnUBCs* gene sequences was extracted and submitted to PlantCARE (Lescot et al., 2002) for the detection of the *cis*-elements.

Phylogenetic and synteny analysis of *BnUBC* proteins

To better understand the evolutionary relationships among the UBC of *B. rapa*, *B. oleracea*, *B. napus* and *A. thaliana*, a

phylogenetic analysis was performed. A phylogenetic tree was constructed applying MEGA 5.2 software (Tamura et al., 2011), based on the neighbor-joining (NJ) method and 1,000 bootstrap replications. The online software Interactive Tree Of Life (iTOL, http://itol.embl.de/) (Letunic and Bork, 2016) was used to decorate this phylogenetic tree. Genes orthologous between *B. napus* and its ancestors (*B. rapa*, *B. oleracea*, and *A. thaliana*) were identified by BLASTn searches of their CDSs based on two criteria: coverage of sequence length > 80%, and identity of aligned regions > 80% (Kong et al., 2013). In addition, DupGen_finder (https://github.com/qiao-xin/DupGen_finder) (Qiao et al., 2019) was employed to identify the modes of gene duplication of *UBC* paralogous genes in *A. thaliana, B. napus, B. rapa* and *B. oleracea*. The syntenic relationship among these paralogs was presented using the fmsb package of the R software.

Ka/Ks calculation

The KaKs calculator (Wang et al., 2010) was applied to estimate the divergence between pairwise nonsynonymous substitution rates (Ka) and synonymous substitution rates (Ks). The evolutionary constraint (Ka/Ks) of *UBC* orthologous genes between *B. napus* and the other three species (*A. thaliana*, *B. rapa* and *B. oleracea*) was calculated according to their CDSs. In addition, to minimize errors, only the gene pairs with Ks < 1 were remained for further analysis (Guo et al., 2017; Wang T, et al., 2020).

Prediction of miRNAs targeting *BnUBCs*

The psRNATarget Server (Dai et al., 2018) was used to predict the miRNAs potentially targeting *BnUBCs*. The gene sequences of *BnUBCs* were submitted as candidates to search against the available sequences of reference *B. napus* miRNA with default arguments. A network of interactions between the predicted miRNAs and their target *UBC* genes in *B. napus* was visualized using the Cytoscape software (Shannon et al., 2003).

Expression profiles and gene ontology enrichment analysis of *BnUBCs*

The publicly available RNA-seq datasets of five different *B. napus* tissues (callus, root, leaf, bud and silique) were collected from the NCBI SRA database (accession no. SRP136038) (Yao et al., 2020). The expression levels of *BnUBCs* were quantified based on their fragments per kilobase of exon per million reads mapped (FPKM) values using Cufflinks with default parameters (Trapnell et al., 2012). Furthermore, the expression levels of *BnUBC* genes were used to construct clustered heatmaps with TBtools (Chen et al., 2020). Then, GO enrichment analysis of *BnUBC* genes was performed using the clusterProfiler package of the R software.

Association mapping of *UBC* genes in a natural population of *B. napus*

The agronomic traits of 324 worldwide accessions comprising a natural population of *B. napus* were applied to detect the natural sequence variations of *BnUBCs* (Tang, 2019; Xie et al., 2022). Single nucleotide polymorphisms (SNPs) in *BnUBCs* were annotated using the SnpEff software (Cingolani et al., 2012). Moreover, agricultural traits involving the primary flowering time (PFT), full flowering time (FFT1), final flowering time (FFT2), plant height (PH), thousand seed weight (TSW), main inflorescence silique density (MISD), main inflorescence silique number (MISN), oil content (OC), protein content (PC) and main inflorescence length (MIL) were used to conduct associated mapping (Tang, 2019). The EMMAX software was applied to perform family-based association mapping analysis with a mixed linear model (Kang et al., 2010). Then, the visualization of haplotype blocks and linkage disequilibrium was conducted using the LDBlockShow software (Dong et al., 2021b). The interaction networks of *B. napus* proteins were obtained from the STRING database (Damian et al., 2020).

Results

Genome-wide identification and characterization of *BnUBCs*

To identify the members in the *UBC* family, the annotation file of UBC domain obtained from the Pfam database (http://pfam.xfam.org/) was applied as a query for searching against the protein dataset of *B. napus*. The peptides of assumed *BnUBCs* showing hits of both *AtUBCs* and *OsUBCs* were further employed to confirm the presence of the UBC domains by searching in the Pfam, SMART and CDD databases. In total, 200 *BnUBCs* were identified in *B. napus* (Supplementary Table 1). The full-length transcript sequences of *BnUBCs* ranged from 168 bp (*BnaA08g09170D*) to 5,656 bp (*BnaC07g05960D*) with the deduced amino acid sequences varying from 55 aa (*BnaA08g09170D*) to 1,649 aa (*BnaA07g03330D*) (Supplementary Table 1). The predicted MW of the 200 deduced BnUBC proteins ranged from 6.10 kDa (*BnaA08g09170D*) to 184.41 kDa (*BnaA07g03330D*), and their GRAVY and pIs ranged from -0.904 (*BnaC04g10080D*) to 0.112 (*BnaA08g13930D*) and from 4.19 (*BnaA04g11090D*) to 9.57 (*BnaA09g09660D*), respectively (Supplementary Table 1). As previously reported in other plant species (Die et al., 2018; Zhou et al., 2018), a wide range of pIs suggested that BnUBCs function in different subcellular environments. Furthermore, the instability index of 38 of the 200 BnUBC proteins was less than 39, and these proteins were classified as stable. In addition, the prediction of their subcellular localization indicated that 172 BnUBCs were located in the nucleus, 27 in the cytoplasm and the other one in the cell membrane. In addition, to explore the evolutionary relationships among members of this family across Brassicaceae crops, 93 *BraUBCs* and 95 *BolUBCs*

were identified in the reference genomes of *B. rapa* and *B. oleracea* respectively, following the same identification pipeline (Supplementary Table 2). Details of these *BraUBCs* and *BolUBCs* were listed in Supplementary Table 2.

Chromosomal distribution analysis of *BnUBCs*

The chromosomal positions of detected *BnUBCs* were drafted to corresponding chromosomes with the RIdeogram package of R software. The 200 *BnUBCs* were unevenly distributed across all 19 chromosomes with 97 and 103 genes located in the A_n and C_n subgenomes, respectively (Figure 1). The A_n subgenome carried, on average 9.7 *BnUBCs* on 10 chromosomes; A10 had the lowest number of *BnUBCs* (6), while A03 and A05 harbored the highest of *BnUBCs* (11 each). The average number of *BnUBCs* in the C_n subgenome was 11.4, with the lowest number (5) on C09 and the highest number on C03. Hence, no biased tendency was observed between the two subgenomes. Furthermore, on each chromosome, *BnUBCs* were unevenly distributed, especially, since the genes on A06 and A08 were located at the ends.

Phylogenetic relationships analysis of *BnUBCs*

To explore the evolutionary relationships among the *UBC* genes in *A. thaliana*, *B. rapa*, *B. oleracea*, and *B. napus*, a tree of phylogenetic was constructed based on the multiple sequence alignments of 48 *AtUBCs*, 93 *BraUBCs*, 95 *BolUBCs*, and 200 *BnUBCs*. All 200 *BnUBCs* could be categorized into 18 subfamilies and the number of groups was consistent with a previous report (Jia et al., 2019) (Figure 2). Among them, Group XVIII possessed the highest number (37) of the *BnUBCss*. Furthermore, group III (two *BnUBCs*) and group VIII (ten *BnUBCs*) members were involved in RUB and SUMO conjugation pathways, respectively, and other three subgroups (XVI, XV, VI) contained the members lacking the Cys active-site (UEVs).

Gene structure, conserved motifs as well as *cis*-acting regulatory elements analysis of *BnUBCs*

Concerning the evolution of multigene families, the variety of gene structures provided an important resource (Ischoff et al., 2002; Yunpeng et al., 2016; Pellicer et al., 2018). For exploring the multiplicity of the gene structure of *BnUBCs* among the different groups, the intron-exon structure of *BnUBCs* was compared based on the phylogenetic relationships of these genes (Figure 3A, Supplementary Table 3). The results indicated that the gene architecture of *BnUBCs* was relatively complex (Figure 3B). Only five genes (*BnaA10g10530D*, *BnaA05g17560D*, *BnaC04g10080D*, *BnaC07g46220D* and *BnaA08g09170D*) contained a single intron, while the others harbored multiple introns. *BnaA07g03330D* contained the most introns and exons. Besides, 33 of 200 *BnUBCs* lacked untranslated regions (UTR) at both ends or one end because of inaccurate annotation. The number of introns among the *BnUBCs* ranged from 1 to 26, and the length of introns ranged from 14 to 6,632 bp. Moreover, exon numbers in*BnUBCs* also varied widely across distinct subfamilies, ranging from 2 to 26. However, the majority of the members within the same subfamily tended to show similar intron-exon distribution patterns. For example, members of Group X generally possessed five exons, whereas those of Group XV, harbored eight exons. Therefore, the organization of introns and exons presented valuable evidence for the phylogenetic relationship among members of this gene family.

Furthermore, a total of 10 conserved motifs were recognized in 200 *BnUBC* genes (Figure 3C). Conserved motifs ranged in length from 11 to 74 aa, and among 10 motifs motif 3 as well as motif 1 were the most abundant. Moreover, the motif 1-4 were significantly matched to the UQ_con domain (Pfam: PF00179). Furthermore, a motif sequence search in the CDD database revealed that motifs 5 and 6 belonged to UBCc family. In addition, motifs 7-8 and motif 10 were matched to UBQ-conjugating_enzyme/RWD based on the search in the InterPro database. Motif composition varied across distinct subfamilies, whereas the conserved character of motif distributions in the same subcategory was similar, which emphasized their phylogenetic relationship. For example, all Group IX members possessed motif 1, motif 4-6 and motif 10, whereas in Group X, members generally contained motifs 1-6 and motif 10. In addition, three structural properties were detected and refined in BnUBC proteins (Supplementary Figure 1).

The *cis*-acting elements in the promoter region of a gene can modulate the initiation and efficiency of gene expression through the binding of specific transcription factors, associated with plant development, growth and stress response (Ibraheem et al., 2010). To better understand the potential functions and transcriptional regulation of *BnUBCs*, we isolated the sequences 2-kb upstream of the *BnUBC* CDSs from the *B. napus* genome sequence to search for *cis*-elements (Supplementary Table 4) (Chalhoub et al., 2014). A total of 115 functional *cis*-elements were detected in the promoter regions of *BnUBCs* (Supplementary Figure 2, Supplementary Table 5). Among these *cis*-elements, many were light-responsive and hormone-related related elements. Besides, majority of *BnUBCs* (197/200, 98.50%) contained both CAAT-box and TATA-box elements, which usually existed in eukaryon. Moreover, the MYB element, which participates in the regulation of phenylpropanoid secondary metabolism in plants was identified in 192 out of 200 *BnUBCs*. The MYC element was identified in 183 of the 200, (91.50%) *BnUBCs*, which were involved in the regulation of plant growth and development as well as resistance to environmental

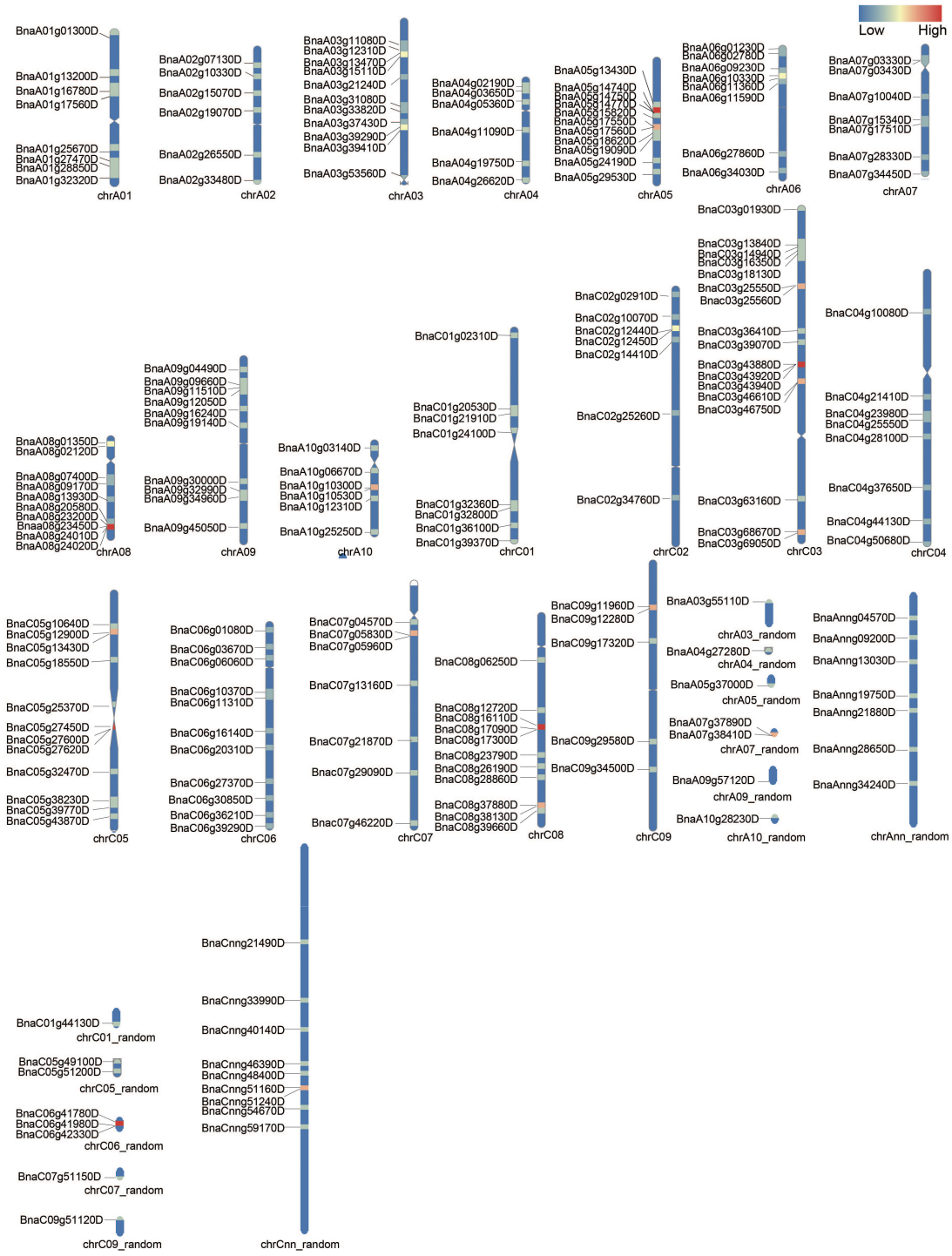


FIGURE 1
Chromosomal distribution of *BnUBCs*. The *BnUBCs*' genome locations are plotted based on the location of genes, length of chromosomes, and positions of centromeres. The chromosome name is marked at the bottom of each bar. Heatmap represents the gene density (number of genes per Mb) of each chromosome.

stress and secondary metabolite synthesis. Furthermore, many *BnUBCs* contained light-responsive elements such as Box 4 element (part of a conserved DNA module involved in the light response), G-box element (*cis*-acting regulatory element involved in the light response), GT1-motif (light responsive element), and TCT-motif (part of a light responsive element) (Supplementary Figure 2, Supplementary Table 5). Additionally, the stress-related elements, including the LTR element and TC-rich repeats were also

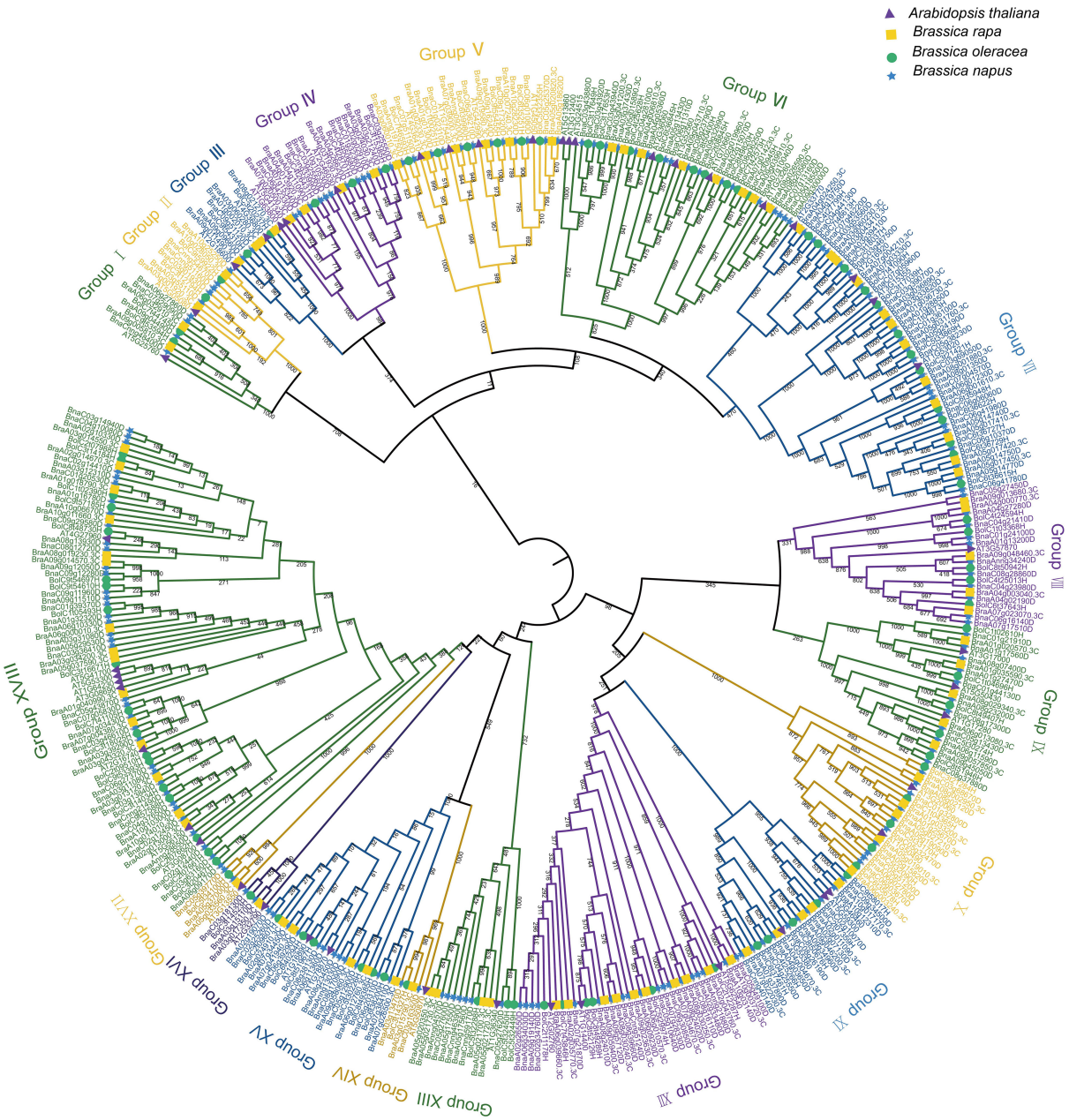


FIGURE 2
The phylogenetic relationship of *UBC* genes among *A. thaliana*, *B. rapa*, *B. oleracea* and *B. napus*. The construction of a phylogenetic tree using the *UBC* gene family among *A. thaliana*, *B. rapa*, *B. oleracea*, and *B. napus*. MEGA7.0 generated the phylogenetic tree topology using the neighbor-joining method with 1,000 bootstrap replicates. The members belonged to the same subgroup according to the node and branch they located and the characteristic of the tree. The different subgrouops were depicted in various colors.

detected in *BnUBCs*, suggesting that the *BnUBCs* played an important role in plant development and growth as well as in response to biotic and abiotic stresses.

Synteny and gene duplication analysis of *UBCs* in *B. napus, A. thaliana, B. rapa* and *B. oleracea*

Gene family expansion occurred mainly *via* the duplication of genes through whole-genome duplication (WGD), tandem duplication as well as segmental duplication events (Bodt et al., 2005; Mun et al., 2009). Moreover, the duplication of ubiquitin members through WGD events plays important role in plants (Hua et al., 2018). To further explore the evolutionary history of *BnUBC* gene family expansion we performed a syntenic comparison of genome sequences between *A. thaliana*, *B. napus*, *B. rapa* and *B. oleracea*. *A. thaliana* is the ancestor of *Brassica* species, whose structural genes have been identified and functionally annotated, therefore it is served as a prominent model system for the investigation of the evolutionary history of *Brassica*. homologous genes from inter-species comparison and paralogous genes from intra-species comparison were identified to

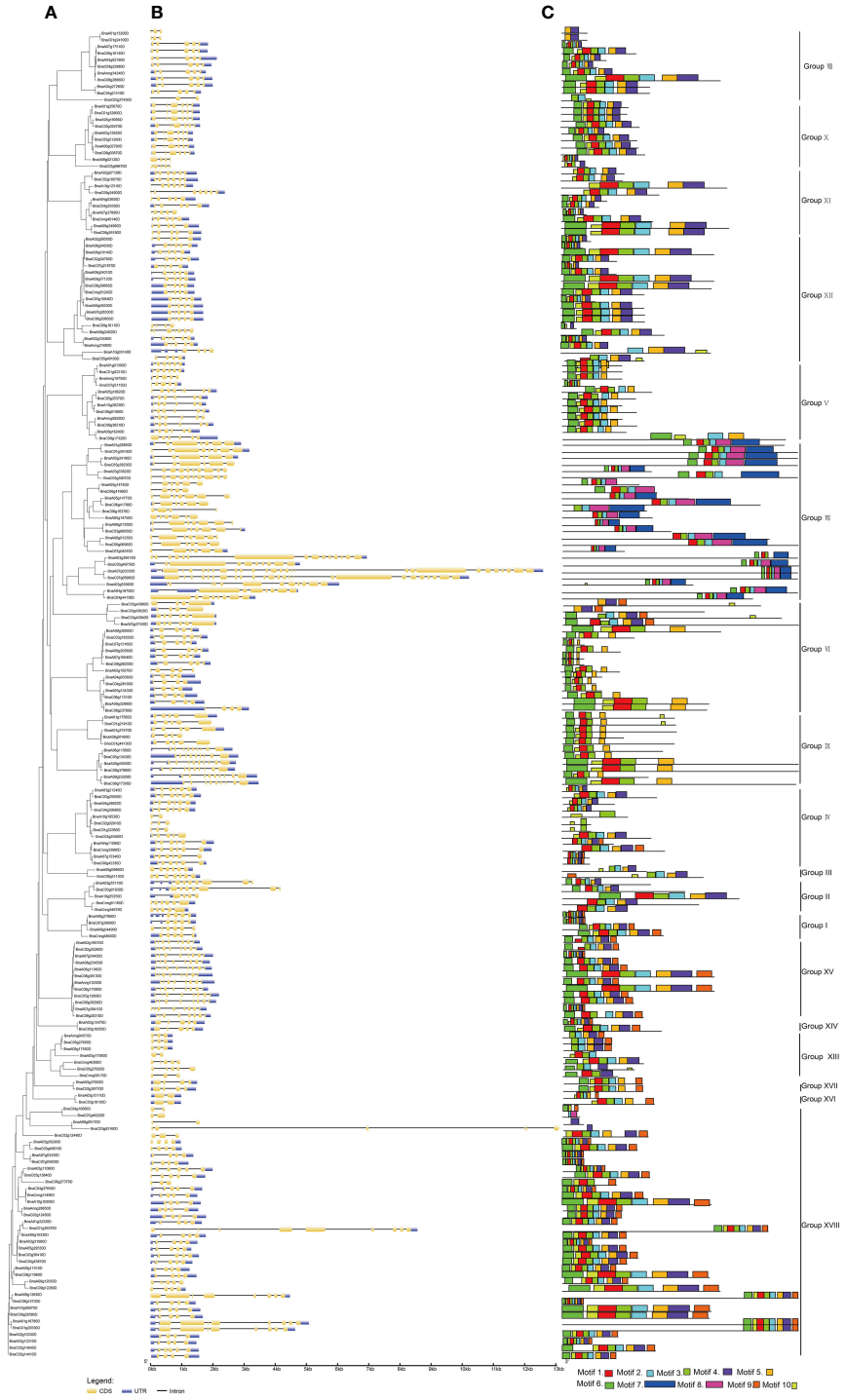


FIGURE 3
Gene structure and motif analysis of *B. napus UBC* genes according to their phylogenetic relationships. **(A)** Phylogenetic analysis. **(B)** Gene structures of *BnUBCs*. Yellow and blue-filled boxes represented CDS and UTR, respectively. Black lines indicated introns. **(C)** Conserved motifs in BnUBC proteins were detected by MEME analysis.

trace the duplicated gene pairs of *BnUBCs*. We identified 268, 104, 86 and 29 paralogous *UBC* gene pairs of within *B. napus*, *B. rapa*, *B. oleracea* and *A. thaliana*, respectively (Figure 4B). Out of 286 paralogous *UBC* gene pairs in *B. napus*, 50 and 44 pairs were detected in the A_n subgenome, and C_n subgenome, respectively, while the remaining 174 pairs were identified across the A_n and C_n subgenomes. Moreover, we classified these paralogous gene pairs into five types: dispersed, proximal, tandem, WGD and transposed (Table 1, Supplementary Table 6). In *B. napus*, 195 genes, accounting for 97.5% of all *BnUBCs*, were derived from gene

10.3389/fpls.2023.1118339

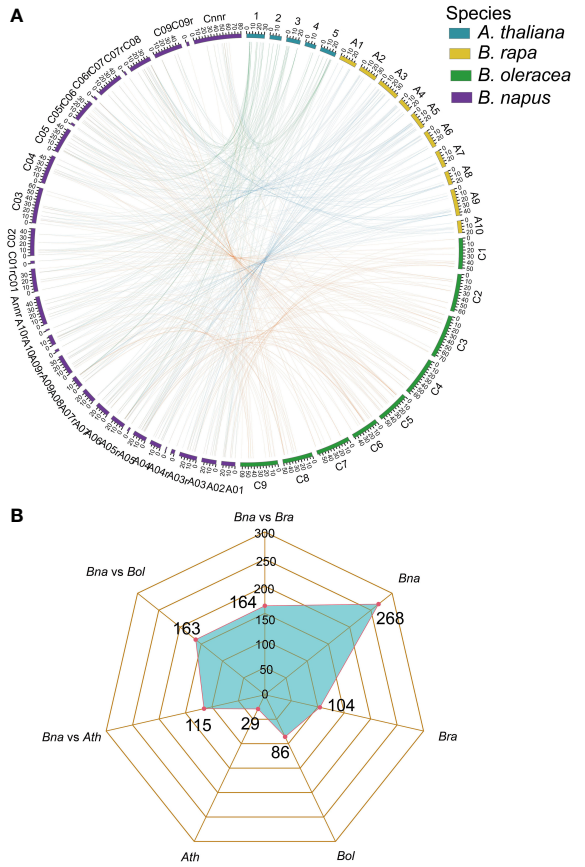


FIGURE 4
Genome-wide synteny analysis of *UBC* genes among *A. thaliana*, *B. napus*, *B. rapa*, and *B. oleracea* **(A)** Collinearity among the *UBC* genes in *A. thaliana*, *B. rapa*, *B. oleracea*, and *B. napus*. The chromosomes of *A. thaliana*, *B. rapa*, *B. oleracea* and *B. napus* were indicated with different colors corresponding to the color legend at the top right. Green lines represented the collinearity between *A. thaliana* and *B. napus*. The blue lines represented the collinearity between *B. rapa* and *B. napus*. The orange lines represented the collinearity between *B. oleracea* and *B. napus*. The visualization of the collinear correlations was performed by CIRCOS software. **(B)** Radar charts showed the number of orthologous and paralogous gene pairs of *UBC*s across four plant species.

duplication, of which WGD (64.5%) and transposed (17.4%) gene duplication events were primarily responsible for gene family expansion (Table 1, Supplementary Table 6).

Furthermore, synteny comparison between *B. napus* and the other three species revealed that 180 *BnUBCs* were collinear to the orthologous genomic regions in one or more of the other three species, 164 orthologous gene pairs were identified between *B. napus* and *B. rapa*, 163 between *B. napus* and *B. oleracea* as well as 115 between *B. napus* and *A. thaliana* (Figures 4A, B, Supplementary Table 7). It was obvious that majority of the *BnUBC*s were inherited from their progenitors. In addition, *Brassica* species experienced an extra whole-genome triplication (WGT) event after divergence from *A. thaliana* (Feng et al., 2014). Therefore, a single *A. thaliana* gene was inherited as three copies in *B. oleracea* as well as *B. rapa* and as six copies in *B. napus*, if no gene loss occurred after the WGT event. We found that only 25.9% of the *AtUBC*s corresponded to six copies in *B. napus*, suggesting that a substantial gene loss occurred during the process of polyploidization. To estimate the selection pressure on *BnUBC*s, the non-synonymous to synonymous substitution ratios (Ka/Ks) of *BnUBC* orthologous genes to *BraUBC*s, *BolUBC*s, and *AtUBC*s were calculated. The Ka/Ks less than one represents purifying selection, whereas Ka/Ks equal one indicates neutral evolution and Ka/Ks more than one represents positive selection (Nekrutenko, 2002). Thus, the Ka/Ks ratio can predict the pressure of selection on each duplicated pair throughout evolution as well as their divergence time. The Ka/Ks ratios of all the *BnUBCs* paralogous gene pairs, except two duplicated gene pairs (*BnaA07g03430D*/*BnaC07g05830D*, *BnaC02g02910D*/*BnaC03g25550D*) were less than one, suggesting that these genes were subjected to strong purifying selection (Supplementary Table 8). Moreover, the duplicated gene pair *BnaA07g03430D*/BnaC07g05830D had Ka/Ks ratio considerably larger than 1, suggesting these genes were subjected to strong evolutionary pressure under positive selection. Furthermore, the Ka and Ks values of the identified orthologous gene pairs between *A. thaliana* and *B. napus* were calculated to estimate the selection pressure on orthologous gene pairs and the divergence time of the two species. The Ka/Ks ratios of all the orthologous gene pairs ranged from 0.001 to 0.25 with an average of 0.074, which indicated these genes were under purifying selection (Supplementary Table 9). The orthologous gene pairs among *A. thaliana* and *B. napus* diverged around 16 million years ago (MYA) based on the mutational rate estimate, $R = 1.5 \times 10^{-8}$ synonymous substitutions per site per year (Koch et al., 1999; Koch et al., 2001). This result was consistent with the previously estimated divergence time (14–24 MYA) between the *A. thaliana* and *B. napus* lineages (Koch et al., 2000; Cheung et al., 2009). In addition, the Ks values of the orthologous genes between *B. napus* and its progenitors were also calculated (Supplementary Table 9). The Ka/Ks ratio of the orthologous gene pairs between *B. napus* and *B. oleracea* was significantly higher than that between *B. napus* and *B. rapa*, showing that genes in the C_n subgenome underwent relatively weaker selection pressure in *B. napus* during the process of evolution (Figure 5). In addition, the Ka/Ks values of the orthologous gene pairs between *B. napus* and *Arabidopsis* were significantly lower than those of the orthologous gene pairs identified between *B. napus* and *B. oleracea*

TABLE 1 The identification of duplicated type for *UBC* genes in *B. napus* and other three Brassicaceae species.

Species	WGD	Tandem	Proximal	Transposed	Dispersed	Total
B. napus	221	1	1	45	0	268
B. rapa	82	1	2	19	0	104
B. oleracea	66	1	2	17	0	86
A. thaliana	13	1	0	14	1	29

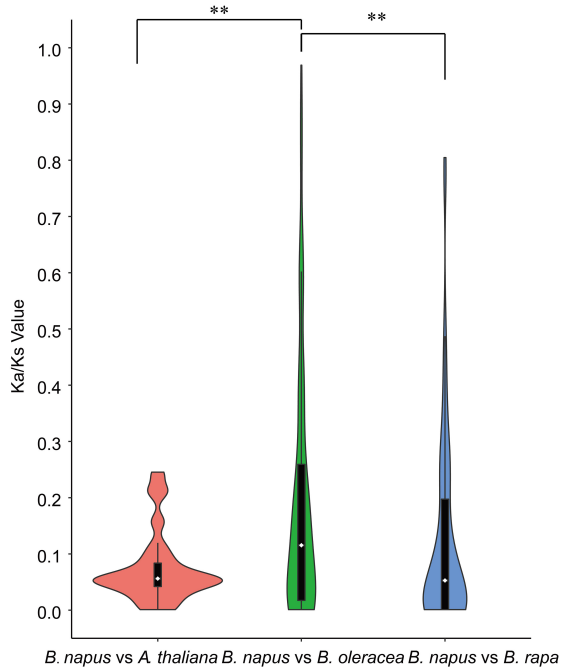


FIGURE 5
Violinplot of the pairwise Ka/Ks ratios of syntenic orthologous gene pairs. Only orthologous gene pairs with Ks <1 were considered. Wilcoxon Rank-Sum Tests were performed between different types of orthologous gene groups (**P < 0.01). Dotted lines within the violin plots represented the first, second, and third quartiles.

and between *B. napus* and *B. rapa*, indicating that orthologous gene pairs between *B. napus* and *A. thaliana* underwent intense purifying selection. Moreover, the sequences of their proteins might maintain more consistent characteristics during evolution (Figure 5).

Comprehensive analysis of miRNA-targeting *BnUBC*s

Many miRNAs were involved in regulating the function of their targeted genes in plants (Jones-Rhoades and Bartel, 2004; Yu et al., 2017; Song et al., 2019). To explore the miRNAs regulating *BnUBCs*, 106 *BnUBC* targeting genes of 71 putative miRNAs were identified, and their relationship network was constructed by Cytoscape software (Figure 6, Supplementary Table 10). It was found that *BnaC06g03670D* was targeted by the *B. napus* miRNA169 family based on an interaction network. A recent study showedthat miR169 regulated the function of the anaphase-promoting complex/cyclosome (APC/C), an essential ubiquitin-protein ligase, by targeting DUO POLLEN1 (DUO1). DUO1 upregulated the expression level of APC/C, which stimulates the production of miR159 (Zheng et al., 2011). Moreover, members of the miRNA156 family were found to regulate most of the *BnUBCs*. In a recent study, overexpression of miR156 inhibited gibberellic acid (GA)-induced and ubiquitination-mediated degradation of DELLA (Jerome Jeyakumar et al., 2020). In addition, the remaining *BnUBCs* were presumably targeted by members of the other 26 miRNA families (Figure 6).

The expression patterns of *BnUBC*s in various tissues and gene ontology analysis of *BnUBC* proteins

Several studies reported that members of the *UBC* family mediated crosstalk among diverse signaling pathways for various abiotic stress responses (Gao et al., 2017; Jue et al., 2018; Jia et al., 2019). Besides, *UBCs* were predicted to play an important role in protein and ion binding (Liu et al., 2019). To predict the putative functions of *BnUBCs*, GO enrichment analysis was performed. The GO terms were divided into three categories: biological process (BP), molecular function (MF), and cellular component (CC) (Supplementary Table 11). The majority of enriched GO terms, such as postreplication repair, belonged to the BP category (Supplementary Figure 3). The remaining GO terms were broadly associated with vegetative growth (negative regulation of flower development, root epidermal cell differentiation), response to stress (response to iron ion, cellular response to water deprivation and response to gibberellin), cell growth and protein process. The CC category included UBC13–MMS2 complex and the perinuclear region of the cytoplasm. Moreover, various MF were detected in this analysis, which included acid-amino acid ligase activity, SUMO transferase activity, ubiquitin protein ligase binding and endopeptidase activity. The results of GO enrichment analysis suggested that *BnUBCs* play crucial roles in plant development, growth, cell protein quality regulation and response to stress, consistent with the findings of previous studies (Dreher and Callis, 2007; Sadanandom et al., 2012; Jue et al., 2018).

Moreover, the expression profiles of *BnUBC* genes were examined across five major tissues (callus, leaf, root, bud and silique) based on the previously published RNA-seq datasets of *B. napus* (Yao et al., 2020). The expression levels of *BnUBC*s were estimated with FPKM and displayed as a heatmap (Figure 7A). A total of 38 *BnUBCs* showed relatively weak expression (FPKM < 1) and 24 *BnUBC*s were silience in any of the five tissues. The remaining *BnUBCs* showed high expression levels (FPKM ≥ 1), the majority of which expressed at a specific organ or differentially expressed in various tissues (Figure 7B). *BnaC06g16140D* (ortholog of *AT3G57870*) presented the highest expression in all tissues, suggesting its critical roles in plant growth. Genes showing low expression levels (FPKM <1) across all five tissues could be regarded as pseudogenes (Figure 7, Supplementary Table 12). Notably, in our study, most of *BnUBC*s exhibited relatively lower expression levels in the leaf than in other tissues (Supplementary Table 12).

Furthermore, the expression divergence between the duplicated genes in *B. napus* investigated. For example, *BnaA06g10330D* was silence across all tissues, whereas its paralog *BnaA03g31080D* presented a high expression level in various tissues. Similarly, *BnaA04g02190D* expressed in all tissues, however, its paralogs *BnaA01g13200D* and *BnaC01g24100D* didn't express in any

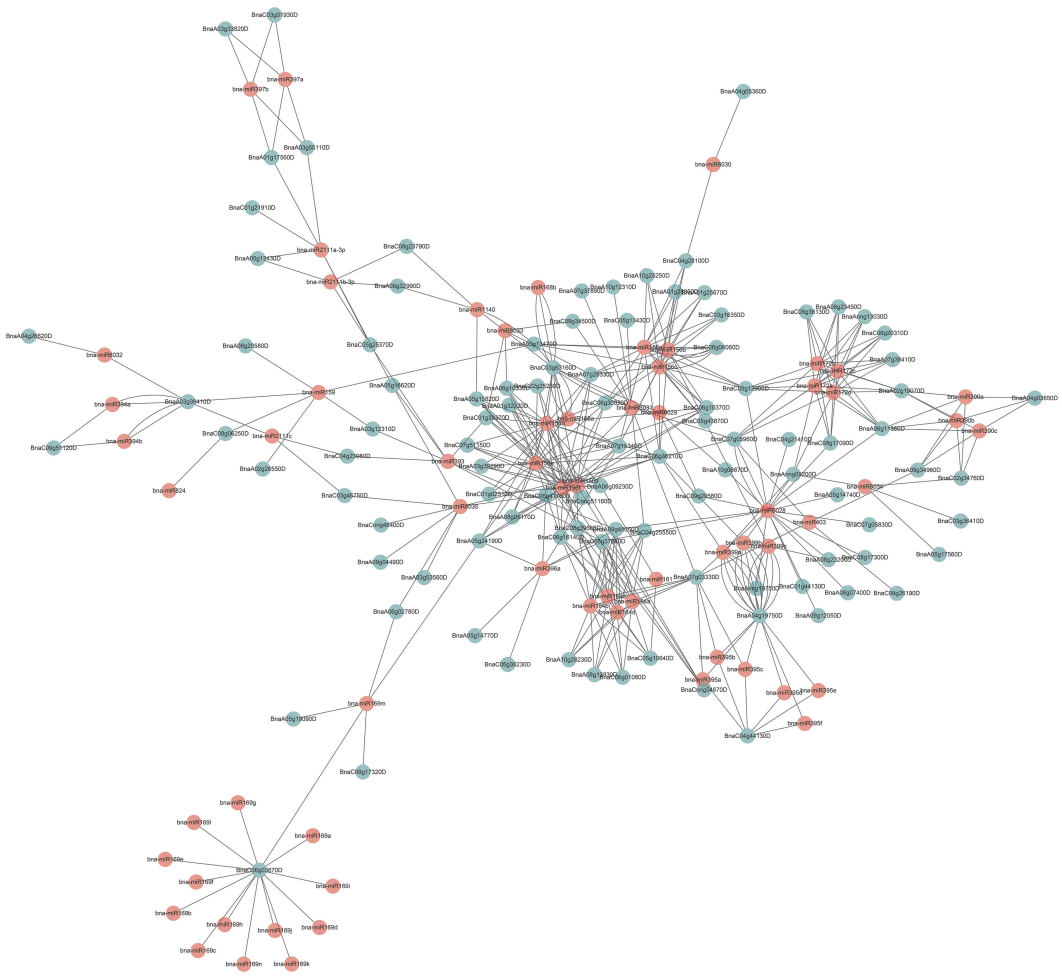


FIGURE 6
Schematic representation of the interaction network between the miRNAs and their putative *BnUBC* targets.

tissue. All these results could provide a clue for the functional divergence between duplicated gene pairs in *BnUBCs*.

Genetic effects of *BnUBC*s on agronomic traits

SNPs in a natural population with 324 *B. napus* accessions collected from worldwide countries were identified (Tang, 2019) (Supplementary Table 13) for investigating the genetic variants of *BnUBCs* in germplasm. Each *BnUBC* gene contained 34 SNPs on average, which was close to the 36 SNPs in each gene across the entire genome. Taking genome size into account, the number of SNP in each kilobase (kb) among the *BnUBCs* was calculated (17 SNPs/kb), whereas 13 SNPs/kb for the genes across entire genome. *BnUBCs* in the A_n subgenome contained slightly higher SNP density (20 SNPs/kb) than that in the C_n subgenome (14 SNPs/kb). Moreover, variations in SNP number between some paralogous gene pairs of *BnUBCs* were observed. For example, *BnaA03g11080D* possessed 65 SNPs, however, its paralog *BnaC02g12440D* contained no SNP. Moreover, the paralogs *BnaA07g03330D*/*BnaC03g46750D*, contained 235 and 33 SNPs, respectively. Based on the paired t-test, no significant difference was detected in SNP density between the *BnUBC* paralogs. Additionally, SNP annotation presented that 1,671 SNPs existed in the exon regions of *BnUBCs* and 600 SNPs lead to missense mutations.

Ubiquitination is a well-characterized post-translational that regulates plant growth and developmental processes contributing to diverse phenotypes, and affect agronomic traits (Dreher and Callis, 2007; Vierstra, 2009; Sadanandom et al., 2012; Linden and Callis, 2020; Yu and Hua, 2022). Association mapping analysis was performed to explore the impact of *BnUBCs* on agronomic phenotype. In total, 96 SNPs in 33 *BnUBC* genes (Supplementary Table 13) were significantly associated with the investigated agronomic traits ($p < 0.001$). Moreover, *BnaC02g25260D* was significantly associated with plant height (Figure 8A, B). Furthermore, the two genotypes of accessions were divided according to the most significantly associated SNP of *BnaC02g25260D*, and the plant height between the two genotypes was statistically significant according to the t-test ($p < 2.7$ e–4) (Figure 8C). The protein interaction network of *BnaC02g25260D* was obtained from STRING (Damian et al., 2020) (Figure 8D) and

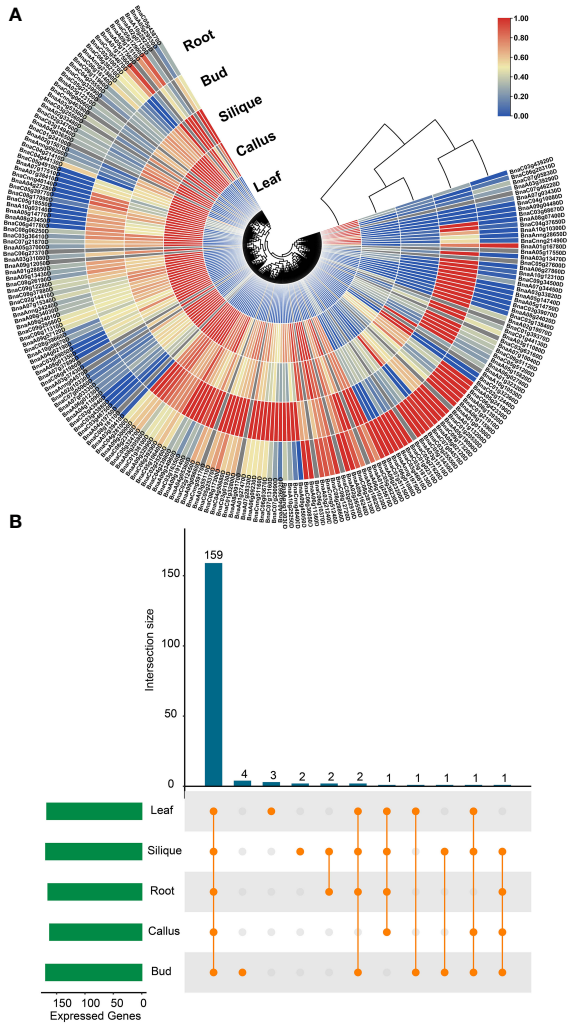


FIGURE 7
Expression profiles of *BnUBCs* across various tissues. **(A)** Expression data were represented by values of FPKM, and different colors showed different expression levels. **(B)** Number of *BnUBCs* expressed across various tissues.

the interacted genes were enriched in the regulation of embryonic development (GO:0045995), positive regulation of cell size (GO:0045793), cellulose biosynthetic process (GO:0030244), circadian rhythm (GO:0007623), positive regulation of flower development (GO:0009911), detection of visible light (GO:0009584), leaf development (GO:0048366), regulation of hormone levels (GO:0010817) and chlorophyll biosynthetic process (GO:0015995) (Figure 8E). Overall, the above categories were related to plant development and growth, the regulation of hormones and oil content, which affected the plant height and the cell size in seed formation (Wang X, et al., 2020; Dong et al., 2021a). Furthermore, SNPs distributed in *BnaAnng34240D* were significantly associated with flowering time and silique density (Supplementary Figure 4A-D, G-H). The protein interaction networks of *BnaAnng34240D* were also obtained from STRING (Supplementary Figure 5A), and its interacted proteins enriched in the protein modification process, regulation of transcription and so on (Supplementary Figure 5B). Besides, *BnaA07g28330D* was significantly associated with TSW, which affect the crop yield (Supplementary Figure 4E, F). Furthermore, we found that *BnaA07g28330D* exhibited a higher expression level in the silique of high oil content accession than in that in the silique of low oil content accession (https://bnaomics.ocri-genomics.net/tools/exp-view/index.php) (Tang, 2019).

Discussion

Ubiquitin-conjugating enzymes transfer ubiquitin from ubiquitin-activating enzymes to ubiquitin ligases, which is a key step in protein ubiquitination (Bae and Kim, 2013). In plants, UBC proteins are involved in multiple crucial processes, including growth, development and abiotic stress response (Welchman et al., 2005; Dreher and Callis, 2007; Sadanandom et al., 2012). However, only a few members of the UBC family have been characterized in plants such as *Arabidopsis* (Kraft et al., 2005), rice (E et al., 2015), maize (Jue et al., 2015) and tomato (Sharma and Bhatt, 2017) on the genome-wide level. *B. napus* is a valuable oil crop and an ideal polyploid model for studying the evolution, domestication and genetics of polyploids. However, a systematical investigation of the *BnUBC* family *s* has not been conducted to date. In this study, the availability of the *B. napus* whole-genome sequence provides opportunities for characterizing the *BnUBC* family and revealing its genetic effects on agronomic traits.

In this study, 200 *UBC* genes including two RUB genes, ten SUMO and 31 *UEV* genes, encoding UBC domain-containing ubiquitin-conjugating enzymes were detected in *B. napus*. The number of *UBC* gene in *A. thaliana*, maize, tomato, rice and *B. napus* (48, 75, 59, 39 and 200, respectively) are not correlated with the genome size (~125, ~466, ~2300, ~900, and ~825 Mb, respectively). Increasing evidence suggest that segmental duplications may be the main factor responsible for the expansion of gene families in plants (Wang et al., 2015; Liu et al., 2017; Wang et al., 2017). The loss and gain of genes during the process of polyploidization determine the genetic variability, which ultimately affects the function of the protein (Albalat and Cañestro, 2016). Analysis of gene structures and motifs of *BnUBCs* revealed that variation in the number of exons and introns as well as motif constitution may lead to their functional diversity.

Cis-acting elements in promoter regions regulate the transcription of genes that participate in various physiological processes (responses to hormone stress, plant growth and development) and that form the basic functional links among the complex regulatory networks (Nuruzzaman et al., 2014). Abundant *cis*-acting elements which were light-responsive and related to the hormone, plant growth and development were detected in the promoter regions of *BnUBCs*. Two *cis*-acting elements (TATA-box and CAAT-box) were presented in most *BnUBCs* (Supplementary Figure 2). Some *cis*-acting elements widely distributed in the eukaryotes, such as TATA-box and CAAT-box were also found in *BnUBCs*. These *cis*-acting elements constitute the RNA transcription factors binding sites (Laloum et al., 2013) and regulate the gene transcription process in conjunction with its binding factor. Moreover, other *cis*-acting elements such as MYB,

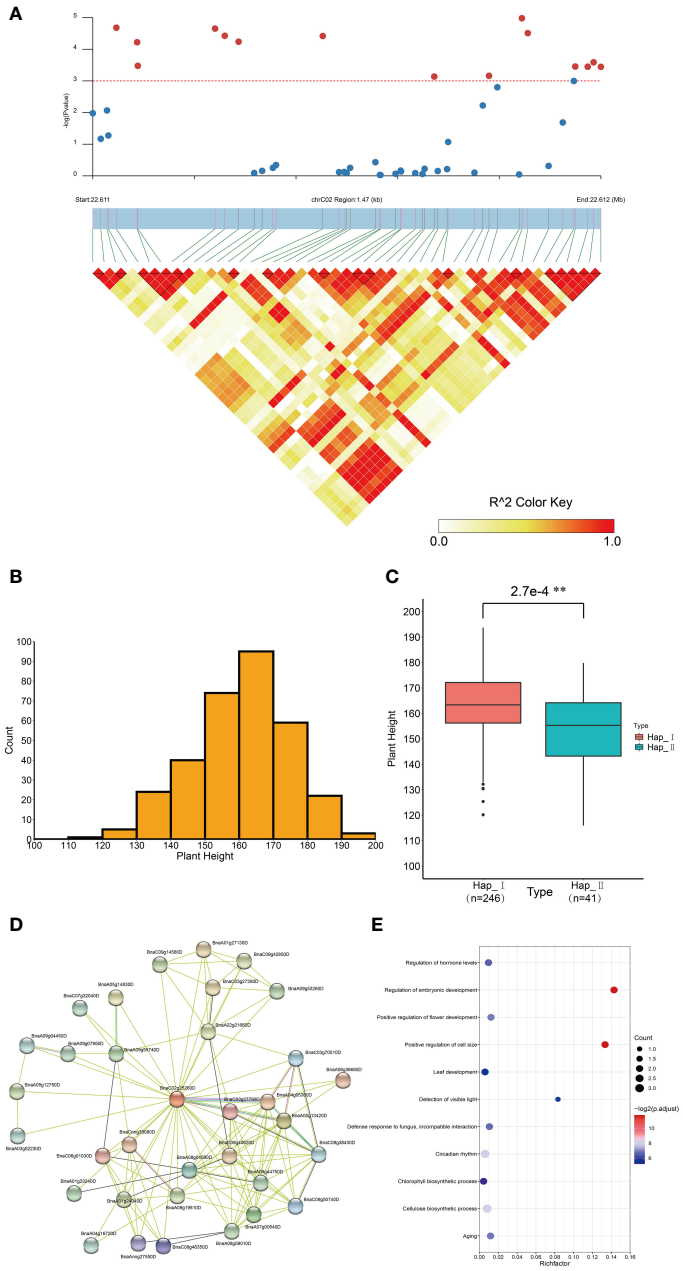


FIGURE 8
Association mapping analysis of *BnaC02g25260D* in *B. napus* germplasm with 324 core collections. **(A)** Significant association of *BnaC02g25260D* with tplant height. **(B)** Distribution of the plant height of 324 accessions. **(C)** Box plot showed the comparison of plant height between two haplotypes divided based on the most significantly associated SNP in *BnaC04g00810D*. **(D)** Protein-protein interaction network of BnaC02g25260D. **(E)** GO enrichment analysis of BnaC02g25260D interacting proteins. **p < 0.01.

MYC, ARE and ABRE, which are related to plant growth and development, were also found in the majority of *BnUBC*s. For example, MYB transcription factors function in plant breeding and response to stresses (Li et al., 2019) and MYC proteins function as transcriptional activators in abscisic acid (ABA) signaling (Abe et al., 2003). Besides, the abscisic-responsive element (ARE) *cis*-acting element is necessary for the anaerobic induction and the ABA-responsive element (ABRE) *cis*-acting element functioned in the response to ABA stress (Narusaka et al., 2003). Furthermore, many *cis*-acting elements involved in response to light were found in the *BnUBC* promoter regions, suggesting their potential role in regulating the pathways associated with light responsiveness. Overall, the *cis*-acting elements detected in *BnUBC* promoters suggested that they played crucial roles in plant growth, development and responses to abiotic stress and light.

Phylogenetic analysis divided BnUBC proteins into 18 groups, which is similar to the previous studies (Kraft et al., 2005; Dong et al., 2016; Sharma and Bhatt, 2017). In addition, the UBC domains showed uniform distribution in all the *UBC*s and gene expansion caused the variation of this core element during the evolution.

Group XVIII possessed the largest number of *BnUBC*s with the most variation in gene structure and motifs, indicating that members in this group evolved more diverse functions than the other groups. Moreover, the phylogenetic classification of *BnUBCs* was further supported by the analyses of gene architecture and conserved motifs. Additionally, most of the members in the same subgroup shared similar gene structures and conserved motifs, but showed obvious different in *cis*-acting regulatory elements (Supplementary Figure 2), indicating their potential function diversity (Zou et al., 2011; Oudelaar and Higgs, 2021). Furthermore, the phylogenetic relationships of *AtUBC*s and *BnUBC*s suggested that gene duplication was likely a major factor in the diversification of *BnUBC* genes during evolution.

Gene duplication is considered a major mechanism leading to gene family expansion and functional diversification (Lynch and Conery, 2000; Prince and Pickett, 2002; Bianconi et al., 2018). In plants, three types of genome duplication have been reported, including WGD, tandem duplication and chromosomal segmental duplication (Ramsey and Schemske, 1998). According to many previous studies, WGD and segmental duplication are important factors for genome duplication and gene expansion (Ma et al., 2017; Wu et al., 2018; Zhu et al., 2020). After divergence from the ancestor of *Arabidopsis* lineage, *Brassica* species underwent WGT of approximately 13 MYA. Allotetraploid *B. napus* was formed by the interspecific hybridization between *B. rapa* and *B. oleracea* (Allender and King, 2010; Chalhoub et al., 2014). Thus, the genome size of *B. napus* was expanded and a single gene in *A. thaliana* corresponded to six copies in *B. napus* (Lysak et al., 2005). According to collinearity analysis, WGD showed a large contribution to genome expansion in *Brassica* species, followed by transposed duplication (Table 1). In this study, the number of *BnUBC*s (200) was approximately 4-fold higher than that of *AtUBC*s, while the number of *BraUBC* and *BolUBC* genes was <3-fold higher than that of *AtUBC*s, indicating gene loss in Brassica species during evolution (Albalat and Cañestro, 2016). Moreover, the numbers of *BnUBCs* in the A_n and C_n subgenomes were 97 and 103, respectively, which was almost consistent with that in the genomes of its progenitors *B. rapa* and *B. oleracea* genomes. These results demonstrate that most gene loss of *UBC*s occurred after whole genome triplication in *Brassica* ancestors and *BnUBC*s were mainly inherited from their progenitors (Chalhoub et al., 2014). Notably, several orthologs of *AtUBCs* (*AT3G57870*, *AT1G14400*, *AT5G56150*, *AT1G23260*, *AT1G17280*, *AT2G36060* and *AT3G08690*) maintained six copies in *B. napus* and were not loss after WGD, implying their crucial roles in plant growth and development. In *B. napus*, the Ka/Ks ratios between *UBC* paralogs were obviously less than one, indicating that these genes were undergone purifying selection during evolution.

The majority of *BnUBCs* localized to the nucleus, and only a few localized to the cytoplasm (Supplementary Table 1). This is consistent with a previous study, which showed their existence in the nuclear region for their involvement in fruit ripening (Wang et al., 2014). Moreover, these findings were also consistent with the results of GO enrichment analysis (Supplementary Figure 3). GO enrichment analyses predicted that *BnUBC*s are mainly involved in ubiquitination, cell growth and response to stresses (Supplementary Figure 3, Supplementary Table 11). However, the functional analysis of *BnUBC*s had been lacking. Therefore, we analyzed the expression profiles of these *BnUBC*s in various tissues (bud, callus, root, silique, and leaf) to predict their possible functions. The RNA-seq data (Yao et al., 2020) showed that the expression levels of *BnUBC*s varied across various tissues, consistent with previous studies (Cui et al., 2012; Jeon et al., 2012; E et al., 2015; Jue et al., 2018). For instance, *AtUBC1* and *AtUBC2* expressed in root, leaf, flower and seedling (Xu et al., 2009) And their orthologs in *B. napus* showed relatively high expression in bud, root and silique. The average expression level of *BnUBCs* was low in the leaf and the least number of genes expressed in callus (Figure 7, Supplementary Table 12). Moreover, diverse expression patterns were found between several duplicated genes, suggesting their divergence through subfunctionalization, neofunctionalization or pseudogenization in *B. napus* polyploidization (Chaudhary et al., 2009). Moreover, the difference of *cis*-acting regulatory elements in the promoter regions between *BnUBC*s may contribute to their expression level and function divergences.

To reveal the genetic effects of *BnUBC*s on agronomic traits, SNPs of *BnUBC*s were detected in accessions comprising a *B. napus* natural population (Tang, 2019) (Supplementary Table 13). The average SNP density of Bn*UBC*s (17 SNPs/kb) was slightly higher than that of genes in the entire genome (13 SNPs/kb), suggesting that a large number of polymorphisms accumulated in *BnUBC*s during evolution. The SNP density of *BnUBCs* in the A_n subgenome was higher than that in the C_n subgenome, which is consistent with some other gene families in *B. napus* (Zhu et al., 2020; Wahid et al., 2022; Xie et al., 2022). Classically, the difference in genetic variations between paralogs may lead to subfunctionalization, pseudogenization or neofunctionalization (Schiessl et al., 2017). Several duplicated *BnUBC* gene pairs, such as *BnaA03g11080D*/*BnaC02g12440D* and B*naA07g03330D*/*BnaC03g46750D*, showed significant different in the number of SNPs. Differences were also observed in the expression levels of these duplicated genes, implying their functional divergence. Furthermore, association mapping analysis of *BnUBC*s, based on SNPs, revealed their genetic effects on agronomic traits. For example, *BnaC02g25260D* was significantly associated with plant height (Figure 8). According to the results of GO enrichment analysis, *BnaC02g25260D* -interacting protein were likely involved in the regulation of plant cell growth, hormone levels and chlorophyll biosynthetic process, and could eventually influence plant height (Wang X, et al., 2020; Dong et al., 2021a). In addition, *BnaAnng34240D* was significantly associated with agronomic traits such as flowering time and main inflorescence silique density (Supplementary Figure 4). Ubiquitination, as reported in previous studies, is one of the crucial mechanisms that control the photoperiodic regulation of floral organ development (Sadanandom et al., 2012; Piñeiro and Jarillo, 2013). Moreover, the interacting proteins of BnaAnng34240D were involved in protein modification, regulation of transcription and so on. *AtSUMO1/2* and S *AtSCE1a* (ortholog of *BnaAnng34240D*) are implicated in ABA responses and the ubiquitin-like SUMO protease 1 (ULP1) gene *AtESD4* is involved in flowering time regulation (Miura et al., 2007). In addition, *BnaA07g28330D* was significantly associated with TSW, an

important trait affecting crop yield. The obviously different expression levels of *BnaA07g28330D* between high and low oil content accessions in silique, implying its effects on oil content. Thus, the results of this analysis in the present study provided a valuable resource including candidate *UBC* genes affecting agronomic traits in *B. napus*.

To date, the *UBC* family were reported in several plant species. In this study, the characteristics of the *BnUBC* were compared with those of the *UBC* family in other plant species to help broaden our understanding of the differences or similarities in E2s across species. When compared with other species, this family in *B. napus* was largest and contained most subgroups suggesting the complexity of the allotetraploid genome. Moreover, majority of the *UBC*s were derived from WGD in *B. napus* and *Vitis vinifera* (Gao et al., 2017), whereas by segmental duplication in rice (E et al., 2015), tomato (Sharma and Bhatt, 2017), maize (Jue et al., 2015) and banana (Dong et al., 2016). The physicochemical properties of BnUBCs were similar to those in potato (Liu et al., 2019), maize (Jue et al., 2015) and banana (Dong et al., 2016). Additionally, and the chromosomal locations of the *BnUBC*s predicted in this study were partially consistent with previous reports (Miura et al., 2011; Cui et al., 2012; Cheng et al., 2017).

Our study focused on the *UBC* family, which represents only a part of the UPS, which generally contains the sequential action of E1, E2, and E3 enzymes. To gain better insight into the regulatory functions of the UPS, we compared our results with the other family in UPS. WGD played a predominant role in the expansion of the ubiquitin family (Hua et al., 2018), which is consistent with *BnUBCs*. Different from the similarity of physicochemical properties within the *UBC* family across species, significant sequence diversities were discovered between Poaceae and Brassicaceae (Hua and Yu, 2019). Only two members of the *E1* family in *A. thaliana* were detected with experimental verification (Hatfield et al., 1997). In addition, these two genes did not exhibit different expression patterns and the encoded E1 proteins showed no significant difference in enzymatic activities. However, the *BnUBC*s differentially expressed in various tissues and displayed various gene structures and motifs between different subgroups. With regard to the *E3* family, it can be divided into three types according to their subunit composition and functional mechanism, including homologous to E6-AP carboxyl terminus (HECT), Really Interesting New Gene (RING) and U- box type and cullin-RING ligase (CRL) (Vierstra, 2009). Thus, the *E3* family was more complicated than the *UBC* family based on the classification of its domains. The U-box type *E3* family contains 79 members in maize (Li et al., 2022), 68 in sorghum (Fang et al., 2022) and 62 in tomato (Sharma and Taganna, 2020); this family is slightly larger in size than the *E2* family. Furthermore, the *E3* genes are involved in plant development, growth and abiotic stress response, which is consistent with the functions of *E2* genes reported previously (Sharma and Taganna, 2020; Wang et al., 2021; Fang et al., 2022; Li et al., 2022; Lin et al., 2022).

Conclusion

In the present research, we conducted a comprehensive investigation of the *UBC* family in *B. napus*. A total of 200 *BnUBC*s were identified and classified into 18 groups. The gene structures and motifs were highly conserved among members of the same phylogenetic subgroup. Moreover, *cis*-acting elements found in the *BnUBC*s promoters and the expression patterns of *BnUBC*s in diverse tissues demonstrated that these genes play a crucial role in plant growth and development. In addition, synteny analysis of the *UBC*s between *B. napus* and three ancestors (*A. thaliana*, *B. oleracea* and *B. rapa*) revealed the expansion history of the *BnUBC* gene family. Furthermore, genetic variations identification and association mapping analyses of *BnUBC*s uncovered their potential genetic effects on agronomic traits related to oil content and yield in *B. napus*. Overall, this study provided useful information about *BnUBCs* and will facilitate functional studies as well as the genetic breeding of *B. napus* in the future.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

SY, CT, and XY designed the research. SY, HW, XL and PH performed the experiments. SY, MX, MH and XC analyzed the data. SY, CT, and XY wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The project was supported by the National Natural Science Foundation of China (nos 32201220, 31770250) and the Young Top-notch Talent Cultivation Program of Hubei Province for CB.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1118339/full#supplementary-material

References

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* 15 (1), 63–78. doi: 10.1105/tpc.006130

Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. *Nat. Rev. Genet.* 17 (7), 379–391. doi: 10.1038/nrg.2016.39

Allender, C. J., and King, G. J. (2010). Origins of the amphiploid species *Brassica napus* l. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biol.* 10, 54. doi: 10.1186/1471-2229-10-54

Arrigoni, A., Grillo, B., Vitriolo, A., De Gioia, L., and Papaleo, E. (2012). C-terminal acidic domain of ubiquitin-conjugating enzymes: a multi-functional conserved intrinsically disordered domain in family 3 of E2 enzymes. *J. Struct. Biol.* 178 (3), 245–259. doi: 10.1016/j.jsb.2012.04.003

Aung, K., Lin, S. I., Wu, C. C., Huang, Y. T., Su, C. L., and Chiou, T. J. (2006). pho2, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. *Plant Physiol.* 141 (3), 1000–1011. doi: 10.1104/pp.106.078063

Bae, H., and Kim, W. T. (2013). The n-terminal tetra-peptide (IPDE) short extension of the U-box motif in rice SPL11 E3 is essential for the interaction with E2 and ubiquitin-ligase activity. *Biochem. Biophys. Res. Commun.* 433 (2), 266–271. doi: 10.1016/j.bbrc.2013.03.005

Bae, H., and Kim, W. T. (2014). Classification and interaction modes of 40 rice E2 ubiquitin-conjugating enzymes with 17 rice ARM-u-box E3 ubiquitin ligases. *Biochem. Biophys. Res. Commun.* 444 (4), 575–580. doi: 10.1016/j.bbrc.2014.01.098

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37 (Web Server issue), W202–W208. doi: 10.1093/nar/gkp335

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4 (11), 879–887. doi: 10.1038/s41477-018-0289-4

Bianconi, M. E., Dunning, L. T., Moreno-Villena, J. J., Osborne, C. P., and Christin, P. A. (2018). Gene duplication and dosage effects during the early emergence of C4 photosynthesis in the grass genus alloteropsis. *J. Exp. Bot.* 69 (8), 1967–1980. doi: 10.1093/jxb/ery029

Bodt, S. D., Maere, S., and Peer, Y. (2005). Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* 20 (11), 591–597. doi: 10.1016/j.tree.2005.07.008

Burroughs, A. M., Jaffee, M., Iyer, L. M., and Aravind, L. (2008). Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *J. Struct. Biol.* 162 (2), 205–218. doi: 10.1016/j.jsb.2007.12.006

Callis, J. (2014). The ubiquitination machinery of the ubiquitin system. *Arabidopsis Book* 12, e0174. doi: 10.1199/tab.0174

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., et al. (2014). Plant genetics. early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345 (6199), 950–953. doi: 10.1126/science.1253435

Chaudhary, B., Flagel, L., Stupar, R. M., Udall, J. A., Verma, N., Springer, N. M., et al. (2009). Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (gossypium). *Genetics* 182 (2), 503–517. doi: 10.1534/genetics.109.102608

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009

Cheng, M. C., Kuo, W. C., Wang, Y. M., Chen, H. Y., and Lin, T. P. (2017). UBC18 mediates ERF1 degradation under light-dark cycles. *New Phytol.* 213 (3), 1156–1167. doi: 10.1111/nph.14272

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* 4 (5), 258–268. doi: 10.1038/s41477-018-0136-7

Cheung, F., Trick, M., Drou, N., Lim, Y. P., Park, J. Y., Kwon, S. J., et al. (2009). Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. *Plant Cell* 21 (7), 1912–1928. doi: 10.1105/tpc.108.060376

Chou, K. C., and Shen, H. B. (2010). Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PloS One* 5 (6), e11335. doi: 10.1371/journal.pone.0011335

Christensen, D. E., Brzovic, P. S., and Klevit, R. E. (2007). E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat. Struct. Mol. Biol.* 14 (10), 941–948. doi: 10.1038/nsmb1295

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6 (2), 80–92. doi: 10.4161/fly.19695

Criqui, M. C., de Almeida Engler, J., Camasses, A., Capron, A., Parmentier, Y., Inzé, D., et al. (2002). Molecular characterization of plant ubiquitin-conjugating enzymes belonging to the UbcP4/E2-C/UBCx/UbcH10 gene family. *Plant Physiol.* 130 (3), 1230–1240. doi: 10.1104/pp.011353

Cui, F., Liu, L., Zhao, Q., Zhang, Z., Li, Q., Lin, B., et al. (2012). *Arabidopsis* ubiquitin conjugase UBC32 is an ERAD component that functions in brassinosteroid-mediated salt stress tolerance. *Plant Cell* 24 (1), 233–244. doi: 10.1105/tpc.111.093062

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server, (2017 release). *Nucleic Acids Res.* 46 (W1), W49–W54. doi: 10.1093/nar/gky316

Damian, S., Gable, A. L., Nastou, K. C., David, L., Rebecca, K., Sampo, P., et al. (2020). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49 (D1), D605-D612. doi: 10.1093/nar/gkaa1074

Die, J. V., Gil, J., and Millan, T. (2018). Genome-wide identification of the auxin response factor gene family in cicer arietinum. *BMC Genomics* 19 (1). doi: 10.1186/s12864-018-4695-9

Dong, Z., Alam, M. K., Xie, M., Yang, L., and Liu, S. (2021a). Mapping of a major QTL controlling plant height using a high-density genetic map and QTL-seq methods based on whole-genome resequencing in *Brassica napus*. *G3-Genes Genomes Genet.* 11 (7), jkab118. doi: 10.1093/g3journal/jkab118

Dong, S. S., He, W. M., Ji, J. J., Zhang, C., and Yang, T. L. (2021b). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings Bioinf.* 22 (4), bbaa227. doi: 10.1101/2020.06.14.151332

Dong, C., Hu, H., Jue, D., Zhao, Q., Chen, H., Xie, J., et al. (2016). The banana E2 gene family: Genomic identification, characterization, expression profiling analysis. *Plant Sci.* 245, 11–24. doi: 10.1016/j.plantsci.2016.01.003

Dong, B., Rengel, Z., and Delhaize, E. (1998). Uptake and translocation of phosphate by pho2 mutant and wild-type seedlings of *Arabidopsis thaliana*. *Planta* 205 (2), 251–256. doi: 10.1007/s004250050318

Doroodian, P., and Hua, Z. (2021). The ubiquitin switch in plant stress response. *Plants (Basel)* 10 (2). doi: 10.3390/plants10020246

Dreher, K., and Callis, J. (2007). Ubiquitin, hormones and biotic stress in plants. *Ann. Bot.* 99 (5), 787–822. doi: 10.1093/aob/mcl255

E, Z., Zhang, Y., Li, T., Wang, L., and Zhao, H. (2015). Characterization of the ubiquitin-conjugating enzyme gene family in rice and evaluation of expression profiles under abiotic stresses and hormone treatments. *PloS One* 10 (4), e0122621. doi: 10.1371/journal.pone.0122621

Eddy, S. R. (2011). Accelerated profile HMM searches. *PloS Comput. Biol.* 7 (10), e1002195. doi: 10.1371/journal.pcbi.1002195

Fang, Y., Du, Q., Yang, Q., Jiang, J., Hou, X., Yang, Z., et al. (2022). Identification, characterization, and expression profiling of the putative U-box E3 ubiquitin ligase gene family in *Sorghum bicolor*. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.942302

Fang, S., and Weissman, A. M. (2004). A field guide to ubiquitylation. *Cell Mol. Life Sci.* 61 (13), 1546–1561. doi: 10.1007/s00018-004-4129-5

Feng, C., Jian, W., and Wang, X. (2014). Genome triplication drove the diversification of brassica plants. *Horticult Res.* 1, 14024. doi: 10.1038/hortres.2014.24

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39 (Web Server issue), W29–W37. doi: 10.1093/nar/gkr367

Gao, Y., Wang, Y., Xin, H., Li, S., and Liang, Z. (2017). Involvement of ubiquitin-conjugating enzyme (E2 gene family) in ripening process and response to cold and heat stress of *Vitis vinifera*. *Sci. Rep.* 7 (1), 13290. doi: 10.1038/s41598-017-13513-x

Glickman, M. H., and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol. Rev.* 82 (2), 373–428. doi: 10.1152/physrev.00027.2001

Guo, Y., Liu, J., Zhang, J., Liu, S., and Du, J. (2017). Selective modes determine evolutionary rates, gene compactness and expression patterns in *Brassica*. *Plant J.* 91 (1), 34–44. doi: 10.1111/tpj.13541

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., et al. (2020). RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.* 6, e251. doi: 10.7717/peerj-cs.251

Hatfield, P. M., Gosink, M. M., Carpenter, T. B., and Vierstra, R. D. (1997). The ubiquitin-activating enzyme (E1) gene family in *Arabidopsis thaliana*. *Plant J.* 11 (2), 213–226. doi: 10.1046/j.1365-313x.1997.11020213.x

Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system. resolution, affinity purification, and role in protein breakdown. *J. Biol. Chem.* 258 (13), 8206–8214.

Hodson, C., Purkiss, A., Miles, J. A., and Walden, H. (2014). Structure of the human FANCL RING-Ube2T complex reveals determinants of cognate E3-E2 selection. *Structure* 22 (2), 337–344. doi: 10.1016/j.str.2013.12.004

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31 (8), 1296–1297. doi: 10.1093/bioinformatics/btu817

Hua, Z., Doroodian, P., and Vu, W. (2018). Contrasting duplication patterns reflect functional diversities of ubiquitin and ubiquitin-like protein modifiers in plants. *Plant J.* 95 (2), 296–311. doi: 10.1111/tpj.13951

Hua, Z., and Yu, P. (2019). Diversifying evolution of the ubiquitin-26S proteasome system in brassicaceae and poaceae. *Int. J. Mol. Sci.* 20 (13). doi: 10.3390/ijms20133226

Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M., et al. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* 286 (5443), 1321–1326. doi: 10.1126/science.286.5443.1321

Ibraheem, O., Botha, C. E., and Bradley, G. (2010). In silico analysis of cis-acting regulatory elements in 5' regulatory regions of sucrose transporter gene families in rice (Oryza sativa japonica) and *Arabidopsis thaliana*. *Comput. Biol. Chem.* 34 (5-6), 268–283. doi: 10.1016/j.compbiolchem.2010.09.003

Ischoff, E. B., Barale, J. C., and Mercereaupuijalon, O. (2002). Three multigene families in plasmodium parasites: facts and questions. *Int. J. Parasitol.* 32 (11), 1323–1344. doi: 10.1016/s0020-7519(02)00111-x

Jeon, E. H., Pak, J. H., Kim, M. J., Kim, H. J., Shin, S. H., Lee, J. H., et al. (2012). Ectopic expression of ubiquitin-conjugating enzyme gene from wild rice, OgUBC1, confers resistance against UV-b radiation and botrytis infection in *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* 427 (2), 309–314. doi: 10.1016/j.bbrc.2012.09.048

Jerome Jeyakumar, J. M., Ali, A., Wang, W. M., and Thiruvengadam, M. (2020). Characterizing the role of the miR156-SPL network in plant development and stress response. *Plants (Basel)* 9 (9). doi: 10.3390/plants9091206

Jia, L., Zhao, Q., and Chen, S. (2019). Evolution and expression analysis of the sorghum ubiquitin-conjugating enzyme family. *Funct. Plant Biol.* 46 (3), 236–247. doi: 10.1071/FP18184

Jones-Rhoades, M. W., and Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* 14 (6), 787–799. doi: 10.1016/j.molcel.2004.05.027

Jue, D., Sang, X., Liu, L., Shu, B., Wang, Y., Xie, J., et al. (2018). The ubiquitin-conjugating enzyme gene family in longan (Dimocarpus longan lour.): Genome-wide identification and gene expression during flower induction and abiotic stress responses. *Molecules* 23 (3). doi: 10.3390/molecules23030662

Jue, D., Sang, X., Lu, S., Dong, C., Zhao, Q., Chen, H., et al. (2015). Genome-wide identification, phylogenetic and expression analyses of the ubiquitin-conjugating enzyme gene family in maize. *PloS One* 10 (11), e0143488. doi: 10.1371/journal.pone.0143488

Jue, D., Sang, X., Shu, B., Liu, L., Wang, Y., Jia, Z., et al. (2017). Characterization and expression analysis of genes encoding ubiquitin conjugating domain-containing enzymes in carica papaya. *PloS One* 12 (2), e0171357. doi: 10.1371/journal.pone.0171357

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42 (4), 348–354. doi: 10.1038/ng.548

Kim, D. Y., Scalf, M., Smith, L. M., and Vierstra, R. D. (2013). Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in *Arabidopsis*. *Plant Cell* 25 (5), 1523–1540. doi: 10.1105/tpc.112.108613

Koch, M., Bishop, J., and Mitchell-Olds, T. (1999). Molecular systematics and evolution of arabidopsis and *Arabis*. *Plant Biol. (Stuttg)* 1 (05), 529–537. doi: 10.1111/j.1438-8677.1999.tb00779.x

Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17 (10), 1483–1498. doi: 10.1093/oxfordjournals.molbev.a026248

Koch, M., Haubold, B., and Mitchell-Olds, T. (2001). Molecular systematics of the brassicaceae: evidence from coding plastidic matK and nuclear chs sequences. *Am. J. Bot.* 88 (3), 534–544. doi: 10.2307/2657117

Kong, X., Lv, W., Jiang, S., Zhang, D., Cai, G., Pan, J., et al. (2013). Genome-wide identification and expression analysis of calcium-dependent protein kinase in maize. *BMC Genomics* 14 (433). doi: 10.1186/1471-2164-14-433

Kraft, E., Stone, S. L., Ma, L., Su, N., Gao, Y., Lau, O. S., et al. (2005). Genome analysis and functional characterization of the E2 and RING-type E3 ligase ubiquitination enzymes of arabidopsis. *Plant Physiol.* 139 (4), 1597–1611. doi: 10.1104/pp.105.067983

Kravic, B., Behrends, C., and Meyer, H. (2020). Regulation of lysosome integrity and lysophagy by the ubiquitin-conjugating enzyme UBE2QL1. *Autophagy* 16 (1), 179–180. doi: 10.1080/15548627.2019.1687217

Laloum, T., De Mita, S., Gamas, P., Baudin, M., and Niebel, A. (2013). CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci.* 18 (3), 157–166. doi: 10.1016/j.tplants.2012.07.004

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Peer, Y., et al. (2002). PlantCARE, a database of plant cis -acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30 (1), 325–327. doi: 10.1093/nar/30.1.325

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44 (W1), W242–W245. doi: 10.1093/nar/gkw290

Li, J., Han, G., Sun, C., and Sui, N. (2019). Research advances of MYB transcription factors in plant stress resistance and breeding. *Plant Signal Behav.* 14 (8), 1613131. doi: 10.1080/15592324.2019.1613131

Li, W., and Schmidt, W. (2010). A lysine-63-linked ubiquitin chain-forming conjugase, UBC13, promotes the developmental responses to iron deficiency in *Arabidopsis* roots. *Plant J.* 62 (2), 330–343. doi: 10.1111/j.1365-313X.2010.04150.x

Li, X., Zhu, L., Wu, Z., Chen, J., Wang, T., Zhang, X., et al. (2022). Classification and expression profile of the U-box E3 ubiquitin ligase enzyme gene family in maize (Zea mays l.). *Plants (Basel)* 11 (19). doi: 10.3390/plants11192459

Lin, Y., Hwang, W. C., and Basavappa, R. (2002). Structural and functional analysis of the human mitotic-specific ubiquitin-conjugating enzyme, UbcH10. *J. Biol. Chem.* 277 (24), 21913–21921. doi: 10.1074/jbc.M109398200

Lin, W. Y., Lin, S. I., and Chiou, T. J. (2009). Molecular regulators of phosphate homeostasis in plants. *J. Exp. Bot.* 60 (5), 1427–1438. doi: 10.1093/jxb/ern303

Lin, Z., Nie, H., Zhang, Y., Yin, Z., and Yan, X. (2022). Genome-wide identification and analysis of HECT E3 ubiquitin ligase gene family in *Ruditapes philippinarum* and their involvement in the response to heat stress and vibrio anguillarum infection. *Comp. Biochem. Physiol. Part D Genomics Proteomics* 43, 101012. doi: 10.1016/j.cbd.2022.101012

Linden, K. J., and Callis, J. (2020). The ubiquitin system affects agronomic plant traits. *J. Biol. Chem.* 295 (40), 13940–13955. doi: 10.1074/jbc.REV120.011303

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A., et al. (2014). The brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5, 3930. doi: 10.1038/ncomms4930

Liu, W., Tang, X., Zhu, X., Qi, X., Zhang, N., and Si, H. (2019). Genome-wide identification and expression analysis of the E2 gene family in potato. *Mol. Biol. Rep.* 46 (1), 777–791. doi: 10.1007/s11033-018-4533-9

Liu, C., Xie, T., Chen, C., Luan, A., Long, J., Li, C., et al. (2017). Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (*Ananas comosus*). *BMC Genomics* 18 (1), 503. doi: 10.1186/s12864-017-3896-y

Liu, F., Zhu, C., Gao, P., Zheng, S., and Li, C. (2020). Ubiquitin-conjugating enzyme E2T regulates cell proliferation and migration in cholangiocarcinoma. *Anticancer Drugs* 31 (8), 836–846. doi: 10.1097/CAD.0000000000000955

Lu, C., Napier, J. A., Clemente, T. E., and Cahoon, E. B. (2011). New frontiers in oilseed biotechnology: meeting the global demand for vegetable oils for food, feed, biofuel, and industrial applications. *Curr. Opin. Biotechnol.* 22 (2), 252–259. doi: 10.1016/j.copbio.2010.11.006

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., et al. (2019). Whole-genome resequencing reveals brassica napus origin and genetic loci involved in its improvement. *Nat. Commun.* 10 (1), 1154. doi: 10.1038/s41467-019-09134-9

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290 (5494), 1151–1155. doi: 10.1126/science.290.5494.1151

Lysak, M. A., Koch, M. A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe brassiceae. *Genome Res.* 15 (4), 516–525. doi: 10.1101/gr.3531105

Ma, J. Q., Jian, H. J., Yang, B., Lu, K., Zhang, A. X., Liu, P., et al. (2017). Genome-wide analysis and expression profiling of the GRF gene family in oilseed rape (*Brassica napus* l.). *Gene* 620, 36–45. doi: 10.1016/j.gene.2017.03.030

Michelle, C., Vourc'h, P., Mignon, L., and Andres, C. R. (2009). What was the set of ubiquitin and ubiquitin-like conjugating enzymes in the eukaryote common ancestor? *J. Mol. Evol.* 68 (6), 616–628. doi: 10.1007/s00239-009-9225-6

Millyard, L., Lee, J., Zhang, C., Yates, G., and Sadanandom, A. (2016). The ubiquitin conjugating enzyme, TaU4 regulates wheat defence against the phytopathogen *Zymoseptoria tritici*. *Sci. Rep.* 6, 35683. doi: 10.1038/srep35683

Miura, K., Jin, J. B., and Hasegawa, P. M. (2007). Sumoylation, a post-translational regulatory process in plants. *Curr. Opin. Plant Biol.* 10 (5), 495–502. doi: 10.1016/j.pbi.2007.07.002

Miura, K., Sato, A., Ohta, M., and Furukawa, J. (2011). Increased tolerance to salt stress in the phosphate-accumulating *Arabidopsis* mutants siz1 and pho2. *Planta* 234 (6), 1191–1199. doi: 10.1007/s00425-011-1476-y

Mun, J. H., Kwon, S. J., Yang, T. J., Seol, Y. J., and Jin, M. (2009). Genome-wide comparative analysis of the brassica rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* 10 (10), R111. doi: 10.1186/gb-2009-10-10-r111

Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H., et al. (2003). Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis* rd29A gene in response to dehydration and high-salinity stresses. *Plant J.* 34 (2), 137–148. doi: 10.1046/j.1365-313x.2003.01708.x

Nekrutenko, A. (2002). The K A/K s ratio test for assessing the protein-coding potential of genomic regions: An empirical and simulation study. *Genome Res.* 12 (1), 198–202. doi: 10.1101/gr.200901

Nuruzzaman, M., Sharoni, A. M., Satoh, K., Kumar, A., Leung, H., and Kikuchi, S. (2014). Comparative transcriptome profiles of the WRKY gene family under control, hormone-treated, and drought conditions in near-isogenic rice lines reveal differential, tissue specific gene activation. *J. Plant Physiol.* 171 (1), 2–13. doi: 10.1016/j.jplph.2013.09.010

Oudelaar, A. M., and Higgs, D. R. (2021). The relationship between genome structure and function. *Nat. Rev. Genet.* 22 (3), 154–168. doi: 10.1038/s41576-020-00303-x

Ozkan, E., Yu, H., and Deisenhofer, J. (2005). Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proc. Natl. Acad. Sci. U.S.A.* 102 (52), 18890–18895. doi: 10.1073/pnas.0509418102

Papaleo, E., Casiraghi, N., Arrigoni, A., Vanoni, M., Coccetti, P., and De Gioia, L. (2012). Loop 7 of E2 enzymes: an ancestral conserved functional motif involved in the E2-mediated steps of the ubiquitination cascade. *PloS One* 7 (7), e40786. doi: 10.1371/journal.pone.0040786

Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)* 9 (2), 88. doi: 10.3390/genes9020088

Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* 70, 503–533. doi: 10.1146/annurev.biochem.70.1.503

Piñeiro, M., and Jarillo, J. A. (2013). Ubiquitination in the control of photoperiodic flowering. *Plant Sci.* 198, 98–109. doi: 10.1016/j.plantsci.2012.10.005

Poyurovsky, M. V., Priest, C., Kentsis, A., Borden, K. L., Pan, Z. Q., Pavletich, N., et al. (2007). The Mdm2 RING domain c-terminus is required for supramolecular assembly and ubiquitin ligase activity. *EMBO J.* 26 (1), 90–101. doi: 10.1038/sj.emboj.7601465

Prince, V. E., and Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* 3 (11), 827–837. doi: 10.1038/nrg928

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* 20 (1), 38. doi: 10.1186/s13059-019-1650-2

Ramadan, A., Nemoto, K., Seki, M., Shinozaki, K., Takeda, H., Takahashi, H., et al. (2015). Wheat germ-based protein libraries for the functional characterisation of the *Arabidopsis* E2 ubiquitin conjugating enzymes and the RING-type E3 ubiquitin ligase enzymes. *BMC Plant Biol.* 15, 275. doi: 10.1186/s12870-015-0660-9

Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29 (1), 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Rana, D., van den Boogaart, T., O'Neill, C. M., Hynes, L., Bent, E., Macpherson, L., et al. (2004). Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. *Plant J.* 40 (5), 725–733. doi: 10.1111/j.1365-313X.2004.02244.x

Sadanandom, A., Bailey, M., Ewan, R., Lee, J., and Nelis, S. (2012). The ubiquitin-proteasome system: central modifier of plant signalling. *New Phytol.* 196 (1), 13–28. doi: 10.1111/j.1469-8137.2012.04266.x

Schiessl, S., Huettel, B., Kuehn, D., Reinhardt, R., and Snowdon, R. (2017). Post-polyploidisation morphotype diversification associates with gene copy number variation. *Sci. Rep.* 7, 41845. doi: 10.1038/srep41845

Schumacher, F. R., Wilson, G., and Day, C. L. (2013). The n-terminal extension of UBE2E ubiquitin-conjugating enzymes limits chain assembly. *J. Mol. Biol.* 425 (22), 4099–4111. doi: 10.1016/j.jmb.2013.06.039

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Sharma, B., and Bhatt, T. K. (2017). Genome-wide identification and expression analysis of E2 ubiquitin-conjugating enzymes in tomato. *Sci. Rep.* 7 (1), 8613. doi: 10.1038/s41598-017-09121-4

Sharma, B., and Taganna, J. (2020). Genome-wide analysis of the U-box E3 ubiquitin ligase enzyme gene family in tomato. *Sci. Rep.* 10 (1), 9581. doi: 10.1038/s41598-020-66553-1

Silver, E. T., Gwozd, T. J., Ptak, C., Goebl, M., and Ellison, M. J. (1992). A chimeric ubiquitin conjugating enzyme that combines the cell cycle properties of CDC34 (UBC3) and the DNA repair properties of RAD6 (UBC2): implications for the structure, function and evolution of the E2s. *EMBO J.* 11 (8), 3091–3098. doi: 10.1002/j.1460-2075.1992.tb05381.x

Smalle, J., and Vierstra, R. D. (2004). The ubiquitin 26S proteasome proteolytic pathway. *Annu. Rev. Plant Biol.* 55, 555–590. doi: 10.1146/annurev.arplant.55.031903.141801

Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and their regulatory roles in plant-environment interactions. *Annu. Rev. Plant Biol.* 70, 489–525. doi: 10.1146/annurev-arplant-050718-100334

Sun, L., and Chen, Z. J. (2004). The novel functions of ubiquitination in signaling. *Curr. Opin. Cell Biol.* 16 (2), 119–126. doi: 10.1016/j.ceb.2004.02.005

Sun, F., Fan, G., Hu, Q., Zhou, Y., Guan, M., Tong, C., et al. (2017). The high-quality genome of *Brassica napus* cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. *Plant J.* 92 (3), 452–468. doi: 10.1111/tpj.13669

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28 (10), 2731–2739. doi: 10.1093/molbev/msr121

Tang, M. (2019). *Population genome variations and subgenome asymmetry in brassica napus l* (Huazhong: Huazhong Agricultural University).

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7 (3), 562–578. doi: 10.1038/nprot.2012.016

van Wijk, S. J., and Timmers, H. T. (2010). The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. *FASEB J.* 24 (4), 981–993. doi: 10.1096/fj.09-136259

Vierstra, R. D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* 10 (6), 385–397. doi: 10.1038/nrm2688

Wahid, S., Xie, M., Sarfraz, S., Liu, J., Zhao, C., Bai, Z., et al. (2022). Genome-wide identification and analysis of Ariadne gene family reveal its genetic effects on agronomic traits of brassica napus. *Int. J. Mol. Sci.* 23 (11). doi: 10.3390/ijms23116265

Wang, S., Cao, L., and Wang, H. (2016). *Arabidopsis* ubiquitin-conjugating enzyme UBC22 is required for female gametophyte development and likely involved in Lys11-linked ubiquitination. *J. Exp. Bot.* 67 (11), 3277–3288. doi: 10.1093/jxb/erw142

Wang, T., Hu, J., Ma, X., Li, C., Yang, Q., Feng, S., et al. (2020). Identification, evolution and expression analyses of whole genome-wide TLP gene family in *Brassica napus*. *BMC Genomics* 21 (1), 264. doi: 10.1186/s12864-020-6678-x

Wang, W., Jiang, W., Liu, J., Li, Y., Gai, J., and Li, Y. (2017). Genome-wide characterization of the aldehyde dehydrogenase gene superfamily in soybean and its potential role in drought stress response. *BMC Genomics* 18 (1), 518. doi: 10.1186/s12864-017-3908-y

Wang, C., Song, B., Dai, Y., Zhang, S., and Huang, X. (2021). Genome-wide identification and functional analysis of U-box E3 ubiquitin ligases gene family related to drought stress response in Chinese white pear (Pyrus bretschneideri). *BMC Plant Biol.* 21 (1), 235. doi: 10.1186/s12870-021-03024-3

Wang, Y., Wang, W., Cai, J., Zhang, Y., Qin, G., and Tian, S. (2014). Tomato nuclear proteome reveals the involvement of specific E2 ubiquitin-conjugating enzymes in fruit ripening. *Genome Biol.* 15 (12), 548. doi: 10.1186/s13059-014-0548-2

Wang, Y., Wang, Q., Zhao, Y., Han, G., and Zhu, S. (2015). Systematic analysis of maize class III peroxidase gene family reveals a conserved subfamily involved in abiotic stress response. *Gene* 566 (1), 95–108. doi: 10.1016/j.gene.2015.04.041

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinf.* 8 (1), 77–80. doi: 10.1016/s1672-0229(10)60008-3

Wang, X., Zheng, M., Liu, H., Zhang, L., and Hua, W. (2020). Fine-mapping and transcriptome analysis of a candidate gene controlling plant height in *Brassica napus* l. *Biotechnol. Biofuels* 13 (42). doi: 10.1186/s13068-020-01687-y

Welchman, R. L., Gordon, C., and Mayer, R. J. (2005). Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat. Rev. Mol. Cell Biol.* 6 (8), 599–609. doi: 10.1038/nrm1700

Wen, R., Newton, L., Li, G., Wang, H., and Xiao, W. (2006). Arabidopsis thaliana UBC13: implication of error-free DNA damage tolerance and Lys63-linked polyubiquitylation in plants. *Plant Mol. Biol.* 61 (1-2), 241–253. doi: 10.1007/s11103-006-0007-x

Wen, R., Torres-Acosta, J. A., Pastushok, L., Lai, X., Pelzer, L., Wang, H., et al. (2008). Arabidopsis UEV1D promotes lysine-63-linked polyubiquitination and is involved in DNA damage response. *Plant Cell* 20 (1), 213–227. doi: 10.1105/tpc.107.051862

Wenzel, D. M., Stoll, K. E., and Klevit, R. E. (2011). E2s: structurally economical and functionally replete. *Biochem. J.* 433 (1), 31–42. doi: 10.1042/bj20100985

Wu, Y., Ke, Y., Wen, J., Guo, P., Ran, F., Wang, M., et al. (2018). Evolution and expression analyses of the MADS-box gene family in *Brassica napus*. *PloS One* 13 (7), e0200762. doi: 10.1371/journal.pone.0200762

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., et al. (2019). Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Mol. Plant* 12 (1), 30–43. doi: 10.1016/j.molp.2018.11.007

Xie, M., Zuo, R., Bai, Z., Yang, L., Zhao, C., Gao, F., et al. (2022). Genome-wide characterization of Serine/Arginine-rich gene family and its genetic effects on agronomic traits of *Brassica napus*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.829668

Xu, L., Ménard, R., Berr, A., Fuchs, J., Cognat, V., Meyer, D., et al. (2009). The E2 ubiquitin-conjugating enzymes, AtUBC1 and AtUBC2, play redundant roles and are involved in activation of FLC expression and repression of flowering in *Arabidopsis thaliana*. *Plant J.* 57 (2), 279–288. doi: 10.1111/j.1365-313X.2008.03684.x

Yao, S., Liang, F., Gill, R. A., Huang, J., Cheng, X., Liu, Y., et al. (2020). A global survey of the transcriptome of allopolyploid *Brassica napus* based on single-molecule long-read isoform sequencing and illumina-based RNA sequencing data. *Plant J.* 103 (2), 843–857. doi: 10.1111/tpj.14754

Ye, Y., and Rape, M. (2009). Building ubiquitin chains: E2 enzymes at work. *Nat. Rev. Mol. Cell Biol.* 10 (11), 755–764. doi: 10.1038/nrm2780

Yu, P., and Hua, Z. (2022). The ubiquitin-26S proteasome system and autophagy relay proteome homeostasis regulation during silique development. *Plant J.* 111 (5), 1324–1339. doi: 10.1111/tpj.15891

Yu, Y., Jia, T., and Chen, X. (2017). The 'how' and 'where' of plant microRNAs. *New Phytol.* 216 (4), 1002–1017. doi: 10.1111/nph.14834

Yunpeng, C., Yahui, H., Dandan, M., Dahui, L., Qing, J., Yi, L., et al. (2016). Structural, evolutionary, and functional analysis of the class III peroxidase gene family in Chinese pear (Pyrus bretschneideri). *Front. Plant Sci.* 7, 1874. doi: 10.3389/fpls.2016.01874

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., et al. (2018). Improved *Brassica rapa* reference genome by single-molecule sequencing and

chromosome conformation capture technologies. *Hortic. Res.* 5, 50. doi: 10.1038/s41438-018-0071-9

Zheng, B., Chen, X., and McCormick, S. (2011). The anaphase-promoting complex is a dual integrator that regulates both MicroRNA-mediated transcriptional regulation of cyclin B1 and degradation of cyclin B1 during *Arabidopsis* male gametophyte development. *Plant Cell* 23 (3), 1033–1046. doi: 10.1105/tpc.111.083980

Zhou, X., Wu, X., Li, T., Jia, M., Liu, X., Zou, Y., et al. (2018). Identification, characterization, and expression analysis of auxin response factor (ARF) gene family in *Brachypodium distachyon*. *Funct. Integr. Genomics* 18 (6), 709–724. doi: 10.1007/s10142-018-0622-z

Zhu, W., Guo, Y., Chen, Y., Wu, D., and Jiang, L. (2020). Genome-wide identification, phylogenetic and expression pattern analysis of GATA family genes in *Brassica napus*. *BMC Plant Biol.* 20 (1), 543. doi: 10.1186/s12870-020-02752-2

Zolman, B. K., Monroe-Augustus, M., Silva, I. D., and Bartel, B. (2005). Identification and functional characterization of arabidopsis PEROXIN4 and the interacting protein PEROXIN22. *Plant Cell* 17 (12), 3422–3435. doi: 10.1105/tpc.105.035691

Zou, C., Sun, K., Mackaluso, J. D., Seddon, A. E., Jin, R., Thomashow, M. F., et al. (2011). Cis-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 108 (36), 14992–14997. doi: 10.1073/pnas.1103202108

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 March 2023
DOI 10.3389/fpls.2023.1157836

OPEN ACCESS

EDITED BY
Hongbo Chao,
Zhengzhou University, China

REVIEWED BY
Zhen Huang,
Northwest A&F University, China
Jing Wen,
Huazhong Agricultural University, China

*CORRESPONDENCE
Yourong Chai
chaiyourong@163.com

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 03 February 2023
ACCEPTED 07 March 2023
PUBLISHED 29 March 2023

CITATION
Lin N, Wang M, Jiang J, Zhou Q, Yin J, Li J, Lian J, Xue Y and Chai Y (2023) Downregulation of *Brassica napus MYB69* (*BnMYB69*) increases biomass growth and disease susceptibility *via* remodeling phytohormone, chlorophyll, shikimate and lignin levels.
Front. Plant Sci. 14:1157836.
doi: 10.3389/fpls.2023.1157836

Downregulation of *Brassica napus MYB69* (*BnMYB69*) increases biomass growth and disease susceptibility *via* remodeling phytohormone, chlorophyll, shikimate and lignin levels

Na Lin[1], Mu Wang[1], Jiayi Jiang[1], Qinyuan Zhou[1], Jiaming Yin[1], Jiana Li[1,2], Jianping Lian[1], Yufei Xue[1] and Yourong Chai[1,2]*

[1]Integrative Science Center of Germplasm Creation in Western China (CHONGQING) Science City and Southwest University, College of Agronomy and Biotechnology, Southwest University, Chongqing, China, [2]Engineering Research Center of South Upland Agriculture, Ministry of Education, Academy of Agricultural Science, Southwest University, Chongqing, China

MYB transcription factors are major actors regulating plant development and adaptability. *Brassica napus* is a staple oil crop and is hampered by lodging and diseases. Here, four *B. napus MYB69* (*BnMYB69s*) genes were cloned and functionally characterized. They were dominantly expressed in stems during lignification. *BnMYB69* RNA interference (*BnMYB69i*) plants showed considerable changes in morphology, anatomy, metabolism and gene expression. Stem diameter, leaves, roots and total biomass were distinctly larger, but plant height was significantly reduced. Contents of lignin, cellulose and protopectin in stems were significantly reduced, accompanied with decrease in bending resistance and *Sclerotinia sclerotiorum* resistance. Anatomical detection observed perturbation in vascular and fiber differentiation in stems, but promotion in parenchyma growth, accompanied with changes in cell size and cell number. In shoots, contents of IAA, shikimates and proanthocyanidin were reduced, while contents of ABA, BL and leaf chlorophyll were increased. qRT-PCR revealed changes in multiple pathways of primary and secondary metabolisms. IAA treatment could recover many phenotypes and metabolisms of *BnMYB69i* plants. However, roots showed trends opposite to shoots in most cases, and *BnMYB69i* phenotypes were light-sensitive. Conclusively, BnMYB69s might be light-regulated positive regulators of shikimates-related metabolisms, and exert profound influences on various internal and external plant traits.

KEYWORDS

***Brassica napus*, MYB69, shikimates, lignin, *Sclerotinia sclerotiorum*, IAA, chlorophyll, biomass**

Introduction

MYB is a large family of transcription factors (TFs) (Feller et al., 2011). It has been a hotspot in the study of the plant transcription factor function because of its large number of genes and different functions. The first identified MYB TF in plants was COLORED1 (C1), which is involved in the biosynthesis of anthocyanins in *Zea mays* (Paz-Ares et al., 1987). MYB TFs contain a highly conserved DNA binding domain (DBD) located at the N-terminus (Baldoni et al., 2015; Wei et al., 2017; Zhuang et al., 2021). The MYB domain is composed of 1-4 imperfect repeats (R1-R4) each containing approximately 50-53 amino acids (aa). Each repeat consists of three α-helices, of which the second and third ones form a helix-turn-helix (HTH) motif (Baldoni et al., 2015). The HTH intercalates with the major groove of DNA. MYB TFs have three spaced Trp residues forming a hydrophobic core that stabilizes the structure (Ogata et al., 1995). The C-terminus is responsible for the distinct regulatory activities of MYB TFs (Jin and Martin, 1999).

MYB TFs are divided into four groups, 1R-MYB, R2R3-MYB, 3R-MYB, and 4R-MYB, based on the number of repeats (Dubos et al., 2010). R2R3-MYB transcription factors (TFs) have been shown to play important roles in plants, including cell fate and identity, developmental processes, and responses to biotic and abiotic stresses (Hajiebrahimi et al., 2017). Although members of the *MYB* superfamily have been annotated in *Arabidopsis thaliana* and many other plants, most of them have not been functionally characterized.

In the past decade, *R2R3-MYB* genes have been extensively studied. The R2R3-MYB proteins play important roles in diverse biological processes including growth and development, primary and secondary metabolism such as flavonoid and anthocyanin biosynthesis as well as abiotic and biotic stress responses (Yao et al., 2020). Plant development and secondary metabolisms are related to cell differentiation including the change of cell size, cell number and cell wall. It was well known that the plant cell wall is a dynamic barrier against pathogens invasion (Underwood, 2012). The cell walls of vascular bundles deposit a large amount of cellulose, lignin and pectin, which are closely related to the support and resistance of plants. Lignin plays an important role in plant growth and development, lodging resistance and structural support (Jiang et al., 2020). The deposition of unpolymerized lignin monomers is considered to present a physical barrier that prevents infection in the interaction between plant and pathogen, and its metabolism is also actively involved in plant lodging resistance and in response to various environmental stresses (Jiang et al., 2020).

Many studies have shown that MYB69, which is from subgroup S21, is a SCW (secondary cell wall) activator (Kranz et al., 1998; Zhong et al., 2007; Dubos et al., 2010; Zhu et al., 2017). It has been demonstrated by transcriptional activation analysis that MYB69 and other transcription factors are able to activate the expression of the *HIS3* and *b-GAL* reporter genes in yeast, confirming that they are indeed transcriptional activators (Zhong et al., 2008). AtMYB69 is a positive regulator dedicated to cell wall thickening of fiber cells in *A. thaliana* (Dubos et al., 2010). Zhao and Bartley (2014) have shown that AtMYB69 is an activator whose dominant repression reduces SCW thickening of both interfascicular fibers and xylem fibers in stems. These studies indicate that *MYB69* may be related to the growth of plant cell. Studies have also shown by cluster analysis that *MYB105, MYB110, MYB117*, *MYB56*, *MYB52*, *MYB54* and *MYB69* belong to the same subgroup, so it is probable that their functions are similar (Kranz et al., 1998; Dubos et al., 2010; Zhu et al., 2017). *MYB20*, *MYB63*, *MYB69*, *SND2* and *SND3* are also hypothesized to be part of the SND1/NST1 transcriptional network regulating secondary wall biosynthesis (McCarthy et al., 2010). Xu et al. (2021) reported that *FvMYB3, FvMYB9, FvMYB11, FvMYB22, FvMYB64*, and *FvMYB105* mostly expressed at green stage of fruit development in woodland strawberry, aligned with proanthocyanidins accumulation. Wang et al. (2022) found that both *35S::MYB10* and *MYB110* can upregulate anthocyanin biosynthesis in *Actinidia chinensis* fruit. Hong et al. (2021) have identified MYB117 as a negative regulator in controlling flowering time through regulating the expression of *Flowering Locus T* in *Arabidopsis*. Other researchers showed that *SiMYB56* enhanced ABA synthesis under drought conditions and activated the ABA signaling pathway, contributing to the enhanced drought tolerance of transgenic rice (Xu et al., 2020). It has also been demonstrated that dominant repression of *MYB69* and *SND2*, *SND3*, *MYB103*, *MYB85*, *MYB52*, *MYB54*, and *KNAT7* causes a severe reduction in secondary wall thickening of both interfascicular fibers and xylary fibers in inflorescence stems (Zhong et al., 2008). It was reported that ERF4 and MYB52 transcription factors played antagonistic roles in regulating homogalacturonan de-methylesterification in *Arabidopsis* seed coat mucilage (Ding et al., 2021). For MYB46 downstream transcription factors, *MYB43, MYB52, MYB54, MYB58*, *MYB63* and *KNAT7* are direct targets of MYB46 (Zhong and Ye, 2012), which was a master transcriptional regulator of secondary wall biosynthesis (Kim et al., 2012). And it was reported that MYB46 is negatively regulated by MPK6 during salt stress (Im et al., 2021). These studies have revealed that the function of *MYB69* was not only related to regulating secondary wall biosynthesis but also probably related with some metabolisms and resistances in plant.

B. napus is an important oil crop worldwide and is negatively impacted by a variety of biotic and abiotic stresses, particularly lodging and *S. sclerotiorum* stem rot, which seriously reduce the yield and affect the quality of *B. napus* seeds. Lodging is related to stem and root lodging (Li and Qi, 2017; Zhang et al., 2021). The flexibility of stems and roots is related to secondary wall thickening and lignin accumulation. Studying the function of *MYB69* in *B. napus* can both elucidate whether it contributes to lodging resistance and antimicrobial resistance in *B. napus* and systematically analyze how *MYB69* regulates plant cell growth and development.

In this work, downregulation of *BnMYB69* causes an increase in the biomass and susceptibility to *S. sclerotiorum*, and related biochemical, physiological and gene expression characterization provides insight into *BnMYB69* function in *B. napus*.

Materials and methods

Plant materials

The material of gene cloning and expression characterization was cv. Zhongyou 821 (ZY821) of *B. napus*. The explant material of

genetic transformation is cv. Zhongshuang 10 (ZS10) of *B. napus*. These plants were from growth chambers, green houses or isolated field farms at Southwest University and cultivated by standard agronomic management. *Nicotiana benthamiana* plants for protein localization study were cultured in room at 28°C under a 16 h/8 h light/dark cycle.

Fungal strains

The strain of *S. sclerotiorum* in this experiment was provided by Dr. Jiaqin Mei at Southwest University. *S. sclerotiorum* was grown up on potato media (PDA) at 22°C with 85% humidity.

Cloning of *BnMYB69*s

Total RNA was isolated from ZY821 organs using EASYspin plant RNA rapid extraction kit (Biomed, China). Total DNA was extracted from ZY821 leaves using traditional CTAB method. The quality and quantity of RNA and DNA were detected by agarose gel electrophoresis and NanoDrop 2000c (Thermo Fisher, USA).

In-silico cloning of *Brassica napus MYB69*s was performed on NCBI (https://blast.ncbi.nlm.nih.gov) using *A. thaliana MYB69* (*AtMYB69*) mRNA as Query in BLASTn against *B. napus* (taxid:3708) nr/nt, refseq_genomes, refseq_rna, est, and TSA databases. Sequences orthologous to *AtMYB69* were downloaded and multi-aligned on Vector NTI Advance 11.51, and primers for RACE (rapid-amplification of cDNA ends) cloning were designed (Table S1).

Using the SMARTer RACE Amplification Kit (Clontech, USA), 1 μg of total RNA mixture from various organs of ZY821 was used for synthesis of first-strand total cDNA. The primer pairs RBnMYB69-51 + LUPM and RBnMYB69-52 + NUP were used for the primary and the nested PCRs in 5'-RACE of *BnMYB69*s, while primer pairs FBnMYB69-31 + LUPM and FBnMYB69-32 + NUP were used for primary and nested PCRs in 3'-RACE of *BnMYB69*s. Based on sequencing results of 5'- and 3'-RACE colonies, primers FBnMYB69-1 + RBnMYB69-1, FBnMYB69-2 + RBnMYB69-2, FBnMYB69-3 + RBnMYB69-3 and FBnMYB69-4 + RBnMYB69-4 were used to amplify full-length cDNAs and gDNAs of *BnMYB69-1*, *BnMYB69-2*, *BnMYB69-3* and *BnMYB69-4*, respectively. Electrophoresis of PCR products, gel recovery, TA cloning, *Escherichia coli* transformation, colony culture and other experiments abided by routine methods. Bioinformatics Analyses of *BnMYB69*s genes and proteins were performed on local Vector NTI Advance 11.51 and websites of NCBI, Expasy, Softberry, etc. (See Supplementary Tables and Figures for more information).

Quantitative RT-PCR (qRT-PCR) analysis

The total RNA of each organ or tissue sample from material was reverse transcribed by the PrimeScript 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China) to obtain the corresponding first-strand total cDNA. All cDNA samples were diluted 20-fold using sterile water for qRT-PCR. The overall and member-specific expression levels of *B. napus* genes were detected by Gene-specific primers for qRT-PCR, respectively.

Primers are listed in Table S1, and the *25SrRNA* gene was used for the internal standard. The CFX Connect Real-Time PCR Detection System (Bio-Rad, Berkeley, CA, USA) with FastStart Universal SYBR Green Master reagents (Roche, Basel, Switzerland) was performed for qRT–PCR analysis. The program was that firstly, the sample was put in 95°C 10 min, followed by 45 cycles of amplification (95°C 10 s, 58-64°C 30 s). Secondly, the temperature was raised from 65 to 95°C, and the specificity of the amplification was confirmed by the melting curve. The detection was repeated three times. Thirdly, CFX Manager 3.1 (Bio-Rad, Berkeley, CA, USA) by the $2^{-\triangle\triangle CT}$ method analyzed all data.

Subcellular localization

The coding sequence of *BnMYB69-1*(*BnaA01.MYB69*) was amplified using the primer pair FBnMYB69SL + RBnMYB69SL (Table S1), and it was cloned and designated *BnMYB69SL*. Then *BnMYB69SL* was subcloned into pEGAD by *Eco*RI+*Bam*HI double digestion, and enhanced green fluorescent protein (EGFP) was joined into the N-terminus to form the subcellular localization expression vector pEGAD-*BnMYB69SL* (Figure S4A). The vector was introduced in *Agrobacteriu tumefaciens* strain LBA4404 and transformed into tobacco leaves. Then the plant was cultured at 28°C for two days, stained its leaves with nuclear dye (DAPI) and photographed with a confocal microscope (LSM 800, ZEISS, Jena, Germany).

RNAi vector construction and genetic transformation

Using the primer pair FBnMYB69I + RBnMYB69I (Table S1), a 429 bp C-terminus coding region of cDNA for the *BnMYB69* family (*BnMYB69s)* was amplified from the cDNA library. This cDNA fragment had >86.7% identity match with all *BnMYB69* genes, and no identity match with other sequences which were found by BLASTing the entire *B. napus* genome (http://www.ncbi.nlm.nih.gov/genome/?term=Brassica+napus). An antisense fragment of *BnMYB69I* (*BnMYB69IA*) was subcloned into pFGC5941M using *Nco*I+*Aat*II double digestion to form the intermediate vector pFGC5941M-*BnMYB69IA* (Figure S4B). And a sense fragment of *BnMYB69I* (*BnMYB69IS*) was subcloned into pFGC5941M-*BnMYB69IA* using *Bam*HI+*Xba*I double digestion to form the RNAi vector pFGC5941M-*BnMYB69I*. *CaMV35S* was the promoter of the vector. pFGC5941M-*BnMYB69I* plasmids were introduced into LBA4404 *(A. tumefaciens)* by the freeze-thaw transformation method. The engineering strain harboring pFGC5941M-*BnMYB69I* was transformed into hypocotyl segments of ZS10 using the method described by Cardoza and Stewart (2003). The positive transgenic plants were detected by DNA electrophoresis after PCR amplification (Figure S4C).

Measurement of agronomic traits

Transgenic and nontransgenic *B. napus* (WT) plants were grown in a greenhouse, agronomic traits were investigated at the bolting stage, full flowering stage and pod setting stage, and dry matter weight, chlorophyll content, and 1000 seed weight per plant were measured at the harvest stage.

Assessment of resistance to *S. sclerotiorum*

The method of leaf inoculation was from Godoy et al. (1990). The third or fourth leaf from the top in each plant at the 9 to 12-leaf stage was excised with at least 6 leaves per different line. The excisions of these leaves were wrapped with wet paper to keep moisture, and were placed into porcelain plate paved with wet filter paper. 7-mm-diameter PDA discs containing *S. sclerotiorum* hyphae were inoculated at the center of the left and right abdomens in leaf. Then, the whole plate was covered with cling film and kept at 22°C. After 48 h, the long diameter (a) and the short diameter (b) of the lesion were measured. The lesion area was gained by this calculation formula: $S = (\pi \times a \times b)/4$.

The method of stem inoculation was from Mei et al. (2015). The stems were inoculated at the full-bloom stage with 6 plants per different line. Stem fragments (45 cm long) were cut off at 20 cm from the ground, and both ends also were wrapped with wet paper and covered with cling film. 7-mm-diameter PDA discs containing *S. sclerotiorum* hyphae were inoculated at two points distanced with 10 cm on the stem segment. Lesion lengths were measured after 72 h.

During *S. sclerotiorum* infection, leaves were taken at different time points for the determination of relevant biochemical components and key enzymes, and RNA was extracted for qRT –PCR (as described above).

Microscopic observation of tissue section

The methods of Mäule staining, phloroglucinol-HCl staining and fluorescent brightener dyeing of lignin were from the literature (Zhong et al., 2007; Yin et al., 2017). Cellulose staining with Fast Green FCF and pectin staining by the hydroxylamine method (Leagene, Beijing, China) were performed on the stems and roots of transgenic and WT plants. Stained sections were observed with a stereo light microscope (Olympus SZX2-FOA, Tokyo, Japan) and a fluorescence microscope (Nikon Eclipse E600 W, Tokyo, Japan). The tissues and organs were observed by scanning electron microscopy (Hitachi SU3500, Japan).

Determination of cell wall and primary metabolism composition

At the harvest stage, the middle part of mature stems was cut off and dried at 70°C to constant weight. Then, these stems were powdered with a grinder, and screened with a 100-mesh sieve. The cellulose, lignin and pectin contents from the cell walls of transgenic and WT plants were extracted and determined. The method of extraction and determination of cellulose and lignin was from Foster et al. (Foster et al., 1996). The method of extraction and quantification of pectin was from Blumenkrantz and Asboe-Hansen (1973). The absorbance of cellulose and lignin was determined by an Infinite M2000 Pro Microplate Reader (Tecan, Männedorf, Switzerland). The contents protopectin and soluble pectin were determined according to the instructions of the Protopectin and Soluble Pectin Content Determination kits from Shanghai Haling Biotechnology Co., Ltd.

At bolting stage, the middle-part samples of stems, leaves and roots were fetched and stored in a -80°C freezer. The frozen samples were ground into powder with liquid nitrogen to test the activity of CAT, MDA, SOD and POD, as well as the contents of chlorophyll, reducing sugars, and total amino acids. The extraction and determination of these biochemical components followed the kit instructions from Beijing Leagene Biotech Co., Ltd.

Determination of hormone components

At the bolting stage, the mid-region fragments of stems and leaves were sampled and stored in a -80°C refrigerator. Frozen samples were sent to Qingdao Kechuang Quality Inspection Company and Nanjing Zhongding Biotechnology Co., Ltd. to detect the changes in hormone components using gas chromatography or liquid chromatography, respectively.

Hormone treatment

Cultured in 1/3 X Hoagland nutrient solution, WT and *BnMYB69i* lines were treated with 100 μmol/L IAA, with non-addition as the experimental control (CK). They were cultured on Petri dishes (light 16 h/22°C; dark 8 h/18°C) and then in large pots after two weeks with 40 seedlings per pot. The water was changed every 10 days. There were 2 lines, 2 treatments and 3 repetitions. The morphological, physiological and biochemical indices were measured at the cotyledon stage (20 days), 3 true leaves stage (60 days) and 5 true leaves stage (120 days). Three plants were randomly selected from each line for the measurement of plant height, stem diameter and physiological and biochemical indices, and the average values of each parameter were calculated. Plant height was defined as the length from the cotyledonary node to the highest part of the entire plant (Gai, 2005); stem diameter of the stem base at the cotyledon mark was measured with a vernier caliper.

WT and transgenic plants were cultured in high light (8000 lx) and low light (4000 lx) incubators. They grew to the three-leaf stage under light for 16 h at 22°C and under darkness for 8 h at 18°C, and the chlorophyll content of the third true leaf was measured. The methods were performed according to instructions of the Beijing Leagene Biotechnology kit and the experimental guidance of Plant Physiology handbook (College of Agriculture/Department of Plant Physiology, Shanxi Agricultural University).

The transgenic and control plants were treated in the dark three times for 20 min each at different stages. The chlorophyll fluorescence kinetic parameters were determined with a Spectral Fluorescence Imager and Fluor Scan software.

A 0.1 g sample of *B. napus* leaves was frozen in liquid nitrogen and 300 μL of 1% hydrochloric acid methanol solution was added after proofing, mixed evenly and shaded at 4°C overnight. The next day, 200 μL distilled water was added along with 200μL chloroform, the mixture was vortexed for 20 s at room temperature and then centrifuged for 5 min at 17000 g. The extracted supernatant was added into a 96-well plate and determined using an Enzyme Labeling Instrument at 530 nm and 657 nm. The proanthocyanidin content (mg/g) was calculated as (A530-A657)/fresh weight (g).

The IAA content was determined according to the instructions of the plant indole-3-acetic acid ELISA kit (Nanjing Camilo Bioengineering Co., Ltd.), and the method of quantifying protopectin and amino acids was as described above.

Statistical analysis

In this study, at least three biological replications (n≥3) were performed for each experiment. All values are presented as the mean ± standard erro (SE). Asterisks indicate significant or extremely significant differences from WT (*0.01≤ $p < 0.05$, ** for $p < 0.01$) using a one-way ANOVA.

Results

Characterization of *MYB69s* from *B. napus*

Using mixture total RNA from various organs of ZY821 as start material, our RACE-PCR based cloning isolated four *MYB69* genes from *B. napus*, i.e., *BnMYB69-1 (BnaA01.MYB69)*, *BnMYB69-2 (BnaC01.MYB69)*, *BnMYB69-3 (BnaA08.MYB69)* and *BnMYB69-4 (BnaC03.MYB69)*, with respective gene lengths of 1110, 1418, 1296 and 1215 bp, mRNA lengths (not including polyA) of 1015, 1320, 1191 and 1110 bp, and coding region (open reading frame, ORF, including stop codon) lengths of 756, 753, 732, and 756 bp, respectively. *BnMYB69-1*, *BnMYB69-2*, *BnMYB69-3* and *BnMYB69-4* showed high homologies to *AtMYB69*, with gene level identities of 84.2%, 78.3%, 77.8% and 80.0%, mRNA level identities of 85.5%, 80.0%, 79.5% and 80.8%, and ORF level identities of 86.5%, 86.2%, 84.8 and 87.0%, respectively (Table S2). Only one intron exists in each of the four *BnMYB69* genes. 5'UTR length of *BnMYB69* genes is 28-132 bp, and their 3'UTR length is 127-494 bp. Every *BnMYB69* gene shows alternative transcription start sites (TSS) and alternative polyadenylation sites, thus multiple versions of mRNAs could be cloned with differed length of 5'UTRs and 3'UTRs. One feature is that in most cases the TSS base is a purine (A or G), and in most cases the base before the polyA tail is a pyrimidine (T or C). Another feature is that as *BnMYB69-1* could generate long 5'UTRs, its 3'UTRs are relatively shorter, while the other three *BnMYB69* genes have the opposite trend (Table S2).

The predicted BnMYB69-1, BnMYB69-2, BnMYB69-3, BnMYB69-4 proteins have a length of 251, 250, 243, 251 aa, a MW of 29.01, 28.90, 28.16, 29.04 kD, and a *pI* of 9.58, 9.70, 9.72, 9.56 implying that they are basic proteins. All of them have no signal peptide and no transmembrane helix, and are predicted to be located in the nuclear. Their predicted phosphorylation sites are 19-22 for S (serine), 8-9 for T (threonine), and 1-2 for Y (tyrosine). The N-terminal half of each BnMYB69 protein has two predicted HTH MYB-type domain regions at 14-65/64 and 66-120, on which there are two HTH motif DNA-binding regions at 42-63 and 93-116 (Table S3). The front HTH model belongs to the R2 subdomain and has three conserved tryptophan residues (W), which are separated by 18 or 19 amino acids, and the latter HTH model belongs to the R3 subdomain (Ogata et al., 1994). In the secondary structure of BnMYB69s, α-helix, extended strand, β-turn, and random coil occupy 23.11%-31.87%, 6.77%-12.75%, 4.38%-6.40%, and 54.40%-59.76%, respectively (Figure S1). SWISS-MODEL-predicted tertiary structures of BnMYB69s conform to conserved R2R3-MYB domain (Figure S2). Thus, the four *BnMYB69* genes cloned here most probably encode typical R2R3-MYB transcription factors.

In multiple alignment, BnMYB69 proteins have high similarities to AtMYB69 on the whole protein level. Though they show high similarities to MYB52, MYB54, MYB56, MYB105, MYB117, MYB110 and MYB89 of S21 in the N-terminal half, the similarities in the C-terminal half are distinctly lower with conservation only at a C-terminal motif and several dispersed residues (Figure 1A). Phylogenetic analysis showed that the distances among BnMYB69 proteins are small, and they clustered tightly with AtMYB69. BnMYB69-1 and BnMYB69-2 are close, while BnMYB69-3 and BnMYB69-4 are close, but these two sister pairs are a little far from each other. In the tree, the distances of BnMYB69s to other S21 AtMYB proteins are much larger than to AtMYB69 (Figure 1B). These results clearly show that all *BnMYB69* genes isolated here belong to S21 MYBs and are orthologous genes of *AtMYB69*.

To confirm the subcellular localization of BnMYB69s, experimental validation was carried out. Consistent with the bioinformatics prediction and its predicted role as a transcription factor, the expression of enhanced green fluorescent protein (EGFP)-labeled BnMYB69-1 and blue fluorescent nucleic acid dye (DAPI) coincided in tobacco leaves, showing that they were both localized in the nucleus (Figure 1C).

In qRT-PCR results, overall *BnMYB69* expression could be detected in all organs of *B. napus* but varied greatly among them. It was highly expressed in stems, moderately expressed in pod peels, leaves and roots, and lowly expressed in seeds, flowers, buds, cotyledons, and hypocotyls (Figure 1D). At bolting stage, different stem sections were detected for *BnMYB69* expression, which showed an increase trend in parallel with the stem vascularization/lignification process, with little expression in growth cone tip and highest expression in semi-lignified main-stem. However, its expression in the completely lignified main-stem dropped down (Figure 1E). We also detected the organ-specificity of the four

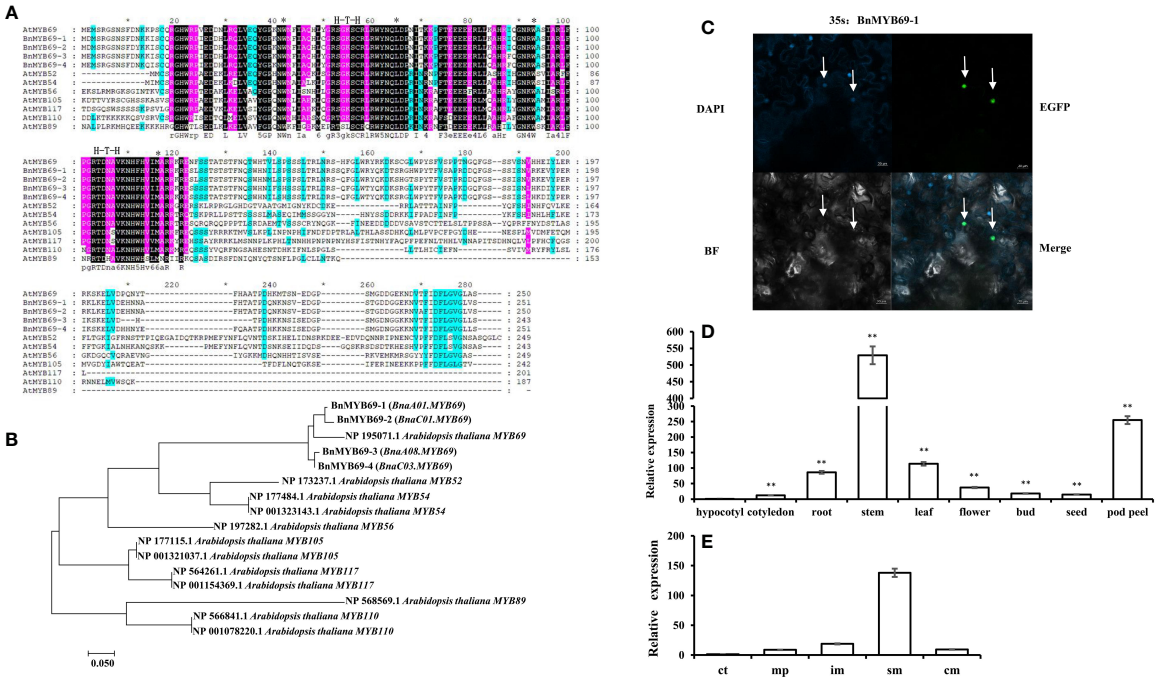


FIGURE 1
Cloning and analysis of *BMYB69s*. **(A)** Sequence alignment analysis of AtMYB69 and S21-AtMYB proteins. (*Between them are amino acid residues with H-T-H motif structure) **(B)** Molecular phylogenetic analysis by maximum likelihood method of BnMYB69s with S21 AtMYBs. **(C)** Subcellular localization of BnMYB69-1. (The arrow refers to the nucleus) **(D)** *BnMYB69s* overall expression in different organs of *B. napus*. **(E)** Expression of *BnMYB69s* in different parts of the main stem of *B. napus*. Note, (ct) growth cone tip. (mp) main-stem primordium. (im) initially lignified main-stem, (sm) semi-lignified main-stem. (cm) completely lignified main-stem. At least three biological replications (n≥3) were conducted for each experiment. All values are presented as the mean ± standard deviation (SD). Asterisks indicate significant or extremely significant differences from the first series (*0.01<p < 0.05; **p< 0.01) using one-way ANOVA.

individual member genes, and all of them showed typical stem-specific expression patterns (Figure S3). Overlapping result of the four genes' expression patterns was not perfectly coincided to the overall expression pattern, suggesting that there might be other member genes that have not been cloned in this study. We further checked the BLAST results, and found that the *B. napus* genome might contain another two *BnMYB69* genes beyond these four genes.

Silencing of *BnMYB69* increased biomass and altered chlorophyll and phytohormone profiles

Compared with the WT plants, the *BnMYB69i* plants had larger leaves, a higher chlorophyll content, heavier 1000-seed weight and shorter height when grown in the field (Figures 2, 3). The leaf area of transgenic plants at the bolting stage was significantly larger than that of the wild type (Figure 2C). The dry weight of *BnMYB69i* plants at the early bolting stage, full bloom stage and harvest stage increased by 36.4%, 44.5% and 53.2%, respectively, compared with those of the WT (Figure 2D). The 1000 seed weight of transgenic plants increased by 7.35% to 55.57% compared with that of WT plants (Figure 2E).

The chlorophyll fluorescence analysis showed that *BnMYB69*-silenced plants were very sensitive to light intensity (Figures 3A, B). Furthermore, the analysis of the chlorophyll content of *BnMYB69i* plants under different light conditions showed that chlorophyll synthesis in *BnMYB69i* plants may be related to light intensity (Figures 3C, D).

The hormone auxin (IAA) and 6-deoxocastasterone (6DS) in stems decreased significantly (Tables 1A, B); in contrast, abscisic acid (ABA) and brassinolide (BL) increased significantly (Tables 1A, B). In addition, IAA decreased and ABA increased in both stems and leaves, but the opposite was true in roots (Figure 3E). These results may explain the changes in morphological characteristics of transgenic plants.

Silencing of *BnMYB69* impaired vascular and fiber development, secondary wall lignification, and mechanical strength of the stem

BnMYB69 is mainly expressed in vascular organs, particularly xylem (Figure 4A). *BnMYB69i* plants showed a phenotypic change related to vascular organ growth retardation (Figure 4C). The stem circumference of the transgenic plant was significantly larger than that of the WT (Figure 2C), and the staining of xylem and the vascular bundle showed that the total number of xylem and vascular bundles in the cross-section of the stem increased (Figures 4Cb, c, e, f), but the number per unit area decreased, and this structural change caused the stem to become loose and soft (Figures 4B, C). The xylem in the longitudinal section of the stem decreased significantly, and the width was only 30% of that of the WT (Figures 4B, Ci, l). The number of pith

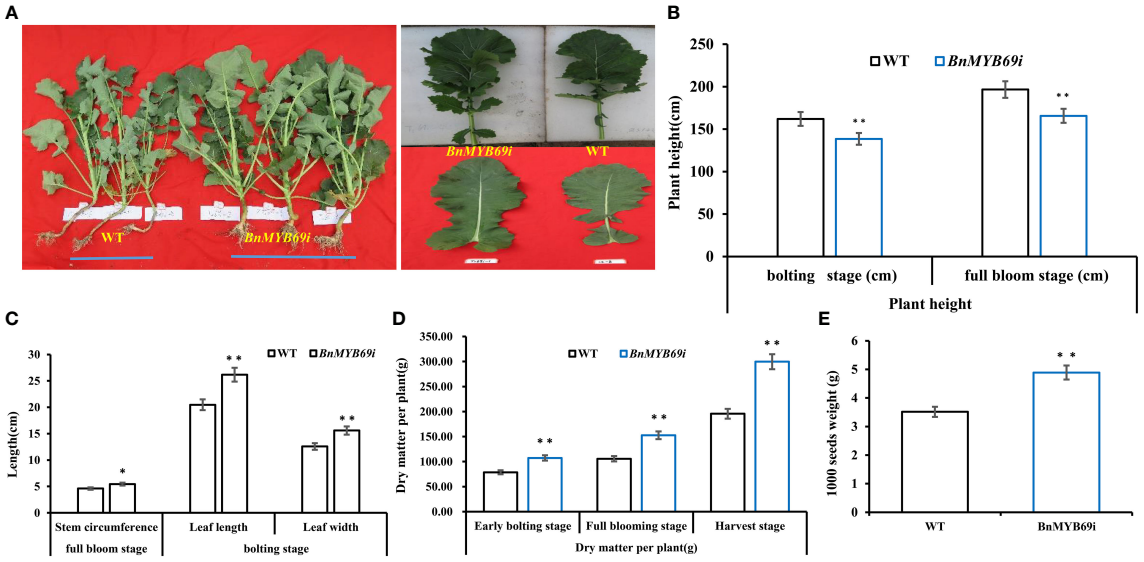


FIGURE 2
Morphological observation and dry matter weight of transgenic and WT plants. **(A)** Comparison of plant morphology between transgenic and WT plants in each generation. **(B)** Comparison of plant height of transgenic and WT plants at different stages. **(C)** Comparison of stem circumference and leaf of transgenic and WT plants. **(D)** Comparison of dry matter weight of transgenic and WT plants at three growth stages. **(E)** Comparison of 1000 seeds weight of transgenic and WT plants. Note, (WT) wild type *Brassica napus* cv. Zhongshuang 10 (BnMYB69i) *BnMYB69* RNA interference rapeseed (the same below). Asterisks indicate significant or extremely significant differences from WT.

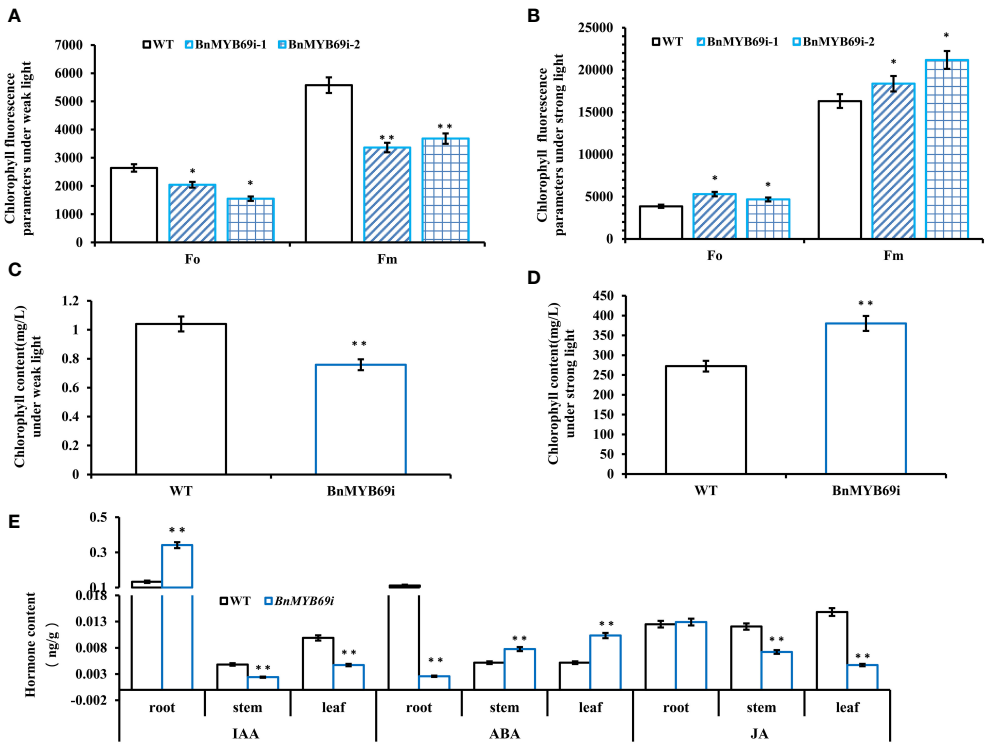


FIGURE 3
Analysis of chlorophyll content in different materials under different light intensities. **(A)** Chlorophyll fluorescence parameters under weak light. **(B)** Chlorophyll fluorescence parameters under strong light. **(C)** Chlorophyll content under weak light. **(D)** Chlorophyll content under strong light. **(E)** Comparison of three hormones in different organs between transgenic and WT plants. (root) hormone content of root. (stem) hormone content of stem. (leaf) hormone content of leaf. Asterisks indicate significant or extremely significant differences from WT (*0.01≤p < 0.05; **p < 0.01) using one-way ANOVA.

TABLE 1A Comparison of thirteen hormones in stems of transgenic and WT plants.

Material	Plant hormones (ng/g)						
	IAA	ABA	IBA	GA1	GA3	GA4	GA7
WT	19.81 ± 0.9892	5.69 ± 1.2799	0.05 ± 0.0070	0.07 ± 0.0014	0.07 ± 0.0131	0.11 ± 0.0267	0.06 ± 0.0114
BnMYB69i	7.19 ± 0.7958**	24.56 ± 3.1883**	0.12 ± 0.0318	0.12 ± 0.0219	0.05 ± 0.0057	0.23 ± 0.0652	0.03 ± 0.0168
Material	TZR	IP	IPA	SA	JA	MeJA	
WT	0.53 ± 0.1228	0.02 ± 0.0031	0.62 ± 0.0861	12.97 ± 2.2208	23.57 ± 2.0096	0.12 ± 0.0073	
BnMYB69i	0.82 ± 0.1236	0.01 ± 0.0032	0.97 ± 0.1517	9.36 ± 0.4046	23.73 ± 2.9764	0.28 ± 0.0641	

IAA, indole-3-acetic acid (auxin); ABA, Abscisic acid; IBA, indolylbutyric acid; GA1, gibberellin 1; GA3, gibberellin3; GA4, gibberellin 4; GA7, gibberellin 7; TZR, Trans Zeatin Riboside; IP, indolepro pionic acid; SA, Salicylic acid; JA, Jasmonic acid; MeJA, methyl jasmonate.

TABLE 1B Comparison of three hormones in transgenic and WT plants.

Material	Rapeseed hormones (ng/g)		
	BL	6DS	CS
WT	0.04 ± 0.0057	186.73 ± 12.8313	1.05 ± 0.1266
BnMYB69i	0.19 ± 0.0233**	123.83 ± 3.5775**	1.09 ± 0.0926

BL, brassinolide; 6DS, 6-deoxocastasterone; CS, castasterone. Because WT was too small, the contents of rapeseed hormone strigolactone (5DS) was not determined. (WT) ZS10 inbred pure material (explant of transgenic study). (BnMYB69i) Transgenic plants with the third generation of selfing. All values are presented as the mean ±standard error (SE) (n=3).

cells increased, and the number per unit area increased both horizontally and vertically (Figures 4Ca, d, i, l). In addition, compared with the WT, the root xylem of transgenic plants was stained deeper, indicating that the lignin content was increased (Figure 4E).

Statistical analysis showed that the number of pith cells in the unit area of transgenic plant stems increased by 14%, and the size and number of cells in cross section changed significantly (Figure 4B). In transgenic plants, the single xylem layer was loose and was approximately 43% thinner than that in the WT (Figure 4B), and the number of vessels in the vascular bundle of a microscopic vision of the cross section was also reduced by approximately 57% (Figure 4B). More significantly, the number of interfascicular fibers of transgenic plants in the cross-section decreased by 47.7%, and the number of interfascicular fibers between two vascular bundles decreased by 30% (Figure 4B). In addition, the number of pith cells in the cross-section of stem of transgenic plants increased extremely significantly by 95.6% (Figure 4B). The total number of vascular bundles in a microscopic vision of the cross section decreased significantly by 50.1% (Figure 4B). Furthermore, the sieve tube number in a microscopic vision of the cross section also decreased significantly in *BnMYB69i* plants (Figure 4B). Due to the changes in the number and size of xylem and pith cells in the cross-section of the stem, the cell cycle-related genes of the stem were analyzed by RT-PCR. The cell cycle-related genes *CAK1AT*, *CDKB1;1*, *CDKB2;1*, *CDT1A*, *CYCB1;1*, *CYCB2;1*, *CYCD2;1*, *CYCD3;1*, *CYCD3;3*, *DEL1*, *E2F1* and *ICK6* were significantly downregulated (Figure 4D).

Because of the changes in the stem xylem of transgenic plants, the resistance to *S. sclerotiorum* and stem strength were tested. After *S. sclerotiorum* inoculation, *BnMYB69* expression was upregulated at 3 and 9 h, distinctly upregulated at 24 h, and returned to a slightly higher than base level at 48 h, showing a dynamic change during the interaction between *B. napus* and *S. sclerotiorum* (Figure 5). This implies that the *BnMYB69* gene family may be involved in the interaction with *S. sclerotiorum* in oilseed rape. *BnMYB69i* plants were evaluated for decreasing *S. sclerotiorum* resistance. The stem and leaf lesion sizes were increased by 81.94% and 65.91%, respectively, compared with those of the WT, implying a substantial decrease in resistance to *S. sclerotiorum* stem rot disease (Figures 5B, C). It is clear that the suppression of *BnMYB69* significantly but solely weakened the resistance to *S. sclerotiorum* in vascular tissues in which innate *BnMYB69* is dominantly expressed.

In addition, the breaking resistance of the upper, middle and lower stems of transgenic plants was decreased by 47.31%, 11.93% and 37.65%, respectively, with an average value of 32.3% for the entire stem (Table 2), which indicates that the stem is fragile and the lodging resistance is reduced. In the stems of transgenic plants, compared with those of WT plants, the contents of neutral detergent fiber (NDF), acid detergent fiber (ADF) and lignin (ADL) decreased significantly by 17.17%, 13.48% and 30.04%, respectively (Table 3), whereas the contents of protopectin and soluble pectin decreased by 19.75% and 42.4%, respectively (Table 4). There was no significant difference in the pectin content of leaves between transgenic plants and WT plants (Table 4). In contrast, in the roots of transgenic plants, the contents of neutral detergent fiber (NDF), acid detergent fiber (ADF) and lignin (ADL) were significantly increased by 10.8%, 11.5%, and 9.3%, respectively (Table 3).

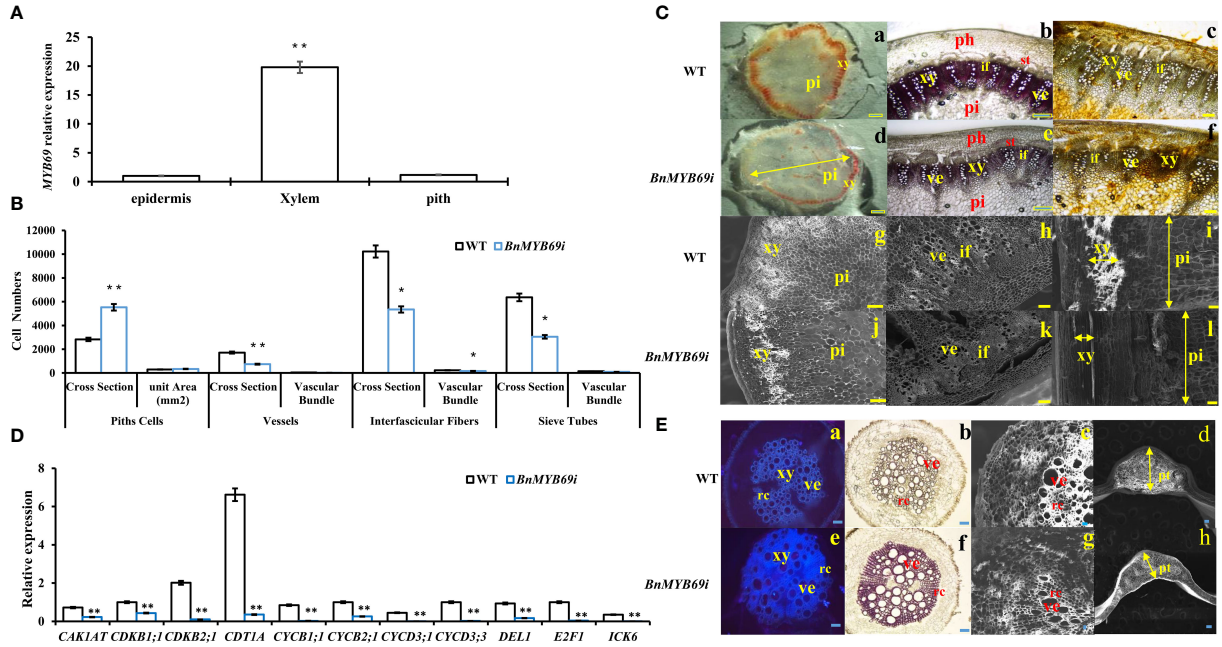


FIGURE 4
*BnMYB69*s Expression and analysis in stem tissues and microscopic observation of stem and root. **(A)** Expression of *BnMYB69* in different parts of main stem of rapeseed. **(B)** Number of cells of piths, sclerenchyma, vessels, interfascicular fibers and sieve tubes of stems cross-section of *BnMYB69I* transgenic plants. Cross Section is the whole cross section for Pith Cells, and is a microscopic vision of the cross section for Vessels, Interfascicular Fibers and Sieve Tubes. **(C)** Histochemical staining and electron microscopic observation of cross-section of stem. **(a-c)** Histochemical staining of stems of WT. **(d-f)** Histochemical staining of stems of *BnMYB69* silenced plants. **(a, d)** Mäule staining of lignin in whole cross-sections. **(b, e)** Phloroglucinol-HCl staining of lignin in partial cross-sections. **(c, f)** Hydroxylammonium staining of pectin in partial cross-sections. **(g-h)** Electron microscopic cross-section of stems of WT. **(i)** Electron microscopic longitudinal section of stem of WT. **(j-k)** Electron microscopic cross-section of stems of *BnMYB69* silenced plants, **(l)** Electron microscopic longitudinal section of stem of *BnMYB69* silenced plants. pi, Piths; xy, Xylem; ph, Phloem; ve, Vessel; if, Interfascicular fibers; st, Sieve tube. Bar, 2 mm **(a,d)**, 1 mm **(b, c, e, f, g-l)**. **(D)** The cell cycle related genes expression in stems of transgenic plant and WT. **(E)** Comparison of root, petiole and leaf morphology between transgenic and WT. **(a, e)** Fluorescent brightener dyeing of lignin in whole root cross-sections, **(b, f)** Phloroglucinol-HCl staining of lignin in whole root cross-sections, **(c, g)** Electron microscopic observation of root partial cross-sections, **(d, h)** Electron microscopic observation of petiole cross-section of transgenic and control plants. xy, Xylem; ve, Vessel; rc, Ray cell; pt, Petiole thickness. Bar, 1 mm **(a, b, e, f)**, 200 μm **(c, g)**,1 mm **(d, h)**. Asterisks indicate significant or extremely significant differences from WT (*0.01≤ p < 0.05; **p< 0.01) using one-way ANOVA.

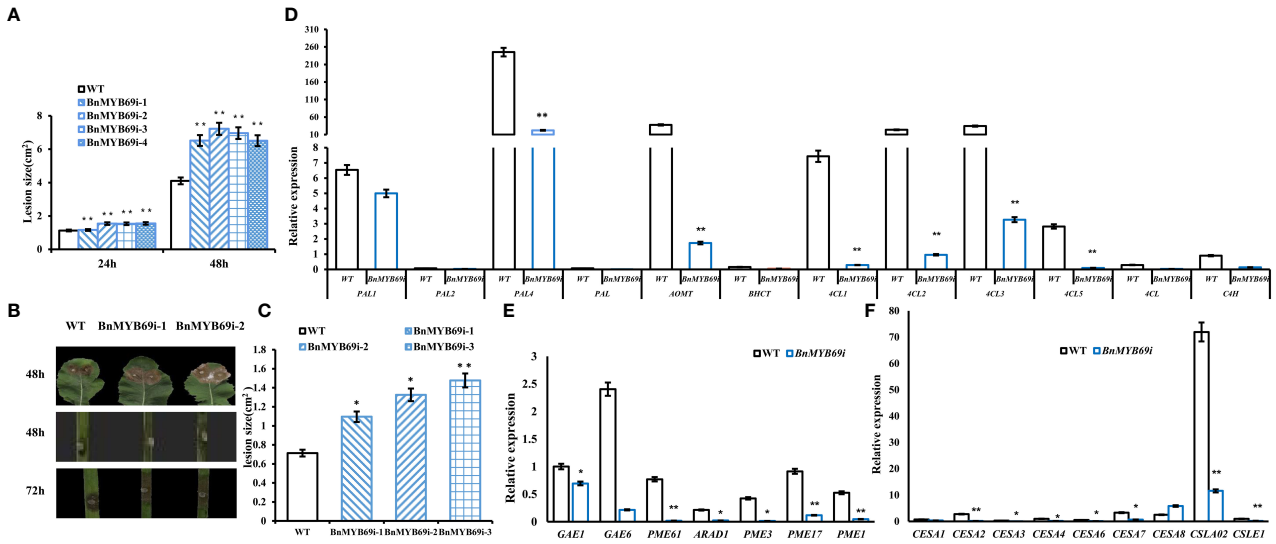


FIGURE 5
Expression of *BnMYB69*s and secondary wall biosynthesis genes in *B. napus* leaf and stem after *S. sclerotiorum* inoculation. **(A)** Statistics of leaf lesion area at different times after inoculation. **(B)** Lesion plaques on leaves and stems after *S. sclerotiorum* inoculation. **(C)** Statistics of stem lesion area at 48h after inoculation. **(D)** Lignin biosynthesis related genes expression in stems of transgenic plant and WT. **(E)** Pectin biosynthesis related genes expression in stems of transgenic plant and WT. **(F)** Cellulose biosynthesis related genes expression in stems of transgenic plant and WT. Asterisks indicate significant or extremely significant differences from WT (*0.01≤ p < 0.05; **p< 0.01) using one-way ANOVA.

TABLE 2 Comparison of bending-resistance between transgenic and WT plants.

Material	Bending resistance (N)		
	Lower stem	Middle stem	Upper stem
WT	291.1±19.23	111.1±3.45	61.7±4.89
BnMYB69i	183.0±8.87**	90.8±4.54**	30.7±2.25**

Asterisks indicate significant or extremely significant differences from WT (*0.01≤ p < 0.05; **p< 0.01) using one-way ANOVA.

Rapeseed stems mainly contain S and G lignin types, and we found that the ratio of S/G increased significantly by 7.35% (Table 3), implying G lignin biosynthesis was more serviously impaired than S lignin biosynthesis. However, the S/G ratio of roots in transgenic plants decreased significantly, implying a trend opposite to the stem. These results showed that the silencing of *BnMYB69* significantly reduced the differentiation of lignification tissues, the deposition of stem secondary cell wall ingredients (such as lignin, cellulose and protopectin), and the mechanical strength in the stem, with little influence on leaf lignification and an increase in root lignification (Table 3).

The lignin biosynthesis-related genes *PAL1*, *PAL2, PAL4, PAL*, *C4H*, *4CL1*, *4CL2*, *4CL3*, *4CL5*, *4CL*, *COMT*, and *HCT* were downregulated in the stems of transgenic plants (Figure 5D). The pectin biosynthesis-related genes *GAE1*, *GAE6*, and *ARAD1* were significantly downregulated. Additionally, the pectin methylesterase genes *PME1*, *PME3*, *PME17*, and *PME61* were significantly downregulated (Figure 5E). In transgenic plants, the genes *CESA1*, *CESA2*, *CESA3*, *CESA4*, *CESA6*, *CESA7*, *CSLE1* and *CSLA02*, which are related to cellulose biosynthesis, were significantly downregulated (Figure 5F), and only *CESA8* was upregulated. The results showed that *BnMYB69* regulated the synthesis of lignin, cellulose and pectin, resulting in changes in breaking resistance and disease resistance.

IAA treatment of transgenic *BnMYB69i* plants recovered many phenotypes and increased chlorophyll synthesis

Phenotypic observation of the growth changes of *BnMYB69i* plants at the cotyledon stage and the three- and five-leaf stages showed that IAA treatment significantly increased the plant height of transgenic plants and inhibited the increase in stem circumference. The *BnMYB69i* plants recovered to the same size as the WT plants at the seedling stage in the growth chamber (Figures 6A, B).

qRT-PCR analysis of the expression of key genes in the IAA synthesis pathway showed that the expression of the *TAA1*, *YUC2*, *TAR1*, *TAR2*, *CYP79B3*, *NIT2*, *AMI1*, *AAO1* and *AAO2* genes decreased significantly in the untreated *BnMYB69i* plants ($P < 0.05$), while the expression of *CYP79B2* increased significantly ($P < 0.05$). The expression levels of *TAR1*, *TAR2*, *YUC2*, and *NIT2* in *BnMYB69i* plants treated with IAA were significantly higher than those in untreated plants, but there was no significant difference in *TAA1*, *CYP79B2*, or *CYP79B3*. *AMI1*, *AAO1*, and *AAO2* were significantly lower than in untreated plants ($P < 0.05$) (Figure 7A).

The chlorophyll content of various lines under strong light was determined. The results showed that the total chlorophyll (Ct) of

TABLE 3 Cell wall ingredients determination in stem and root by NIRs.

Material	Stem CWR (%)			
	NDF	ADF	ADL	S / G ratio
WT	71.25±1.45	56.05±0.55	12.70±0.46	0.351±0.001
BnMYB69i	66.70±0.51*	49.41±0.05*	8.90±0.02**	0.377±0.009*
Material	**Root CWR (%)**			
	NDF	ADF	ADL	S / G ratio
WT	76.02±0.30	>57.15±0.83	13.02±0.19	0.343±0.0012
BnMYB69i	84.25±2.47*	63.72±1.77*	14.22±0.10**	0.335±0.0015*

NDF, Neutral detergent fiber; ADF, Acid detergent fiber; ADL, Acid detergent lignin; S/G, syringyl(S)/guaiacyl (G) ratio.
Asterisks indicate significant or extremely significant differences from WT (*0.01≤ p < 0.05; **p< 0.01) using one-way ANOVA.

TABLE 4 Protopectin content determination in stem and leaf.

Material	Protopectin (μmol/g)		Soluble pectin (μmol/g)	
	stem	leaf	stem	leaf
WT	171.31±10.14	264.31±2.01	79.50±20.03	56.18±7.29
BnMYB69i	137.64±3.86*	256.25±1.15	45.86±22.03*	60.86±2.87

Asterisks indicate significant or extremely significant differences from WT (*0.01≤ p < 0.05; **p< 0.01) using one-way ANOVA.

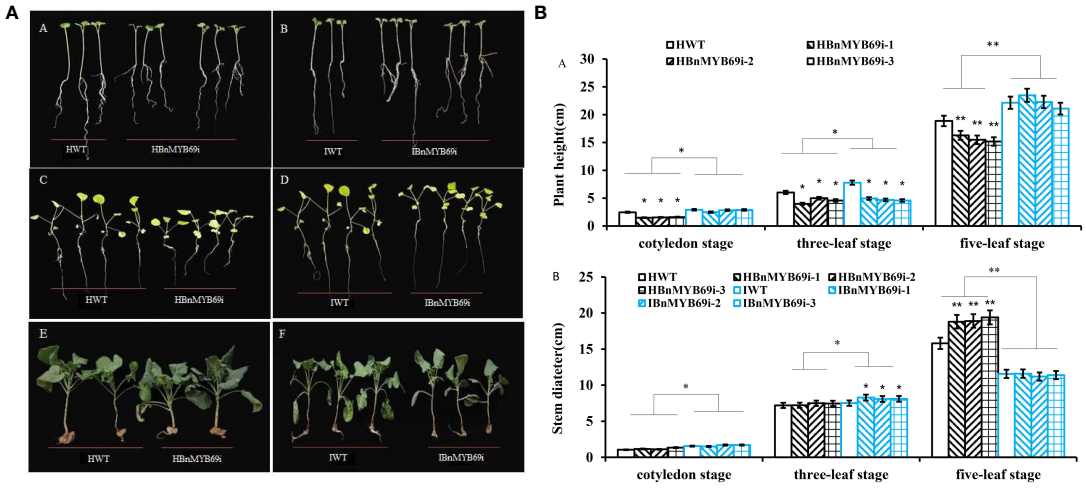


FIGURE 6
Phenotypic observation of transgenic plants and WT treated with H_2O and IAA. **(A)** Growth of *BnMYB69* transgenic plants and WT under different treatments at different stages. **(A, B)** cotyledon stage. **(C, D)** three-leaf stage. **(E, F)** five-leaf stage. HWT, hydroponic WT plant; HBnMYB69i, hydroponic *BnMYB69* interfering plant; IWT, WT plant with IAA treatment; IBnMYB69i, *BnMYB69* interfering plant with IAA treatment (the same below). **(B)** Plant height and stem diameter of RNAi plants and WT under different treatments at different periods. **(A)** Change of plant height. **(B)** Change of stem diameter. Asterisks indicate significant or extremely significant differences from WT (*0.01≤ p < 0.05; **p< 0.01) using one-way ANOVA.

BnMYB69i plants increased significantly under strong light with an increase mainly in leaf chlorophyll a and a decrease in chlorophyll b (Figure 7B).

The expression of chlorophyll synthesis-related genes under IAA treatment was checked with qRT-PCR (Figure 7C). The expression levels of *HEMA1*, *HEMG1*, *CHLM*, *CRD1*, and *CHLG* in transgenic plants were significantly downregulated after treatment ($P < 0.05$), but the expression levels of *GSA1*, *PORA* and *CAO* were significantly upregulated in the transgenic plants ($P < 0.05$) (Figure 7C). There was no significant difference in *HEMF1* gene expression (Figure 7C). It was also shown that *BnMYB69* probably regulates the gene expression of *HEMA1*, *HEMG1*, *CHLM*, *CRD1*, and *CHLG* in some way.

Silencing of *BnMYB69* affected shikimic acid metabolism and flavonoid-proanthocyanidin biosynthesis

Our above results indicated that *BnMYB69* is involved in the regulation of various metabolites, such as lignin, cellulose, pectin, phytohormones, and chlorophylls. The upstream reasons should be diverse and complicated, in this study we first detected the changes of shikimate-related pathways, because shikimate is directly upstream to the lignin pathway which is the major target influenced by *BnMYB69* downregulation.

qRT-PCR showed that the expression of key genes of the glycolytic pathway (*HKL1* and *PKT7*) was significantly upregulated in transgenic plants, and IAA treatment completely mitigated this effect ($P < 0.01$) (Figure 8A). The expression of key genes of the pentose phosphate pathway (*GP6D5* and *tktA*) (Figure 8B) and shikimic acid pathway (*ADT2*) were significantly downregulated in *BnMYB69i* plants ($P < 0.01$) (Figure 8C). IAA treatment completely recovered expression levels of *ADT2* and *GP6D5*, over-compensated the expression of *DHD/SDH* and *SK2*, but had little recovery on *tktA* (Figure 8C). These results show that the pentose phosphate pathway and shikimic acid pathway, in coordination with the upstream glucose metabolic pathway, are most likely to be regulated by *BnMYB69* together with IAA signaling, and different loci have differed responses.

Downstream the shikimate pathway, flavonoid-proanthocyanidin (PA) pathway is neighboring to lignin pathway and these two pathways have competition in metabolic flux. Thus, we detected the influence of *BnMYB69* downregulation on the flavonoid-PA pathway. Like lignin content, the PA content in the leaves of *BnMYB69i* plants was also significantly lower than WT, and IAA treatment exerted a little compensation (Figure 9A). It was confirmed that shikimates, lignins, and PAs are all decreased by *BnMYB69* downregulation.

The expression of key genes involved in PA biosynthesis was also analyzed using qRT-PCR (Figures 9B, C). In *BnMYB69i* plants, expression levels of core flavonoid biosynthesis genes *CHS*, *F3H* and *F3'H*, anthocyanin biosynthesis gene *ANS*, and PA pathway regulatory genes *TT1*, *TT2*, TT8 and *TTG2* were all significantly lower than those of WT plants ($P < 0.05$). IAA treatment did not exert recovery effect on most of these loci, but *TT2* expression was completely recovered. These results showed that *BnMYB69* regulated the shikimate-flavonoid-PA compound pathway, independent of or downstream IAA signaling, but IAA is involved in *BnMYB69* regulation on *TT2* expression.

Discussion

BnMYB69 is a S21 R2R3-type transcription factor

The R2R3-MYB family is the largest and most widely studied family of MYB transcription factors. In addition to the similar

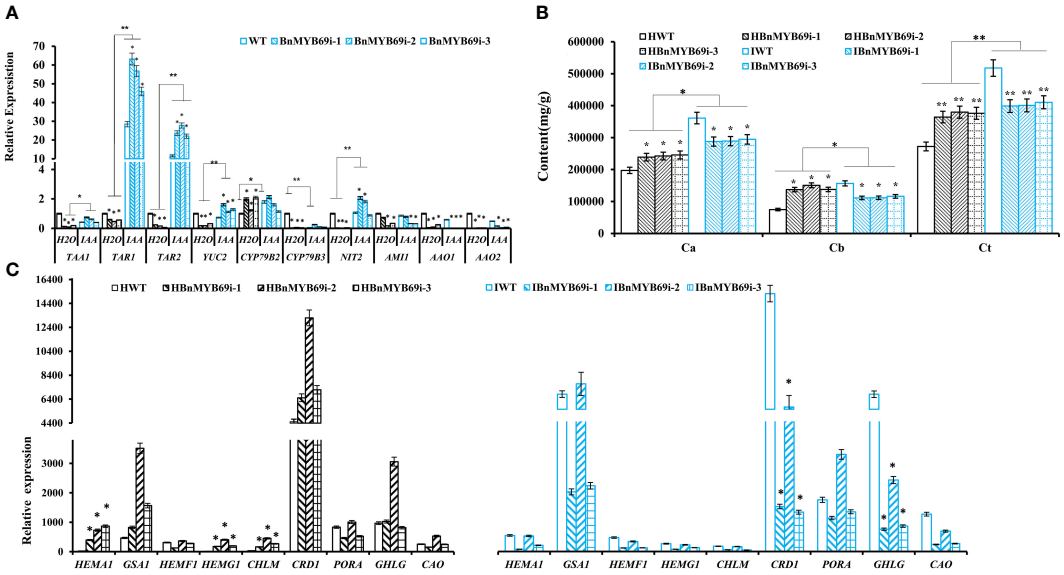


FIGURE 7
Changes of IAA and chlorophyll synthesis in transgenic and WT plants under different treatments. **(A)** Expression analysis of genes related to IAA synthesis pathway in transgenic and WT plants under different treatment. **(B)** Analysis of chlorophyll content difference of *BnMYB69* transgenic plants under different treatments. Ca, chlorophyll a; Cb, chlorophyll b; Ct, total chlorophyll. **(C)** Differential expression of key enzymes in chlorophyll synthesis pathway under different treatments. (H_2O) Control plant with water instead of IAA. (IAA) plant with IAA treatment (the same below). Asterisks indicate significant or extremely significant differences from WT (*0.01≤ p < 0.05; **p< 0.01) using one-way ANOVA.

functions of 1R-MYB family members, the members of this family are widely involved in plant life processes such as primary metabolism and biological and abiotic stress responses (Ito, 2005). Researchers believe that most members of the R2R3 MYB protein family are involved in regulating plant-specific physiological and biochemical processes such as plant secondary metabolism and cell morphogenesis (Wan and Li, 2002). In addition, many members of the R2R3 MYB gene family isolated from different species regulate anthocyanin biosynthesis (Czemmel et al., 2009; Machado et al., 2009; Feng et al., 2010; Lin-Wang et al., 2011). However, most *MYB* transcription factors have been cloned, analyzed and studied in *Arabidopsis*, some researchers have found that MYB TFs play vital

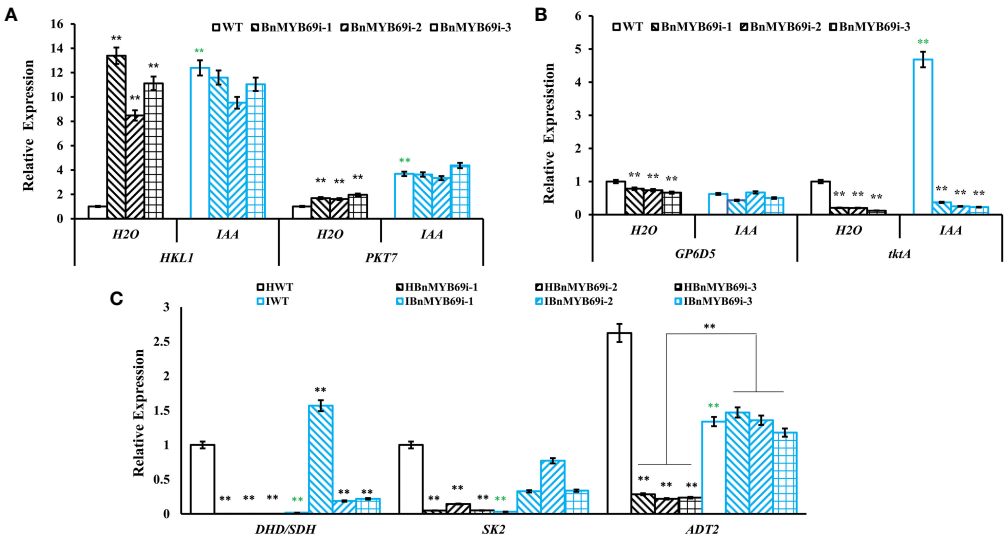


FIGURE 8
Expression of key genes of related metabolic pathways under different treatments. **(A)** Expression of key genes of glycolysis pathway in different materials. **(B)** Expression of key genes of pentose phosphate pathway in different materials. **(C)** Expression of *DHD/SDH, SK2* and *ADT2* in shikimate pathway in different materials. Asterisks * or ** indicate significant or extremely significant differences from HWT (*0.01<p < 0.05; **p< 0.01) using one-way ANOVA.

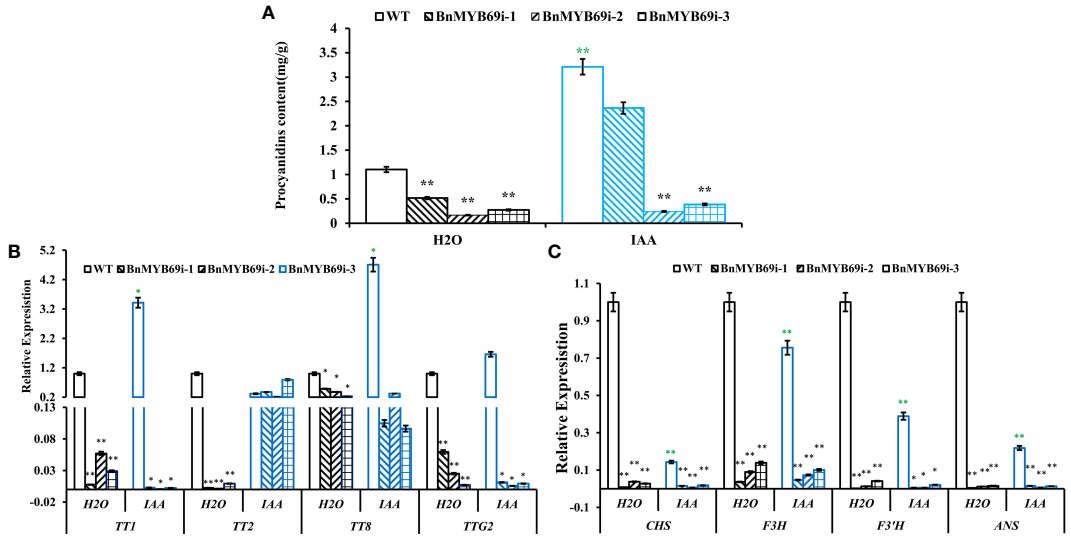


FIGURE 9
Changes of proanthocyanidin (PA) synthesis genes in transgenic and WT plants under different treatments. **(A)** Changes of PA content under different treatments. **(B, C)** Expression of PA synthesis related genes of different materials under different treatments. Asterisks indicate significant or extremely significant differences from WT (*0.01≤ p < 0.05; **p< 0.01) using one-way ANOVA.

roles in plant development and anthocyanin metabolism, and *PAP1/2* can promote the expression of anthocyanin biosynthesis genes (Muleke et al., 2021). In fact, MYB transcription factor members play a role in regulating growth and development in different parts of different time periods. For example, *AtMYB115* and *AtMYB188/PGA37* participate in regulating the seed germination rate at the seed germination stage (Wang et al., 2009); *AtMYB30*, *AtMYB38*, and *AtMYB18/LAF1* regulate hypocotyl elongation (Yang et al., 2009; Froidure et al., 2010); and *AtMYB21*, *AtMYB24*, *AtMYB26*, *ATMYB35*, *AtMYB57*, *GhMYB80*, and *AtMYB99* participate in anther development (Millar and Gubler, 2005; Xu et al., 2014). There are also some MYB transcription factors, such as *BnMYB43* and *GbMYBR1*, involved in lateral organ development and stem morphogenesis, which regulate the number of plant branches and the development of plant higher organs (Jiang et al., 2020; Su et al., 2020).

In this study, four *BnMYB69* genes were isolated, characterized, and functionally elucidated. Results of systemic analysis on the structure and homology features of their genes and proteins all indicated that they encoded a S21 R2R3-MYB transcription factor, orthologous to AtMYB69. There are few reports on the function of MYB69. Previous studies prove that AtMYB69 is a transcriptional activator dedicated to secondary cell wall thickening of fiber cells (Dubos et al., 2010). According to other reports on the function of MYB transcription factors similar to ATMYB69 protein, these transcription factors are not only related to secondary wall synthesis (Zhong et al., 2008; Kim et al., 2012), but also related to anthocyanin metabolism (Peng et al., 2019; Xu et al., 2021), plant resistance (Xu et al., 2020; Im et al., 2021), etc. Therefore, systematic study of the function of *MYB69* is helpful to providing insight into the biofunction of *MYB* gene family.

Silencing of *BnMYB69* affected lodging resistance and *S. sclerotiorum* resistance

The *MYB69* transcription factor is related to xylem formation and is strongly expressed in stems. Many researchers have found that the expression of key genes in lignin synthesis is related to lodging traits, particularly *COMT* (Ma, 2009; Li et al., 2013), *F5H* (Li and Qi, 2011; Li et al., 2013), *4CL* (Li and Qi, 2011; Huang et al., 2013; Li et al., 2013) and *PAL* (Huang et al., 2013). Our results showed that silencing *BnMYB69* reduced the lodging resistance of rapeseed. Cytological observation of cross-sections and longitudinal sections of stems and the expression of key genes of lignin metabolism in transgenic plants revealed a positive correlation with lignin synthesis. This shows that BnMYB69 is an activator of lignin synthesis in *B. napus*.

In addition to lignin, plant cell wall components also contain cellulose and pectin. Cellulose and pectin content analysis showed that the cell wall of *BnMYB69i* plants developed poorly. This might be the reason that the resistance to *S. sclerotiorum* in *BnMYB69i* plants was significantly decreased compared with that in WT plants.

The effect of *BnMYB69*s on roots is quite different from or even opposite to its effect on stems, particularly on monolignol types. The proportion of S/G in *BnMYB69i* plants increases in stems and decreases in roots, which may contribute to the different physical characteristics of the roots and the stems. Furthermore, lignin, cellulose and pectin increased in roots, though decreased in the stems, indicating that there was a significant difference between the aboveground and underground parts of *BnMYB69i* plants. Since the basic difference between aboveground and underground is dark or in light, this phenomenon implies that *BnMYB69i* functioning depends on light condition.

BnMYB69 silencing leads to changes in the cell cycle, resulting in plant dwarfing

The offspring of *BnMYB69i* were generally shorter than those of WT plants. Pith cells and unit volume were greater based on electron microscopy and staining, and the expression of most cell cycle-related genes in the stems of transgenic plants was significantly lower than that in the stems of WT plants. This indicates that cell division in transgenic plants was not vigorous and rapid, and the total number of cells should be reduced, which was consistent with the results of cells in the previous cross-cut stems. In addition, the IAA content in the stems of transgenic plants was lower than that of WT plants, which confirmed that plant cell division was not vigorous. Furthermore, the increase in ABA content led to plant dwarfing and root development changes. Antisense transgenic plants and overexpression plants of the rice dwarf gene *OsDWARF48* were analyzed, and the IAA content of the dwarf plants was lower than that of WT plants (Li et al., 2018).

qRT-PCR results confirmed that auxin can also promote cell division by degrading the expression of CDK inhibitor protein Kip-related proteins (KRPs). The statistical results of the number of cells (Figure 4B) in the stems of transgenic plants also show that the change in cell division caused a change in the number of xylem cells and then caused a change in stem height and stem circumference. Since *BnMYB69* is mainly expressed in tissues under lignification, thus it is logical that the confinement of *BnMYB69* downregulation on cell cycle and cell division is mainly on the secondary growth and plant height, and the promotion effect on parenchyma cells, lateral growth and total biomass might be because of more flux branch-off from suppressed secondary metabolism.

BnMYB69 silencing affects IAA synthesis, photosynthetic reaction, and glucose metabolism

The results of the hormone content tests on the stems and leaves of the offspring of transgenic plants showed that the content of IAA was significantly lower than that of WT, indicating that *BnMYB69* affected the synthesis of IAA. The precursor of IAA is tryptophan, which regulates the growth, development and resistance of plants. Researchers have found that expression of the *BnMYB193* gene in *B. napus* was also induced by IAA, ABA and other hormones to promote the development of lateral roots of rapeseed and improve the lodging resistance of plants (Wu et al., 2018).

In the IAA compensation experiment, *BnMYB69* affected the shikimic acid pathway upstream of IAA synthesis. In addition, the pentose phosphate pathway upstream of the shikimic acid pathway was also affected by *BnMYB69*. It can be inferred that *BnMYB69* has regulatory functions on metabolic pathways since sugar metabolism, although sugar synthesis is affected by photosynthesis. The up and down regulation of the hexokinase gene in plants leads to extensive plant phenotypic changes such as overexpression of *AtHKL1* in tomato, inhibition of plant growth, reduction of chlorophyll content and photosynthetic rate in leaves, and photochemical quantum efficiency reduction (Dai et al., 1999). Inhibiting *AtG6PD5* expression causes plants to have short stems and leaves, curly leaves and light leaf color (Wang, 2013). *BnMYB69i* transgenic plants indeed had lower plant height, which also confirmed the regulatory function of *BnMYB69*.

In a study of *B. napus*, *TT1* gene silencing led to the downregulation of flavonoid pathway genes such as *TT3*, *TT4*, *TT5*, *TT6*, *TT7* and *TT18*. The interference of *BrTT1* gene expression in yellow seed *B. napus* causes the abnormal or absent expression of several important *BrTT1* genes in the flavonoid biosynthesis pathway, which then blocks the accumulation of PAs resulting in a transparent seed coat. The *BnTT1* gene in *B. napus* also regulates the PA and flavonoid metabolism pathways (Lian et al., 2017).

ANS is the most critical gene in the anthocyanin biosynthesis pathway. *UFGT* and *ANS* cooperate to transform unstable anthocyanins into stable anthocyanins in the cytoplasm and vacuole. Among the five genes of the anthocyanin biosynthesis pathway, only *ANS* and *DFR* are regulated by *TTG2*. Studies on tobacco show that *TTG2* and *ARF8* work together to control the degree of flower color by regulating the expression of *ANS* and *DFR* in the anthocyanin biosynthesis pathway (Li et al., 2017). *ARF8* relies on *TTG2* to regulate the expression of *ANS* and *DFR*, which affects anthocyanin accumulation. *TTG2* inhibition prevents plant trichome formation, anthocyanin accumulation and seed color pigmentation. Similarly, *AtTT8* is also involved in the regulation of anthocyanin and PA biosynthesis in *Arabidopsis* (Baudry et al., 2006; Xu et al., 2013). Complementation analysis in *Arabidopsis TT8* mutant showed that *NnTT8* in *Nelumbo nucifera* could function similarly to *AtTT8* in regulating anthocyanin and proanthocyanin biosynthesis (Deng et al., 2021). The downregulated expression of *TT1*, *TTG2, TT8, ANS* and *DFR* in *BnMYB69i* transgenic plants also confirmed its impact on flavonoid compounds and the anthocyanin synthesis pathway, which explains the increase in surface trichomes of transgenic plants.

Chlorophyll synthesis in *BnMYB69*-silenced plants was more sensitive to light intensity

The entire process of chlorophyll synthesis requires 15 enzymatic reactions. Among them, *HEMA* is the coding gene that catalyzes ALA synthesis. The *HEMA* family has three members in *Arabidopsis*, and the expression of *HEMA1* is induced by light (McCormac et al., 2001), whereas *HEMA2* and *HEMA3* are not induced by light, and the *HEMA3* gene is not even expressed. However, the *HEMA1* gene is expressed in all plant tissues, and its expression is higher in photosynthetic tissues (Tanaka et al., 1996). CHLM and CRD can catalyze the transformation of protoporphyrin IX into chlorophyll ester a, and knock out of *CRD* in *Arabidopsis* resulted in slow plant growth and development and chlorosis of leaves (Bang et al., 2008). CHLG acts to form chlorophyll a, and CAO catalyzes Chla

oxidation to form Chlb. Under strong light, chlorophyll synthesis increases, and the PSII reaction intensifies, such that the expression of the two coding genes increases. The *PORA* gene, a member of the *POR* family, is a coding gene for protochlorophyll ester oxidoreductase. The *PORA* gene has a photoinduction effect, and it is also the main enzyme of plant leaf chlorophyll accumulation and the only confirmed gene in the phytochrome A-mediated mechanism for the greening blocks (Sperling et al., 1997). Some researchers have shown that *PORA* overexpression is sufficient to guide the synthesis of a large amount of chlorophyll and the development of normal plants in the absence of *PORB* and *PORC* (Paddock et al., 2010). However, the expression of *PORA* decreased in that experiment. It was speculated that *PORA* was affected by the *BnMYB69* gene, and the expressions of *PORB* and *PORC* may be increased. Mechanism of BnMYB69 effect on chlorophyll synthesis and light reaction will be deeply studied in the future.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

YC: Designing the research, analyzing some data, and revising the manuscript. NL: Designing and doing some experiments, analyzing most data, and writing the original manuscript. MW and JJ: Doing some experiments. QZ and JY: Providing writing suggestions. JLi: Providing some plant material. JLian and YX: Providing some technologies of the experiment. All authors: reading this manuscript and agreeing with publication. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (NSFC, with grant numbers of 31701335, 31171177 and 32272015).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1157836/full#supplementary-material

References

Baldoni, E., Genga, A., and Cominelli, E. (2015). Plant MYB transcription factors: Their role in drought response mechanisms. *Int. J. Mol. Sci.* 16 (7), 15811–15851. doi: 10.3390/ijms160715811

Bang, W. Y., Jeong, I. S., Kim, D. W., Im, C. H., Ji, C., Hwang, S. M., et al. (2008). Role of *Arabidopsis* CHL27 protein for photosynthesis, chloroplast development and gene expression profiling. *Plant Cell Physiol.* 49 (9), 1350–1363. doi: 10.1093/pcp/pcn111

Baudry, A., Caboche, M., and Lepiniec, L. (2006). TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. *Plant J.* 46 (5), 768–779. doi: 10.1111/j.1365-313X.2006.02733.x

Blumenkrantz, N., and Asboe-Hansen, G. (1973). New method for quantitative determination of uronic acids. *Anal. Biochem.* 54 (2), 484–489. doi: 10.1016/0003-2697(73)90377-1

Cardoza, V., and Stewart, C. N. (2003). Increased *Agrobacterium*-mediated transformation and rooting efficiencies in canola (*Brassica napus* L.) from hypocotyl segment explants. *Plant Cell Rep.* 21 (6), 599–604. doi: 10.1007/s00299-002-0560-y

Czemmel, S., Stracke, R., Weisshaar, B., Cordon, N., Harris, N. N., Walker, A. R., et al. (2009). The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. *Plant Physiol.* 151 (3), 1513–1530. doi: 10.1104/pp.109.142059

Dai, N., Schaffer, A., Petreikov, M., Shahak, Y., Giller, Y., Ratner, K., et al. (1999). Overexpression of *Arabidopsis* hexokinase in tomato plants inhibits growth, reduces photosynthesis, and induces rapid senescence. *Plant Cell* 11 (7), 1253–1266. doi: 10.2307/3870747

Deng, J., Li, J., Su, M., Lin, Z., Chen, L., and Yang, P. (2021). A bHLH gene *NnTT8* of *Nelumbo nucifera* regulates anthocyanin biosynthesis. *Plant Physiol. Biochem.* 158, 518–523. doi: 10.1016/j.plaphy.2020.11.038

Ding, A., Tang, X., Yang, D., Wang, M., Ren, A., Xu, Z., et al. (2021). ERF4 and MYB52 transcription factors play antagonistic roles in regulating homogalacturonan de-methylesterification in *Arabidopsis* seed coat mucilage. *Plant Cell* 33 (2), 381–403. doi: 10.1093/plcell/koaa031

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15 (10), 573–581. doi: 10.1016/j.tplants.2010.06.005

Feller, A., Machemer, K., Braun, E. L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* 66 (1), 94–116. doi: 10.1111/j.1365-313X.2010.04459.x

Feng, S., Wang, Y., Yang, S., Xu, Y., and Chen, X. (2010). Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10. *Planta* 232 (1), 245–255. doi: 10.1007/s00425-010-1170-5

Foster, T. J., Ablett, S., McCann, M. C., and Gidley, M. J. (1996). Mobility-resolved ^{13}C-NMR spectroscopy of primary plant cell walls. *Biopolymers* 39 (1), 51–66. doi: 10.1002/(sici)1097-0282(199607)39:1<51::Aid-bip6>3.0.Co;2-u

Froidure, S., Roby, D., and Rivas, S. (2010). Expression of the *Arabidopsis* transcription factor AtMYB30 is post-transcriptionally regulated. *Plant Physiol. Biochem.* 48 (8), 735–739. doi: 10.1016/j.plaphy.2010.04.012

Gai, J. Y. (2005). Segregation analysis of genetic system of quantitative traits in plants. *Yi Chuan* (in Chinese) 27 (1), 130–136. doi: 10.1007/s11515-005-0001-8

Godoy, G., Steadman, J. R., Dickman, M. B., and Dam, R. (1990). Use of mutants to demonstrate the role of oxalic acid in pathogenicity of *Sclerotinia sclerotiorum* on *Phaseolus vulgaris*. *Physiol. Mol. Plant Pathol.* 37 (3), 179–191. doi: 10.1016/0885-5765(90)90010-U

Hajiebrahimi, A., Owji, H., and Hemmati, S. (2017). Genome-wide identification, functional prediction, and evolutionary analysis of the R2R3-MYB superfamily in *Brassica napus*. *Genome* 60 (10), 797–814. doi: 10.1139/gen-2017-0059

Hong, L., Niu, F., Lin, Y., Wang, S., Chen, L., and Jiang, L. (2021). MYB117 is a negative regulator of flowering time in *Arabidopsis*. *Plant Signaling Behav.* 16 (5), e1901448. doi: 10.1080/15592324.2021.1901448

Huang, J.-H., LI, W., Qu, C.-M., Liu, L.-Z., Xu, X.-F., Wang, R., et al. (2013). Expression characteristics of key genes in lignin pathway among different lodging resistance lines of *Brassica napus* L. *Acta Agronomica Sin.* 39 (8), 1339–1344. doi: 10.3724/sp.j.1006.2013.01339

Im, J. H., Ko, J.-H., Kim, W.-C., Crain, B., Keathley, D., and Han, K.-H. (2021). Mitogen-activated protein kinase 6 negatively regulates secondary wall biosynthesis by modulating MYB46 protein stability in *Arabidopsis thaliana*. *PloS Genet.* 17 (4), e1009510. doi: 10.1371/journal.pgen.1009510

Ito, M. (2005). Conservation and diversification of three-repeat myb transcription factors in plants. *J. Plant Res.* 118 (1), 61–69. doi: 10.1007/s10265-005-0192-8

Jiang, J., Liao, X., Jin, X., Tan, L., Lu, Q., Yuan, C., et al. (2020). *MYB43* in oilseed rape (*Brassica napus*) positively regulates vascular lignification, plant morphology and yield potential but negatively affects resistance to *Sclerotinia sclerotiorum*. *Genes (Basel)* 11 (5), 581. doi: 10.3390/genes11050581

Jin, H., and Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.* 41 (5), 577–585. doi: 10.1023/a:1006319732410

Kim, W. C., Ko, J. H., and Han, K. H. (2012). Identification of a cis-acting regulatory motif recognized by MYB46, a master transcriptional regulator of secondary wall biosynthesis. *Plant Mol. Biol.* 78 (4-5), 489–501. doi: 10.1007/s11103-012-9880-7

Kranz, H. D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., et al. (1998). Towards functional characterisation of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. *Plant J.* 16 (2), 263–276. doi: 10.1046/j.1365-313x.1998.00278.x

Li, P., Chen, X., Sun, F., and Dong, H. (2017). Tobacco *TTG2* and *ARF8* function concomitantly to control flower colouring by regulating anthocyanin synthesis genes. *Plant Biol.* 19 (4), 525–532. doi: 10.1111/plb.12560

Li, Y., H.i, G., and Cun-kou, Q. (2013). Expression of lignin synthesis key genes of *F5H,4CL* and *COMT* on lodging resistance of *Brassica napusror* L. *Chin. J. Oil Crop Sci.* (in Chinese) 35 (1), 24–28. doi: 10.7505/j.issn.1007-9084.2013.01.004

Li, Y. C., and Qi, C. K. (2011). Lignin content and key gene expression in lignin synthesis of *Brassica napus* L. with lodging resistance. *Jiangsu J. Agric. Sci.* (in Chinese) 3), 481–487. doi: 10.3969/j.issn.1000-4440.2011.03.004

Li, J. B., and Qi, H. X. (2017). The advancement of genetic research for lodging resistance in rice. *Hubei Agri. Sci.* (in Chinese) 56 (23), 4450–4453. doi: 10.14088/j.cnki.issn.0439-8114.2017.23.005

Li, Z., Wang, Q. G., Pan, J. W., and Liu, W. (2018). Functional analysis of height regulation of cyptochrome P450 gene *OsDWARF48* in rice. *Shandong Agric. Sci.* (in Chinese) 50 (11), 1–9. doi: 10.14083/j.issn.1001-4942.2018.11.001

Lian, J., Lu, X., Yin, N., Ma, L., Lu, J., Liu, X., et al. (2017). Silencing of *BnTT1* family genes affects seed flavonoid biosynthesis and alters seed fatty acid composition in *Brassica napus*. *Plant Sci.* 254, 32–47. doi: 10.1016/j.plantsci.2016.10.012

Lin-Wang, K., Micheletti, D., Palmer, J., Volz, R., Lozano, L., Espley, R., et al. (2011). High temperature reduces apple fruit colour *via* modulation of the anthocyanin regulatory complex. *Plant Cell Environ.* 34 (7), 1176–1190. doi: 10.1111/j.1365-3040.2011.02316.x

Ma, Q. H. (2009). The expression of caffeic acid 3-o-methyltransferase in two wheat genotypes differing in lodging resistance. *J. Exp. Bot.* 60 (9), 2763–2771. doi: 10.1093/jxb/erp132

Machado, A., Wu, Y., Yang, Y., Llewellyn, D. J., and Dennis, E. S. (2009). The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *Plant J.* 59 (1), 52–62. doi: 10.1111/j.1365-313X.2009.03847.x

McCarthy, R. L., Zhong, R., Fowler, S., Lyskowski, D., Piyasena, H., Carleton, K., et al. (2010). The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. *Plant Cell Physiol.* 51 (6), 1084–1090. doi: 10.1093/pcp/pcq064

McCormac, A. C., Fischer, A., Kumar, A. M., Soll, D., and Terry, M. J. (2001). Regulation of *HEMA1* expression by phytochrome and a plastid signal during de-etiolation in *Arabidopsis thaliana*. *Plant J.* 25 (5), 549–561. doi: 10.1046/j.1365-313x.2001.00986.x

Mei, J., Liu, Y., Wei, D., Wittkop, B., Ding, Y., Li, Q., et al. (2015). Transfer of sclerotinia resistance from wild relative of *Brassica oleracea* into *Brassica napus* using a hexaploidy step. *Theor. Appl. Genet.* 128 (4), 639–644. doi: 10.1007/s00122-015-2459-3

Millar, A. A., and Gubler, F. (2005). The *Arabidopsis GAMYB-like* genes, *MYB33* and *MYB65*, are MicroRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* 17 (3), 705–721. doi: 10.1105/tpc.104.027920

Muleke, E., Wang, Y., Zhang, W.-t., Xu, L., Ying, J.-l, Karanja, B. K., et al. (2021). Genome-wide identification and expression profiling of MYB transcription factor genes in radish (*Raphanus sativus* L.). *J. Integ. Agr.* 20 (11), V–V. doi: 10.1016/S2095-3119(20)63308-1

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., et al. (1995). Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-myb. *Nat. Struct. Biol.* 2 (4), 309–320. doi: 10.1038/nsb0495-309

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., et al. (1994). Solutin structure of a specific DNA complex of the MYB DNA-binding domain with cooperative recognition helices. *Cell* 79 (4), 639–648. doi: 10.1016/0092-8674(94)90549-5

Paddock, T. N., Mason, M. E., Lima, D. F., and Armstrong, G. A. (2010). *Arabidopsis* protochlorophyllide oxidoreductase a (*PORA*) restores bulk chlorophyll synthesis and normal development to a porB porC double mutant. *Plant Mol. Biol.* 72 (4-5), 445–457. doi: 10.1007/s11103-009-9582-y

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A., and Saedler, H. (1987). The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* 6 (12), 3553–3558. doi: 10.1002/j.1460-2075.1987.tb02684.x

Peng, Y., Kui, L.-W., Cooney, J. M., Wang, T., Espley, R. V., and Allan, A. C. (2019). Differential regulation of the anthocyanin profile in purple kiwifruit (*Actinidia species*). *Hortic. Res.* 6, 3. doi: 10.1038/s41438-018-0076-4

Sperling, U., van Cleve, B., Frick, G., Apel, K., and Armstrong, G. A. (1997). Overexpression of light-dependent *PORA* or *PORB* in plants depleted of endogenous POR by far-red light enhances seedling survival in white light and protects against photooxidative damage. *Plant J.* 12 (3), 649–658. doi: 10.1046/j.1365-313x.1997.00649.x

Su, X., Xia, Y., Jiang, W., Shen, G., and Pang, Y. (2020). *GbMYBR1* from ginkgo biloba represses phenylpropanoid biosynthesis and trichome development in *Arabidopsis*. *Planta* 252 (4), 68. doi: 10.1007/s00425-020-03476-1

Tanaka, R., Yoshida, K., Nakayashiki, T., Masuda, T., Tsuji, H., Inokuchi, H., et al. (1996). Differential expression of two *HEMA* mRNAs encoding glutamyl-tRNA reductase proteins in greening cucumber seedlings. *Plant Physiol.* 110 (4), 1223–1230. doi: 10.1104/pp.110.4.1223

Underwood, W. (2012). The plant cell wall: A dynamic barrier against pathogen invasion. *Front. Plant Sci.* 3. doi: 10.3389/fpls.2012.00085

Wan, X. R., and Li, L. (2002). MYB proteins in plants. *Plant Physiol. Commun.* (in Chinese) 38 (2), 165–170. doi: 10.13592/j.cnki.ppj.2002.02.031

Wang, W. Q., Moss, S. M. A., Zeng, L., Espley, R. V., Wang, T., Kui, L.-W., et al. (2022). The red flesh of kiwifruit is differentially controlled by specific activation-repression systems. *New Phytol.* 235 (2), 630–645. doi: 10.1111/nph.18122

Wang, X. (2013). Cloning and functional analysis of Arabidopsis glucose-6-phosphate 1-dehydrogenase gene (*G6PD5*). [master's thesis]. [Haerbing]: Heilongjiang University. doi: 10.7666/d.Y2322537

Wang, X., Niu, Q. W., Teng, C., Li, C., Mu, J., Chua, N. H., et al. (2009). Overexpression of *PGA37/MYB118* and *MYB115* promotes vegetative-to-embryonic transition in arabidopsis. *Cell Res.* 19 (2), 224–235. doi: 10.1038/cr.2008.276

Wei, X., Shan, T., Hong, Y., Xu, H., Liu, X., and Zhang, Z. (2017). TaPIMP2, a pathogen-induced MYB protein in wheat, contributes to host resistance to common root rot caused by *Bipolaris sorokinian*. *Sci. Rep.* 7, 1754. doi: 10.1038/s41598-017-01918-7

Wu, Y. W., Ke, Y. Z., Ran, F., Wen, J., Guo, P. C., Sun, L. P., et al. (2018). Cloning and expression analysis of *BnMYB193* gene in *Brassica napus*. *J. Agric. Biotechnol.* (in Chinese) 26 (7), 1155–1163. doi: 10.3969/j.issn.1674-7968.2018.07.007

Xu, W., Grain, D., Le Gourrierec, J., Harscoet, E., Berger, A., Jauvion, V., et al. (2013). Regulation of flavonoid biosynthesis involves an unexpected complex transcriptional regulation of TT8 expression, in *Arabidopsis*. *New Phytol.* 198 (1), 59–70. doi: 10.1111/nph.12142

Xu, Y., Iacuone, S., Li, S. F., and Parish, R. W. (2014). *MYB80* homologues in *Arabidopsis*, cotton and *Brassica*: Regulation and functional conservation in tapetal and pollen development. *BMC Plant Biol.* 14, 278. doi: 10.1186/s12870-014-0278-3

Xu, W., Tang, W., Wang, C., Ge, L., Sun, J., Qi, X., et al. (2020). *SiMYB56* confers drought stress tolerance in transgenic rice by regulating lignin biosynthesis and ABA signaling pathway. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00785

Xu, P., Wu, L., Cao, M., Ma, C., Xiao, K., Li, Y., et al. (2021). Identification of MBW complex components implicated in the biosynthesis of flavonoids in woodland strawberry. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.774943

Yang, S. W., Jang, I. C., Henriques, R., and Chua, N. H. (2009). FAR-RED ELONGATED HYPOCOTYL1 and FHY1-LIKE associate with the *Arabidopsis* transcription factors LAF1 and HFR1 to transmit phytochrome a signals for inhibition of hypocotyl elongation. *Plant Cell* 21 (5), 1341–1359. doi: 10.1105/tpc.109.067215

Yao, L., Yang, B., Xian, B., Chen, B., Yan, J., Chen, Q., et al. (2020). The R2R3-MYB transcription factor BnaMYB111L from rapeseed modulates reactive oxygen species accumulation and hypersensitive-like cell death. *Plant Physiol. Bioch.* 147, 280–288. doi: 10.1016/j.plaphy.2019.12.027

Yin, N. W., Li, J. N., Liu, X., Lian, J. P., Fu, C., Li, W., et al. (2017). Lignification response and the difference between stem and root of *Brassica napus* under heat and drought compound stress. *Acta Agronomica Sin.* (in Chinese) 43 (11), 1689–1695. doi: 10.3724/SP.J.1006.2017.01689

Zhang, M., Mo, Z., Liao, J., Pan, S., Chen, X., Zheng, L., et al. (2021). Lodging resistance related to root traits for mechanized wet-seeding of two super rice cultivars. *Rice Sci.* 28 (2), 200–208. doi: 10.1016/j.rsci.2021.01.009

Zhao, K., and Bartley, L. E. (2014). Comparative genomic analysis of the R2R3 MYB secondary cell wall regulators of *Arabidopsis*, poplar, rice, maize, and switchgrass. *BMC Plant Biol.* 14, 135. doi: 10.1186/1471-2229-14-135

Zhong, R., Lee, C., Zhou, J., McCarthy, R. L., and Ye, Z. H. (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell* 20 (10), 2763–2782. doi: 10.1105/tpc.108.061325

Zhong, R., Richardson, E. A., and Ye, Z. H. (2007). The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. *Plant Cell* 19 (9), 2776–2792. doi: 10.1105/tpc.107.053678

Zhong, R., and Ye, Z.-H. (2012). MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. *Plant Cell Physiol.* 53 (2), 368–380. doi: 10.1093/pcp/pcr185

Zhu, X. B., Zhang, G. F., and Chen, P. (2017). Research progress in the transcriptional regulation of secondary cell wall thickening(Review). *Plant Physiol. J.* (in Chinese) 53 (9), 1598–1608. doi: 10.13592/j.cnki.ppj.2017.0203

Zhuang, W., Shu, X., Lu, X., Wang, T., Zhang, F., Wang, N., et al. (2021). Genome-wide analysis and expression profiles of PdeMYB transcription factors in colored-leaf poplar (*Populus deltoids*). *BMC Plant Biol.* 21 (1), 1–432. doi: 10.1186/s12870-021-03212-1

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 June 2023
DOI 10.3389/fpls.2023.1211040

OPEN ACCESS

EDITED BY
Liezhao Liu,
Southwest University, China

REVIEWED BY
Sayedehsaba Bilgrami,
Southwest University, China
Chaobo Tong,
Oil Crops Research Institute, Chinese Academy of Agricultural Science, China

*CORRESPONDENCE
Junsheng Zhao
zhaojunshengsd@163.com
Xiurong Zhang
zhangxr@oilcrops.cn

RECEIVED 24 April 2023
ACCEPTED 07 June 2023
PUBLISHED 22 June 2023

CITATION
Zhang Y, Gong H, Cui X, Gao C, Li N, Pu Y, Zhang X and Zhao J (2023) Integrated lipidomic and transcriptomic analyses reveal the mechanism of lipid biosynthesis and accumulation during seed development in sesame.
Front. Plant Sci. 14:1211040.
doi: 10.3389/fpls.2023.1211040

Integrated lipidomic and transcriptomic analyses reveal the mechanism of lipid biosynthesis and accumulation during seed development in sesame

Yujuan Zhang[1], Huihui Gong[1], Xinxiao Cui[1], Chunhua Gao[1], Nana Li[2], Yanyan Pu[2], Xiurong Zhang[1*] and Junsheng Zhao[1*]

[1]Institute of Industrial Crops, Shandong Academy of Agricultural Sciences, Jinan, China, [2]Institute of Crop Germplasm Resources, Shandong Academy of Agricultural Sciences, Jinan, China

Sesame is one of the most important oilseed crops and attracts significant attention because of its huge nutritional capacity. However, the molecular mechanisms underlying oil accumulation in sesame remains poorly understood. In this study, lipidomic and transcriptomic analyses in different stages of sesame seed (Luzhi No.1, seed oil content 56%) development were performed to gain insight into the regulatory mechanisms that govern differences in lipid composition, content, biosynthesis, and transport. In total, 481 lipids, including fatty acids (FAs, 38 species), triacylglycerol (TAG, 127 species), ceramide (33 species), phosphatidic acid (20 species), and diacylglycerol (17 species), were detected in developing sesame seed using gas and liquid chromatography-mass spectrometry. Most FAs and other lipids accumulated 21–33 days after flowering. RNA-sequence profiling in developing seed highlighted the enhanced expression of genes involved in the biosynthesis and transport of FAs, TAGs, and membrane lipids, which was similar to that seen during lipid accumulation. Through the differential expression analysis of genes involved in lipid biosynthesis and metabolism during seed development, several candidate genes were found to affect the oil content and FA composition of sesame seed, including *ACCase*, *FAD2*, *DGAT*, *G3PDH*, *PEPCase*, *WRI1* and *WRI1-like* genes. Our study reveals the patterns of lipid accumulation and biosynthesis-related gene expression and lays an important foundation for the further exploration of sesame seed lipid biosynthesis and accumulation.

KEYWORDS

sesame, seed development, lipidome, transcriptome, fatty acid, lipid biosynthesis

Introduction

Vegetable oils are not only an essential part of the human daily diet and play an important role in nutrition but are also important contributors to chemical feedstocks and renewable biofuels (Dyer et al., 2008; Lummiss et al., 2012). In nature, some plants can synthesize a large amount of oil in seed, pollen, some fleshy fruits, and other organs, and store them in the form of triacylglycerol (TAG), which mainly include oilseed crops (e.g., soybean, rapeseed, peanut, sesame, and sunflower), perennial woody oil plants (e.g., oil palm, coconut, oil-tea camellia, and walnut), and some fiber crops such as cotton, flax, and hemp, among others (Zhang et al., 2009). Recently, with the improvement in living standards and the development of the economy, the worldwide consumption of vegetable oils is increasing year by year and there is more demand for high-quality edible oils (Lu et al., 2011). Therefore, it would be of great scientific and practical significance to elucidate the molecular mechanism of the biosynthesis and regulation of oil in plants.

In seed, oil mostly occurs in the form of triacylglycerol (TAG) which is composed primarily of three fatty acids (FAs) esterified to a glycerol backbone and stored in cytosolic oil bodies or liposomes (Kelly et al., 2013b; Bansal et al., 2018). In recent years, the molecular mechanism of the biosynthesis and regulation of oil in plants have been studied intensively (Harwood, 1996; Bates et al., 2013; Li-Beisson et al., 2013; He and Ding, 2020; Yang et al., 2022). Lipid biosynthesis and accumulation is a complex process in plant cells. Lipids are formed by *de novo* synthesis in the plastid by elongation and desaturation of FAs; assembly of TAGs occurs in the endoplasmic reticulum (ER); and finally, the oil body forms in the seed (Harwood, 1996; Bates et al., 2013; He and Ding, 2020). Many genes and gene products involved in controlling oil biosynthesis, including enzymes, directly participate in FA and TAG biosynthetic pathways. In *Arabidopsis*, TAG biosynthesis and its regulation has been well characterized. Acyl-lipid metabolism requires 120 enzymatic reactions and more than 600 genes to encode proteins and regulatory factors (Li-Beisson et al., 2013). Based on the genes that encode the key enzymes that participate in lipid biosynthetic and metabolic pathways in *Arabidopsis*, many homologous genes have been cloned and validated their functions in oilseed crops, such as *ACS* (Ding et al., 2020), *SAD* (Knutzon et al., 1992), *LPAAT* (Bourgis et al., 1999), *DATG* (Chen et al., 2016), and *FAD* (Haun et al., 2014). In addition, several crucial transcription factors (TFs), have been identified and play important roles in the process of seed maturation and lipid accumulation. Examples include the LEC (Tan et al., 2011), WRI (Chen et al., 2020), MYB (Liu et al., 2014), bZIP (Song et al., 2013), DOF (Wang et al., 2007), as well as NFYA TF families (Kelly et al., 2013a). Moreover, several new genes that regulate lipid biosynthesis in oilseed crops have also been identified (Elhiti et al., 2012; Liu et al., 2012; Kelly et al., 2013a).

Sesame (*Sesamum indicum* L.) is an important oil and cash crop worldwide, and is grown widely in several countries in Asia and Africa such as China, India, and Sudan due to its high oil content and quality (Zhang et al., 2013; Qureshi et al., 2022). Their seed oil contains a huge amount of unsaturated FAs, such as linoleic acid, oleic acid, and alpha-linolenic acid, which determine their important nutritional and economic values (Wei et al., 2015; Melo et al., 2021). Previous research has found that sesame oil contains a stably high proportion of unsaturated FAs, ranging from 83% to 87% (Wei et al., 2015). Furthermore, as a rich source of natural antioxidants, such as sesamin and sesamol, sesame seed is an important source of oil, with high nutritional value, and the global demand for various sesame products is increasing (Figueiredo et al., 2017). Therefore, improving the oil content and quality of sesame seed is an important and ongoing objective for sesame breeding. *SiNST1* and *SiPPO*, as well as several other genes (including *SiACNA*, *SiDGAT2*, *SiFATA*, *SiFATB*, and *SiSAD*), detected as candidate genes that control oil content in sesame were found to be involved in lipid metabolism (Wei et al., 2015). Despite this, no specific function has been clarified for these genes, and our knowledge of the biosynthesis and regulation of oil in sesame is still limited. The correlation between changes in transcriptome and lipid accumulation during seed development in sesame has not been explored.

Here, to understand the lipid composition during sesame seed development and the regulatory mechanism that governs the process of lipid accumulation in sesame, both transcriptomic and lipidomic profiling of sesame seed at three different developmental stages were conducted. This study aimed to investigate the network and candidate genes highly associated with sesame seed oil synthesis and to provide a better theoretical basis for the molecular improvement and genetic breeding of sesame and other oilseed crops.

Materials and methods

Sample preparation

Sesame accession Luzhi No.1, which has a seed oil content of 56%, was used in this study. The sesame was grown in the experimental field of the Shandong Academy of Agricultural Sciences. Seeds from capsules at the middle part of the sesame plants were harvested at 9, 21, and 33 days (d) after flowering, representing early, middle, and late stages of Luzhi No.1 seed maturation, respectively. All seed samples with three biological replicates were flash-frozen in liquid nitrogen and stored at −80°C.

Fatty acid determination

Determination of FA samples of Luzhi No.1 at 9, 21, and 33 d was performed by gas chromatography coupled with mass spectrometry (GC-MS) at Shanghai Applied Protein Technology Company, according to a method described previously (Zhu et al., 2019). In brief, seed samples (50 mg) were frozen and ground in liquid nitrogen. A solution of chloroform:methanol (3:2, v/v) was used to extract FAs. After ultrasonication for 30 min, 2 mL of 1% sulfuric acid in methanol were added to the supernatant. For methyl esterification, the mixture was incubated in an 80°C water bath for 30 min to achieve FA esterification. Then, 1 mL n-hexane and 5 mL

water were added and vortex-mixed. The supernatant (500 µL) was spiked with an internal standard (25 µL of 500ppm methyl salicylate), mixed, and subjected to GS-MS using an Agilent Model 7890A/5975C GC-MS system with an Agilent DB-WAX capillary GC column (30 m × 0.25 mm × 0.25 µm).

Conditions for chromatography were as follows: the initial temperature was 50°C and was maintained for 3 min. The temperature was then increased to 220°C at 10°C/min, and maintained at 220°C for 20 min. The carrier gas was helium (1.0 mL/min). A quality-control sample was used for testing and evaluating the stability and repeatability of the system. The temperatures of the injection port and transmission line were 280°C and 250°C, respectively. The electron bombardment ionization source, SIM (Selected ion Monitor) scanning mode, and electron energy were 70 eV. To quantify medium- and long-chain FA, MSD ChemStation software was used to extract retention time and chromatographic peak area, and Supelco 37-component FAME (FA methyl ester) mix (Sigma-Aldrich) was used to construct a calibration curve for the concentration range of 0.5–1000 mg/L. FA content was expressed as µg g^{-1} fresh weight (FW). One-way ANOVA was used to compare data between multiple groups, followed by the Student *t*-test to compare data between two groups. *P*-values below 0.05 were deemed statistically significant.

Lipid extraction and targeted lipidomic analysis

Total lipid was extracted from samples of Luzhi No.1 at 9, 21, and 33 d, which represents early, mid, and late phases of seed development in sesame, according to a method reported previously (Taguchi and Ishikawa, 2010). Lipid measurements and data analysis were performed by high-performance liquid chromatography-mass spectrometry (LC-MS), according to the Shanghai Applied Protein Technology Company (Shanghai, China) standard procedures (Zhu et al., 2019). A methyl tert-butyl ether method was used for sample preparation and lipid extraction. In brief, seed material (40 mg) was accurately weighed and spiked with appropriate amounts of internal lipid standards (SPLASH® LIPIDOMIX MASS SPRC STANDARD, AVANTI,330707-1EA). After homogenizing with 200 uL distilled water, 240 uL pre-cooled methanol, and 800 uL methyl tert-butyl ether, a gentle stream of nitrogen was used to dry the total lipid extract. Quality control (QC) samples were prepared by pooling and mixing 10 uL of each sample. These were run at the beginning of the sample queue to condition the column, and every 10 injections thereafter to assess inconsistencies, which are particularly evident in large batch acquisitions in terms of retention time drifts and variation in ion intensity over time.

For LC-MS/MS, reverse phase chromatography was selected for LC separation using a CSH C18 column (1.7 µm, 2.1 mm × 100 mm, Waters). The lipid extracts were re-dissolved in 200 µL of 90% isopropanol/acetonitrile, centrifuged at 14,000× g for 15 min, and finally 3 µL of sample was injected. Solvent A was acetonitrile–water (6:4, vol/vol) containing 0.1% formic acid and 0.1 mM ammonium formate, and solvent B was acetonitrile–isopropanol (1:9, vol/vol) containing 0.1% formic acid and 0.1 mM ammonium formate. The gradient elution was programmed as follows: the initial mobile phase contained 30% solvent B and 70% solvent A at a flow rate of 300 µL/min. It was held for 2 min, and then solvent B was linearly increased to 100%, while solvent A was decreased to 0% from 2–25 min. This was followed by equilibration in 5% solvent B and 95% solvent A from 25–35 min. Mass spectra were acquired using a Q-Exactive Plus system in positive and negative modes, respectively. ESI (Electron Spray Ionization) parameters were optimized and preset for all measurements as follows: source temperature, 300°C; capillary temperature, 350°C; ion spray voltage, 3000 V, S-Lens RF Level was set at 50%. The MS1 scan range was 200-1800 m/z. The mass:charge ratios were determined for lipid molecules and fragments; 10 fragments (MS2 scan, HCD) were collected after each full scan. MS1 occurred at a resolution of 70 000 at m/z 200 and MS2 occurred at a resolution of 17 500 at m/z 200.

Lipidsearch (Thermo Fisher Scientific, USA) was used to extract and identify the lipid molecule and internal standard peaks. The main parameters were as follows: precursor tolerance: 5 ppm; product tolerance: 5 ppm; and product ion threshold: 5%.

Data processing and statistical analysis

Lipid species were identified using LipidSearch software version 4.2 (Thermo Scientific™) to process the raw data and for peak alignment, retention time correction, and peak area. LipidSearch contains data on more than 30 lipid classes, with information on more than 1,700,000 ion fragments. In the data extracted from LipidSearch, the ion peak with > 50% missing values were removed, and the Perato scaling method was used for normalization (Want et al., 2010). The absolute amount of each lipid ion was calculated using a single-point calibration method, as described previously (Want et al., 2010). Lipid content was expressed as µg g-1 FW. One-way ANOVA was used to compare data between multiple groups, followed by Student's *t*-test to compare data between two groups. *P*-values below 0.05 were deemed statistically significant.

RNA extraction and transcriptomic analysis

Total RNA from fresh seed of Luzhi No.1 was extracted using an EASYspin Plus kit (Aidlab, Beijing, China), according to the manufacturer's instructions. RNA quality was assessed using the Agilent Bioanalyzer. RNA-Seq libraries were prepared and sequenced on an Illumina Hiseq2500 platform at Shanghai Applied Protein Technology Company and subsequent data processing was performed as described previously (Zhu et al., 2019). Raw reads of fastq format were processed through in-house perl scripts and clean reads were obtained by removing those containing adapters, poly-Ns, and low-quality reads from raw data. The clean reads were mapped to the sesame reference genome v.1.0 (https://www.ncbi.nlm.nih.gov/genome/?term=sesamum) using HISAT2 (v2.0.5). Gene expression level for each sample was normalized by calculating fragments per kilobase of transcript per million fragments mapped (FPKM). The difference

between transcriptome samples was performed using the DESeq2 and genes with adjusted *P*-value < 0.05 and |log2foldchange| >1 were assigned as differentially expressed. The KEGG pathway enrichment analysis was performed for differentially expressed genes (DEGs) using the clusterProfiler R package with a corrected *P*-value < 0.05 regarded as significantly enriched.

qRT-PCR

The qRT-PCR was performed with ChamQTM SYBR qPCR Master Mix (Vazyme Biotech, Nanjing, China) on a LightCycler480 Real-Time PCR System, according to a method described previously (Zhang et al., 2019). The gene primers for qRT-PCR are listed in Supplementary Table S1. The sesame *actin7* gene was used as an endogenous control gene to normalize all target gene expression (Wei et al., 2015).

Statistical analysis

For statistical analysis, all experiments were performed at least three times. Principal component analysis (PCA) and Pearson correlation were performed using the R statistical environment (version 3.5.1).

Results

Patterns of FA accumulation in developing sesame seed

The oil content of the sesame seed at 9, 21, and 33 days after flowering was approximately 6%, 45%, and 56%, respectively. It is therefore crucial to understand the characteristics of FA accumulation and identify the correlations between the FAs during seed development in sesame. The changes in FA composition and level in developing seeds were analyzed by GC-MS. In total, 38 FA species, including medium- and long-chain FAs were identified. FA composition did not change significantly across developmental stages (Table S2). At the beginning of storage, all FA species showed lower levels, however, 16 and 26 FA species significantly increased at 21 and 33 d, respectively (Table S2). At both 21 and 33 d, palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1n9), and linoleic acid (C18:2n6) were the most abundant FAs, accounting for approximately 95% of the total FAs (Figures 1A, B). Phenomenally, linoleic acid (C18:2n6) and stearic acid (C18:0) were increased (percentage of total FAs) during seed development, while palmitic acid (C16:0) and linolenic acid C18:3n3 showed the opposite trend (Figure 1B). Although the content of oleic acid (C18:1n9) in 21 and 33 d continued to increase, its relative ratio of total FAs did not significantly change during the developmental period examined (Figure 1B). Moreover, it was observed that more than 72% of the total FAs, including monounsaturated FA (MUFA), polyunsaturated FA (PUFA), and saturated FA (SFA), accumulated strongly from 21 to 33 days after flowering, which suggested the critical period for rapid accumulation of FAs is in the 21–33 d period of storage (Figure 1C).

Altered profiles of lipids across developmental stages of sesame seed

To further understand the changes in other lipids during seed development, LC-MS was used to obtain lipidomic profiling of the sesame seed at 9, 21, and 33 d after flowering. The PCA of the lipidomic data showed that biological replicate samples were closely clustered together, indicating the experiment had good repeatability (Figure S1). The LC-MS analysis indicated that a total of 25 lipid classes containing 443 lipid species were detected in all samples of developing seed. Among them, 358 lipid species were differentially accumulated at 21 or 33 d, however, their composition was not changed from 9 to 33 d after flowering (Table S3). Compared to 9 d, quantification data reveal that the content of most lipid classes increased at 21 and 33 d, among which were triacylglycerol (TAG, 127 species), ceramide (Cer, 33 species), phosphatidic acid (PA, 20 species), diacylglycerol (DAG, 17 species), and so on (Table S3: Figure 2). In particular, the contents of total PA at 9, 21, and 33 d accounted for 81%, 62%, and 50% of the total lipids respectively, which were higher than other lipid classes (Figure 2).

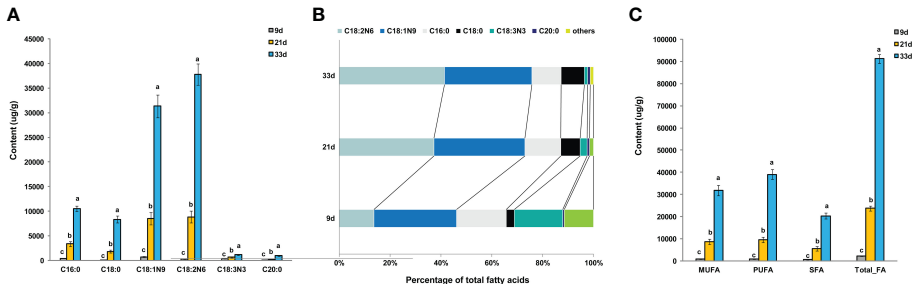


FIGURE 1
Analysis of FAs in developing sesame seed. **(A)** The levels of the most abundant FAs species. **(B)** Percentage composition of the main FA species at different developmental periods. **(C)** Saturated and unsaturated FA levels. n3, n6, and n9, the first double bonds in unsaturated FAs, occur at the 3, 6, and 9 positions of the methyl end of the carbon chain, respectively; FA, fatty acid; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids; SFA, saturated fatty acid. Different lowercase letters indicate a significant difference between the groups according to Student's *t*-test ($P < 0.05$).

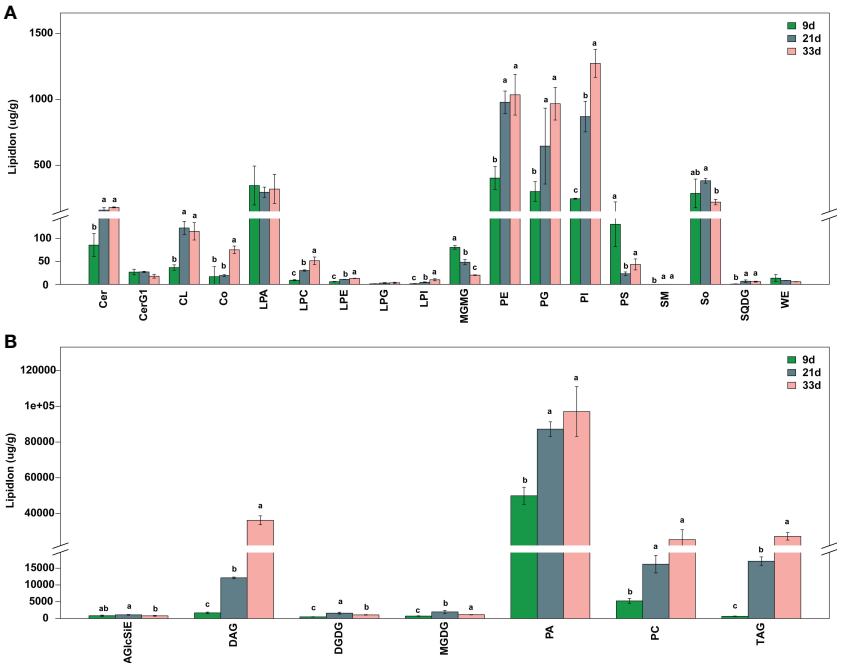


FIGURE 2
Total contents of lipid classes in seed at different developmental stages. **(A)** Low-abundance lipid species. **(B)** High-abundance lipid species. AGlcSiE, acylglcsitosterol ester; Cer, ceramide; CerG1, simple glc series; CL, cardiolipin; Co, coenzyme; DAG, diglyceride; DGDG, digalactosyldiacylglycerol; LPA, lysophosphatidic acid; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; LPG, lysophosphatidylglycerol; LPI, lysophosphatidylinositol; MGDG, monogalactosyldiacylglycerol; MGMG, monogalactosyldiacylglycerol; PA, phosphatidic acid; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PS, phosphatidylserine; SM, sphingomyelin; So, sphingosine; SQDG, sulfoquinovosyldiacylglycerol; TAG, triacylglycerol; WE, wax esters. Different lowercase letters indicate a significant difference between the groups according to Student's *t*-test ($P < 0.05$).

The accumulation of lipid species in the developing sesame seed was complicated, and the significant changes in pairwise comparisons between three developmental stages are shown in Figure 3. The accumulation of most phospholipids rapidly increased at 21 d; however, some phospholipids decreased or did not change at 33 d (Figure 3; Table S3). Among 35 differentially accumulated PAs, 3 PA species showed higher levels (>10000 μg g^{-1}) in at least one developmental period of sesame seed, including 16:0/18:1, 16:0/18:2, and 18:2/18:2 PAs (Figure 3; Table S3). The contents of 16:0/18:1, 16:1/18:1, 18:0/16:1, 33:4 and 36:1 phosphatidylcholines (PCs) increased significantly at 21 and 33 d (Figure 3). In particular, 16:0/18:1 PC showed the highest values in all samples and was a major contributor to the rise in total PCs (Table S3). Compared to PAs and PCs, the levels of phosphatidylethanolamines (PEs), phosphatidylglycerols (PGs), phosphatidylinositols (PIs), and phosphatidylserines (PSs) were lower in developing seed (Table S3). For the DAGs and TAGs, the contents of most species steadily increased across developmental stages (Figure 3; Table S3). A total of 24 DAGs and 132 TAGs species differentially accumulated during seed development. In particular, the accumulation of 18:2/18:2, 18:1/18:2 DAG, and 18:1/18:2/18:2, 18:2/18:2/18:2 TAG was prominent at 21 and 33 d, which showed higher levels in developing seed (Figure 3; Table S3). Although 56 Cer, 9 AcylGlcSitosterol esters (AGlcSiEs) and 10 and 13 major galactolipid species, digalactosyldiacylglycerols (DGDGs) and monogalactosyldiacylglycerol (MGMG), respectively, showed significant changes, their levels were lower during sesame seed development (Figure 3; Table S3).

Transcriptome analysis of developing sesame seed

During sesame seed development, a transcriptome profile was examined for developing seed at 9, 21, and 33 d after flowering with three biological replicates. For each library, raw reads, clean reads, clean bases, Q20 (%), Q30 (%), GC (%), and the mapped percentage were recorded in Table S4. The Pearson correlation analysis showed that all libraries from the biological replicates were highly related, therefore, the quality of the samples in this sequencing was high enough for gene expression analysis (Figure S2). To investigate the dynamic changes in sesame transcriptome during seed development, pairwise comparisons between three developmental stages were performed. The gene expression levels were assessed by FPKM. A total of 10932 DEGs were identified between three developmental stages, and 6667, 9523, and 4551 DEGs were found at 21 vs. 9 d, 33 vs. 9 d, and 33 vs. 21 d, respectively (Figure 4). It was notable that the comparison of 33 vs. 9 d had the largest number of DEGs, with 4173 and 5350 genes exhibiting significantly up or down-regulated expression, respectively (Figure 4A). Interestingly, out of all the DEGs, 1926 DEGs were significantly differentially expressed in the two pairwise comparisons between the three developmental stages. KEGG pathway enrichment analysis showed that DEGs were mainly enriched in lipid biosynthesis and metabolic, carbohydrate, and energy metabolism pathways (Table S5), in particular, in "fatty acid biosynthesis (sind00061)", "fatty acid elongation (sind00062)",

10.3389/fpls.2023.1211040

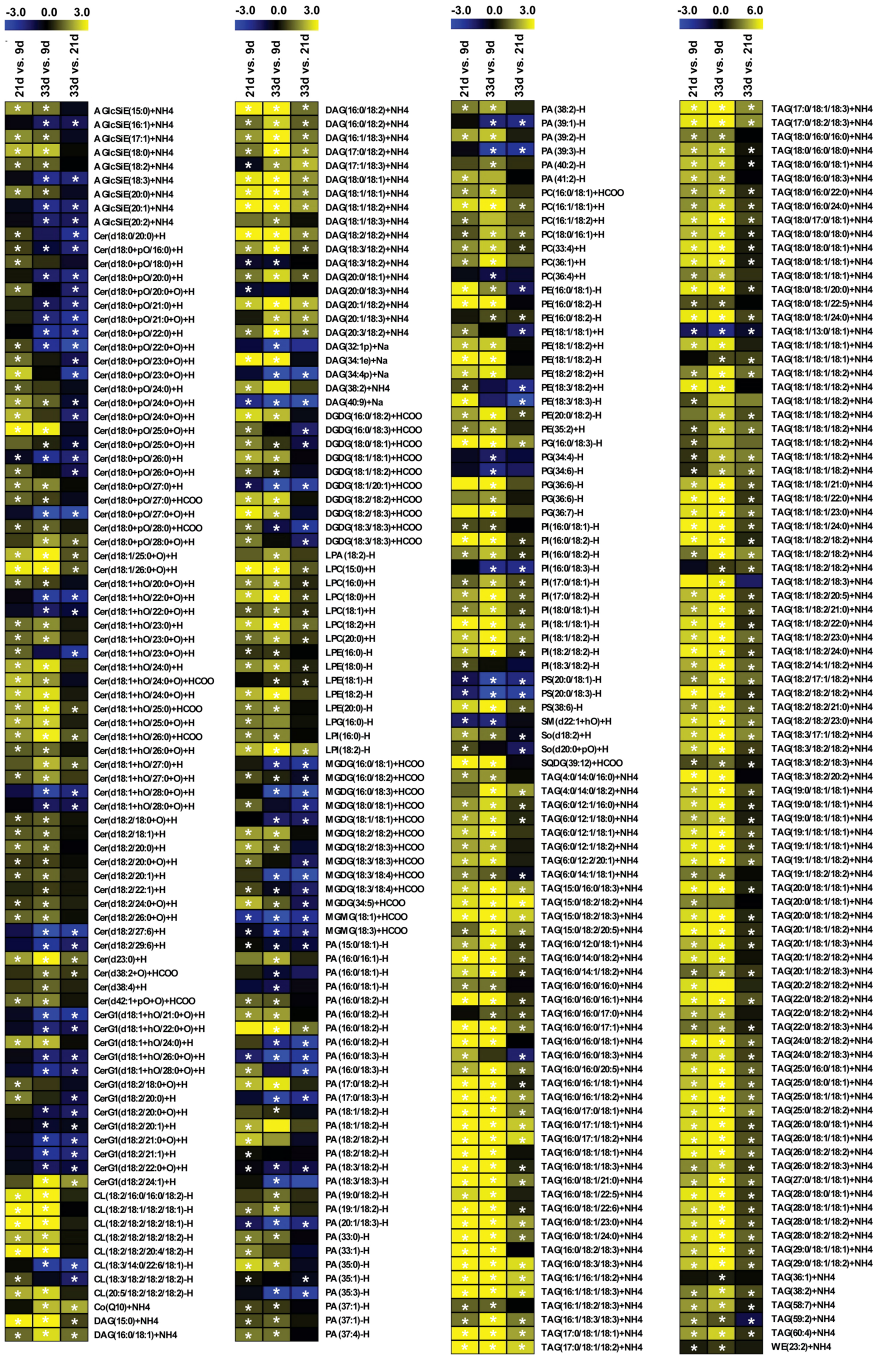


FIGURE 3
Heatmaps of content changes of lipids in developing seed. Yellow denotes increased lipid content and blue denotes decreased lipid content. AGlcSiE, acylglcsitosterol ester; Cer, ceramide; CerG1, simple glc series; CL, cardiolipin; Co, coenzyme; DAG, diglyceride; DGDG, digalactosyldiacylglycerol; LPA, lysophosphatidic acid; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; LPG, lysophosphatidylglycerol; LPI, lysophosphatidylinositol; MGDG, monogalactosyldiacylglycerol; MGMG, monogalactosyldiacylglycerol; PA, phosphatidic acid; PC, phosphatidyl choline; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PS, phosphatidylserine; SM, sphingomyelin; So, sphingosine; SQDG, sulfoquinovosyldiacylglycerol; TAG, triacylglycerol; WE, wax esters. Heat maps were constructed based on the log_2FoldChange levels for each lipid between 21, 33, and 9 d. Asterisks (*) indicate that mean values are significantly different between 9, 21, and 33 d, according to Student's *t*-test ($P < 0.05$).

"fatty acid metabolism (sind01212)", "biosynthesis of unsaturated fatty acids (sind01040)", "alpha-linolenic acid metabolism (sind00592)", "glycerolipid metabolism (sind00561)", "phosphatidylinositol signaling system (sind04070)", "glycosphingolipid biosynthesis (sind00603)", and "glycolysis/gluconeogenesis (sind00010)".

Analyses by qRT-PCR were performed on 10 selected DEGs related to lipid biosynthesis and metabolism. In all these genes,

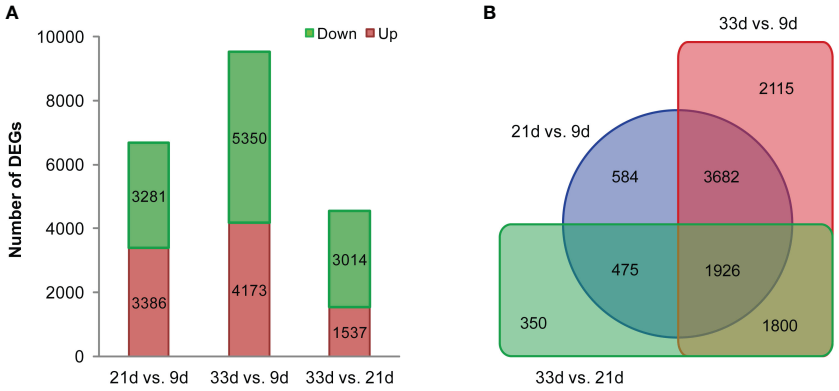


FIGURE 4
Numbers of differentially expressed genes (DEGs). **(A)** Pairwise comparisons for each stage of development. **(B)** Venn diagram of DEGs during seed development in sesame.

there was a very good correlation between RNA-seq and qRT-PCR data, indicating that the RNA-seq data are reliable (R^2 = 0.83, Figure S3).

Expression pattern of genes related to FA biosynthesis and TAG assembly

Schematic and expression patterns of DEGs involved in FA biosynthesis and TAG assembly in sesame seed at different developmental stages are shown in Figure 5 and Table S6, respectively. Acetyl-CoA carboxylase (ACCase, EC:6.4.1.2), the first key rate-limiting enzyme in FA biosynthesis, catalyzes the formation of malonyl-CoA from acetyl-CoA in plastids (Figure 5). In this study, nine putative members of *ACCase* gene showed the highest expression level at 21 d of seed development (Table S6). Note that another *ACCase* gene, LOC105161564, was highly expressed continuously during seed development (Table S6). The soluble stearoyl-ACP desaturase (SAD, EC:1.14.19.2) is membrane-bound and catalyzes the key step in unsaturated fat biosynthesis, and we found that the expression of five *SADs* was dramatically up-regulated in 21 and 33 d samples (Figure 5; Table S6). The FA desaturase (FAD, EC:1.14.19.6/1.14.19.22), catalyzes the introduction of double bonds at specific positions of FAs and thus the ratio of saturated to unsaturated FAs in seed (Figure 5). Expression of *FAD2* (LOC105159686) was significantly down-regulated at 33 d while two *FAD3* members (LOC105160895, LOC105167042) were significantly down-regulated at 21 and 33 d of seed development, which is consistent with the C18:2 and C18:3 accumulation patterns in developing sesame seed (Table S6; Figure 1B). These findings revealed that the regulation of *SAD* and *FAD2* levels during the middle and later stages of seed development may have a crucial role in the accumulation of oleic acid (C18:1n9) and linoleic acid (C18:2n6). Generally, the majority of the genes involved in FA biosynthesis showed higher levels of expression at 21 and 33 d after flowering, which is consistent with the changes in the accumulation of non-esterified FAs (Table S6). In addition, other important genes encoding enzymes involved in TAG biosynthesis, such as glycerol-3-phosphate acyltransferase (GPAT, EC:2.3.1.15), lysophospholipid acyltransferase (LPAT, EC:2.3.1.51), phosphatidate phosphatase (PAP, EC:3.1.3.81/3.1.3.4), and diacylglycerol acyltransferase (DGAT, EC:2.3.1.15), have also been detected in this study (Figure 5; Table S6). Higher transcript levels of *GPATs*, *PAPs*, *DGATs*, and two *LPAT* members were observed at 33 d after flowering (Table S6) and matched well with the change in the content of some DAG and TAG species (Figure 3).

Expression pattern of genes related to carbohydrate metabolism

Glycerol-3-phosphate (G-3-P) and acetyl-CoA produced by glycolysis/gluconeogenesis are the raw materials for FA biosynthesis and TAG assembly, indicating that carbohydrate metabolism may have important implications for the improvement of oil content in plants. According to KEGG pathway enrichment analysis, several DEGs involved in carbon metabolism (sind01200), glycolysis/gluconeogenesis (sind00010), galactose metabolism (sind00052), starch and sucrose metabolism (sind00500), and pyruvate metabolism (sind00620) were identified during sesame seed development (Table S5). In plants, glycerol-3-phosphate dehydrogenase (G3PDH, EC:1.1.1.8) catalyzes dihydroxyacetone phosphate to form G-3-P. In general, the expression of four *G3PDH* members was distinctly up-regulated at 21 and 33 d, with LOC105176298 showing the highest transcript level (Table S7). Seven genes with generally higher transcript abundance in seed 21 d after flowering encode pyruvate dehydrogenases (PDHCs), which catalyze the synthesis of acetyl-CoA from pyruvate (Table S7). Notably, four genes encoding phosphoenolpyruvate carboxylase (PEPCase, EC:4.1.1.31) were found to be differentially expressed during sesame seed development, with *PEPCase 2* and *PEPCase 4* significantly down-regulated while *PEPCase* and *PEPCase 1-like* were significantly up-regulated (Table S7). Pyruvate kinase (PK, EC:2.7.1.40) is a key glycolytic regulatory enzyme, converting phosphoenol-pyruvate

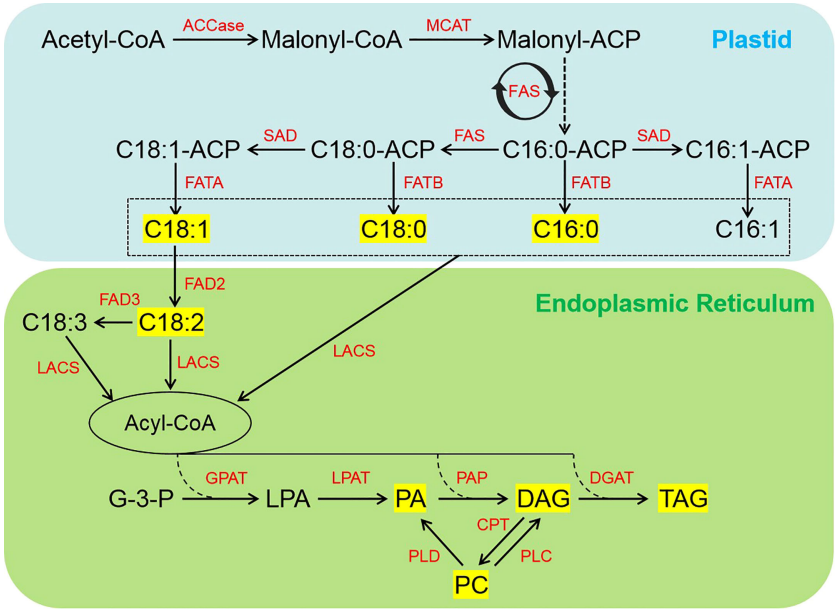


FIGURE 5
Simplified schematic of FA biosynthesis and TAG assembly pathways during seed development in sesame. The yellow box indicates lipids with content significantly higher in developing sesame seed. ACCase, acetyl-CoA carboxylase; CPT, diacylglycerol cholinephosphotransferase; DAG, diacylglycerol; DGAT, diacylglycerol acyltransferase; FA, fatty acid; FAS, fatty acid synthase; FATA, acyl-ACP thioesterase A; FATB, acyl-ACP thioesterase B; FAD2, fatty acid desaturase 2; FAD3, fatty acid desaturase 3; G-3-P, glycerol-3-phosphate; GPAT, glycerol-3-phosphate acyltransferase; LACS, long-chain acyl-CoA synthetase; LPA, lysophosphatidic acid; LPAT, lysophosphatidic acid acyltransferase; MCAT, malonyl CoA-ACP transacylase; PA, phosphatidic acid; PAP, phosphatidic acid phosphatase; PC, phosphatidylcholine; PLC, phospholipase C; PLD, phospholipase D; SAD, stearoyl-ACP desaturase; TAG, triacylglycerol.

into pyruvate. The expression of a gene (LOC105161300) encoding PK was found to be increased persistently during sesame seed development, whereas four other *PK* genes were significantly up-regulated only in 21 d (Table S7).

Differentially expressed TFs genes involved in lipid biosynthesis

To obtain a comprehensive understanding of the transcriptional regulation of genes involved in sesame seed development, the differentially expressed TFs related to lipid biosynthesis were identified using Plant TFDB and Pfam databases (Figure 6A; Table S8). In this study, a total of 492 TF genes belonging to 33 TF families were identified and the top 15 TF families are shown in Figure 6A. Of these, the AP2/ERF family had the largest number of members (109, 22%), followed by bHLH (67, 14%), MYB (49, 10%), and WRKY (49, 10%; Figure 6A). Our findings also showed that there were five TF family members with high expression levels (FPKM>1000) at least at the early, mid, or late stages of seed development (Table S8). Notably, there were a total of 11 AP2/ERF members, including two WRI1 genes, that showed higher expression levels at 21 or 33 d, and their expression pattern was quite similar to many genes related to lipid biosynthesis (Figure 6B). Moreover, three MYB and 13 WRKY genes showed higher transcript abundance at 9 d, while lower transcript levels were observed in developing sesame seed at 21 and 33 d (Figure 6B).

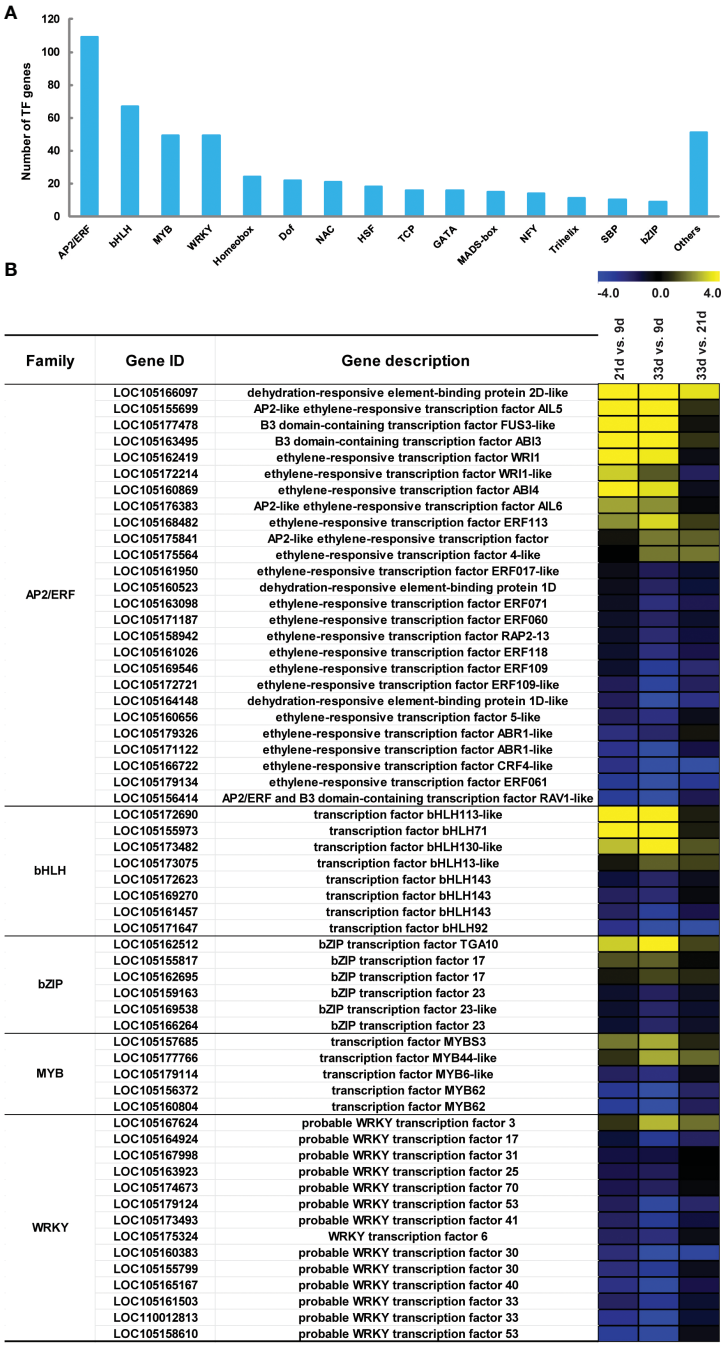
Discussion

As oils are an important source of high-density energy and a source of essential FAs in oilseed crops, they are mainly stored in seed. In comparison to other oilseed crops, sesame seed has a relatively high oil content. Lipid accumulation during sesame seed development is a complex process that requires further investigation to uncover the underlying mechanisms (Wang et al., 2014). This study integrated lipidomic and transcriptomic analyses of developing sesame seed to uncover the overall lipid biosynthesis process and several important biosynthesis pathways and candidate genes involved in the lipid composition and content of sesame seed.

Characteristics of lipids accumulation in developing sesame seed

Oil seed plants accumulate stored lipids primarily through FA synthesis in plastids and TAG synthesis in their ERs through a series of enzyme-catalyzed reactions (Harwood, 1996; Bates et al., 2013). When we measured sesame seed FA composition and content during seed development by GC-MS, we found significant differences in FA content. In line with previous findings in *Brassica napus*, *Glycine max* and *Camellia oleifera*, FA accumulation was slower at the early stage and quicker at the middle and late stages (Woodfield et al., 2017; Lin et al., 2018; Woodfield et al., 2018; Yang et al., 2019). For instance, palmitic acid, stearic acid, oleic acid, and linoleic acid in the

10.3389/fpls.2023.1211040



Family	Gene ID	Gene description
AP2/ERF	LOC105166097	dehydration-responsive element-binding protein 2D-like
	LOC105155699	AP2-like ethylene-responsive transcription factor AIL5
	LOC105177478	B3 domain-containing transcription factor FUS3-like
	LOC105163495	B3 domain-containing transcription factor ABI3
	LOC105162419	ethylene-responsive transcription factor WRI1
	LOC105172214	ethylene-responsive transcription factor WRI1-like
	LOC105160869	ethylene-responsive transcription factor ABI4
	LOC105176383	AP2-like ethylene-responsive transcription factor AIL6
	LOC105168482	ethylene-responsive transcription factor ERF113
	LOC105175841	AP2-like ethylene-responsive transcription factor
	LOC105175564	ethylene-responsive transcription factor 4-like
	LOC105161950	ethylene-responsive transcription factor ERF017-like
	LOC105160523	dehydration-responsive element-binding protein 1D
	LOC105163098	ethylene-responsive transcription factor ERF071
	LOC105171187	ethylene-responsive transcription factor ERF060
	LOC105158942	ethylene-responsive transcription factor RAP2-13
	LOC105161026	ethylene-responsive transcription factor ERF118
	LOC105169546	ethylene-responsive transcription factor ERF109
	LOC105172721	ethylene-responsive transcription factor ERF109-like
	LOC105164148	dehydration-responsive element-binding protein 1D-like
	LOC105160656	ethylene-responsive transcription factor 5-like
	LOC105179326	ethylene-responsive transcription factor ABR1-like
	LOC105171122	ethylene-responsive transcription factor ABR1-like
	LOC105166722	ethylene-responsive transcription factor CRF4-like
	LOC105179134	ethylene-responsive transcription factor ERF061
	LOC105156414	AP2/ERF and B3 domain-containing transcription factor RAV1-like
bHLH	LOC105172690	transcription factor bHLH113-like
	LOC105155973	transcription factor bHLH71
	LOC105173482	transcription factor bHLH130-like
	LOC105173075	transcription factor bHLH13-like
	LOC105172623	transcription factor bHLH143
	LOC105169270	transcription factor bHLH143
	LOC105161457	transcription factor bHLH143
	LOC105171647	transcription factor bHLH92
bZIP	LOC105162512	bZIP transcription factor TGA10
	LOC105155817	bZIP transcription factor 17
	LOC105162695	bZIP transcription factor 17
	LOC105159163	bZIP transcription factor 23
	LOC105169538	bZIP transcription factor 23-like
	LOC105166264	bZIP transcription factor 23
MYB	LOC105157685	transcription factor MYBS3
	LOC105177766	transcription factor MYB44-like
	LOC105179114	transcription factor MYB6-like
	LOC105156372	transcription factor MYB62
	LOC105160804	transcription factor MYB62
WRKY	LOC105167624	probable WRKY transcription factor 3
	LOC105164924	probable WRKY transcription factor 17
	LOC105167998	probable WRKY transcription factor 31
	LOC105163923	probable WRKY transcription factor 25
	LOC105174673	probable WRKY transcription factor 70
	LOC105179124	probable WRKY transcription factor 53
	LOC105173493	probable WRKY transcription factor 41
	LOC105175324	WRKY transcription factor 6
	LOC105160383	probable WRKY transcription factor 30
	LOC105155799	probable WRKY transcription factor 30
	LOC105165167	probable WRKY transcription factor 40
	LOC105161503	probable WRKY transcription factor 33
	LOC110012813	probable WRKY transcription factor 33
	LOC105158610	probable WRKY transcription factor 53

FIGURE 6
Differential expression of transcription factors (TFs) related to lipid biosynthesis. **(A)** The top 15 differentially expressed TF families. **(B)** Expression changes of the major TF genes with FPKM>1000 at either time point of seed development. The blue and yellow denote down-regulation and up-regulation, respectively.

current study were the main FAs in developing sesame seed, and their critical period for rapid accumulation was also in the middle and later period of storage. Oleic and linoleic acids are unsaturated FAs and two of the most important components of sesame seed oil, and our findings also show that oleic and linoleic acids are the most abundant at 21d and 33d which determine the quality of the seed oil. The observed alterations in FAs content indicate that certain pivotal genes underwent regulation of expression during sesame seed development. Using the LC-MS method, we found that the lipids DAG and TAG displayed a similar pattern as the main FA accumulation. Remarkably, the accumulation of total PAs was rapidly increasing aftert 21 days, and their proportions remained relatively steady at the late stage which was significantly higher than all other lipids (Figure 2). In TAG assembly pathways, the main molecular species of PA were the immediate precursors to DAG, and then DGAT catalyzes the conjugation of a fatty acyl-CoA to DAG, to form TAG. Therefore, the rapid accumulation of DAG and TAG at the late maturation stage of sesame seed development requires

substantial amounts of PAs, which suggests PAs have an important role during the accumulation of oil in sesame. The findings suggest that the significant alterations in lipids accumulation are controlled by specific pathways, and the DEGs identified in this study hold potential for identifying regulators of lipid biosynthesis.

FA biosynthesis and TAG assembly

Although there has been extensive research on the biochemistry of lipid biosynthesis in a wide range of plants, the biosynthesis, transport, and storage of lipids constitute complex processes associated with lipid accumulation and its related pathways are not completely understood (Bates et al., 2013; Chen et al., 2015a; Bates, 2016; He and Ding, 2020). In sesame, it is very evident that our understanding of lipid biosynthesis and accumulation, and particularly its regulation, is incomplete following the discovery of new genes involved in oil production and quality (Wei et al., 2015). In this study, several DEGs associated with FA biosynthesis and TAG assembly pathways were identified among developmental stages, including *ACCase*, *MCAT*, *FAS*, *FAD2*, *FAD3*, *GPAT*, and *DGAT*, which was similar to the findings from a previous study on sesame (Wang et al., 2014). There was, however, a significant difference between these DEGs in terms of their expression patterns during seed development (Table S6). For example, the ACCase enzyme is responsible for the first committed step of the FA *de novo* synthesis pathway and overexpression of ACCase genes would lead to an increase in the oil content of plants (Klaus et al., 2004; Chen et al., 2019; Chaturvedi et al., 2021). In addition, during the last and highly important step in TAG biosynthesis, DGAT plays a crucial role in oil accumulation (Weselake et al., 2008; Banas et al., 2013). In a previous study, the DGAT gene was found to be highly expressed in seed and its overexpression in *Arabidopsis* seed increased TAG production (Chen et al., 2016). It is worth noting that several *ACCase* genes showed the highest expression level after 21 d of sesame seed development, whereas two *DGAT* genes were highly expressed at 33 d (Table S6). Additionally, FAD2 synthesizes C18:2 from C18:1 and FAD3 converts C18:2 into C18:3 in the ER (Pham et al., 2012; Dar et al., 2017). Previous studies have shown that the knockout of *FAD2* in soybean (Do et al., 2019) and cotton (Chen et al., 2021) and silencing of *FAD2* in flax (Chen et al., 2015b), oilseed rape (Peng et al., 2010), and other plants can dramatically increase C18:1 content. The transcripts of *FAD2* and *FAD3* genes were identified to be significantly decreased at 21 or 33 d, suggesting that 18:2 and C18:3 accumulation may be dependent on the regulation of *FAD2* and *FAD3* expression levels during mid- and late-stage seed development in sesame. These results indicate that genes involved in FA biosynthesis and TAG assembly were important for controlling sesame seed oil content and FA composition, which should be investigated in future research.

Carbohydrate metabolism

Carbohydrate metabolism pathways, including carbon metabolism, glycolysis/gluconeogenesis, pentose phosphate pathway, and citrate cycle pathway, could provide carbon sources, ATP, and NADPH for lipid biosynthesis, thus indicating that lipid biosynthesis is highly dependent on carbohydrate metabolism (Perry and Harwood, 1994; Hutchings et al., 2005; Furumoto et al., 2011). It was previously demonstrated that 14C-labeled pyruvate can be converted into FAs in isolated plastids from *B. napus* embryos, which suggested that FA biosynthesis could be carried out directly with pyruvate (Eastmond and Rawsthorne, 2000). Several genes encoding important enzymes for the regulation of carbohydrate metabolism have been found to regulate seed oil content, especially *G3PDH*, *PEPCase*, and *PDHC* genes (Yui et al., 2003; Vigeolas et al., 2007; Xu et al., 2016) For instance, *G3PDH* overexpression in rapeseed increased the oil content of transgenic rapeseed to 40% (Vigeolas et al., 2007), while silencing *PEPCase* in cotton by RNAi decreased oil content (Xu et al., 2016), compared to wild type plants. Interestingly, the transcriptomic results in our present work suggest that many DEGs are involved in carbohydrate metabolism pathways, especially *G3PDH*, *PDHC*, *PEPCase*, and *PK* genes, which were found to be significantly differentially expressed during sesame seed development (Table S7). Clarifying the regulation of carbohydrate metabolism might contribute more carbon sources to lipid biosynthesis and provide an important reference to increasing the oil content in oilseed crops through breeding. Therefore, our findings reveal that these putative genes may play important roles in the process of lipid biosynthesis and accumulation during sesame seed development. More studies are needed to verify the gene functions and explore their associated regulatory mechanisms.

TFs involved in the regulation of lipid biosynthesis and accumulation

Plant seeds accumulate lipids through a complex and well-organized biological process involving carbohydrate metabolism, FA biosynthesis, and TAG assembly pathways, and the up-regulation of multiple related genes is likely required for greater lipid accumulation. Researchers have discovered that several important TFs are involved in the regulation of lipid biosynthesis and accumulation in recent years (Tan et al., 2011; Liu et al., 2014; Chen et al., 2020). Changing the expression level of *BnGRF2* (Liu et al., 2012), *WRI1* (Chen et al., 2020), *MYB73* (Liu et al., 2014), *GmbZIP123* (Song et al., 2013), and *DOF1/4* (Wang et al., 2007) in *Arabidopsis* and *BnLEC1* (Tan et al., 2011), NFYA (Lu et al., 2016), and *GmWRI1s* (Chen et al., 2020; Guo et al., 2020) in oilseed crops can significantly improve the oil content in plant seed. Through genetic engineering, oil synthesis-related TFs may thus represent promising targets for enhancing oil yields in oilseed crops. In this study, a systematic search for sesame TFs involved in lipid accumulation was conducted by identifying genes expressed more strongly during rapid lipid synthesis. Our work showed that a total of 33 TF families were involved in lipid accumulation and five TF family members showed high expression levels (FPKM > 1000 at least at one time point) during seed development in sesame, including AP2/ERF, bHLH, bZIP, MYB, and 13 WRKY TF family. Many studies have demonstrated that *WRI1* encodes an AP2/ERF TF, which regulates

carbon partitioning into oils and proteins and controls the expression of many genes essential for FA synthesis (Sakuma et al., 2002; Baud et al., 2007; Kong and Ma, 2018; Kong and Ma, 2018; Lu et al., 2021). Remarkably, there were several AP2/ERF TF members, especially WRI1 and WRI1-like genes in sesame that showed higher levels of expression at 21 or 33 d, with expression patterns similar to many genes involved in lipid biosynthesis, suggesting their important position in the process of sesame seed oil accumulation. Additionally, in addition to AP2/ERFTF expression, we proposed that the high transcript levels of other TFs suggest these have a role in oil accumulation in sesame seed as well.

Conclusions

In the present work, we describe lipidomic and transcriptomic analyses of major lipids and genes related to lipid biosynthesis and its accumulation in the oil-rich crop, *Sesamum indicum* L. A targeted lipidomic analysis showed a variety of molecular species in each class of lipid, and the critical period for the rapid accumulation of most lipids was found to be during the middle and later periods of storage in developing sesame seed. Furthermore, the transcriptomic analysis showed that a large number of DEGs were enriched in lipid biosynthesis and metabolic, carbohydrate, and energy metabolism pathways, and a number of candidate genes (such as *ACCase*, *FAD2*, and *DGAT*) were identified as having an impact on sesame seed oil content and FA composition. In conclusion, these results will improve our understanding of the molecular basis of lipid biosynthesis and accumulation, which provides good theoretical guidance and technical support for the molecular improvement of sesame.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://ngdc.cncb.ac.cn/gsa/browse/CRA008917, Genome Sequence Archive.

Author contributions

YZ, JZ and XZ designed and conceived the study. HG, XC, CG, NL and YP performed the experiments and the data analysis. YZ, JZ and XZ wrote and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Technological Innovation Project of the Shandong Academy of Agricultural Sciences (CXGC2023A45, CXGC2022B01, CXGC2022F18) and the Shandong Science and Technology SME Innovation Capacity Improvement Project (2021TSGC1209). The funding bodies was not involved in data collection, analysis, interpretation, and writing of the manuscript.

Acknowledgments

We are grateful to Shanghai Applied Protein Technology Company (Shanghai, China) for technical support and data analyses and thank International Science Editing for language editing.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1211040/full#supplementary-material

References

Banas, W., Sanchez Garcia, A., Banas, A., and Stymne, S. (2013). Activities of acyl-CoA:diacylglycerol acyltransferase (DGAT) and phospholipid:diacylglycerol acyltransferase (PDAT) in microsomal preparations of developing sunflower and safflower seeds. *Planta* 237 (6), 1627–1636. doi: 10.1007/s00425-013-1870-8

Bansal, S., Kim, H. J., Na, G., Hamilton, M. E., Cahoon, E. B., Lu, C., et al. (2018). Towards the synthetic design of camelina oil enriched in tailored acetyl-triacylglycerols with medium-chain fatty acids. *J. Exp. Bot.* 69 (18), 4395–4402. doi: 10.1093/jxb/ery225

Bates, P. D. (2016). Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis. *Biochim. Biophys. Acta* 1861(9 Pt B)), 1214–1225. doi: 10.1016/j.bbalip.2016.03.021

Bates, P. D., Stymne, S., and Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Curr. Opin. Plant Biol.* 16 (3), 358–364. doi: 10.1016/j.pbi.2013.02.015

Baud, S., Mendoza, M. S., To, A., Harscoet, E., Lepiniec, L., and Dubreucq, B. (2007). WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty

acid metabolism during seed maturation in *Arabidopsis*. *Plant J.* 50 (5), 825–838. doi: 10.1111/j.1365-313X.2007.03092.x

Bourgis, F., Kader, J. C., Barret, P., Renard, M., Robinson, D., Robinson, C., et al. (1999). A plastidial lysophosphatidic acid acyltransferase from oilseed rape. *Plant Physiol.* 120 (3), 913–922. doi: 10.1104/pp.120.3.913

Chaturvedi, S., Gupta, A. K., Bhattacharya, A., Dutta, T., Nain, L., and Khare, S. K. (2021). Overexpression and repression of key rate-limiting enzymes (acetyl CoA carboxylase and HMG reductase) to enhance fatty acid production from rhodotorula mucilaginosa. *J. Basic Microbiol.* 61 (1), 4–14. doi: 10.1002/jobm.202000407

Chen, Y., Fu, M., Li, H., Wang, L., Liu, R., Liu, Z., et al. (2021). High-oleic acid content, nontransgenic allotetraploid cotton (*Gossypium hirsutum* l.) generated by knockout of *GhFAD2* genes with CRISPR/Cas9 system. *Plant Biotechnol. J.* 19 (3), 424–426. doi: 10.1111/pbi.13507

Chen, B., Wang, J., Zhang, G., Liu, J., Manan, S., Hu, H., et al. (2016). Two types of soybean diacylglycerol acyltransferases are differentially involved in triacylglycerol biosynthesis and response to environmental stresses and hormones. *Sci. Rep.* 6, 28541. doi: 10.1038/srep28541

Chen, G., Woodfield, H. K., Pan, X., Harwood, J. L., and Weselake, R. J. (2015a). Acyl-trafficking during plant oil accumulation. *Lipids* 50 (11), 1057–1068. doi: 10.1007/s11745-015-4069-x

Chen, D., Yuan, X., Liang, L., Liu, K., Ye, H., Liu, Z., et al. (2019). Overexpression of *acetyl-CoA carboxylase* increases fatty acid production in the green alga *Chlamydomonas reinhardtii*. *Biotechnol. Lett.* 41 (10), 1133–1145. doi: 10.1007/s10529-019-02715-0

Chen, B., Zhang, G., Li, P., Yang, J., Guo, L., Benning, C., et al. (2020). Multiple GmWRI1s are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (*Glycine max*). *Plant Biotechnol. J.* 18 (1), 155–171. doi: 10.1111/pbi.13183

Chen, Y., Zhou, X. R., Zhang, Z. J., Dribnenki, P., Singh, S., and Green, A. (2015b). Development of high oleic oil crop platform in flax through RNAi-mediated multiple FAD2 gene silencing. *Plant Cell Rep.* 34 (4), 643–653. doi: 10.1007/s00299-015-1737-5

Dar, A. A., Choudhury, A. R., Kancharla, P. K., and Arumugam, N. (2017). The *FAD2* gene in plants: occurrence, regulation, and role. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01789

Ding, L. N., Gu, S. L., Zhu, F. G., Ma, Z. Y., Li, J., Li, M., et al. (2020). Long-chain acyl-CoA synthetase 2 is involved in seed oil production in brassica napus. *BMC Plant Biol.* 20 (1), 21. doi: 10.1186/s12870-020-2240-x

Do, P. T., Nguyen, C. X., Bui, H. T., Tran, L. T. N., Stacey, G., Gillman, J. D., et al. (2019). Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous *GmFAD2-1A* and *GmFAD2-1B* genes to yield a high oleic, low linoleic and alpha-linolenic acid phenotype in soybean. *BMC Plant Biol.* 19 (1), 311. doi: 10.1186/s12870-019-1906-8

Dyer, J. M., Stymne, S., Green, A. G., and Carlsson, A. S. (2008). High-value oils from plants. *Plant J.* 54, 640–655. doi: 10.1111/j.1365-313X.2008.03430.x

Eastmond, P. J., and Rawsthorne, S. (2000). Coordinate changes in carbon partitioning and plastidial metabolism during the development of oilseed rape embryos. *Plant Physiol.* 122 (3), 767–774. doi: 10.1104/pp.122.3.767

Elhiti, M., Yang, C., Chan, A., Durnin, D. C., Belmonte, M. F., Ayele, B. T., et al. (2012). Altered seed oil and glucosinolate levels in transgenic plants overexpressing the *Brassica napus SHOOTMERISTEMLESS* gene. *J. Exp. Bot.* 63 (12), 4447–4461. doi: 10.1093/jxb/ers125

Figueiredo, P. S., Candido, C. J., Jaques, J. A., Nunes, A. A., Caires, A. R., Michels, F. S., et al. (2017). Oxidative stability of sesame and flaxseed oils and their effects on morphometric and biochemical parameters in an animal model. *J. Sci. Food Agric.* 97 (10), 3359–3364. doi: 10.1002/jsfa.8186

Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., et al. (2011). A plastidial sodium-dependent pyruvate transporter. *Nature* 476 (7361), 472–475. doi: 10.1038/nature10250

Guo, W., Chen, L., Chen, H., Yang, H., You, Q., Bao, A., et al. (2020). Overexpression of *GmWRI1b* in soybean stably improves plant architecture and associated yield parameters, and increases total seed oil production under field conditions. *Plant Biotechnol. J.* 18 (8), 1639–1641. doi: 10.1111/pbi.13324

Harwood, J. L. (1996). Recent advances in the biosynthesis of plant fatty acids. *Biochim. Biophys. Acta* 1301 (1-2), 7–56. doi: 10.1016/0005-2760(95)00242-1

Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the *fatty acid desaturase 2* gene family. *Plant Biotechnol. J.* 12 (7), 934–940. doi: 10.1111/pbi.12201

He, M., and Ding, N. Z. (2020). Plant unsaturated fatty acids: multiple roles in stress response. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.562785

Hutchings, D., Rawsthorne, S., and Emes, M. J. (2005). Fatty acid synthesis and the oxidative pentose phosphate pathway in developing embryos of oilseed rape (*Brassica napus* l.). *J. Exp. Bot.* 56 (412), 577–585. doi: 10.1093/jxb/eri046

Kelly, A. A., Shaw, E., Powers, S. J., Kurup, S., and Eastmond, P. J. (2013a). Suppression of the SUGAR-DEPENDENT1 triacylglycerol lipase family during seed development enhances oil yield in oilseed rape (*Brassica napus* l.). *Plant Biotechnol. J.* 11 (3), 355–361. doi: 10.1111/pbi.12021

Kelly, A. A., van Erp, H., Quettier, A. L., Shaw, E., Menard, G., Kurup, S., et al. (2013b). The sugar-dependent1 lipase limits triacylglycerol accumulation in vegetative tissues of *Arabidopsis*. *Plant Physiol.* 162 (3), 1282–1289. doi: 10.1104/pp.113.219840

Klaus, D., Ohlrogge, J. B., Neuhaus, H. E., and Dormann, P. (2004). Increased fatty acid production in potato by engineering of acetyl-CoA carboxylase. *Planta* 219 (3), 389–396. doi: 10.1007/s00425-004-1236-3

Knutzon, D. S., Thompson, G. A., Radke, S. E., Johnson, W. B., Knauf, V. C., and Kridl, J. C. (1992). Modification of brassica seed oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene. *Proc. Natl. Acad. Sci. U.S.A.* 89 (7), 2624–2628. doi: 10.1073/pnas.89.7.2624

Kong, Q., and Ma, W. (2018). WRINKLED1 transcription factor: how much do we know about its regulatory mechanism? *Plant Sci.* 272, 153–156. doi: 10.1016/j.plantsci.2018.04.013

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-lipid metabolism. *Arabidopsis book* 11, e0161. doi: 10.1199/tab.0161

Lin, P., Wang, K., Zhou, C., Xie, Y., Yao, X., and Yin, H. (2018). Seed transcriptomics analysis in *Camellia oleifera* uncovers genes associated with oil content and fatty acid composition. *Int. J. Mol. Sci.* 19 (1), 118. doi: 10.3390/ijms19010118

Liu, J., Hua, W., Yang, H. L., Zhan, G. M., Li, R. J., Deng, L. B., et al. (2012). The *BnGRF2* gene (GRF2-like gene from *Brassica napus*) enhances seed oil production through regulating cell number and plant photosynthesis. *J. Exp. Bot.* 63 (10), 3727–3740. doi: 10.1093/jxb/ers066

Liu, Y. F., Li, Q. T., Lu, X., Song, Q. X., Lam, S. M., Zhang, W. K., et al. (2014). Soybean GmMYB73 promotes lipid accumulation in transgenic plants. *BMC Plant Biol.* 14, 73. doi: 10.1186/1471-2229-14-73

Lu, X., Li, Q. T., Xiong, Q., Li, W., Bi, Y. D., Lai, Y. C., et al. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *Plant J.* 86 (6), 530–544. doi: 10.1111/tpj.13181

Lu, C., Napier, J. A., Clemente, T. E., and Cahoon, E. B. (2011). New frontiers in oilseed biotechnology: meeting the global demand for vegetable oils for food, feed, biofuel, and industrial applications. *Curr. Opin. Biotechnol.* 22 (2), 252–259. doi: 10.1016/j.copbio.2010.11.006

Lu, L., Wei, W., Li, Q. T., Bian, X. H., Lu, X., Hu, Y., et al. (2021). A transcriptional regulatory module controls lipid accumulation in soybean. *New Phytol.* 231 (2), 661–678. doi: 10.1111/nph.17401

Lummiss, J. A., Oliveira, K. C., Pranckevicius, A. M., Santos, A. G., dos Santos, E. N., and Fogg, D. E. (2012). Chemical plants: high-value molecules from essential oils. *J. Am. Chem. Soc.134* 46), 18889–18891. doi: 10.1021/ja310054d

Melo, D., Alvarez-Orti, M., Nunes, M. A., Costa, A. S. G., Machado, S., and Alves, R. C. (2021). Whole or defatted sesame seeds (*Sesamum indicum* l.)? the effect of cold pressing on oil and cake quality. *Foods* 10 (9), 2108. doi: 10.3390/foods10092108

Peng, Q., Hu, Y., Wei, R., Zhang, Y., Guan, C., Ruan, Y., et al. (2010). Simultaneous silencing of *FAD2* and *FAE1* genes affects both oleic acid and erucic acid contents in *Brassica napus* seeds. *Plant Cell Rep.* 29, 317–325. doi: 10.3390/foods10092108

Perry, H. J., and Harwood, J. L. (1994). The carbon flux to triacylglycerol in maturing oilseed rape embryos. *Biochem. Soc Trans.* 22 (2), 203S. doi: 10.1042/bst022203s

Pham, A. T., Shannon, J. G., and Bilyeu, K. D. (2012). Combinations of mutant *FAD2* and *FAD3* genes to produce high oleic acid and low linolenic acid soybean oil. *TAG. Theor. Appl. Genet. Theor. Appl. Genet.* 125 (3), 503–515. doi: 10.1007/s00122-012-1849-z

Qureshi, M., Langham, D. R., Lucas, S. J., Uzun, B., and Yol, E. (2022). Breeding history for shattering trait in sesame: classic to genomic approach. *Mol. Biol. Rep.* 49 (7), 7185–7194. doi: 10.1007/s11033-022-07636-2

Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). DNA-Binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* 290 (3), 998–1009. doi: 10.1006/bbrc.2001.6299

Song, Q. X., Li, Q. T., Liu, Y. F., Zhang, F. X., Ma, B., Zhang, W. K., et al. (2013). Soybean *GmbZIP123* gene enhances lipid content in the seeds of transgenic *Arabidopsis* plants. *J. Exp. Bot.* 64 (14), 4329–4341. doi: 10.1093/jxb/ert238

Taguchi, R. , and Ishikawa, M. (2010). Precise and global identification of phospholipid molecular species by an orbitrap mass spectrometer and automated search engine lipid search. *J. Chromatogr. A.* 1217(25), 4229–4239. doi: 10.1016/j.chroma.2010.04.034

Tan, H., Yang, X., Zhang, F., Zheng, X., Qu, C., Mu, J., et al. (2011). Enhanced seed oil production in canola by conditional expression of *Brassica napus LEAFY COTYLEDON1* and *LEC1-LIKE* in developing seeds. *Plant Physiol.* 156 (3), 1577–1588. doi: 10.1104/pp.111.175000

Vigeolas, H., Waldeck, P., Zank, T., and Geigenberger, P. (2007). Increasing seed oil content in oil-seed rape (*Brassica napus* l.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. *Plant Biotechnol. J.* 5 (3), 431–441. doi: 10.1111/j.1467-7652.2007.00252.x

Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., et al. (2014). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 15, R39. doi: 10.1186/gb-2014-15-2-r39

Wang, H. W., Zhang, B., Hao, Y. J., Huang, J., Tian, A. G., Liao, Y., et al. (2007). The soybean dof-type transcription factor genes, *GmDof4* and *GmDof11*, enhance lipid

content in the seeds of transgenic *Arabidopsis* plants. *Plant J.* 52 (4), 716–729. doi: 10.1111/j.1365-313X.2007.03268.x

Want, E. J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S., Shockcor, J., et al. (2010). Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* 5 (6), 1005–1018. doi: 10.1038/nprot.2010.50

Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., et al. (2015). Genetic discovery for oil production and quality in sesame. *Nat. Commun.* 6, 8609. doi: 10.1038/ncomms9609

Weselake, R. J., Shah, S., Tang, M., Quant, P. A., Snyder, C. L., Furukawa-Stoffer, T. L., et al. (2008). Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. *J. Exp. Bot.* 59 (13), 3543–3549. doi: 10.1093/jxb/ern206

Woodfield, H. K., Cazenave-Gassiot, A., Haslam, R. P., Guschina, I. A., Wenk, M. R., and Harwood, J. L. (2018). Using lipidomics to reveal details of lipid accumulation in developing seeds from oilseed rape (*Brassica napus* l.). *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1863(3), 339–348. doi: 10.1016/j.bbalip.2017.12.010

Woodfield, H. K., Sturtevant, D., Borisjuk, L., Munz, E., Guschina, I. A., Chapman, K., et al. (2017). Spatial and temporal mapping of key lipid species in *Brassica napus* seeds. *Plant Physiol.* 173 (4), 1998–2009. doi: 10.1104/pp.16.01705

Xu, Z., Li, J., Guo, X., Jin, S., and Zhang, X. (2016). Metabolic engineering of cottonseed oil biosynthesis pathway *via* RNA interference. *Sci. Rep.* 6, 33342. doi: 10.1038/srep33342

Yang, Y., Kong, Q., Lim, A. R. Q., Lu, S., Zhao, H., Guo, L., et al. (2022). Transcriptional regulation of oil biosynthesis in seed plants: current understanding, applications, and perspectives. *Plant Commun.* 3 (5), 100328. doi: 10.1016/j.xplc.2022.100328

Yang, S., Miao, L., He, J., Zhang, K., Li, Y., and Gai, J. (2019). Dynamic transcriptome changes related to oil accumulation in developing soybean seeds. *Int. J. Mol. Sci.* 20 (9), 2202. doi: 10.3390/ijms20092202

Yui, R., Iketani, S., Mikami, T., and Kubo, T. (2003). Antisense inhibition of mitochondrial pyruvate dehydrogenase E1alpha subunit in anther tapetum causes male sterility. *Plant J.* 34 (1), 57–66. doi: 10.1046/j.1365-313x.2003.01704.x

Zhang, M., Fan, J., Taylor, D. C., and Ohlrogge, J. B. (2009). DGAT1 and PDAT1 acyltransferases have overlapping functions in *Arabidopsis* triacylglycerol biosynthesis and are essential for normal pollen and seed development. *Plant Cell* 21 (12), 3885–3901. doi: 10.1105/tpc.109.071795

Zhang, Y., Li, D., Zhou, R., Wang, X., Dossa, K., Wang, L., et al. (2019). Transcriptome and metabolome analyses of two contrasting sesame genotypes reveal the crucial biological pathways involved in rapid adaptive response to salt stress. *BMC Plant Biol.* 19 (1), 66. doi: 10.1186/s12870-019-1665-6

Zhang, H., Miao, H., Wang, L., Qu, L., Liu, H., Wang, Q., et al. (2013). Genome sequencing of the important oilseed crop *Sesamum indicum* l. *Genome Biol.* 14 (1), 401. doi: 10.1186/gb-2013-14-1-401

Zhu, Y., Wang, K., Wu, C., Zhao, Y., Yin, X., Zhang, B., et al. (2019). Effect of ethylene on cell wall and lipid metabolism during alleviation of postharvest chilling injury in peach. *Cells* 8 (12), 1612. doi: 10.3390/cells8121612

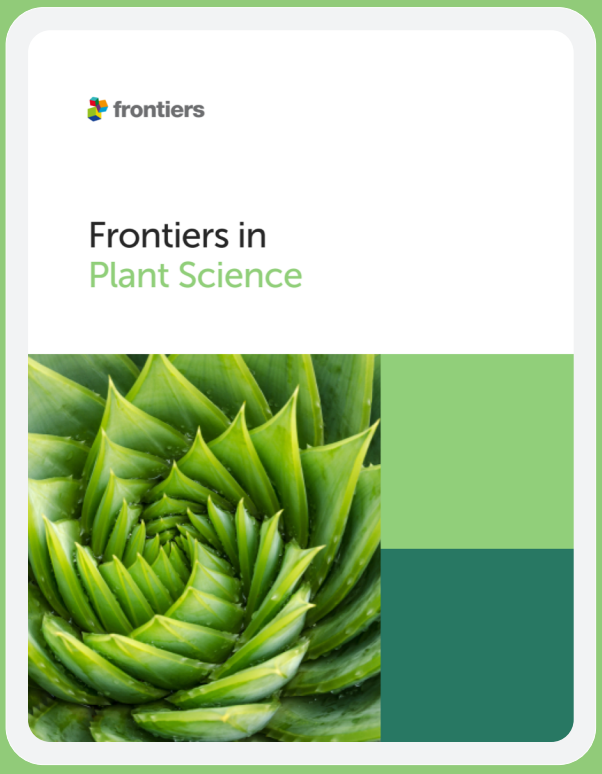
Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact